Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

BIOLOGIA CELLULARE E MOLECOLARE

Ciclo XXVI

Settore Scientifico Disciplinare: 05/E2
Settore Concorsuale di afferenza: BIO/11

TITOLO TESI

Polymerizing activity and regulation of *group B Streptococcus* pilus 2a sortase C1

Presentata da: **Francesca Zerbini**

Coordinatore Dottorato

Chiar.mo Prof.
Vincenzo Scarlato

Relatore

Chiar.mo Prof.
Vincenzo Scarlato

Co-relatore

Dott.ssa
Roberta Cozzi

Esame finale anno 2014

Oggetto del mio progetto di dottorato, presentato in questo lavoro di tesi, è stato lo studio del meccanismo di assemblaggio del pilo 2a di *Streptococcus agalactiae* (Streptococco di gruppo B, GBS), focalizzandomi soprattutto sull'attività e la regolazione della sortasi C1.

Il lavoro svolto durante lo svolgimento di questo progetto di dottorato è stato oggetto delle seguenti pubblicazioni:

- Cozzi R*, **Zerbini F*,** Assfalg M, D'Onofrio M, Biagini M, Martinelli M, Nuccitelli A, Norais N, Telford JL, Maione D, Rinaudo CD. *Group B Streptococcus pilus sortase regulation: a single mutation in the lid region induces pilin protein polymerization in vitro.* FASEB J. 2013 Aug;27(8):3144-54. Epub 2013 Apr 30.

  * These authors contributed equally to this paper

- Cozzi R, Nuccitelli A, D'Onofrio M, Necchi F, Rosini R, **Zerbini F**, Biagini M, Norais N, Beier C, Telford JL, Grandi G, Assfalg M, Zacharias M, Maione D, Rinaudo CD. *New insights into the role of the glutamic acid of the E-box motif in group B Streptococcus pilus 2a assembly.* FASEB J. 2012 May;26(5): 2008-18.

Table of Contents

Abstract

Group B *Streptococcus* [GBS; *Streptococcus agalactiae*] is the leading cause of life-threatening diseases in newborn and is also becoming a common cause of invasive diseases in non-pregnant, elderly and immune-compromised adults. Pili, long filamentous fibers protruding from the bacterial surface, have been discovered in GBS, as important virulence factors and vaccine candidates. Gram-positive bacteria build pili on their cell surface via a class C sortase-catalyzed transpeptidation mechanism from pilin protein substrates. Despite the availability of several crystal structures, pilus-related C sortases remain poorly characterized to date and their mechanisms of transpeptidation and regulation need to be further investigated. The available three-dimensional structures of these enzymes reveal a typical sortase fold except for the presence of a unique feature represented by an N-terminal highly flexible loop, known as the "lid". This region interacts with the residues composing the catalytic triad and covers the active site, thus maintaining the enzyme in an auto-inhibited state and preventing the accessibility to the substrate. It is believed that enzyme activation may occur only after lid displacement from the catalytic domain. In this work we provide the first direct evidence of the regulatory role of the lid, demonstrating that it is possible to obtain *in vitro* an efficient polymerization of pilin subunits using an active C sortase lid mutant carrying a single residue mutation in the lid region. Moreover, biochemical analyses of this recombinant mutant reveal that the lid confers thermodynamic and proteolytic stability to the enzyme. A further characterization of this sortase active mutant showed promiscuity in the substrate recognition, as it is able to polymerize different LPXTG-proteins *in vitro*.

Chapter 1. Introduction

1.1 Structure and assembly of Gram-positive bacterial pili

Pili, or fimbriae, are protein polymers that form long, filamentous structures that extend from bacterial cells, mediating adhesion to host cells, colonization, biofilm formation and sometimes motility (Proft and Baker 2009). Pili of pathogenic organisms are also highly immunogenic, making them attractive for vaccine development. The best-known and characterized pili are those of Gram-negative bacteria: the Type I and Type P pili of Escherichia coli, and the Type IV pili of Neisseria species (Waksman and Hultgren 2009), which form rod-like bundles of non-covalently assembled subunits. In contrast, the pili on Gram-positive bacteria are fundamentally different. They are long (2–5 μm) but extremely thin (about 3 nm), assembled by enzymes called sortases, and they are rare examples of covalent polymers (Fig.1). Despite many years of study of Gram-positive bacterial pili, they remained largely unnoticed until very recently (Kang and Baker). Their characterization followed the discovery of sortases and the availability of genome sequences (Kang and Baker). The assay generally used to determine the expression of pilus structures is to subject the total bacterial cell lysate to boiling in SDS followed by SDS-PAGE. A protein that is part of a pilus will appear as a high molecular weight (HMW) ladder in immunoblot. Another method used to detect pili is visualization by negative staining, or, more specifically, by immunogold electron microscopy (IEM), which can reveal the localization of a protein within the pilus structure. Gram-positive pili are composed of multiple copies of a single pilin shaft, other than additional proteins

associated with the shaft, but not required for the integrity or synthesis of the pilus (Ton-That and Schneewind 2003).

Early data from studies of oral Gram-positive pathogens indicated that such structures are involved in adhesion and attachment to host cell, in the interaction with components of the extracellular matrix (ECM), and in biofilm formation (Konto-Ghiorghi, Mairey et al. 2009). Additionally, a recent study provided evidence for an active role of *S. agalactiae* pilus proteins in the newly discovered paracellular translocation through the epithelial barrier, during host colonization (Soriani, Santi et al. 2006). Gram-positive pili could be considered important virulence factors for several diseases (Nallapareddy, Singh et al. 2006), in particular infections of the urinary, genital and gastrointestinal tracts. Furthermore, in pathogenic *Streptococcus* species pili are reported to be also promising vaccine candidates (Maione, Margarit et al. 2005).

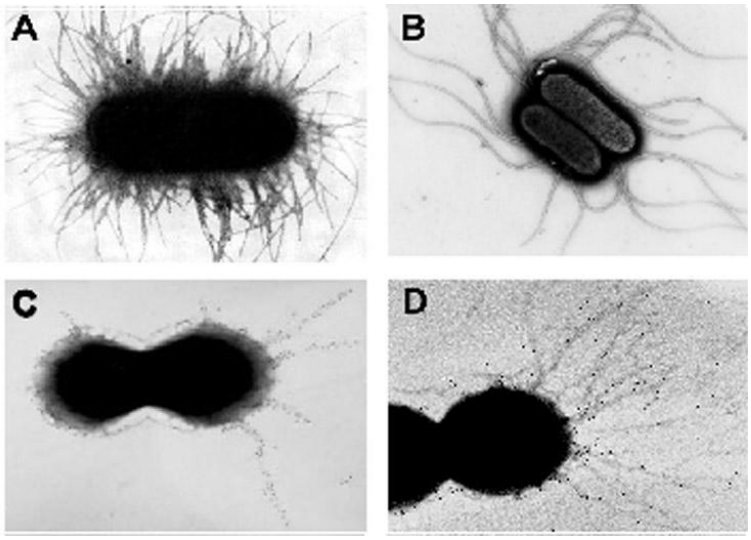


Figure 1. Different examples of pilus-like structures in Gram-negative and Gram-positive bacteria. Electron micrographs of fimbriae in Gram-negative organisms : *E. coli* (A) and *Salmonella enterica* (B). Electron microscopy of two different types of pili in

Gram-positive bacteria: fibrils in *Streptococcus salivarius* (C) and pili in *Streptococcus agalactiae* (D) stained by immunogold labeling (Telford, Barocchi et al. 2006).

Thon-That and Schneewind, working on Corynebacterium diphteriae as a model, have provided the first insights into the assembly mechanism of Gram-positive pili (Ton-That and Schneewind 2003). The three pilus proteins together with genes coding for sortases, that are required for pilus assembly, are encoded in a small gene cluster within pathogenicity islands which are known as Pilus Islands (PIs). The genes are transcribed in the same direction, indicating that they are part of an operon. The three pilus components are characterized by the presence of an N-terminal signal peptide together with a C-terminal cell-wall sorting signal (CWSS), that is found in many surface proteins and is required for the attachment to the peptidoglycan of the cell wall. The CWSS comprises the amino acid sequence "LPXTG" (where X denotes any amino acid) or a variation of this motif (such as VV/PXTG in the case of the main pilin subunit of Group A streptococcus, Cpa), followed by a hydrophobic membrane-spanning domain and a positively charged tail. This motif is targeted by sortase enzymes, which are membrane-bound transpeptidases catalysing the covalent linkage of LPXTG motif proteins to the peptidoglycan. During pilus formation, specific pilus-related sortases catalyse the covalent attachment of the pilin subunits to each other and to the peptidoglycan cell wall (Telford, Barocchi et al. 2006). Immunogold electron microscopy (IEM) using antisera specific for the three pilus components revealed that pilus shaft is a polymer of one pilin called backbone protein (BP), and the other two components are ancillary proteins (AP). Backbone protein specific

antisera stain the whole length of the pilus structure (Telford, Barocchi et al. 2006).

The first insights into the assembly mechanism of Gram-positive pili were provided by a study performed on *Corynebacterium diphteriae* (Ton-That, Marraffini et al. 2004).

Initially, the three pilus components containing an LPXTG motif are secreted in a Sec-dependent way (Telford, Barocchi et al. 2006). Each component remains anchored to the cell membrane, owing to the presence of the C-terminal transmembrane domain.

The second step involves a sortase-dependent reaction in which the membrane-anchored proteins are cleaved at the LPXTG motif, between the threonine (T) and glycine (G) residue. This reaction leads to the formation of acyl-enzyme intermediates in which a covalent thioester bond is formed between the thiol group of the cysteine residue located in the catalytic pocket of the sortase and the carboxyl group of the threonine residue in the LPXTG motif of the pilin protein (Telford, Barocchi et al. 2006). Because sortases are membrane-associated enzymes, the acyl-enzyme derivatives that are formed are retained on the external side of the membrane (Fig. 2).

The following steps of the assembly process involve the oligomerization of the pilus protein subunits and the anchoring of the oligomerized structure to the cell wall.

These steps require the nucleophilic attack of the thioester bond in the acyl-enzyme intermediate. During pilus polymerization the nucleophile is provided by the ε-amino group of a specific lysine (K) residue within the "pilin motif",

WXXXVXVYPKN (where X denotes any amino acid), which has been found in most of the pilin subunits that have been characterized (Ton-That and Schneewind 2003). The nucleophilic attack results in cleavage of the thioester bond and concomitant formation of an amide bond between the carbonyl-group carbon of the threonine residue of the pilin subunit (present in the catalytic pocket of the sortase) and the lysine side-chain (ε-amino group) of the pilin motif of the neighboring pilin subunit. This leads to the formation of a membrane-associated covalently linked dimer with a pilin motif that can interact with other sortase-associated pilin subunits, forming an elongated pilus fiber. Ton-That and co-workers have shown that replacing the lysine residue in the pilin motif with an alanine residue abolishes the polymerization process, highlighting the importance of this conserved sequence in pilus formation (Telford, Barocchi et al. 2006).

According to this model, pilus growth occurs by subunit addition at the base of the pilus (Fig. 2), and the length of the pilus depends on the relative abundance of the pilus subunits that are coupled to the membrane-associated sortases (Telford, Barocchi et al. 2006). Finally, the association of the membrane-proximal pilus subunit with the cell wall occurs when the thioester bond between the subunit and the sortase is subject to nucleophilic attack by the amino group in the cross-bridge of the peptidoglycan precursor lipid II (Ton-That and Schneewind 2004), and this leads to the formation of an amide bond between the basal subunit and the bacterial cell wall.

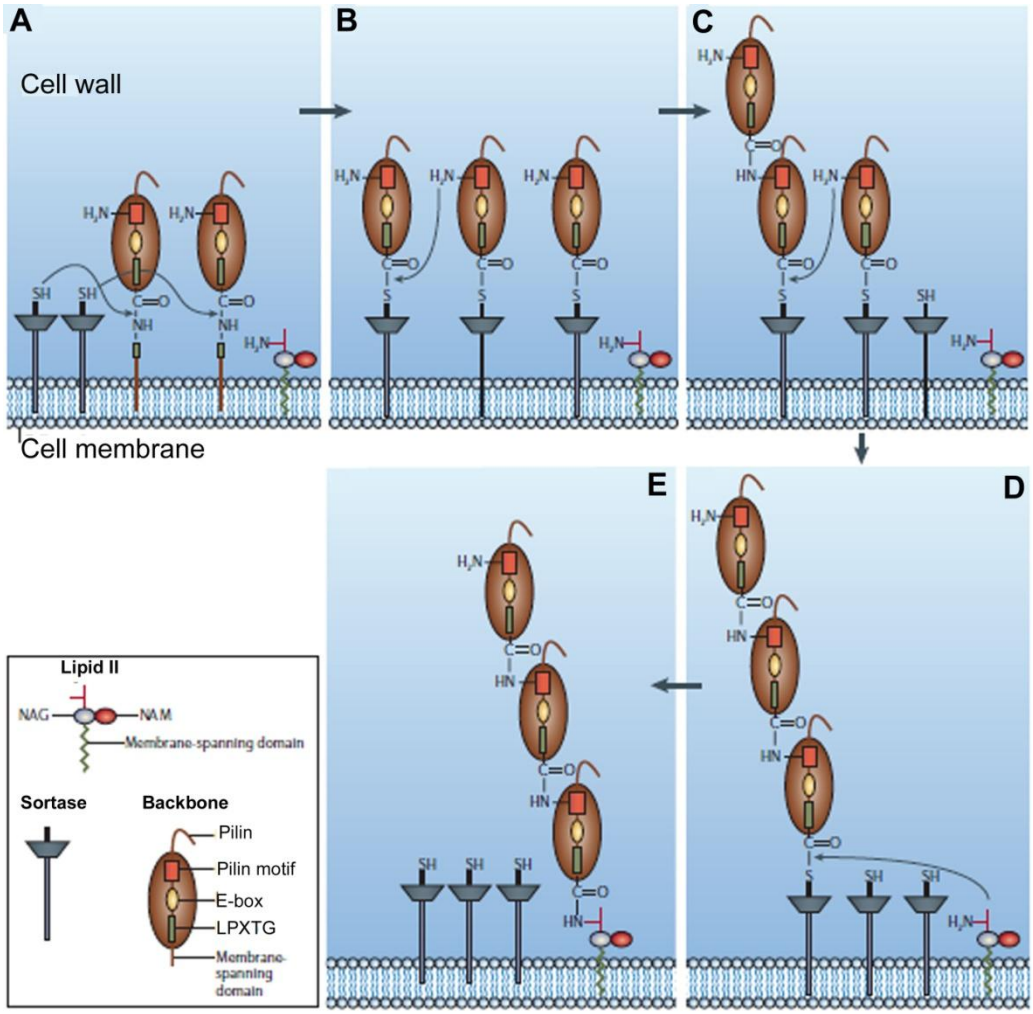


Figure 2. General model for pilus assembly in Gram-positive bacteria (Telford, Barocchi et al. 2006). (A) In the first step, proteins that contain the amino-acid motif LPXTG are targeted to the cell membrane by Sec-dependent secretion (not shown). This is followed by a sortase-mediated reaction (indicated by the arrows) in which the LPXTG motif is cleaved between the threonine (T) and glycine (G) residues. **(B)** The reaction leads to the formation of an acyl-enzyme intermediate in which a covalent thioester bond is formed between the thiol group of a cysteine residue in the sortase and the carboxyl group of the pilin threonine residue. **(C)** Oligomerization occurs after the nucleophilic attack provided by the e-amino group of the lysine residue in the pilin motif on the cysteine residue of the sortase. **(D)**The thioester bond between the pilin subunit and the sortase is targeted by the amino group of the pentapeptide of lipid II, the precursor of peptidoglycan. **(E)** This leads to the formation of an elongated pilus covalently linked to the cell wall peptidoglycan. NAG, *N*-acetyl glucosamine; NAM, N-acetyl muramic acid (Telford, Barocchi et al. 2006).

It has been suggested that another conserved aminoacidic sequence in the backbone subunit, called the "E-box" (consensus YxLxETxAPxGY), due to a highly conserved glutamic acid residue, plays a role in pilus polymerization (Telford, Barocchi et al. 2006).

Despite low sequence similarities, the pilin subunits of gram-positive bacteria show very similar tridimensional structure comprising immunoglobulin G (IgG)-like domains of shared evolutionary origin. Each pilin subunit is stabilized by intramolecular isopeptide bonds, and all contain sequence elements and/or residues that are essential for pilus assembly, and which are conserved among pilin subunits in different bacteria (Rosini, Rinaudo et al. 2006). Such motives include the above mentioned pilin motif, the cell-wall sorting signal (CWSS) containing the sortase recognition site LPxTG motif, and the E-box motif as assigned for the first time in the major pilin subunit SpaA of *Corynebacterium diphtheriae* (Ton-That, Marraffini et al. 2004) and subsequently in other bacterial pilins (Mandlik, Swierczynski et al. 2008). The E-box contains a conserved glutamic acid residue, which in *C. diphtheria* SpaA (Glu-446) has been demonstrated to be essential for the incorporation of the minor pilins SpaB and SpaC (Ton-That, Marraffini et al. 2004). Intriguingly, in SpaA, this glutamate is the catalytic residue that mediates the formation of the Lys-363–Asn-462 intramolecular isopeptide bond (Kang, Paterson et al. 2009), similar to the role assigned to Glu-258 in *GAS* Spy0128, in which this residue was shown to be essential for the corresponding intramolecular reaction to occur (Kang, Coulibaly et al. 2007). Moreover, several X-ray crystal structures of backbone pilins have shown that the E-box domain is involved in the formation of such isopeptide

bonds and that these linkages confer higher stability to the monomeric subunit (Hendrickx, Budzik et al.; Kang and Baker; Kang, Coulibaly et al. 2007). Recently, the X-ray crystal structure of the shaft-forming backbone protein of *S. agalactiae* pilus 2a (BP-2a) was solved (Nuccitelli, Cozzi et al.). The 3-D structure revealed an IgG-like fold domains organization, comprising 4 structural units, designated D1–D4. The domains D2, D3, and D4 are each stabilized by an intramolecular Lys-Asn isopeptide bond, located in a largely hydrophobic pocket, comprising several aromatic residues, including a bond-catalyzing aspartyl or glutamyl residue (Fig.3) (Nuccitelli, Cozzi et al.). However, the role of intramolecular isopeptide bonds and of the E-box motif in pilus assembly still needs to be clarified (Cozzi, Nuccitelli et al. 2012).

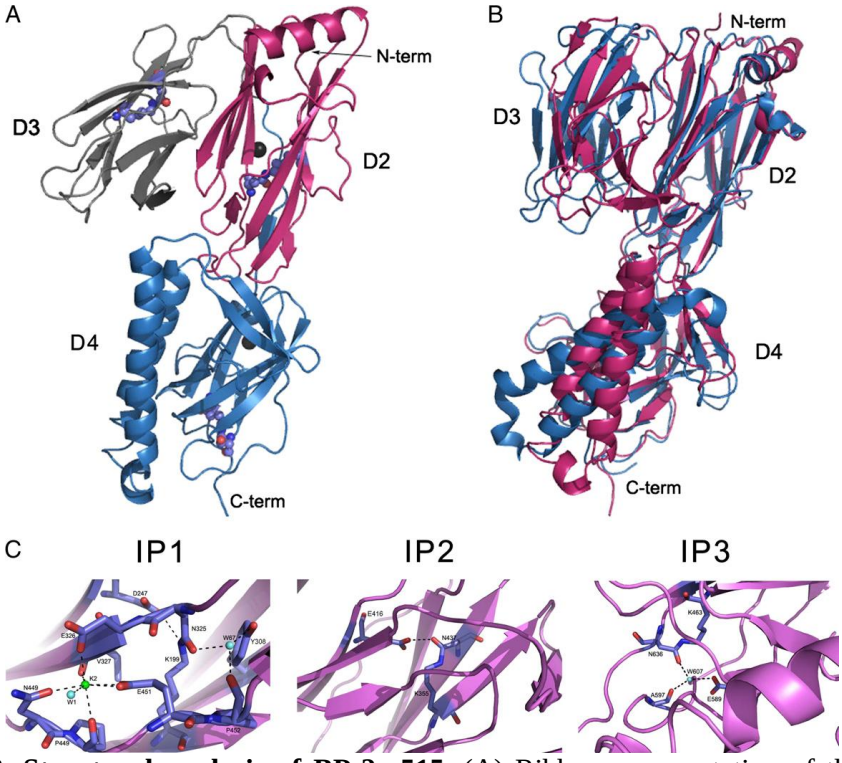


Figure 3. Structural analysis of BP-2a-515. (A) Ribbon representation of the crystal structure of BP-2a-515 (residues 190–640), illustrating the N and C termini, domains D2, D3, and D4, two potassium ions (blue spheres), and the three intramolecular isopeptide bonds (spheres). (B) Superimposition of BP-2a-515 (purple) with RrgB from

Streptococcus pneumoniae (blue), highlighting the structural similarity between the two proteins. (C) Structural details of the D2, D3, and D4 domains in the regions involved in isopeptide bond formation. All images were generated using Pymol Version 1.1r1 (www.pymol.org) (Nuccitelli, Cozzi et al. 2011).

In conclusion, pilus assembly in Gram-positive bacteria seems to occur by a universal mechanism of ordered cross-linking of precursor proteins, whose multiple conserved features are recognized by designated sortase enzymes (Ton-That and Schneewind 2003; Ton-That, Marraffini et al. 2004).

1.2 Pili in Group B Streptococcus

1.2.1 *Streptococcus agalactiae* (Group B streptococcus, GBS)

Streptococcus agalactiae (commonly referred to as Group B *Streptococcus* or GBS) is an encapsulated Gram-positive coccus, catalase negative and facultatively anaerobic. It generally grows in pairs or in long chains of spherical bacteria, less than 2 μm in size (Fig.4A). It displays beta-hemolysis when cultured on blood agar plates and produces zones of hemolysis that are only slightly larger than the colonies themselves (Fig.4B) (Gibbs, Schrag et al. 2004). GBS strains are classified into nine serotypes according to immunogenic characteristics of the capsule polysaccharides (Ia, Ib, II, III, IV, V, VI, VII, VIII and IX). Approximately, 10% of serotypes are non-typeable (Kogan, Uhrin et al. 1996).

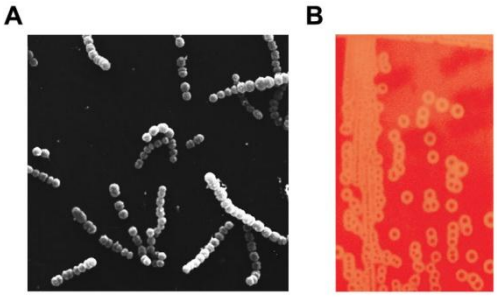


Figure 4. *Streptococcus agalactiae*. (A) Scanning Electron Microscopy (SEM) of *Streptococcus agalactiae*. **(B)** Colonies of *Streptococcus agalactiae* on a blood agar plate. Note the zone of clear haemolysis.

Consistent with other streptococcal species (Mitchell 2003), *Streptococcus agalactiae* is present on the mucosal surfaces of animals and humans. In fact, GBS can usually colonize as a normal commensal the intestinal and vaginal tract but also the pharyngeal mucosa of human adults (Baker 1997) and 20–40% of healthy women carry GBS (Baker 1997; Hansen, Uldbjerg et al. 2004; Yamamoto, Pargade et al. 2006).

Invasive group B streptococcal disease emerged in the 1970s as a leading cause of neonatal morbidity and mortality in the United States (McCracken 1973), and represents the most common etiological agent of invasive bacterial infections (pneumonia, septicaemia and meningitis) in human neonates (Nizet, Gibson et al. 1996; Davies, Adair et al. 2001; Gibbs, Schrag et al. 2004). Most infections and colonization of newborns are due to aspiration of contaminated amniotic and vaginal fluid before or during delivery (Doran and Nizet 2004).

Streptococcus agalactiae is also associated to a number of postpartum sequelae, such as urinary tract infections, amnionitis, endometritis, as well as to wound

infection and mortality or morbidity in immunocompromised adults (Schuchat 1998).

Among them, pili have been recently implicated in mediating attachment to human epithelial cells (Dramsi, Caliot et al. 2006), and in the binding and invasion of brain microvascular endothelial cells (Maisey, Hensler et al. 2007).

1.2.2 **Identification of novel genomic islands coding for pilus-like structures in *Streptococcus agalactiae***

A Reverse Vaccinology approach (De Groot and Rappuoli 2004) has been used to identify protective antigens for inclusion in a vaccine against GBS. Five proteins were found to elicit protection against GBS in a mouse maternal immunization assay (Maione, Margarit et al. 2005). Furthermore, analysis of the eight sequenced genomes of GBS has shown that four of these five protective antigens, GBS80 (TIGR annotation SAG0645), GBS104 (SAG0649), GBS67 (SAG1408) and GBS59 (SAG1407), are located in tandem in two different genomic islands that belongs to the "dispensable genome" of GBS (Tettelin, Masignani et al. 2005) (Lauer, Rinaudo et al. 2005; Rosini, Rinaudo et al. 2006). The genes coding for GBS80 and GBS104 are localized in a genomic island, named Pilus Island 1 (PI-1), containing genes coding for three LPXTG proteins and two sortases with similar organization to the genes coding for pilus-like structures in *C. diptheriae* (Fig.5A) (Ton-That and Schneewind 2003).

The genes are transcribed in the same direction, indicating that they are part of an operon. GBS80, GBS104, and GBS52 (SAG0646), represent the three LPXTG motif containing proteins of this island. The other two genes (SAG0647 an

SAG0648) code for sortase enzymes, which are known to catalyse the covalent linkage of LPXTG motif proteins to the peptidoglycan (Fig.5A).

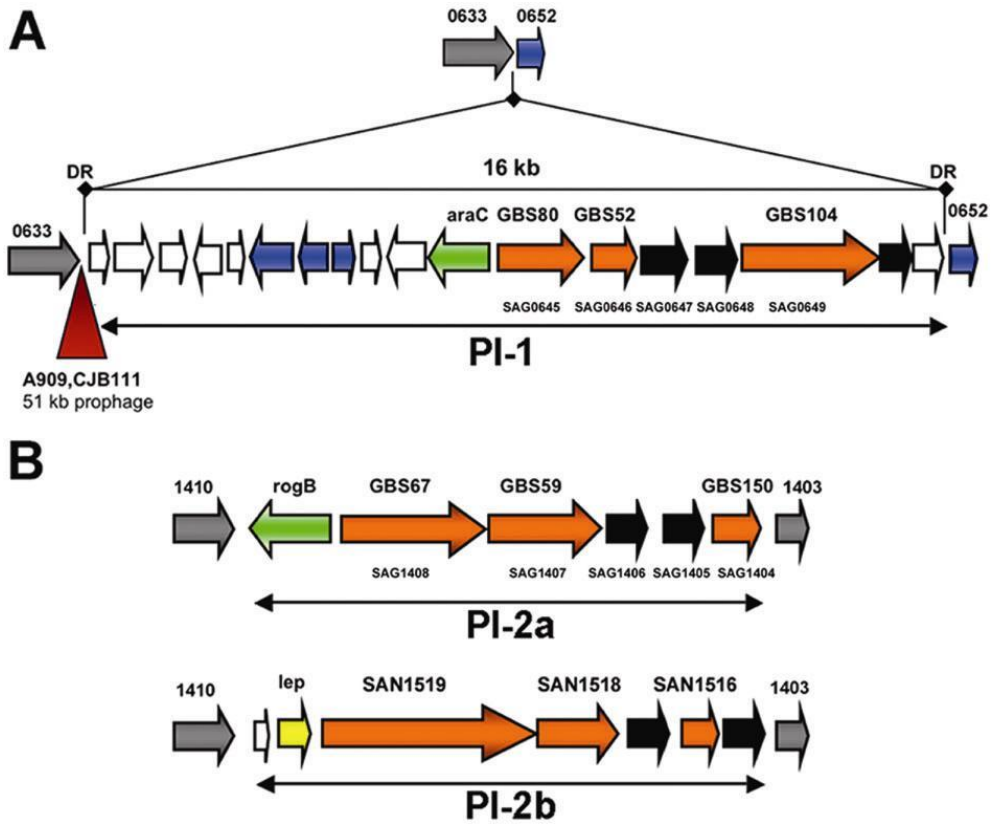


Figure 5. Schematic representation of GBS pilus-island regions. (A. pilus island 1; B. pilus island 2) Genes coding for LPXTG-containing proteins are represented with orange arrows, whereas transcriptional regulators are in green and conserved flanking genes are in grey. At least two sortases are present in each PI (black arrows), while a signal peptidase is present in PI-2b (yellow arrow). In PI-1, transposable elements are also present (blue arrows), as well as interrupted or frame-shifted genes (white arrows). The insertion site for the 51 kb prophage in PI-1 of strains A909 and CJB111 is shown. For PI-1 and PI-2a, gene numbers are relative to the database annotation for strain 2603 V/R, while for PI-2b, gene numbers are relative to COH1 strain. DR: direct repeat (Rosini, Rinaudo et al. 2006).

PI-1 consists of an approximately 16 kbp-long DNA region flanked by 11 bp of direct repeats, and it has been found in ≈ 70% of the GBS strains that have been analysed (Tettelin, Masignani et al. 2005; Margarit, Rinaudo et al. 2009). Two conserved genes (*sag0633* and *sag0652*), that are present in all GBS strains that have been analysed, flank this DNA region. In strains that lack the region, the flanking genes are contiguous. In addition to the pilus genes, the genomic island contains a gene that encodes an AraC-type transcriptional regulator, as well as a gene (*spy0123*) that encodes a heat-shock protein (Hsp33) and remnants of transposase-like genes. Two strains, A909 and CJB111, contain an insertion of a 51.2 kb-long prophage at one end of the 16 kbp-long island (Fig. 5A) (Rosini, Rinaudo et al. 2006). The overall organization of this genomic region suggests that the complete island may have been acquired by horizontal DNA transfer.

The other two protective antigens, GBS67 and GBS59, are located in a second island with a similar organization to Pilus Island-1 and for this reason named Pilus Island 2 (PI-2) (Fig. 5B). As PI-1, the second pilus locus is located in a variable region of the genome and contains genes coding for three LPXTG proteins (GBS67, GBS59, and GBS150) and two sortases (Fig. 5B).

There are two variants of this region (PI-2a and PI-2b), which differ in an 11-kb segment of DNA that is flanked by identical conserved genes (*sag1403* and *sag1410*). The two variants encode for distinct pili that have only limited amino-acid sequence similarity. PI-2a contains, in addition to pilus genes, a gene that encodes for a RogB-type transcriptional regulator. PI-2b lacks the transcriptional regulator but contains a gene that encodes for a protein similar to the LepA-type signal peptidase of Gram-negative bacteria (Fig.5B).

In summary, there are three genomic islands in GBS that are found at two different genomic locations. The three islands are similar in organization but poorly conserved among different isolates. All strains analyzed carried at least 1 of the islands, and 94% expressed pili on their surface (Margarit, Rinaudo et al. 2009). PCR and FACS analysis on a wide panel of GBS clinical isolates revealed that pilus 2a is the most represented and surface exposed among the three pilus types (Margarit, Rinaudo et al. 2009).

Immunoblot analysis, using sera raised against the three LPXTG proteins present in each island, showed that all proteins were part of high molecular weight (HMW) covalently-linked polymers (Fig.6A, C and E). Immunogold electron microscopy (IEM), using antibodies raised against GBS80 (for PI-1), GBS59 (for PI-2a) and GBS1518 (for PI-2b) showed that these polymers constitute pilus-like structures extending beyond the bacterial surface (Lauer, Rinaudo et al. 2005) (Rosini, Rinaudo et al. 2006) (Fig. 6B, D and F).

Each PI of GBS contains two genes encoding SrtC transpeptidases. Generation of deletion mutants showed that both enzymes are capable of polymerizing the backbone pilus subunit, but each preferentially incorporates one of the two ancillary proteins (Rosini, Rinaudo et al. 2006).

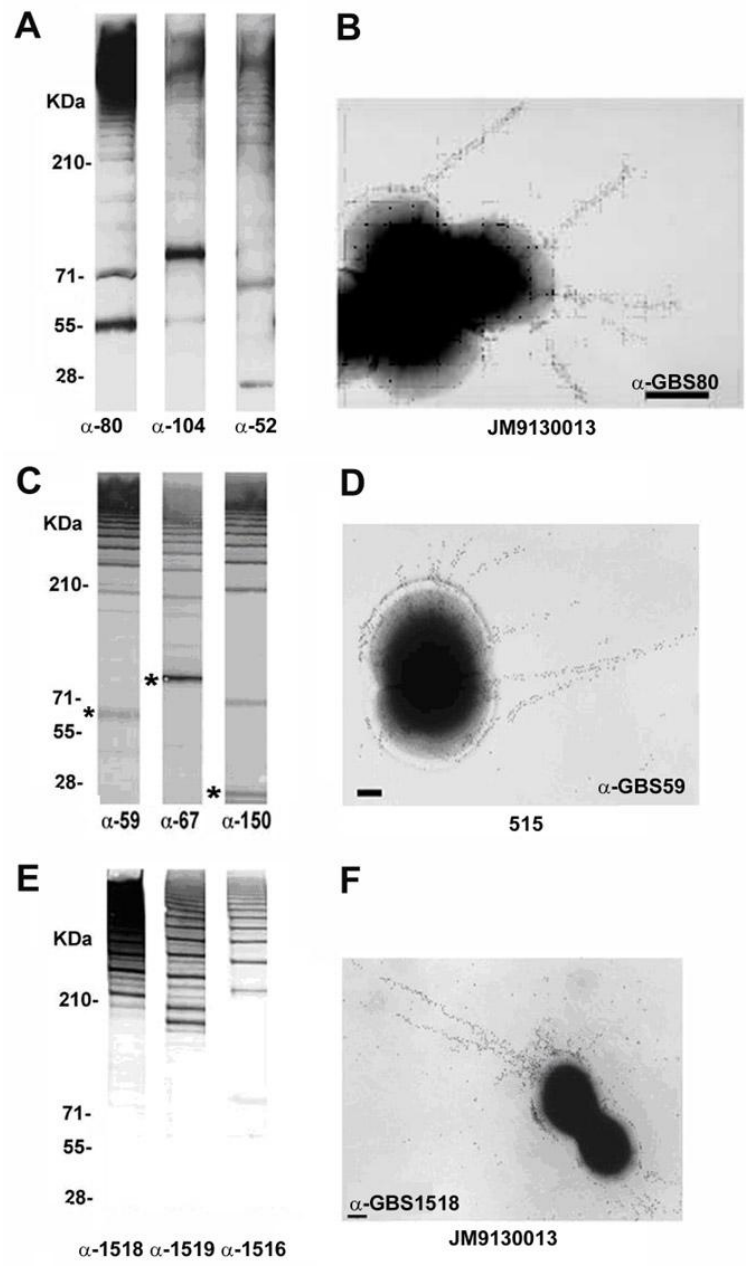


Figure 6. Novel genomic islands code for pilus-like structures. (A) Immunoblots of total protein extracts from JM9130013 strain probed with antisera specific for PI-1 proteins GBS80 (α-80), GBS104 (α-104) and GBS52 (α-52). **(B)** Immunogold labeling and transmission electron microscopy of GBS80 in strain JM9130013, showing long pilus-like structures.**(C)** Immunoblots of total protein extracts from 515 strain probed with antisera specific for PI-2a proteins GBS59 (α-59), GBS67 (α-67) and GBS150 (α-150). Asterisks (*) indicate the monomeric form of GBS59, GBS67 and GBS150. **(D)** Immunogold electron microscopy of 515 strain incubated with sera raised against GBS59 protein and labeled with secondary antibodies conjugated with 10nm gold particles. **(E)** Immunoblots of total protein extracts from JM9130013 strain probed with antisera specific for PI-2b proteins SAN1518 (α-1518), SAN1519 (α-1519) and SAN1516 (α-1516). **(F)** Immunogold electron microscopy of JM9130013 wt strain incubated with sera

raised against GBS1518 protein and labeled with secondary antibodies conjugated with 10nm gold particles (Rosini, Rinaudo et al. 2006).

There is growing evidence that, in addition to the SrtC transpeptidases, the housekeeping SrtA may play a role in GBS pilus assembly. Indeed, a study based on the generation of a knock-out strain for *srtA* gene revealed that the enzyme is not involved in pilus polymerization, but it is essential for the permanent anchoring of GBS pilus 2a to the cell wall (Nobbs, Rosini et al. 2008). Moreover, a detailed analysis of PI-2a identified the ancillary protein GBS150 as the substrate for SrtA.

1.3 Sortase enzyme in Gram-positive bacteria

In Gram-positive bacteria, a class of surface proteins are covalently anchored on the cell wall by a transpeptidase, which has been called sortase (Srt) (Paterson and Mitchell 2004) (Ton-That, Marraffini et al. 2004) (Clancy, Melvin et al.). Sortases are positioned at the cytoplasmic membrane via a membrane anchor located either at the N- or C-terminus, contain the active site, LxTC motif (Marraffini, Dedent et al. 2006), of which cystein is essential for the sortase activity (Ton-That, Liu et al. 1999) and recognize their substrate proteins via a common C-terminal pentapeptide sequence, which acts as a cell wall sorting signal.

So far, more than 700 putative sortase substrates encoded by more than 50 different prokaryotic genomes have been identified (Nguyen, Phan et al.).

These enzymes have also been developed into powerful molecular biology reagents to site-specifically attach proteins to a variety of biomolecules (Tsukiji and Nagamune 2009) (Popp and Ploegh). Although they are not essential for bacterial viability when cells are grown in rich media, sortases can be important virulence factors as they display surface proteins that mediate bacterial adhesion to host tissues, host cell entry, evasion and suppression of the immune response and acquisition of essential nutrients. The sorting reaction catalyzed by the sortase A protein from *Staphylococcus aureus* (Sa-SrtA) is the best understood and begins when a full-length precursor protein containing an amino terminal leader peptide is exported from the cytoplasm through the secretory pathway (Fig.7). The C-terminal CWSS is then processed by Sa-SrtA. The CWSS consists of a

LPXTG motif, followed by a segment of hydrophobic amino acids, and a tail composed primarily of positively charged residues. The C-terminal charged tail presumably retards export, positioning the protein for processing by the extracellular membrane associated Sa-SrtA enzyme. A highly conserved active site cysteine residue in Sa-SrtA then nucleophilically attacks the backbone carbonyl carbon of the threonine residue in the LP*X*TG motif, breaking the threonine and glycine peptide bond and creating a sortase-protein complex in which the components are linked via a thioacyl bond. The protein is then transferred by Sa-SrtA to the cell wall precursor lipid II, when the amino group in this molecule nucleophilically attacks the thioacyl linkage to create an isopeptide linked protein-lipid II product. Transglycosylation and transpeptidation reactions that synthesize the cell wall then incorporate this product into the peptidoglycan, where it is covalently linked to the cross-bridge peptide. Other sortases catalyse a similar transpeptidation reaction, but join remarkably different LPXTG motifs and amino groups. Since the discovery of Sa-SrtA a little more than decade ago by Schneewind and colleagues (Mazmanian, Liu et al. 1999), over 800 genes encoding related proteins have been identified in ~260 distinct bacterial species (Finn, Mistry et al.). The vast majority of sortases is found in Gram-positive bacteria that contain a conventional cell wall (they are absent in Mollicutes) (Pallen, Lam et al. 2001). Most bacterial species contain multiple sortase enzymes that have been named in an *ad hoc* manner (e.g. SrtA, SrtB, SrtC, etc.). To provide a framework in which to discuss their functions, the sortases from Gram-positive bacteria were grouped into families based upon their primary sequences (Fig.8). Approximately 60% of all sortase proteins can be partitioned into six

distinct families of enzymes that share related amino acid sequences, these include class A to F enzymes (Comfort and Clubb 2004) (Dramsi, Caliot et al. 2006). Experimental and bioinformatics analyses indicate members of each group recognize distinct CWSSs in which the LPXTG sequence is varied (hereafter called sorting signal motifs). Class A enzymes are present in Firmicutes and have been studied extensively. They appear to perform a housekeeping role in the cell as members of this group are capable of anchoring a large number of functionally distinct proteins to the cell wall. Class B enzymes are also present in Firmicutes and can have distinct functions. Some members of this group attach haem-receptors to the peptidoglycan, while others assemble pili. Most surface proteins attached by class A enzymes contain a canonical LPXTG motif within their CWSS and have diverse functions that can promote bacterial adhesion, nutrient acquisition, host cell invasion, and immune evasion. Class A enzymes have attracted significant interest as potential drug targets because a number of clinically important pathogens use these sortases to display virulence factors and they are attenuated in their virulence if their *srtA* gene is eliminated (*S. aureus, L. monocytogenes, Streptococcus pyogenes* and *Streptococcus pneumoniae* among others) (Naik, Suree et al. 2006) (Maresso, Chapa et al. 2006).

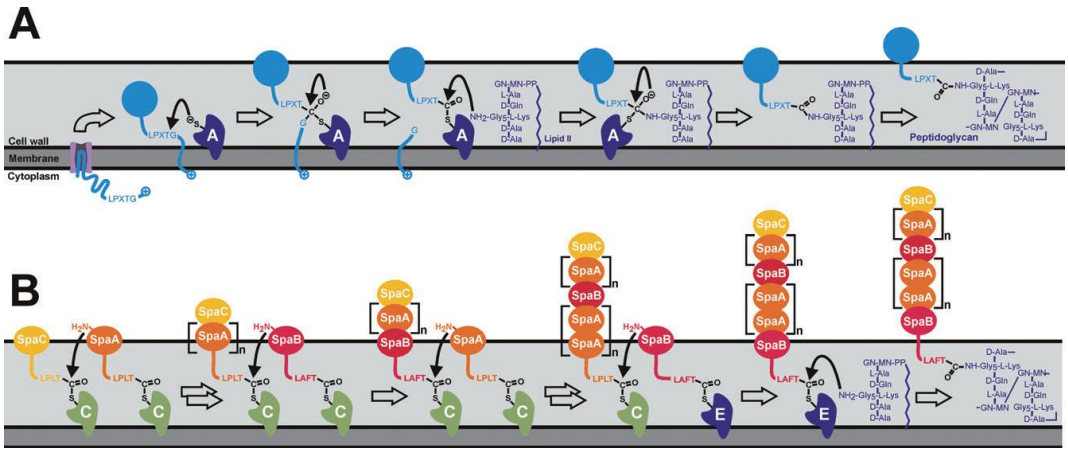


Figure 7. Mechanisms of sortase mediated attachment of surface proteins and pilus assembly at the bacterial cell wall.

(A) The *S. aureus* housekeeping sortase A anchors surface proteins to the peptidoglycan. The precursor protein containing an amino terminal leader peptide is secreted across the membrane through the Sec pathway. The exported protein (light blue) is processed by the sortase enzyme (dark blue, labelled 'A'), which recognizes the LP*X*TG sequence and cleaves the surface proteins between the threonine and glycine residues of the motif. The enzyme then recognizes the pentaglycine cross-bridge peptide of lipid II as the second substrate. Subsequent formation of a peptide bond between the carbonyl of the threonine and the free amino group of the cross-bridge peptide results in covalent attachment of the protein to lipid II. The surface protein is then fully incorporated into the cross-linked peptidoglycan via the transglycosylation and transpeptidation reactions during the bacterial cell wall synthesis. The sphere coloured light blue represents the folded form of the cell surface displayed protein. **(B)** Pilin-specific and housekeeping sortases assemble the SpaA pilus in *C. diphtheria*. The formation of complexes between the pilus-specific sortase C (light green) and the tip protein SpaC (light orange) initiates pilus assembly. The class C enzyme also recognizes the main pilin subunit SpaA (orange) forming SrtC–SpaA complexes. Nucleophilic attack by the free amino group originating from a lysine residue present in SpaA results in dissolution of the sortase–SpaC intermediate and the formation of a sortase–SpaA–SpaC complex. Repetition of this transpeptidation reaction results in pilus elongation. The class C sortase also incorporates the minor pilin SpaB (red) into the growing shaft by an analogous mechanism. Termination of pilus biogenesis is presumably initiated when the pilin polymer is transferred to the class E typehousekeeping sortase (dark blue), which subsequently catalyses the nucleophilic attack by the amino group within lipid II. In the final assembly step the lipid II linked pilus is incorporated in the murein sacculus via normal cell wall biosynthesis (Spirig, Weiner et al.).

Class C enzymes are broadly distributed in Gram-positive bacteria and function as pilin polymerases that construct pili. Class D enzymes predominate in Bacilli and in *Bacillus anthracis;* this type of enzyme anchors proteins to the cell wall that facilitate sporulation. Actinobacteria contain class E and F enzymes whose functions are largely unknown. In *Corynebacterium diphtheriae* a class E enzyme appears to perform a housekeeping function similar to class A enzymes (Ton-That and Schneewind 2003), while class F enzymes have yet to be studied. Sortases are also present in a few Gram-negative and archaebacterial species, but the functions of these enzymes are unknown (Pallen, Lam et al. 2001; Pallen, Chaudhuri et al. 2003; Comfort and Clubb 2004).

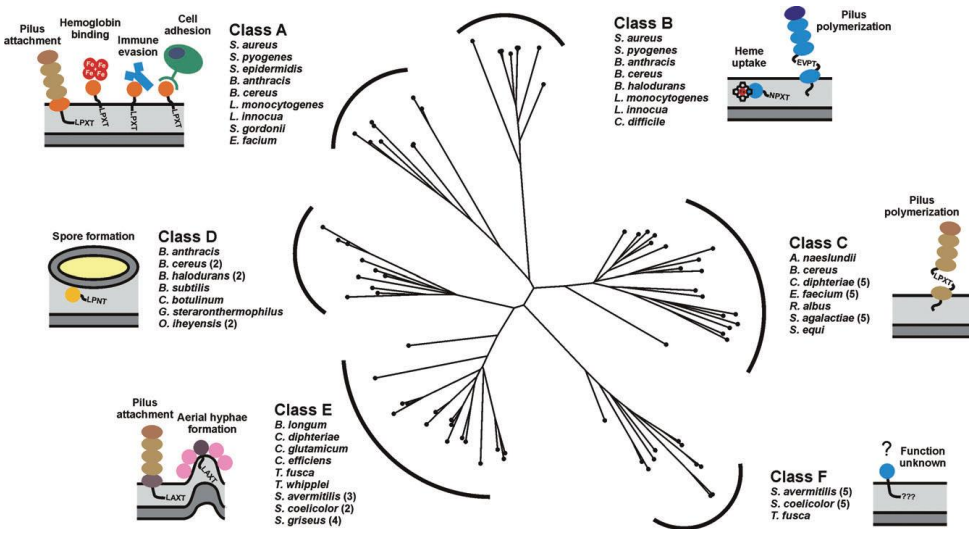


Figure 8. Phylogenic tree showing the relationships among the six classes of sortases from Gram-positive bacteria. A multiple sequence alignment based on pairwise constraints of a selected set of 73 sortase proteins was generated using the program COBALT and a phylogenetic tree constructed using the neighbour joining method (Papadopoulos and Agarwala, 2007). The analysed sortases can be partitioned into six distinct subfamilies based on their primary sequences. It should be noted that the class D and E enzymes described here are collectively referred to as a class D enzymes by Bierne and colleagues (Dramsi *et al*., 2005). Class D and E enzymes have also previously been

referred to as subfamily-4 and -5 enzymes (Comfort and Clubb, 2004). The bacterial species associated with the enzyme classes A–F are listed and schematic representations of the main biological function of their corresponding sortase substrates are illustrated (Spirig T. et al, Molecular Microbiology, 2011).

1.4 The sortase A class

Members of this subfamily play a pivotal role in the cell, anchoring a large number of diverse proteins to the cell wall. The majority of surface proteins (a total of 511) are predicted to be anchored by SrtA-type sortases, which are distributed in a wide range of Gram-positive bacterial genera (*Bacillus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Listeria*, *Staphylococcus*, and *Streptococcus*). The prototype SrtA from *S. aureus* is included in this subfamily. Bacteria always encode only a single SrtA-type homolog, which on average is predicted to anchor a large number of proteins (≈ 12 substrates). The genes of the target proteins are never proximal to the gene encoding SrtA-type enzyme. The analysis of their predicted substrates suggests that members of this subfamily target the sequence LPXTG, in which X is often a lysine, a glutamate, an asparagine, a glutamine or an alanine (Fig. 9). A Pfam (Protein Family database) analysis of the predicted substrates indicates that they are functionally diverse (Bateman, Birney et al. 2000).

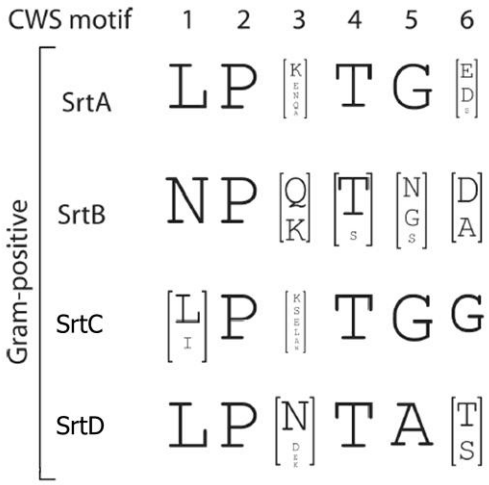


Figure 9. Sorting signals categorized by subfamily type. The figure shows the position-specific frequency of amino acids within the sorting signals of different types of sortases. The one-letter symbol for the amino acid residue is given for each position in the six-residue motif. The font size of each letter is proportional to the frequency with which an amino acid occurs. If an amino acid appears in fewer than 8% of the substrates, then the letter does not appear in the figure. When one type of amino acid is completely conserved at a particular position of the sorting signal motif or when one type of amino acid occurs in more than 92% of the CWS-containing proteins, then only one letter is present in a position. When no amino acid type is predominant in a given position of the motif, then the amino acid types found in the motif are given in brackets (Comfort and Clubb 2004).

Sortase A harbors an N-terminal hydrophobic segment that functions as a signal peptide for secretion and as a stop transfer signal for membrane anchoring (Fig. 10). The enzymes belonging to this subfamily adopts a type II membrane topology, with the N-terminus inside the cytoplasm and the C-terminal enzymatic portion located across the plasma membrane.

SrtA-type homologs have a low percentage of aa sequence identity (about 30% of *S. aureus* SrtA with those of other Gram-positive). This may suggest that sortases have coevolved with their substrates. Importantly, Gram-positive bacteria display significant differences at the third position of the stem peptide of a peptidoglycan subunit that can be substituted by variable side chains. Similarly, it has become

increasingly apparent that variation within the CWSS motif exists (Comfort and Clubb 2004). The amino acid composition and length of the transmembrane part or the charged tail constituting the CWSS vary between different Gram-positive bacteria. These observations suggest a coevolution of substrate(s)–enzyme pairs.

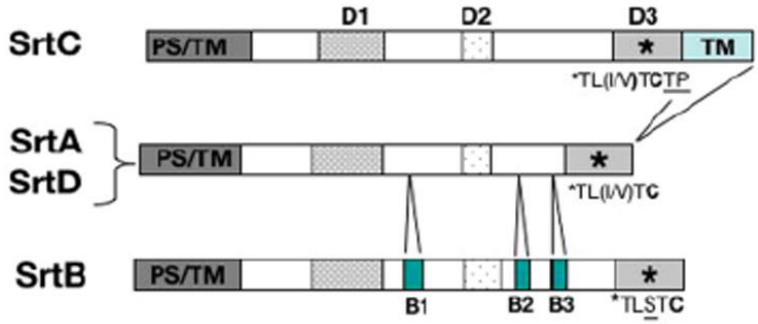


Figure 10. Four structural classes of sortases in Gram-positive bacteria. All sortases possess at their N-terminus the signal peptide and three conserved domains D1, D2 and D3. The two key amino acids forming the catalytic site are found in domains D2 (His120) and D3 (Cys184) of all sortases (numbering is according to the canonical Staphylococcus aureus SrtA sequence). Each class of sortases also possesses a specific pattern of conserved amino acids (Dramsi et al., 2005). The sortase B class (SrtB) possesses three additional amino acid segments (B1, B2, B3), which are not found in SrtA and the TLXTC motif, in which X is often a serine residue. The sortase C class (SrtC) possesses a typical C-terminal hydrophobic domain (TM) and a conserved proline residue located after the catalytic site TLXTC (Dramsi, Magnet et al. 2008).

The structure of *S. aureus* sortase A (206 amino acids) has been studied by nuclear magnetic resonance (NMR) and X-ray analysis.

A truncated version that lacked the first 59 amino acids retained the ability to cleave the LPXTG peptide and the transpeptidase activity *in vitro* (Ton-That, Liu et al. 1999). This enzyme adopts a unique eight-stranded β-barrel fold, which contains several short helices and loops (Fig. 11) (Ilangovan, Iwahara et al. 2001; Ilangovan, Ton-That et al. 2001; Zong, Bice et al. 2004).

The active site was found within an elongated hydrophobic groove formed by the β4, β7 and β8 strands. Two conserved residues, His120 and Arg197, are positioned in close proximity to the active site sulphydryl of Cys184 (Ilangovan, Ton-That et al. 2001; Zong, Bice et al. 2004).

Mutagenesis studies revealed that both His120 and Arg197 are involved in catalysis. Replacement of Cys184 by Ala completely abolished sortase activity both *in vitro* and *in vivo* and replacement of His120 or Arg197 by Ala drastically reduced the enzymatic activity (Ton-That, Mazmanian et al. 2002; Marraffini, Ton-That et al. 2004; Frankel, Tong et al. 2007). Cocrystals of sortase and LPETG peptide revealed that Cys184 and Arg197 reside between the side chains of the scissile T-G peptide bond (Zong, Bice et al. 2004).

Arg197 presumably stabilizes the binding of the substrate in the active site by donating hydrogen bonds from its guanidine group to the backbone carbonyl oxygens of the leucine and proline residues of the sorting signal (Suree, Liew et al. 2009).

Meanwhile, the histidine residue in a charged state may act as a general acid to protonate the leaving amide group of the scissile bond, facilitating collapse of the tetrahedral intermediate.

In the NMR structure, the β3/β4 and β6/β7 loops contain a set of acidic residues involved in calcium binding. Addition of Ca^{2+} in the reaction stimulates sortase activity eightfold, probably by a mechanism that may facilitate substrate binding (Naik, Suree et al. 2006). It has been shown that calcium ions are involved in structural rearrangements of a disordered loop (β6/β7 loop) covering the active site.

Regarding the LPXTG-binding site, mutagenesis and NMR studies revealed the importance of β6/β7 loop in determining substrate selectivity. Particularly, Val168 and Leu169 are important for binding the Leu-Pro region of the LPXTG peptide (Bentley, Lamb et al. 2008). Recently NMR structure of the covalent SrtA-substrate complex identified the LPXTG binding site in a more large groove, formed by β4 and β7 strands, together with β7/β8, β3/β4 and β2/H1 loops, other than the β6/β7 loop already identified (Suree, Liew et al. 2009).

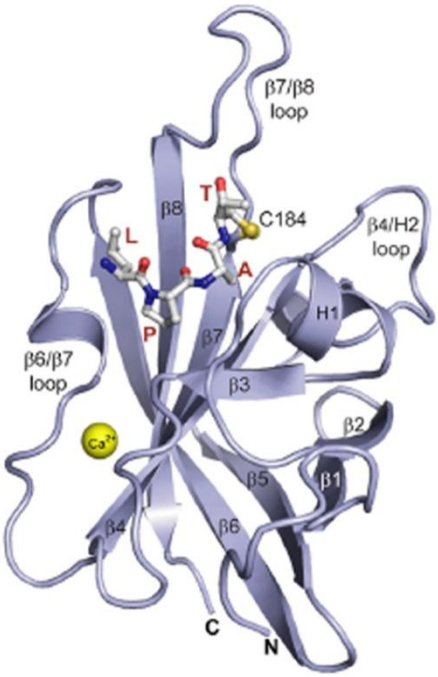


Figure 11. NMR solution structure of the *S. aureus* SrtA$_{\Delta N59}$-LPAT* complex. *Ribbon drawing* of the structure of the SrtA$_{\Delta N59}$-LPAT* complex. The covalently bound peptide is shown in a *red ball-and-stick representation* with its amino acids labeled. A *yellow sphere* represents the calcium ion. The core of SrtA$_{\Delta N59}$ is an 8-strand β-barrel. β4, β7 and β8 form a concave β-sheet, surrounded by some loop regions. The three important catalytic residues Cys184, His120 and Arg197 are located in the middle of the β-sheet (cys184 is labeled). (Suree, Liew et al. 2009).

1.5 Protein engineering using sortase enzymes

1.5.1 Engineering of bacterial surfaces

The sortase-mediated system of anchoring proteins to the cell wall of Gram-positive bacteria was first exploited to decorate these microbes with heterologous proteins. Such experiments require the creation of a genetic fusion of the heterologous protein to the sorting motif. The heterologous protein is then

expressed and directed to the surface though the normal cell-wall sorting pathway. In this manner, the enzyme alkaline phosphatase has been anchored to the cell wall of Staphylococcus aureus (Schneewind, Model et al. 1992), the E7 protein of Human papilloma virus 16 (HPV16) has been displayed on Streptococcus gordonii, a commensal microbe in the oral cavity (Pozzi, Contorni et al. 1992), and a-amalyase has been affixed to the peptidoglycan of Bacillus subtilis, helped by coexpression of the sortase gene from Listeria monocytogenes (Nguyen and Schumann 2006). The peptidoglycan cell wall can even be decorated with non-natural entities (fluorescein, biotin, azide) by incubating dividing S. aureus cultures with chemical probes appended to the N terminus of an LPXTG peptide (Nelson, Chamessian et al.). The incorporation of what are in essence N-terminal labeling probes occurs through use of the endogenous sortase enzyme and anchors the exogenously provided probes onto available pentaglycine side chains of the cell wall (Popp and Ploegh).

1.5.2 C-Terminal and N-Terminal

The ability of sortase to recognize the sorting motif when transplanted onto recombinantly expressed proteins allows the site-specific incorporation of moieties and functional groups that cannot be encoded genetically (Fig.12). This method requires only that the LPXTG motif be solvent exposed and usually results in high yields of the desired transpeptidation product. Indeed, many substrate proteins have now been labeled with probes bearing a wide range of functionalities, including biotin, fluorophores, cross-linkers, and multifunctional

probes (Popp, Antos et al. 2009). The labeling of recombinant proteins by sortase A requires no sophisticated synthetic chemistry; most of the probes are readily accessible by standard peptide synthesis, using off-the shelf reagents. The production and folding of recombinant substrate proteins is not usually compromised by the presence of the small C-term LPXTG tag. Since all transformations are carried out using sortase under physiological buffer conditions (pH, ionic strength, ionic requirements) on substrates whose proper folding and activity status can be ascertained prior to starting the reaction, loss of biological activity is rarely, if ever, observed for the final product. The ability to engage in a sortase-catalyzed transacylation appears to be determined solely by the accessibility and flexibility of the sorting motif. The utility of the sortase labeling method stems from the fact that the enzyme tolerates substrates unrelated in structure and sequence immediately upstream from the cleavage site. This property is not unexpected, given the role of sortase in anchoring a broad range of protein substrates to the cell wall (Popp and Ploegh).

Protein labeling at the N terminus can be accomplished simply by moving the placement of the sortase recognition element from the protein to the short peptide probe and by inclusion of a suitable number of glycine residues at the N terminus of the target protein (Fig.12). Both methyl ester mimetics of the sortase motif (Antos, Chew et al. 2009) as well as the complete LPXTG sortase recognition motif can be used as scaffolds for such probes (Yamamoto and Nagamune 2009). Conceptually, this labeling technique is analogous to the C-terminal labeling, except the acyl-enzyme intermediate is generated between sortase and the peptide probe, and the protein to be labeled bears several glycine residues at the N

terminus, the NH_2 group of which serves as the nucleophile. This strategy was used to install fluorescent probes at the N terminus of membrane proteins in living mammalian cells after a clever initial unmasking step by sortase itself to expose the nucleophilic glycine (Yamamoto and Nagamune 2009).

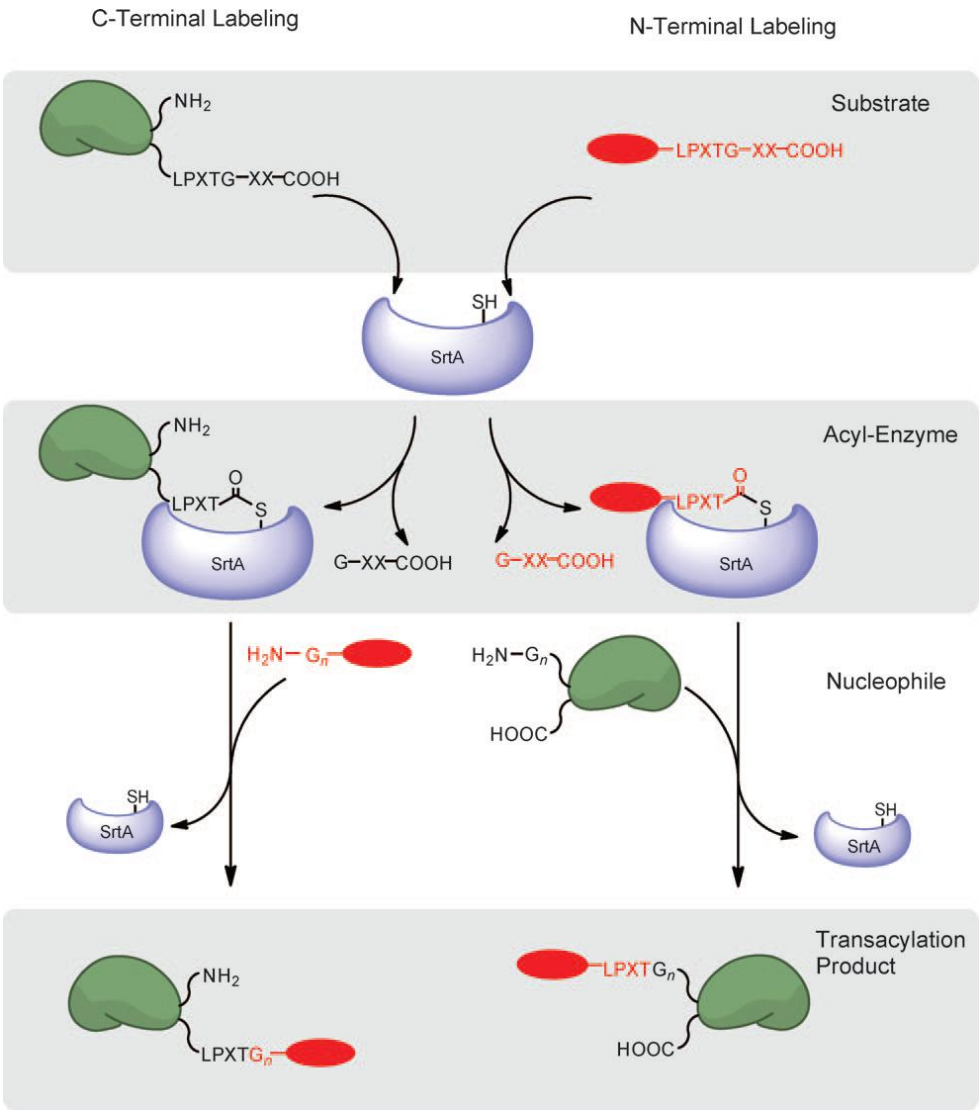


Figure 12. Site-specific C- and N-terminal labeling scheme using sortase A. C-Terminal labeling (left) and N-terminal labeling (right) proceed through a substrate-recognition step (top), followed by generation of a thioacyl intermediate (middle) and resolution of the acylated enzyme by an exogenously added nucleophile (bottom) (Popp and Ploegh).

1.5.3 Other sortases application as protein engineered

Sortase methods allow the production of homogeneous recombinant protein preparations that are modified with nongenetically templated post-translational modifications. Glycoproteins, normally elaborated by a complex set of enzymatic events in the secretory pathway, can thus be constructed. LPXTG-tagged proteins and peptides can be modified with 6-aminohexose-based sugar nucleophiles, including aminoglycoside antibiotics and their analogues (Samantaray, Marathe et al. 2008). Glycosylphosphatidylinositol (GPI) anchors, normally attached at the C terminus of proteins, can be phenocopied by ligation of LPXTG peptides to synthetic glycine nucleophiles, which in turn are linked to the phosphoethanolamine moiety on a GPI derivative (Guo, Wang et al. 2009). Lipidation of proteins is yet another important post-translational modification that has been poorly studied because of the lack of tools available to obtain homogeneous preparations of lipoproteins. Sortase has been used to fill this void (Antos, Miller et al. 2008). A glycine-based scaffold was modified with a panel of linear alkyl chains (C12–C24) as well as with cholesterol or adamantane, and then used to modify a suitably LPETG-tagged version of eGFP. These eGFP lipoproteins associated with the plasma membranes of living cells in a chain-length-dependent fashion (the optimum being a C22 chain), from where they gained access to the endosomal compartment.

Moreover, covalent immobilization of proteins onto solid supports has been accomplished by sortase. A major advantage of the method is that the specificity of the enzyme enables proteins to be immobilized uniformly and in a defined

orientation on the solid surface for subsequent exposure to the analyte of interest (Popp and Ploegh).

1.6 Sortases that assemble pili: class C enzymes

Gram-positive bacteria use class C enzymes to build pili that promote microbial adhesion and biofilm formation. First, one or more class C enzymes form the long thin shaft of the pilus by linking together pilin subunits via isopeptide bonds. The base of the pilus is then anchored to the cell wall by a housekeeping sortase or, in some cases, the class C enzyme itself (Spirig, Weiner et al.).

Recently, an extensive characterization of pilus-associated sortases from *Streptococcus pneumoniae* pilus 1 (SrtC-1, SrtC-2, and SrtC-3) was performed, and the X-ray structures of all 3 SrtC enzymes have been solved (Neiers, Madhurantakam et al. 2009) (Manzano, Izore et al. 2009). The overall fold of all three enzymes is very similar to other known sortases, corresponding to a β-barrel structure, composed of eight anti-parallel β-strands linked by multiple helices. The catalytic triad (constituted of His131, Cys193, Arg202 in SrtC1; His159, Cys221, Arg230 in SrtC2; His144, Cys206, Arg215 in SrtC3) within the substrate binding region is encapsulated by the lid, which maintain the active site in a closed conformation in the absence of substrate. The lid anchoring within the active site is through multiple interactions with key catalytic residues (Manzano, Contreras-Martel et al. 2008; Neiers, Madhurantakam et al. 2009). Structural comparison of the three pilus-associated sortases revealed some slight differences in terms of flexibility, positioning and number of residues of the lid and B-factor

values of the N-terminal helices. An additional helix in the C-terminal region is only present in SrtC-3. Some structural differences suggested a molecular explanation for the functional differences observed among these sortases, in terms of substrate specificity and incorporation of the ancillary pilins into pili.

Manzano *et al.* have also showed that site-specific mutations of the anchor residues in the lid region did not affect backbone protein recognition or the formation of the acyl-intermediate; however, the stability and the efficiency of the enzyme were negatively affected (Manzano, Izore et al. 2009). While the catalytic triad of Cys, His, and Arg side chains within the active site cleft is absolutely conserved among different classes of sortases (Zong, Bice et al. 2004) (Zong, Mazmanian et al. 2004), including SrtA from *Staphylococcus aureus*, the region corresponding to the lid is thus far found only in X-ray diffraction solved crystal structures of pilus-related C sortases in Gram-positive bacteria (Manzano, Contreras-Martel et al. 2008) (Weiner, Robson et al.).

The crystal structures of several other pilin-related class C sortases, including AcSrtC-1 from Actinomyces oris (Persson 2011), SrtC1 from S. suis (Lu, Qi et al. 2011) and GBS (Cozzi, Malito et al. 2011; Khare, Fu et al. 2011; Khare, Krishnan et al. 2011), have been reported. These structures all reveal a core 8-stranded β-barrel, with the catalytic triad (His, Cys, Arg) situated in the active site at the end of a groove along one side of the β-barrel. The GBS and S. suis SrtC1 structures were determined with the active-site in the 'open' conformation, while the other structures showed the active site occluded by a loop region, termed the lid. The lid in SrtC1 from GBS PI-2a (SrtC1-2a) and Actinomyces oris SrtC2 is dispensable for sortase activity *in vivo* (Wu, Mishra et al.; Cozzi, Malito et al. 2011).

1.7 Class C sortases in GBS

Sequence comparison by multiple alignment and phylogenetic analysis permitted the identification of 3 major clusters, corresponding to class C sortases of PI-1, PI-2a, and PI-2b, with amino acid identities ranging from 15 to 60% (Fig. 13). While both variants of PI-2 (PI-2a and PI-2b) contain 2 sortase genes (SrtC1 and SrtC2), PI-1 carries a third gene (SrtC3) predicted to code for a C sortase not directly involved in pilus polymerization (Rinaudo, Rosini et al.; Buccato, Maione et al. 2006; Rosini, Rinaudo et al. 2006). Moreover, the crystal structure of the sortase SrtC1 of PI-2a (SAL_1484) was solved at high resolution (Cozzi, Malito et al. 2011).

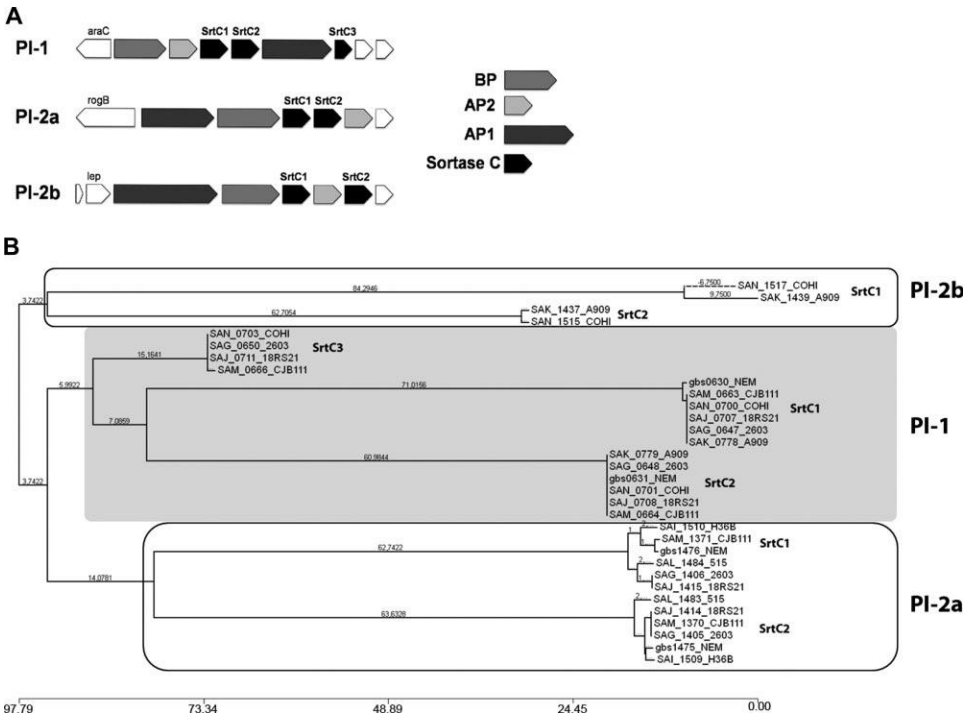


Figure 13. Class C sortases in PIs of GBS. (A) Schematic representation of GBS PIs. **(*B*)** Phylogenetic tree inferred from the alignment by the neighbor-joining distance-based method of C sortases from the available genomes of GBS. Single sortases are indicated by TIGR annotation. The 3 major clusters, highlighted in the boxes, include C sortases of each PI (Cozzi, Malito et al. 2011).

1.7.1 Structural organization and biochemical characterization of PI-1 and PI-2a sortase C enzymes

The ectodomain of sortase SrtC1 of PI-2a (residues 43–254) was crystallized and the structure was solved by molecular replacement, using as the initial search model the coordinates of sortase C1 of *S. pneumoniae* (PDB 2W1J).

The overall folding of GBS SrtC1 is highly similar to the folding of previously determined pilus-associated sortases. A β-barrel made of 9 antiparallel β-strands forms the core of the enzyme; a so-called roof made of 3 α-helices positioned above the β-barrel and a loop (known as the "mobile lid") that covers the active site (Fig.14A). This last one is positioned on one inner side of the β-barrel core and is made of the catalytic triad His157-Cys219-Arg228. The lid of SrtC1 harbors 3 residues, Asp84, Pro85, and Tyr86 ,which make interactions with residues of the active site and surroundings. While Asp84 and Pro85 are highly conserved, an aromatic residue (Tyr86 in SAL_1484) generally occupies the third position. The carboxylate group of Asp84 forms a salt bridge with the side chain of the conserved catalytic residue Arg228 (Fig. 14B) and with a water molecule (W76). The ring of Tyr86 is positioned in a pocket lined by highly conserved hydrophobic residues (Leu131, Leu138, Val153, Leu217) on one side and by the catalytic residue His157 on the other side. The residue Pro85 points toward the same hydrophobic pocket. The aromatic benzene ring of Tyr86 is close enough to the catalytic Cys219 side chain to make an aromatic-sulfur interaction (Cozzi, Malito et al. 2011). As shown previously, this sulfur-aromatic interaction is conserved in other sortases (SrtB and SrtD), and this finding suggests that this serves as a general mechanism of anchoring the lid within the active site (Viguera

and Serrano 1995) (Fig.14B). This sulfur-aromatic interaction has been postulated to strengthen the anchoring of the lid within the active site (Neiers, Madhurantakam et al. 2009). In addition, the hydroxyl group of Tyr86 makes H-bond interactions with the hydroxyl side chain of the highly conserved Thr155 and with the backbone amino group of the conserved Ala156 (Fig.14B). The aromatic ring of Tyr86 is also positioned in a hydrophobic environment where it potentially can be involved in CH-π weak polar interactions. This network of interactions between catalytic residues and those located on the lid (Asp84 and Try86) is postulated to regulate the movement of the lid and therefore the access of LPXTG substrates to the active site (Manzano, Izore et al. 2009; Cozzi, Malito et al. 2011).

The active site of GBS SrtC1 is made of the highly conserved catalytic triad His157, Cys219, Arg228 (Fig.14) and through site-directed mutagenesis and *in vivo* complementation studies, it has been demonstrated that each residue in the catalytic triad is essential for pilus polymerization; these data confirm the relationship between GBS C sortases and other members of sortase family. However previously reported NMR data on SrtA of *S. aureus* describe large chemical shift changes in the amide nitrogen and proton atoms of residues localized in specific loops on calcium ion addition, leading to the prediction that those residues form a structurally ordered calcium-binding site (Ilangovan, Ton-That et al. 2001; Cozzi, Malito et al. 2011). The absence of significant NMR chemical shift change on addition of EDTA or $CaCl_2$ to GBS SrtC1 indicates that, while calcium binding is required for the activity of the housekeeping SrtA in *S. aureus*, GBS SrtC1 does not bind any calcium ion (Cozzi, Malito et al. 2011).

The crystal structure of GBS SrtC1 showed that catalytic residues are not accessible to pilus substrates (Fig.14), as they are locked by the lid. Moreover, the mutations of key residues Asp84 and Tyr86 in the lid region or the deletion of the entire lid region had no effect on pilus protein polymerization (Fig.15A) (Cozzi, Malito et al. 2011).

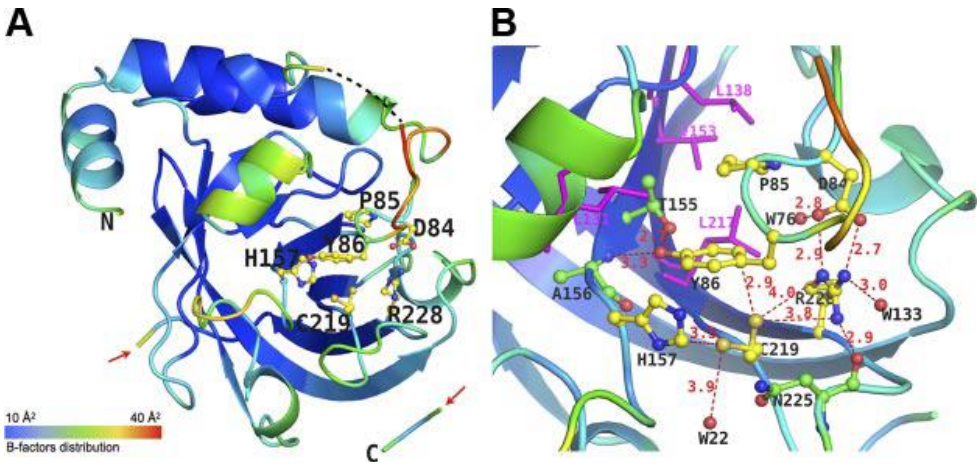


Figure 14. Overall folding of SAL_1484 and active site organization. *A*) Overall folding and *B* factors of SAL_1484. SAL_1484 is represented as a cartoon, colored according to *B*-factor distribution, from low (blue) to high (red). Residues forming the mobile lid and the active site are shown as balls and sticks and are labeled. N and C termini are labeled. Carbon, oxygen, an nitrogen atoms are depicted in yellow, red, and blue, respectively. Position of residues 92–93 of the mobile lid, missing from the model because of poor electron density, is indicated by black dashes. Red arrows indicate the gap in the C-terminal region, fragment of residues 240–249. *B*) Active site of SAL_1484. Residues forming the mobile lid (Asp84, Tyr86) and the active site (His157, Cys219, Arg228) are shown as balls and sticks, with carbon, oxygen, and nitrogen atoms in yellow, red, and blue, respectively. Conserved surrounding and interacting residues (Thr155, Ala156, Asn225) are shown as balls and sticks, with carbon, oxygen, and nitrogen atoms in green, red, and blue, respectively. Conserved hydrophobic residues are shown as magenta sticks and labeled in magenta. Distances between atoms are labeled and shown as red dashes. Water molecules are shown as red spheres. Background cartoon representation of SAL_1484 is colored according to *B* factors as in panel *A* (Cozzi, Malito et al. 2011).

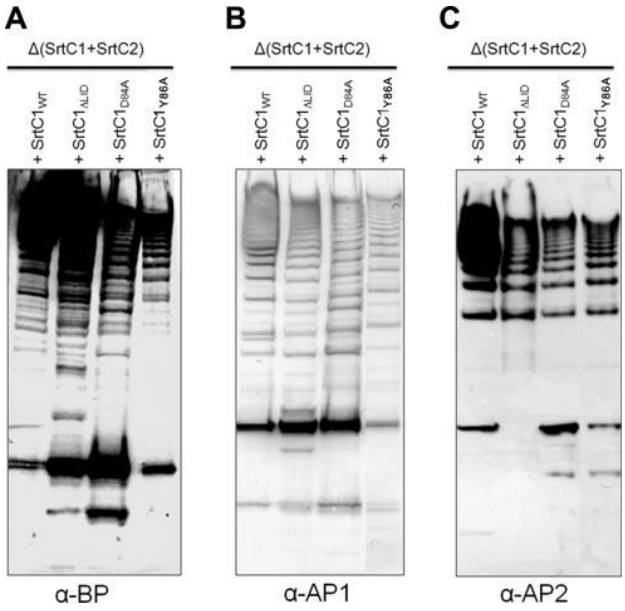


Figure 15. Lid region is not essential for pilus protein polymerization. Immunoblots of total protein extracts from 515 mutant strain of both sortase C genes _(SrtC1_SrtC2) complemented by plasmids expressing SrtC1 wild-type ($SrtC1_{WT}$) or SrtC1 carrying the mutation D84A ($SrtC1_{D84A}$) or Y86A ($SrtC1_{Y86A}$) or the deletion of the entire lid region (SrtC1Δlid). Nitrocellulose membranes were probed with antisera specific for the BP (*A*) and the ancillary proteins, AP1 (*B*) and AP2 (*C*) (Cozzi, Malito et al. 2011).

To better investigate the role of this region in catalysis, *in vitro* measurements of the kinetic properties of recombinant $SrtC1_{Y86A}$ and $SrtC1_{\Delta LID}$ in comparison with the $SrtC1_{WT}$ were performed. Accordingly to the *in vivo* data, the lid mutants can efficiently cleave the substrate peptide, and the rate of peptide cleavage by lid mutant variants was even higher than that obtained with the wild-type (Fig.16) (Cozzi, Malito et al. 2011).

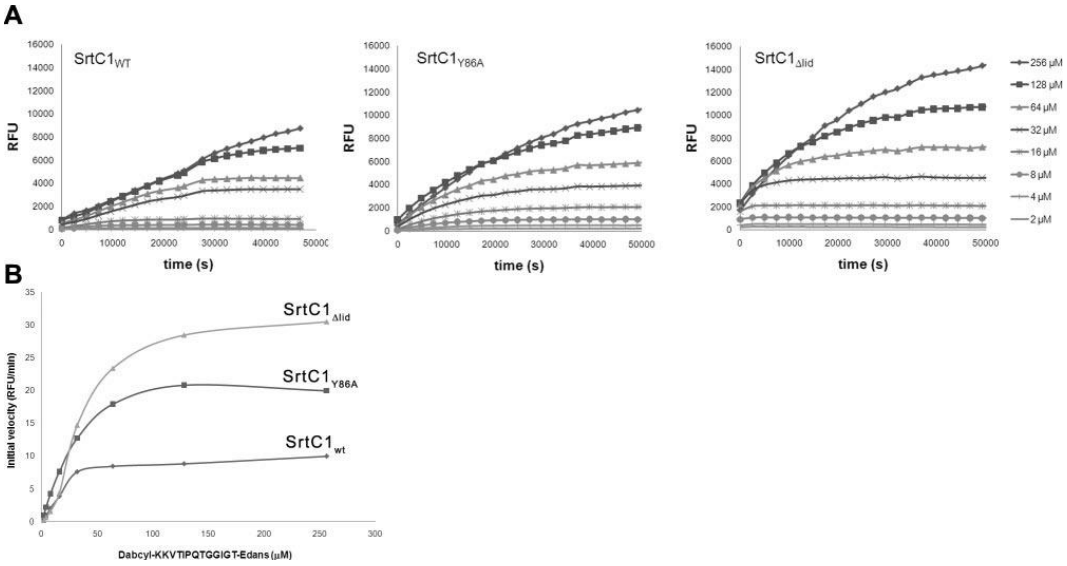


Figure 16. FRET assay with wild-type SrtC1 and lid mutants. (A) Progress curves of the cleavage reaction of PI-2a BP fluorescent peptide catalyzed by recombinant SrtC1 wild-type ($SrtC1_{WT}$), SrtC1 carrying the mutation Y84A ($SrtC1_{Y86A}$) and the deletion of the entire lid region ($SrtC1_{\Delta lid}$). Reactions containing 25µM of enzyme and from 2 to 256 µM of fluorescent peptide were performed at 37°C in 20 mM Tris (pH 7.5), 75 mM NaCl, and 1 mM DTT. ***(B)*** Rate [relative fluorescence units (RFU)/min] *vs.* concentration of substrate (Cozzi, Malito et al. 2011).

These results fit with the role of the lid suggested by the crystal structure, in which the lid covers the active site and sterically blocks the access of substrate. Structural and biochemical data suggest that the lid maintains the enzyme in an inactive closed conformation and that, for the enzyme activation, the lid needs to move. Therefore the deletion of the lid region does not abrogate pilus protein polymerization because its role is not catalytic; rather, it is a catalytic cleft-blocking loop, and only its movement can activate the enzyme *in vivo* (Cozzi, Malito et al. 2011).

The main questions remain to understand how this movement can be regulated by the interaction with the pilus proteins and to identify which are the residues involved in stabilizing the active open lid conformation of the enzyme. Based on

these analyses, the SrtC enzymes can be considered as having two functional domains: (i) an N-terminal regulatory region that contains the flexible inhibitory, pseudo-substrate lid, involved in enzyme regulation and probably specificity; and (ii) an enzymatic region, the b-barrel core that contains the catalytic triad (Cozzi, Prigozhin et al. 2012).

Moreover, the predicted C- and N-terminal TM domains of GBS SrtC1 are absolutely required for sortase biological function (Cozzi, Malito et al. 2011). The importance of TM domains for the enzyme activity has been recently reported for the pilus-associated sortase of *Corynebacterium diphtheriae*. Ton-That and co-workers (Guttilla, Gaspar et al. 2009) showed that the predicted C-terminal TM domain of pilus-associated sortase SrtA is essential for efficient pilus polymerization in *C. diphtheriae*. In addition, the evidence that the substitution in GBS SrtC1 of the C-terminal TM region with the corresponding hydrophobic helix of the backbone subunit of pilus type 2a could not restore enzyme activity strengthens the view that this region could play a key role in enzyme function. Finally, also in an *in vitro* FRET assay using fluorescent peptides mimicking the natural LPXTG substrates, the activity of SrtC enzymes was detected only when the enzyme contains the C-terminal TM region. These data suggest that the N-terminal hydrophobic helix could have a role in protein anchoring to the membrane, while the C-terminal TM region could be also involved in enzyme activity (Cozzi, Malito et al. 2011).

The crystal structures of GBS PI-1SrtC2 and SrtC1 were determined (Fig.17) . In both enzymes, the catalytic residues are not accessible to pilin substrates, suggesting that the enzymes cannot bind substrates in this conformation. Also

these sortase C enzymes contain an additional N-terminal extension of approximately 50 residues, composed of one or two α-helices and a lid that blocks the access of substrates to the active site. Ligand-free SrtC structures are more similar to the peptide-bound SrtA structure than to apo-SrtA. The structural similarity between the LPXTG peptide in the active site of SrtA suggests that the conserved residues in the lid that interact with the active site of GBS sortase act as a pseudo-substrate (Cozzi, Prigozhin et al. 2012). This observation further supports the already proposed regulatory role played by the lid in restricting the access of the pilin substrates to the catalytic cleft (Manzano, Contreras-Martel et al. 2008; Neiers, Madhurantakam et al. 2009; Cozzi, Malito et al. 2011). Moreover, structural analysis combined with *in vitro* experiments performed with fluorogenic peptides and with N-terminal deletion mutants of SrtC1 and SrtC2 show that the entire N-terminus, and not just the lid, as shown for GBS PI-2a SrtC1 (Cozzi, Malito et al. 2011), is disposable for catalysis (Cozzi, Prigozhin et al. 2012). Thus, the minimum active sortase region is the β-sheet core seen in the S. aureus SrtA structure and common to all sortase family members. The N-terminal extension is a unique feature of class C sortases and appears to function as a regulatory motif. Both class A and class C sortases cleave LPXTG-like motifs, but only sortase C can polymerize the pilus proteins to form high molecular weight structures. Hence, the different function of SrtC compared to SrtA, in terms of regulation, specificity or localization, may be due to the presence in this specific class of enzymes of a highly specialized N-terminal segment (Cozzi, Prigozhin et al. 2012).

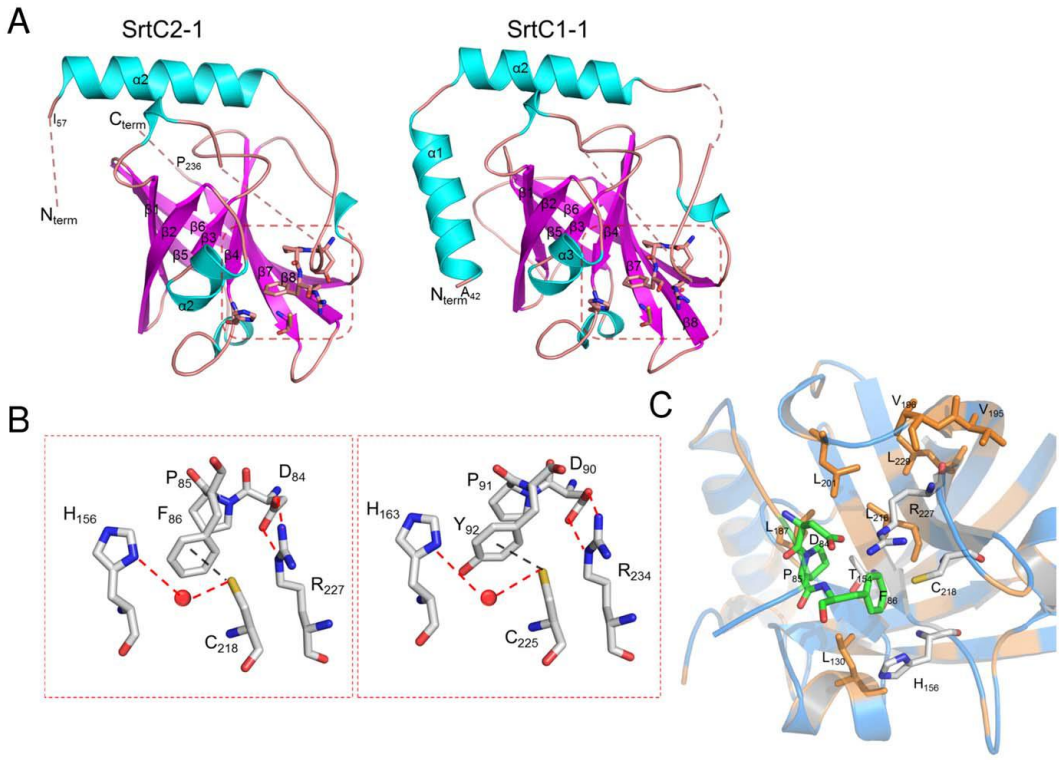


Figure 17. Overall fold of GBS PI-1 SrtC1 and SrtC2 and active site organization. (A) Overall fold of SrtC2 and SrtC1. Residues linking the mobile lid to the second helix and to the first beta-strand are missing in the final structures because of poor electron density, and are shown here as dashed lines. (B) Active sites of SrtC2 and SrtC1. Residues forming the mobile lid (Asp84-Phe86 in SrtC2 and Asp90-Tyr92 in SrtC1) and the active site (H156, C218, R227 in SrtC2 and H163, C225, R234 in SrtC1) are shown as sticks where sulfur, oxygen, and nitrogen atoms, are depicted as yellow, red, and blue, respectively. Water molecules are shown as red spheres. (C) The DPX motif is proximal to the catalytic triad of SrtC2, which is surrounded by conserved hydrophobic residues shown as sticks, where carbon, oxygen, and nitrogen atoms, are depicted as salmon, red, and blue, respectively (Cozzi, Prigozhin et al. 2012).

In vitro and *in vivo* complementation studies revealed that both GBS PI-1 C sortases can cleave all the LPXTG-like peptides tested and exhibit a functional promiscuity for pilin subunit incorporation into pili (Fig.18).

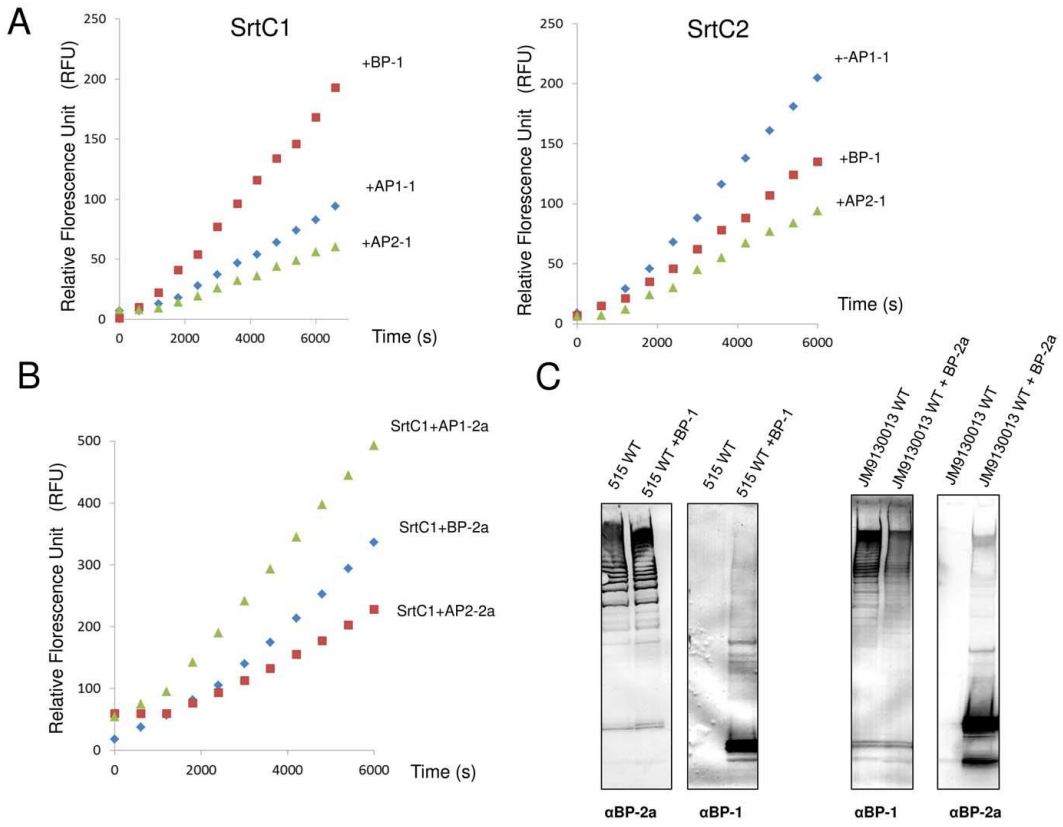


Figure 18. FRET assay with PI-1 and PI-2a peptides for substrate specificity analysis of PI-1 SrtC1 and SrtC2. (A) The reaction solutions contained 128 mM of PI-1 fluorescent peptides and 25 mM of SrtC1-TM (left panel) or SrtC2-TM (right panel). The reactions were performed at 37°C in the assay buffer containing 25 mM HEPES pH 7.5, 100 mM NaCl and 1 mM DTT. Fluorescence emission was monitored every 10 minutes and an increase in fluorescence intensity was observed in the presence of BP, AP1 and AP2 peptides mimicking the LPXTG motif of PI-1 pilus proteins. (B) Reactions were performed with PI-2a peptides and 25 mM of SrtC1-TM in the same conditions described above. (C) *In vivo* substrate specificity analysis. Immunoblots of total protein extracts from GBS 515 (containing SrtC1 and SrtC2 of pilus island 2a) and JM9130013 (containing SrtC1 and SrtC2 of pilus islands 1 and 2b) wild-type and complemented strains with plasmids expressing the backbone proteins of PI-1 (BP-1) and PI-2a (BP-2a), respectively. The nitrocellulose membranes were probed with antisera specific for BP-1 and BP-2a (Cozzi, Prigozhin et al. 2012).

Each class C enzyme, although clearly exhibited redundant functions, predominantly incorporated into pili one of the two ancillary subunits. Taken together, these studies suggest that the promiscuous action shown by class C

sortases on distinct substrates originates from their ability to cleave variable LPXTG-like motifs. On the other hand, the preferential ancillary protein incorporation observed *in vivo*, not apparent in the cleavage reactions with peptides *in vitro*, suggests that the substrate specificity of C sortases may be due to recognition of more extensive structural determinants rather than a few specific residues. Interestingly, the lower enzymatic activity displayed in *in vitro* assays by both SrtC1 and SrtC2 on the AP2 peptide can be explained as the LPXTG-like motif of the minor ancillary protein has been demonstrated to be substrate of SrtA for anchoring the entire polymerized pilus to the cell wall (Necchi, Nardi-Dei et al.; Nobbs, Rosini et al. 2008; Cozzi, Prigozhin et al. 2012). In this context, the specificity of SrtC1 for AP2 observed *in vivo* (Rosini, Rinaudo et al. 2006), but not detectable with the LPXTG-peptides based assay, can be due to the specific joint of the lysine residue in the AP2 pilin motif to the threonine residue within the LPXTG-like motif of the BP. These data suggest that *in vitro* experiments involving only the sortases in combination with LPXTG-like mimicking peptides are likely to be insufficient to define the determinants of sortase C enzyme specificity. There are obviously other factors, *in vivo*, in addition to the LPXTG-like motif, that guide sortase C specific substrate recognition. The crystal structures of the two PI-1 SrtC enzymes suggest that the hydrolysis of different LPXTG-like peptides may be a consequence of the conservation of the residues and the β-sheet fold of the catalytic domain and of the flexibility of the entire N-terminal domain that could allow LPXTG-like peptides to bind productively to the catalytic cleft (Cozzi, Prigozhin et al. 2012).

Multiple sequence alignment of all GBS sortase C enzymes and structural homology modeling, showed that, in contrast with highly similar SrtC enzymes of PI-1 and PI-2a that also contain the lid region, the pilus-associated sortases of PI-2b are shorter. In addition, even if the catalytic triad is conserved, SrtC1 does not contain the conserved motif DPY(F/W) in the lid, and SrtC2 completely lacks the lid region. PI-2b in GBS has a similar genetic organization to group A *Streptococcus* (GAS) FCT-3 pilus, and like GAS, it contains the LepA gene required in GAS for pilus polymerization (Nakata, Koller et al. 2009). Further investigations will be needed for a better understanding of PI-2b class C sortases in pilus biogenesis (Cozzi, Malito et al. 2011).

Aim of the thesis

Gram-positive bacteria build pili on their cell surface *via* a class C sortase-catalyzed transpeptidation mechanism from pilin protein substrates. Pilus-related sortases recognize specific sequence motifs in the pilin subunits, elongate the pilus polymer, and then tether the entire assembled structure to the cell wall peptidoglycan. Despite the availability of several crystal structures, pilus-related C sortases remain poorly characterized to date and their mechanisms of transpeptidation and regulation need to be further investigated. The aim of this thesis work was the understanding of the activity and the regulation of the pilus 2a sortase C1 (SrtC1) in Group B *Sreptococcus* (GBS).

We investigated the SrtC1 activity and the substrate specificity *in vivo* by complementation of KO GBS strains. Moreover, we tested the enzymatic activity on recombinant proteins, setting up the condition for the *in vitro* polymerization assay.

Performing biochemical analysis we further studied the regulation of the SrtC1 providing the first direct evidence of the regulatory role of the so called *lid* region.

Chapter 2. Results

2.1 Recombinant S. agalactiae SrtC1 of PI-2a production

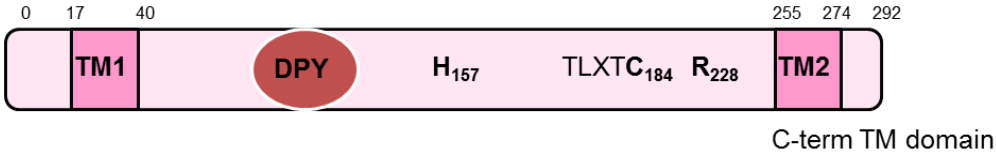
Sequence analysis of GBS SrtC1, identified four distinct regions features: a signal peptide at its N terminus followed by a predicted TM domain, the catalytic domain, and finally, an additional C terminal transmembrane domain followed by a positive tail at its C-terminus.

Based on previous studies (Cozzi, Malito et al. 2011) a recombinant SrtC1 ($SrtC1_{43–292}$) lacking the signal peptide and the N-terminal TM domain, was expressed and purified for the following experiments.

$SrtC1_{43–292}$ containing the C-terminal TM domain, was expressed as an N-terminal 6xHis-MBP TEV cleavable fusion protein (Fig.19A), to increase protein expression and solubility.

A highly purified form of the sortase enzyme was obtained from the total soluble extract of recombinant E. coli cells after two chromatographic steps, involving Ni(II)-chelate affinity chromatography (IMAC), TEV cleavage to remove the 6XHis-MBP-tag a second "subtractive" Ni(II)-chelate affinity chromatography plus an MBP-trap column and size-exclusion chromatography (Fig.19B). The recombinant enzyme was purified in the presence of reducing agent (1 mM DTT) to avoid enzyme dimerization. The mutants of the lid region of recombinant $SrtC1_{43–292}$, $SrtC1_{\Delta NT}$, $SrtC1_{Y86A}$, were generated by PIPE (Polymerase Incomplete Primer Extension) method.

A



B

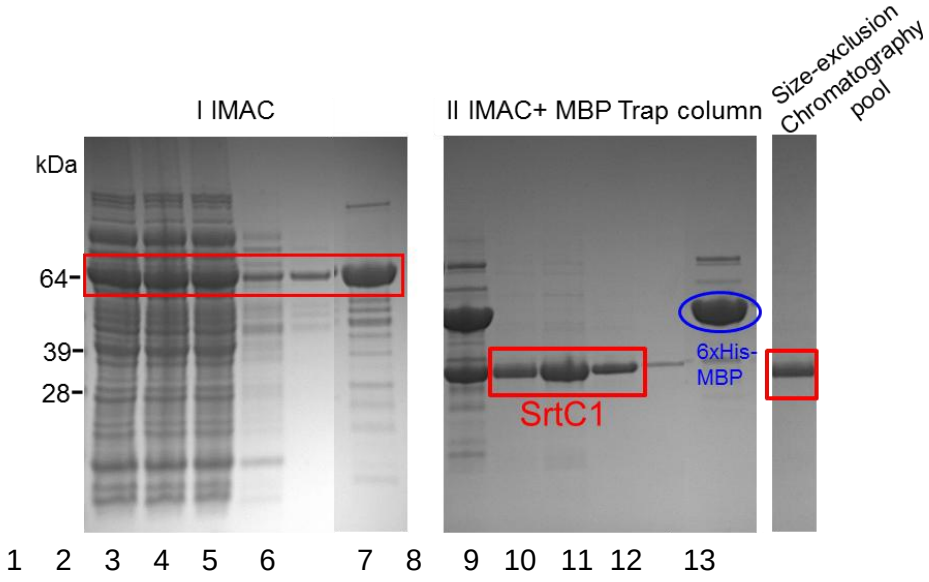


Figure 19. (A) Schematic representation of SrtC1 domains. Four distinct regions can be identified: 1 a signal peptide at its N terminus followed by 2 a predicted TM domain, a catalytic domain of sortase family made of His197, Cys184, Arg228, 3 an additional C terminal transmembrane domain followed by 4 a positive tail at its C-terminus.
(B) SDS-PAGE analysis of the purification of SrtC1$_{43-292}$. lane 1, cell lysate clarified by centrifugation; lane 2, flow through; lane 3-5, wash; lane 6, recombinant protein obtained from the first Ni(II)-chelate affinity chromatography; lane 7, TEV cleavage mixture reload in the second Ni(II)-chelate affinity chromatography plus an MBP-Trap column; lane 8-11, unbound protein after HIS tag cleavage collected and injected in the size-exclusion column; lane 12, elution of the cleaved 6xHis-MBP tag and HIS-tagged TEV protease; lane 13, size-exclusion chromatography pool.

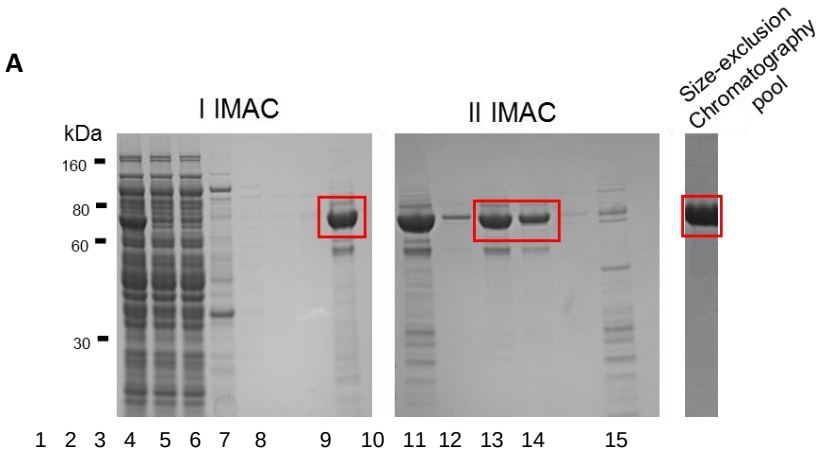
2.2 Recombinant *S. agalactiae* and *S.pneumoniae* backbone proteins production

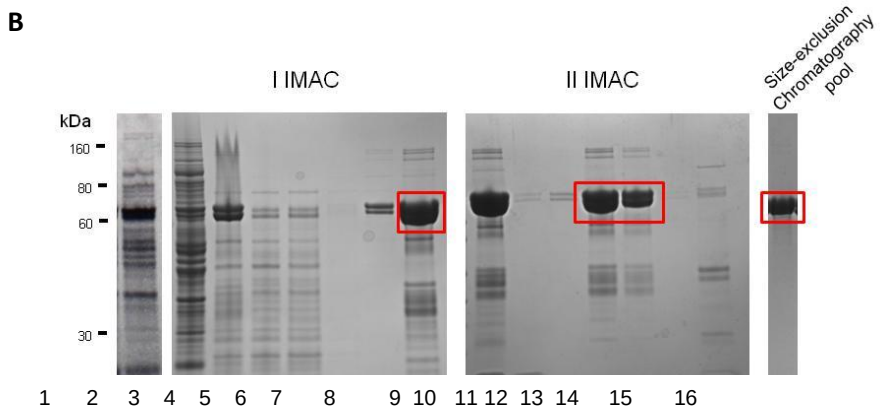
The backbone subunit of type 2a pilus (BP-2a) is present in six main immunogenically different but structurally similar variants (Margarit, Rinaudo et al. 2009). In this work the GBS PI-2a (BP_{30-649}) variants 515, H36B, CJB111 were cloned,expressed and purified containing both the pilin motif and the sorting signal.

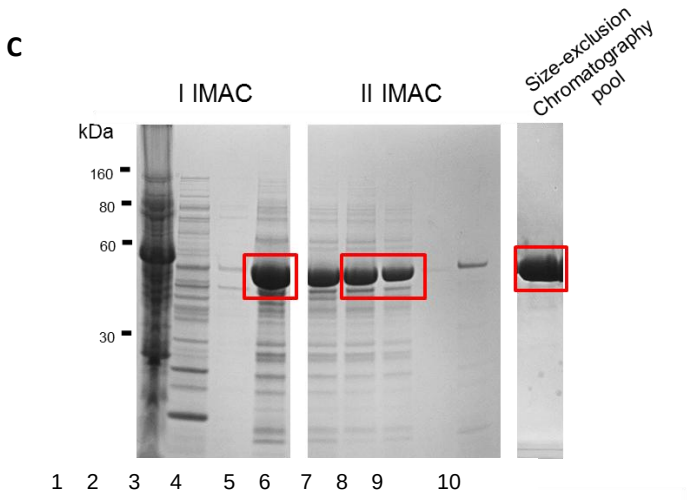
The mutants BP_{K189A}, lacking the Lys residue of the pilin motif, and the BP_{30-640}, lacking the C-terminal LPxTG motif, of variant 515 were generated by PIPE site-directed mutagenesis using wild type BP_{30-649} as template. These recombinant proteins were expressed and purified as N terminal His-tag, TEV cleavable, fusion protein.

Genes coding for pilus 1 and pilus 2b backbone protein, BP-1 and BP-2b (TIGR annotation SAG_0645 and SAN_1518), were PCR amplified from respectively GBS strains 2603V/R and COH1. The gene coding for RrgB (TIGR annotation SP_0463), *S. pneumoniae* backbone protein protein, was PCR amplified from the pneumococcal strain TIGR4.

A highly purified form of all these proteins was obtained from the total soluble extract of recombinant E. coli cells after two chromatographic steps, involving Ni(II)-chelate affinity chromatography (IMAC), TEV cleavage for the 6XHis-tag a second “subtractive” Ni(II)-chelate affinity chromatography and size-exclusion chromatography (Fig.20 A, B, C, D).


A
I IMAC
II IMAC
Size-exclusion Chromatography pool
kDa
160
80
60
30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15


B
I IMAC
II IMAC
Size-exclusion Chromatography pool
kDa
160
80
60
30
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16


C
I IMAC
II IMAC
Size-exclusion Chromatography pool
kDa
160
80
60
30
1 2 3 4 5 6 7 8 9 10

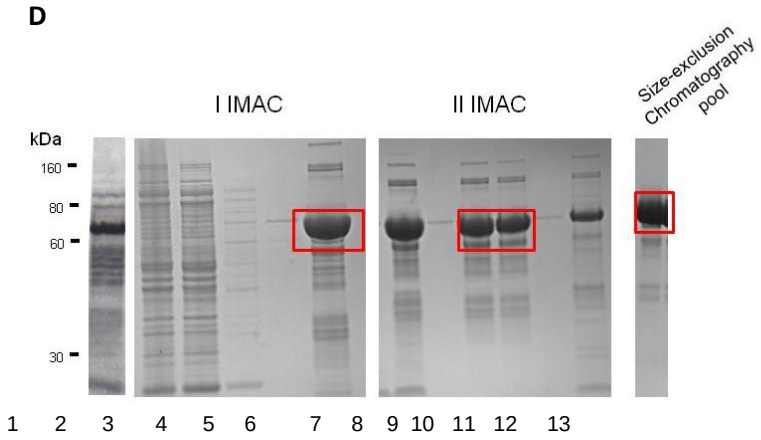


Figure 20. (A) SDS-PAGE analysis of the purification of BP-2a. lane 1, cell lysate clarified by centrifugation; lane 2, flow through; lane 3-7, wash; lane 8, recombinant protein obtained from the first Ni(II)-chelate affinity chromatography; lane 9, TEV cleavage mixture reload in the second Ni(II)-chelate affinity chromatography; lane 7-10-13, unbound protein after HIS tag cleavage collected and injected in the size-exclusion column; lane 14, elution of the cleaved 6xHistag and HIS-tagged TEV protease; lane 15, size-exclusion chromatography pool.
(B) SDS-PAGE analysis of the purification of BP-1. Lane 1, cell lysate clarified by centrifugation; lane 2, flow through; lane 3-6, wash; lane 7-8, recombinant protein obtained from the first Ni(II)-chelate affinity chromatography; lane 9, TEV cleavage mixture reload in the second Ni(II)-chelate affinity chromatography; lane 10-14, unbound protein after HIS tag cleavage collected and injected in the size-exclusion column; lane 15, elution of the cleaved 6xHis tag and HIS-tagged TEV protease; lane 16, size-exclusion chromatography pool.
(C) SDS-PAGE analysis of the purification of BP-2b. Lane 1, cell lysate clarified by centrifugation; lane 2, flow through; lane 3, wash; lane 4, recombinant protein obtained from the first Ni(II)-chelate affinity chromatography; lane 5, TEV cleavage mixture reload in the second Ni(II)-chelate affinity chromatography; lane 6-8, unbound protein after HIS tag cleavage collected and injected in the size-exclusion column; lane 9, elution of the cleaved 6xHis tag and HIS-tagged TEV protease; lane 10, size-exclusion chromatography pool.
(D) SDS-PAGE analysis of the purification of *S.pneumoniae* RrgB. Lane 1, cell lysate clarified by centrifugation; lane 2, flow through; lane 3-5, wash; lane 6, recombinant protein obtained from the first Ni(II)-chelate affinity chromatography; lane 7, TEV cleavage mixture reload in the second Ni(II)-chelate affinity chromatography; lane 8-11 unbound protein after HIS tag cleavage collected and injected in the size-exclusion column; lane 12, elution of the cleaved 6xHis tag and HIS-tagged TEV protease; lane 13, size-exclusion chromatography pool.

2.3 Wild-type SrtC1 is not able to induce recombinant BP polymerization *in vitro*

The crystal structure of SrtC1 from GBS pilus 2a (SrtC1, PDB ID 3O0P) shows that the catalytic cysteine (Cys219) is locked by the aromatic ring of Tyr86 present in the lid, suggesting that this interaction maintains the enzyme in an inactive conformation (Cozzi, Malito et al. 2011).

To investigate the hypothesis that the presence of the lid covering the active site inhibits the catalytic activity of the enzyme, we tested the ability of the wild type SrtC1 ($SrtC1_{WT}$) to polymerize the recombinant full-length PI-2a backbone protein (BP-2a). It is well-known that at least two components are necessary and sufficient for pili polymerization *in vivo*, the backbone protein (BP), forming the pilus shaft, and at least one sortase C coded by the same genomic pilus locus (Rosini, Rinaudo et al. 2006). The presence of covalently linked pili on the GBS surface can be detected by SDS-PAGE immunoblot analysis of cell-wall preparations through the identification of a ladder of high-molecular-weight (HMW) bands (Rosini, Rinaudo et al. 2006). Thus, we set up *in vitro* reactions in which different concentrations (5, 25, 50 and 100 μM) of purified recombinant $SrtC1_{WT}$ were mixed with a fixed concentration (100 μM) of the recombinant BP-2a (BP_{30-649}), and incubated at 37°C for up to 72 hours. The reactions were analyzed by searching for a pattern of HMW bands by SDS-PAGE. In all the conditions used, no HMW bands were observed (Fig.21). However, the presence of some bands at a higher molecular weight than that of the monomeric protein was detected. These bands, analyzed by Mass Spectrometry (MS), corresponded to the formation of a hetero-dimer formed by $SrtC1_{WT}$ and BP-2a and a BP-BP

dimer (data not shown), whose formation was not promoted by the sortase since it was visible also in absence of $SrtC1_{WT}$ (Fig.21 and 22A). Identical results were obtained with different concentrations of BP-2a substrate (25, 50, 100 and 200 µM) incubated with a fixed amount of 25 µM of $SrtC1_{WT}$ (Fig.22B). Moreover, we also tested other reaction conditions, e.g. different pHs (Fig.22C), but we did not observed effects on the enzymatic activity and the $SrtC1_{WT}$ was still unable to polymerize the BP-2a *in vitro*.

These data suggest that the interaction between the wild type form of SrtC1 and its BP-2a substrate is not alone sufficient to induce lid displacement from the catalytic pocket and to promote *in vitro* pilin subunit polymerization (Cozzi, Zerbini et al.).

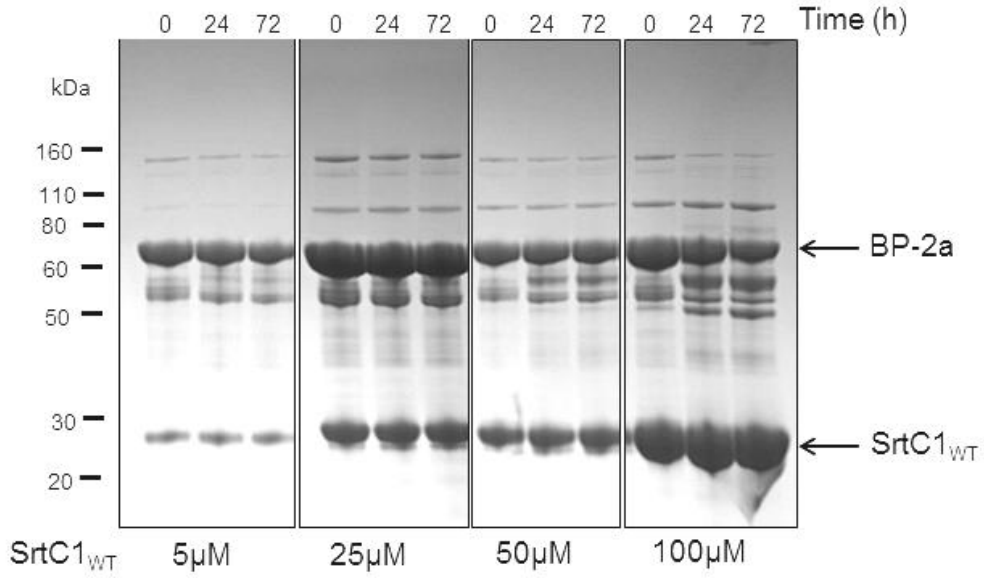
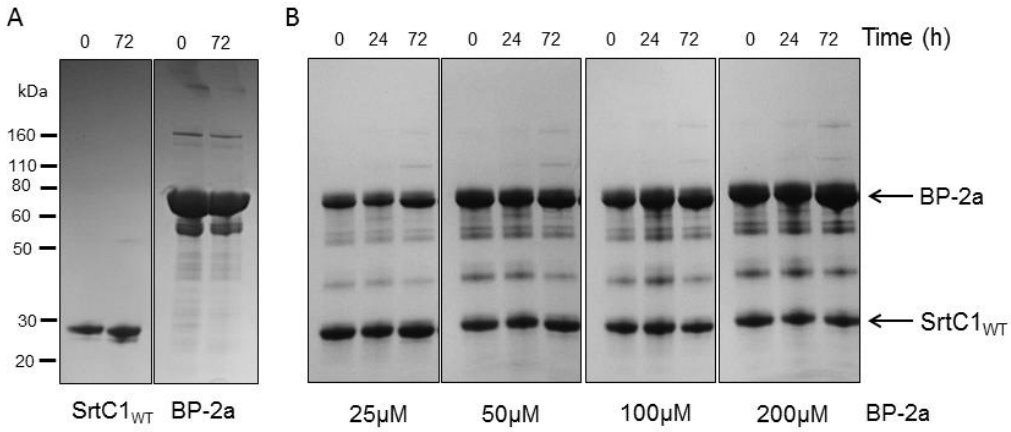


Figure 21. $SrtC1_{WT}$ does not polymerize BP-2a *in vitro*. Time course of reaction between $SrtC1_{WT}$ and BP-2a followed by SDS-PAGE at different times (0-24h-72h).

Different $SrtC1_{WT}$ concentrations (5-25-50-100µM) were mixed with a fixed concentration of recombinant BP-2a (100µM). No high molecular weight structures (HMW) could be identified on SDS-PAGE, in all the conditions tested (Cozzi, Zerbini et al.).



C

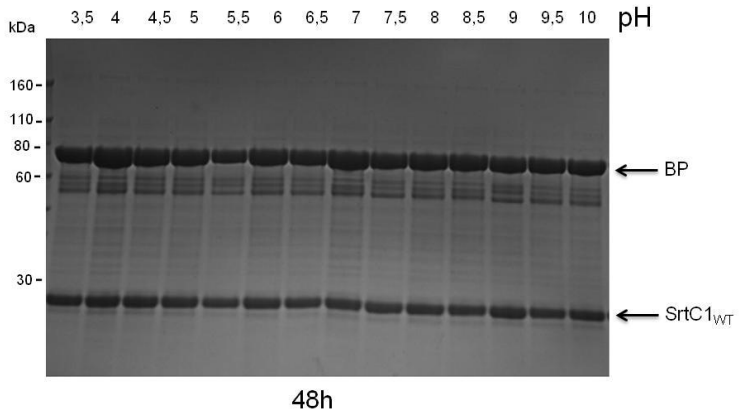
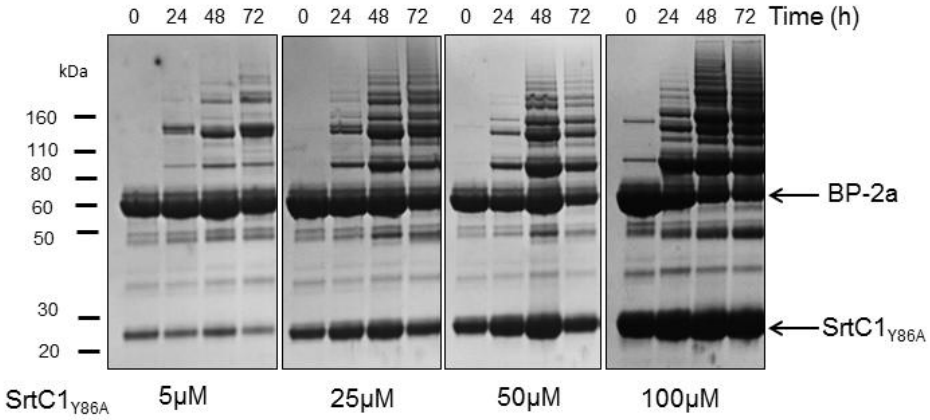


Figure 22. (A) *In vitro* pilus polymerization negative controls. $SrtC1_{WT}$ (25µM) and BP-2a (100µM) were incubated alone in the same reaction conditions (72h at 37°C). No high molecular weight structures (HMW) could be identified on SDS-PAGE. **(B) $SrtC1_{WT}$ does not polymerize BP-2a *in vitro*.** Time course of reaction between $SrtC1_{WT}$ and BP-2a followed by SDS-PAGE at different times (0-24h-72h). Different BP-2a concentrations (25-50-100-200 µM) were mixed with a fixed concentration of $SrtC1_{WT}$ (25 µM). **(C) Different pHs does not have effects on the $SrtC1_{WT}$ activity.** $SrtC1_{WT}$ (25µM) and BP-2a (100µM) were incubated 37°C in buffers with pH ranging from 3,5 to 10 for 48h. No high molecular weight structures (HMW) could be identified on SDS-PAGE, in all the conditions tested (Cozzi, Zerbini et al.).

2.4 BP-2a high molecular weight structures can be assembled *in vitro* by recombinant SrtC1 lid mutant

To investigate whether the mutant enzyme $SrtC1_{Y86A}$ in which the active site is more accessible for accommodating the substrate, could be efficient in polymerizing BP-2a monomers in HMW structures, we performed *in vitro* polymerization experiments in the conditions described above, but using $SrtC1_{Y86A}$. Thus, 100 µM of the purified backbone protein (BP-2a) were incubated at 37°C with the recombinant $SrtC1_{Y86A}$ mutant at different concentrations (5, 25, 50 and 100 µM). Samples from the single reactions were collected after 24, 48 and 72 hours, treated with SDS, in a sample buffer containing also a reducing agent, and analyzed by SDS-PAGE. After Coomassie-staining, the formation of a typical pili pattern of bands with molecular weights above 260 kDa was observed in all reactions performed, providing also a clear evidence of the covalent nature of the polymerized pilus structures (Fig.23A). We observed that also at the highest enzyme concentration some of the BP-2a monomer still remained unprocessed and a complete conversion of monomeric BP-2a in polymeric structures could not be achieved (Fig.23A). Furthermore, the rate of recombinant BP polymers formation did not change also when different concentrations of the BP-2a substrate were used in the reaction, starting from 25 to 200 µM mixed with a fixed concentration (25 µM) of the enzyme (Fig.23B) (Cozzi, Zerbini et al.). However, in the presence of HMW polymeric structures, monomeric forms of the pilin subunits can always be detected also in total proteins prepared from GBS strains (Rosini, Rinaudo et al. 2006).

A



B

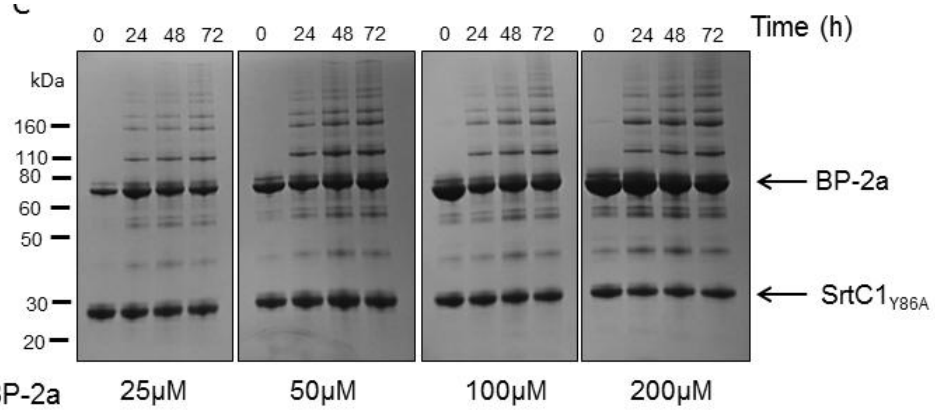


Figure 23. (A) SrtC1$_{Y86A}$ polymerizes BP-2a *in vitro*. Time course of the reaction between SrtC1$_{Y86A}$ and BP-2a followed by SDS-PAGE at different times (0, 24h, 48h and 72h). Different SrtC1$_{Y86A}$ concentrations (5-25-50-100μM) were mixed with a fixed concentration of recombinant BP-2a (100μM). In all the reactions a pattern of high molecular weight structures could be identified. **(B) The SrtC1$_{Y86A}$ can polymerize the BP-2a *in vitro*.** Time course of the reaction between SrtC1$_{Y86A}$ and BP-2a analysed by SDS-PAGE at different time points (0, 24h, 48h and 72h). Different BP-2a concentrations (25-50-100-200μM) were mixed with a fixed SrtC1$_{Y86A}$ concentration (25μM). In all the reactions a pattern of high molecular weight structures could be identified (Cozzi, Zerbini et al.).

2.5 Lysine 189 in the putative pilin motif and the IPQTG sorting signal of BP-2a are essential for pilus formation *in vivo*

To better characterize the HMW pilus-like structures obtained by polymerization of BP-2a monomers mediated by $SrtC1_{Y86A}$ mutant, we investigated the role of specific residues/motives of the BP-2a sequence in pilus 2a polymerization. To identify these sequence elements we first performed sequence comparisons of the homologous pilin subunits in different Gram-positive bacteria. In the BP-2a sequence (strain 515, TIGR annotation SAL_1486) we identified a putative pilin motif containing a highly conserved lysine residue (Lys189), and the IPQTG motif in position 641-645 as the C-terminal sorting motif. To demonstrate the specific contribution in pilus assembly of both the Lys189 in the pilin motif and the IPQTG motif, we used site-specific mutagenesis and complementation studies. The plasmid (pAM-BP_{K189A}) expressing a mutated backbone protein carrying a mutation of the pilin motif lysine residue into alanine and the second plasmid (pAM-$BP_{\Delta IPQTG}$) carrying a deletion of the entire IPQTG sorting signal were used to transform the GBS knock-out (KO) mutant strain lacking the BP-2a gene (515ΔBP-2a) (Rosini, Rinaudo et al. 2006; Cozzi, Nuccitelli et al. 2012). After complementation, the effects of each mutation/deletion on pilus formation were assessed by Western blot analysis, using total proteins extracted from each complemented strain and sera specific for the pilin subunits. As expected, both the K189 residue and the C-terminal IPQTG of BP-2a were absolutely required for pilin protein incorporation into the HMW structures *in vivo* (Fig.24). When the K189 was mutated into an alanine, only the monomer form of the BP-2a could be identified, whereas when the sorting signal IPQTG was deleted in the BP-2a, in

addition to the monomeric form of BP-2a a higher molecular weight band was also observed. Immunoblot analysis performed with antibodies raised against the backbone subunit (α-BP) and the major ancillary pilin (α-AP1) showed that this higher molecular weight band, resistant to SDS treatment, contained both the backbone protein (BP) and the major ancillary protein (AP1). Indeed, the polymerization of the BP-2a cannot occur as its sorting signal is deleted, but the pilin motif of the BP-2a is still available to form a covalent bond between the BP-2a pilin motif and the AP1-2a sorting signal (Fig.24) (Cozzi, Zerbini et al.).

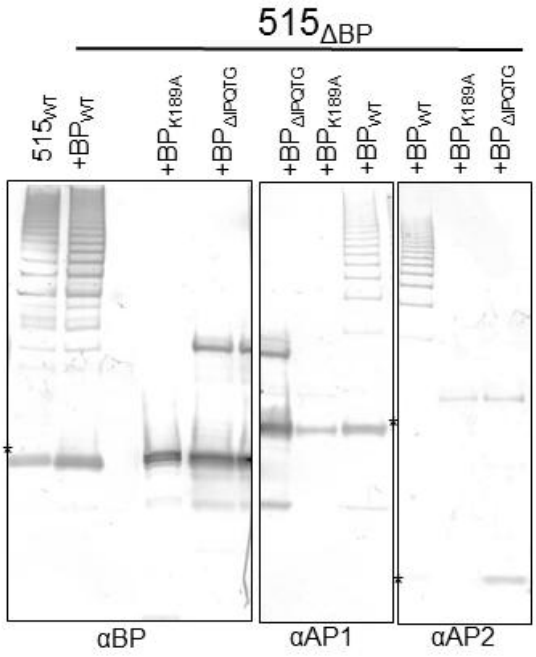


Figure 24. The IPQTG motif and the lysine 189 are essential for BP-2a polymerization and ancillary proteins incorporation. Immunoblots of total protein extracts from 515 strain wild type and mutant strain of BP complemented by plasmids expressing BP wild-type or BP carrying the mutation K189A (BP_{K189A}) or the deletion of LPXTG motif ($BP_{\Delta IPQTG}$). Nitrocellulose membranes were probed with antisera specific for the BP and the ancillary proteins (Cozzi, Zerbini et al.).

2.6 The IPQTG sorting signal is essential for the transpeptidation reaction mediated *in vitro* by the $SrtC1_{Y86A}$ mutant

To investigate the specific contribution of the Lys189 in the pilin motif and the IPQTG sorting signal in the *in vitro* polymerization reaction, we expressed in *E. coli* and purified mutated forms of the BP-2a protein, $BP_{\Delta IPQTG}$ and BP_{K189A}, carrying the deletion of the IPQTG region and the substitution of the Lys189 with an alanine, respectively. After mixing the active $SrtC1_{Y86A}$ with the recombinant $BP_{\Delta IPQTG}$ mutant, HMW polymers could not be detected, confirming that the polymerization reaction occurs through the cleavage of the sorting signal and the formation of the acyl-intermediate between $SrtC1_{Y86A}$ and the IPQTG motif

(Fig.25A). On the contrary, in the reactions in which the active $SrtC1_{Y86A}$ was incubated with BP_{K189A} HMW polymers could be observed, indicating that the Lys residue of the pilin motif (K189), differently from what happens *in vivo*, is not essential for *in vitro* polymerization (Fig.25B). Moreover, when $SrtC1_{Y86A}$ was mixed with recombinant forms of the ancillary proteins (AP1-2a and AP2-2a), that *in vivo* can be polymerized only in the presence of the BP-2a protein (Cozzi, Malito et al. 2011), some polymers were formed (Fig.25C). These data suggest that to perform protein polymerization *in vitro,* $SrtC1_{Y86A}$ can use different nucleophile/s to resolve the acyl-intermediate between the enzyme and the LPXTG-like sorting signal (Cozzi, Zerbini et al.).

To further confirm this hypothesis some bands at high molecular weight, corresponding to polymers of BP-2a proteins wild type and BP_{K189A} were excised from the gel and were subjected to trypsin digestion for mass-spectrometry-based analysis. In all the bands analyzed we found either the tryptic peptide KVTIPQTGEEAK or VTIPQTGEEAK which is consistent with a C-term (VTIPQT) to N-term (GEEAK) bond. This result showed that the nucleophile of the polymerization reaction *in vitro* is the amine group of the BP N-term glycine, both in the wild type and in the $BP_{K189A.}$

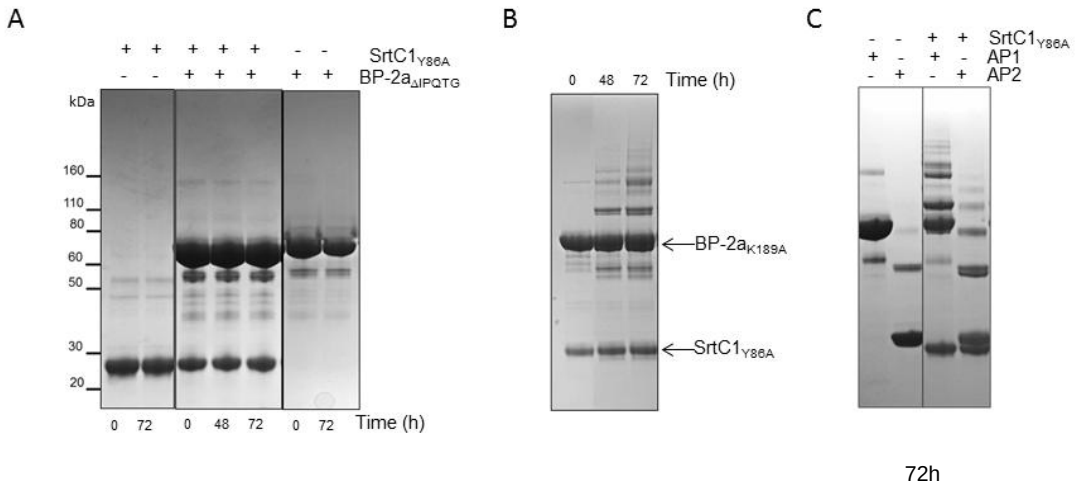


Figure 25. (A) The IPQTG motif is essential for *in vitro* pilus polymerization. SDS-PAGE of the reaction between the $SrtC1_{Y86A}$ and recombinant $BP_{\Delta IPQTG}$ at different time points (0, 48h and 72h) at 37°C. The concentrations of both $SrtC1_{Y86A}$ and $BP_{\Delta IPQTG}$ were fixed at 25µM and 100µM, respectively. No formation of high molecular weight pattern could be identified, showing that the LPXTG like-motif is necessary for the BP polymerization. As controls the $SrtC1_{Y86A}$ (on the left) and BP-$2a_{\Delta IPQTG}$ (on the right) were incubated alone in the same conditions. **(B) The lysine of pilin motif is not essential for *in vitro* pilus polymerization.** The $SrtC1_{Y86A}$ (25µM) and the recombinant BP_{K189A} (100µM) were mixed at 37°C and at different time points (0, 48h and 72h) the reactions were analysed by SDS-PAGE. A weaker patter of high molecular weight structures could be identified. **(C)** The $SrtC1_{Y86A}$ can polymerize the ancillary proteins, harboring the LPXTG like motif. $SrtC1_{Y86A}$ was incubated with recombinant AP1 or AP2 and after 72h a pattern of HMW structure was visible on SDS PAGE. Recombinant AP1 and AP2 were incubated alone in the same conditions as controls (Cozzi, Zerbini et al.).

2.7 The $SrtC1_{Y86A}$ active mutant is able to polymerize *in vitro* backbone proteins of other GBS pili and/or pathogens

The polymerizing activity of $SrtC1_{Y86A}$ was also tested *in vitro* using backbone proteins of different GBS pilus 2a variants and also the backbone proteins of pilus 1 and 2b. In particular, we selected the BP variants from strains H36B and CJB111, which are the most evolutionary distant and closest variants, among the other five variants of the 515 allele, respectively (Margarit, Rinaudo et al. 2009).

In detail, 100µM of each BP-2a protein was mixed with a fixed concentration (25 µM) of SrtC1$_{Y86A}$ and after 72 hours the reaction was analyzed by SDS-PAGE. The formation of a typical pili pattern of bands with molecular weights above 260 kDa was observed in all reactions performed using the different GBS BP-2a variants (Fig.26A and B).

We also further investigated if the active SrtC1$_{Y86A}$ could polymerize the BP protein of *S.pneumoniae*, RrgB, which has a percentage of similarity with BP-2a 515 of 55%. Therefore, we performed the reaction *in vitro* for 72h with 100µM of recombinant RrgB and 25µM of SrtC1$_{Y86A}$ and the coomassie staining of the reaction loaded into the SDS-PAGE showed a typical pattern of HMW structures, indicative that the SrtC1$_{Y86A}$ is also able to recognize and polymerize backbone proteins of other pathogens (Fig.26B)

All the backbone proteins, with the corresponding LPXTG motif, tested in the polymerization reaction, tested in this study, are listed in table 1.

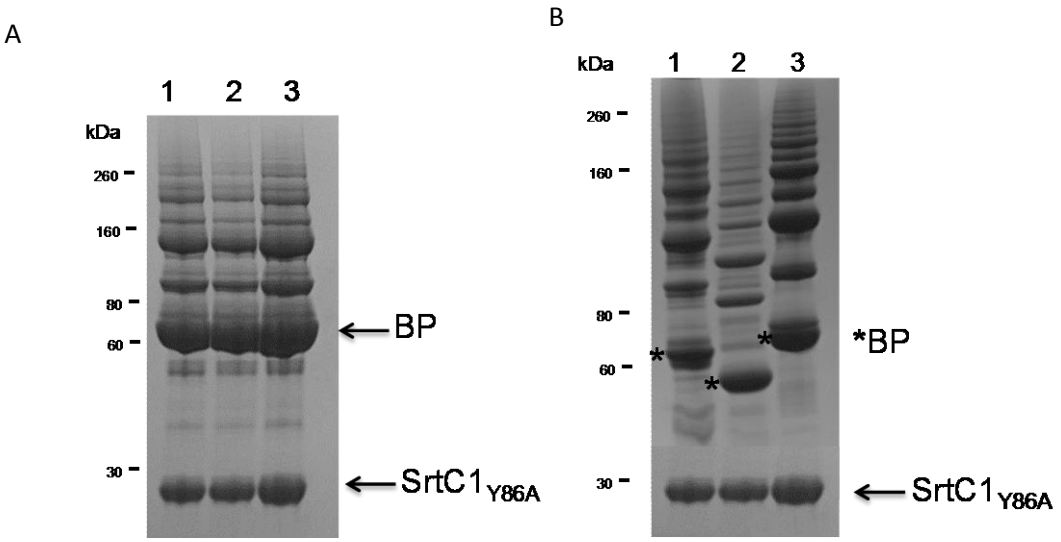


Figure 26. (A) SrtC1$_{Y86A}$ polymerizes different variants of BP-2a *in vitro*. SDS-PAGE after 72h of the reaction between SrtC1$_{Y86A}$ and BP-2a, 515 variant lane 1, H36B lane 2

and CJB111 lane 3. 25µM of $SrtC1_{Y86A}$ were mixed with 100µM of each BP-2a variant. In all the reactions a pattern of high molecular weight structures could be identified. **(B) $SrtC1_{Y86A}$ polymerizes *in vitro* BP-1, BP-2b and *S.pneumoniae* RrgB.** In the reaction *in vitro* were mixed 25µM of $SrtC1_{Y86A}$ with 100µM of each BP proteins. The SDS-PAGE of all the reactions showed a pattern of high molecular weight structures. Lane 1 reaction with BP-1; lane 2 reaction with RrgB; lane 3 reaction with BP-2b.

PROTEIN	LPXTG motif
GBS BP-2a	IPQTG
GBS AP2-2a	LPKTG
GBS AP1-2a	IPKTG
GBS BP-1	IPNTG
GBS BP-2b	LPSTG
S. Pneumoniae RrgB	IPQTG

Table 1. $SrtC1_{Y86A}$ polymerizes *in vitro* BP proteins with different LPXTG motifs.

2.8 The GFP protein containing a C-term LPXTG-motif is polymerized *in vitro* by $SrtC1_{Y86A}$

The data showing that the $SrtC1_{Y86A}$ can be promiscuous in substrate recognition and cleavage, prompted the usage of the active $SrtC1_{Y86A}$ to polymerize non-pilus correlated proteins after the introduction of an N terminal C-terminal LPXTG motif. To this end, the Green Flourescence Protein (GFP) was used as a template non pilus-related protein. The $GFP\text{-}_{IPQTG}$ protein, containing the C-terminal IPQTG motif, (the LPXTG motif of GBS BP-2a. the SrtC1 natural substrate) was expressed and purified in E. coli. Recombinant $GFP\text{-}_{IPQTG}$ was mixed with $SrtC1_{Y86A}$ active mutant and after 72 hours of incubation at 37°C we analyzed the reaction on SDS-PAGE. As showed in fig.27 the $SrtC1_{Y86A}$ is able to polymerize

in vitro the GFP-IPQTG, thus demonstrating that adding the LPXTG motif at the C-term is sufficient to induce the polymerization activity of $SrtC1_{Y86A.}$

In order to investigate the nature of GFP-GFP cross-links, we performed mass spectrometry analysis and we identified a tryptic peptide (IPQTGMVSK) which is consistent with a C-term (IPQT) to N-term (GMVSK) bond

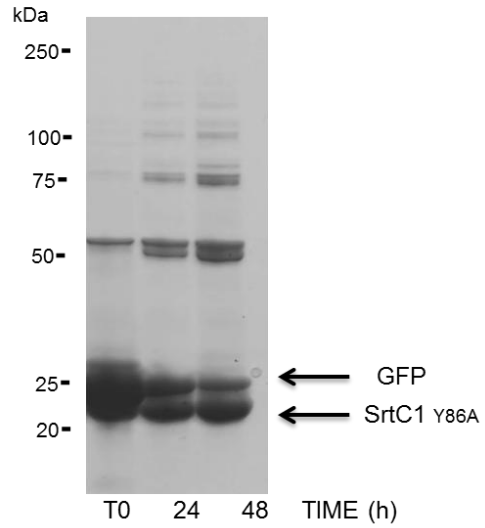


Figure 27. $SrtC1_{Y86A}$ polymerizes the GFP-IPQTG *in vitro*. Time course of the reaction between $SrtC1_{Y86A}$ (25µM) and GFP-IPQTG (100µM). The formation of oligomers could be identified after coomassie staining.

2.9 Biochemical characterization of SrtC1-2a wild type and active mutant reveals that the lid is involved in protein stability

To explore the effects of the lid anchoring residue Tyr86 and the entire lid region on the stabilization of SrtC1-2a enzyme, recombinant forms of $SrtC1_{WT}$, $SrtC1_{Y86A}$ and $SrtC1_{\Delta LID}$ were studied by differential scanning fluorimetry (DSF). DSF is a technique used to monitor the thermal unfolding of proteins in the

presence of a fluorescent dye (Sypro® Orange) that displays weak fluorescence in hydrophilic environments and it becomes highly fluorescent in a hydrophobic environment such as that provided by the exposed side chains of hydrophobic residues of an unfolded protein (Ericsson, Hallberg et al. 2006). The thermal denaturation curves, shown in figure 28A, indicate that the melting temperature, T_m, for $SrtC1_{Y86A}$ is reduced by almost 10°C compared to that of the wild-type enzyme (52°C versus 61°C). $SrtC1_{\Delta LID}$ shows an even lower T_m (47°C) and high pre-transition baseline fluorescence suggestive of a partial exposition of the sortase hydrophobic core or partial aggregation.

The structural integrity of the produced mutants was verified by 2D-NMR spectroscopy. ^{1}H-^{15}N HSQC spectra display signals for all backbone HN atoms in a two-dimensional array defined by a proton and a nitrogen frequency dimension. Due to high sensitivity of peak positions to the proximal chemical environment of the corresponding atoms, a well-dispersed spectrum is characteristic of a folded protein. The ^{1}H-^{15}N HSQC peaks of both the $SrtC1_{WT}$ and $SrtC1_{Y86A}$ are distributed over a large spectral range (Fig.28B, ~5 ppm, ^{1}H, and ~30 ppm, ^{15}N) confirming the expected globular fold of the soluble domain, also in the presence of an extended trans-membrane stretch at the C-terminus. It can be observed that many of the isolated peaks have unchanged positions for $SrtC1_{Y86A}$ compared to $SrtC1_{WT}$, however major perturbations in some areas of the spectrum are indicative of a non-negligible structural rearrangement, most likely affecting the lid region. Minor protein aggregation is suggested by the broader signals of the mutant particularly in the centre of the spectrum. The latter phenomenon was also analysed and confirmed by measurement of the average transverse proton spin

relaxation time (T_2), a sensitive indicator of the overall tumbling rate of the molecule. The T_2 value for $SrtC1_{Y86A}$ was determined as slightly shorter compared to the wild type protein. While the conformational transition induced by the Y86A mutation apparently preserves the overall enzyme fold, producing only localized conformational changes, the ^{1}H-^{1}D and the ^{1}H-^{15}N HSQC spectrum (Fig.29) of the recombinant $SrtC1_{\Delta LID}$ mutant show largely broadened peaks with reduced frequency dispersion, indicative of protein aggregation probably due to the exposition of the hydrophobic core covered by the lid in $SrtC1_{WT}$ (Cozzi, Zerbini et al.).

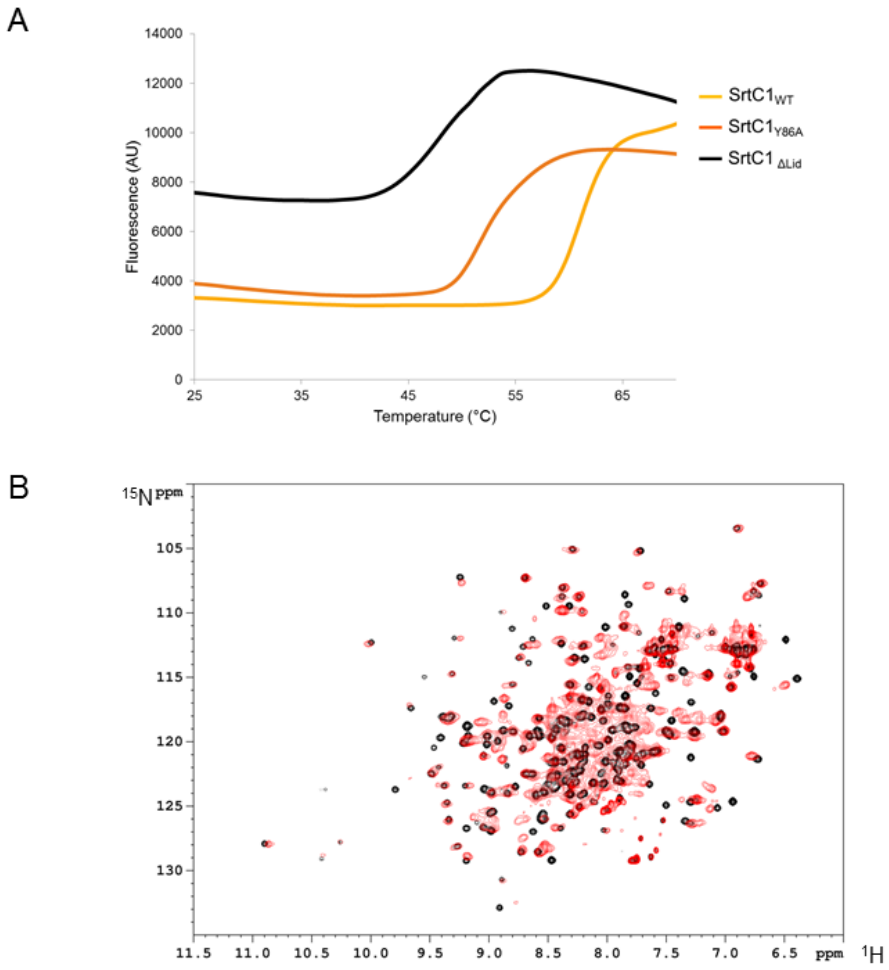


Figure 28. DSF and NMR analysis of SrtC1$_{WT}$ and mutants. DSF analysis of SrtC1$_{WT}$, SrtC1$_{Y86A}$, SrtC1$_{\Delta LID}$ in presence of SYPRO orange showed different thermal stabilities. Fluorescence intensity versus temperature for the unfolding of SrtC1$_{WT}$, SrtC1$_{Y86A}$, SrtC1$_{\Delta LID}$ in 25mM Tris-HCl,100mM NaCl pH=7.5. The melting temperature of SrtC1$_{WT}$ (orange line), SrtC1 $_{Y86A}$(red line), SrtC1 $_{\Delta LID}$ (black line) resulted 61°C, 52°C, 47°C respectively. (B) $^{1}H,^{15}N$-HSQC spectra superimposition for SrtC1$_{Y86A}$ (red) and SrtC1$_{WT}$ (black) enzymes. The chemical shifts of the residues in both proteins produced a similar pattern, indicating that both the proteins are folded (Cozzi, Zerbini et al.).

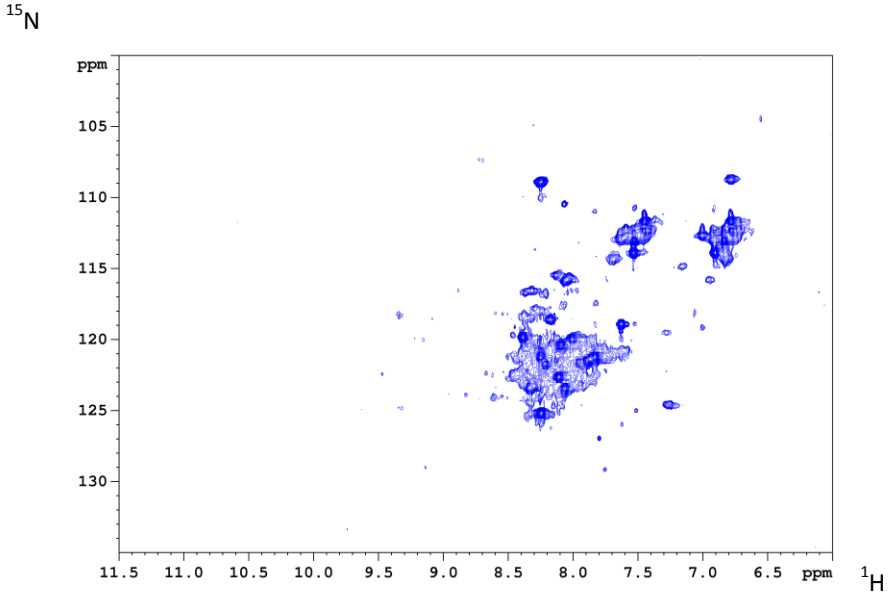


Figure 29. NMR ^{1}H,^{15}N-HSQC spectrum of [^{15}N]-SrtC1$_{\Delta LID}$ (blue). ^{1}H and ^{15}N are on the x and y axes, respectively. The protein sample is in 50 mM phosphate buffer, pH 6.5 and the spectrum was acquired at 25°C (Cozzi, Zerbini et al.).

2.10 Lid anchoring to the active site leads to an overall protection of SrtC1 from proteolysis

To investigate the folding differences between the recombinant non-polymerizing SrtC1$_{WT}$ and the polymerizing SrtC1$_{Y86A}$ mutant in solution, we probed the sensitivity of SrtC1$_{WT}$ and SrtC1$_{Y86A}$ proteins to limited proteolysis. Trypsin-SrtC1digestion was tested at different time points, the reaction was quenched by adding 0.1% formic acid and analysed by SDS-PAGE. The observed digestion patterns of SrtC1$_{WT}$ and SrtC1$_{Y86A}$ were different, indicating that the two enzymes have different susceptibility to proteolysis (Fig.30A). SDS-PAGE analysis of the proteolytic products indicated that SrtC1$_{WT}$ presented two predominant bands of digestion, approximately at 17 and 6 kDa, respectively, still detectable after 60 minutes (Fig.30A). On the contrary, the mutant SrtC1$_{Y86A}$ after only 10 min of

trypsin digestion generated a series of fragments, including a fragment at around 17kDa that, however, was completely degraded after 20 minutes (Fig.30A). Intact mass measurement by ESI-Q-TOF of the 5 min digestion mixture with $SrtC1_{WT}$ revealed that the most intense product of digestion contained two polipeptidic species, the most abundant of 16986.45±0.54Da (91-245 AA) and a less abundant of 16616.88±0.47Da (94-245), instead the lowest molecular weight proteolysis product showed a mass of 6082.56±0.54 (1-52), corresponding to the N-terminus of the recombinant wild type SrtC1 (Fig.30B). The residues of undigested SrtC1 were not detected in ESI spectrum. The primary sequence obtained from intact measurement was also confirmed by Peptide Mass Fingerprint (PMF) of SDS-PAGE bands (data not shown). Taken together, MS data showed that the trypsin cleavage site occurred around at the position 90 generating a fragment that is the N-terminal region of the sortase, including the lid residues, and an equimolar fragment that is the sortase β-barrel catalytic core (Fig.30C). Therefore, the susceptibility to proteolytic attack of the catalytic β-barrel core was greatly increased in the mutant enzyme where Tyr86 was replaced with alanine, suggesting a role of the lid in providing enzyme stability and proteolysis resistance.

To better investigate if the major proteolytic resistance of $SrtC1_{WT}$ compared to $SrtC1_{Y86A}$ lid mutant was due to the interaction between Tyr86 and the catalytic C219 in the active site, interaction that persists after the cleavage of the N-terminal region, we analysed by analytic size-exclusion chromatography the $SrtC1_{WT}$-trypsin reaction mixture after 5 min of digestion (Fig.31A). The fractions corresponding to major peaks were collected and analysed by SDS-PAGE

(Fig.31B). The fractions 15 and 16 (predominant peak) contained both undigested $SrtC1_{WT}$ and two new bands, the proteolytic resistant fragment (17kDa) and the N-terminal fragment (6kDa). This result suggests that the lid still interacts with the catalytic cleft, contributing to the further resistance of the sortase core domain to digestion.

By taking together the NMR, DSF, and proteolysis assays, it can be concluded that lid mutants, even when preserving a globular fold, sample 'open' conformational states at lower energy compared to the wild type protein, suggesting that the Tyr86-Cys219 interaction plays an important role in the thermodynamic and structural stabilization of SrtC1 (Cozzi, Zerbini et al.).

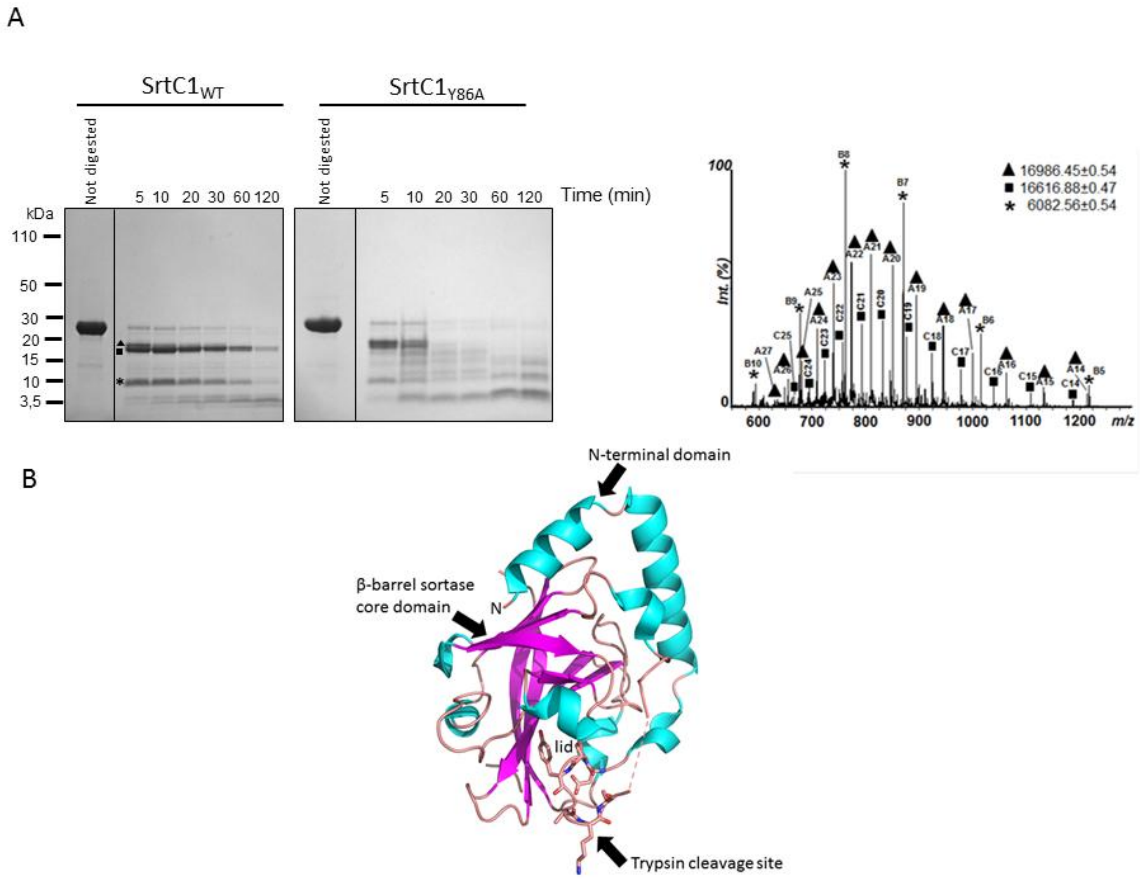


Figure 30. Lid protects SrtC1 from proteolysis. (**A**, *left panel*) Time course of trypsin-SrtC1$_{WT}$ and trypsin-SrtC$_{Y86A}$ proteolysis reactions at 37°C analyzed by SDS-PAGE. The generation of the 6 and 17 kDa fragments, (asterisks, triangle and square, respectively) due to the cleavage in the loop separating the N-terminal region and the sortase β-barrel core, is avoided by the presence of the lid in SrtC1$_{WT}$. (A, *right panel*) Intact mass measurement by ESI-Q-TOF of the trypsin-SrtC1$_{WT}$ proteolysis after 5 min of reaction: all the detected polypeptides and the corresponding charge states are reported **(B)** Structure of GBS SrtC1$_{WT}$ (Protein Data Bank accession 3O0P) is shown, with the N terminal region, the lid loop followed by the trypsin cleavage site R90 in sticks and the β-barrel core indicated (Cozzi, Zerbini et al.).

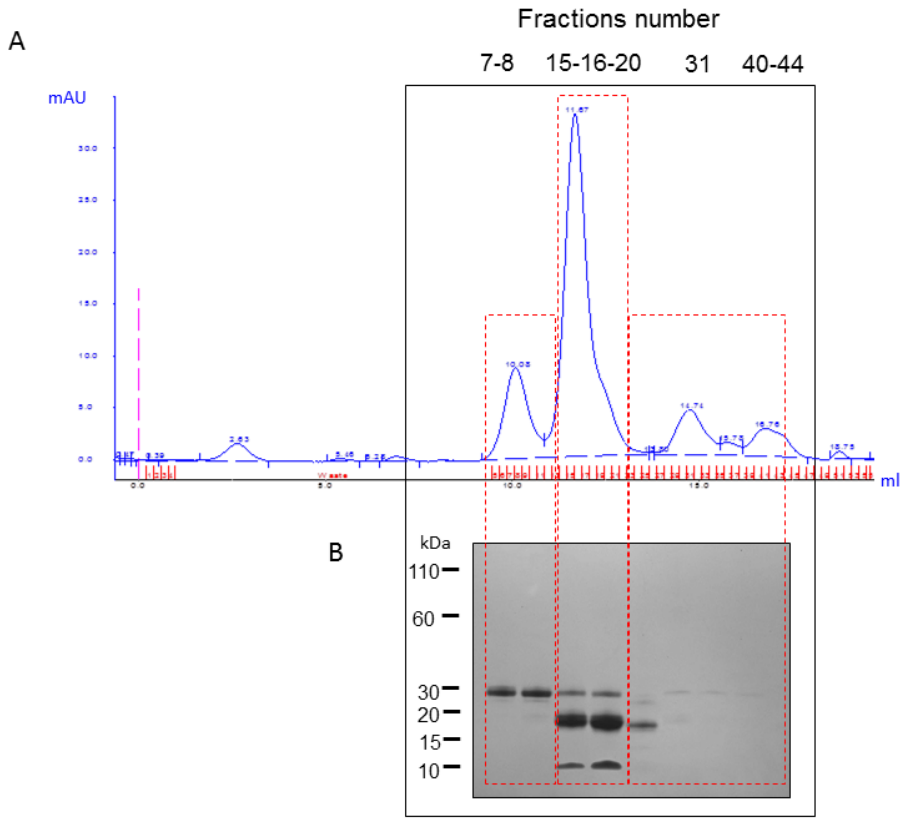


Figure 31. Interaction of the N-terminal domain with the sortase β-barrel. **(A)** SEC profile of the mixture trypsin-SrtC1$_{WT.}$ The digestion mixture trypsin–SrtC1$_{WT}$ (ratio 1:100) in buffer 25mM Tris-HCl-150mM NaCl was loaded in a superdex75 10/300 column. The fractions corresponding to the major packs were collected and analyzed by SDS-PAGE. **(B)** SDS-PAGE of the fractions collected from the size exclusion chromatography (7- 8- 15- 16- 20- 31- 40- 44) showing the co-elution of the N-terminal region and the sortase core of SrtC1$_{WT}$ (Cozzi, Zerbini et al.).

2.11 SrtC1 enzyme deleted of the entire N-terminal region is active in polymerizing BP *in vitro*

To investigate if the sortase resistant core identified by MS is the minimal catalytic domain sufficient for sortase activity, a truncated construct, SrtC1$_{\Delta NT}$, containing only the core domain and deleted of the N-terminal region including the lid loop, was designed. Recombinant SrtC1$_{\Delta NT}$ was recovered with a lower yield than the wild type enzyme; however, a sufficient amount was tested in the *in*

vitro polymerization assay in comparison with $SrtC1_{WT}$ and $SrtC1_{Y86A}$. SDS-PAGE analysis of the reactions showed that $SrtC1_{\Delta NT}$ was able to polymerize the BP-2a, with a pattern of HMW structures similar to that obtained with the single lid mutant $SrtC1_{Y86A}$ (Fig.32). This data suggest that the N-terminal domain is not essential for BP-2a polymerization *in vitro*, further supporting its regulatory function (Cozzi, Zerbini et al.).

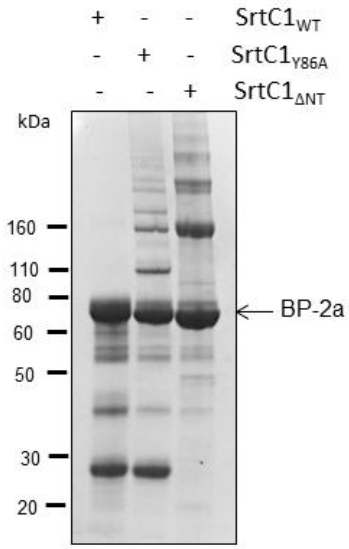


Figure 32. N-terminal truncated SrtC1 polymerizes the BP-2a *in vitro*. SDS-PAGE analysis of the reaction between BP-2a and $SrtC1_{WT}$, $SrtC1_{Y86A}$ and $SrtC1_{\Delta NT}$ after 72h. BP-2a and SrtC1 enzymes concentrations are fixed at 100µM and 25µM respectively. A pattern of high molecular weight structures could be identified in presence of $SrtC1_{Y86A}$ and $SrtC1_{\Delta NT}$ (Cozzi, Zerbini et al.).

Chapter 3. Discussion

Since long time sortases have attracted great interest as potential drug targets as they are responsible for covalently anchoring to the cell wall envelope of a variety of surface proteins that are essential virulence factors. Particularly, the housekeeping *S. aureus* sortase A is a powerful tool for several purposes in protein engineering, as obtaining fluorescent labeled proteins, circularized proteins and covalently linked proteins (Popp and Ploegh). Pilin-specific class C sortases that build structurally complex pili on the surface of Gram-positive bacteria are poorly characterized and much more needs to be investigated about their catalytic activity, specificity and regulation. Nevertheless several crystal structures from different microorganisms have been published in the last years (Manzano, Contreras-Martel et al. 2008; Manzano, Izore et al. 2009; Neiers, Madhurantakam et al. 2009; Cozzi, Malito et al. 2011; Khare, Fu et al. 2011; Khare, Krishnan et al. 2011; Lu, Qi et al. 2011; Persson 2011; Cozzi, Prigozhin et al. 2012). These structures revealed that these enzymes have a typical sortase fold with a β-barrel structure that houses three conserved active site residues that are essential for the enzyme catalytic activity and a unique feature represented by an N-terminal region containing a loop, called "lid", that masks the catalytic triad, suggesting a regulatory role of this loop for sortases activation.

The data presented in this thesis work represent the first direct experimental evidence of the lid function, thus providing novel insights into class C sortases regulation and activation (Cozzi, Zerbini et al.).

By using an *in vitro* assay we demonstrate that an efficient polymerization of pilin proteins in high molecular weight (HMW) complexes can be achieved by using a recombinant sortase C lid mutant, expressed in soluble form and purified from E. Coli (Cozzi, Zerbini et al.). This mutant was generated based on structural analysis of the three-dimensional structure of SrtC1 from GBS pilus island 2a (SrtC1-2a) (Cozzi, Malito et al. 2011). The SrtC1-2a crystal structure showed that the aromatic ring of Tyr86 in the N-terminal lid region is positioned in a highly conserved hydrophobic environment (Leu131, Leu138, Val153, Leu217) and can potentially be involved in CH-π weak polar interactions with specific residues, including the catalytic Cys219 of the enzyme active site (Cozzi, Malito et al. 2011). This kind of interaction has been also observed in other pilus-related sortase structures, suggesting that the lid closes the active site contributing to an overall stability of the protein (Manzano, Contreras-Martel et al. 2008; Manzano, Izore et al. 2009; Cozzi, Malito et al. 2011; Persson 2011). Starting from these observations, we performed *in vitro* experiments by using or the recombinant GBS $SrtC1_{WT}$ or the lid mutant $SrtC1_{Y86A}$ mixed with the purified recombinant backbone protein from pilus 2a (BP-2a), carrying the sequence elements (as the pilin motif and the IPQTG sorting signal) absolutely required for pilus formation *in vivo*. We observed that, while the wild type enzyme was totally inactive, the lid mutant $SrtC1_{Y86A}$ was able to efficiently assembly the backbone subunit in HMW polymers, clearly detectable by SDS-PAGE of the reaction mixtures and Coomassie-staining. These data represent the first direct experimental evidence that a single residue in the lid can regulate the enzyme catalytic activity (Cozzi, Zerbini et al.). However, we have previously observed that $SrtC1_{WT}$ is able to

cleave the fluorogenic peptides carrying the LPXTG motifs of the substrate proteins in *in vitro* FRET-based assays, probably because the used peptides are small enough to enter in the catalytic pocket (Cozzi, Malito et al. 2011; Cozzi, Prigozhin et al. 2012). The lid, interacting with the catalytic cysteine through a specific residue (i.e. Tyr86 in GBS SrtC1-2a), blocks the enzyme in a closed conformation, thus preventing the accessibility to the substrate. The mutation of this key residue might break this interaction making the active site available for substrate binding. Moreover, the destruction of the aromatic-sulfur interaction with the catalytic Cys219 side chain might induce a perturbation in the N-terminal flexible region (helix α1-α2) and/or increase the exposition of the β-barrel core to proteases.

We also evaluated specificity of the $SrtC1_{Y86A}$ transpeptidation activity by using as substrates BP proteins of other GBS pili and/or pathogens. Interestingly we showed that this mutant is able to efficiently polymerize different substrates, showing that the *in vitro* specificity of $SrtC1_{Y86A}$ for the LPXTG motif is not restrictive and therefore this mutant seems to be promiscuous in substrate recognition.

This is further confirmed also by using as substrate for the polymerization reaction a protein non-pilus correlated, the GFP containing the LPXTG motif at the C-term domain, which is polymerized *in vitro* by the active sortase lid mutant.

As soon as we have shown that the $SrtC1_{Y86A}$ can polymerize *in vitro* BP_{K189A} and a non pilus-related protein, as the GFP, only added of the C-terminal LPXTG motif, we have thus investigated on the possible nucleophile residue involved in the reaction *in vitro* by mass spectrometry, analyzing the HMW bands observed in

the reactions performed with BP-2a wild type,BP_{K189A}, and the GFP-IPQTG as substrates.

The cross-linked peptides identified by mass spectrometry, showed that the nucleophile residue employed by $SrtC1_{Y86A}$ is the N-terminal glycine, thus demonstrating that the polymerization *in vitro* occurs through a different mechanism respect the *in vivo* reaction, in which is the lysine residue in the pilin motif the only residue that can resolve the sortase-LPXTG acyl-intermediate.

Moreover, by biochemical characterization assays we observed that the presence of the lid in $SrtC1_{WT}$ confers to the enzyme stability and resistance to proteolysis. As observed in the proteolysis experiments the β-barrel core becomes more susceptible to trypsin digestion, when the Tyr86-Cys219 binding is missed and the active site is not protected by the lid. This observation leads to the hypothesis that the lid is part of a pro-domain, made of the entire N-terminal region and that the activation of the enzyme may occur via proteolytic processing producing an activated form, that consists just of the catalytic domain. In support to this hypothesis we observed that the recombinant $SrtC1_{\Delta NT}$ mutant, lacking the entire N-terminal domain, still retains transpeptidation activity *in vitro* (Fig.32), according also with previous data showing that GBS pilus 1 SrtC mutants, deleted of the entire N-terminus, are active in *in vitro* FRET-based assays (Cozzi, Prigozhin et al. 2012). The proteolytic removal of the N-terminal segment is not the only possible mechanism for sortase C enzyme activation on the bacterial surface and, of course, needs to be confirmed *in vivo* as well as the function of the entire N-terminal domain must be further elucidated (Cozzi, Zerbini et al.). Certainly, our data are in agreement with the available structural data which

indicate that the conformation of the N-terminal segment is not compatible with substrate binding, meaning that a conformational change that relocates the lid must occur in order to make the active-site groove accessible. Moreover, our data support the hypothesis of sortase C enzymes having two functional domains, in which the entire N-terminal domain (not only the lid region) is involved in enzyme activity regulation (Cozzi, Prigozhin et al. 2012).

Another interesting evidence showed in this work is that, differently from what observed *in vivo* in native conditions, $SrtC1_{Y86A}$ mutant can polymerize *in vitro* the pilus backbone protein even when the Lys residue of the pilin motif is mutated. Therefore, we consider plausible that the flexible N-terminal domain, changing its conformation over the catalytic groove, could be involved both in the regulation of the enzyme activity and in the specific recognition of the right pilin motif during the transpeptidation reaction. Also we cannot exclude that on the bacterial surface additional factors could be involved in the sortase activity regulation as well as in the substrate specificity. Further efforts will be necessary to explore what exactly happens *in vivo* and to understand the differences at molecular level between the catalytic mechanisms of transpeptidation occurring *in vitro* vs *in vivo.*

An additional open question is if the regulation role played by the N-terminal domain, including the lid, is common for all pilus-related sortases, including those sortases that do not contain a canonical lid motif (Kang, Coulibaly et al. 2011; Lu, Qi et al. 2011).

Finally, another important message conveyed by this study regards the potential use of activated forms of pilus-associated sortases as a protein engineering to obtain *in vitro* polymers of proteins containing an LPXTG motif at the C-terminal

domain and a glycine residue at the N-terminal domain. Thus, the findings reported in this study open a new field to produce polymeric complexes through an *in vitro* reaction. The design of other mutations into the lid region and their further biochemical and kinetic characterization could be useful in order to increase the activity of this active enzyme and further optimize the *in vitro* reaction conditions.

Chapter 4. Experimental procedures

4.1 Materials and reagents

Buffer salts were purchased from Sigma.

PCR amplifications were performed using PrimeSTAR HS DNA Polymerase (Takara). PCR products were purified using the Wizard SV Gel/PCR Clean-Up System (Promega). Plasmids were purified from E. coli cells using a E.Z.N.A.® HP Plasmid Mini Kit I - Omega Bio-Tek.

Protein purification was performed on AKTA Purifier chromatography system (GE-Healthcare).

4.2 Bionformatics

Prediction of transmembrane helices and membrane topology of protein sequences was obtained using *TMHMM*. Multiple sequence alignments were performed using *ClustalW* program.

4.3 Bacterial Strains, Media and Growth Conditions

Group B *Streptococcus* (GBS) strain used in this work, 515 (serotype Ia), H36B (serotype Ib), CJB111 (serotype V) were a kind gift of D. Kasper. GBS cells were grown in Todd Hewitt Broth (THB thereafter) or in Trypticase soy agar (TSA thereafter) supplemented with 5% sheep blood at 37°C.

4.4 PI-2a SrtC1 recombinant cloning and expression

Genomic DNA isolated from the GBS strain 515 (serotype Ia) was used as template for cloning the recombinant sortase C1.

PCR fragment encoding residues 43 to 292 of the SrtC1 protein ($SrtC1_{WT}$) (TIGR annotation SAL_1484) was first cloned into pMAL-c4X Vector in house modified to generate N-terminal His-MBP, TEV cleavable, fusion proteins. The recombinant mutants $SrtC1_{Y86A}$, $SrtC1_{\Delta LID}$ without the lid region (residues 81–96) and the construct without the N-terminus domain, $SrtC1_{96\text{-}292}$ ($SrtC1_{\Delta NT}$), corresponding to the β-barrel core and the C-terminal trans-membrane domain, were generated by PIPE site-directed mutagenesis using wild type $SrtC1_{43\text{-}292}$ (PIPE ref). The resulting constructs was analyzed by DNA sequencing and expressed into E. coli BL21(DE3) cells resistant to phage T1 (NEB). The cells were grown in EnBase (Biosilta) medium in shaking flasks at 30 °C for 16 hours of cultivation, until OD(600) about 15, and then induced with 1mM IPTG 24h at 25 °C. Afterwards the cells were harvested by centrifugation at 6000rpm for 30minutes. The soluble proteins were extracted by sonication in binding buffer (50mM Tris-HCl (pH 7.5), 400mM NaCl, 20mM imidazole, lysozyme and DNAse) and centrifuged 40min at 9000rpm to remove the cell debris. Protein purification was performed loading the sample on a FF-Crude His-Trap HP nickel chelating column (GE Healthcare) pre-equilibrated with the binding buffer. The proteins were eluted with 300mM imidazole; at the end of the IMAC all the fractions obtained were analyzed by SDS-PAGE. The eluted protein was concentrated by ultrafiltration with amicon (Millipore) and the buffer was exchanged using a PD-10 desalting column (GE Healthcare) equilibrated with

TEV cleavage buffer (Tris-HCl 50mM (pH 8), DTT 1mM, EDTA 0.5mM). His-MBP-tag was cleaved by incubation with TEV protease and then removed performing a "subtractive" IMAC purification step, using a FF-Crude His-Trap HP nickel chelating column and MBP-Trap (GE Healthcare). The protein was quantified with the BCA assay (Pierce).

For the expression of ^{15}N labelled samples the constructs transformed into E. coli BL21 (DE3) cells resistant to phage T1 (NEB) were grown in M9 minimal medium (containing 1g/l of $(^{15}NH_4)_2SO_4$), in shaking flasks at 37 °C until OD(600) about 0.7 and then induced with 1mM IPTG for 5 h at 25 °C. Afterwards the cells were harvested by centrifugation and the proteins purification was performed as described above.

4.5 Recombinant backbone proteins cloning and expression

Recombinant PI-2a backbone protein BP_{30-649}, BP-2a strain 515 (TIGR annotation SAL_1486) was cloned in the SpeedET vector; the other variants amplified from strainsH36B and CJB111 (TIGR annotation SAI_1511, SAM_1372 respectively), were cloned in pET15 vector (adapted for PIPE cloning). The recombinant BP_{K189A} and the BP_{30-640}, lacking the C-terminal IPQTG motif was generated by PIPE site-directed mutagenesis using wild type BP_{30-649} as template.

Genes coding for BP-1 and BP-2b, pilus 1 and pilus 2b backbone protein, (TIGR annotation SAG_0645 and SAN_1518 respectively), were PCR amplified from GBS strains 2603V/R and COH1, respectively. The gene coding for RrgB (TIGR

annotation SP_0463), *S. pneumoniae* backbone protein, was PCR amplified from the pneumococcal strain TIGR4.

PCR fragments encoding BP-$1_{30\text{-}529}$ and $RrgB_{30\text{-}636}$ were cloned using PIPE method into pET15 vector (adapted for PIPE cloning), while BP-$2b_{30\text{-}476}$ was cloned in pET54-DEST (Novagen) vector using the Gateway cloning method.

All the resulting constructs were analyzed by DNA sequencing and expressed into E. coli BL21(DE3) cells resistant to phage T1 (NEB) as N-terminal His-tag, TEV cleavable, fusion proteins. The cells were grown in EnBase (Biosilta) medium in shaking flasks at 30 °C for 16 hours of cultivation, until OD(600) about 15, and then induced with 1mM IPTG 24h at 25 °C. Afterwards the cells were harvested by centrifugation at 6000rpm for 30minutes. The soluble proteins were extracted by sonication in binding buffer (50mM Tris-HCl (pH 7.5), 300mM NaCl, 20mM imidazole, lysozyme and DNAse) and centrifuged 40min at 9000rpm to remove the cell debris. Protein purification was performed loading the sample on a FF-Crude His-Trap HP nickel chelating column (GE Healthcare) pre-equilibrated with the binding buffer. The proteins were eluted with 300mM imidazole; at the end of the IMAC all the fractions obtained were analyzed by SDS-PAGE. The eluted proteins were concentrated by ultrafiltration with amicon (Millipore) and the buffer was exchanged using a PD-10 desalting column (GE Healthcare) equilibrated with TEV cleavage buffer (Tris-HCl 50mM (pH 8), DTT 1mM, EDTA 0.5mM). His-tag was cleaved by incubation with TEV protease and then removed performing a “subtractive” IMAC purification step, using a FF-Crude His-Trap HP nickel chelating column. A final step of purification was performed by size exclusion using HiLoad 26/60 Superdex 200 (GE Healthcare)

equilibrated in 25 mM Hepes, 75 mM NaCl pH=7.5. The fractions containing the pure protein, which showed a single component by SDS-PAGE, were quantified with the BCA assay (Pierce).

4.6 *In vitro* pilus polymerization

The *in vitro* polymerization assay was performed by mixing different concentrations of GBS SrtC1$_{WT}$ and SrtC1$_{Y86A}$ (1-5-10-25-100 µM) and GBS BP-2a (25-50-100-200 µM). The polymerization reactions with other GBS BP proteins, *S.pneumoniae* RrgB and GFP$_{IPQTG}$ were set-up by mixing a fix concentration of SrtC1$_{Y86A}$ (25µM) and 100µM of each pilins. The volume of reaction was 50µl in buffer 25mM of Tris-HCl, 100 mM NaCl, 1 mM DTT, pH=7.5. The incubation was performed at 37°C in a thermo-mixer, and the reaction was analyzed by SDS-PAGE at different time points (up to 3 days). The SDS-PAGE analysis was performed using 4-12% Criterion XT Bis-Tris Precast Gels (Bio-Rad) with MES running buffer and stained with Comassie blue.

4.7 Differential scanning fluorimetry (DSF)

The thermal stability of recombinant GBS PI-2a SrtC1$_{WT}$, SrtC1$_{\Delta LID}$ and SrtC1$_{Y86A}$ was investigated by differential scanning fluorimetry (DSF) analysis. In a 96-well plate (Thermo Fast 96-ABgene) 40ul samples containing 25 µM of the enzyme and 5X Sypro Orange (sigma Aldrich) in 25 mM Tris-HCl, 100 mM NaCl pH=7.5 buffer were analyzed. The unfolding profile and the melting

temperature were monitored by a quantitative PCR thermo cycler (Stratagene) as already reported (Ericsson, Hallberg et al. 2006).

4.8 Antisera

Antisera specific for the BP-2a, AP1-2a and AP2-2a proteins were produced by immunizing CD1 mice with the purified recombinant proteins (Maione, Margarit et al. 2005; Nuccitelli, Cozzi et al. 2011).

4.9 Bacterial strains and growth conditions

GBS 515 strain and mutants were grown in Todd Hewitt Broth (THB) or in Trypticase soy agar (TSA) supplemented with 5% sheep blood at 37°C.

4.10 Construction of complementation vectors

GBS knock-out (KO) mutant strain for BP, was generated as previously reported (Rosini, Rinaudo et al. 2006). For the generation of complementation vectors DNA fragments corresponding to wild type BP (SAL_1486), gene was PCR amplified from GBS 515 genome and the product was cloned into the *E. coli*-streptococcal shuttle vector pAM401/gbs80P+T, previously described (Rosini, Rinaudo et al. 2006) and containing the promoter and terminator regions of the *gbs80* gene (TIGR annotation SAG_0645). Site-directed mutagenesis of pAM_BP was performed using the PIPE (Polymerase Incomplete Primer Extension) method. As template for the introduction by PCR of specific mutations and/or deletion we

used the previously generated (Rosini, Rinaudo et al. 2006; Cozzi, Nuccitelli et al. 2012) complementation vector carrying the BP-2a gene (pAM-BP-2a), and the replacement of selected amino acid.

The complementation vectors pAM_$BP_{\Delta IPQTG}$ and pAM_BP_{K189A} were electroporated into the knock out strain 515ΔBP-2a. Complementation was confirmed by checking BP-2a expression by Western Blotting.

4.11 Western Blot Analysis

Mid-exponential phase bacterial cells were resuspended in 50mM Tris-HCl containing 400U of mutanolysin (Sigma-Aldrich) and COMPLETE protease inhibitors (Roche). The mixtures were then incubated at 37°C for 1h and cells lysed by three cycles of freeze-thawing. Cellular debris were removed by centrifugation and protein concentration was determined using BCA protein assay (Pierce, Rockford, IL).

Total protein extracts (20 µg) or recombinant pili were resolved on 3-8% or 4-12% NuPAGE gels (Invitrogen) by SDS-PAGE and transferred to nitrocellulose. Membranes were probed with mouse antiserum directed against BP and AP1 proteins (1:1,000 dilution) followed by a rabbit anti-mouse horseradish peroxidase-conjugated secondary antibody (Dako, Glostrup, Denmark). Bands were then visualized using an Opti-4CN substrate kit (Bio-Rad).

4.12 Limited proteolysis assay

Sequencing grade Trypsin (Promega) was dissolved in the buffer provided to a final concentration of 0.2 µg/µl and was activated 10 minutes at 37°C. 2 µg of trypsin and 200 µg of $SrtC1_{WT}$ and $SrtC1_{Y86A}$ were mixed in a final volume of 100 µl in buffer 25 mM Tris-HCl, 100 mM NaCl pH=7.5. The reactions were incubated in thermomixer at 37°C. Samples were collected after 5 - 10 - 20 - 30 - 60 and 120 minutes, the proteolysis was then quenched by adding final 0.1% formic acid (vol/vol) and analyzed by SDS-PAGE.

4.13 Intact mass determination by ESI-Q-TOF

The trypsin digestion mixture was diluted in 0.1% formic acid. The acidifed protein solutions were loaded onto a ProteinMicrotrap cartridge (from 60 to 100 pmols), desalted for 2 min with 0.1% formic acid at a flow rate of 200 mL/min and eluted directly into the mass spectrometer using a step gradient of acetonitrile (55% acetonitrile, 0.1% formic acid). Spectra were acquired in positive mode on a SynaptG2 HDMS mass spectrometer equipped with a Z-spray ESI source. The quadrupole profile was optimized to ensure the best transmission of all ions generated during the ionization process.

4.14 Analytic size-exclusion chromatography

$SrtC1_{WT}$ was subjected to limited proteolysis for 30 minutes at 37°C, using the same amount and ratio of trypsin described above. The digestion mix was

promptly loaded in Superdex 75 10/300 (GE) equilibrated with 25 mM Tris-HCl pH 7.5 and 150 mM NaCl, at the flow rate of 0.8 ml/min. The chromatography was performed using a UV detector monitoring at 280 nm and 0.2 ml/ tube fractions were collected. The protein fractions, identified by the chromatogram peaks, were analyzed by SDS-PAGE. A Bio-Rad's gel filtration standard (catalog number 151-1901) was used for the calibration of standard peaks.

4.15 NMR spectroscopy

Nitrogen-15-labelled recombinant proteins were expressed in cells grown in M9 minimal medium containing 1 g/L of $(^{15}NH_4)_2SO_4$ and following otherwise the same protocols as for unlabeled samples. The protein buffers were exchanged using a PD-10 desalting column (Amersham Biosciences), equilibrated with 50 mM phosphate buffer, pH 6.5, and finally concentrated by ultrafiltration to ~0.2 mM. ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were recorded at 25°C on a Bruker Avance III spectrometer operating at 600.13 MHz proton Larmor frequency, equipped with a cryogenic probe. A standard ^{1}H-^{15}N HSQC pulse sequence was used, with pulsed field gradients for suppression of the solvent signal and cancellation of spectral artifacts. 2048 (^{1}H) × 256 (^{15}N) complex data points were acquired with spectral windows of 9515.385 Hz (^{1}H) × 2432.718 Hz (^{15}N), 8 transients, and 1.2 s relaxation delay. Proton T_2 measurements were performed with the 1D oneone echo sequence (Sklenar and Bax, Journal of Magnetic Resonance, 74:469, 1987) using variable delays of 0.2 and 5.2 ms and evaluating the corresponding signal intensities (T_2=2×(5.2-

0.2)/ln($I_{0.2}/I_{5.2}$). Processing of all the spectra was performed with Topspin2.1 (Bruker, Karlsruhe).

Bibliography

Antos, J. M., G. L. Chew, et al. (2009). "Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity." J Am Chem Soc 131(31): 10800-1.

Antos, J. M., G. M. Miller, et al. (2008). "Lipid modification of proteins through sortase-catalyzed transpeptidation." J Am Chem Soc 130(48): 16338-43.

Baker, C. J. (1997). "Group B streptococcal infections." Clin Perinatol 24(1): 59-70.

Bateman, A., E. Birney, et al. (2000). "The Pfam protein families database." Nucleic Acids Res 28(1): 263-6.

Bentley, M. L., E. C. Lamb, et al. (2008). "Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from Staphylococcus aureus." J Biol Chem 283(21): 14762-71.

Buccato, S., D. Maione, et al. (2006). "Use of Lactococcus lactis expressing pili from group B Streptococcus as a broad-coverage vaccine against streptococcal disease." J Infect Dis 194(3): 331-40.

Clancy, K. W., J. A. Melvin, et al. "Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition." Biopolymers 94(4): 385-96.

Comfort, D. and R. T. Clubb (2004). "A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria." Infect Immun 72(5): 2710-22.

Cozzi, R., E. Malito, et al. (2011). "Structure analysis and site-directed mutagenesis of defined key residues and motives for pilus-related sortase C1 in group B Streptococcus." FASEB J 25(6): 1874-86.

Cozzi, R., A. Nuccitelli, et al. (2012). "New insights into the role of the glutamic acid of the E-box motif in group B Streptococcus pilus 2a assembly." FASEB J 26(5): 2008-18.

Cozzi, R., A. Nuccitelli, et al. (2012). "New insights into the role of the glutamic acid of the E-box motif in group B Streptococcus pilus 2a assembly." FASEB J.

Cozzi, R., D. Prigozhin, et al. (2012). "Structural basis for group B streptococcus pilus 1 sortases C regulation and specificity." PLoS One 7(11): e49048.

Cozzi, R., F. Zerbini, et al. "Group B Streptococcus pilus sortase regulation: a single mutation in the lid region induces pilin protein polymerization in vitro." FASEB J 27(8): 3144-54.

Davies, H. D., C. Adair, et al. (2001). "Antibodies to capsular polysaccharides of group B Streptococcus in pregnant Canadian women: relationship to colonization status and infection in the neonate." J Infect Dis 184(3): 285-91.

De Groot, A. S. and R. Rappuoli (2004). "Genome-derived vaccines." Expert Rev Vaccines 3(1): 59-76.

Doran, K. S. and V. Nizet (2004). "Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy." Mol Microbiol 54(1): 23-31.

Dramsi, S., E. Caliot, et al. (2006). "Assembly and role of pili in group B streptococci." Mol Microbiol 60(6): 1401-13.

Dramsi, S., S. Magnet, et al. (2008). "Covalent attachment of proteins to peptidoglycan." FEMS Microbiol Rev 32(2): 307-20.

Ericsson, U. B., B. M. Hallberg, et al. (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem 357(2): 289-98.

Finn, R. D., J. Mistry, et al. "The Pfam protein families database." Nucleic Acids Res 38(Database issue): D211-22.

Frankel, B. A., Y. Tong, et al. (2007). "Mutational analysis of active site residues in the Staphylococcus aureus transpeptidase SrtA." Biochemistry 46(24): 7269-78.

Gibbs, R. S., S. Schrag, et al. (2004). "Perinatal infections due to group B streptococci." Obstet Gynecol 104(5 Pt 1): 1062-76.

Guo, X., Q. Wang, et al. (2009). "Sortase-catalyzed peptide-glycosylphosphatidylinositol analogue ligation." J Am Chem Soc 131(29): 9878-9.

Guttilla, I. K., A. H. Gaspar, et al. (2009). "Acyl enzyme intermediates in sortase-catalyzed pilus morphogenesis in gram-positive bacteria." J Bacteriol 191(18): 5603-12.

Hansen, S. M., N. Uldbjerg, et al. (2004). "Dynamics of Streptococcus agalactiae colonization in women during and after pregnancy and in their infants." J Clin Microbiol 42(1): 83-9.

Hendrickx, A. P., J. M. Budzik, et al. "Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds." Nat Rev Microbiol 9(3): 166-76.

Ilangovan, U., J. Iwahara, et al. (2001). "Assignment of the 1H, 13C and 15N signals of Sortase." J Biomol NMR 19(4): 379-80.

Ilangovan, U., H. Ton-That, et al. (2001). "Structure of sortase, the transpeptidase that anchors proteins to the cell wall of Staphylococcus aureus." Proc Natl Acad Sci U S A 98(11): 6056-61.

Kang, H. J. and E. N. Baker "Intramolecular isopeptide bonds: protein crosslinks built for stress?" Trends Biochem Sci 36(4): 229-37.

Kang, H. J. and E. N. Baker "Structure and assembly of Gram-positive bacterial pili: unique covalent polymers." Curr Opin Struct Biol 22(2): 200-7.

Kang, H. J., F. Coulibaly, et al. (2007). "Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure." Science 318(5856): 1625-8.

Kang, H. J., F. Coulibaly, et al. (2011). "Crystal structure of Spy0129, a Streptococcus pyogenes class B sortase involved in pilus assembly." PLoS One 6(1): e15969.

Kang, H. J., N. G. Paterson, et al. (2009). "The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds." Proc Natl Acad Sci U S A 106(40): 16967-71.

Khare, B., Z. Q. Fu, et al. (2011). "The crystal structure analysis of group B Streptococcus sortase C1: a model for the "lid" movement upon substrate binding." J Mol Biol 414(4): 563-77.

Khare, B., V. Krishnan, et al. (2011). "Structural differences between the Streptococcus agalactiae housekeeping and pilus-specific sortases: SrtA and SrtC1." PLoS One 6(8): e22995.

Kogan, G., D. Uhrin, et al. (1996). "Structural and immunochemical characterization of the type VIII group B Streptococcus capsular polysaccharide." J Biol Chem 271(15): 8786-90.

Konto-Ghiorghi, Y., E. Mairey, et al. (2009). "Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae." PLoS Pathog 5(5): e1000422.

Lauer, P., C. D. Rinaudo, et al. (2005). "Genome Analysis Reveals Pili in Group B Streptococcus." Science 309(5731): 105-.

Lu, G., J. Qi, et al. (2011). "A novel "open-form" structure of sortaseC from Streptococcus suis." Proteins 79(9): 2764-9.

Maione, D., I. Margarit, et al. (2005). "Identification of a Universal Group B Streptococcus Vaccine by Multiple Genome Screen." Science 309(5731): 148-150.

Maisey, H. C., M. Hensler, et al. (2007). "Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells." J Bacteriol 189(4): 1464-7.

Mandlik, A., A. Swierczynski, et al. (2008). "Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development." Trends Microbiol 16(1): 33-40.

Manzano, C., C. Contreras-Martel, et al. (2008). "Sortase-mediated pilus fiber biogenesis in Streptococcus pneumoniae." Structure 16(12): 1838-48.

Manzano, C., T. Izore, et al. (2009). "Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens." Biochemistry 48(44): 10549-57.

Maresso, A. W., T. J. Chapa, et al. (2006). "Surface protein IsdC and Sortase B are required for heme-iron scavenging of Bacillus anthracis." J Bacteriol 188(23): 8145-52.

Margarit, I., C. D. Rinaudo, et al. (2009). "Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm." J Infect Dis 199(1): 108-15.

Marraffini, L. A., A. C. Dedent, et al. (2006). "Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria." Microbiol Mol Biol Rev 70(1): 192-221.

Marraffini, L. A., H. Ton-That, et al. (2004). "Anchoring of surface proteins to the cell wall of Staphylococcus aureus. A conserved arginine residue is required for efficient catalysis of sortase A." J Biol Chem 279(36): 37763-70.

Mazmanian, S. K., G. Liu, et al. (1999). "Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall." Science 285(5428): 760-3.

McCracken, G. H., Jr. (1973). "Group B streptococci: the new challenge in neonatal infections." J Pediatr 82(4): 703-6.

Mitchell, T. J. (2003). "The pathogenesis of streptococcal infections: from tooth decay to meningitis." Nat Rev Microbiol 1(3): 219-30.

Naik, M. T., N. Suree, et al. (2006). "Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop." J Biol Chem 281(3): 1817-26.

Nakata, M., T. Koller, et al. (2009). "Mode of expression and functional characterization of FCT-3 pilus region-encoded proteins in Streptococcus pyogenes serotype M49." Infect Immun 77(1): 32-44.

Nallapareddy, S. R., K. V. Singh, et al. (2006). "Endocarditis and biofilm-associated pili of Enterococcus faecalis." J Clin Invest 116(10): 2799-807.

Necchi, F., V. Nardi-Dei, et al. "Sortase A substrate specificity in GBS pilus 2a cell wall anchoring." PLoS One 6(10): e25300.

Neiers, F., C. Madhurantakam, et al. (2009). "Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity." J Mol Biol 393(3): 704-16.

Nelson, J. W., A. G. Chamessian, et al. "A biosynthetic strategy for re-engineering the Staphylococcus aureus cell wall with non-native small molecules." ACS Chem Biol 5(12): 1147-55.

Nguyen, H. D., T. T. Phan, et al. "Analysis and application of Bacillus subtilis sortases to anchor recombinant proteins on the cell wall." AMB Express 1(1): 22.

Nguyen, H. D. and W. Schumann (2006). "Establishment of an experimental system allowing immobilization of proteins on the surface of Bacillus subtilis cells." J Biotechnol 122(4): 473-82.

Nizet, V., R. L. Gibson, et al. (1996). "Group B streptococcal beta-hemolysin expression is associated with injury of lung epithelial cells." Infect Immun 64(9): 3818-26.

Nobbs, A. H., R. Rosini, et al. (2008). "Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in Streptococcus agalactiae." Infect Immun 76(8): 3550-60.

Nuccitelli, A., R. Cozzi, et al. "Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections." Proc Natl Acad Sci U S A 108(25): 10278-83.

Nuccitelli, A., R. Cozzi, et al. (2011). "Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B Streptococcus infections." Proc Natl Acad Sci U S A 108(25): 10278-83.

Pallen, M. J., R. R. Chaudhuri, et al. (2003). "Genomic analysis of secretion systems." Curr Opin Microbiol 6(5): 519-27.

Pallen, M. J., A. C. Lam, et al. (2001). "An embarrassment of sortases - a richness of substrates?" Trends Microbiol 9(3): 97-102.

Paterson, G. K. and T. J. Mitchell (2004). "The biology of Gram-positive sortase enzymes." Trends Microbiol 12(2): 89-95.

Persson, K. (2011). "Structure of the sortase AcSrtC-1 from Actinomyces oris." Acta Crystallogr D Biol Crystallogr 67(Pt 3): 212-7.

Popp, M. W., J. M. Antos, et al. (2009). "Site-specific protein labeling via sortase-mediated transpeptidation." Curr Protoc Protein Sci Chapter 15: Unit 15 3.

Popp, M. W. and H. L. Ploegh "Making and breaking peptide bonds: protein engineering using sortase." Angew Chem Int Ed Engl 50(22): 5024-32.

Pozzi, G., M. Contorni, et al. (1992). "Delivery and expression of a heterologous antigen on the surface of streptococci." Infect Immun 60(5): 1902-7.

Proft, T. and E. N. Baker (2009). "Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease." Cell Mol Life Sci 66(4): 613-35.

Rinaudo, C. D., R. Rosini, et al. "Specific involvement of pilus type 2a in biofilm formation in group B Streptococcus." PLoS One 5(2): e9216.

Rosini, R., C. D. Rinaudo, et al. (2006). "Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae." Mol Microbiol 61(1): 126-41.

Samantaray, S., U. Marathe, et al. (2008). "Peptide-sugar ligation catalyzed by transpeptidase sortase: a facile approach to neoglycoconjugate synthesis." J Am Chem Soc 130(7): 2132-3.

Schneewind, O., P. Model, et al. (1992). "Sorting of protein A to the staphylococcal cell wall." Cell 70(2): 267-81.

Schuchat, A. (1998). "Epidemiology of group B streptococcal disease in the United States: shifting paradigms." Clin Microbiol Rev 11(3): 497-513.

Soriani, M., I. Santi, et al. (2006). "Group B Streptococcus crosses human epithelial cells by a paracellular route." J Infect Dis 193(2): 241-50.

Spirig, T., E. M. Weiner, et al. "Sortase enzymes in Gram-positive bacteria." Mol Microbiol 82(5): 1044-59.

Suree, N., C. K. Liew, et al. (2009). "The structure of the Staphylococcus aureus sortase-substrate complex reveals how the universally conserved LPXTG sorting signal is recognized." J Biol Chem 284(36): 24465-77.

Telford, J. L., M. A. Barocchi, et al. (2006). "Pili in gram-positive pathogens." Nat Rev Microbiol 4(7): 509-19.

Tettelin, H., V. Masignani, et al. (2005). "Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome"." Proc Natl Acad Sci U S A 102(39): 13950-5.

Ton-That, H., G. Liu, et al. (1999). "Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif." Proc Natl Acad Sci U S A 96(22): 12424-9.

Ton-That, H., L. A. Marraffini, et al. (2004). "Protein sorting to the cell wall envelope of Gram-positive bacteria." Biochim Biophys Acta 1694(1-3): 269-78.

Ton-That, H., L. A. Marraffini, et al. (2004). "Sortases and pilin elements involved in pilus assembly of Corynebacterium diphtheriae." Mol Microbiol 53(1): 251-61.

Ton-That, H., S. K. Mazmanian, et al. (2002). "Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Cysteine 184 and histidine 120 of sortase form a thiolate-imidazolium ion pair for catalysis." J Biol Chem 277(9): 7447-52.

Ton-That, H. and O. Schneewind (2003). "Assembly of pili on the surface of Corynebacterium diphtheriae." Mol Microbiol 50(4): 1429-38.

Ton-That, H. and O. Schneewind (2004). "Assembly of pili in Gram-positive bacteria." Trends Microbiol 12(5): 228-34.

Tsukiji, S. and T. Nagamune (2009). "Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering." Chembiochem 10(5): 787-98.

Viguera, A. R. and L. Serrano (1995). "Side-chain interactions between sulfur-containing amino acids and phenylalanine in alpha-helices." Biochemistry 34(27): 8771-9.

Waksman, G. and S. J. Hultgren (2009). "Structural biology of the chaperone-usher pathway of pilus biogenesis." Nat Rev Microbiol 7(11): 765-74.

Weiner, E. M., S. Robson, et al. "The Sortase A enzyme that attaches proteins to the cell wall of Bacillus anthracis contains an unusual active site architecture." J Biol Chem 285(30): 23433-43.

Wu, C., A. Mishra, et al. "Structural determinants of Actinomyces sortase SrtC2 required for membrane localization and assembly of type 2 fimbriae for interbacterial coaggregation and oral biofilm formation." J Bacteriol 194(10): 2531-9.

Yamamoto, T. and T. Nagamune (2009). "Expansion of the sortase-mediated labeling method for site-specific N-terminal labeling of cell surface proteins on living cells." Chem Commun (Camb)(9): 1022-4.

Yamamoto, Y., V. Pargade, et al. (2006). "The Group B Streptococcus NADH oxidase Nox-2 is involved in fatty acid biosynthesis during aerobic growth and contributes to virulence." Mol Microbiol.

Zong, Y., T. W. Bice, et al. (2004). "Crystal structures of Staphylococcus aureus sortase A and its substrate complex." J Biol Chem 279(30): 31383-9.

Zong, Y., S. K. Mazmanian, et al. (2004). "The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the Staphylococcus aureus cell wall." Structure 12(1): 105-12.