Covalent attachment of proteins to peptidoglycan

Shaynoor Dramsi1,2, Sophie Magnet3,4,5, Sophie Davison2,6 & Michel Arthur3,4,5

1Unité de Biologie des Bactéries Pathogènes à Gram-positif, Institut Pasteur, Paris, France; 2CNRS, URA 2172, Paris, France; 3INSERM, U872, LRMA, Centre de Recherches des Cordeliers, Paris, France; 4Université Pierre et Marie Curie – Paris 6, UMR S872, Paris, France; 5Université Paris Descartes, UMR 872, Paris, France; and 6Unité des Toxines et Pathogénie Bactérienne, Institut Pasteur, Paris, France

Correspondence: Shaynoor Dramsi, Institut Pasteur, 25, rue du Dr Roux, 75724 Paris Cedex 15, France. Tel.:+33 1 44 38 94 87; fax:+33 1 45 68 89 38; e-mail: sdramsi@pasteur.fr

Received 27 July 2007; revised 5 November 2007; accepted 19 December 2007. First published online February 2008.

DOI:10.1111/j.1574-6976.2008.00102.x

Editor: Jacques Coyette

Keywords
surface proteins; Braun lipoprotein; sortase; transpeptidase; peptidoglycan.

Abstract

Bacterial surface proteins are key players in host–symbiont or host–pathogen interactions. How these proteins are targeted and displayed at the cell surface are challenging issues of both fundamental and clinical relevance. While surface proteins of Gram-negative bacteria are assembled in the outer membrane, Gram-positive bacteria predominantly utilize their thick cell wall as a platform to anchor their surface proteins. This surface display involves both covalent and noncovalent interactions with either the peptidoglycan or secondary wall polymers such as teichoic acid or lipoteichoic acid. This review focuses on the role of enzymes that covalently link surface proteins to the peptidoglycan, the well-known sortases in Gram-positive bacteria, and the recently characterized L,D-transpeptidases in Gram-negative bacteria.

Introduction

Bacterial pathogens display proteins on their surface that may interact with their hosts in order to mount successful infections. While surface proteins of Gram-negative bacteria are assembled in the outer membrane, Gram-positive bacteria predominantly utilize their thick cell wall as a platform for display of surface proteins. Although the primary function of the peptidoglycan is to provide a physical barrier for protection against both mechanical and osmotic stresses, it also serves as a scaffold to anchor external structures such as the outer cell membrane in Escherichia coli or the capsule in Bacillus anthracis. Over the past 20 years, it has become apparent that Gram-positive bacteria have evolved a variety of unique mechanisms by which proteins are displayed on the cell surface. This surface display involves both covalent and noncovalent interactions with either the peptidoglycan or secondary wall polymers such as teichoic or lipoteichoic acids. In this review, we will focus exclusively on the role of enzymes that covalently link surface proteins to the peptidoglycan, the well-known sortases in Gram-positive bacteria, and the recently characterized L,D-transpeptidases, which anchor lipoproteins to the peptidoglycan in Gram-negative bacteria. For detailed mechanisms on the display of surface proteins in Gram-positive bacteria, we refer the reader to a number of excellent reviews (Navarre & Schneewind, 1999; Mazmanian et al., 2001; Ton-That et al., 2004a; Marraffini et al., 2006).

The Gram-positive cell wall: a platform for covalent anchoring of LPXTG proteins

Sjöquist et al. (1972) presented the first evidence that the protein A of Staphylococcus aureus was covalently linked to the peptidoglycan. These authors showed that this protein could be released from the bacterial surface by treatment of staphylococci with lysostaphin, a glycyl–glycine endopeptidase that cleaves the pentaglycyl cross-bridge of the cell wall. The M protein of Streptococcus pyogenes was the second protein shown to be assembled on the cell wall. Pancholi and Fischetti elegantly showed that anchoring of M protein to the cell wall occurred at the C-terminal part after the LPXTG sequence. After trypsin treatment of whole bacteria, the M protein fragment buried in the cell wall and hence protected from the proteolytic digestion was released by phage lysin, and sequenced by Edman degradation. The resulting protein sequence consists of the C-terminal hydrophobic domain and the last residues of the M protein (Pancholi & Fischetti, 1988). A year later, they obtained the first hint about the enzyme responsible for cell wall anchoring of M protein:
a thiol-dependent membrane-bound protein of *S. pyogenes* (Pancholi & Fischetti, 1989). Furthermore, a bioinformatic analysis of the primary sequences (both at the protein and at the DNA level) of 11 surface proteins from Gram-positive cocci revealed the existence of a conserved C-terminal hexapeptide LPXTGE potentially important for cell wall localization (Fischetti *et al*., 1990).

**Cell wall sorting signal (CWS)**

The CWS consists of an LPXTG sequence, where X is any amino acid, followed by a hydrophobic stretch of amino acids and a short positively charged tail, all three of which are necessary for efficient sorting of a protein to the cell wall (Schneewind *et al*., 1992) (Fig. 1). Staphylococcal protein A lacking the entire 35-residue long CWS was secreted into the medium. Deletion of either the charged tail or the hydrophobic domain also caused protein A secretion. In contrast, deletion of the LPXTG motif resulted in abnormal localization of protein A in the bacterial membrane. It is believed that the hydrophobic domain and positively charged tail slow down protein secretion within the plasma membrane whereas the LPXTG motif is critical for protein sorting to the cell wall. The CWS alone is sufficient for cell wall anchoring of reporter proteins possessing an N-terminal signal peptide (Schneewind *et al*., 1993). Sorting signals have been observed in a plethora of predicted gene products through sequencing of the genomes of Gram-positive bacteria and were found mainly at the C-terminus of surface proteins. A reported exception is an IgA protease of *Streptococcus pneumoniae* described recently, in which the sorting signal is located near the N-terminus (Bender & Weiser, 2006).

Surface proteins contain an N-terminal signal peptide that promotes their translocation across the bacterial membrane through the Sec pathway. The *Bacillus subtilis* Sec pathway is structurally and functionally related, but not identical, to the well-known Sec machinery of *E. coli* because homologs of SecA, SecD, SecE, SecF, SecY and YajC have been identified in the genome of *B. subtilis* whereas the SecB homolog was not detected (Kunst *et al*., 1997). The protein translocation pore of *E. coli* is composed of three membrane proteins, SecYEG, which can accept substrates in two ways: (1) via a cotranslational mechanism involving the signal recognition particle (SRP) or (2) maintained unfolded by binding to chaperones like SecB. Distinct features of the *B. subtilis* protein secretion machinery include the presence of an SRP that is more closely related to human SRP than to...
its E. coli counterpart (Yamane et al., 2004). The translocated protein is transferred to SecA, an ATPase that undergoes conformational rearrangements upon interacting with the secretion machinery, a process that is thought to push precursor proteins through the translocation pore (Economou & Wickner, 1994). Following translocation of the complete polypeptide, a type I signal peptidase removes the signal peptide from the precursor, and mature protein is released into the periplasmic space (Dalbey & Wickner, 1985). Bacillus subtilis produces five type I signal peptidases (SipS, T, U, V and W) whereas E. coli produces only one: lepB. Upon complete membrane translocation, the protein that emerges in an unfolded state has to be rapidly and correctly folded into its native conformation to avoid degradation by proteases in the cell wall or in the extracellular milieu (Sibbald et al., 2006). The translocated and processed proteins that lack specific retention signals for the membrane or the cell wall are secreted into the growth medium (for reviews, see Tjalsma et al., 2004; Sibbald et al., 2006).

Signal peptides of Gram-positive bacteria are generally longer, more hydrophobic and more charged at their N-terminus than their counterparts in Gram-negative bacteria (von Heijne, 1990). The signal peptides of Gram-positive organisms have generally been found to be functional in Gram-negative bacteria, whereas the reverse has not been observed (Schneewind et al., 1992). Refined sequence analysis of the signal peptides revealed the presence of a subsequence YSIRK-G/S, which is found in many LPXTG proteins from streptococci and staphylococci but is not conserved in bacilli and listeriae. Deletion of this sequence reduces the efficiency of signal peptidase, resulting in accumulation of precursors in the membrane and thus reduces the efficiency of signal peptidase, resulting in degradation by proteases in the cell wall or in the extracellular milieu (Schneewind et al., 1992). Refined sequence analysis of the signal peptides revealed the presence of a subsequence YSIRK-G/S, which is found in many LPXTG proteins from streptococci and staphylococci but is not conserved in bacilli and listeriae. Deletion of this sequence reduces the efficiency of signal peptidase, resulting in accumulation of precursors in the membrane and thus impacting on the secretion of LPXTG proteins (Bae & Schneewind, 2003). Very recently, Carlsson et al. analyzed the localized secretion of two cell wall-anchored proteins in S. pyogenes, i.e. M protein and protein F, and found that the signal sequences direct secretion of these LPXTG proteins to different regions of the bacterium. Indeed, the signal sequence of M protein promotes secretion at the bacterial septum, whereas that of PrtF preferentially promotes secretion at the old pole (Carlsson et al., 2006). These results suggest the existence of specialized regions involved in secretion (for a recent review, see Buist et al., 2006).

**Identification of sortase**

Identification of the gene encoding the ‘LPXTG’ cell wall-anchoring enzyme arose from a genetic screen for S. aureus mutants that failed to anchor a reporter protein to the bacterial cell wall (Mazmanian et al., 1999). The gene was named srtA and encodes the sortase A (SrtA) of S. aureus. This prototype SrtA is a 206 amino acid polypeptide harboring an N-terminal hydrophobic segment that functions as both a signal peptide for secretion and as a stop transfer signal for membrane anchoring. It was demonstrated that SrtA is membrane associated with its N-terminus inside the cytoplasm and the C-terminal catalytic portion protruding in the cell wall (Mazmanian et al., 2000).

The three-dimensional structure of SrtA alone or in complex with its substrate revealed a unique fold consisting of an eight-stranded β-barrel aligned in an antiparallel and parallel fashion (Ilangovan et al., 2001; Zong et al., 2004a). The active site was found within an elongated hydrophobic groove that can accommodate unfolded proteins, suggesting a scanning mechanism of newly translocated proteins for the LPXTG motif. The nuclear magnetic resonance (NMR) structure showed that the catalytic site residue Cysteine184 is in close proximity with Histidine120 within the active site. Replacement of Cys by Ala completely abolished sortase activity both in vitro and in vivo and replacement of His or Arginine197 by Ala drastically reduced the enzymatic activity (Ton-That et al., 2002; Marraffini et al., 2004; Frankel et al., 2007). In the NMR structure, the β3–β4 and β6–β7 loops contain a set of acidic residues involved in calcium binding. Addition of Ca²⁺ in the reaction activates sortase activity eightfold probably by a mechanism that may facilitate substrate binding (Naik et al., 2006). The LPXTG-binding site has been localized by NMR studies in the C-terminal region of the β2 strand (Thr₁⁸⁰ and Ile₁⁹²) and to the vicinity of the loop connecting β3–β4 (Ala₁¹⁸). Mutations of these residues significantly impaired sortase activity in vitro (Liew et al., 2004; Frankel et al., 2007). Determination of the anchor structure of LPXTG proteins was deduced from biochemical analysis of an engineered recombinant substrate in which the protein A CWS was fused to the C-terminal end of E. coli maltose-binding protein (MalE). Cell wall-anchored MalE was released with lysostaphin from the staphylococcal envelope, purified and cleaved with trypsin, and C-terminal peptides were analyzed by Edman degradation and mass spectrometry, which revealed the sequence LPET-Gly₄, LPET-Gly₃ and LPET-Gly₂, indicating that the CWS of protein A is cleaved between the threonine and the glycine and the carboxyl group of threonine residue is then amide linked to the pentaglycine side chain of the peptidoglycan subunits (Schneewind et al., 1995).

**Two-step enzymatic reaction catalyzed by sortase A**

In a first step, sortase A cleaves the LPXTG motif between the threonine and the glycine residues both in vivo and in vitro, capturing cleaved polypeptide as a thioester-linked acyl-enzyme at its active cysteine residue (acylation step) (Ton-That et al., 1999). In a second step, covalent
attachment of the protein to the peptidoglycan is completed by the nucleophilic attack of the amino group of the side chain of a lipid II peptidoglycan precursor on the thioester (deacylation step) (Perry et al., 2002; Ruzin et al., 2002; Ton-That et al., 2002). This two-step reaction can be summarized as follows: first acylation step: \( R_1-LPXT(CO-NH)\text{-G}-R_2 + E-SH \rightarrow R_1-LPXT(CO-S)\text{-E}+NH_2-G-R_2 \) and second deacylation step: \( R_1-LPXT(CO-S)\text{-E}+NH_2-Gly_3-R_3 \rightarrow R_1-LPXT(CO-NH)\text{-Gly}_3-R_3 + E-SH \). 

The two-step reaction catalyzed by sortase A can be reconstituted in vitro. Fluorescence resonance emission transfer substrates tethered to LPETG peptide (Abz-LPETG-Dpn) have allowed determination of the kinetic parameters of the sortase A reaction \( (K_{cat}, K_m) \) as substrate cleavage is accompanied by an increase in fluorescence (Ton-That et al., 2000; Kruger et al., 2004). In the absence of the receiver substrate, recombinant sortase A with a six-histidyl affinity tag replacement of the N-terminal membrane anchor catalyzes a slow hydrolysis of peptides, cleaving the peptide bond between the threonine and the glycine of the LPETG motif. The addition of specific nucleophiles that mimic the physiological peptidoglycan side chain (tri- and pentaglycine in \( S. aureus \)) leads to an increase in the catalytic rate and results in amide bond formation between the carboxyl group of threonine and the amino group of the N-terminal glycine of the acceptor peptide (Ton-That et al., 2000; Huang et al., 2003). Recently, Aulabaugh et al. (2007) were able to show the formation of a competent acyl-enzyme intermediate during the overall transpeptidation reaction using a reverse-HPLC assay. Further detailed kinetic studies are consistent with a ping-pong hydrolytic shunt mechanism that invokes reverse protonation of His\(^{120}\) and Cys\(^{184}\) for the catalytic mechanism of SrtA acylation. These data support a mechanistic model in which the nucleophilic Cys\(^{184}\) thiolate attacks on the scissile peptide bond between threonine and glycine of the LPXTG motif, resulting in a short-lived tetrahedral intermediate. His\(^{120}\) is hypothesized to protonate the substrate leaving group, facilitating the collapse of the tetrahedral intermediate and formation of the acyl-enzyme, while Arg\(^{197}\) was proposed to stabilize the oxyanion in the transition state (Frankel et al., 2005, 2007). The acyl-enzyme intermediate is then resolved by the nucleophilic attack of the amino group of the pentaglycine side chain, thereby regenerating the enzyme-active site and tethering surface protein to cell wall fragments.

**Multiple sortases, substrate specificity and protein function**

Sortase homologs have been found in all available Gram-positive bacterial genomes, and in most cases, more than one sortase gene has been identified (Pallen et al., 2001; Comfort & Clubb, 2004; Dramsi et al., 2005). Our previous bioinformatic analysis of 61 sortases from complete Gram-positive genomes suggested the existence of four distinct classes of sortases named A, B, C and D (Dramsi et al., 2005). Multiple sequence alignments using the CLUSTAL X program revealed that the three amino acid residues thought to be involved in the formation of the catalytic site (His\(^{120}\)-Cys\(^{184}\)-Arg\(^{197}\)) were conserved in all 61 sortases (numbering is according to the \( S. aureus \) SrtA sequence). All 61 sortases also possess a signal peptide and each class of sortases possesses a specific pattern of conserved amino acids (Fig. 2) (Dramsi et al., 2005). Sortase homologs have also been found in five species of Gram-negative bacteria: *Colwellia psychrerythraea*, *Microbulbifer degradans*, *Bradyrhizobium japonicum*, *Shewanella oneidensis* and *Shewanella putrefaciens* (Comfort & Clubb, 2004). These bacteria encode a single sortase homolog and a single CWS-containing substrate. Both genes are adjacent and located in a mobile element in *Shewanella*, suggesting a possible acquisition from another organism.

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. This may explain the low percentage of identity within the same group of sortases and may suggest that sortases have coevolved with the peptidoglycan of a given bacterium. For example, the SrtA from \( S. aureus \) has been shown to be specific for diglycine nucleophile (Huang et al., 2003), whereas in *Listeria monocytogenes* the LPXTG proteins are attached to meso-diaminopimelic acid (mesoDAP) residue within its cell wall (Dhar et al., 2000). Similarly, since the discovery of the CWS, it has become increasingly

**Fig. 2.** 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 (His\(^{120}\)) and D3 (Cys\(^{184}\)) 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.
apparent that variation within this motif exists (Comfort & Clubb, 2004). In addition, the amino acid composition and length of the transmembrane part or the charged tail constituting the CWS vary between different Gram-positive bacteria. These observations suggest a coevolution of substrate(s)–enzyme pairs.

The sortase A class

The sortase A class that includes the prototype SrtA from *S. aureus* is constituted of unique SrtA representatives from nearly all low GC% Gram-positive bacteria with two exceptions: *Enterococcus faecalis* strain V583, in which two copies of srtA have been found (Dramsi et al., 2005), and *Streptococcus thermophilus* strains LMG18311 and CNRZ1066, which do not contain a functional srtA gene. Interestingly, the genome of *S. thermophilus* strain LMG18311 encodes at least 14 pseudogenes of orthologous LPXTG proteins, suggesting an ongoing regressive evolution process in this specialized bacterium dedicated to grow in milk (Hols et al., 2005). All SrtA sortases contain the catalytic TLXTC signature sequence where X is often a valine, an isoleucine or a threonine. The gene-encoding *srtA* is generally not clustered with any of its substrates and is constitutively expressed. SrtA is a broad-range enzyme required for anchoring the majority or all of the LPXTG-containing proteins of a given bacteria.

The sortase B class

The sortase B class, which includes the prototype SrtB from *S. aureus*, is the smallest group with unique representatives in *L. monocytogenes*, *B. anthracis*, *Bacillus cereus*, *S. pyogenes* and *Clostridium perfringens*. All SrtB sortases contain three specific amino acid segments (B1, B2 and B3; Fig. 2) that are not present in SrtA enzymes and the catalytic TLXTC motif, in which X is usually a serine (Dramsi et al., 2005). SrtB functions analogously to SrtA with two key differences in *S. aureus*. First, SrtB has a specificity profile different from that of SrtA. The SrtB of *S. aureus* selectively anchors the unique NQPTN-containing protein IsdC, and does not anchor LPXTG-containing proteins (Mazmanian et al., 2002). Second, the acceptor structure for SrtB is not Lipid II (a peptidoglycan precursor) but is the uncross-linked Gly3 side chains in mature peptidoglycan from the staphylococcal cell wall (Marraffini & Schneewind, 2005). Another difference is that the gene-encoding *srtB* and its putative substrate(s) are often part of the same operon. In *L. monocytogenes*, the SrtB-containing operon has been characterized and two surface proteins were identified in *srl* as putative SrtB substrates: SvpA (IsdC homolog) and Lmo2186. SvpA has been shown to be anchored to the cell wall by SrtB (Bierne et al., 2004). A nongel proteomic analysis elegantly identified SvpA and Lmo2186 as SrtB substrates in *L. monocytogenes*. In addition, the analysis of the peptides identified in these proteins suggests that SrtB of *L. monocytogenes* may recognize two different sorting motifs: NXZTN and NPKXZ (Pucciarelli et al., 2005). Recently, the srtB–*isd* locus in *B. anthracis* has been characterized (Maresso et al., 2006). SrtB of *B. anthracis* was shown to be required for cell wall anchoring of IsdC by recognizing its NPKTG sorting motif (Maresso et al., 2006). Thus, the organization of the *srtB–isd* locus of *S. aureus* appears to be partially conserved in the pathogenic *L. monocytogenes*, *B. anthracis*, *B. cereus*, in the nonpathogenic *Listeria innocua*, as well as in the Gram-positive extremophile *Bacillus halodurans*. In all these cases, the putative SrtB targets contain one or several NEAT domains that are expected to play a role in iron transport. The role of sortase B and its substrate in iron acquisition has been demonstrated in both *S. aureus* and *B. anthracis* (Mazmanian et al., 2003; Maresso et al., 2006). In contrast, the svpA–srtB locus plays no role in iron utilization by *L. monocytogenes* (Newton et al., 2005; Jin et al., 2006). Of note, the *srtB* genes of *S. pyogenes* and *C. perfringens* are apparently not linked physically to iron transport-related genes. Finally, all *Isd* loci characterized to date contain several substrates for the house-keeping class A sortase and one or two substrate(s) for the class B sortase, suggesting that sortase A and sortase B may work in tandem.

The crystal structures of sortase B from *S. aureus* and *B. anthracis* revealed an overall structure close to that of SrtA (Zhang et al., 2004; Zong et al., 2004b). Interestingly, replacement of the β6/β7 loop in SrtA with the corresponding loop from SrtB was found to be sufficient to change the specificity profile of SrtA, i.e. allowing NQPTN recognition and the acylation step. However, this engineered enzyme was unable to perform the transpeptidation step, suggesting that additional specificity determinants are required for the complete reaction (Bentley et al., 2007).

The sortase C class

The sortase C class is the largest and most heterogenous group of accessory sortases. These sortases are often present in several copies in the genome, clustered with their substrates and associated with mobile genetic elements. Remarkably, only sortases of class C displayed a hydrophobic C-terminal domain (Dramsi et al., 2005). Sortases C are found in both low GC% (streptococci, bacilli, enterococci) and high GC% Gram-positive bacteria (*Actinomyces* or corynebacteria). It has been known for a long time that *Actinomyces* species elaborate two different types of fimbriae composed of either FimA or FimP, two LPXTG-containing proteins (Wu & Fives-Taylor, 2001). The genes encoding the fimbrial proteins are always found adjacent to a gene encoding a sortase C enzyme. Knockout mutations of...
**Actinomyces naeslundii** srt homologs abolish fimbriae formation. These early studies provided the first clue that the fimbriae/pilus assembly in Gram-positive bacteria might be catalyzed by sortase.

Unlike SrtA and SrtB, SrtC are polymerizing enzymes that are involved in the formation of pili or fimbria in several Gram-positive bacteria. SrtC catalyzes the cross-linking between two pilin subunits. This is the first example of sortase involving only polypeptides as substrates of the enzymatic reaction instead of the protein–peptidoglycan pair for class A and class B sortases. The role of class C sortases has first been demonstrated in *Corynebacterium diphteriae* (Ton-That & Schneewind, 2003). *Corynebacterium diphteriae* strain NCTC13129 possesses three distinct operons involved in pilus biosynthesis, all displaying similar organization with three LPXTG-containing proteins and one or two class C sortases (Ton-That & Schneewind, 2004). The first operon involved in pilus biosynthesis encodes three LPXTG-containing proteins named SpaA, SpaB and SpaC and one class C sortase unfortunately named SrtA. Immunogold labeling of *C. diphteriae* strain NCTC13129 with specific antibodies raised against recombinant SpaA, SpaB and SpaC, followed by electron microscopy revealed pili on the bacterial surface. SpaA represents the major pilin whereas SpaB and SpaC are minor pilins that are incorporated into the SpaA pilus shaft, SpaB being found at regular intervals and SpaC being found at the tip. Deletion of the major pilin gene or the sortase gene completely abolished the assembly of SpaA pili as well as staining with SpaB and SpaC antibodies. In contrast, deletion of the minor pilin genes (*spaC and spaB*) did not abrogate SpaA pilus formation. Western blotting of cell wall protein extracts with anti-SpaA antibody revealed the presence of high-molecular-weight species that could be separated on 3–8% gradient sodium dodecyl sulfate-polyacrylamide gel electrophoresis as well as that of monomeric pilin. In the extracts from the sortase mutant, only the monomeric pilin could be detected. Similar results were found with antibodies against SpaB and SpaC, the minor pilins.

Pairwise comparison of the amino acid sequence of SpaA, FimA and FimP identified two conserved elements: (1) the sequence WxxxVxVYPK named the ‘pilin motif’ and (2) the sequence YxLxETxAPxGY designated the ‘E-box’ (Ton-That et al., 2004b). Alanine substitution experiments revealed that the lysine 490 of the SpaA pilin motif is absolutely critical for SpaA polymerization. Mutations of the SpaA CWS LPLTG resulted in a similar phenotype. Amino acid substitutions of the E-box did not abrogate the formation of SpaA pilus but prevented the incorporation of SpaB into the SpaA pilus. As mentioned previously, a reaction catalyzed by sortases involves a nucleophilic attack of amino groups on thioester acyl-enzyme intermediates to form a transpeptide bond between its two substrates. The absolute requirement for the conserved lysine residue suggested the following model to explain the covalent assembly of pilin subunits. The pilin-specific sortase first cleaves the LPXTG motif of the first pilin monomer and forms an acyl-enzyme intermediate, which is in turn resolved by the nucleophilic attack of the free amino group of the conserved lysine provided by the second monomer, thereby cross-linking two adjacent pilin subunits via a transpeptide bond. It was elegantly shown that a chimeric protein, here staphylococcal enterotoxin B, containing at its N-terminus the first 200 amino acid sequence of SpaA providing the signal peptide and pilin motif and at its C-terminus the CWS of SpaA, is able to form polymers when expressed in corynebacteria. Thus, the pilin motif and the sorting signal represent the two topogenenic elements that are necessary for pilin polymerization by class C sortases (Ton-That et al., 2004b).

This mechanism of pilus assembly catalyzed by class C sortases has now been demonstrated in several Gram-positive pathogens using similar genetic and biochemical analyses (Telford et al., 2006). The existence of pili has been demonstrated in *Streptococcus agalactiae* (Group B Streptococci) (Lauer et al., 2005; Dramsi et al., 2006; Rosini et al., 2006; Maisey et al., 2007), S. pneumoniae (Barocchi et al., 2006; LeMieux et al., 2006), *S. pyogenes* [Group A Streptococci (GAS)] (Mora et al., 2005; Abbot et al., 2007; Manetti et al., 2007), E. faecalis (Nallapreddy et al., 2006; Tendolkar et al., 2006) and *A. naeslundii* (Mishra et al., 2007).

A singular case is that of *S. pyogenes* (GAS). In this human pathogen, pili components are encoded in the structurally and sequentially variable FCT (for fibronectin-, collagen-binding proteins and T-antigen) region (Bessen & Kalia, 2002; Mora et al., 2005). It is now shown that the T-antigen constitutes the pilus shaft protein. Recently, it was shown that *S. pyogenes* strains can be divided into at least six different groups based on their FCT regions (Kratovac et al., 2007). In all FCT groups, except FCT-1 (M6 strain), the gene encoding the pilus shaft protein (T-antigen) displayed a QVPTG or EVPTG sorting motif that is reminiscent of that found in SrtB substrates (Barnett et al., 2004; Scott & Zahner, 2006). **BLAST** analysis of the pilus-associated sortases, named *srtC1* or *srtC2*, revealed their belonging to class B sortases (our unpublished observations). In contrast, in M6 strain, the *tee6* gene encoding the major pilin possessed the canonical LPXTG sorting motif, and the pilus-specific sortase, unfortunately named SrtB, is a typical class C sortase (Barnett & Scott, 2002). Thus, *S. pyogenes* appears as an exception to the rule in which class B sortases are responsible for pilus polymerization. Interestingly, in GAS M49 strain, an additional gene named *lepA*, encoding a protein homologous to signal peptidase I, is found clustered with the pilus and the pilus-specific sortase genes. Interestingly, this gene is also found in the second pilus operon PI-2b of *S. agalactiae* strain A909 (Rosini et al., 2006; Telford...
et al., 2006). Very recently, Nakata et al. (2007) showed that leptA, like the pilin-specific sortase, is crucial for pilus assembly (ASM 2007, poster D126). In both S. agalactiae PI-2b and 4 in the S. pyogenes FCT-3 region, the genes encoding the major pilins, sbp1 and fcta, respectively, do not contain the canonical pilin motif. It is therefore possible that pilus formation occurs through a different mechanism in these cases.

The sortase D class

The sortase D class includes 14 sortases originating from high and low GC% Gram-positive bacteria and it can be divided into three subclusters reflecting host phylogeny (Bacilli, Clostridia and Actinomycetales). Two studies in the high GC% Gram-positive Streptomyces coelicolor suggest that these enzymes are functional and play a role under specific developmental conditions, i.e. mycelium formation in S. coelicolor. Indeed, three proteins containing the LPXTG motif in S. coelicolor belong to a novel class of cell-surface proteins named ‘chaplins.’ These hydrophobic proteins are thought to lower the water surface tension, thus allowing the emergence of an aerial hyphae (Claessen et al., 2003; Elliot et al., 2003). A recent study confirms the role of a class D sortase in a developmental process in B. anthracis (Marraffini & Schneewind, 2006). In addition to the class A (srtA) and class B (srtB) sortases, a third sortase-like gene, called srtC and belonging to the class D sortases, was located in the four gene operon basl–srtC–sctR–sctS, where basl encodes for a surface protein with an LPNTA sorting motif and sctR–sctS encode a regulatory two-component regulatory system. Another gene, basH, encoding a surface protein with an LPNTA sorting motif, is found elsewhere in the B. anthracis genome. Marraffini and Schneewind showed that SrtC can cleave BasH and Basl at the LPNTA motif, thereby targeting both polypeptides at the cell wall of sporulating bacilli. The sortase mutant displayed a defect in spore formation in both tissues and blood. Thus, the class D sortase of B. anthracis plays a role in a developmental process, here sporulation.

Anchoring of the Braun lipoprotein to the peptidoglycan of E. coli

Structure and localization of the Braun lipoprotein

In E. coli, the peptidoglycan and the outer membrane are linked by one of the major lipoproteins known as the Braun lipoprotein, which is also referred to as the murein lipoprotein or Lpp (Braun & Rehn, 1969; Braun, 1975). The prolipoprotein is composed of 78 amino acids, which include an N-terminal peptide signal of 20 amino acids. As most lipoproteins, Lpp harbors a particular signal sequence containing at its C-terminus the conserved lipobox (Leu–3-Ser/Ala–2-Ala/Gly–1-Cys–1). The first step in the biosynthesis of lipoproteins is the transfer of the diacylglycerol moiety from phosphatidylglycerol to the sulfhydryl group of the invariant cysteine residue. This reaction is catalyzed by lipoprotein diacylglycerol transferase encoded by the lgt gene. The lipoproteins are translocated across the cytoplasmic membrane through the Sec pathway. The specific lipoprotein signal peptidase (Lsp), also known as signal peptide II, cleaves the peptide bond between the amino acid at position −1 and the lipid-modified cysteine residue (Tokunaga et al., 1982). Lgt and Lsp are highly conserved enzymes in bacteria. Interestingly, both enzymes appear to be essential in Gram-negative bacteria but dispensable in Gram-positive bacteria (Leskela et al., 1999; Tjalsma et al., 1999).

Lipoproteins are further processed in Gram-negative bacteria by a third enzyme designated lipoprotein N-acyltransferase Lnt. Lnt catalyzes the addition of an N-acyl group to the diacylglycerol cysteine. This modification is necessary for efficient recognition of outer membrane lipoproteins by the Lol system, which transports them from the plasma membrane to the outer membrane (Narita et al., 2004; Robichon et al., 2005). The residue at position +1 of the mature 58-amino acid protein is thus a glycerylcysteine that carries three fatty acids (Hantke & Braun, 1973). The 1.9 Å resolution crystal structure of the polypeptide comprising residues 2–57 of the mature lipoprotein folds into a long parallel coiled-coil trimer with distinctive amino- and carboxy-terminal capping structures (Shu et al., 2000). The trimeric structure was also demonstrated in vivo based on chemical cross-linking of the native lipoprotein with several reagents (Choi et al., 1986). The full-length mature lipoprotein is localized in the outer membrane but it remains unclear how the trimer penetrates the lipid bilayer. The N-terminal glycerylcysteine residue is essential for membrane association (Choi et al., 1986), and the serine residue present at position +2 functions as the outer membrane sorting signal (Yakushi et al., 1997). The Braun lipoprotein is immunogenic and antigenic only in cells exhibiting an abnormal outer membrane structure (Braun, 1975), suggesting that the protein does not extend at the bacterial surface. Interestingly, the Braun lipoprotein is found either freely associated to the outer membrane or covalently bound via its C-terminal residue to the peptidoglycan, approximately at a ratio of one covalently linked molecule to two freely associated molecules (Hirashima et al., 1973; Braun, 1975). This ratio varies during bacterial growth, the bound form increasing during the stationary phase (Pisabarro et al., 1985).

Role of the Braun lipoprotein

Deletion of the lpp gene is not lethal in E. coli but leads to defects in the permeability barrier of the outer membrane
Covalent attachment to the peptidoglycan

In Gram-positive bacteria, $\alpha,\beta$-transpeptidases catalyze formation of 3-3 cross-links between peptidoglycan stem peptides (Fig. 3a) instead of the classical 4-3 cross-links formed by the $\alpha,\beta$-transpeptidase activity of penicillin-binding proteins (PBPs) (Mainardi et al., 2008). The peptidoglycan cross-linking activity of $\alpha,\beta$-transpeptidases has been originally identified in Enterococcus faecium as a bypass of the PBPs that confers a high level of $\beta$-lactam resistance (Mainardi et al., 2005). The catalytic domain of the $\alpha,\beta$-transpeptidase from E. faecium (Ldt$_{fae}$) is the first functionally characterized member of a conserved protein family designated ErfK–Ycfs–YhnG or Pfam 03734 in databases. Additional members of this family from E. faecalis (Ldt$_{ba}$) and B. subtilis (Ldt$_{sub}$) were also subsequently shown to catalyze peptidoglycan cross-linking in vitro. Determination of the substrate specificity of these three enzymes indicated that diversification of the structure of peptidoglycan precursors associated with speciation has led to a parallel evolution of the substrate specificity of the $\alpha,\beta$-transpeptidases, affecting mainly the recognition of the acyl acceptor (Magné et al., 2007a). The ErfK–Ycfs–YhnG conserved domain is widely spread among bacteria because it is found in proteins from more than 300 species belonging to various bacterial phyla, including the major Proteobacteria, Firmicutes and Actinobacteria, but their functions remain to be explored.

The covalent linkage between the Braun lipoprotein and the peptidoglycan has been characterized decades ago from analysis of peptidoglycan fragments (muuropeptides) generated by lysozyme and pronase digestion (Braun & Wolff, 1970). The peptide bond links the $\varepsilon$-amino group of the C-terminal lysine residue of the Braun lipoprotein to the $\alpha$-carboxyl of mesoDAP present at the third position of the stem peptide of a peptidoglycan subunit (Fig. 3b). Recently, in an attempt to identify the enzyme responsible for 3-3 cross-links ([mesoDAP$\rightarrow$mesoDAP] formation), four genes encoding homologs of the $\alpha,\beta$-transpeptidases from Gram-positive bacteria have been identified and deleted from the E. coli chromosome (Magné et al., 2007b). Surprisingly, the peptidoglycan of the resulting quadruple mutant still contained 3-3 cross-links whereas covalent anchoring of the Braun lipoprotein was abolished. Genetic transcomplementation showed that three out of the four deleted genes, erfK, ycfs and ybiS, were able to restore independently the anchoring of the Braun lipoprotein to the peptidoglycan, indicating that three enzymes can catalyze the transpeptidation reaction in vivo. However, this function appears to be mainly performed by YbiS in wild-type E. coli because deletion of the ybiS gene alone abolished covalent linkage of the Braun lipoprotein to peptidoglycan almost completely (Magné et al., 2007b). MesoDAP$\rightarrow$mesoDAP cross-links were increased upon production of the remaining protein, YhnG, suggesting that it catalyzes peptidoglycan cross-linking in addition to the unknown protein active in the quadruple mutant.

A two-step reaction has been proposed for erfK, ycfs and ybiS. In the first step, the enzymes cleave the peptide bond between mesoDAP at position 3 and D-Ala at position 4 of a
Peptidoglycan-linked proteins

(a) **Acyl acceptor**
Peptidoglycan subunit

- NAG -NAM
- L-Ala
- d-Gln
- L-Lys - d-iAsx
- d-Ala

(b) **Acyl acceptor**
Braun lipoprotein

- FA 2'-gliceryl -Cys 1
- Ser 2
- (residues 3 - 57)
- Lys 58

(c) **Acyl acceptor**
Peptidoglycan precursor

- C55
- NAG -NAM
- L-Ala
- d-Glu
- L-Lys - (Gly)₅
- d-Ala
- d-Ala

(d) **Acyl acceptor**
Pilin

- N-ter
- Lys
- Leu
- Pro
- X
- Thr
- Thr
- Gly
- C-ter

**Acyl donor**
Peptidoglycan subunit

- NAG -NAM
- L-Ala
- d-Gln
- L-Lys - d-iAsx
- d-Ala

- Lys - d-iAsx
- d-Ala

- Lys - d-iAsx
- d-Ala

- (NH₂ - CO)

- + d-Ala

- Lys - d-iAsx
- d-Ala

- (NH₂ - CO)

- + d-Ala

**Acyl donor**
Peptidoglycan subunit

- FA 2'-gliceryl -Cys 1
- Ser 2
- (residues 3 - 57)
- Lys 58
- mesoDAP
- d-Ala

**Acyl donor**
Peptidoglycan subunit

- FA 2'-gliceryl -Cys 1
- Ser 2
- (residues 3 - 57)
- Lys 58
- mesoDAP
- d-Ala

**Acyl donor**
Surface protein

- N-ter
- Leu
- Pro
- X
- Thr
- Gly
- C-ter

**Acyl donor**
Pilin

- N-ter
- Lys
- Leu
- Pro
- X
- Thr
- Gly
- C-ter

**SrtA** from S. aureus

- N-ter
- L-Ala
- d-Glu
- L-Lys - (Gly)₅
- d-Ala

- (NH₂ - CO)

- + Gly

**SrtC**

- N-ter
- Lys
- Leu
- Pro
- X
- Thr
- Gly
- C-ter
peptidoglycan stem peptide that acts as the acyl donor (Fig. 3b). In the second step, a peptide bond is formed between the α-carbonyl of mesoDAP and the side chain amine of the C-terminal Lys⁸⁹ of the Braun lipoprotein.

The crystal structures of the catalytic domain of Ldtfm and Ldtbg (also named YkuD) have been solved and define a new globular fold formed by two α-helices and eight β-strands (Biarrotte-Sorin et al., 2006; Bielnicki et al., 2006). Three residues, Cys⁴⁴², His⁴²¹ and Asp⁴⁵² (numbering according to the sequence of Ldtfm), form the putative catalytic triad of the 1,3-transpeptidase (Biarrotte-Sorin et al., 2006). The catalytic cysteine and histidine residues are conserved in the four 1,3-transpeptidase homologs of E. coli.

Various experimental approaches have shown that sortases anchor surface proteins to lipid intermediate II before the transglycosylation reaction (Fig. 3c). In contrast, it is not known whether the Braun lipoprotein is attached to the lipid intermediate or to polymerized disaccharide–peptide subunits. For this protein, the whole sequence order of events leading to oligomerization, outer membrane localization and covalent anchoring to the peptidoglycan remains undetermined. These events can occur, at least in part, independently, because trimer formation does not require posttranslational modification, membrane association or covalent linkage to the peptidoglycan (Choi et al., 1986). In addition, amino acid substitutions in the sorting signal do not abolish covalent anchoring to the peptidoglycan although they lead to mislocalization into the inner membrane (Ichihara et al., 1982; Yakushi et al., 1997).

Homologs of the Braun lipoprotein are present in most enterobacteria. To date, it is the only protein that has been shown to be linked covalently to the peptidoglycan in Gram-negative bacteria. However, the redundancy of the 1,3-transpeptidases in E. coli suggests the existence of other targets in vivo.

Conclusions

The formation of 3-3 cross-links in the peptidoglycan of Gram-positive bacteria and the anchoring of the Braun lipoprotein to the peptidoglycan of E. coli are both catalyzed by related 1,3-transpeptidases. The two reactions involve a similar acyl donor substrate, the peptide stem of a peptidoglycan subunit that is cleaved between the L-diamino acid at position 3 and D-alanine at position 4, but involve different amino acyl acceptor substrates (Fig. 3a and b). For the crosslinking enzymes, the acyl acceptor is the amino group at the third position of another peptidoglycan subunit whereas for YbiS it is the amino group of the C-terminal L-Lys of the Braun lipoprotein. The acyl donor of the transpeptidation reaction catalyzed by the sortases is the surface protein that will be anchored to the peptidoglycan after cleavage of the sorting signal (Fig. 3c). The reacting amino group is located at the third position of a peptidoglycan precursor that plays the role of the acyl acceptor for class A and class B sortases, whereas a second protein can play the role of the acyl acceptor for class C sortases (Fig. 3d). Strikingly, both the sortases and the 1,3-transpeptidases use intermediate covalent enzyme thioesters to activate the acyl donor before transfer although they are structurally unrelated.

Acknowledgements

We thank Agnès Fouet and Patrick Trieu-Cuot for critically reading the manuscript. This work was supported by the program ‘ACI Microbiologie 2003’ from the Fonds National de la Science (Grant ACIM-4-9), the European Community (COBRA, contract LSHM-CT-2003-503335, 6th PCRD), the National Institute of Allergy and Infectious Diseases (Grant R01 AI45626), the Pasteur Institute (GPH No. 9) and the Centre National de la Recherche Scientifique (CNRS).

References


Bessen DE & Kalai A (2002) Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-


Ton-That H, Liu G, Mazmanian SK, Faull KF & Schneewind O (1999) Purification and characterization of sortase, the
transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci USA* **96**: 12424–12429.


