Current Topics in Microbiology and Immunology

Fabio Bagnoli
Rino Rappuoli *Editors*

Protein and Sugar Export and Assembly in Gram-positive Bacteria

Springer

Current Topics in Microbiology and Immunology

Volume 404

More information about this series at http://www.springer.com/series/82

Fabio Bagnoli · Rino Rappuoli
Editors

Protein and Sugar Export and Assembly in Gram-positive Bacteria

Responsible Series Editor: Rino Rappuoli

Springer

Editors
Fabio Bagnoli
GSK Vaccines
Siena
Italy

Rino Rappuoli
GSK Vaccines
Siena
Italy

ISSN 0070-217X ISSN 2196-9965 (electronic)
Current Topics in Microbiology and Immunology
ISBN 978-3-319-56012-0 ISBN 978-3-319-56014-4 (eBook)
DOI 10.1007/978-3-319-56014-4

Library of Congress Control Number: 2017934881

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Declaration of Interest

Fabio Bagnoli and Rino Rappuoli are employees of GSK Vaccines and own GSK stocks and patents on vaccines against Gram positive pathogens. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Authorship

Fabio Bagnoli and Rino Rappuoli were involved in the conception and design of the book and approved its content before publication.

Preface

This book provides an overview of the current knowledge on the envelope structures of Gram-positive bacteria, their biosynthesis and assembly, their functions as well as their role as antibacterial targets and in biotechnology applications. This is a concise volume containing eleven chapters, from renowned experts in the field, reviewing recent findings and knowledge on very diverse arguments and at the same time linked to each other. That is the uniqueness behind a book like this and the added value towards a search in literature databases.

The cell envelope of these bacteria includes surface proteins, capsular polysaccharides, peptidoglycan, teichoic acids, and phospholipids. These components play key roles in cell viability, virulence and evasion of host defences. Many virulence factors of pathogenic species reside on the bacterial surface. Surface proteins have very diverse functions (e.g., adhesion, invasion, signalling, conjugation, interaction with the environment and immune-evasion). On the other hand, polysaccharides often play a mechanical protective role for the bacterium and the remarkable structural diversity in capsular polysaccharides favours immune evasion. Peptidoglycan is a well-established target for antibiotics and can undergo modification to decrease susceptibility to the drugs.

Both surface proteins and sugars, being the most exterior components, are also accessible to antibodies and represent important vaccine targets. Certain proteins assemble into complexes forming secretion apparatuses, such as the type VII secretion system, pili (or fimbrae) and flagella. These macromolecular structures have very diverse functions, which include secretion, conjugation, adhesion, biofilm formation and motility. Obviously, different species have different envelope structures and the knowledge on most important species (e.g., Actinomyces spp., Bacillus spp., Clostridium spp., Enterococcus spp., Streptococcus spp., and Staphylococcus spp.) is rapidly increasing.

Given the complexity and breath of the literature behind this argument we decided to write this book in the attempt to give an overview of the current knowledge on the envelope structures of Gram-positive bacteria, their biosynthesis, and functions. Secretion systems, spatial organization of cell wall-anchored proteins and bioinformatic algorithms for predicting subcellular localization of proteins are

explained in a simple but detailed fashion. Assembly mechanisms of structures such as pili and sugar polymers are described along with the recently discovered Type VII secretion system. The latter one has been described in low-GC Gram-positive bacteria and they can show a very complex organization with up to five chromosomal-encoded systems (ESX-1 to ESX-5) in mycobacteria to a much simpler organization in Firmicutes.

Finally, relevant examples of applied science which exploit knowledge on Gram-positive bacteria are also included. Possible targets for new antimicrobials are noted. We highlighted the development of the Twin-arginine protein translocation system (Tat) for the biotechnological secretion of fully folded and co-factor-containing proteins and its potential use as an anti-microbial drug target. The use of these bacteria in biotechnology for the production of heterologous proteins and methodologies for analyzing surface and secreted proteins with a particular emphasis to vaccine antigen discovery are also discussed.

In conclusion, this book is useful to any researcher, clinician or technician who is involved with basic or applied science projects on Gram-positive bacteria.

Siena, Italy

Fabio Bagnoli
Rino Rappuoli

Contents

Envelope Structures of Gram-Positive Bacteria

Mithila Rajagopal and Suzanne Walker

Abstract Gram-positive organisms, including the pathogens *Staphylococcus aureus, Streptococcus pneumoniae,* and *Enterococcus faecalis,* have dynamic cell envelopes that mediate interactions with the environment and serve as the first line of defense against toxic molecules. Major components of the cell envelope include peptidoglycan (PG), which is a well-established target for antibiotics, teichoic acids (TAs), capsular polysaccharides (CPS), surface proteins, and phospholipids. These components can undergo modification to promote pathogenesis, decrease susceptibility to antibiotics and host immune defenses, and enhance survival in hostile environments. This chapter will cover the structure, biosynthesis, and important functions of major cell envelope components in gram-positive bacteria. Possible targets for new antimicrobials will be noted.

Abbreviations

PG	Peptidoglycan
MRSA	Methcillin-resistant *Staphylococcus aureus*
GlcNAc	*N*-acetylglucosamine
GalNAc	*N*-acetylgalactosamine
MurNAc	*N*-acetlymuramic acid
PBP	Penicillin-binding protein
PGT	Peptidoglycan glycosyltransferase
Und-P	Undecaprenyl phosphate
TA	Teichoic acid

M. Rajagopal · S. Walker (✉)
Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA
e-mail: suzanne_walker@hms.harvard.edu

M. Rajagopal
Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

Current Topics in Microbiology and Immunology (2017) 404:1–44
DOI 10.1007/82_2015_5021

Published Online: 16 February 16

WTA	Wall teichoic acid
LTA	Lipoteichoic acid
CPS	Capsular polysaccharides
PIA	Polysaccharide intercellular adhesin

Contents

1 Introduction

The cell envelope is a complex, dynamic, multilayered structure that serves to protect bacteria from their unpredictable and often hostile surroundings. The cell envelopes of most bacteria fall into one of two major groups. Gram-negative bacteria have an inner, cytoplasmic membrane surrounded by a thin layer of peptidoglycan (PG) and an outer membrane containing lipopolysaccharide. The outer membrane functions as a permeability barrier to control the influx and egress of ions, nutrients, and environmental toxins, and it also contributes to osmoprotection. Gram-positive bacteria lack a protective outer membrane but the PG layers are many times thicker than those in gram-negative organisms (Silhavy et al. 2010; Vollmer et al. 2008). Embedded in the inner membrane and attached to the PG layers are long anionic polymers called teichoic acids (TAs), which play multiple roles in cell envelope physiology as well as pathogenesis (Brown et al. 2013; Percy and Gründling 2014; Schneewind and Missiakas 2014). Membrane-embedded and wall-associated

proteins serve as environmental sensors, regulate passage of nutrients and ions across the cytoplasmic membrane, facilitate efflux of toxins and other molecules, modulate surface adhesion, and participate in enzymatic synthesis, degradation, and remodeling of the cell envelope during growth and division, and in response to environmental stress (Buist et al. 2008; Kovacs-Simon et al. 2011; Navarre and Schneewind 1999; Stock et al. 2000; Zhen et al. 2009). Other important cell envelope components in gram-positive organisms include capsular polysaccharides (CPS), which are covalently attached to PG, and extracellular polysaccharides, which form an amorphous outer layer (Arciola et al. 2015; Yother 2011).

The importance of the cell envelope for bacterial survival makes it a target for antibiotics, and several classes of clinically used antibiotics inhibit biosynthesis of PG, resulting in osmotic rupture. Other antibiotics damage the membrane barrier (Walsh 2003). Because resistance to clinically used antibiotics has become widespread, there is a push to better understand cell envelope biogenesis and regulation, and to identify new cell envelope targets that can be exploited in the development of next-generation antibiotics. In this chapter, we will focus on important cell envelope components of gram-positive pathogens using *Staphylococcus aureus* as a focal point, except where other gram-positive pathogens are better studied. Attention will also be given to the nonpathogenic *Bacillus subtilis* because its genetic tractability and other biological characteristics have led to its adoption as the principal gram-positive model organism.

2 Cell Membrane

Gram-positive organisms are surrounded by bilayer membranes that can vary substantially in composition but typically include large amounts of phosphatidylglycerol and cardiolipin. In *Bacillus* species, phosphatidylethanolamine is abundant as well (Clejan et al. 1986; Haque and Russell 2004; Minnikin and Abdolraimzadeh 1974). Many gram-positive species express at least one type of aminoacylated phosphatidylglycerol (Epand et al. 2007; Parsons and Rock 2014). For example, in *S. aureus*, lysyl-phosphatidylglycerol is found in significant amounts, particularly during logarithmic growth (Ernst et al. 2009). This phospholipid is synthesized by a polytopic membrane protein, MprF, which catalyzes the transfer of lysine from lysyl-tRNA to phosphatidylglycerol on the inner leaflet of the membrane and then translocates this species to the outer leaflet of the membrane (Ernst et al. 2009; Kristian et al. 2003). Lysyl-phosphatidylglycerol reduces susceptibility to antimicrobial peptides produced during host infection (Peschel et al. 2001) and also provides protection against aminoglycosides, bacitracin, daptomycin, and some β-lactams (Nishi et al. 2004; Komatsuzawa et al. 2001). Daptomycin-resistant *S. aureus* clinical isolates frequently contain mutations that increase MprF expression or translocase activity (Friedman et al. 2006; Julian et al. 2007; Jones et al. 2008; Yang et al. 2009b). Other species of gram-positive bacteria have MprF homologs that have been implicated in similar functions (Ernst and Peschel 2011). It is thought that the positive

charges of lysyl-phosphatidylglycerol serve to repel positively charged antibiotics or antibiotic-metal complexes (Ernst and Peschel 2011; Nishi et al. 2004).

The composition of both the head groups and the fatty acyl chains in membrane phospholipids can change rapidly in response to environmental conditions, such as low pH, osmotic stress, or temperature extremes (Zhang and Rock 2008). For example, branched chain fatty acid content in membranes can vary substantially depending on growth conditions. Membrane lipid composition affects membrane viscosity, which modulates membrane permeability and can influence both solute transport and protein interactions. Membrane lipid homeostasis is thus a crucial process and interfering with it can compromise viability (de Mendoza 2014; Zhang and Rock 2008).

In addition to the lipid components, the cell membrane contains the lipid anchor component of lipoteichoic acid (LTA) and includes numerous transmembrane and

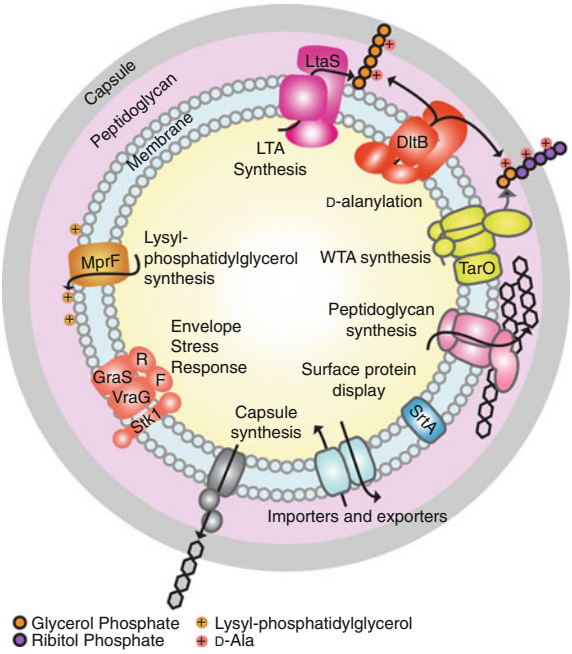


Fig. 1 The gram-positive cell envelope. The complex gram-positive cell envelope is the first line of defense for the organism. Here, the *S. aureus* envelope is shown as an example. Major pathways involved in the synthesis of the cell envelope include capsule, PG, and TA syntheses. TAs can be modified by D-alanylation. D-alanylation and lysyl-phosphatidylglycerol synthesis are known factors for antibiotic resistance. Envelope stress response regulators modulate the organism's response to toxic molecules or conditions that perturb the cell envelope. Importers and exporters, ubiquitously present among bacteria, serve the necessary role of channeling in nutrients and pumping out the toxic molecules. Finally, surface protein display systems function to tether proteins to the cell membrane or cell wall, which perform important roles in adhesion and interaction with the environment

lipoproteins with functions in cell envelope synthesis, transport of cell envelope precursors and nutrients, and export of toxic compounds (Fig. 1). Among these, transmembrane proteins are the sensory components of several two-component sensing systems that regulate the cell's response to external stimuli, including cell density and presence of damaging toxins. For instance, the amount of lysyl-phosphatidylglycerol in *S. aureus* is regulated by a complex of proteins that includes a two-component signaling system, GraRS, and a two-component ABC-transporter-like system, VraFG. This complex, which senses and responds to a variety of stimuli, including the presence of antimicrobial peptides, also regulates D-alanylation of TAs (Falord et al. 2011; Li et al. 2007a, b; Yang et al. 2012). Modulating the negative charge density of the cell envelope through lysinylation of phosphatidylglycerol and D-alanylation of TAs decreases susceptibility of *S. aureus* to antimicrobial peptides produced during host infection and increases resistance to cationic antibiotics administered to treat infection (Ernst and Peschel 2011; Brown et al. 2013; Revilla-Guarinos et al. 2014; Bayer et al. 2013).

3 Peptidoglycan

Gram-positive bacteria are surrounded by many layers of peptidoglycan (PG), which form a protective shell that is 30–100 nm thick (Silhavy et al. 2010). The PG layers are covalently modified with carbohydrate polymers including wall teichoic acids (WTAs) or functionally related anionic glycopolymers as well as CPS. The PG layers also scaffold numerous proteins, some of which are bound non-covalently through interactions with PG-binding modules such as LysM domains (Buist et al. 2008) while others are covalently attached by sortases (Schneewind and Missiakas 2012). Some wall-associated proteins play important roles in cell envelope remodeling during growth and division, whereas others scavenge nutrients and metals from the environment or serve as adhesins that promote surface binding and colonization (Navarre and Schneewind 1999). PG has numerous important functions but perhaps the most important is that it stabilizes the cell membrane, enabling it to withstand high internal osmotic pressures. This function is critical for cell survival because the turgor pressure pushing against the cell membrane can reach 20 atmospheres in some gram-positive bacteria (Mitchell and Moyle 1956; Norris and Sweeney 1993). Since PG is essential for viability and the biosynthetic pathway is highly conserved in gram-positive and gram-negative organisms, PG biosynthesis is a target for many clinically used antibiotics, including β-lactams, which are the most successful class of antibiotics in history, and vancomycin, which is still widely used to treat serious gram-positive infections, including methicillin-resistant *S. aureus* (MRSA) infections.

3.1 Peptidoglycan Structure

PG is composed of linear chains of repeating disaccharide units cross-linked via peptide side chains (Fig. 2). The disaccharide subunit is completely conserved and consists of *N*-acetylglucosamine (GlcNAc) coupled through a β-1,4-linkage to *N*-acetylmuramic acid (MurNAc) (Schleifer and Kandler 1972). The average chain length of the glycan strands can vary considerably across species. In *S. aureus*, the glycan strands are relatively short, averaging 6–18 disaccharide units (Boneca et al. 2000; Ward 1973) while in *B. subtilis,* the glycan chains are much longer. Early measurements of *B. subtilis* glycan strands indicated an average chain length of 54–96 disaccharide units, but more recent experiments using atomic force microscopy to probe size exclusion-purified glycan strands have suggested that glycan chains can reach 5000 disaccharide units in length (Hayhurst et al. 2008; Ward 1973). The longer glycan chains found in *B. subtilis* may be a result of the cylindrical shape, which results in a substantially greater stress imparted on the cylindrical walls compared with the poles (Hayhurst et al. 2008).

MurNAc, a sugar unique to bacteria, contains a C3 lactate group. In nascent (uncross-linked) PG of gram-positive organisms, this group is bonded to the N-terminus of a linear peptide consisting of five amino acids. The first, L-alanine, is typically followed by D-isoglutamine, and the terminal dipeptide is D-Ala-D-Ala.

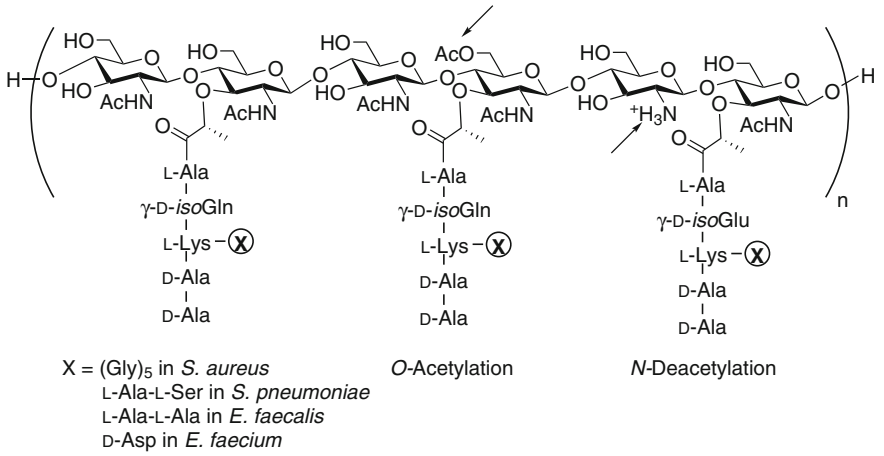


Fig. 2 PG structure and common variations. PG consists of chains of alternating GlcNAc and MurNAc residues. The MurNAc residues are functionalized with pentapeptide units which are cross-linked via the substituents on L-Lys to generate the mature PG. The linear glycan chain is highly conserved across both gram-positives and gram-negatives. The stem pentapeptide is well conserved across gram-positives, aside from *B. subtilis* which contains *meso*-diaminopimelic acid instead of L-Lysine at position 3 of the stem pentapeptide. There is considerable variation in the substituents on the L-Lys across gram-positive species as indicated. PG can be modified by *O*-acetylation of MurNAc or *N*-deacetylation of GlcNAc moieties in response to challenge from antimicrobials such as lysozyme

Position 3 of the pentapeptide chain is either L-lysine or *meso*-diaminopimelic acid (*m*-DAP), with the former being found in *S. aureus*, *Streptococcus pneumoniae*, *Enterococcus faecalis*, and *Enterococcus faecium*, and the latter being found in *B. subtilis* (Schleifer and Kandler 1972). The ε-amino group of L-Lys is typically coupled to one or more additional amino acids. In *S. aureus*, for example, L-lysine is coupled to pentaglycine, although serine can also be incorporated in some strains (De Jonge et al. 1993; Schleifer and Kandler 1972). *S. pneumoniae* and *E. faecalis* contain dipeptide substituents consisting of L-Ala-L-Ser or L-Ala-L-Ala, respectively (De Jonge et al. 1996; Schleifer and Kandler 1972; Severin and Tomasz 1996). *S. pneumoniae* PG is unusual in that it can be a mixture of either dipeptide-substituted or un-substituted stem peptides (Garcia-Bustos et al. 1987; Severin and Tomasz 1996). *E. faecium* contains a D-aspartate substituent (Patti et al. 2008; Vollmer et al. 2008). Canonical glycan strand cross-linking occurs via formation of an amide bond between the side chain or branching peptide on amino acid 3 of one stem peptide and the backbone carbonyl of amino acid 4 on another stem peptide, with the loss of the terminal D-ala (Schleifer and Kandler 1972). Cross-links can also form to the carbonyl of amino acid 3 in some species of gram-positive organisms (Lavollay et al. 2008, 2011; Mainardi et al. 2000; Schleifer and Kandler 1972).

3.2 Peptidoglycan Biosynthesis

PG biosynthesis takes place in distinct stages, the first of which involves assembly of a UDP-MurNAc pentapeptide in the cytoplasm. This stage is followed by coupling of the phospho-MurNAc pentapeptide to the undecaprenyl phosphate (Und-P) "carrier lipid" embedded in the membrane to form a lipid-linked monosaccharide known as Lipid I, which is glycosylated to form the disaccharide Lipid II. Additional amino acids, if any, are appended to the pentapeptide chain at this point and then Lipid II is translocated across the membrane. In the final stage of PG biosynthesis, Lipid II is polymerized and the resulting glycan strands are cross-linked to give mature PG. The lipid carrier released during glycan chain polymerization is recycled back into the cell to continue synthesis. Most of the enzymatic steps for the majority of the biosynthetic pathway are well-conserved across both gram-negative and gram-positive bacteria (Fig. 3).

Assembly of Lipid II:

The first committed step in PG synthesis involves the MurA-catalyzed transfer of enolpyruvate from phosphoenolpyruvate to the C3 hydroxyl of UDP-GlcNAc (Marquardt et al. 1992). Some low GC gram-positive organisms, including *S. aureus, S. pneumoniae,* and *B. subtilis,* contain two *murA* alleles, which are differently regulated (Blake et al. 2009; Du et al. 2000; Kock et al. 2004). The secondary *murA* allele may allow for increased flux into the PG biosynthetic pathway in response to cell wall stress (Blake et al. 2009). MurB reduces the C3 enolate to the lactate,

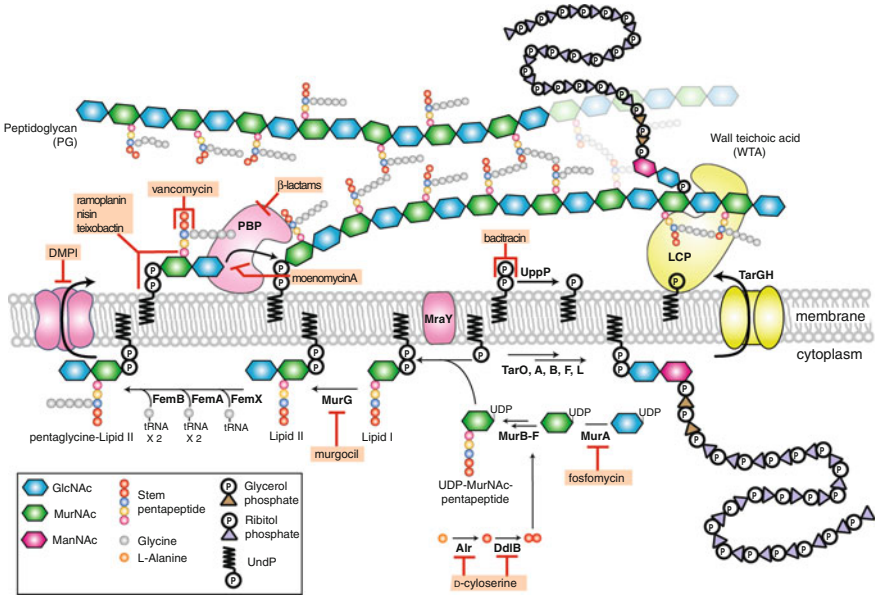


Fig. 3 Synthesis of PG and antibiotics that target PG synthesis. The enzymatic steps for PG synthesis are well conserved across species. Here, the biosynthesis of *S. aureus* PG is shown as an example. The synthesis begins with the assembly of the GlcNAc-MurNAc-pentapeptide and its attachment to carrier lipid Und-P in the cell membrane. After this point, the L-Lysine at position 3 is substituted with additional amino acids and then flipped to the outside of the cell where it is cross-linked by PBPs. The same lipid carrier is also utilized for WTA (shown here) and capsule synthesis. The synthesis of PG is crucial to the cell and over time, several antibiotics have been discovered that target various steps in PG biosynthesis

resulting in the formation of UDP-MurNAc (Benson et al. 1993). The pentapeptide chain is then coupled in a stepwise manner, with MurC, MurD, MurE adding L-alanine, D-glutamic acid, and L-lysine (or *m*-DAP), respectively. Using D-Ala produced from L-Ala by D-alanine racemase (Alr), D-Ala-D-Ala ligase (Ddl) makes the dipeptide, which is then added to the UDP-MurNAc-tripeptide by MurF. Since peptide bond formation is thermodynamically unfavorable, the ligases use ATP to activate the amino acids and provide a driving force for coupling (Bouhss et al. 1997; Patin et al. 2010; Walsh 1989) (Fig. 3).

The next stage of PG synthesis begins with the transfer of phospho-MurNAc pentapeptide to a lipid carrier in the bacterial membrane, typically Und-P, although *Mycobacterium smegmatis* uses decaprenylphosphate (Mahapatra et al. 2005). This step is catalyzed by MraY (Bouhss et al. 2004; Chung et al. 2013; Pless and Neuhaus 1973) and produces the first lipid-linked intermediate, Lipid I. Finally, MurG catalyzes the addition of GlcNAc to give Lipid II (Hu et al. 2003a; Mengin-Lecreulx et al. 1991). Amidation of the α-carboxylate of *iso*-glutamic acid at position 2 of the peptide chain, which is observed in many organisms (Vollmer et al. 2008), most likely occurs intracellularly after lipid-linked PG precursors are

formed. The enzymes involved in this modification were recently identified in *S. aureus* as MurT and GatD (Figueiredo et al. 2012; Münch et al. 2012).

When a peptide branch is present, the required amino acids are usually added to the completed Lipid II moiety. One exception is *Lactobacillus viridescens* where the first amino acid of the L-Ala-L-Ser bridge is added to the UDP-*N*-acetylmuramylpentapeptide (Rogers et al. 1980). In *S. aureus*, the pentaglycine is assembled by FemX, FemA, and FemB, which sequentially add one, two, and two glycines, respectively. These enzymes utilize glycyl-tRNA donors (Henze et al. 1993; Maidhof et al. 1991; Rohrer et al. 1999; Schneider et al. 2004). Serines rather than glycines are incorporated in a similar manner in other staphylococcal strains (Thumm and Götz 1997; Tschierske et al. 1997). This incorporation of serine contributes to resistance to lysostaphin, a glycylglycine endopeptidase (Thumm and Götz 1997). The corresponding enzymes in *E. faecalis* and *S. pneumoniae* have also been identified (Bouhss et al. 2002; Filipe et al. 2000). It is interesting that the Mur ligases use ATP-activated amino acids directly, but the enzymes that assemble the branching peptides use charged tRNAs. When tRNAs were found to be the aminoacyl donors for PG precursors in the 1960s, it caused some excitement because tRNAs were previously known only for their involvement in protein synthesis (Kresge et al. 2007). It is now known that phospholipids as well as PG precursors are aminoacylated by acyl-tRNAs (see above).

The final step in the cytoplasmic phase of PG synthesis involves the translocation of Lipid II across the membrane. This is accomplished by a flippase called MurJ, which was identified only recently (Ruiz 2008, 2009; Sham et al. 2014). In *B. subtilis*, there is also a secondary Lipid II flippase, Amj, that enables survival when MurJ (YtgP) is deleted (Meeske et al. 2015). The complete story of the discovery of the Lipid II flippase has been well-described in the chapter by Lam and coworkers in this volume.

Glycan polymerization and cross-linking:

Once Lipid II is on the outside of the cell, it is polymerized and cross-linked. Glycan polymerization is accomplished by peptidoglycan glycosyltransferases (PGTs; also known as synthetic transglycosylases), while cross-linking is accomplished by transpeptidases. These activities are often found as domains in a single protein, but monofunctional variants of both enzyme classes exist. The nomenclature of PG biosynthetic enzymes is somewhat confusing as many are designated as penicillin-binding proteins, which highlights the fact that they covalently bind β-lactams (Blumberg and Strominger 1974), but obscures their catalytic function, which vary. There are two main categories of PBPs—high-molecular mass PBPs that contain a second domain and low-molecular mass PBPs. The high-molecular mass PBPs are further divided into Class A and Class B PBPs, with the Class A PBPs distinguished by the presence of an N-terminal PGT domain and the Class B PBPs distinguished by the presence of an N-terminal domain of unknown function. The penicillin-binding domains found in both Class A and Class B PBPs functions as transpeptidase domains, serving to cross-link glycan strands. The low-molecular mass PBPs, sometimes called Class C PBPs, typically function as D,D-

carboxypeptidases, serving to hydrolyze the terminal D-alanine of the stem peptide (Ghuysen 1991; Sauvage et al. 2008; Waxman and Strominger 1983). Some organisms including *S. aureus* contain low-molecular mass PBPs that function as transpeptidases, rather than carboxypeptidases. Methicillin-sensitive *S. aureus* (MSSA) strains contain four PBPs. PBP1 and PBP3 are Class B PBPs (Pinho et al. 2000; Wada and Watanabe 1998), PBP2 is a Class A PBP (Pinho et al. 2001a), and PBP4 is a low-molecular weight PBP that acts as a transpeptidase to form additional cross-links in PG (Kozarich and Strominger 1978; Qiao et al. 2014; Wyke et al. 1981). MRSA strains contain an additional PBP, PBP2A, that is highly resistant to β-lactams. PBP2A serves to cross-link PG when the other PBPs have been inactivated by β-lactams (Hartman and Tomasz 1984; Lim and Strynadka 2002). In addition to these enzymes, *S. aureus* also contains two monofunctional transglycosylases, SgtA and MGT (Heaslet et al. 2009; Reed et al. 2011; Terrak and Nguyen-Distèche 2006). Under optimal laboratory growth conditions, only PBP1 and PBP2 are essential for viability (Pinho et al. 2001b; Reed et al. 2015; Wada and Watanabe 1998). It is typical for bacteria to contain multiple PBPs and PGTs, with some essential and others important for survival under stressful conditions. In part, this redundancy reflects the central importance of PG for viability. Rod-shaped organisms such as *B. subtilis* typically have more PBPs than cocci such as *S. aureus* (Zapun et al. 2008). In *B. subtilis,* PG synthesis occurs both at the septum during cell division and along the cylindrical walls during cell elongation, and there is considerable evidence that different biosynthetic machines are involved in these different modes of PG synthesis (Claessen et al. 2008; Daniel et al. 2000; Spratt 1975; Zapun et al. 2008). Deconvoluting the cellular functions of PBPs and other cell wall biosynthetic enzymes has been a major challenge due to redundancy and possible interdependency (Reed et al. 2015; Scheffers and Pinho 2005).

Recycling of carrier lipid:

The Und-P carrier lipid is present in limited amounts in bacterial membranes. In addition to serving as a carrier lipid for PG synthesis, Und-P is a carrier for WTA precursors as well as CPS precursors. To ensure an ongoing supply of all these cell wall precursors, the carrier lipid must be rapidly recycled. Hence, once Lipid II has reacted to form the glycan strands of PG, the undecaprenyl pyrophosphate released is converted to Und-P by UppP and other phosphatases (Bouhss et al. 2008; El Ghachi et al. 2004, 2005), and Und-P is flipped back inside the cell by an unknown mechanism to enable another round of precursor synthesis.

3.3 Tailoring Modifications of Peptidoglycan

Tailoring modifications of PG subunits modulate the properties of the cell envelope and may protect bacteria from antimicrobial peptides and proteins (Fig. 2). There are a number of tailoring modifications found in gram-positive bacteria. These include *N*-deacetylation, the removal of C2-acetyl groups from GlcNAc and/or

MurNAc sugars, and *O*-acetylation of the MurNAc C6 hydroxyl (Davis and Weiser 2011; Moynihan et al. 2014).

N-deacetylation has been shown to protect bacteria from lysozyme, a host muramidase that can cleave the glycosidic bond between GlcNAc and MurNAc residues (Ohno et al. 1982). Some gram-positive organisms including *S. pneumoniae*, *Bacillus anthracis*, *B. subtilis,* and other *Bacillus* species are naturally lysozyme resistant and contain a high proportion of *N*-deacetylated sugars in their cell wall (Hayashi et al. 1973; Vollmer and Tomasz 2000; Zipperle et al. 1984). In *S. pneumoniae*, approximately 80 % of the glucosamine residues and 10 % of the muramic acid residues are *N*-deacetylated (Vollmer and Tomasz 2000). This is comparable to the 88 and 34 %, respectively, observed in *B. anthracis* (Zipperle et al. 1984). The enzyme responsible for GlcNAc deacetylation, PgdA, was first identified in *S. pneumoniae* (Vollmer and Tomasz 2000). PdaA, a MurNAc deacetylase (Fukushima et al. 2005), as well as a second MurNAc deacetylase, PdaC, which also has chitin deacetylase activity (Kobayashi et al. 2012), have been identified in *B. subtilis.* The *pgdA* mutant in *S. pneumoniae* was shown to have attenuated virulence (Vollmer and Tomasz 2002) and the *pdaA* mutant in *B. subtilis* is unable to germinate (Fukushima et al. 2002), indicating the possibility of other roles of *N*-deacetylation.

O-acetylation of the MurNAc moiety has been observed in several gram-positive and gram-negative species in variable amounts. In some strains of *S. aureus,* for example, 60 % of MurNAc residues are *O*-acetylated (Clarke and Dupont 1992). *O*-acetylation has been shown to be important for lysozyme resistance and the gene responsible was identified as *oatA* in *S. aureus* (Bera et al. 2005). Homologs of OatA have also been identified in other gram-positive organisms, including *S. pneumoniae* (Crisóstomo et al. 2006) and *E. faecalis* (Hebert et al. 2007). Interestingly, while most gram-positive organisms use OatA homologs for *O*-acetylation, gram-negative organisms use proteins of a different family called Pat. *B. anthracis* produces both kinds of acetyltransferases, and the Pat transferases have been implicated in acetylation of secondary cell wall polysaccharide (Laaberki et al. 2011; Lunderberg et al. 2013). In addition to resistance to lysozyme, *O*-acetylation has been shown to play a role in β-lactam resistance in *S. pneumoniae* and *Listeria monocytogenes* (Aubry et al. 2011; Crisóstomo et al. 2006), and in pathogenesis and immune evasion in *S. aureus* (Bera et al. 2006; Shimada et al. 2010). *O*-acetylation is critical for infection by *L. monocytogenes* and is reported to decrease cytokine production during early stages of infection of mice (Aubry et al. 2011). GlcNAc residues in PG can also be *O*-acetylated but this is more unusual. In *Lactobacillus plantarum,* GlcNAc *O*-acetylation plays a role in inhibiting *L. plantarum'*s major autolysin (Bernard et al. 2011).

In addition to these modifications, PG can be modified at the MurNAc C6 position with different glycopolymers including TAs, teichuronic acids, and CPS. Proteins are also covalently attached to the pentaglycine branch of stem peptides of PG by sortases (Schneewind and Missiakas 2012). In *S. aureus*, sortase-mediated protein attachment is thought to occur on the outside of the cell before Lipid II is polymerized (Perry et al. 2002; Ruzin et al. 2002).

4 Teichoic Acids

The cell envelopes of gram-positive bacteria are rich in teichoic acids (TAs). There are two major classes of TAs: lipoteichoic acids (LTAs), which are anchored to a lipid embedded in the cell membrane, and wall teichoic acids (WTAs), which are covalently attached to PG. LTAs are believed to be present in all gram-positive bacteria with the exception of some *Micrococcus* strains (Powell et al. 1975); WTAs are found in many, including *B. subtilis*, *S. aureus, Staphylococcus epidermidis, S. pneumoniae* and enterococcal species. In organisms where canonical WTAs are not found, other anionic glycopolymers are attached to PG and may play analogous roles (Neuhaus and Baddiley 2003). Under phosphate-limiting conditions, some *B. subtilis* strains produce teichuronic acids instead of WTAs. Teichuronic acids are described in greater detail in the chapter by Lam and coworkers. It is estimated that WTAs and other polyanionic polymers comprise up to 60 % of the cell wall mass (Hancock 1997). Along with LTAs, these polymers play central roles in numerous cellular processes. Some of these functions are covered in detail below.

4.1 Wall Teichoic Acid Structure

WTAs typically consist of a disaccharide linkage unit that is connected at the reducing end to PG via a phosphodiester linkage and at the non-reducing end to a main chain polymer. The structure of the main chain can vary considerably across species but always contains phosphodiester linkages that impart anionic charges to the cell wall (Fig. 4). In *S. aureus* and *B. subtilis,* WTA main chains are composed of glycerol-phosphate or ribitol-phosphate repeats. The WTA main chains are coupled through a disaccharide linkage unit to PG (Armstrong et al. 1960; Brown et al. 2013; Kojima et al. 1985; Neuhaus and Baddiley 2003).

In *S. pneumoniae,* the main chain repeat is composed of 2-acetamido-4-amino-2,4,6-trideoxygalactose, glucose, ribitol phosphate, and two GalNAc moieties, each decorated with phosphorylcholine. The incorporation of phosphorylcholine in WTAs is extremely rare and appears to be exclusive to *S. pneumoniae* (Denapaite et al. 2012; Fischer et al. 1993). In *E. faecalis* 12030, the repeating unit contains D-glucose, D-galactose, 2-acetamido-2-deoxy-D-galactose, 2-acetamido-2-deoxy-D-glucose, and ribitol phosphate (Theilacker et al. 2012). In *E. faecium* U0317, the WTA polymer is simpler, consisting of repeating units of two residues of 2-acetamido-2-deoxy-D-galactose and glycerol phosphate (Bychowska et al. 2011).

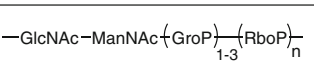
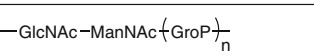
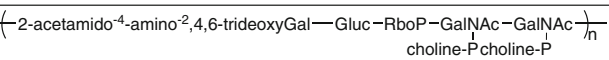
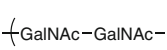
Strain	WTA structure	Reference
S. aureus *B. subtilis* W23		Brown et al. 2010 Brown et al. 2013
B. subtilis 168		Pereria and Brown, 2009
S. pneumoniae		Denapite et al. 2012
E. faecalis 12030		Theilacker et al. 2012
E. faecium U0317		Bychowska et al. 2012

Fig. 4 WTA structure and common variations. WTAs are anionic polymers with a sugar-phosphate backbone attached to the C6 position of MurNAc in PG. The structure of WTAs is highly variable across gram-positive species. WTA polymer structures for specific strains are indicated here with the following abbreviations: glycerol phosphate (GroP), ribitol phosphate (RboP), *N*-acetylmannosamine (ManNAc), galactose (Gal), glucose (Gluc), *N*-acetylgalactosamine (GalNAc), *N*-acetylglucosamine (GlcNAc), phosphorylcholine (choline-P)

4.2 Wall Teichoic Acid Biosynthesis

The biosynthetic pathways for WTA assembly in *B. subtilis* and *S. aureus* have been well established (Brown et al. 2008, 2010; Lazarevic et al. 2002; Mauël et al. 1991) and are covered in the chapter by Lam and coworkers. The assembly begins in a similar manner to PG assembly. Briefly, phospho-GlcNAC is transferred from UDP-GlcNAC to the Und-P lipid carrier and then this "starter unit" is further elaborated by a series of intracellular enzymes to assemble the full polymeric precursor. While the structures of the main chains made in *B. subtilis* and *S. aureus* are similar, particularly in WTAs from *B. subtilis* W23 and *S. aureus*, there are substantial differences in the biosynthetic pathways that were not evident from bioinformatic analysis (Brown et al. 2008, 2010; Meredith et al. 2008; Pereira et al. 2008). It is not yet possible to predict the enzymatic functions of putative TA primases and polymerases accurately. Once the full chain is polymerized inside the cell, it is flipped by a two-component ABC transporter to the surface of the bacterial membrane and ligated to the PG. The pathway in *S. pneumoniae* and other species has not been as well elucidated and most of the enzymes, apart from those responsible for choline uptake, have been deduced by bioinformatic analysis and remain to experimentally validated (Denapaite et al. 2012).

Unlike PG, WTAs are not essential for survival of *S. aureus* in vitro as the first two genes in the pathway can be deleted. However, the subsequent genes in the pathway were identified as essential (Chaudhuri et al. 2009; Kobayashi et al. 2003). This apparent paradox was resolved by studies showing that the downstream genes in the WTA pathway can be deleted as long as one of the first two genes has been disrupted (D'Elia et al. 2006). This finding implied that the essentiality of the downstream genes was conditional on flux into the pathway, and it was suggested

that lethality due to a late block in WTA biosynthesis could arise from accumulation of a toxic metabolite or from sequestration of the Und-P carrier lipid in WTA intermediates, which would lead to inhibition of PG biosynthesis (D'Elia et al. 2009). It was recently shown that inhibiting a late step in WTA biosynthesis results in rapid depletion of the PG precursor Lipid II, consistent with lethality arising from inhibition of PG biosynthesis (Qiao et al. 2014). Other cell envelope polymers such as CPS are synthesized on the Und-P carrier lipid, and the biosynthetic pathways for some of these also contain a mix of nonessential early genes and conditionally essential late genes (Xayarath and Yother 2007). Conditional essentiality of the late genes depends on whether intermediates can be metabolized through an alternative pathway to release the carrier lipid.

The final step of the WTA pathway involves the ligation of WTAs onto PG. The LytR-CpsA-Psr protein family was recently shown to be involved in this process (Kawai et al. 2011; Over et al. 2011; Dengler et al. 2012). *B. subtilis*, *S. aureus*, and *S. pneumoniae* strains have three LytR-CpsA-Psr homologs. In the case of *S. aureus*, one of these homologs has been shown to be involved in the ligation of CPS to PG (Chan et al. 2014). The other two appear to be involved in the ligation of WTAs to PG (Chan et al. 2013), but their cellular functions have not been clearly delineated. No LytR-CpsA-Psr family member has yet been reconstituted in vitro. More details on the discovery of these proteins are provided in the chapter by Lam and coworkers.

4.3 Lipoteichoic Acid Structure

In most organisms, LTAs are synthesized by completely different biosynthetic pathways from WTAs, except in the case of *S. pneumoniae* where the repeating units are structurally identical and are thought to be assembled using the same enzymes (Denapaite et al. 2012; Fischer et al. 1993). The most common LTA structure comprises a polyglycerol-phosphate chain anchored to a glycolipid in the membrane. This type of LTA is found in *S. aureus*, *B. subtilis,* and *L. monocytogenes*. In other species of gram-positive organisms, LTAs contain additional sugar moieties connecting the glycolipid anchor to the polyglycerol-phosphate polymer. The glycolipid anchor is usually diacylglycerol with two glucose moieties (Glc_2DAG), as in *S. aureus* and *B. subtilis*, but it can also contain more than two glucose residues (*Clostridium difficile)* as well as other sugar moieties such as galactose (in *L. monocytogenes*) or GlcNAc (in *Clostridium innocuum*) (Fischer 1988; Percy and Gründling 2014).

4.4 Lipoteichoic Acid Synthesis

LTA synthesis begins in the cytoplasm with the assembly of the glycolipid anchor. In *S. aureus* and *B. subtilis*, YpfP (also called UgtP) is responsible for attaching both glucose units to diacylglycerol (DAG) to give the glycolipid anchor, Glc_2DAG (Jorasch et al. 1998; Kiriukhin et al. 2001), which is then flipped across the membrane by LtaA (Gründling and Schneewind 2007a). LtaS then builds the polymer chain by transferring glycerol phosphate from phosphatidylglycerol to Glc_2DAG (Gründling and Schneewind 2007b). Deleting *ypfP* or *ltaA* does not abolish the synthesis of LTA, but results in polymers with altered structure. Evidently, LTA can be synthesized on DAG, as well as Glc_2DAG (Gründling and Schneewind 2007a). LtaS is a polytopic membrane protein with an extracellular domain. The crystal structure of the extracellular domain of LtaS (eLtaS) bound to glycerol phosphate has been reported and suggests a possible covalent mechanism for LtaS in which an active site threonine reacts with phosphotidylglycerol to form a covalent glycero-phospho-threonine intermediate. This intermediate is resolved by reaction with the hydroxyl group of the growing LTA chain (Lu et al. 2009; Schirner et al. 2009). Some organisms, such as *L. monocytogenes,* contain a two-enzyme pathway to make LTA main chains (Webb et al. 2009). One enzyme, LtaP, functions as a primase to add one unit of glycerol phosphate to the glycolipid anchor. In the case of *L. monocytogenes*, this glycolipid anchor is Gal-Glc-DAG. A polymerase, LtaS, then extends the chain. LtaP is not essential for LTA synthesis; however, LTAs from a *ltaP* null mutant are longer than those from the wild-type strain (Webb et al. 2009), as in a *ltaA* or *ypfP* deletion in *S. aureus*. The mechanistic basis for length differences between "primed" and "unprimed" glycolipid anchors is not understood. A recent crystal structure of LtaS from *L. monocytogenes* reveals a glycerol-phosphate-binding site that may accommodate part of the growing LTA chain (Campeotto et al. 2014). While glycerol-phosphate polymerization activity has not been reconstituted for any LtaS, perhaps because some of the transmembrane helices form part of the active site for polymerization, eLtaS from *S. aureus* was shown to be sufficient for cleavage of the phosphodiester bond in phosphatidylglycerol (Karatsa-Dodgson et al. 2010). The diacylglycerol product released in the LtaS reaction with phosphatidylglycerol is recycled back into the cell and the protein responsible for recycling has been identified as diacylglycerol kinase DgkB (Jerga et al. 2007).

While *S. aureus* contains only one LtaS and *L. monocytogenes* has LtaP and LtaS, *B. subtilis* has four LtaS homologs (Gründling and Schneewind 2007b; Schirner et al. 2009). It has been reported that while three of these homologs—LtaS, YqgS, and YfnI—have LtaS-like activity, one of them, YvgJ, functions as a primase (Wörmann et al. 2011). Unlike in *L. monocytogenes,* the *B. subtilis* primase is not required for normal LTA synthesis, suggesting that the LtaS enzymes are capable of initiating synthesis of LTA polymers efficiently. YfnI has been shown to make LTA polymers that are substantially longer than those produced by LtaS or YqgS (Wörmann et al. 2011). The observation that *yfnI* expression is regulated by

the alternative sigma factor SigM, which responds to stress conditions (Jervis et al. 2007), suggests that certain stresses call for the production of elongated polymers in *B. subtilis* (Wörmann et al. 2011). Phenotypically, *ltaS* mutants show increased cell elongation and chain length, reduced cell diameter, cell bending, lysis, and abnormally thick septa, whereas single deletions of the other three homologs do not have any obvious defects. The *ltaS-yqgS* double mutant has sporulation defects; all other double mutant combinations with *ltaS* can sporulate. These results implicate LtaS and YqgS in sporulation. The quadruple mutant is viable, although it has a more severe phenotype than the single *ltaS* mutant (Schirner et al. 2009). Deletion of *ltaS* in *S. aureus* has also been accomplished, but viable mutants have suppressors that enable growth through a mechanism that involves increased levels of cyclic-di-AMP, which may regulate cell membrane functions (Corrigan et al. 2011, 2013). Even with the suppressor, these mutants have severe cell division defects (Corrigan et al. 2011; Gründling and Schneewind 2007b; Oku et al. 2009). Hence, LTAs are critical even for in vitro growth of many gram-positive organisms.

4.5 Tailoring Modifications of Teichoic Acids

Both LTAs and WTAs are often modified with D-alanine esters to modulate the charges of the cell envelope. They can also be modified with sugar moieties. These tailoring modifications have been implicated in numerous functions in cell physiology and infection.

D-*alanylation*: The ribitol (in WTA) or glycerol (as in *S. aureus* LTA) groups in TAs are frequently decorated with D-alanine moieties, which introduce positive charges to neutralize the negatively charged phosphates in the polymer backbone. On ribitol groups, D-alanylation occurs at the C2 position (Neuhaus and Baddiley 2003). D-alanine moieties are added by four proteins, DltABCD, encoded by the *dlt* operon (Fig. 5). DltA activates D-alanine as the AMP ester and then transfers it to the sulfhydryl group on the phosphopantetheinyl arm of the carrier protein DltC (Heaton and Neuhaus 1992, 1994; Perego et al. 1995; Volkman et al. 2001). DltA is similar to carrier protein ligases found in non-ribosomal peptide synthetases (Brown et al. 2013; Percy and Gründling 2014; Yonus et al. 2008). The next steps are not understood. DltB is a polytopic membrane protein belonging to the mBOAT family (for membrane-bound O-acetyl transferases), which is ubiquitous in all kingdoms of life (Hoffman 2000). DltD contains a single membrane spanning helix and an extracellular domain with predicted esterase/thioesterase activity (Brown et al. 2013; Reichmann et al. 2013). It has been proposed that DltC transfers D-alanine to Und-P to form an acyl-phosphate intermediate, which is then transferred through the membrane by DltB to modify LTAs with the assistance of DltD (Perego et al. 1995; Reichmann et al. 2013). There is no evidence for the proposed acyl-phosphate intermediate and the reaction to form it from the thioester is thermodynamically unfavorable, although it may conceivably be coupled to hydrolysis of the pyrophosphate released during D-alanine activation by DltA. Pulse-chase

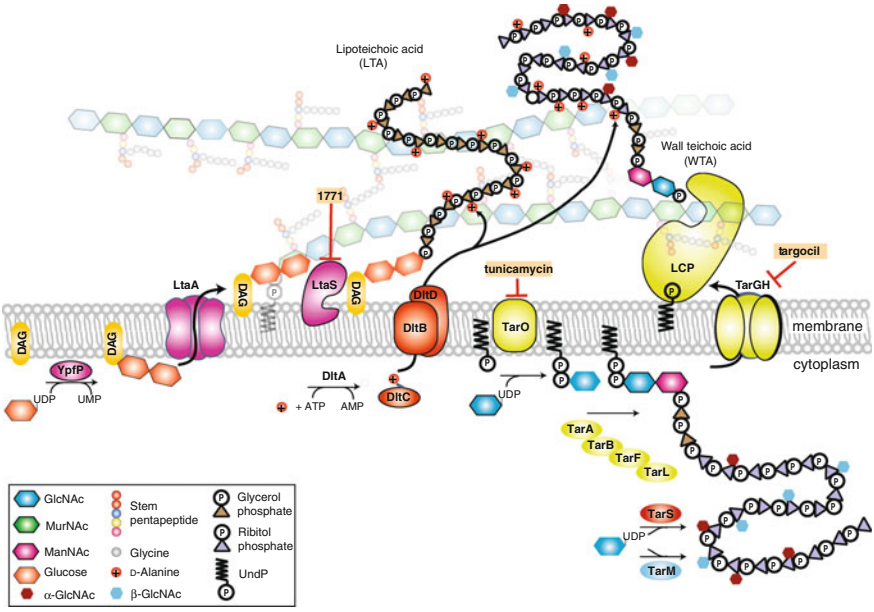


Fig. 5 TA biosynthesis and modification pathways. LTA and WTA biosynthetic pathways in *S. aureus* are shown here. Although both are anionic sugar-phosphate backbones, they are assembled differently by separate biosynthetic pathways in *S. aureus*. TAs are further modified with D-alanine residues by the *dlt* pathway and with α- or β-GlcNAC residues installed by glycosyltransferases TarM and TarS, respectively. TAs perform several functions for the cell including playing roles in biofilm formation, adhesion, phage attachment, virulence, and antibiotic resistance, most notably resistance to β-lactams. D-alanylation has been shown to play an important role in these functions as well. Specifically, the absence of D-alanine modifications sensitizes to cationic antimicrobial peptides, including host defensins. The only known roles for α- and β-GlcNAC modifications are in phage attachment, and for β-GlcNAcs, in β-lactam resistance. Due to its roles in adhesion, virulence, and antibiotic resistance, attempts are being made to target TA biosynthesis and modification pathways. The known compounds targeting these pathways are shown here

experiments have suggested that D-alanines installed on LTAs are subsequently transferred to WTAs (Haas et al. 1984; Koch et al. 1985), but the mechanistic details of the transfer are unclear. In particular, it is not known whether an enzyme is involved in the process.

Glycosylation: The majority of ribitol phosphate groups in WTAs in *S. aureus* are glycosylated with GlcNAc on the ribitol C4 position (Brown et al. 2013). Similarly, LTAs can also be glycosylated with GlcNAc or α-galactose in *B. subtilis* (Percy and Gründling 2014). In *S. pneumoniae*, LTA can be glycosylated with GalNAc (Draing et al. 2006). In staphylococci, it has been shown that D-alanylation and glycosylation compete for the same position on LTAs. Approximately 70 % of the glycerol phosphates carry D-alanines while 15 % carry GlcNAc moieties (Schneewind and Missiakas 2014). WTA precursors are glycosylated intracellularly and the enzymes responsible for glycosylation have been identified in a number of

organisms. In *B. subtilis* 168, TagE attaches α-glucosyl units to the polyglycerol-phosphate WTA chains (Allison et al. 2011); in *B. subtilis* W23, TarQ attaches β-glucosyl units to the polyribitol-phosphate WTA chains (Brown et al. 2012). In *S. aureus*, TarM attaches α-GlcNAc residues while TarS attaches β-GlcNAc residues (Shobanifar et al. 2015; Xia et al. 2010; Brown et al. 2012). No enzymes responsible for LTA glycosylation, which occurs extracellularly, have yet been identified. It is likely that these enzymes use membrane-anchored sugar substrates that cannot diffuse away from the cell and therefore do not resemble the nucleotide-diphosphate sugar transferases that glycosylate WTA precursors inside the cell.

4.6 Roles of Teichoic Acids and Their Tailoring Modifications in Cell Physiology and Immune Evasion

Roles in cell division and morphology: TAs perform several crucial functions for the cell. In *B. subtilis*, WTAs are required to maintain the rod-shaped morphology (Boylan et al. 1972; Pollack and Neuhaus 1994; Schirner et al. 2015). In the quadruple mutant lacking all four LtaS homologs, there are severe cell division and septation defects that cause filamenting and clumping of cells and the mutant grows very slowly, indicating that LTAs are required for proper cell division (Schirner et al. 2009). Disruption of YpfP caused the rod-shaped cells to become bent and distended, and also disrupted the localization of the cytoskeletal protein MreB, important for the rod-shape in *B. subtilis* (Matsuoka et al. 2011). Interestingly, YpfP has also been implicated in a metabolic sensing role, localizing to the division site in a nutrient-dependent manner and inhibiting the assembly of FtsZ. It is important that the number of Z-rings to cell length is maintained at a constant ratio so cells do not initiate division before reaching the correct cell mass. Thus, YpfP could play a significant role in cell cycle events (Weart et al. 2007). In *S. aureus,* both LTAs and WTAs have been implicated in cell division: mutants defective in either LTA or WTA biosynthesis have major septal defects, including placing new septa at angles non-orthogonal to previous septa and forming multiple septa almost simultaneously. These mutants are also impaired in separation after division (Campbell et al. 2011; Gründling and Schneewind 2007b; Oku et al. 2009). In *S. aureus*, LTAs are more critical to the cell than WTAs in vitro as evidenced by the fact that the *ltaS* deletion strain is viable only in the presence of suppressors (Corrigan et al. 2011), whereas *tarO* mutants grow fairly well. WTAs, however, become very important in vivo (Valentino et al. 2014; Wang et al. 2013; Weidenmaier et al. 2005). Simultaneous disruption of WTAs and LTAs is lethal in both *S. aureus* and *B. subtilis* (Oku et al. 2009; Santa Maria et al. 2014; Schirner et al. 2009). In *S. aureus,* cells lacking both polymers are unable to form the essential division ring (Z-ring) (Santa Maria et al. 2014). Interestingly, in the absence of WTAs, D-alanyl modifications on LTAs become essential. Both WTAs and D-alanylation have been

implicated in autolysin regulation, and when WTAs and D-alanines are both missing, cells lyse rapidly. The evidence suggests that LTAs and WTAs have overlapping but not fully redundant roles in cell division and autolysin regulation (Santa Maria et al. 2014).

Roles in ligand binding and scaffolding: TAs have been implicated in binding cations, and this correlates inversely with D-alanylation levels (Archibald et al. 1973; Neuhaus and Baddiley 2003). Cation homeostasis is thus an important function of TAs that can be regulated through D-alanylation. WTAs also serve as phage receptors in *S. aureus* (Brown et al. 2012; Chatterjee 1969; Xia et al. 2010; Young 1967). Phage binding is mediated by the GlcNAc modifications added on to WTAs (Brown et al. 2012; Xia et al. 2010). Requirement for glucose in TAs for phage adsorption has been shown in *B. subtilis* 168 as well (Young 1967; Allison et al. 2011). WTAs have also been implicated in other protein scaffolding roles. For instance, in *S. aureus*, FmtA, a protein that plays a role in methicillin resistance in MRSA strains, was shown to bind to WTAs (Qamar and Golemi-Kotra 2012). In *S. pneumoniae*, several proteins bind specifically to the choline moieties on TAs. These proteins, which include the highly studied virulence protein PspA, have been implicated in numerous functions from adhesion to virulence, and cell wall hydrolysis (Fischer 2000; Giudicelli and Tomasz 1984; Gosink et al. 2000; Hakenbeck et al. 2009; Rosenow et al. 1997). In *L. monocytogenes*, InlB, a protein that promotes entry into mammalian cells, is shown to interact with LTAs (Jonquières et al. 1999). The domain necessary for interaction with LTAs in this protein contains GW modules (conserved modules of ~80 amino acids which have the dipeptide Gly-Trp). These modules have also been identified in Ami, a *L. monocytogenes* autolysin and the *S. aureus* autolysin Atl (Cabanes et al. 2002). Autolysins are hydrolases that degrade PG and thus play an essential role in cell division and separation. In *S. aureus*, WTA plays a role in Atl localization. While Atl is usually localized to the cross-wall, it is mislocalized across the cell surface in WTA-deficient strains. Mislocalization of autolysins could be one reason WTA-deficient mutants are prone to autolysis (Schlag et al. 2010). It has been suggested that D-alanylation is also involved in autolysin regulation (Peschel et al. 2000). Similarly, PBP4 in *S. aureus* is also mislocalized when WTAs are absent (Atilano et al. 2010), indicating a role for WTAs in the localization of PG biosynthetic machinery.

Roles in antibiotic resistance and virulence: In MRSA, the lack of WTAs dramatically reduces the organism's resistance to β-lactams, indicating that WTAs play a major role in methicillin resistance of *S. aureus* (Campbell et al. 2011). The influence of WTAs on resistance has been traced specifically to the β-GlcNAc modification on WTAs, which suggests that β-GlcNAcylated WTAs scaffold a factor required for β-lactam resistance (Brown et al. 2012). In *S. aureus,* WTAs also provide resistance to antimicrobial fatty acids on the skin during skin colonization (Kohler et al. 2009). D-alanylation plays an important role in modulating resistance to certain antibiotics. It is very important for repelling cationic antimicrobial peptides (CAMPs), a crucial part of host immune response (Collins et al. 2002; Kristian et al. 2005; Peschel et al. 1999). This has been observed in several gram-positive

species including *S. aureus* (Peschel et al. 1999), *S. pneumoniae* (Kovács et al. 2006), and *E. faecalis* (Fabretti et al. 2006). An increase in D-alanylation is also observed in mutants resistant to daptomycin, an antibiotic used to treat MRSA (Yang et al. 2009a). Antimicrobial resistance due to D-alanylation has been attributed to its functions in imparting positive charges to the cell surface and its contributions to changes in the biophysical aspects of the cell envelope (Mishra et al. 2014; Saar-Dover et al. 2012).

TAs in their D-alanylated form play a major role in biofilm formation, adhesion to the surface of cells and medical devices, colonization of host tissue, and virulence, likely due to surface charge effects (Brown et al. 2013; Gross et al. 2001; Jett et al. 1994; Neuhaus and Baddiley 2003; Percy and Gründling 2014). Biofilms, which consist of viable cells held together by an extracellular matrix of DNA and proteins from lysed cells as well as extracellular polysaccharides and other polymers, form on surfaces of medical instruments or in hosts, and enable the organism to evade both natural and synthetic antimicrobials (Hall-Stoodley et al. 2004; Sutherland 2001; Abee et al. 2011). Thus, adhesion and biofilm formation are key tools in a pathogen's arsenal. The role of TAs in adhesion and effective host colonization has been well established in several gram-positive organisms (Aly et al. 1980; Baur et al. 2014; Fabretti et al. 2006; Weidenmaier et al. 2004). In *S. aureus*, WTA glycosylation has specifically been implicated in adhesion (Winstel et al. 2015). For all these reasons, TAs are potent virulence factors and mutants lacking TAs or D-alanylation have highly attenuated virulence (Abachin et al. 2002; Collins et al. 2002; Fittipaldi et al. 2008; Suzuki et al. 2011a; Weidenmaier et al. 2005; Xu et al. 2015). As mentioned above, several choline-binding proteins in *S. pneumoniae* have roles in virulence and mutants made to grow independent of choline have highly attenuated virulence (Kharat and Tomasz 2006).

LTAs contribute to the immune response generated during infection by gram-positive bacteria (Ginsburg 2002). Although there was some controversies concerning whether the immunomodulation arises from LTAs or from lipoproteins that are often copurified (Hashimoto et al. 2006a, b), evidence suggests that LTAs likely affect the immune system response on their own as well (Bunk et al. 2010; Mohamadzadeh et al. 2011; von Aulock et al. 2007). LTAs are reported to stimulate the production of cytokines (Bhakdi et al. 1991; Draing et al. 2008; Ray et al. 2013) and those from *S. pneumoniae* and *S. aureus* can activate immune cells via toll-like receptor 2, lipopolysaccharide binding protein and CD14 (Ryu et al. 2009; Schröder et al. 2003). They also activate the complement system of the immune response (Fiedel and Jackson 1978; Loos et al. 1986) and can affect other macrophage parameters, including secretion of tumor necrosis factor α and nitrite (Keller et al. 1992). Antibodies have been identified that are directed toward non-D-alanylated LTAs in *E. faecalis* (Theilacker et al. 2006). Due to this ability to modify host immunity, efforts are ongoing to develop LTA-conjugated vaccines against gram-positive bacteria (Percy and Gründling 2014). The choline-binding proteins anchored to TAs in *S. pneumoniae* could be used as vaccine candidates as well (Jedrzejas 2001; Rosenow et al. 1997).

5 Capsular Polysaccharides

Capsular polysaccharides (CPS) are highly variable glycopolymers that are anchored to PG (Chan et al. 2014; Sorensen et al. 1990; Xayarath and Yother 2007; Yother 2011). They extend above the cell wall and have been implicated in phage resistance and immune evasion (O'Riordan and Lee 2004; Roberts 1996). Although not present in all gram-positive organisms, encapsulation is observed in most highly pathogenic strains. The synthesis of CPS is covered in the chapter by Lam and coworkers. Since CPS is best studied in *S. pneumoniae*, we will focus on the structural diversity in CPS in *S. pneumoniae* and their function in immune evasion.

5.1 Structural Diversity of CPS

A phenomenal 93 different serotypes of pneumococcal capsule have been identified over the years and most of the serotypes can cause infection (Kalin 1998; Yother 2011). Recombinational exchanges at the CPS biosynthetic locus can result in a large amount of variation in capsular type (Coffey et al. 1998). Disruption and sequence changes in the genes of the CPS cluster occurring naturally can change the CPS serotype from one to another (Calix et al. 2014; Calix and Nahm 2010; van Selm et al. 2003) contributing to the diversity of pneumococcal capsules. These differences are usually observed in the gene responsible for modifying sugar moieties in CPS with *O*-acetyl groups. In fact, in vivo switching from one capsule type to another has been observed (Venkateswaran et al. 1983). This switch has been attributed to a change in the number of short tandem TA nucleotide repeats in the putative *O*-acetyltransferase gene, which could explain reversible switching between serotypes that might occur in vivo (van Selm et al. 2003).

CPS is made of long chains of repeating oligomeric units and the repeating units vary between serotypes. As an example, the repeat unit of *S. pneumoniae* serotype 2 is made of a backbone with glucose-rhamnose-rhamnose-rhamnose unit and a glucose-glucuronic acid side chain (Kenne et al. 1975). Recently, serotypes of *S. pneunomiae* that have CPS containing two different repeat units have been described (Oliver et al. 2013a, b). There are multiple different serotypes in *S. aureus* as well. Out of the 11 serotypes described for *S. aureus*, serotypes 5 and 8 are responsible for the majority of human infections (O'Riordan and Lee 2004).

5.2 CPS, Host Immunity, and Vaccine Development

It has long been known that CPS reduces the ability of bacteriophage to interact with the cell surface (Wilkinson and Holmes 1979). CPS plays a major role in virulence of bacterial pathogens and capsule mutants are avirulent. Capsule has

been shown to facilitate abscess formation by activating T cells in the host immune system (Tzianabos et al. 2001). The complement system is important in immune response activation and clearing an infection. Capsule is able to mask the binding of opsonic C3 fragments to the complement receptor, thus decreasing opsonization and phagocytosis by leukocytes (Cunnion et al. 2003; Peterson et al. 1978). This has also been demonstrated in *E. faecalis*, where capsule masks C3 deposits and LTAs from detection by the host immune system, thereby decreasing tumor necrosis factor α production (Thurlow et al. 2009). In Group B *Streptococcus*, the terminal sialic acid groups on capsules have been shown to interact with Siglecs on human leukocytes. They are suggested to mimic the human cell surface glycans, reducing the activation of innate immune response (Carlin et al. 2007, 2009).

Due to the high immunomodulatory ability of CPS, it has been explored for vaccine development. It has been known for a long time that immunization with polyvalent pneumococcal polysaccharide is effective as a vaccine (MacLeod et al. 1945; Shapiro et al. 1991). It was later shown that conjugating the polysaccharides to a carrier protein resulted in a more effective vaccine (De Velasco et al. 1995). Today, different variations on pneumococcal vaccines are available, incorporating up to 23 polysaccharide variants (PPSV23), or conjugate vaccines incorporating 7 (PCV7) or 13 (PCV13) CPS serotypes (Bogaert et al. 2004; Pilishvili and Bennett 2015; Steens et al. 2014). PCV13 is used for immunization of infants <2 years of age and has recently also been approved for immunizing adults 50 years or older in series with PPSV23. PPSV23, however, is not effective in infant immunization. This is because PPSV23 generates immune responses that are T-cell independent and therefore, poorly supported by the immature immune systems of children <2 years. In contrast, PCV13 generates immune responses that are mediated by T-cell-dependent mechanisms effective in infants (Pilishvili and Bennett 2015). Efforts are being made in improving not only the polysaccharide composition of vaccines but also the carrier protein used to conjugate the polysaccharide. The immunogenic properties of the carrier protein could alter the immune response to the vaccine (Dagan et al. 2010; Pobre et al. 2014). There is a concern that pneumococcal conjugate vaccines select for non-vaccine serotypes. Pelton et al. reported that immunization with PCV7 during 2000–2003 reduced vaccine serotypes from 22 to 2 % but increased the incidence of non-vaccine serotypes from 7 to 16 % (Pelton et al. 2004). With over 90 different serotypes of *S. pneumoniae,* this is an important concern, and studies are ongoing to resolve this issue (Jefferies et al. 2011; Nurhonen and Auranen 2014).

Capsular conjugate vaccines against serotypes 5 and 8 of *S. aureus* have also been explored (Creech et al. 2009; Fattom et al. 2004; Robbins et al. 2004). However, these vaccines have so far not passed clinical trials (Bagnoli et al. 2012; Cook et al. 2009), and evidence has emerged that this reduced efficacy could be due to interference from natural non-opsonic antibodies to PNAG, the *S. aureus* exopolysaccharide, present in human serum (Skurnik et al. 2012).

6 Exopolysaccharides and Biofilm Formation

Apart from these major cell envelope structures, other glycopolymers called exopolysaccharides are secreted by cells as well. These exopolysaccharides are long chains that associate with each other to form the biofilm matrix (Sutherland 2001; Otto 2008; Vlamakis et al. 2013). Polysaccharide intercellular adhesin (PIA) in *S. epidermidis* is a well-studied component of biofilms (Mack et al. 1996; Itoh et al. 2005). It is a linear polymer of β-1,6-linked GlcNAc moieties, although some residues can be *N*-deacetylated. PIA/PNAG is suggested to be held to the cell surface by ionic interactions of the positively charged, un-acetylated moieties of the polymer, so *N*-deacetylation is important for surface localization of PIA (Vuong et al. 2004). PIA is synthesized by the *icaADBC* operon in *S. epidermidis,* and homologs have been identified in other species including *S. aureus* (Gerke et al. 1998; Heilmann et al. 1996; Mack et al. 1996; Rohde et al. 2010). In *S. aureus,* this high-molecular mass exopolysaccharide termed PNAG is produced by biofilm-forming strains. Due to its role in modulating immune responses, vaccines using conjugated PNAG are also being explored (Maira-Litrán et al. 2012). Its role in biofilm formation has created interest in the study of the role of each enzyme in the *icaADBC* operon and how it is regulated (Arciola et al. 2015; O'Gara 2007). There are also *ica*-independent methods for biofilm formation which include roles by TAs and cell-surface-associated proteins. The mechanism for biofilm formation in MRSA appears to be *ica*-independent, whereas it is *ica*-dependent in the sensitive strains (O'Gara 2007). Biofilm formation is thus a complex and highly regulated system.

7 Antibiotics Targeting the Cell Envelope

Due to the crucial importance of the cell envelope to cell survival, many antibiotics that target cell envelope synthesis have been developed over the years (Fig. 3) (Walsh 2003). There are some antibiotics that target the intracellular steps of PG synthesis, including fosfomycin, which inhibits MurA, the first committed step of PG synthesis (Kahan et al. 1974). However, the greatest clinical successes have been achieved by those antibiotics that target the extracellular steps of cell wall synthesis. These include the unusual substrate-binding antibiotics, which form complexes with cell wall precursors instead of the enzymes that process them. Binding to these precursors prevents their use and results in inhibition of cell wall synthesis. Vancomycin, a glycopeptide antibiotic used to treat MRSA, belongs to the substrate-binding class of antibiotics. It binds to the D-Ala-D-Ala motif of the stem peptide in Lipid II and nascent PG, thereby interfering with both Lipid II polymerization to form PG strands and with subsequent cross-linking of the strands (Anderson et al. 1967; Perkins and Nieto 1974; Perkins 1969; Reynolds 1989). Binding to and sequestering Lipid II has been established as the mechanism of

action of some other antibiotics including ramoplanin, a cyclic lipoglycodepsipeptide antibiotic (Lo et al. 2000; Hu et al. 2003b), nisin and other lantibiotics (Brötz et al. 2002; Hsu et al. 2004; Oman et al. 2011; Patton and van der Donk 2005), and the recently discovered teixobactin (Ling et al. 2015). All these compounds recognize the pyrophosphate-sugar moiety of Lipid II. Plectasin, a fungal defensin, also acts by binding to Lipid II (Schneider et al. 2010). Human defensins have also been shown to interact with Lipid II (Sass et al. 2010; De Leeuw et al. 2010). It is interesting that antimicrobial peptides produced by the host as part of the innate immune response use Lipid II binding to counteract bacterial threats. The structural diversity of the compounds that bind Lipid II is truly astonishing and indicates that this cell wall precursor is an exceptional target.

Development of resistance to compounds which bind to essential substrates is particularly slow for several reasons. They typically act on the extracellular surface of the membrane and are not subject to efflux pump-mediated resistance mechanisms. Moreover, because they do not bind to a protein target, a single mutation in the gene encoding the target cannot confer high-level resistance (Wright 2011). In the case of vancomycin, intermediate resistance can arise through multiple mutations that modify the envelope, but high-level resistance only arises due to the modification of the structure of the target substrate (Gardete and Tomasz 2014; Walsh and Howe 2002; Healy et al. 2000). The modification, which involves replacing D-Ala-D-Ala with a dipeptide to which vancomycin cannot bind, requires several enzymes, as well as a two-component sensing system, and the genes encoding these enzymes are encoded on a cassette that is transferred between organisms (Arthur and Courvalin 1993; Palmer et al. 2010). Glycopeptide resistance genes originated in a glycopeptide producer as a means of self-immunity, but now have spread widely, particularly in enterococcal strains (Marshall et al. 1998). D-Ala-D-Lac, synthesized by the *vanA* cassette, is the most common replacement for D-Ala-D-Ala in vancomycin-resistant strains. Vancomycin has a thousand-fold lower affinity for D-Ala-D-Lac because a crucial hydrogen bond between the drug and the target can no longer be formed (Arthur and Courvalin 1993; Handwerger et al. 1992; Bugg et al. 1991). A change from D-Ala-D-Ala to D-Ala-D-Ser in Lipid II can also cause moderate resistance to vancomycin (Depardieu et al. 2007; Lebreton et al. 2011). Although high-level vancomycin resistance is common in enterococci (VRE), it has not yet emerged as a major problem in *S. aureus*, likely due to reduced frequency of transfer of the resistance cassette between enterococci and staphylococci (Palmer et al. 2010; Périchon and Courvalin 2009). The several cases where vancomycin-resistant S. aureus (VRSA) have been identified have involved coinfection with VRE (Weigel et al. 2003; Zhu et al. 2008; Sievert et al. 2008; Chang et al. 2003; Whitener et al. 2004). The barriers that prevent facile transfer of *vanA* resistance into *S. aureus* are not well understood, and there is a concern that these barriers may be overcome with continued evolution. While there is an interest in substrate binders as a class, none of the ones that recognizes the sugar pyrophosphate portion of Lipid II has been developed for clinical use, although ramoplanin is in clinical trials (Paknikar and Narayana 2012). As with vancomycin, high-level resistance to ramoplanin does not develop spontaneously. Moderate

ramoplanin resistance develops after multiple passaging and involves cell envelope modifications that may impede access to the Lipid II target on the cell surface (Schimdt et al. 2010). If any Lipid II binders come to be used clinically, resistance genes from the producing organisms may eventually find their way into relevant pathogens, like in the case of vancomycin.

β-lactams, a remarkably successful class of antibiotics, are also among the extracellular PG synthesis inhibitors. β-lactams are proposed structural mimics of D-Ala-D-Ala and inhibit the transpeptidase activity of PBPs by acylating the active site, preventing the cross-linking of stem peptides (Yocum et al. 1979, 1980). Widespread resistance to β-lactams first emerged in the form of β-lactamases, which degrade β-lactams (Gutkind et al. 2013). Combination antibiotics of β-lactams with β-lactamase inhibitors are used to treat many β-lactam-resistant infections. One example is Augmentin, a combination of amoxicillin and clavulanic acid (Drawz et al. 2014; Reading and Cole 1977; White et al. 2004). While β-lactamases continue to be a major concern in gram-negative organisms such as *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* (Hong et al. 2015; Pitout et al. 2015), some gram-positive organisms have acquired a different mechanism of resistance. Methicillin-resistant *S. aureus* (MRSA) expresses a penicillin-binding protein (PBP2A) that has reduced affinity for β-lactams (Hartman and Tomasz 1984; Lim and Strynadka 2002; Fuda et al. 2004). When native PBPs are inhibited by β-lactams, PBP2A can continue to cross-link PG. Due to the growing concern about the spread of MRSA, a significant amount of time has been invested in designing next-generation β-lactams that can target the resistant PBP, including ceftobiprole (Davies et al. 2007) and ceftaroline (Moisan et al. 2010). In addition, other classes of antibiotics have been developed to treat MRSA, including daptomycin, tedizolid, linezolid, and the glycopeptide analog oritavancin (Hall and Michaels 2015; Holmes and Howden 2014; Leach et al. 2011; McDaneld et al. 2013; Mitra et al. 2015).

8 The Quest for Novel Antibiotic Targets

Resistance to antibiotics of all classes is a serious concern for the future of human health, and efforts should be made to identify novel pathways that can be targeted by new antibiotics or whose inhibition can potentiate the effects of existing antibiotics in resistant strains. Efforts are ongoing to identify and target the multiple other steps involved in the PG biosynthetic pathway. For instance, inhibitors of the Lipid II flippase in *S. aureus*, DMPI and CDFI, have been identified (Huber et al. 2009). Targeting pathways that contribute to resistance to current antibiotics are also being explored as a viable option. Apart from the β-lactamases described above, the potential for targeting such auxiliary proteins and pathways is immense, particularly in the case of MRSA, where many cellular factors contribute to β-lactam resistance (Berger-bächi and Rohrer 2002). For instance, changes to the stem peptide and interpeptide bridge resensitize MRSA to β-lactams (Ludovice et al. 1998; De Jonge et al. 1993; Maidhof et al. 1991; Tschierske et al. 1997). This has

also been observed in *S. pneumoniae* (Weber et al. 2000). In *S. aureus*, inactivation of one of the PBPs involved in cross-linking of stem peptides, PBP4, is shown to play a role in resistance to β-lactams (Memmi et al. 2008). This has also been shown for the inhibition of PG amidation (Figueiredo et al. 2012). Inactivation of *tarO*, encoding the first step in WTA biosynthesis, also sensitizes MRSA to β-lactams (Campbell et al. 2011). Finally, factors affecting methicillin resistance also include proteins of hitherto unknown functions. FmtA is an example of one such protein factor (Komatsuzawa et al. 1997). Further understanding of the roles and identification of compounds that target these auxiliary factors could be useful in designing effective combination therapies with β-lactams to treat MRSA.

Since TAs and their modifications perform such important functions in cell survival, virulence, and β-lactam resistance, they are being investigated for their potential in combination therapies and as anti-virulence targets (Fig. 5). Tunicamycin, a well-known natural product inhibitor of the first step for WTA synthesis (Hancock et al. 1976), has been shown to restore β-lactam susceptibility in MRSA (Campbell et al. 2011). Although tunicamycin is toxic to eukaryotes, potent, non-toxic TarO inhibitors could have great potential (Farha et al. 2014). In addition, the conditionally essential nature of the WTA pathway has been exploited in a pathway-specific screen to identify downstream inhibitors with antibiotic activity (Swoboda et al. 2009). Targocil and several other downstream inhibitors of the ABC transporter (TarGH) that exports WTA polymers have been reported (Lee et al. 2010; Campbell et al. 2012; Suzuki et al. 2011b; Wang et al. 2013). An inhibitor of LTA polymerization (compound 1771, [2-oxo-2-(5-phenyl-1,3,4-oxodiazol-2ylamino-ethyl-2-naphtho[2,1-b] furan-1-ylacetate]) was also described recently (Richter et al. 2013). Finally, due to its numerous roles in adhesion, virulence, and biofilm formation, the D-alanylation pathway is a potential candidate for anti-virulence therapy. A compound that inhibits the first enzyme in the pathway has been reported (May et al. 2005), but has not been shown to inhibit D-alanylation in cells. Agents that inhibit biofilm formation and adhesion mediated by other factors are being actively investigated as well (Chen et al. 2013). Inhibitors of TAs and their modifications are yet to make it to the clinic (Silver 2013), although late stage WTA inhibitors have shown some efficacy in combination with MRSA in animal models (Wang et al. 2013).

Acknowledgements Suzanne Walker is supported by funds from the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award numbers R01AI099144 and P01AI083214.

References

Abachin E, Poyart C, Pellegrini E, Milohanic E, Fiedler F, Berche P, Trieu-Cuot P (2002) Formation of D-alanyl-lipoteichoic acid is required for adhesion and virulene of *Listeria monocytogenes*. Mol Microbiol 43:1–14

Abee T, Kovács AT, Kuipers OP, van der Veen A (2011) Biofilm formation and dispersal in Gram-positive bacteria. Curr Opin Biotechnol 22:172–179

Allison SE, D'Elia MA, Arar S, Monteiro MA, Brown ED (2011) Studies of the genetics, function, and kinetic mechanism of TagE, the wall teichoic acid glycosyltransferase in *Bacillus subtilis* 168. J Biol Chem 286(27):23708–23716. http://doi.org/10.1074/jbc.M111.241265

Aly R, Shinefield HR, Litz C, Maibach HI (1980) Role of teichoic acid in the binding of *Staphylococcus aureus* to nasal epithelial cells. J Infect Dis 141(4):463–465

Anderson JS, Matsuhashi M, Haskin MA, Strominger JL (1967) Biosynthesis of the peptidoglycan of bacterial cell walls. II. Phospholipid carriers in the reaction sequence. J Biol Chem 242 (13):3180–3190

Archibald AR, Baddiley J, Heptinstall S (1973) The alanine ester content and magneisum binding capacity of walls of *Staphylococcus aurues* H grown at different pH values. Biochim Biophys Acta 291(3):629–634

Arciola CR, Campoccia D, Ravaioli S, Montanaro L (2015) Polysaccharide intercellular adhesin in biofilm: structural and regulatory aspects. Front Cell Infect Microbiol 5(7):1–10

Armstrong JJ, Baddiley J, Buchanan JG (1960) Structure of the ribitol teichoic acid from the walls of *Bacillus subtilis*. Biochem J 76(3):610–621

Arthur M, Courvalin P (1993) Genetics and mechanisms of glycopeptide resistance in enterococci. Antimicrob Agents Chemother 37(8):1563–1571

Atilano ML, Pereira PM, Yates J, Reed P, Veiga H, Pinho MG, Filipe SR (2010) Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in *Staphylococcus aureus*. Proc Natl Acad Sci USA 107(44):18991–18996

Aubry C, Goulard C, Nahori MA, Cayet N, Decalf J, Sachse M, Boneca IG, Cossart P, Dussurget O (2011) OatA, a peptidoglycan *O*-acetyltransferase involved in *Listeria monocytogenes* immune escape, is critical for virulence. J Infect Dis 204(5):731–740

Bagnoli F, Bertholet S, Grandi G (2012) Inferring reasons for the failure of *Staphylococcus aureus* vaccines in clinical trials. Front Cell Infect Microbiol 2(16):1–4

Baur S, Rautenberg M, Faulstich M, Grau T, Severin Y, Unger C, Hoffmann WH, Rudel T, Autenrieth IB, Weidenmaier C (2014) A nasal epithelial receptor for *Staphylococcus aureus* WTA governs adhesion to epithelial cells and modulates nasal colonization. PLoS Pathog 10 (5):e1004089

Bayer AS, Schneider T, Sahl HG (2013) Mechanisms of daptomycin resistance in *Staphylococcus aureus*: role of the cell membrane and cell wall. Ann NY Acad Sci 1277:139–158

Benson TE, Marquardt JL, Marquardt AC, Etzkorn FA, Walsh CT (1993) Overexpresson, purification, and mechanistic study of UDP-*N*-acetylenolpyruvylglucosamine reductase. Biochemistry 32(8):2024–2030

Bera A, Herbert S, Jakob A, Vollmer W, Götz F (2005) Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan *O*-acetyltransferase OatA is the major determinant for lysozyme resistance of *Staphylococcus aureus*. Mol Microbiol 55(3):778–787

Bera A, Biswas R, Herbert S, Götz F (2006) The presence of peptidogylcan *O*-acetyltransferase in various staphylococcal species correlates with lysozyme resistance and pathogenicity. Infect Immun 74(8):4598–4604

Berger-Bächi B, Rohrer S (2002) Factors influencing methicillin resistance in staphylococci. Arch Microbiol 178(3):165–171

Bernard E, Rolain T, Courtin P, Guillot A, Langella P, Hols P, Chapot-Chartier M (2011) Characterization of *O*-acetylation of *N*-acetylglucosamine: a novel structural variation of bacterial peptidoglycan. J Biol Chem 286(27):23950–23958

Bhakdi S, Klonisch T, Nuber P, Fischer W (1991) Stimulation of monokine production by lipoteichoic acids. Infect Immun 59(12):4614–4620

Blake KL, O'Neill AJ, Mengin-Lecreulx D, Henderson PJ, Bostock JM, Dunsmore CJ, Simmons KJ, Fishwick CW, Leeds JA, Chopra I (2009) The nature of *Staphylococcus aureus* MurA and MurZ and approaches for detection of peptidoglycan biosynthesis inhibitors. Mol Microbiol 72(2):335–343

Blumberg PM, Strominger JL (1974) Interaction of penicillin with the bacterial cell: penicillin-binding proteins and penicillin-sensitive enzymes. Bacteriol Rev 38(3):291–335

Bogaert D, Hermans PW, Adrian PV, Rümke HC, de Groot R (2004) Pneumococcal vaccines: an update on current strategies. Vaccine 22(17–18):2209–2220

Boneca IG, Huang ZH, Gage DA, Tomasz A (2000) Characterization of *Staphylococcus aureus* cell wall glycan strands, evidence for a new β-*N*-acetylglucosaminidase activity. J Biol Chem 275(14):9910–9918

Bouhss A, Mengin-Lecreulx D, Blanot D, van Heijenoort J, Parquet C (1997) Invariant amino acids in the Mur peptide synthetases of bacterial peptidoglycan synthesis and their modification by site-directed mutagenesis in the UDP-MurNAc:L-alanine ligase from *Escherichia coli*. Biochemistry 36(39):11556–11563

Bouhss A, Josseaume N, Severin A, Tabei K, Hugonnet JE, Shlaes D, Mengin-Lecreulx D, van Heijenoort J, Arthur M (2002) Synthesis of the L-alanyl-L-alanyl cross-bridge of *Enterococcus faecalis* peptidoglycan. J Biol Chem 277(48):45935–45941

Bouhss A, Crouvoisier M, Blanot D, Mengin-Lecreulx D (2004) Purification and characteriztaion of the bacterial MraY translocase catalyzing the first membrane step of peptidoglycan synthesis. J Biol Chem 279:29974–29980

Bouhss A, Trunkfield AE, Bugg TD, Mengin-Lecreulx D (2008) The biosynthesis of peptidoglycan lipid-linked intermediates. FEMS Microbiol Rev 32(2):208–233

Boylan RJ, Mendelson NH, Brooks D, Young FE (1972) Regulation of the bacterial cell wall: analysis of a mutant of *Bacillus subtilis* defective in biosynthesis of teichoic acid. J Bacteriol 110(1):281–290

Brötz H, Josten M, Wiedemann I, Schneider U, Götz F, Bierbaum G, Sahl HB (2002) Role of lipid-bound peptidoglycan precursors in the formation of pores by nisin, epidermin and other lantibiotics. Mol Microbiol 30(2):317–327

Brown S, Zhang YH, Walker S (2008) A revised pathway proposed for *Staphylococcus aureus* wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps. Chem Biol 346(17):2816–2819

Brown S, Meredith T, Swoboda J, Walker S (2010) *Staphylococcus aureus* and *Bacillus subtilis* W23 make polyribitol wall teichoic acids using different enzymatic pathways. Chem Biol 17 (10):1101–1110

Brown S, Xia G, Luhachack LG, Campbell J, Meredith TC, Chen C, Winstel V, Gekeler C, Irazoqui JE, Peschel A, Walker S (2012) Methicillin resistance in *Staphylococcus aureus* requires glycosylated wall teichoic acids. Proc Natl Acad Sci USA 109(46):18909–18914

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

Bugg TD, Wright GD, Dutka-Malen S, Arthur M, Courvalin P, Walsh CT (1991) Molecular basis for vancomycin resistance in *Ente*rococcus faecium BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. Biochem 30:10408–10415

Buist G, Steen A, Kok J, Kuipers OP (2008) LysM, a widely distributed protein motif for binding to (peptido) glycans. Mol Microbiol 68(4):838–847

Bunk S, Sigel S, Metzdorf D, Sharif O, Triantafilou K, Triantafilou M, Hartung T, Knapp S, von Aulock S (2010) Internalization and coreceptor expression are critical for TLR2-mediated recognition of lipoteichoic acid in human peripheral blood. J Immunol 185(6):3708–3717

Bychowska A, Theilacker C, Czerwicka M, Marszewska K, Huebner J, Holst O, Stepnowski P, Kaczyński Z (2011) Chemical structure of wall teichoic acid isolated from *Enterococcus faecium* strain U0317. Carbohydr Res 346(17):2816–2819

Cabanes D, Dehoux P, Dussurget O, Frangeul L, Cossart P (2002) Surface proteins and the pathogenic potential of *Listeria monocytogenes*. Trends Microbiol 10(5):238–245

Calix JJ, Nahm MH (2010) A new pneumococcal serotype, 11E, has variably inactivated *wcjE* gene. J Infect Dis 202(1):29–38

Calix JJ, Brady AM, Du VY, Saad JS, Nahm MH (2014) Spectrum of pneumococcal serotype 11A variants results from incomplete loss of capsule *O*-acetylation. J Clin Microbiol 52(3):758–765

Campbell J, Singh AK, Santa Maria JP, Kim Y, Brown S, Swoboda JG, Mylonakis E, Wilkinson BJ, Walker S (2011) Synthetic lethal compound combinations reveal a fundamental

connection between wall teichoic acid and peptidoglycan biosyntheses in *Staphylococcus aureus*. ACS Chem Biol 6(1):106–116
Campbell J, Singh AK, Swoboda JG, Gilmore MS, Wilkinson BJ, Walker S (2012) An antibiotic that inhibits a late step in wall teichoic acid biosynthesis induces the cell wall stress stimulon in *Staphylococcus aureus*. Antimicrob Agents Chemother 56(4):1810–1820
Campeotto I, Percy MG, MacDonald JT, Förster A, Freemont PS, Gründling A (2014) Structural and mechanistic insight into the *Listeria monocytogenes* two-enzyme lipoteichoic acid synthesis system. J Biol Chem 289(41):28054–28069
Carlin AF, Lewis AL, Varki A, Nizet V (2007) Group B Streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. J Bacteriol 189(4):1231–1237
Carlin AF, Uchiyama S, Chang YC, Lewis AL, Nizet V, Varki A (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. Blood 113(14):3333–3336
Chan YG, Frankel MB, Dengler V, Schneewind O, Missiakas D (2013) *Staphylococcus aureus* mutants lacking the LytR-CpsA-Psr family of enzymes release cell wall teichoic acids into the extracellular medium. J Bacteriol 195(20):4650–4659
Chan YG, Kim HK, Schneewind O, Missiakas D (2014) The capsular polysaccharide of *Staphyloccocus aureus* is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes. J Biol Chem 289(22):15680–15690
Chang S, Sievert DM, Hageman JC, Boulton ML, Tenover FC, Downes FP et al (2003) Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. N Engl J Med 348:1342–1347
Chatterjee AN (1969) Use of bacteriophage-resistant mutants to study the nature of the bacteriophage receptor site of *Staphylococcus aureus*. J Bacteriol 98(2):519–527
Chaudhuri RR, Allen AG, Owen PJ, Shalom G et al (2009). Comprehensive dentification of essential *Staphylococcus aureus* genes using transpson-mediated differential hybridisation (TMDH). BMC Genomics 10(291). doi:10.1186/1471-2164-10-291
Chen M, Yu Q, Sun H (2013) Novel strategies for the prevention and treatment of biofilm related infections. Int J Mol Sci 14(9):18488–18501
Chung BC, Zhao J, Gillespie RA, Kwon DY, Guan Z, Hong J, Zhou P, Lee SY (2013) Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. Science 341 (6149):1012–1016
Claessen D, Emmins R, Hamoen LW, Daniel RA, Errington J, Edwards DH (2008) Control of the cell elongation-division cycle by shuttling of PBP1 protein in *Bacillus subtilis*. Mol Microbiol 68(4):1029–1046
Clarke AJ, Dupont C (1992) *O*-acetylated peptidoglycan: its occurrence, pathobiological significance, and biosynthesis. Can J Microbiol 38(2):85–91
Clejan S, Krulwich TA, Mondrus KR, Seto-Young D (1986) Membrane lipid composition of obligatively and facultatively alkalophilic strains of *Bacillus* spp. J Bacteriol 168(1):334–340
Coffey TJ, Enright MC, Daniels M, Morona JK, Morona R, Hryniewicz W, Paton JC, Spratt BG (1998) Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. Mol Microbiol 27(1):73–83
Collins LV, Kristian SA, Weidenmaier C, Faigle M, van Kessel KP, van Strijp JA, Götz F, Neumeister B, Peschel A (2002) *Staphylococcus aureus* strains lacking D-alanine modifications of teichoic acids are highly susceptible to human neutrophil killing and are virulence attenuated in mice. J Infect Dis 186(2):214–219
Cook J, Hepler R, Pancari G, Kuklin N, Fan H, Wang XM, Cope L, Tan C, Joyce J, Onishi J, Montogomery D, Anderson A, McNeely T (2009) *Staphylcoccus aurues* capsule type 8 antibodies provide inconsistent efficacy in murine models of staphylococcal infection. Hum Vaccin 5(4):254–263

Corrigan RM, Abbott JC, Burhenne H, Kaever V, Gründling A (2011) c-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. PLoS Pathog 7(9):e1002217

Corrigan RM, Campeotto I, Jeganathan T, Roelofs KG, Lee VT, Gründling A (2013) Systematic identification of conserved bacterial c-di-AMP receptor proteins. Proc Natl Acad Sci USA 110 (22):9084–9089

Creech CB, Johnson BG, Alsentzer AR, Hohenboken M, Edwards KM, Talbot TR 3rd (2009) Vaccination as infection control: a pilot study to determine the impact of *Staphylococcus aureus* vaccination on nasal carriage. Vaccine 28(1):256–260

Crisóstomo MI, Vollmer W, Kharat AS, Inhülsen S, Gehre F, Buckenmaier S, Tomasz A (2006) Attenuation of penicillin resistance in a peptidoglycan *O*-acetyltransferase mutant of *Streptococcus pneumoniae*. Mol Microbiol 61(6):1497–1509

Cunnion KM, Zhang HM, Frank MM (2003) Availability of complement bound to *Staphylococcus aureus* to interact with membrane complement receptors influences efficiency of phagocytosis. Infect Immun 71(2):656–662

D'Elia MA, Pereira MP, Chung YS, Zhao W, Chau A, Kenney TJ, Sulavik MC, Black TA, Brown ED (2006) Lesions in teichoic acid biosynthesis in *Staphylococcus aureus* lead to a lethal gain of function in the otherwise dispensable pathway. J Bacteriol 188(12):4183–4189

D'Elia MA, Millar KE, Bhavsar AP, Tomljenovic AM, Hutter B, Schaab C, Moreno-Hagelsieb G, Brown ED (2009) Probing teichoic acid genetics with bioactive molecules reveals new interactions among diverse processes in bacterial cell wall biogenesis. Chem Biol 16 (5):548–556

Dagan R, Poolman J, Siegrist CA (2010) Glycoconjugate vaccines and immune interference: A review. Vaccine 28(34):5513–5523

Daniel RA, Harry EJ, Errington J (2000) Role of penicillin-binding protein PBP 2B in assembly and functioning of the division machinery of *Bacillus subtilis*. Mol Microbiol 35(2):299–311

Davies TA, Page MG, Shang W, Andrew T, Kania M, Bush K (2007) Binding of ceftobiprole and comparators to the penicillin-binding proteins of *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, and *Streptococcus pneumoniae*. Antimicrob Agents Chemother 51 (7):2621–2624

Davis KM, Weiser JN (2011) Modifications to the peptidoglycan backbone help bacteria to establish infection. Infect Immun 79(2):562–570

De Jonge BL, Sidow T, Chang Y, Labischinski H, Berger-Bächi B, Gage DA, Tomasz A (1993) Altered muropeptide composition in *Staphylococcus aureus* strains with an inactivated *femA* Locus. J Bacteriol 175(9):2779–2782

De Jonge BL, Handwerger S, Gage D (1996) Altered peptidoglycan composition in vancomycin-resistant *Enterococcus faecalis*. Antimicrob Agents Chemother 40(4):863–869

De Leeuw E, Li C, Zeng P, Li C, Diepeveen-de Buin M, Lu WY, Breukink E, Lu W (2010) Functional interaction of human neutrophil peptide-1 with the cell wall precursor Lipid II. FEBS Lett 584(8):1543–1548

De Mendoza D (2014) Temperature sensing by membranes. Annu Rev Microbiol 68:101–116

De Velasco EA, Merkus D, Anderton S, Verheul AF, Lizzio EF, Van der Zee R, Van Eden W, Hoffman T, Verhoef J, Snippe H (1995) Synthetic peptides representing T-cell epitopes act as carriers in pneumococcal polysacchride conjugate vaccines. Infect Immun 63(3):961–968

Denapaite D, Brückner R, Hakenbeck R, Vollmer W (2012) Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: lessons from genomes. Microb Drug Resist 18(3):344–358

Dengler V, Meier PS, Heusser R, Kupferschmied P, Fazekas J, Friebe S, Staufer SB, Majcherczyk PA, Moreillon P, Berger-Bächi McCallum N (2012) Deletion of hypothetical wall teichoic acid ligases in *Staphylococcus aureus* activates the cell wall stress response. FEMS Microbiol Lett 333(2):09–120

Depardieu F, Podglajen I, Leclercq R, Collatz E, Courvalin P (2007) Modes and modulations of antibiotic resistance gene expression. Clin Microbiol Rev 20(1):79–114

Draing C, Pfitzemaier M, Zummo S, Mancuso G, Geyer A, Hartung T, von Aulock S (2006) Comparison of lipoteichoic acid from different serotypes of *Streptococcus pneumoniae*. J Biol Chem 281(45):33849–33859

Draing C, Sigel S, Deininger S, Traub S, Munke R, Mayer C, Hareng L, Hartung T, von Aulock S, Hermann C (2008) Cytokine induction by Gram-positive bacteria. Immunobiology 213(3–4):285–296

Drawz SM, Papp-Wallace KM, Bonomo RA (2014) New β-lactamase inhibitors: a therapeutic renaissance in an MDR world. Antimicrob Agents Chemother 58(4):1835–1846

Du W, Brown JR, Sylvester DR, Huang J, Chalker AF, So CY, Holmes DJ, Payne DJ, Wallis NG (2000) Two active forms of UDP-*N*-acetylglucosamine enolpyruvyl transferase in Gram-positive bacteria. J Bacteriol 182(15):4146–4152

El Ghachi M, Bouhss A, Blanot D, Mengin-Lecreulx D (2004) The *bacA* gene of *Escherichia coli* encodes an undecaprenyl pyrophosphate phosphatase activity. J Biol Chem 279(29):30106–30113

El Ghachi M, Derbise A, Bouhss A, Mengin-Lecreulx D (2005) Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in *Escherichia coli*. J Biol Chem 280(19):18689–18695

Epand RF, Savage PB, Epand RM (2007) Bacterial lipid composition and the antimicrobial efficacy of cationic steroid compounds (ceragenins). Biochim Biophys Acta 1768(10):2500–2509

Ernst CM, Peschel A (2011) Broad-spectrum antimicrobial peptide resistance by MprF-mediated aminoacylation and flipping of phospholipids. Mol Microbiol 80(2):290–299

Ernst CM, Staubitz P, Mishra NN, Yang SJ, Hornig G, Kalbacher H, Bayer AS, Kraus D, Peschel A (2009) The bacterial defensin resistance protein MprF consists of separable domains for lipid lysinylation and antimicrobial peptide repulsion. PLoS Pathog 5(11):e1000660

Fabretti F, Theilacker C, Baldassarri L, Kaczynski Z, Kropec A, Holst O, Huebner J (2006) Alanine esters of enterococcal lipoteichoic acid play a role in biofilm formation and resistance to antimicrobial peptides. Infect Immun 74(7):4164–4171

Falord M, Mäder U, Hiron A, Débarbouillé M, Msadek T (2011) Investigation of the *Staphylococcus aureus* GraSR regulon reveals novel links to virulence, stress response and cell wall signal transduction pathways. PLoS ONE 6(7):e21323

Farha MA, Koteva K, Gale RT, Sewell EW, Wright GD, Brown ED (2014) Designing analogs of ticlopidine, a wall teichoic acid inhibitor, to avoid formation of its oxidative metabolites. Bioorg Med Chem Lett 24(3):905–910

Fattom AI, Horwith G, Fuller S, Propst M, Naso R (2004) Development of StaphVAX, a polysaccharide conjugate vaccine against *S. aureus* infection: from lab bench to phase III clinical trials. Vaccine 22(7):880–887

Fiedel BA, Jackson RW (1978) Activation of the alternative complement pathway by a streptococcal lipoteichoi acid. Infect Immun 22(1):286–287

Figueiredo TA, Sobral RG, Ludovice AM, Almeida JM, Bui NK, Vollmer W, de Lancastre H, Tomasz A (2012) Identification of genetic determinants and enzymes involved with the amidation of glutamic acid residues in the peptidoglycan of *Staphylococcus aureus*. PLoS Pathog 8(1):e1002508

Filipe SR, Pinho MG, Tomasz A (2000) Characterization of the *murMN* operon involved in the synthesis of branched peptidoglycan peptides in *Streptococcus pneumoniae*. J Biol Chem 275 (36):22768–27774

Fischer W (1988) Physiology of lipoteichoic acids in bacteria. Adv Microb Phsyiol 29:233–302

Fischer W (2000) Phosphocholine of pneumococcal teichoic acids: role in bacterial physiology and pneumococcal infection. Res Microbiol 151(6):421–427

Fischer W, Behr T, Hartmann R, Peter-Katalinić J, Egge H (1993) Teichoic acids and lipoteichoic acid of Streptococcus pneumoniae possess identical chain structures. Eur J Biochem 215 (3):851–857

Fittipaldi N, Sekizaki T, Takamatsu D, Harel J, MdeL Domínguez-Punaro, von Aulock SV, Draing C, Marois C, Kobisch M, Gottschalk M (2008) D-alanylation of lipoteichoic acid contributes to the virulence of *Streptococcus suis*. Infect Immun 76(8):3587–3594

Friedman L, Alder JD, Silverman JA (2006) Genetic changes that correlate with reduced susceptibility to Daptomycin in *Staphylococcus aureus*. Antimicrob Agents Chemother 50 (6):2137–2145

Fuda C, Suvorov M, Vakulenko SB, Mobashery S (2004) The basis for resistance to β-lactam antibiotics by penicillin-binding protein 2A of Methicillin-resistant *Staphylococcus aureus*. J Biol Chem 279:40802–40806

Fukushima T, Yamamoto H, Atrih A, Foster SJ, Sekiguchi J (2002) A polysaccharide deacetylase gene (*pdaA*) is required for germination and for production of muramic δ-lactam residues in the spore cortex of *Bacillus subtilis*. J Bacteriol 184(21):6007–6015

Fukushima T, Kitajima T, Sekiguchi J (2005) A polysaccharide deacetylase homologue, PdaA, in *Bacillus subtilis* acts as an *N*-acetylmuramic acid deacetylase *in vitro*. J Bacteriol 187 (4):1287–1292

Garcia-Bustos JF, Chait BT, Tomasz A (1987) Structure of the peptide network of pneumococcal peptidoglycan. J Biol Chem 262(32):15400–15405

Gardete S, Tomasz A (2014) Mechanisms of vancomycin resistance in *Staphylococcus aureus*. J Clin Invest 124(7):2836–2840

Gerke C, Kraft A, Süssmuth R, Schweitzer O, Götz F (1998) Characterization of the *N*-acetylglucosaminyltransferase activity involved in the biosynthesis of the *Staphylococcus epidermidis* polysaccharide intercellular adhesin. J Biol Chem 273(29):18586–18593

Ghuysen JM (1991) Serine β-lactamases and penicillin-binding proteins. Annu Rev Microbiol 45:37–67

Ginsburg I (2002) Role of lipoteichoic acid in infection and inflammation. Lancet Infect Dis 2 (3):171–179

Giudicelli S, Tomasz A (1984) Attachment of pneumococcal autolysin to wall teichoic acids, an essential step in enzymatic wall degradation. J Bacteriol 158(3):1188–1190

Gosink KK, Mann ER, Guglielmo C, Tuomanen EI, Masure HR (2000) Role of novel choline binding proteins in virulence of *Streptococcus pneumoniae*. Infect Immun 68(10):5690–5695

Gross M, Cramton SE, Götz F, Peschel A (2001) Key role of teichoic acid net charge in *Staphylococcus aureus* colonization of artificial surfaces. Infect Immun 69(5):3423–3426

Gründling A, Schneewind O (2007a) Genes required for glycolipid synthesis and lipoteichoic acid anchoring in *Staphylococcus aureus*. J Bacteriol 189(6):2521–2530

Gründling A, Schneewind O (2007b) Synthesis of glycerol phosphate lipoteichoic acid in *Staphylococcus aureus*. Proc Natl Acad Sci USA 104(20):8478–8483

Gutkind GO, Di Conza J, Power P, Radice M (2013) β-lactamase-mediated resistance: a biochemical, epidemiological and genetic overview. Curr Pharm Des 19(2):164–208

Haas R, Koch HU, Fischer W (1984) Alanyl turnover from lipoteichoic acid to teichoic acid in *Staphylococcus aureus*. FEMS Microbiol Lett 21(1):27–31

Hakenbeck R, Madhour A, Denapaite D, Brückner R (2009) Versatality of choline metabolism and choline-binding proteins in *Streptococcus pneumoniae* and commensal streptococci. FEMS Microbiol Rev 33(3):572–586

Hall RG, Michaels HN (2015) Profile of tedizolid phosphate and its potential in the treatment of acute bacterial skin and skin structure infections. Infect Drug Resist 8:75–82

Hall-Stoodley L, Costerton W, Stoodley P (2004) Bacterial biofilms: from the natural enviroment to infectious diseases. Nat Rev Microbiol 2:95–108

Hancock IC (1997) Bacterial cell surface carbohydrates: structure and assembly. Biochem Soc Trans 25(1):183–187

Hancock IC, Wiseman G, Baddiley J (1976) Biosynthesis of the unit that links teichoic acid to the bacterial wall: inhibition by tunicamycin. FEBS Lett 69(1):75–80

Handwerger S, Pucci MJ, Volk KJ, Liu J, Lee MS (1992) The cytoplasmic peptidoglycan precursor of vancomycin-resistant *Enterococcus faecalis* terminates in lactate. J Bacteriol 174 (18):5982–5984

Haque MA, Russell NJ (2004) Strains of *Bacillus cereus* vary in the phenotypic adaptation of their membrane lipid composition in response to low water activity, reduced temperature and growth in rice starch. Microbiol 150(Pt 5):1397–1404

Hartman BJ, Tomasz A (1984) Low-affinity penicillin-binding protein associated with β-lactam resistance in *Staphylococcus aurues*. J Bacteriol 158(2):513–516

Hashimoto M, Tawaratsmida K, Kariya H, Aoyama K, Tamura T, Suda Y (2006a) Lipoprotein is a dominant Toll-like receptor 2 ligand in *Staphylococcus aureus* cell wall components. Int Immunol 18(2):355–362

Hashimoto M, Tawaratsmida K, Kariya H, Kiyohara A, Suda Y, Krikae F, Kirikae T, Götz F (2006b) Not lipoteichoic acid but lipoprotein appear to be the dominant immunobiologically active compounds in *Staphylococcus aureus*. J Immunol 177(5):3162–3169

Hayashi H, Araki Y, Ito E (1973) Occurence of glucosamine residues with free amino groups in cell wall peptidoglycan from bacilli as a factor responsible for resistance to lysozyme. J Bacteriol 113(2):592–598

Hayhurst EJ, Kailas L, Hobbs JK, Foster SJ (2008) Cell wall peptidoglycan architecture in *Bacillus subtilis*. Proc Natl Acad Sci USA 105(38):14603–14608

Healy VL, Lessard IA, Roper DI, Knox JR, Walsh CT (2000) Vancomycin resistance in enterococci: reprogramming of the D-ala-D-ala ligases in bacterial peptidoglycan biosynthesis. Chem Biol 7(5):109–119

Heaslet H, Shaw B, Mistry A, Miller A (2009) Characterization of the active site of *S. aureus* monofunctional transglycosylase (Mtg) by site-directed mutation and structural analysis of the protein complexed with moenomycin. J Struct Biol 167(2):129–135

Heaton MP, Neuhaus FC (1992) Biosynthesis of D-alanyl-lipoteichoic acid: cloning, nucleotide sequence, and expression of the *Lactobacillus casei* gene for the D-alanine-activating enzyme. J Bacteriol 174(14):4707–4717

Heaton MP, Neuhaus FC (1994) Roles of the D-alanyl carrier protein in the biosynthesis of D-alanyl-lipoteichoic acid. J Bacteriol 176(3):681–690

Hebert L, Courtin P, Torelli R, Sanguinetti M, Chapot-Chartier MP, Auffray Y, Benachour A (2007) *Enterococcus faecalis* constitutes an unusual bacterial model in lysozyme resistance. Infect Immun 75(11):5390–5398

Heilmann C, Schweitzer O, Gerke C, Vanittanakom N, Mack D, Götz F (1996) Molecular basis of intercellular adhesion in the biofilm-forming *Staphylococcus epidermidis*. Mol Microbiol 20 (5):1083–1091

Henze U, Sidow T, Wecke J, Labischinski H, Berger-Bächi B (1993) Influence of *femB* on methicillin resistance and peptidoglycan metabolism in *Staphylococcus aureus*. J Bacteriol 175 (6):1612–1620

Hoffman K (2000) A superfamily of membrane-bound *O*-acetyltansferases with implications for wnt signalling. Trends Biochem Sci 25(3):111–112

Holmes NE, Howden BP (2014) What's new in the treatment of serious MRSA infection? Curr Opin Infect Dis 27(6):471–478

Hong DJ, Bae IK, Jang IH, Jeong SH, Kang HK, Lee K (2015) Epidemiology and characteristics of metall-β-lactamase-producing *Pseudomonas aeruginosa*. Infect Chemother 47(2):81–97

Hsu ST, Breukink E, Tischenko E, Lutters MA, de Kruijiff B, Kaptein R, Bonvin AM, van Nuland NA (2004) The nisin–lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. Nat Struct Mol Biol 11(10):963–967

Hu Y, Chen L, Ha S, Gross B, Falcone B, Walker D, Mokhtarzadeh M, Walker S (2003a) Crystal structure of the MurG: UDP-GlcNAc complex reveals common structural principles of a superfamily of glycosyltransferases. Proc Natl Acad Sci USA 100(3):845–849

Hu Y, Helm JS, Chen L, Ya XY, Walker S (2003b) Ramoplanin inhibits bacterial transglycosylases by binding as a dimer to Lipid II. J Am Chem Soc 125(29):8736–8737

Huber J, Donald RG, Lee SH, Jarantow LW, Salvatore MJ, Meng X, Painter R, Onishi RH, Occi J, Dorso K, Young K, Park YW, Skwish S, Szymonifka MJ, Waddell TS, Miesel L, Phillips JW, Roemer T (2009) Chemical genetic identification of peptidoglycan inhibtors potentiating carbapenem activity against methicillin-resistant *Staphylococcus aureus*. Chem Biol 16:837–848

Itoh Y, Wang X, Hinnebusch BJ, Preston JF, Romeo T (2005) Depolymerization of β-1,6-*N*-acetyl-D-glucosamine disrupts the integrity of diverse bacterial biofilms. J Bacteriol 187 (1):382–387

Jedrzejas MJ (2001) Pneumococcal virulence factors: structure and function. Microbiol Mol Biol Rev 65(2):187–207

Jefferies JM, Clarke SC, Webb JS, Kraaijeveld AR (2011) Risk of red queen dynamics in pneumococcal vaccine strategy. Trends Microbiol 19(8):377–381

Jerga A, Lu YJ, Schujman GE, de Mendoza D, Rock CO (2007) Identification of a soluble diacylglycerol kinase required for lipoteichoic acid production in *Bacillus subtilis*. J Biol Chem 282(30):21738–21745

Jervis AJ, Thackray PD, Houston CW, Horsburgh MJ, Moir A (2007) SigM-responsive genes of *Bacillus subtilis* and their promoters. J Bacteriol 189(12):4534–4538

Jett BD, Huycke MM, Gilmore MS (1994) Virulence of enterococci. Clin Microbiol Rev 7 (4):462–478

Jones T, Yeaman MR, Sakoulas G, Yang SJ, Proctor RA, Sahl HG, Schrenzel J, Xiong YQ, Bayer AS (2008) Failures in clinical treatment of *Staphylococcus aureus* infection with daptomycin are associated with alterations in surface charge, membrane phospholipid asymmetry, and drug binding. Antimicrob Agents Chemother 52(1):269–278

Jonquières R, Bierne H, Fiedler F, Gounon P, Cossart P (1999) Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of Gram-positive bacteria. Mol Microbiol 34(5):902–914

Jorasch P, Wolter FP, Zähringer U, Heinz E (1998) A UDP glucosyltransferase from *Bacillus subtilis* successively transfer up to four glucose residues to 1,2-diacylglycerol: expression of *ypfP* in *Escherichia coli* and structural analysis of its reaction product. Mol Microbiol 29 (2):419–430

Julian K, Kosowska-Shick K, Whitener C, Roos M, Labischinski H, Rubio A, Parent L, Ednie L, Koeth L, Bogdanovich T, Appelbaum PC (2007) Characterization of a daptomycin-nonsusceptible vancomycin-intermediate *Staphylococcus aureus* strain in a patient with endocarditis. Antimicrob Agents Chemother 51(9):3445–3448

Kahan FM, Kahan JS, Cassidy PJ, Kropp H (1974) The mechanism of action of fosfomycin (phosphonomycin). Ann NY Acad Sci 235:364–386

Kalin M (1998) Pneumococcal serotypes and their clinical relevance. Thorax 53(3):159–162

Karatsa-Dodgson M, Wörmann ME, Gründling A (2010) *In vitro* anaylsis of the *Staphylococcus aureus* lipoteichoic acid synthase enzyme using fluorescently labeled lipids. J Bacteriol 192 (20):5341–5349

Kawai Y, Marles-Wright J, Cleverley RM, Emmins R, Ishikawa S, Kuwano M, Heinz N, Bui NK, Hoyland CN, Ogasawara N, Lewis RJ, Vollmer W, Daniel RA, Errington J (2011) A widespread family of bacterial cell wall assembly proteins. EMBO J 30(24):4931–4941

Keller R, Fischer W, Keist R, Bassetti S (1992) Macrophage response to bacteria: induction of marked secretory and cellular activities by lipoteichoic acids. Infect Immun 60(9):3664–3672

Kenne L, Lindberg B, Svensson S (1975) The structure of capsular polysaccharide of the pneumococcus type II. Carbohydr Res 40:69–75

Kharat AS, Tomasz A (2006) Drastic reduction in the virulence of *Streptococcus pneumoniae* expressing type 2 capsular polysaccharide but lacking choline residues in the cell wall. Mol Microbiol 60(1):93–107

Kiriukhin MY, Debabov DV, Shinabarger DL, Neuhaus FC (2001) Biosynthesis of the glycolipid anchor in lipoteichoic acid of *Staphylococcus aureus* RN4220: role of YpfP, the diglucosyl-diacylglycerol synthase. J Bacteriol 183(11):3506–3514

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M et al (2003) Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A 100(8):4678–4683

Kobayashi K, Sudiarta P, Kodoma T, Fukushima T, Ara K, Ozaki K, Sekiguchi J (2012) Identification and characterization of a novel polysaccharide deacetylase C (PdaC) from *Bacillus subtilis*. J Biol Chem 287(13):9765–9776

Koch HU, Döker R, Fischer W (1985) Maintenance of D-alanine ester substitution of lipoteichoic acid by reesterification in *Staphylococcus aureus*. J Bacteriol 164(3):1211–1217

Kock H, Gerth U, Hecker M (2004) MurAA, catalyzing the first committed step in peptidoglycan biosynthesis, is a target for Clp-dependent proteolysis in *Bacillus subtilis*. Mol Microbiol 51 (4):1087–1102

Kohler T, Weidenmaier C, Peschel A (2009) Wall teichoic acid protects *Staphylococcus aureus* against antimicrobial fatty acids from human skin. J Bacteriol 191(13):4482–4484

Kojima N, Araki Y, Ito E (1985) Structure of the linkage units between ribitol teichoic acids and peptidoglycan. J Bacteriol 161(1):299–306

Komatsuzawa H, Sugai M, Ohta K, Fujiwara T, Nakashima S, Lee CY, Suginaka H (1997) Cloning and characterization of the *fmt* gene which affects the methicillin-resistance level and autolysis in the presence of triton X-100 in methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother 41(11):2355–2361

Komatsuzawa H, Ohta K, Fujiwara T, Choi GH, Labischinski H, Sugai M (2001) Cloning and sequencing of the gene, *fmtC*, which affects oxacillin resistance in methicillin-resistant *Staphylococcus aureus*. FEMS Microbiol Lett 203(1):49–54

Kovács M, Halfmann A, Fedtke I, Heintz M, Peschel A, Vollmer W, Hakenbeck R, Brückner R (2006) A functional *dlt* operon, encoding proteins required for incorporation of D-alanine in teichoic acids in Gram-positive bacteria, confers resistance to cationic antimicrobial peptides in *Streptococcus pneumoniae*. J Bacteriol 188(16):5797–5805

Kovacs-Simon A, Titball RW, Michell SL (2011) Lipoproteins of bacterial pathogens. Infect Immun 79(2):548–561

Kozarich JW, Strominger JL (1978) A membrane enzyme from *Staphylococcus aureus* which catalyzes transpeptidase, carboxypeptidase, and penicillinase activities. J Biol Chem 253 (4):1272–1278

Kresge N, Simoni RD, Hill RL (2007) t-RNA involvement in peptidoglycan synthesis: the work of Dieter Söll. J Biol Chem 282(26):e20–e21

Kristian SA, Dürr M, Van Strijp JA, Neumeister B, Peschel A (2003) MprF-mediated lysinylation of phospholipids in *Staphylococcus aureus* leads to protection against oxgen-independent neutrophil killing. Infect Immun 71(1):546–549

Kristian SA, Vivekanand D, Wiedenmaier C, Kansal R, Fedtke I, Peschel A, Gallo RL, Nizet V (2005) D-alanylation of teichoic acids promotes Group A *Streptococcus* antimicrobial peptide resistance, neutrophil survival, and epithelial cell invasion. J Bacteriol 187(19):6719–6725

Laaberki MH, Pfeffer J, Clarke AJ, Dworkin J (2011) *O*-acetylation of peptidoglycan is required for proper cell separation and S-layer anchoring in *Bacillus anthracis*. J Biol Chem 286 (7):5278–5288

Lavollay M, Arthur M, Fourgeaud M, Dubost L, Marie A, Veziris N, Blanot D, Gutmann L, Mainardi J (2008) The peptidoglycan of stationary-phase *Mycobacterium tuberculosis* predominantly contains cross-links generated by L,D-transpeptidation. J Bacteriol 190 (12):4360–4366

Lavollay M, Fourgeaud M, Hermann JL, Dubost L, Marie A, Gutmann L, Arthur M, Mainardi J (2011) The peptidoglycan of *Mycobacterium abscessus* is predominantly cross-linked by L,D-transpeptidases. J Bacteriol 193(3):778–782

Lazarevic V, Abellan FX, Möller SB, Karamata D, Mauël C (2002) Comparison of ribitol and glycerol teichoic acid genes in *Bacillus subtilis* W23 and 168: identical function, similar divergent organization, but different regulation. Microbiology 148(Pt 3):815–824

Leach KL, Brickner SJ, Noe MC, Miller PF (2011) Linezolid, the first ozazolidinone antibacterial agent. Ann NY Acad Sci 1222:49–54

Lebreton F, Depardieu F, Bourdon N, Fines-Guyon M, Berger P, Camiade S, Leclerq R, Courvalin P, Cattoir V (2011) D-ala-D-ser VanN-type transferable vancomycin resistance in *Enterococcus faecium*. Antimicrob Agents Chemother 55(10):4606–4612

Lee K, Campbell J, Swoboda JG, Cuny GD, Walker S (2010) Development of improved inhibitors of wall teichoic acid biosynthesis with potent activity against *Staphylococcus aureus*. Bioorg Med Chem Lett 20:1767–1770

Li M, Cha DJ, Lai Y, Villaruz AE, Sturdevant DE, Otto M (2007a) The antimicrobial peptide-sensing system *aps* of *Staphylococcus aureus*. Mol Microbiol 66(5):1136–1147

Li M, Lai Y, Villaruz AE, Cha DJ, Sturdevant DE, Otto M (2007b) Gram-positive three-component antimicrobial peptide-sensing system. Proc Natl Acad of Sci USA 104 (22):9469–9474

Lim D, Strynadka NC (2002) Structural basis for the β-lactam resistance of PBP2a from methicillin-resistant *Staphylocccus aureus*. Nat Struct Biol 9(11):870–876

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon P, Mueller A et al (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517(7535):455–459

Lo MC, Men H, Branstrom A, Helm J, Yao N, Goldman R, Walker S (2000) A new mechanism of action proposed for ramoplanin. J Am Chem Soc 122:3540–3541

Loos M, Clas F, Fisher W (1986) Interaction of purified lipoteichoic acid with classical complement pathway. Infect Immun 53(3):595–599

Lu D, Wörmann ME, Zhang X, Schneewind O, Gründling A, Freemont PS (2009) Structure-based mechanism of lipoteichoic acid synthesis of *Staphylococcus aureus* LtaS. Proc Natl Acad Sci USA 106(5):1584–1589

Ludovice AM, Wu SW, De Lencastre H (1998) Molecular cloning and DNA sequencing of the *Staphylococcus aureus* UDP-*N*-acetylmuramyl tripeptide synthetase (*murE*) gene, essential for the optimal expression of methicillin resistance. Microb Drug Resist 4(2):85–90

Lunderberg JM, Nguyen-Mau SM, Richter GS, Wang YT, Dworkin J, Missiakas DM, Schneewind O (2013) *Bacillus anthracis* acetyltransferases PatA1 and PatA2 modify the secondary cell wall polysaccharide and affect the assembly of S-layer proteins. J Bacteriol 195 (5):977–989

Mack D, Fischer W, Krokotsch A, Leopold K, Hartmann R, Egge H, Laufs R (1996) The intercellular adhesin involved in biofilm accumulation of *Staphylococcus epidermidis* is a linear β-1,6-linked glucosaminoglycan: purification and structural analysis. J Bacteriol 178 (1):175–183

MacLeod CM, Hodges RG, Heidelbrger M, Bernhard WG (1945) Prevention of pneumococcal pneumonia by immunization with specific capsular polysaccharides. J Exp Med 82(6):445–465

Mahapatra S, Yagi T, Belisle JT, Espinosa BJ, Hill PJ, McNeil MR, Brennan PJ, Crick D (2005) Mycobaterial Lipid II is composed of a complex mixture of muramyl and peptide moieties linked to a decaprenly phosphate. J Bacteriol 187(8):2747–2757

Maidhof H, Reinicke B, Blümel P, Berger-Bächi B, Labischinski H (1991) *femA*, which encodes a factor essential for methicillin resistance, affects glycine content of peptidoglycan in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains. J Bacteriol 173(11):3507–3513

Mainardi JL, Legrand R, Arthur M, Schoot B, van Heijenoort J, Gutmann L (2000) Novel mechanism of beta-lactam resistance due to bypass of Dᴅ-transpeptidation in *Enterococcus faecium*. J Biol Chem 275(22):16490–16496

Maira-Litrán T, Bentancor LV, Bozkurt-Guzel C, O'Malley JM, Cywes-Bentley C, Pier GB (2012) Synthesis and evaluation of a conjugate vaccine composed of *Staphylococcus aureus* poly-*N*-acetylglucosamine and clumping factor A. PLoS ONE 7(9):e43813

Marquardt JL, Siegele DA, Kolter R, Walsh CT (1992) Cloning and sequencing of *Escherichia coli murZ* and purification of its product, a UDP-*N*-acetylglucosamine enolpyruvyl transferase. J Bacteriol 174(17):5748–5752

Marshall CG, Lessard IA, Park I, Wright GD (1998) Glycopeptide antibiotic resistance genes in glycopeptide-producing organisms. Antimicrob Agents Chemother 42(9):2215–2220

Matsuoka S, Chiba M, Tanimura Y, Hashimoto M, Hara H, Matsumoto K (2011) Abnormal morphology of *Bacillus subtilis ugtP* mutant cells lacking glucolipids. Genes Genet Syst 86 (5):295–304

Mauël C, Young M, Karamata D (1991) Genes concerned with synthesis of poly(glycerol phosphate), the essential teichoic acid in *Bacillus subtilis* strain 168, are organized in two divergent transcription units. J Gen Microbiol 137(4):929–941

May JJ, Finking R, Wiegshoff F, Weber TT, Bandur N, Koert U, Marahiel MA (2005) Inhibition of the D-alanine:D-alanyl carrier protein ligase from *Bacillus subtilis* increases the bacterium's susceptibility to antibiotics that target the cell wall. FEBS J 272(12):2993–3003

McDaneld PM, Spooner LM, Mohr JF, Belliveau PP (2013) Use of daptomycin to treat infections with methicillin-resistant *Staphylococcus aureus* having vancomycin minimum inhibitory concentrations of 1.5 to 2 μg/mL. Ann Pharmacother 47(12):1654–1665

Meeske AJ, Sham LT, Kimsey H, Koo BM, Gross CA, Bernhardt TG, Rudner DZ (2015) MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis*. Proc Natl Acad Sci USA 112(20):6437–6442

Memmi G, Filipe SR, Pinho MG, Fu Z, Cheung A (2008) *Staphylococcus aureus* PBP4 is essential for β-lactam resistance in community-acquired methicillin-resistant strains. Antimicrob Agents Chemother 52(11):3955–3966

Mengin-Lecreulx D, Texier L, Rousseau M, van Heijenoort J (1991) The *murG* gene of *Escherichia coli* codes for the UDP-*N*-acetylglucosamine: *N*-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. J Bacteriol 173(15):4625–4636

Meredith TC, Swoboda JG, Walker S (2008) Late-stage polyribitol phosphate wall teichoic acid biosynthesis in *Staphylococcus aureus*. J Bacteriol 190(8):3046–3056

Minnikin DE, Abdolraimzadeh H (1974) Effect of pH on the proportions of polar lipds, in chemostat cultures of *Bacillus subtilis*. J Bacteriol 120(3):999–10003

Mishra NN, Bayer AS, Weidenmaier C, Grau T, Wanner S, Stefani S, Cafiso V, Bertuccio T, Yeaman MR, Nast CC, Yang SJ (2014) Phenotypic and genotypic characterization of daptomycin-resistant methicillin-resistant *Staphylococcus aureus* strains: relative roles of mprF and dlt operons. PLoS ONE 9(9):e107426

Mitchell P, Moyle J (1956) Osmotic structure and function in bacteria. Symp Soc Gen Microbiol 6:150–180

Mitra S, Saeed U, Havlichek DH, Stein GE (2015) Profile of oritavancin and its potential in the treatment of acute bacterial skin structure infections. Infect Drug Resist 8:189–197

Mohamadzadeh M, Pfeiler EA, Brown JB, Zadeh M, Gramarossa M, Managlia E, Bere P, Sarraj B, Khan MW, Pakanati KC, Ansari MJ, O'Flaherty S, Barrett T, Klaenhammer TR (2011) Regulation of induced colonic inflammation by *Lactobacillus acidophilus* deficient in lipoteichoic acid. Proc Natl Acad Sci USA 108(Supplement 1):4623–4630

Moisan H, Pruneau M, Malouin F (2010) Binding of ceftaroline to penicillin-binding proteins of *Staphylococcus aureus* and *Streptococcus pneumoniae*. J Antimicrob Chemother 5(4):713–716

Moynihan PJ, Sychantha D, Clarke AJ (2014) Chemical biology of peptidoglycan acetylation and deacetylation. Bioorg Chem 54:44–50

Münch D, Roemer T, Lee SH, Engeser M, Sahl HG, Schneider T (2012) Identification and in vitro analysis of the GatD/MurT enzyme-complex catalyzing Lipid II amdiation in *Staphylococcus aureus*. PLoS Pathog 8(1):e1002509

Navarre WW, Schneewind O (1999) Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev 63(1):174–229

Neuhaus FC, Baddiley J (2003) A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in Gram-positive bacteria. Microbiol Mol Biol Rev 67(4):686–723

Nishi H, Komatsuzawa H, Fujiwara T, McCallum N, Sugai M (2004) Reduced content of lysyl-phosphatidylglycerol in the cytoplasmic membrane affects susceptibility to moenomycin, as well as vancomycin, gentamicin, and antimicrobial peptides *Staphylococcus aureus*. Antimicrob Agents Chemother 48(12):4800–4807

Norris V, Sweeney S (1993) Deformations in the cytoplasmic membrane of *Escherichia coli* direct the repair of peptidoglycan. In: de Pedro MA, Höltje JV, Löffelhardt W (eds) Bacterial growth and lysis: metabolism and structure of the bacterial sacculus. Springer, US, pp 375–385

Nurhonen M, Auranen K (2014) Optimal serotype compositions for pneumococcal conjugate vaccination under serotype replacement. PLoS Comput Biol 10(2):e1003477

O'Gara JP (2007) *ica* and beyond: biofilm mechanisms and regulation in *Staphylococcus epidermidis* and *Staphylococcus aureus*. FEMS Microbiol Lett 270(2):179–188

O'Riordan K, Lee JC (2004) *Staphylococcus aureus* capsular polysaccharides. Clin Microbiol Rev 17(1):218–234

Ohno N, Yadomae T, Miyazaki T (1982) Identification of 2-amino-2-deoxyglucose residues in the peptidoglucan of *Streptococcus pneumoniae*. Carbohydr Res 107(1):152–155

Oku Y, Kurokawa K, Matsuo M, Yamada S, Lee BL, Sekimizu K (2009) Pleiotropic roles of polyglycerolphosphate synthase of lipoteichoic acid in growth of *Staphylococcus aureus* cells. J Bacteriol 191(1):141–151

Oliver MB, Jones C, Larson TR, Calix JJ, Zartler ER, Yother J, Nahm MH (2013a) *Streptococcus pneumoniae* serotype 11D has a bi-specific glycosyltransferase and expresses two different capsular polysaccharide repeating units. J Biol Chem 288(30):21945–21954

Oliver MB, van der Linden MP, Küntzel SA, Saad JS, Nahm MH (2013b) Discovery of *S*reptococcus pneumoniae serotype 6 variants with glycosyltransferases synthesizing two different repeating units. J Biol Chem 288(36):25976–25985

Oman TJ, Lupoli TJ, Wang TSA, Kahne D, Walker S, van der Donk WA (2011) Haloduracin α binds the peptidoglycan precursor Lipid II with 2:1 stoichiometry. J Am Chem Soc 133 (44):17544–17547

Otto M (2008) Staphylococcal biofilms. Curr Top Microbiol Immunol 322:207–228

Over B, Heusser R, McCallum N, Schulthess B, Kupferschmied P, Gaiani JM, Sifri CD, Berger-Bächi B, Meier PS (2011) LytR-CpsA-Psr proteins in *Staphylococcus aureus* display partial functional redundancy and the deletion of all three severely impairs spetum placement and cell separation. FEMS Microbiol Lett 320(2):142–151

Paknikar SS, Narayana S (2012) Newer antibacterials in therapy and clinical trials. N Am J Med Sci 4(11):537–547

Palmer KL, Kos VN, Gilmore MS (2010) Horizontal gene transfer and the genomics of enterococcal antibiotic resistance. Curr Opin Microbiol 13(5):632–639

Parsons JB, Rock CO (2014) Bacterial lipids: metabolism and membrane homeostasis. Prog Lipid Res 52(3):249–276

Patin D, Boniface A, Kovač A, Hervé M, Dementin S, Barreteau H, Mengin-Lecreulx D, Blanot D (2010) Purification and biochemical characterization of Mur ligases from *Staphylococcus aureus*. Biochimie 92(12):1793–1800

Patti GJ, Chen J, Schaefer J, Gross ML (2008) Characterization of structural variation in the peptidoglycan of vancomyin susceptible *Enterococcus faecium*: undertanding glycopeptide-antibiotic binding sites using mass spectrometry. J Am Soc Mass Spectrom 19 (10):1467–1475

Patton GC, van der Donk WA (2005) New developments in lantibiotic biosynthesis and mode of action. Curr Opin Microbiol 8:543–551

Pelton SI, Loughlin AM, Marchant CD (2004) Seven valent pneumococcal conjugate vaccine immunization in two Boston communities: changes in serotypes and antimicrobial susceptibility among *Streptococcus pneumoniae* isolates. Pediatr Infect Dis J 23(11):1015–1022

Percy MG, Gründling A (2014) Lipoteichoic acid synthesis and function in Gram-positive bacteria. Annu Rev Microbiol 68:81–100

Perego M, Glaser P, Minutello A, Srauch MA, Leopold K, Fischer W (1995) Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in *Bacillus subtilis*: identification of genes and regulation. J Biol Chem 270(26):15598–15606

Pereira MP, D'Elia MA, Troczynska J, Brown ED (2008) Duplication of teichoic acid biosynthetic genes in *Staphylococcus aureus* leads to functionally redundant poly(ribitol phosphate) polymerases. J Bacteriol 190(16):5642–5649

Périchon B, Courvalin P (2009) VanA-type vancomycin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother 53(11):4580–4587

Perkins HR (1969) Specificity of combination between mucopeptide precursors and vancomycin or ristocetin. Biochem J 111(2):195–205

Perkins HR, Nieto M (1974) The chemical basis for the action of the vancomycin group of antibiotics. Ann NY Acad Sci 235:348–363

Perry AM, Ton-That H, Mazmanian SK, Schneewind O (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. J Biol Chem 277(18):16241–16248

Peschel A, Otto M, Jack RW, Kalbacher H, Jung G, Götz F (1999) Inactivation of the *dlt* operon in *Staphylococcus aureus* confers sensitivity to defensins, protegrins, and other antimicrobial peptides. J Biol Chem 274(13):8405–8410

Peschel A, Vuong C, Otto M, Götz F (2000) The D-alanine residues of *Staphylococcus aureus* teichoic acids alter the susceptibility to vancomycin and the activity of autolytic enzymes. Antimicrob Agents Chemother 44(10):2845–2847

Peschel A, Jack RW, Otto M, Collins LV, Staubitz P, Nicholson G, Kalbacher H, Nieuwenhuizen WF, Jung G, Tarkowski A, van Kessel KP, van Strijp JA (2001) *Staphylococcus aureus* resistance to human defensins and evasion of neutrophil killing via the novel virulence factor *mprF* is based on modification of membrane lipids with L-lysine. J Exp Med 193(1):1067–1076

Peterson PK, Wilkinson BJ, Kim Y, Schmeling D, Quie PG (1978) Influence of encapsulation on staphylococcal opsonization and phagocytosis by human polymorphonuclear leukocytes. Infect Immun 19(3):943–949

Pilishvili T, Bennett NM (2015) Pneumococcal disease prevention among adults: strategies for the use of pneumococcal vaccines. Vaccine. doi:10.1016/j.vaccine.2015.05.102

Pinho MG, de Lancastre H, Tomasz A (2000) Cloning, characterization, and inactivation of the gene *pbpC*, encoding pencillin-binding protein 3 of *Staphylococcus aureus*. J Bacteriol 182 (4):1074–1079

Pinho MG, de Lancastre H, Tomasz A (2001a) An aquired and a native pencillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. Proc Natl Acad Sci USA 98 (19):10886–10891

Pinho MG, Filipe SR, de Lancastre H, Tomasz A (2001b) Complementation of the essential peptidoglycan transpeptidase function of penicillin-binding protein 2 (PBP2) by the drug resistance PBP2A in *Staphylococcus aureus*. J Bacteriol 183(22):6525–6531

Pitout JD, Nordmann P, Poirel L (2015) Carbapenemase-producing *Klebsiella pneumoniae*: a key pathogen set for global nosocomial dominance. Antimicrob Agents Chemother. doi:10.1128/AAC.01019-15

Pless D, Neuhaus FC (1973) Initial membrane reaction in peptidoglycan synthesis. Lipid dependence of phospho-*N*-acetylmuramylpentapeptide translocase (exchange reaction). J Biol Chem 248(5):1568–1576

Pobre K, Tashani M, Ridda I, Rashid H, Wong M, Booy R (2014) Carrier priming or suppression: understanding carrier priming enhancement of anti-polysaccharide antibody response to conjugate vaccines. Vaccine 32(13):1423–1430

Pollack JH, Neuhaus FC (1994) Changes in wall teichoic acid during the rod-sphere transition of *Bacillus subtilis* 168. J Bacteriol 176(23):7252–7259

Powell DA, Duckworth M, Baddiley J (1975) A membrane-associated lipomannan in micrococci. Biochemistry J 151(2):387–397

Qamar A, Golemi-Kotra D (2012) Dual roles of FmtA in *Staphylococcus aureus* cell wall biosynthesis and autolysis. Antimicrob Agents Chemother 56(7):3797–3805

Qiao Y, Lebar MD, Schirner K, Schaefer K, Tsukamoto H, Kahne D, Walker S (2014) Detection of lipid-linked peptidoglycan precursors by exploiting an unexpected transpeptidase reaction. J Am Chem Soc 136(42):14678–14681

Ray A, Cot M, Puzo G, Gilleron M, Nigou J (2013) Bacterial cell wall macroamphiphiles: pathogen-/microbe-associated molecular patterns detected by mammalian innate immune system. Biochimie 95(1):33–42

Reading C, Cole M (1977) Clavulanic acid: a β-lactamase-inhibiting β-lactam from *Streptomyces clavuligerus*. Antimicrob Agents Chemother 11(5):852–857

Reed P, Veiga H, Jorge AM, Terrak M, Pinho MG (2011) Monofunctional transglycosylases are not essential for *Staphylococcus aureus* cell wall synthesis. J Bacteriol 193(10):2549–2556

Reed P, Atilano ML, Alves R, Hoiczyk E, Sher X, Reichmann NT, Pereira PM, Roemer T, Filipe SR, Periera-Leal JB, Ligoxygakis P, Pinho MG (2015) *Staphylococcus aureus* survives with a minimal peptidoglycan synthesis machine but sacrifices virulence and antibiotic resistance. PLoS Pathog 11(5):e1004891. doi:10.1371/journal.ppat.1004891

Reichmann NT, Cassona CP, Gründling A (2013) Revised mechanism of D-alanine incorporation into cell wall polymers in Gram-positive bacteria. Microbiology 159(Pt 9):1868–1877

Revilla-Guarinos A, Gebhard S, Mascher T, Zuniga M (2014) Defence against antimicrobial peptides: different strategies in Firmicutes. Environ Microbiol 16(5):1225–1237

Reynolds PE (1989) Structure, biochemistry and mechanism of action of glycopeptide antibiotics. Eur J Clin Microbiol Infect Dis 8(11):943–950

Richter SG, Eli D, Kim HK, Hendrickx APA, Sorg JA, Schneewind O, Missiakas D (2013) Small molecule inhibitor of lipoteichoic acid synthesis is an antibiotic for Gram-positive bacteria. Proc Natl Acad Sci USA 110(9):3531–3536

Robbins JB, Schneerson R, Horwith G, Naso R, Fattom A (2004) *Staphylococcus aureus* type 5 and 8 capsular polysaccharide-protein conjugate vaccines. Am Heart J 147(4):593–598

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

Rogers HJ, Perkins HR, Ward JB (1980) Biosynthesis of peptidoglycan. Microbial cell walls and membranes. Chapman and Hall, London, pp 239–290

Rohde H, Frankenberger S, Zähringer U, Mack D (2010) Structure, function and contribution of polysaccharide intercellular adhesin (PIA) to *Staphylococcus epidermidis* biofilm formation and pathogenesis of biomaterial-associated infections. Eur J Cell Biol 89(1):103–111

Rohrer S, Ehlert K, Tschierske M, Labischinski H, Berger-Bächi B (1999) The essential *Staphylococcus aureus* gene *fmhB* is involved in the first step of peptidoglycan pentaglycine interpeptide formation. Proc Natl Acad Sci USA 96(16):9351–9356

Rosenow C, Ryan P, Weiser JN, Johnson S, Fontan P, Ortqvist A, Masure HR (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of *Streptococcus pneumoniae*. Mol Microbiol 25(5):819–829

Ruiz N (2008) Bioinformatics identification of MurJ (MviN) as the peptidoglyan lipidII flippase in *Escherichia coli*. Proc Natl Acad Sci USA 105(40):15553–15557

Ruiz N (2009) *Streptococcus pyogenes* YtgP (Spy_0390) complements *Escherichia coli* strains depleted of the putative peptidoglycan flippase MurJ. Antimicrob Agents Chemother 53 (8):3604–3605

Ruzin A, Severin A, Ritacco F, Tabei K, Singh G, Bradford PA, Siegel MM, Projan SJ, Shales DM (2002) Further evidence that a cell wall precursor [C(55)-MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. J Bacteriol 184(8):2141–2147

Ryu YH, Baik JE, Yang JS, Kang SS, Im J, Yun CH, Kim DW, Lee K, Chung DK, Ju HR, Han SH (2009) Differential immunostimulatory effects of Gram-positive bacteria due to their lipoteichoic acids. Int Immunopharmacol 9(1):127–133

Saar-Dover R, Bitler A, Nezer R, Shmuel-Galia L, Firon A, Shimoni E, Trieu-Cuot P, Shai Y (2012) D-alanylation of lipoteichoic acids confer resistance to cationic peptides in group B *Streptococcus* by increasing the cell wall density. PLoS Pathog 8(9):e1002891

Saas V, Schneider T, Wilmes M, Körner C, Tossi A, Novikova N, Shamova O, Sahl HG (2010) Human β-defensin 3 inhibits cell wall biosynthesis in staphylococci. Infect Immun 78(6):2793–2800

Santa Maria JP, Sadaka A, Moussa SH, Brown S, Zhang YJ, Rubin EJ, Gilmore MS, Walker S (2014) Compound-gene interaction mapping reveals distinct roles for Staphylococcus aureus teichoic acids. Proc Natl Acad Sci USA 111(34):12510–12515

Sauvage E, Terrak M, Ayala JA, Charlier P (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev 32(2):234–258

Scheffers DJ, Pinho MG (2005) Bacterial cell wall synthesis: new insights from localization studies. Microbiol Mol Biol Rev 69(4):585–607

Schirner K, Marles-Wright J, Lewis RJ, Errington J (2009) Distinc and essential morphogenic functions for wall- and lipo-teichoic acids in *Bacillus subtilis*. EMBO J 28(7):830–842

Schirner K, Eun YJ, Dion M, Luo Y, Helmann JD, Garner EC, Walker S (2015) Lipid-linked cell wall precursors regulate membrane association of bacterial actin MreB. Nat Chem Biol 11 (1):38–45

Schlag M, Biswas R, Krismer B, Kohler T, Zoll S, Yu W, Schwarz H, Peschel A, Götz F (2010) Role of staphylococcal wall teichoic acid in targeting the major autolysin Atl. Mol Microbiol 75(4):864–873

Schleifer KH, Kandler O (1972) Peptidoglycan. Types of bacterial cell walls and their taxonomic implications. Bacterial Rev 36(4):407–477

Schmidt JW, Greenough A, Burns M, Luteran AE, McCafferty DG (2010) Generation of ramoplanin-resistant *Staphylococcus aureus*. FEMS Microbiol Lett 310(2):104–111

Schneewind O, Missiakas DM (2012) Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc Lond B Biol Sci 367(1592):1123–1139

Schneewind O, Missiakas D (2014) Lipoteichoic acids, phosphate-containing polymers in the envelope of Gram-positive bacteria. J Bacteriol 196(6):1133–1142

Schneider T, Senn MM, Berger-Bächi B, Tossi A, Sahl HG, Wiedermann I (2004) *In vitro* assembly of a complete interpeptide bridge containing cell wall precursor (lipid-II-Gly5) of *Staphylococcus aureus*. Mol Microbiol 53(2):675–685

Schneider T, Kruse T, Wimmer R, Wiedemann I, Saas V, Pag U, Jansen A, Nielsen AK, Mygind PH, Raventós DS, Soren N, Ravn B, Bonvin AM, De Maria L, Andersen AS, Gammelgaard LK, Sahl HG, Kristensen HH (2010) Plectasin, a fungal defensin targets the bacterial cell wall precursor Lipid II. Science 328:1168–1172

Schröder NW, Morath S, Alexander C, Hamann L, Hartung T, Zähringer U, Göbel UB, Weber JR, Schumann RR (2003) Lipoteichoic acid (LTA) of *Streptococcus pneumoniae* and *Staphylococcus aureus* activates immune cells via Toll-like receptor (TLR)-2, lipopolysaccharide binding protein (LBP) and CD14, whereas TLR-4 and MD-2 are not involved. J Biol Chem 278(18):15587–15594

Severin A, Tomasz A (1996) Naturally Occurring Peptidoglycan Variants of *Streptococcus pneumoniae*. J Bacteriol 178(1):168–174

Sham LT, Butler EK, Lebar MD, Kahne D, Bernhardt TG, Ruiz N (2014) Bacterial cell wall. MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. Science 345(6193):220–222

Shapiro ED, Berg AT, Austrian R, Shroeder D, Parcells V, Margolis M, Adair RK, Clemens JD (1991) The protective efficacy of polyvalent pneumococcal polysaccharide vaccine. N Engl J Med 325(21):1453–1460

Shimada T, Park BG, Wolf AJ, Brikos C, Goodridge HS, Becker CA, Reyes CN, Miao EA, Aderem A, Götz F, Liu GY, Underhill DM (2010) *Staphylococcus aureus* evades lysozyme-based peptidoglycan digestion that links phagocytosis, inflammasome activation, and IL-1beta secretion. Cell Host Microbe 7(1):38–49

Sievert DM, Rudrik JT, Patel JB, McDonald LC, Wilkins MJ, Hageman JC (2008) Vancomycin-resistant *Staphylococcus aureus* in the United States, 2002–2006. Clin Infect Dis 46(5):668–674

Silhavy TJ, Kahne D, Walker S (2010) The bacterial cell envelope. Cold Spring Harb Perspect Biol 2:a000414

Silver L (2013) Viable screening targets related to the bacterial cell wall. Ann NY Acad Sci 1277:29–53

Skurnik D, Kropec A, Roux D, Theilacker C, Huebner J, Pier GB (2012) Natural antibodies in normal human serum inhibit *Staphylococcus aureus* capsular polysaccharide vaccine efficacy. Clin Infect Dis 55(9):1188–1197

Sobhanifar S, Worrall LJ, Gruninger RJ, Wasney GA, Blaukopf M, Baumann L, Lameignere E, Solomonson M, Brown ED, Withers SG, Strynadka NC (2015) Structure and mechanism of *Staphylococcus aureus* TarM, the wall teichoic acid α-glycosyltransferase. Proc Natl Acad Sci USA 112(6):E576–E585

Sorensen UB, Henrichsen H, Chen HC, Szu SC (1990) Covalent linkage between the capsular polysaccharide and the cell wall peptidoglycan of *Streptococcus pneumoniae* revealed by immunological methods. Microb Pathog 8(5):325–334

Spratt BG (1975) Distinct penicillin binding proteins involved in the division, elongation, and shape of *Escherichia coli* K12. Proc Natl Acad Sci USA 72(8):2999–3003

Steens A, Vestrheim DF, Aaberge IS, Wiklund BS, Storsaeter J, Riise Bergsaker MA, Ronning K, Furuseth E (2014) A review of the evidence to inform pneumococcal vaccine recommendations for risk groups aged 2 years and older. Epidemeol Infect 142(12):2471–2482

Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. Annu Rev Biochem 69:183–215

Sutherland IW (2001) Biofilm exopolysaccharides: a strong and sticky framework. Microbiology 147(Pt 1):3–9

Suzuki T, Campbell J, Swoboda JG, Walker S, Gilmore MS (2011a) Role of wall teichoic acids in *Staphylococcus aureus* endophthalmitis. Invest Ophthalmol Vis Sci 52(6):3187–3192

Suzuki T, Swoboda JG, Campbell J, Walker S, Gilmore MS (2011b) *In vitro* antimicrobial activity of wall teichoic acid biosynthesis inhibitors against *Staphylococcus aureus* isolates. Antimicrob Agents Chemother 55(2):767–774

Swoboda JG, Meredith TC, Campbell J, Brown S, Suzuki T, Bollenbach T, Malhowski AJ, Kishony R, Gilmore MS, Walker S (2009) Discovery of a small molecule that blocks wall teichoic acid biosynthesis in *Staphylococcus aureus*. ACS Chem Biol 4(10):875–883

Terrak M, Nguyen-Distèche M (2006) Kinetic characterization of the monofunctional glycosyltransferase from *Staphylococcus aureus*. J Bacteriol 188(7):2528–2532

Theilacker C, Kaczyński Z, Kropec A, Fabretti F, Sange T, Holst O, Huebner J (2006) Opsonic antibiodies to *Enterococcus faecalis* strain 12030 are directed against lipoteichoic acid. Infect Immun 74(10):5703–5712

Theilacker C, Holst O, Lindner B, Huebner J, Kacyński Z (2012) The structure of the wall teichoic acid isolated from *Enterococcus faecalis* strain 12030. Carbohydr Res 354:106–109

Thumm G, Götz F (1997) Studies on prolysostaphin processing and characterization of the lysostaphin immunity factor (Lif) of *Staphylococcus simulans* biovar staphylolyticus. Mol Microbiol 23(6):1251–1255

Thurlow LR, Thomas VC, Fleming SD, Hancock LE (2009) *Enterococcus faecalis* capsular polysaccharide serotypes C and D and their contributions to host innate immune evasion. Infect Immun 77(12):5551–5557

Tschierske M, Ehlert K, Strandén AM, Berger-Bächi B (1997) Lif, the lysostaphin immunity factor, complements FemB in staphylococcal peptidoglycan interpeptide bridge formation. FEMS Microbiol Lett 153(2):261–264

Tzianabos AO, Wang JY, Lee JC (2001) Structural rationale for the modulation of abscess formation by *Staphylococcus aureus* capsular polysaccharides. Proc Natl Acad Sci USA 98 (16):9365–9370

Valentino MD, Foulston L, Sadaka A, Kos VN, Villet RA, Santa Maria J Jr, Lazinski DW, Camili A, Walker S, Hooper DC, Gilmore M (2014) Genes contributing to *Staphylococcus aureus* fitness in abscess- and infection-related ecologies. MBio 5(5):e01729–14

Van Selm S, van Cann LM, Kolkman MA, van der Zeijst BA, van Putten JP (2003) Genetic basis for the structural difference between *Streptococcus pneumoniae* serotype 15B and 15C capsular polysaccharides. Infect Immun 71(11):6192–6198

Venkateswaran PS, Stanton N, Austrian R (1983) Type variation of strains of *Streptococcus pneumoniae* in capsular serogroup 15. J Infect Dis 147(6):1041–1054

Vlamakis H, Chai Y, Beauregard P, Losick R, Kolter R (2013) Sticking together: building a biofilm the *Bacillus subtilis* way. Nat Rev Microbiol 11(3):157–168

Volkman BF, Zhang Q, Debabov DV, Rivera E, Kresheck GC, Neuhaus FC (2001) Biosynthesis of D-alanyl-lipoteichoic acid: the tertiary structure of apo-D-alanyl carrier protein. Biochemistry 40(27):7964–7972

Vollmer W, Tomasz A (2000) The *pgdA* gene encodes for a peptidoglycan *N*-acetylglucosamine deacetylase in *Streptococcus pneunomiae*. J Biol Chem 275(27):20496–20501

Vollmer W, Tomasz A (2002) Peptidoglycan *N*-acetylglucosamine deacetylase, a putative virulence factor in *Streptoccus pneumoniae*. Infect Immun 70(12):7176–7178

Vollmer W, Blanot D, de Pedro MA (2008) Peptidoglycan structure and architecture. FEMS Microbiol Rev 32(2):149–167

Von Aulock S, Hartung T, Hermann C (2007) Comment on "not lipoteichoic acid but lipoproteins appear to be the dominant immunobiologically active compound in *Staphylococcus aureus*". J Immunol 178(5):2610–2611

Vuong C, Kocianova S, Voyich JM, Yao Y, Fischer ER, DeLeo FR, Otto M (2004) A critical role for exopolysaccharide modification in bacterial biofilm formation, immune evasion and virulence. J Biol Chem 279(52):54881–54886

Wada A, Watanabe H (1998) Pencillin-binding protein 1 of *Staphylococcus aureus* is essential for growth. J Bacteriol 180(10):2759–2765

Walsh CT (1989) Enzymes in the D-alanine branch of bacterial cell wall peptidogylcan assembly. J Biol Chem 264(5):2393–2396

Walsh C (2003) Antibiotics that act on cell wall biosynthesis. Antibiotics: actions, origins, resistance. ASM Press, Washington D.C., pp 24–49

Walsh CT, Howe RA (2002) The prevalence and mechanisms of vancomycn resistance in *Staphylococcus aureus*. Ann Rev Microbiol 56:657–675

Wang H, Gill CJ, Lee SH, Mann P, Zuck P, Meredith TC, Murgolo N, She X, Kales S, Liang L, Liu J, Wu J, Santa Maria JS, Su J, Pan J, Hailey J, Mcguinness D, Tan CM, Flattery A, Walker S, Black T, Roemer T (2013) Discovery of novel wall teichoic acid inhibitors as effective anti-MRSA β-lactam combination agents. Chem Biol 20(2):272–284

Ward JB (1973) The chain length of the glycans in bacterial cell walls. J Biochem 133(2):395–398

Waxman D, Strominger JL (1983) Penicillin-binding proteins and the mechanism of action of β-lactam antibiotics. Annu Rev Microbiol 52:825–829

Weart RB, Lee AH, Chien AC, Haeusser DP, Hills NS, Levin PA (2007) A metabolic sensor governing cell size in bacteria. Cell 130(2):335–347

Webb AJ, Karatsa-Dodgson M, Gründling A (2009) Two-enzyme systems for glycolipid and polyglycerophosphate lipoteichoic acid synthesis in *Listeria monocytogenes*. Mol Microbiol 74 (2):299–314

Weber B, Ehlert K, Diehl A, Reichmann P, Labischinski H, Hakenbeck R (2000) The fib locus in *Streptococcus pneumoniae* is required for peptidoglycan crosslinking and PBP-mediated β-lactam resistance. FEMS Microbiol Lett 188(1):81–85

Weidenmaier C, Kokai-Kun JF, Kristian SA, Chanturiya T, Kalbacher H, Gross M, Nicholson G, Neumeister B, Mond JJ, Peschel A (2004) Role of teichoic acids in *Staphylococcus aureus* nasal colonization, a major risk factor in nosocomial infections. Nat Med 10(3):243–245

Weidenmaier C, Peschel A, Xiong YQ, Kristian SA, Dietz K, Yeaman MR, Bayer AS (2005) Lack of wall teichoic acids in *Staphylococcus aureus* leads to reduced interactions with endothelial cells and to attenuated virulence in rabbit model endocarditis. J Infect Dis 191(10):1771–1777

Weigel LM, Clewell DB, Gill SR, Clark NC, McDougal LK, Flannagan SE, Kolonay JF, Shetty J, Killgore GE, Tenover FC (2003) Genetic analysis of a high-level vancomycin-resistant isolate of *Staphylococcus aureus*. Science 302(5650):1569–1571

White AR, Kaye C, Poupard J, Pypstra R, Woodnutt G, Wynne B (2004) Augmentin® (amoxicillin/clavulanate) in the treatment of community-acquired respiratory tract infection: a review of the continuing development of an innovative antimicrobial agent. J Antimicrob Chemother 53(Suppl 1):i3–i20

Whitener CJ, Park SY, Browne FA, Parent LJ, Julian K, Bozdogan B et al (2004) Vancomycin-resistant *Staphylococcus aureus* in the absense of vancomycin exposure. Clin Infect Dis 38(8):1049–1055

Wilkinson BJ, Holmes KM (1979) *Staphylococcus aureus* cell surface: capsule as a barrier to bacteriophage adsorption. Infect Immun 23(2):549–552

Winstel V, Kühner P, Salomon F, Larsen J, Skov R, Hoffmann W, Peschel A, Weidenmaier C (2015) Wall teichoic acid glycosylation governs *Staphylococcus aureus* nasal colonization. MBio 6(4):e00632–e00615

Wörmann ME, Corrigan RM, Simpson PJ, Matthews SJ, Gründling A (2011) Enzymatic activities and functional interdependencies of *Bacillus subtilis* lipoteichoic acid synthesis enzymes. Mol Microbiol 79(3):566–583

Wright G (2011) Molecular mechanism of antibiotic resistance. Chem Comm 47:4055–4061

Wyke AW, Ward JB, Hayes MV, Curtis NA (1981) A role *in vivo* for penicillin-binding protein-4 of *Staphylococcus aureus*. Eur J Biochem 119(2):389–393

Xayarath B, Yother J (2007) Mutations blocking side chain assembly, polymerization, or transport of a Wzy-dependent *Streptococcus pneumoniae* capsule are lethal in the absence of suppressor mutations and can affect polymer transfer to the cell wall. J Bacteriol 189(9):3369–3381

Xia G, Maier L, Sanchez-Carballo P, Li M, Otto M, Holst O, Peschel A (2010) Glycosylation of wall teichoic acid in *Staphylococcus aureus* by TarM. J Biol Chem 285(18):13405–13415

Xu H, Wang L, Huang J, Zhang Y, Ma F, Wang J, Xu W, Zhang X, Yin Y, Wu K (2015) Pneumococcal wall teichoic acid is required for pathogenesis of *Streptococcus pneumoniae* in murine models. J Microbiol 53(2):147–154

Yang SJ, Kreiswirth BN, Sakoulas G, Yeaman MR, Xiong YQ, Sawa A, Bayer AS (2009a) Enhanced expression of *dltABCD* is associated with the development of daptomycin nonsusceptibility in a clinical endocarditis isolate of *Staphylococcus aureus*. J Infect Dis 200(12):1916–1920

Yang SJ, Xiong YQ, Dunman PM, Schrenzel J, Francois P, Peschel A, Bayer AS (2009b) Regulation of *mprF* in daptomycin-nonsusceptible *Staphylococcus aureus* strains. Antimicrob Agents Chemother 53(6):2636–2637

Yang SJ, Bayer AS, Mishra NN, Meehl M, Ledala N, Yeaman MR, Xiong YQ, Cheung AL (2012) The *Staphylococcus aureus* two-component sensing system, GraRS, senses and confers resistance to selected cationic antimicrobial peptides. Infect Immun 80(1):74–81

Yocum RR, Waxman DJ, Rasmussen JR, Strominger JL (1979) Mechanism of penicillin action: penicillin and substrate bind covalently to the same active site serine in two bacterial D-alanine carboxypeptidases. Proc Natl Acad Sci USA 76(6):2730–2734

Yocum RR, Rasmussen JR, Strominger JL (1980) The mechanism of action of penicillin. Penicillin acylates the active site of *Bacillus stearothermophilus* D-alanine carboxypeptidase. J Biol Chem 255(9):3977–3986

Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT (2008) Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. J Biol Chem 283(47):32484–32491

Yother J (2011) Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. Annu Rev Microbiol 65:563–581

Young FE (1967) Requirement of glucosylated teichoic acid for adsorption of phage in *Bacillus subtilis* 168. Proc Natl Acad Sci USA 58(6):2377–2384

Zapun A, Vernet T, Pinho MG (2008) The different shapes of cocci. FEMS Microbiol Rev 32 (2):345–360

Zhang YM, Rock CO (2008) Membrane lipid homeostasis in bacteria. Nat Rev Microbiol 6 (3):222–233

Zhen M, Jacobsen FE, Giedroc DP (2009) Metal transporters and metal sensors: how coordination chemistry controls bacterial metal homeostasis. Chem Rev 109(10):4644–4681

Zhu W, Clark NC, McDougal LK, Hageman J, McDonald LC, Patel JB (2008) Vancomycin-resistant *Staphylococcus aureus* isolates associated with Inc18-like *vanA* plasmids in Michigan. Antimicrob Agents Chemother 52(2):452–457

Zipperle GF Jr, Ezzell JW Jr, Doyle RJ (1984) Glucosamine substitution and muramidase susceptibility in *Bacillus anthracis*. Can J Microbiol 30(5):553–559

The Canonical and Accessory Sec System of Gram-positive Bacteria

Irfan Prabudiansyah and Arnold J.M. Driessen

Abstract The Sec system is present in all bacteria and responsible for the translocation of the majority of proteins across the cytoplasmic membrane. The system consists of two principal components: the ATPase motor protein, SecA, and the protein-conducting channel, SecYEG. In addition to this canonical Sec system, several Gram-positive bacteria also possess a so-called accessory Sec system. This is a specialized translocation system that is responsible for the export of a subset of secretory proteins, including virulence factors. The accessory Sec system consists of a second SecA paralog, termed SecA2, with or without a second SecY paralog, termed SecY2. In some bacteria, the accessory Sec system is dependent on the canonical Sec system for functionality, while in other bacteria, they can function independently. In this review, we provide an overview of the current knowledge of the canonical and accessory Sec system of Gram-positive bacteria with a focus on the primary component of the Sec translocase, SecA and SecYEG.

Contents

I. Prabudiansyah · A.J.M. Driessen (✉)
Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, Nijenborgh 7, 9727 Groningen, AG, The Netherlands
e-mail: a.j.m.driessen@rug.nl

Current Topics in Microbiology and Immunology (2017) 404:45–67
DOI 10.1007/82_2016_9

Published Online: 07 May 2016

1 Introduction

Bacteria are favorite model organisms in scientific research to study cellular processes that also occur in multicellular organisms. Cells are equipped with mechanisms to transport proteins from the site of synthesis, i.e., the cytoplasm, to the cell envelope or to the extracellular environment. This process is not only important for the survival of the cell, but also plays important roles in pathogenesis, and for instance the interaction with eukaryotic hosts. The major route for protein transport in bacteria is provided by the general secretion pathway (Sec pathway). The Sec pathway and its functional components have been extensively studied in the Gram-negative bacterium, *Escherichia coli*, and also in Gram-positive bacterium, *Bacillus subtilis* (Bieker et al. 1990; Schatz and Beckwith 1990; Van Wely et al. 2001; De Keyzer et al. 2003; Driessen and Nouwen 2008). Protein transport in this pathway is mediated by the Sec translocase, which in its minimal form consists of a protein-conducting channel formed by the heterotrimeric membrane protein complex, SecYEG, and the essential ATPase SecA that acts as a molecular motor (Brundage et al. 1990).

In general, the Sec system of Gram-negative and Gram-positive bacteria is similar in composition, and main components are highly conserved. In the last decade, however, studies on the Gram-positive Sec system revealed some interesting differences. Besides the canonical Sec components, a large number of Gram-positive bacteria possess accessory Sec components that are not found in Gram-negative bacteria. This concerns a presence of a second SecA paralog, termed SecA2, either with or without a second SecY paralog, termed SecY2 (Braunstein et al. 2001; Bensing and Sullam 2002; Lenz and Portnoy 2002). In contrast to the canonical Sec translocase, which is essential for the translocation of the majority of secretory proteins, the accessories SecA2 and SecY2, in most cases, appear to be not essential (Braunstein et al. 2001; Bensing and Sullam 2002; Lenz and Portnoy 2002). They seem to be especially important for the export of a subset of proteins, which in some bacteria are mostly virulence factors (Lenz et al. 2003; Rigel and Braunstein 2008; Sullivan et al. 2012). In bacterial species that possess both SecA2 and SecY2, e.g., *Streptococcus*, the accessory components form a separate translocation system to export specific substrates independently of the canonical Sec system (Bensing et al. 2014). Interestingly, in species that possess only SecA2, e.g., *Mycobacteria*, SecA2 seems to work together in conjunction with the canonical SecYEG/SecA1 translocase in the export of multiple substrates (Rigel et al. 2009; Feltcher and Braunstein 2012; Freudl 2013; Bensing et al. 2014). Here, we discuss the current knowledge on the canonical and accessory Sec system in

Gram-positive bacteria. We will mainly focus on the primary component of the system, the motor protein SecA and the protein membrane channel SecY, and highlight the potential mechanistic implications.

1.1 Sec Pathway, the Major Route for Protein Secretion

The Sec pathway forms a conserved route for the secretion of housekeeping proteins and for protein secretion in general. Secretory proteins are initially synthesized at the ribosome as unfolded preproteins with a cleavable N-terminal signal peptide. The signal peptide is important to distinguish preproteins from cytoplasmic proteins, adds to the unfolding of the associated mature domain, and targets these proteins to the translocation sites at the membrane (Von Heijne 1990; Dalbey et al. 1997). Targeting is mostly a post-translational event, but some secretory proteins are targeted co-translationally (Josefsson and Randall 1981).

During post-translational targeting, protein synthesis is first completed at the ribosome before the preprotein engages with the Sec system at the membrane. The molecular chaperone stabilizes the preproteins in a translocation-competent state and directs them to the translocation site. At the membrane, the molecular chaperone transfers the preprotein to the SecA motor domain of the Sec translocase. In the next steps, multiple cycles of ATP binding and hydrolysis by SecA result in the stepwise translocation of the unfolded preprotein through the SecYEG channel (Schiebel et al. 1991; Economou and Wickner 1994; Van der Wolk et al. 1997). Another heterotrimeric membrane protein complex, SecDFyajC, stimulates translocation and utilizes the proton motive force (PMF) to facilitate this process (Duong and Wickner 1997; Tsukazaki et al. 2011). Generally, the post-translocation mechanism is similar in Gram-negative and Gram-positive bacteria. One main difference is in the nature of the molecular chaperones. In Gram-negative bacteria, the chaperone function is carried out by SecB (Lecker et al. 1989) or other general chaperones such as trigger factor, DnaK, or GroEL (Castanié-Cornet et al. 2014). In Gram-positive bacteria that lack a SecB homolog, the chaperone CsaA has been implicated in protein translocation (Bron et al. 2000; Müller et al. 2000). It should be noted that the more general chaperones do not entail the specific targeting function of SecB.

During co-translational targeting, preproteins remain bound to the ribosome as a nascent chain and are targeted to the SecYEG channel by signal recognition particle (SRP) and the signal recognition particle receptor FtsY. SRP and FtsY are both GTPases, and release of the nascent chain from SRP to the SecYEG channel is facilitated by heterodimerization of SRP and FtsY and subsequent GTP hydrolysis. Mostly, nascent membrane proteins utilize the aforementioned targeting route for co-translational membrane insertion, albeit there is a subset of secretory proteins that use this pathway as well. However, preprotein translocation is strictly dependent on the ATPase SecA and ATP hydrolysis. After translocation, the signal peptide is cleaved from the preprotein by signal peptidase, and the protein will fold

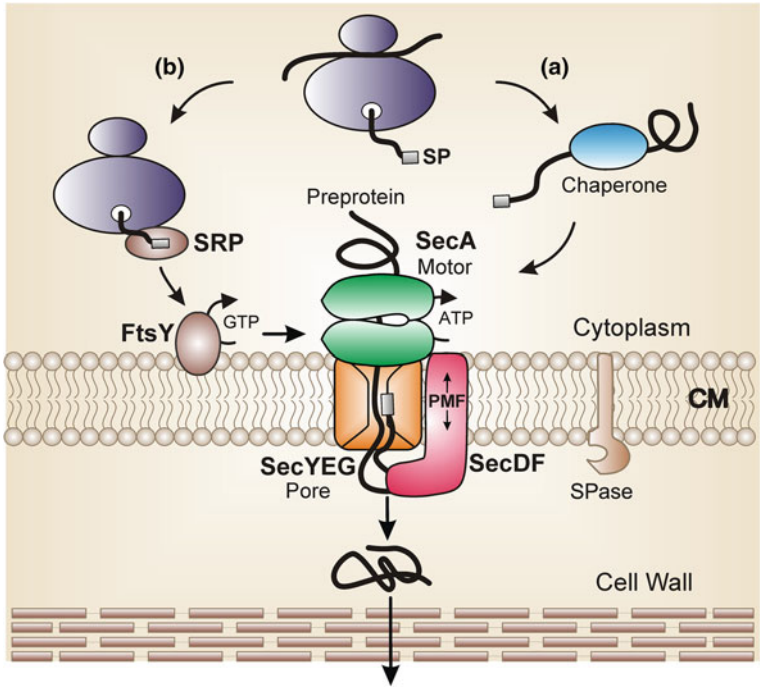


Fig. 1 The Sec pathway of Gram-positive bacteria. **a** Post-translational targeting: preproteins synthesized at the ribosome (*purple*) are captured in an unfolded state by a chaperone protein (*blue*) and targeted to SecA (*green*) that is bound to the SecYEG channel (*orange*) at the cytoplasmic membrane (CM). SecA pushes the preprotein through the SecYEG channel in an ATP-dependent manner. In addition, SecDFyajC (*red*) uses the PMF to pull preproteins into the periplasm. After translocation, the signal peptide (SP) is removed by signal peptidase (SPase) and the mature protein is released. **b** In co-translational targeting, preproteins are targeted to the translocation site as a ribosome–nascent chain complex (RNC) by SRP and FtsY (*brown*) where protein synthesis will commence involving SecA and ATP

on the trans-side of the membrane into its functional conformation or continue its passage guided by chaperones to the outer membrane. In Gram-positive bacteria, some proteins interact with the cell wall, while others pass through the cell wall and are released into the external environment. A schematic representation of the general Sec pathway in Gram-positive bacteria is shown in Fig. 1.

1.2 SecA, the Translocation Motor

SecA functions as an ATP-driven molecular motor to facilitate protein translocation across the SecYEG protein-conducting pore. SecA is a highly conserved bacterial protein, but also presents in chloroplasts of plant cells, where it is needed for protein

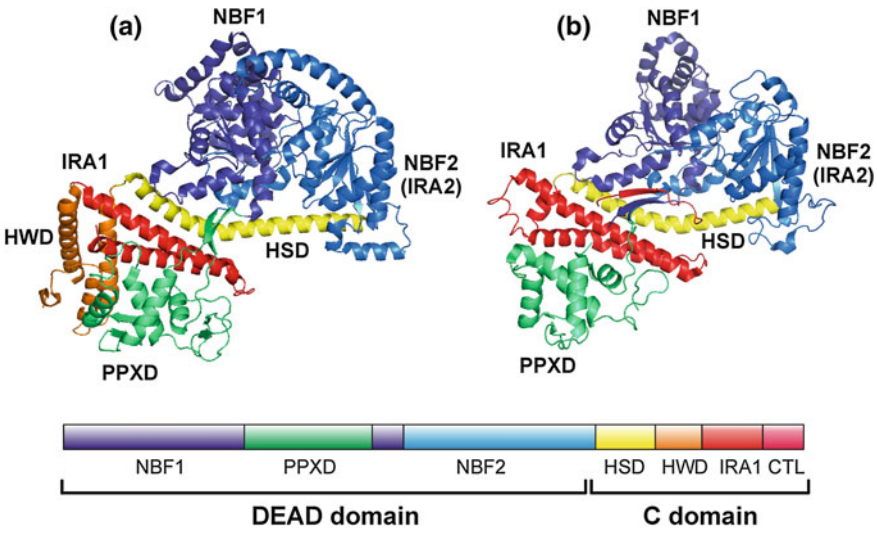


Fig. 2 Structure of SecA, the translocation motor. **a** SecA1 protomer from *Mycobacterium tuberculosis* (PDB accession code: 1NKT) and **b** SecA2 from *M. tuberculosis* (PDB accession code: 4UAQ), showing the different subdomains, which are indicated with different color. NBF1 and NBF2, nucleotide-binding folds 1 and 2; PPXD, preprotein cross-linking domain; HSD, α-helical scaffold domain; HWD, α-helical wing domain; IRA1, the intramolecular region of ATP hydrolysis 1; and CTL, C-terminal linker domain

translocation into the thylakoid. In the cell, SecA exists in a soluble cytosolic form (Oliver and Beckwith 1982) and membrane-bound forms, i.e., associated with SecYEG (Lecker et al. 1990) and with phospholipids (Lill et al. 1990; Cabellis et al. 1991; Hendricks and Wickner 1991). SecA is a homodimeric protein. X-ray crystallographical studies on SecA proteins from different bacteria revealed a dimer organization either with antiparallel (Hunt et al. 2002; Ding et al. 2003; Sharma et al. 2003; Zimmer et al. 2006; Papanikolau et al. 2007) or with parallel (Vassylyev et al. 2006) protomers. The SecA protomer can be divided into several structural domains (Fig. 2). The DEAD motor domain is part of the central core of SecA and consists of two subdomains: the nucleotide-binding fold 1 (NBF1) and NBF2, also called the intramolecular regulator of ATPase activity 2 domain (IRA2). Both NBFs are homologous to the RecA-like nucleotide-binding folds found in DNA/RNA helicases (Tanner and Linder 2001). The interface of the NBF1 and NBF2, which comprise the Walker A and B motifs, is the site for ATP binding and hydrolysis (Osborne et al. 2004; Robson et al. 2007; Zimmer and Rapoport 2009). In addition to the NBFs, SecA contains two substrate specificity domains: the preprotein-binding domain (PBD) or preprotein cross-linking domain (PPXD) and the C domain. The PPXD has been shown to be involved in binding of preproteins (Papanikou et al. 2005; Musial-Siwek et al. 2007; Gelis et al. 2007). The C domain is located at the C terminus of NBF2 and consists of four subdomains: the helical scaffold domain (HSD) that control the opening and closing of the DEAD motor

(Hunt et al. 2002; Mori and Ito 2006); the helical wing domain (HWD); the intramolecular regulator of ATP hydrolysis (IRA1), which acts as an inhibitor of ATP hydrolysis (Karamanou et al. 1999); and the C-terminal linker domain (CTL), which contains a zinc finger which is important for the interaction with the secretion specific chaperone SecB (Fekkes et al. 1997, 1999; Zhou and Xu 2003) and phospholipids (Breukink et al. 1995).

The oligomeric state of SecA in solution and during protein translocation has been extensively studied with various methods. SecA exists in solution in a dimer–monomer equilibrium with a K_d in nanomolar range (Kusters et al. 2011; Wowor et al. 2011). This suggests that most of SecA must be dimeric in the cell, since the cellular concentration of SecA has been suggested to be in the micromolar range (5–8 μM) (Akita et al. 1991; Or et al. 2002). The dimer–monomer equilibrium of SecA is shifted toward the monomer at high ionic strength and low temperature (Woodbury et al. 2002; Kusters et al. 2011; Wowor et al. 2011), suggesting that electrostatic and hydrophobic interactions play a critical role in maintaining the dimer. SecA dimerization is also influenced by ligands, such as phospholipids, signal peptides (Benach et al. 2003; Or et al. 2005; Musial-Siwek et al. 2005), and nucleotides (Bu et al. 2003). The precise functional state of SecA during protein translocation process is still a matter of debate. A multitude of studies demonstrate that SecA functions as a dimer during protein translocation (Driessen 1993; Karamanou et al. 2005; Jilaveanu et al. 2005; De Keyzer et al. 2005; Jilaveanu and Oliver 2006; Das et al. 2008; Wang et al. 2008; Kusters et al. 2011; Gouridis et al. 2013), although other report suggests that monomeric SecA obtained by extensive mutagenesis retains some activity with a hyperactive SecY mutant (Or et al. 2005). Single molecule studies demonstrate that the dimeric SecA binds the SecYEG translocation pore with high affinity, wherein one of the protomers binds SecYEG tightly, whereas the other protomer is bound to the SecYEG-bound SecA (Kusters et al. 2011).

1.3 SecYEG, the Protein-Conducting Channel

The Sec translocon shows a heterotrimeric organization that is highly conserved in three kingdoms of life (Pohlschröder et al. 1997). In bacteria, it consists of three integral membrane proteins SecY, SecE, and SecG (Brundage et al. 1990), which together forms the protein-conducting channel SecYEG complex, that is homologs to Sec61αγβ in eukaryotes and SecYEβ in archaea. The X-ray crystallography structure of *Methanoccoccus jannaschii* SecYEβ provided the first high-resolution insight into the structural organization of the translocation channel (Van Den Berg et al. 2004) (Fig. 3).

The SecY protein forms the actual channel and consists of ten α-helical transmembrane segments (TMSs) that are organized as to sets, i.e., TMSs 1–5 and TMSs 6–10. The N- and C-termini localize to the cytoplasm. The two TM domains are connected by a periplasmic loop between TMS 5 and TMS 6 forming a clamshell

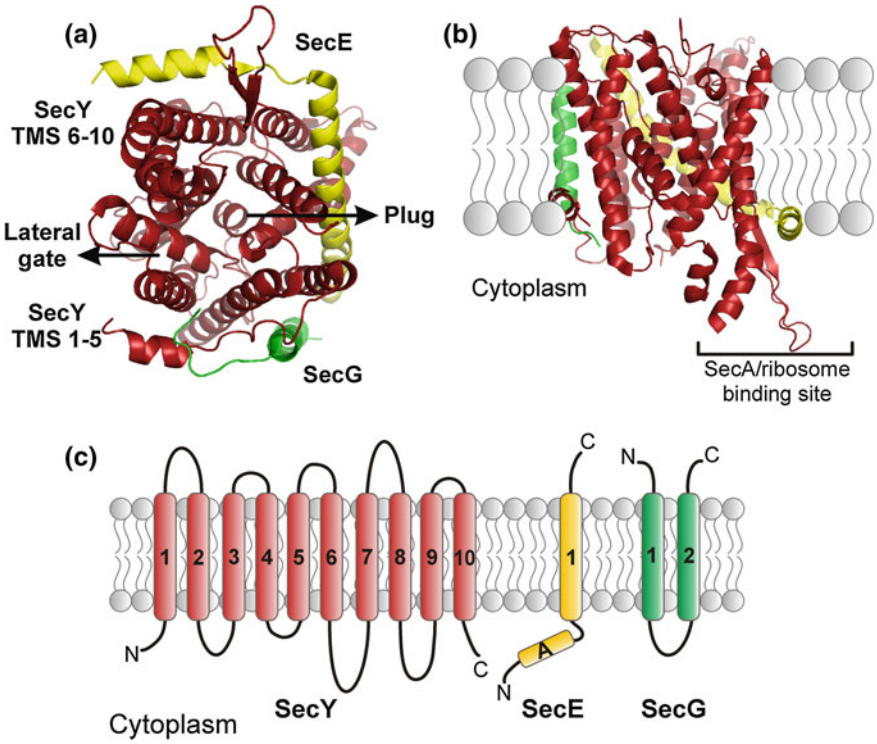


Fig. 3 Structure of SecYEG, the protein-conducting channel. **a** Structure of SecYEβ from *Methanoccoccus jannaschii* (PDB accession code: 1RH5) viewed from the cytoplasm. SecY is colored in *red*, SecE in *yellow*, and SecG in *green*. The lateral gate and the plug are indicated with the *arrow line*. **b** SecYEβ viewed from the side, in position in the lipid bilayer. **c** Secondary structure prediction of SecYEG from Gram-positive bacteria (*B. subtilis*). The TMSs are numbered, and the essential amphipathic helix of SecE is labeled with A

structure. The *E. coli* SecE consists of three TMSs (TMSs 1-3), where the first two TMSs are connected to the third tilted TMS via the amphipathic helix. Only the third TMS and the amphipathic helix are essential for the functionality (Schatz et al. 1991; Nishiyama and Mizushima 1992; Murphy and Beckwith 1994). TMS3 and the amphiphatic helix embrace the SecY clamshell, in which the TMS3 associated with one half of the clamshell and the amphipathic helix associated with the other half. These two SecE domains are the major sites of SecY-SecE interactions and are important for the stability and flexibility of SecY (Lycklama et al. 2012, 2013). SecE of Gram-positive bacteria consists only one TMS and the amphipathic helix, which is homologous to the corresponding functional part of the *E. coli* SecE (Jeong et al. 1993; Cao and Saier 2003). The Secβ protein, which presumably is homologous to the bacterial SecG, is located peripherally in the structure and shows limited contact with SecY. SecG is not essential for the functionality in bacteria, but it increases the efficiency of translocation (Nishiyama and Mizushimal 1993;

Hanada et al. 1994; Nishiyama and Hanada 1994) and has also been suggested to associate with SecA (Suzuki et al. 1998; Mori et al. 1998).

The overall structure of SecYEG channel shows an hourglass shape (Van Den Berg et al. 2004). The central channel is constricted with six isoleucine residues that form a hydrophobic pore ring that acts as a seal to provide a barrier for water and ions (Saparov et al. 2007; Gumbart and Schulten 2008). Below the pore ring is a small helix called the plug domain. The plug domain together with the pore ring prevents ion flow through the channel in the closed conformation (Gumbart and Schulten 2008). The structure also contains a lateral gate between TMS 2 and TMS 7 of SecY, which plays an essential role in the lateral insertion of membrane proteins. The signal sequence of preproteins and nascent TMSs was suggested to bind this lateral gate (Plath et al. 1998, 2004), and this initial insertion into the lateral gate is believed to open the channel. The cytoplasmic side of the SecYEG channel contains several binding sites for the cytoplasmic binding partner, i.e., SecA and ribosome (Ménétret et al. 2007; Zimmer et al. 2008; Frauenfeld et al. 2011), and is important for protein translocation. SecYEG also interacts with other membrane proteins, such as SecDFyajC (Duong and Wickner 1997), and the membrane protein insertase, YidC (Scotti et al. 2000).

2 The Accessory Sec Translocase

2.1 SecA2, the Specialized SecA Protein

The first SecA2 protein was identified fifteen years ago in Mycobacteria (Braunstein et al. 2001; Rigel and Braunstein 2008). Nowadays, it clear that SecA2 is present in a large number of Gram-positive bacteria, but absent in Gram-negative bacteria. Interestingly, these accessory SecA2s are not closely related to each other phylogenetically (Bensing et al. 2014). The sequence similarity between SecA2 and their corresponding SecA homolog (SecA1) varies between bacterial species. In almost all cases, SecA2 proteins are smaller in size, as compared with SecA1s. The first crystal structure of SecA2 was published recently from *Mycobacterium tuberculosis* (Swanson et al. 2016) (Fig. 2). Overall, the structure shows a high similarity to the *M. tuberculosis* SecA1 and other orthologs of the SecA family. Most functional domains are present in the SecA2 structure, including the two NBFs. Despite the similarities, there are some structural differences in SecA2 compared to SecA1. SecA2 is a smaller protein because of several deletions, mostly in the C domain. The main structural difference is the absence of HWD domain in SecA2. The functional significant of the absence of the HWD in SecA2 is unclear, but possibly this hints at a reduced interaction with SecYEG. In SecA, the HWD is important for the interaction with SecYEG (Zimmer et al. 2008; Das and Oliver 2011). In the structure of SecA2, also the orientation of the PPXD domain and the two-helix finger are different from that of SecA1. Additionally, the conserved

tyrosine in the two-helix finger of SecA1, which provides the major contact with the SecA1 substrate (Erlandson et al. 2008), is missing in SecA2. Overall, these structural differences may contribute to the special and distinct role of SecA2 in the protein export.

In general, the accessory SecA2 has a more specific role than the canonical SecA, as discussed in detail elsewhere (Feltcher and Braunstein 2012; Freudl 2013; Bensing et al. 2014). In contrast to SecA1, which is essential and involved in transport of the majority of proteins, SecA2 in most cases is not essential, and has a specialized function for the export of a subset of proteins. Additionally, SecA2 also plays an important role in virulence in some bacteria (Lenz et al. 2003; Rigel and Braunstein 2008; Sullivan et al. 2012). Based on its interacting partner and substrates, it has been proposed that there are two types of SecA2 proteins. The first type is suggested to associate with the canonical Sec system and transports multiple type of substrates, and also called the SecA2-only system (Rigel et al. 2009; Feltcher and Braunstein 2012; Bensing et al. 2014). The second type is suggested to interact with the accessory SecY2 channel, also called the SecA2/SecY2 system, and transport a single specific substrate completely independent from the canonical Sec system, (Bensing et al. 2014). Different types of SecA2-dependent proteins have been identified, and these appear to be involved in different functions and cellular locations. Some of these substrates are cell envelope proteins while others are secreted. Intriguingly, some substrates possess a signal sequence while others do not. However, the mechanisms by which SecA2 selects the substrates for transport is essentially unknown, nor is it clear why the substrates are not recognized by the canonical SecA. One hypothesis is that the absence of the HWD in SecA2 might result in a more solvent-exposed signal peptide binding cleft that could help SecA2 in the recognition of specific SecA2 substrates, including signal peptide-less variants (Swanson et al. 2016).

2.2 *SecY2, the Accessory Membrane Channel*

In addition to SecA2, some Gram-positive bacteria, e.g., *Streptococcus*, also possess the accessory SecY2, that is homologous to SecY (Bensing and Sullam 2002; Bensing et al. 2014). SecY2 is predicted to form an accessory membrane channel that is responsible for the export of specific proteins, which cannot be exported by the canonical SecYEG. In general, SecY2 proteins share a low sequence similarity to SecY (SecY1). However, the predicted membrane topology of SecY2 is identical to that of SecY (Bensing et al. 2014). The conserved residues in cytoplasmic loop 5 of SecY, that are important for SecA interaction (Van der Sluis et al. 2006) are absent in SecY2, possibly suggesting a lack of interaction with the canonical SecA1 protein. In *Streptococcus* and *Staphylococcus*, SecY2 works together with SecA2 to export large serine-rich repeat (SRR) glycoproteins to the cell surface (Bensing and Sullam 2002; Siboo et al. 2008). Therefore, it is suggested that SecY2 directly associates with SecA2 to form the functional Sec translocase. SecY2 is also

suggested to associate with some potential accessory secretion proteins (Asps), i.e., Asp1-Asp5 (Bensing et al. 2014). Asp4 and Asp5 have been suggested to form the membrane channel with SecY2, and thus function similar to SecE and SecG, respectively. However, this Asps are absent in other bacteria that do contain a SecY2, and thus the exact composition of this accessory Sec translocase has remained elusive.

3 The Accessory Sec System in Different Type of Gram-positive Bacteria

Gram-positive bacteria have a less complex cell envelope structure as compared to Gram-negative bacteria. Most of Gram-positive bacteria comprise only a single cytoplasmic membrane followed by a cell wall and defined as monoderm species. However, some Gram-positive bacteria possess an extra membrane, a peptidoglycan–mycolic acid wall structure, thus are diderm species, e.g., Mycobacteria. These are included as well in the discussion in this section. The accessory Sec system has been identified in almost three dozen Gram-positive bacterial species. Some of them have been well characterized and there are some interesting similarities and differences between them. In most cases, the presence of accessory Sec components is closely associated with functions in intracellular survival and virulence. Below, we will discuss in detail of the accessory Sec system in different Gram-positive bacterial species. We will divide the bacterial species into two groups based on their accessory Sec type: SecA2-only systems and SecA2/SecY2 systems.

3.1 Bacterial Species with SecA2-Only System

3.1.1 *Mycobacterium* Species

Mycobacteria possess the Sec transport systems for the translocation of proteins across the cytoplasmic membrane. All essential Sec components are present in Mycobacteria, both in pathogenic species, and non-pathogenic species. Interestingly, all Mycobacteria also possess the accessory SecA2, but they lack a SecY2, and thus, the system is called the SecA2-only system (Braunstein et al. 2001; Rigel and Braunstein 2008). Mycobacterial SecA2 proteins share only about 50 % similarity with their corresponding SecA homolog (SecA1). Structural studies in *M. tuberculosis* show that SecA2 is smaller in size, as compared to SecA1, due to some deletions in C domain as discussed earlier. Functional studies in *M. tuberculosis* and *Mycobacterium smegmatis* show that SecA1 is essential and likely it functions as the housekeeping SecA similar to *E. coli* SecA (Braunstein et al. 2001). In contrast, SecA2 is non-essential since the deletion mutants could be constructed in several

Mycobacteria, including *M. tuberculosis* (Braunstein et al. 2003), *M. smegmatis* (Braunstein et al. 2001), and *Mycobacterium marinum* (Watkins et al. 2012). Cell fractionation studies in *M. smegmatis* show that SecA2 is predominantly cytosolic, while SecA1 is equally distributed between membrane and cytosolic fractions (Rigel et al. 2009). SecA1 and SecA2 have independent functions in protein export (Braunstein et al. 2001) and are present in equivalent amounts in *M. tuberculosis*. The role of SecA2 appears limited to only a subset of proteins. Similar to SecA1, SecA2 also bears ATPase activity which is required for SecA2-mediated protein export in both *M. tuberculosis* (Hou et al. 2008) and *M. smegmatis* (Rigel et al. 2009). SecA2 is also important for the virulence of *M. tuberculosis* (Braunstein et al. 2003; Kurtz et al. 2006) and *M. marinum* (Watkins et al. 2012).

Mycobacterial SecA2 appears to transport different types of substrates. Proteomic studies in the non-pathogenic *M. smegmatis* identified Msmeg1704 and Msmeg1712 as SecA2-dependent proteins (Gibbons et al. 2007). Both are sugar-binding proteins and contain a predicted N-terminal lipoprotein sequence. In *M. tuberculosis*, two SecA2-dependent proteins were identified: superoxide dismutase A (SodA) and catalase-peroxidase (KatG) (Braunstein et al. 2003). These two proteins play a role in surviving oxidative stress (Braunstein et al. 2003). Interestingly, both proteins lack a signal sequence. A proteomic study in *M. marinum* revealed protein kinase G (PknG) as a SecA2-dependent protein which is important for virulence (Van der Woude et al. 2014). Similar to SodA and KatG in *M. tuberculosis*, PknG also does not possess a signal sequence. Although these proteins lacking a signal sequence can be translocated via the Sec pathway (Krehenbrink et al. 2011), it is unclear how SecA2 specifically recognizes these substrates and how it is targeted to the Sec pathway. Recent studies in *M. smegmatis* suggest that protein export by the Mycobacterial SecA2 is determined by the preprotein mature domain instead of a signal sequence (Feltcher et al. 2013). Interestingly, the study also showed that the mature domain of SecA2 substrates can also be exported by the Twin-arginine translocation (Tat) pathway when fused to a signal peptide for the Tat pathway. This suggest that SecA2 substrates may have a tendency to fold prior to export (Feltcher et al. 2013) and that SecA2 facilitates the targeting and export of such unique substrates.

In Mycobacteria and other organisms with SecA2-only system, SecA2 is suggested to work together with the canonical SecYEG protein. A SecA2-SecYEG association has been proposed based on a genetic study in *M. smegmatis* (Ligon et al. 2013). Additionally, structural conservation of SecA–SecY contact sites in SecA2 structure is in line with the expected interaction between SecA2 and SecYEG (Swanson et al. 2016). However, there is no evidence that SecA2 indeed directly interacts with SecYEG. SecA1 depletion studies in *M. smegmatis* indicate that SecA2-dependent proteins also depend on SecA1 (Rigel et al. 2009). Recent in vitro studies show that the *M. tuberculosis* SecA2 interacts with itself to form a homodimer, but that it can also interact with SecA1 to form a heterodimer (Prabudiansyah et al. 2015). Possibly, SecA2 interacts with the SecA1/YEG translocase via an interaction with SecA1, explaining why SecA2 substrates also

require SecA1 for translocation while SecA2 based on cell fractionation studies is mostly cytosolic.

3.1.2 *Listeria* Species

Listeria species possess SecA2 but lack SecY2, similar to Mycobacteria (Lenz and Portnoy 2002). Sequence alignments predict that SecA2 of Listeria harbors all functional domains found in SecA1, including HWD (Bensing et al. 2014). *Listeria monocytogenes* SecA2 is not essential for growth, but involved in virulence and protective immunity (Lenz and Portnoy 2002; Lenz et al. 2003). SecA2 is needed for the export of a large number of proteins, and by proteomic studies, at least 17 substrates were identified, some of which carry a signal sequences, while others do not (Lenz et al. 2003; Archambaud et al. 2006; Renier et al. 2013). Two of SecA2-dependent proteins that contain a signal sequence have been studied in more detail. These are p60 (protein of 60 kDa) also called CwhA (Cell wall hydrolase A) and NamA also called MurA (N-acetylmuramidase A) (Lenz and Portnoy 2002; Lenz et al. 2003; Carroll et al. 2003). SecA2-dependent proteins without a signal sequence include a homolog of SodA, MnSOD (manganese superoxide dismutase) (Lenz et al. 2003; Archambaud et al. 2006), and LAP (Listeria adhesion protein) (Burkholder et al. 2009). All substrates studied thus far appear to function in virulence (Lenz et al. 2003; Carroll et al. 2003; Burkholder et al. 2009).

Recent studies in *L. monocytogenes* revealed that SecA2-dependent protein secretion requires SecA1 (Halbedel et al. 2014). These data support the hypothesis mentioned earlier that in bacteria with SecA2-only system, SecA2 works together with the canonical SecYEG/SecA1 system. Another study shows that the polar-localized cell division protein DivIVA is required for the translocation of the SecA2-dependent proteins p60 and NamA, and it was suggested that DivIVA influences the activity of SecA2 (Halbedel et al. 2012). *S*ecA2 was also identified in non-pathogenic species such as *Listeria innocua*, *Listeria welshimeri*, *Listeria seeligeri*, *Listeria grayi,* and *Listeria marthii*, but the exact role of SecA2 in those species is unknown. In *Listeria innocua*, SecA2 is involved in the translocation of NamA (Mishra et al. 2011). Another study in a non-pathogenic listeria on LAP translocation failed to provide evidence that SecA2 promotes bacterial adhesion. However, the lack of an effect might be due to the low level of expression of LAP in this organism (Jagadeesan et al. 2010).

3.1.3 *Clostridium difficile*

Clostridium difficile is the only clostridial species found to possess the SecA2 protein. *C. difficile* SecA2 is mostly cytosolic, whereas SecA1 localizes to the membrane (Fagan and Fairweather 2011). In contrast to most of bacterial SecA2 which are not essential, SecA2 of *C. difficile* is essential for viability (Fagan and Fairweather 2011). Additionally, the *C. difficile* SecA2 shows high homology to the

SecA1 protein (~ 80 % similarity). The *C. difficile secA2* gene is encoded within a gene cluster encoding surface layer (S-layer) proteins. These proteins form a paracrystalline sheets that assembles at the cell surface. Two proteins have been identified as the major substrates of the *C. difficile* SecA2 system, the S-layer protein (SlpA), which is the main component of the S-layer, and a cell wall protein (CwpV) (Fagan and Fairweather 2011). SlpA and CwpV both contain a signal sequence. SlpA is essential for viability, suggesting the essential function of SecA2 in *C. difficle*.

3.1.4 *Bacillus* Species

The accessory SecA2 protein is present in some Bacillus species such as *Bacillus anthracis, Bacillus cereus, Bacillus thuringiensis, Bacillus smithi,* and *Bacillus methanolicus*, but absent in *Bacillus subtilis*. In *B. anthracis,* SecA2 is encoded within an S-layer gene cluster, similar to the *C. difficle* SecA2. However, the *B. anthracis* SecA2 is not essential for viability, but important for the translocation of a surface array protein (Sap) and the extractable antigen 1 (EA1), which are two major components of the *Bacillus* S-layer (Nguyen-Mau et al. 2012). Sap is encoded in the *secA2* locus, whereas EA1 is encoded in other genomic region. Apparently, SecA2 also is specialized for the translocation of S-layer protein in multiple organisms. *B. anthracis* also possess the accessory SecY2 protein, but this protein appears not essential for the translocation of the aforementioned SecA2 substrates; thus, it belongs to the group of SecA2-only system. *B. anthracis* SecA2 associate with the accessory secretion factors, SlaP and SlaQ (Nguyen-Mau et al. 2012, 2015). These two proteins are encoded immediately downstream of the *secA2* gene. Thus, it appears that SecA2 functions together with SlaP and SlaQ to promote the S-layer assembly in *B. anthracis* (Nguyen-Mau et al. 2012, 2015). SlaP and SlaQ are also present in other pathogenic bacillus species, such as *B. cereus*, and likely fulfill a similar function.

3.2 *Bacterial Species with SecA2/SecY2 System*

3.2.1 *Streptococcus* Species

Streptococcus species possess both the SecA2 and SecY2 proteins that form the SecA2/SecY2 system. These proteins were originally identified and characterized in *Streptococcus gordonii* (Bensing and Sullam 2002) and *Streptococcus parasanguinis* (Chen et al. 2004). The predicted domain organization of Streptococcal SecA2 is similar to SecA2 of other organisms that lack SecY2. However, the Streptococcal SecA2 functions independent of the canonical Sec system and likely forms a separate Sec system with SecY2 (Bensing and Sullam 2002). Cell fractionation studies in *Streptococcus parasanguinis* show that SecA2 is mostly

associated with the membrane (Chen et al. 2006), in contrast to other bacterial SecA2 s which are predominantly cytosolic. SecA2 and SecY2 are not essential for viability or the translocation of most secretory proteins (Bensing et al. 2014). In contrast to the multiple type of substrates in the SecA2-only system, Streptococcal SecA2/SecY2 system exports more specific substrate. The system appear to be specialized in the transport of SRR glycoproteins (Bensing and Sullam 2002; Bensing et al. 2014). SRR glycoproteins are a family of adhesins in Gram-positive bacteria that fulfill an important role in pathogenicity (Zhou and Wu 2009; Lizcano et al. 2012).

In *S. gordonii,* SecA2 and SecY2 play an important role in the translocation of GspB. GspB is a serine-rich glycoprotein that mediates the binding of *S. gordonii* to platelets, and is encoded within the *secA2/secY2* gene cluster. GspB contains a relatively long signal sequence (90 amino acids) and a specific domain called the accessory Sec transport (AST) domain that is essential for targeting to the SecA2/SecY2 translocase (Bensing and Sullam 2010). The *S. gordonii* SecA2/SecY2 system also includes other accessory secretion proteins (Asps) that are all encoded within the same operon: Asp1 to Asp5. Asp4 and Asp5 show homology to the canonical SecE and SecG proteins, respectively (Takamatsu et al. 2005). They are predicted to form the translocation channel together with SecY2, but this remains to be demonstrated. Asp1, Asp2, and Asp3 interact with each other and are required for the translocation of GspB (Seepersaud et al. 2010). Asp2 and Asp3 directly interact with GspB via the SRR domain (Yen et al. 2011). Both proteins also interact with SecA2, which may also be involved in the targeting of GspB preprotein to the translocation site (Seepersaud et al. 2010; Bensing et al. 2012; Yen et al. 2013). Asp2 is important for the correct glycosylation of GspB during translocation (Seepersaud et al. 2012). Several cytosolic glycosylation factors, such as GftA (Gtf1), GftB (Gtf2), Gly, and Nss, are involved in the glycosylation of GspB prior to translocation (Takamatsu et al. 2004a, b; Zhou and Wu 2009). Therefore, one of the key features of this system is that the non-canonical Sec translocase translocated glycosylated precursor proteins.

In *S. parasanguinis,* SecA2/SecY2 are important for the translocation of Fap1, a homolog of GspB (Chen et al. 2004). *S. parasanguinis* possess also homologs of Asp1, Asp2, and Asp3, termed glycosylation-associated proteins 1 (Gap1), Gap2, and Gap3, respectively, but seems to lack the Asp4 and Asp5 proteins. Similar to *S. gordonii* Asps*, S. parasanguinis* Gaps are also encoded in the same operon as *secY2, secA2*, and also *fap1*. Gap1 and Gap3 interact with SecA2 in vitro (Zhou et al. 2011), and both proteins are suggested to be involved in the complete glycosylation of Fap1 (Zhou and Wu 2009). Based on the combined studies in *S. gordonii and S. parasanguinis*, it is suggested that Gap1-3 (Asp1-3) forms a single functional complex with dual functions: First, it is involved in targeting the partially glycosylated substrate to the SecA2/SecY2 translocase, and second, it is needed for the complete glycosylation of the substrate during translocation (Zhou and Wu 2009; Bensing et al. 2014). In vitro study in *S. parasanguinis* shows that SecA2 and Gap3 associate with the canonical SecA1 (Zhou et al. 2011), suggesting cross talk between the accessory and the canonical Sec system. However, the

biological function of this association has not been established. In general, pre-proteins that are exported by the Streptococcal accessory Sec system are glycosylated and cannot be exported by the canonical Sec system (Bensing et al. 2014). However, glycosylation is not required for protein translocation in the accessory Sec system, since the unglycosylated substrate can still be secreted by this system (Bensing et al. 2005). Interestingly, the non-glycosylated versions of GspB and Fap1 can also be translocated by the canonical Sec system (Bensing et al. 2005; Chen et al. 2007), suggesting that glycosylation may take part in the targeting of these substrates to the accessory Sec system.

3.2.2 *Staphylococcus* Species

The Staphylococcal accessory Sec system consists of both SecA2 and SecY2, similar to Streptococcus. The two proteins were discovered in *Staphylococcus aureus* (Siboo et al. 2008) and found to be present also in some of other staphylococcal species: *Staphylococcus epidermidis, Staphylococcus warneri,* and *Staphylococcus carnosus.* In *S. aureus,* SecA2 and SecY2 are not essential for viability, but required for translocation of SraP (serine-rich adhesin for platelets) (Siboo et al. 2008). However, it is still need to be verified whether SraP is selectively translocated via the SecA2/SecY2 system or not. SraP is a homolog of GspB of *S. gordonii* and fulfills a role in binding to human platelets (Siboo et al. 2005). The accessory Sec locus of *S. aureus* is similar to that of *S. gordonii*. It also contains the genes encode Asp1-Asp3 that all are required for the transport of SraP (Sibbald et al. 2010). However, *S. aureus* lacks four proteins present in *S. gordonii*, i.e., Gly, Nss, Asp4, and Asp5. Gly and Nss are important for the glycosylation of GspB in *S. gordonii*, suggesting that there is a difference in the glycosylation mechanism between staphylococcus and streptococcus. Asp4 and Asp5 are predicted to be the part of accessory membrane channel together with SecY2 in *S. gordonii*. The absence of Asp4 and Asp5 in *S. aureus* indicates that the requirements for these Asps are specific for some species, i.e., *S. gordonii,* while other proteins might function as partner subunits of the *S. aureus* SecY2. Interestingly, genetic study suggests that *S. aureus* SecY2 function together with SecG (Sibbald et al. 2010). Since *S. aureus* lacks a second set of *secE* and *secG* genes, these findings suggest that the *S. aureus* SecY2 might form an alternative translocation channel with SecE1 and SecG1.

3.3 Other Gram-positive Bacteria

The accessory Sec system is more widespread in Gram-positive bacteria, but mostly poorly characterized. Among these is *Corynebacterium glutamicum* that possess the SecA2-only system. *C. glutamicum* SecA2 is essential for viability (Caspers and Freudl 2008), but no specific substrates have been identified thus far. The accessory

SecA2 and/or SecY2 proteins are also present in some of *Gordonia*, *Pediococcus*, *Enterococcus*, and *Lactobacillus* species (Bensing et al. 2014), but not further characterized.

4 Conclusions

The canonical Sec translocase is omnipresent in bacteria, and similar systems exist in Gram-negative and positive bacteria. In contrast, accessory Sec components are found only in Gram-positive bacteria where they fulfill diverse functions in protein secretion. During the last fifteen years, a better understanding on Gram-positive bacterial accessory Sec system has been obtained. Since this system also plays an important role in bacterial virulence, the studies also revealed mechanisms of pathogenesis, and potentially, the system may function as a possible drug target. The current model of the Gram-positive accessory Sec system is shown in Fig. 4, which is based on the combined studies discussed above.

Despite the growing knowledge on the system, several important questions remain to be answered. One of the main question is how the accessory SecA2 specifies and distinguishes its substrates from the canonical SecA1. Another

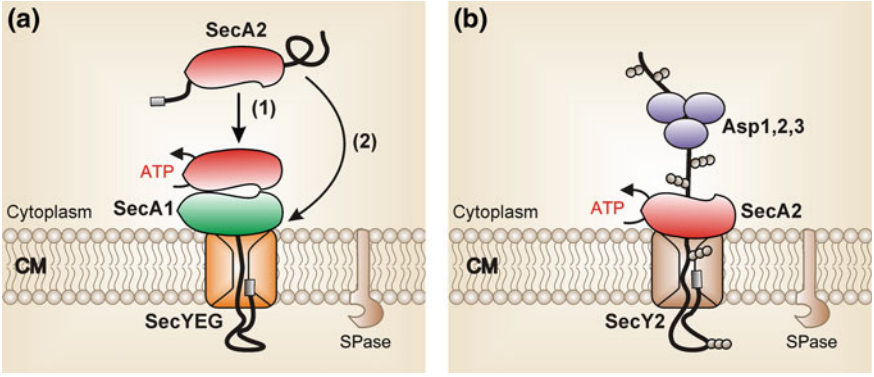


Fig. 4 The accessory Sec system of Gram-positive bacteria. **a** SecA2-only system. SecA2-dependent proteins are recognized by the accessory SecA2 (*red*) and are targeted to the SecYEG channel (*orange*), either via (*1*) the housekeeping SecA1 (*green*) or (2) direct interaction of SecA2 with SecYEG. The ATPase activity of either SecA2 or SecA1 or both provides the energy for the translocation of the proteins through the SecYEG pore. **b** SecA2/SecY2 system. Partially glycosylated preproteins are targeted to the accessory SecA2 protein (*red*) by the Asp1-3 (Gap1-3) complex (*purple*). During the translocation process, the Asp1-3 (Gap1-3) complex modifies the glycan composition and completes the glycosylation of the preproteins (*brown dot*). The ATPase activity of SecA2 provides the energy for the translocation of the fully glycosylated preprotein through the SecY2 membrane channel (*brown*). *SPase,* signal peptidase; *CM,* cytoplasmic membrane

interesting question is whether there is cross talk between the accessory and the canonical Sec system. Future study may yield new insights on the overall Sec system in Gram-positive bacteria.

Acknowledgements This work was supported by Earth and Life Sciences (ALW) which is supported by the Netherlands Organization for Scientific Research (NWO) grant no 854.11.001. The work of I.P. was supported by a Bernoulli grant of the University of Groningen.

References

Akita M, Shinkai A, Matsuyama S, Mizushima S (1991) SecA, an essential component of the secretory machinery of *Escherichia coli*, exists as homodimer. Biochem Biophys Res Commun 174:211–216

Archambaud C, Nahori M-A, Pizarro-Cerda J et al (2006) Control of Listeria superoxide dismutase by phosphorylation. J Biol Chem 281:31812–31822. doi:10.1074/jbc.M606249200

Benach J, Chou Y-T, Fak JJ et al (2003) Phospholipid-induced monomerization and signal-peptide-induced oligomerization of SecA. J Biol Chem 278:3628–3638. doi:10.1074/jbc.M205992200

Bensing B, Sullam PM (2002) An accessory sec locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets. Mol Microbiol 44:1081–1094. doi:10.1046/j.1365-2958.2002.02949.x

Bensing B, Sullam PM (2010) Transport of preproteins by the accessory Sec system requires a specific domain adjacent to the signal peptide. J Bacteriol 192:4223–4232. doi:10.1128/JB.00373-10

Bensing B, Takamatsu D, Sullam PM (2005) Determinants of the streptococcal surface glycoprotein GspB that facilitate export by the accessory Sec system. Mol Microbiol 58:1468–1481. doi:10.1111/j.1365-2958.2005.04919.x

Bensing B, Yen YT, Seepersaud R, Sullam PM (2012) A Specific interaction between SecA2 and a region of the preprotein adjacent to the signal peptide occurs during transport via the accessory Sec system. J Biol Chem 287:24438–24447. doi:10.1074/jbc.M112.378059

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

Bieker KL, Phillips GJ, Silhavy T (1990) The sec and prl genes of *Escherichia coli*. J Bioenerg Biomembr 22:291–310

Braunstein M, Brown AM, Kurtz Jr. S, Jacobs WR (2001) Two nonredundant SecA homologues function in *Mycobacteria*. J Bacteriol 183:6979–6990. doi:10.1128/JB.183.24.6979

Braunstein M, Espinosa BJ, Chan J et al (2003) SecA2 functions in the secretion of superoxide dismutase A and in the virulence of *Mycobacterium tuberculosis*. Mol Microbiol 48:453–464. doi:10.1046/j.1365-2958.2003.03438.x

Breukink E, Nouwen N, van Raalte A et al (1995) The C terminus of SecA is involved in both lipid binding and SecB binding. J Biol Chem 270:7902–7907

Bron S, Venema G, Van Dijl JM (2000) Chaperone-like activities of the CsaA protein of *Bacillus subtilis*. Microbiology 146:77–88

Brundage L, Hendrick JP, Schiebel E et al (1990) The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of precursor protein tiranslocation. Cell 62:649–657

Bu Z, Wang L, Kendall D (2003) Nucleotide binding induces changes in the oligomeric state and conformation of SecA in a lipid environment: a small-angle neutron-scattering study. J Mol Biol 332:23–30. doi:10.1016/S0022-2836(03)00840-4

Burkholder KM, Kim K-P, Mishra KK et al (2009) Expression of LAP, a SecA2-dependent secretory protein, is induced under anaerobic environment. Microbes Infect Pasteur 11:859–867. doi:10.1016/j.micinf.2009.05.006

Cabellis RJ, Dolan KM, Qian L, Oliver DB (1991) Characterization of membrane-associated and soluble states of SecA protein from wild-type and SecA51(TS) mutant strains of *Escherichia coli*. J Biol Chem 266:24420–24427

Cao TB, Saier Jr MHS (2003) The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. Biochim Biophys Acta 1609:115–125

Carroll S a., Hain T, Technow U, et al (2003) Identification and characterization of a peptidoglycan hydrolase, MurA, of *Listeria monocytogenes*, a muramidase needed for cell separation. J Bacteriol 185:6801–6808. doi: 10.1128/JB.185.23.6801-6808.2003

Caspers M, Freudl R (2008) *Corynebacterium glutamicum* possesses two secA homologous genes that are essential for viability. Arch Microbiol 189:605–610. doi:10.1007/s00203-008-0351-0

Castanié-Cornet M-P, Bruel N, Genevaux P (2014) Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. Biochim Biophys Acta 1843:1442–1456. doi:10.1016/j.bbamcr.2013.11.007

Chen Q, Wu H, Fives-Taylor PM (2004) Investigating the role of secA2 in secretion and glycosylation of a fimbrial adhesin in *Streptococcus parasanguis* FW213. Mol Microbiol 53:843–856. doi:10.1111/j.1365-2958.2004.04116.x

Chen Q, Wu H, Kumar R et al (2006) SecA2 is distinct from SecA in immunogenic specificity, subcellular distribution and requirement for membrane anchoring in *Streptococcus parasanguis*. FEMS Microbiol Lett 264:174–181. doi:10.1111/j.1574-6968.2006.00455.x

Chen Q, Sun B, Wu H et al (2007) Differential roles of individual domains in selection of secretion route of a *Streptococcus parasanguinis* serine-rich adhesin, Fap1. J Bacteriol 189:7610–7617. doi:10.1128/JB.00748-07

Dalbey RE, Lively MO, Bron S, van Dijl JM (1997) The chemistry and enzymology of the type I signal peptidases. Protein Sci 6:1129–1138. doi:10.1002/pro.5560060601

Das S, Oliver DB (2011) Mapping of the SecA·SecY and SecA·SecG interfaces by site-directed in vivo photocross-linking. J Biol Chem 286:12371–12380. doi:10.1074/jbc.M110.182931

Das S, Stivison E, Folta-stogniew E, Oliver D (2008) Reexamination of the role of the amino terminus of SecA in promoting its dimerization and functional state. J Bacteriol 190:7302–7307. doi:10.1128/JB.00593-08

De Keyzer J, van der Does C, Driessen JM (2003) The bacterial translocase: a dynamic protein channel complex. Cell Mol Life Sci 60:2034–2052. doi:10.1007/s00018-003-3006-y

De Keyzer J, van der Sluis EO, Spelbrink REJ et al (2005) Covalently dimerized SecA is functional in protein translocation. J Biol Chem 280:35255–35260. doi:10.1074/jbc.M506157200

Ding H, Hunt JF, Mukerji I, Oliver D (2003) *Bacillus subtilis* SecA ATPase exists as an antiparallel dimer in solution. Biochemistry 42:8729–8738

Driessen AJM (1993) SecA, the peripheral subunit of the *Escherichia coli* precursor protein translocase, is functional as a dimer. Biochemistry 32:13190–13197. doi:10.1021/bi00211a030

Driessen AJM, Nouwen N (2008) Protein translocation across the bacterial cytoplasmic membrane. Annu Rev Biochem 77:643–667. doi:10.1146/annurev.biochem.77.061606.160747

Duong F, Wickner W (1997) Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. EMBO J 16:2756–2768

Economou A, Wickner W (1994) SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. Cell 78:835–843. doi:10.1016/S0092-8674(94)90582-7

Erlandson KJ, Miller SBM, Nam Y et al (2008) A role for the two-helix finger of the SecA ATPase in protein translocation. Nature 455:984–987. doi:10.1038/nature07439

Fagan RP, Fairweather NF (2011) *Clostridium difficile* has two parallel and essential Sec secretion systems. J Biol Chem 286:27483–27493. doi:10.1074/jbc.M111.263889

Fekkes P, Van Der Does C, Driessen AJM (1997) The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. EMBO J 16:6105–6113

Fekkes P, de Wit JG, Boorsma A et al (1999) Zinc stabilizes the SecB binding site of SecA. Biochemistry 38:5111–5116. doi:10.1021/bi982818r

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

Feltcher ME, Gibbons HS, Ligon LS, Braunstein M (2013) Protein export by the Mycobacterial SecA2 system is determined by the preprotein mature domain. J Bacteriol 195:672–681. doi:10.1128/JB.02032-12

Frauenfeld J, Gumbart J, Van Der Sluis EO et al (2011) Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nat Struct Mol Biol 18:614–621. doi:10.1038/nsmb.2026

Freudl R (2013) Leaving home ain't easy: protein export systems in Gram-positive bacteria. Res Microbiol 164:664–674. doi:10.1016/j.resmic.2013.03.014

Gelis I, Bonvin AMJJ, Keramisanou D et al (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131:756–769. doi:10.1016/j.cell.2007.09.039

Gibbons HS, Wolschendorf F, Abshire M et al (2007) Identification of two *Mycobacterium smegmatis* lipoproteins exported by a SecA2-dependent pathway. J Bacteriol 189:5090–5100. doi:10.1128/JB.00163-07

Gouridis G, Karamanou S, Sardis MF et al (2013) Quaternary dynamics of the SecA motor drive translocase catalysis. Mol Cell 52:655–666. doi:10.1016/j.molcel.2013.10.036

Gumbart J, Schulten K (2008) The roles of pore ring and plug in the SecY Protein-conducting channel. J Gen Physiol 132:709–719. doi:10.1085/jgp.200810062

Halbedel S, Hahn B, Daniel RA, Flieger A (2012) DivIVA affects secretion of virulence-related autolysins in *Listeria monocytogenes*. Mol Microbiol 83:821–839. doi:10.4161/viru.20747

Halbedel S, Reiss S, Hahn B et al (2014) A systematic proteomic analysis of *Listeria monocytogenes* house-keeping protein secretion systems. Mol Cell Proteomics 13:3063–3081. doi:10.1074/mcp.M114.041327

Hanada M, Nishiyama K, Mizushimas S, Tokudal H (1994) Reconstitution of an efficient protein translocation machinery comprising SecA and the three membrane proteins, SecY, SecE, and SecG (p12)*. J Biol Chem 269:23625–23631

Hendricks JP, Wickner W (1991) SecA protein needs both acidic phospholipids and SecY/E protein for functional high-affinity binding to the *Escherichia coli* plasma membrane. J Biol Chem 266:24596–24600

Hou JM, D'Lima NG, Rigel NW et al (2008) ATPase activity of *Mycobacterium tuberculosis* SecA1 and SecA2 proteins and its importance for SecA2 function in macrophages. J Bacteriol 190:4880–4887. doi:10.1128/JB.00412-08

Hunt JF, Weinkauf S, Henry L et al (2002) Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. Science 297:2018–2026. doi:10.1126/science.1074424

Jagadeesan B, Koo OK, Kim K-P et al (2010) LAP, an alcohol acetaldehyde dehydrogenase enzyme in *Listeria*, promotes bacterial adhesion to enterocyte-like Caco-2 cells only in pathogenic species. Microbiology 156:2782–2795. doi:10.1099/mic.0.036509-0

Jeong SM, Yoshikawa H, Takahashi H (1993) Isolation and characterization of the secE homologue gene of *Bacillus subtilis*. Mol Microbiol 10:133–142. doi:10.1111/j.1365-2958.1993.tb00910.x

Jilaveanu LB, Oliver D (2006) SecA dimer cross-linked at its subunit interface is functional for protein translocation. J Bacteriol 188:335–338. doi:10.1128/JB.188.1.335

Jilaveanu LB, Zito CR, Oliver D (2005) Dimeric SecA is essential for protein translocation. Proc Natl Acad Sci USA 102:7511–7516. doi:10.1073/pnas.0502774102

Josefsson L, Randall LL (1981) Processing in vivo of precursor maltose-binding protein in *Escherichia coli* occurs post-translationally as well as co-translationally. J Biol Chem 256:2504–2507

Karamanou S, Vrontou E, Sianidis G et al (1999) A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. Mol Microbiol 34:1133–1145. doi:10.1046/j.1365-2958.1999.01686.x

Karamanou S, Sianidis G, Gouridis G et al (2005) *Escherichia coli* SecA truncated at its termini is functional and dimeric. FEBS Lett 579:1267–1271. doi:10.1016/j.febslet.2005.01.025

Krehenbrink M, Edwards A, Downie JA (2011) The superoxide dismutase SodA is targeted to the periplasm in a SecA-dependent manner by a novel mechanism. Mol Microbiol 82:164–179. doi:10.1111/j.1365-2958.2011.07803.x

Kurtz S, McKinnon KP, Runge MS et al (2006) The SecA2 secretion factor of *Mycobacterium tuberculosis* promotes growth in macrophages and inhibits the host immune response. Infect Immun 74:6855–6864. doi:10.1128/IAI.01022-06

Kusters I, van den Bogaart G, Kedrov A et al (2011) Quaternary structure of SecA in solution and bound to SecYEG probed at the single molecule level. Structure 19:430–439. doi:10.1016/j.str.2010.12.016

Lecker S, Lill R, Ziegelhoffer T et al (1989) Three pure chaperone proteins of *Escherichia coli*—SecB, trigger factor and GroEL form soluble complexes with precursor proteins in vitro. EMBO J 8:2703–2709

Lecker S, Schiebel E, Hendrick JF, Wickner W (1990) The binding cascade of SecB to SecA to Sey/E mediates preprotein targeting to the *E. coli* plasma membrane. Cell 63:269–279

Lenz LL, Portnoy D (2002) Identification of a second Listeria secA gene associated with protein secretion and the rough phenotype. Mol Microbiol 45:1043–1056. doi:10.1046/j.1365-2958.2002.03072.x

Lenz LL, Mohammadi S, Geissler A, Portnoy D (2003) SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis. Proc Natl Acad Sci USA 100:12432–12437. doi:10.1073/pnas.2133653100

Ligon LS, Rigel NW, Romanchuk A et al (2013) Suppressor analysis reveals a role for SecY in the SecA2-dependent protein export pathway of *Mycobacteria*. J Bacteriol 195:4456–4465. doi:10.1128/JB.00630-13

Lill R, Dowhan W, Wickner W (1990) The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. Cell 60:271–280

Lizcano A, Sanchez CJ, Orihuela CJ (2012) A role for glycosylated Serine-rich repeatproteins in Gram-positive bacterial pathogenesis. Mol Oral Microbiol 27:257–269. doi:10.1111/j.2041-1014.2012.00653.x.A

Lycklama A, Nijeholt J, Driessen AJM (2012) The bacterial Sec-translocase: structure and mechanism. Philos Trans R Soc Lond B Biol Sci 367:1016–1028. doi:10.1098/rstb.2011.0201

Lycklama A, Nijeholt JA, De Keyzer J, Prabudiansyah I, Driessen AJM (2013) Characterization of the supporting role of SecE in protein translocation. FEBS Lett 587:3083–3088

Ménétret J-F, Schaletzky J, Clemons WM et al (2007) Ribosome binding of a single copy of the SecY complex: implications for protein translocation. Mol Cell 28:1083–1092. doi:10.1016/j.molcel.2007.10.034

Mishra KK, Mendonca M, Aroonnual A et al (2011) Genetic organization and molecular characterization of secA2 locus in *Listeria* species. Gene 489:76–85. doi:10.1016/j.gene.2011.08.015

Mori H, Ito K (2006) The long alpha-helix of SecA is important for the ATPase coupling of translocation. J Biol Chem 281:36249–36256. doi:10.1074/jbc.M606906200

Mori H, Sugiyama H, Yamanaka M et al (1998) Amino-terminal region of SecA is involved in the function of SecG for protein translocation into *Escherichia coli* membrane vesicles. J Biochem 124:122–129. doi:10.1093/oxfordjournals.jbchem.a022070

Müller JP, Ozegowski J, Vettermann S et al (2000) Interaction of *Bacillus subtilis* CsaA with SecA and precursor proteins. Biochem J 348:367–373. doi:10.1042/bj3480367

Murphy CK, Beckwith JON (1994) Residues essential for the function of SecE, a membrane component of the *Escherichia coli* secretion apparatus—are located in a conserved cytoplasmic region. Proc Natl Acad Sci USA 91:2557–2561

Musial-Siwek M, Rusch SL, Kendall D (2005) Probing the affinity of SecA for signal peptide in different environments. Biochemistry 44:13987–13996. doi:10.1021/bi050882k

Musial-Siwek M, Rusch SL, Kendall D (2007) Selective photoaffinity labeling identifies the signal peptide binding domain on SecA. J Mol Biol 365:637–648. doi:10.1016/j.jmb.2006.10.027

Nguyen-Mau S-M, Oh S-Y, Kern VJ et al (2012) Secretion genes as determinants of *Bacillus anthracis* chain length. J Bacteriol 194:3841–3850. doi:10.1128/JB.00384-12

Nguyen-Mau S-M, Oh S-Y, Schneewind DI et al (2015) *Bacillus anthracis* SlaQ promotes S-layer protein assembly. J Bacteriol 197:32163. doi:10.1128/JB.00492-15

Nishiyama K, Hanada M (1994) Disruption of the gene encoding p12 (SecG) reveals the direct involvement and important function of SecG in the protein translocation of *Escherichia coli* at low temperature. EMBO J 13:3272–3277

Nishiyama K, Mizushima S (1992) The carboxy-terminal region of SecE interacts with SecY and is functional in the reconstitution of protein translocation activity in *Escherichia coli*. J Biol Chem 267:7170–7176

Nishiyama K, Mizushimal S (1993) A novel membrane protein involved in protein translocation across the cytoplasmic membrane of *Escherichia coli*. EMBO J 12:3409–3415

Oliver DB, Beckwith J (1982) Identification of a new gene (secA) and gene product involved in the secretion of envelope proteins in *Escherichia coli*. J Bacteriol 150:686–691

Or E, Navon A, Rapoport T (2002) Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. EMBO J 21:4470–4479

Or E, Boyd D, Gon S et al (2005) The bacterial ATPase SecA functions as a monomer in protein translocation. J Biol Chem 280:9097–9105. doi:10.1074/jbc.M413947200

Osborne AR, Clemons WM, Rapoport T (2004) A large conformational change of the translocation ATPase SecA. Proc Natl Acad Sci USA 101:10937–10942. doi:10.1073/pnas.0401742101

Papanikolau Y, Papadovasilaki M, Ravelli RBG et al (2007) Structure of dimeric SecA, the *Escherichia coli* preprotein translocase motor. J Mol Biol 366:1545. doi:10.1016/j.jmb.2006.12.049

Papanikou E, Karamanou S, Baud C et al (2005) Identification of the preprotein binding domain of SecA. J Biol Chem 280:43209–43217. doi:10.1074/jbc.M509990200

Plath K, Mothes W, Wilkinson BM et al (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell 94:795–807. doi:10.1016/S0092-8674(00)81738-9

Plath K, Wilkinson BM, Stirling CJ, Rapoport TA (2004) Interactions between Sec complex and prepro-alpha-factor during posttranslational protein transport into the endoplasmic reticulum. Mol Biol Cell 15:1–10. doi:10.1091/mbc.E03

Pohlschröder M, Prinz WA, Hartmann E, Beckwith J (1997) Protein translocation in the three domains of life: variations on a theme. Cell 91:563–566

Prabudiansyah I, Kusters I, Driessen AJM (2015) In vitro interaction of the housekeeping SecA1 with the accessory SecA2 protein of *Mycobacterium tuberculosis*. PLoS ONE 10:e0128788. doi:10.1371/journal.pone.0128788

Renier S, Chambon C, Viala D et al (2013) Exoproteomic analysis of the SecA2-dependent secretion in *Listeria monocytogenes* EGD-e. J Proteomics 80:183–195. doi:10.1016/j.jprot.2012.11.027

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

Rigel NW, Gibbons HS, McCann JR et al (2009) The accessory SecA2 system of *Mycobacteria* requires ATP binding and the canonical SecA1. J Biol Chem 284:9927–9936. doi:10.1074/jbc.M900325200

Robson A, Booth AEG, Gold VM et al (2007) A large conformational change couples the ATP binding site of SecA to the SecY protein channel. J Mol Biol 374:965–976. doi:10.1016/j.jmb.2007.09.086

Saparov SM, Erlandson K, Cannon K et al (2007) Determining the conductance of the SecY protein translocation channel for small molecules. Mol Cell 26:501–509. doi:10.1016/j.molcel.2007.03.022

Schatz P, Beckwith J (1990) Genetic analysis of protein export in *Escherichia coli*. Annu Rev Genet 24:215–248

Schatz PJ, Biekerl KL, Ottemann KM et al (1991) One of three transmembrane stretches is sufficient the functioning of the SecE protein, a membrane component of the *E. coli* secretion machinery. EMBO J 1:1749–1757

Schiebel E, Driessen AJM, Hartl FU, Wickner W (1991) ΔμH + and ATP function at different steps of the catalytic cycle of preprotein translocase. Cell 64:927–939. doi:10.1016/0092-8674(91)90317-R

Scotti P, Urbanus ML, Brunner J et al (2000) YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J 19:542–549. doi:10.1093/emboj/19.4.542

Seepersaud R, Bensing B, Yen YT, Sullam PM (2010) Asp3 mediates multiple protein-protein interactions within the accessory Sec system of *Streptococcus gordonii*. Mol Microbiol 78:490–505. doi:10.1111/j.1365-2958.2010.07346.x

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

Sharma V, Arockiasamy A, Ronning DR et al (2003) Crystal structure of *Mycobacterium tuberculosis* SecA, a preprotein translocating ATPase. Proc Natl Acad Sci USA 100:2243–2248

Sibbald MJJB, Winter T, van der Kooi-Pol MM et al (2010) Synthetic effects of secG and secY2 mutations on exoproteome biogenesis in *Staphylococcus aureus*. J Bacteriol 192:3788–3800. doi:10.1128/JB.01452-09

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

Siboo IR, Chaffin DO, Rubens CE, Sullam PM (2008) Characterization of the accessory Sec system of *Staphylococcus aureus*. J Bacteriol 190:6188–6196. doi:10.1128/JB.00300-08

Sullivan JT, Young EF, McCann JR, Braunstein M (2012) The *Mycobacterium tuberculosis* SecA2 system subverts phagosome maturation to promote growth in macrophages. Infect Immun 80:996–1006. doi:10.1128/IAI.05987-11

Suzuki H, Nishiyama K, Tokuda H (1998) Coupled structure change of SecA and SecG revealed by the synthetic lethality of the secAcsR11 and DeltasecG:kan double mutant. Mol Microbiol 29:331–341. doi:10.1046/j.1365-2958.1998.00937.x

Swanson S, Ioerger TR, Rigel NW et al (2016) Structural similarities and differences between two functionally distinct SecA proteins: the *Mycobacterium tuberculosis* SecA1 and SecA2. J Bacteriol 198:720–730. doi:10.1128/JB.00696-15

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

Takamatsu D, Bensing BA, Sullam PM (2004b) Genes in the accessory sec locus of *Streptococcus gordonii* have three functionally distinct effects on the expression of the platelet-binding protein GspB. Mol Microbiol 52:189–203. doi:10.1111/j.1365-2958.2004.03978.x

Takamatsu D, Bensing B, Sullam PM (2005) Two additional components of the accessory sec system mediating export of the *Streptococcus gordonii* platelet-binding protein GspB. J Bacteriol 187:3878–3883. doi:10.1128/JB.187.11.3878-3883.2005

Tanner NK, Linder P (2001) DExD/H box RNA helicases: from generic motors to specific dissociation functions. Mol Cell 8:251–262

Tsukazaki T, Mori H, Echizen Y et al (2011) Structure and function of a membrane component SecDF that enhances protein export. Nature 474:235–238. doi:10.1038/nature09980

Van Den Berg B, Clemons WM Jr, Collinson I et al (2004) X-ray structure of a protein-conducting channel. Nature 427:36–44

Van der Sluis EO, Nouwen N, Koch J et al (2006) Identification of two interaction sites in SecY that are important for the functional interaction with SecA. J Mol Biol 361:839–849. doi:10.1016/j.jmb.2006.07.017

Van der Wolk JPW, De Wit JG, Driessen AJM (1997) The catalytic cycle of the *Escherichia coli* SecA ATPase comprises two distinct preprotein translocation events. EMBO J 16:7297–7304

Van der Woude AD, Stoop EJM, Stiess M et al (2014) Analysis of SecA2-dependent substrates in *Mycobacterium marinum* identifies protein kinase G (PknG) as a virulence effector. Cell Microbiol 16:280–295. doi:10.1111/cmi.12221

Van Wely KHM, Swaving J, Freudl R, Driessen AJM (2001) Translocation of proteins across the cell envelope of Gram-positive bacteria. FEMS Microbiol Rev 25:437–454. doi:10.1111/j.1574-6976.2001.tb00586.x

Vassylyev DG, Mori H, Vassylyeva MN et al (2006) Crystal structure of the translocation ATPase SecA from *Thermus thermophilus* reveals a parallel, head-to-head dimer. J Mol Biol 364:248–258. doi:10.1016/j.jmb.2006.09.061

Von Heijne G (1990) The signal peptide. J Membr Biol 201:195–201

Wang H, Na B, Yang H, Tai PC (2008) Additional in vitro and in vivo evidence for SecA functioning as dimers in the membrane: dissociation into monomers is not essential for protein translocation in *Escherichia coli*. J Bacteriol 190:1413–1418. doi:10.1128/JB.01633-07

Watkins BY, Joshi S, Ball D et al (2012) *Mycobacterium marinum* SecA2 promotes stable granulomas and induces tumor necrosis factor alpha in vivo. Infect Immun 80:3512–3520. doi:10.1128/IAI.00686-12

Woodbury RL, Hardy SJS, Randall LL (2002) Complex behavior in solution of homodimeric SecA. Protein Sci 11:875–882. doi:10.1110/ps.4090102.Protein

Wowor AJ, Yu D, Kendall D, Cole JL (2011) Energetics of SecA dimerization. J Mol Biol 408:87–98. doi:10.1016/j.jmb.2011.02.006

Yen YT, Seepersaud R, Bensing B, Sullam PM (2011) Asp2 and Asp3 interact directly with GspB, the export substrate of the *Streptococcus gordonii* accessory Sec system. J Bacteriol 193:3165–3174. doi:10.1128/JB.00057-11

Yen YT, Cameron T, Bensing B et al (2013) Differential localization of the streptococcal accessory sec components and implications for substrate export. J Bacteriol 195:682–695. doi:10.1128/JB.01742-12

Zhou M, Wu H (2009) Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. Microbiology 155:317–327. doi:10.1099/mic.0.025221-0

Zhou J, Xu Z (2003) Structural determinants of SecB recognition by SecA in bacterial protein translocation. Nat Struct Biol 10:942–947. doi:10.1038/nsb980

Zhou M, Zhang H, Zhu F, Wu H (2011) Canonical SecA associates with an accessory secretory protein complex involved in biogenesis of a streptococcal serine-rich repeat glycoprotein. J Bacteriol 193:6560–6566. doi:10.1128/JB.05668-11

Zimmer J, Rapoport TA (2009) Conformational flexibility and peptide interaction of the translocation ATPase SecA. J Mol Biol 394:606–612. doi:10.1016/j.jmb.2009.10.024

Zimmer J, Li W, Rapoport T (2006) A novel dimer interface and conformational changes revealed by an X-ray structure of *B. subtilis* SecA. J Mol Biol 364:259–265. doi:10.1016/j.jmb.2006.08.044

Zimmer J, Nam Y, Rapoport T (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 455:936–943. doi:10.1038/nature07335

Twin-Arginine Protein Translocation

Vivianne J. Goosens and Jan Maarten van Dijl

Abstract Twin-arginine protein translocation systems (Tat) translocate fully folded and co-factor-containing proteins across biological membranes. In this review, we focus on the Tat pathway of Gram-positive bacteria. The minimal Tat pathway is composed of two components, namely a TatA and TatC pair, which are often complemented with additional TatA-like proteins. We provide overviews of our current understanding of Tat pathway composition and mechanistic aspects related to Tat-dependent cargo protein translocation. This includes Tat pathway flexibility, requirements for the correct folding and incorporation of co-factors in cargo proteins and the functions of known cargo proteins. Tat pathways of several Gram-positive bacteria are discussed in detail, with emphasis on the Tat pathway of *Bacillus subtilis*. We discuss both shared and unique features of the different Gram-positive bacterial Tat pathways. Lastly, we highlight topics for future research on Tat, including the development of this protein transport pathway for the biotechnological secretion of high-value proteins and its potential applicability as an antimicrobial drug target in pathogens.

Abbreviations

Sec pathway	General Secretory pathway
Tat	Twin-arginine translocation
NMR	Nuclear magnetic resonance
Y2H	Yeast two-hybrid

V.J. Goosens
MRC Centre for Molecular Bacteriology and Infection, Section of Microbiology,
Imperial College London, London SW7 2AZ, UK
e-mail: v.goosens@imperial.ac.uk

J.M. van Dijl (✉)
Department of Medical Microbiology, University of Groningen,
University Medical Center Groningen, Hanzeplein 1, P.O. Box 30001,
9700, RB Groningen, The Netherlands
e-mail: j.m.van.dijl01@umcg.nl

Current Topics in Microbiology and Immunology (2017) 404:69–94
DOI 10.1007/82_2016_7

Published Online: 29 April 2016

Contents

1 The Twin-Arginine Protein Translocation Pathway

The movement of substances across biological membranes is essential for the growth, replication and survival of all living cells. In order to translocate substances through these phospholipid bilayers a variety of diverse import and export systems have evolved. One class of compounds that need to be translocated across membranes are proteins. However, transporting proteins over a membrane is not enough to guarantee their proper function, as proteins require correct folding and often co-factors for full functionality. Consequently, there is an intimate relationship between the protein translocation process and protein folding. Protein translocation pathways ensure that substrates are folded either post-translocationally or pre-translocationally. In the case of the Sec pathway for protein secretion, the translocated proteins are folded after translocation (Tjalsma et al. 2000). In contrast, twin-arginine translocation (Tat) pathways are known specifically for the ability to move pre-folded and co-factor-containing proteins across membranes. Translocation of large globular and tightly folded proteins across a membrane is no small feat as the energy needed and the size of the membrane passage required is far greater than that needed for translocating a loosely folded polypeptide chain by the Sec pathway. Exactly how the Tat pathway is able to transport these globular proteins without breaking the cellular barrier or destroying transmembrane ion gradients (e.g. the proton-motive force) is perplexing and of great fundamental scientific interest.

The Tat pathway is evolutionarily conserved in all the kingdoms of life. It is present in 77 % of bacteria, in archaeal species and in the membranes of thylakoids in plants and cyanobacteria (Chaddock et al. 1995; Hutcheon and Bolhuis 2003; Simone et al. 2013). Remnants of the pathway have even been observed in sponges (Pett and Lavrov 2013a). The focus of this chapter is the Gram-positive bacterial

Tat pathway, where the folded proteins are moved from the cytoplasm into the membrane, cell wall or extracellular milieu. Tat systems have been most extensively studied in the Gram-negative bacterium *Escherichia coli*, but also in the Gram-positive *Bacillus* and *Corynebacterium* species, and in pea thylakoids [reviewed in (Palmer and Berks 2012; Goosens et al. 2014b; Patel et al. 2014)].

What has become evident from comparing the various Tat systems is that they are broadly conserved with a high degree of similarity between proteins and mechanisms. For this reason, this review builds on and refers to observations made in Tat systems from other species in addition to Gram-positive bacteria.

2 Tat-Dependent Cargo

The number of Tat-dependent cargo proteins ranges from over 100 in *Streptomyces* species, to only a few in *B. subtilis* and *Staphylococcus aureus* and none in, for example, *Lactobacillus* species (Schaerlaekens et al. 2001, 2004b; Widdick et al. 2006; Joshi et al. 2010; Yamada et al. 2007; Biswas et al. 2009; Goosens et al. 2013; Bolotin et al. 2001; Kleerebezem et al. 2003). These Tat-dependent cargo proteins include secreted proteins, lipoproteins, cell wall-associated proteins and proteins that form components of larger extracytoplasmic complexes on the membrane surface (Widdick et al. 2011; Keller et al. 2012; Monteferrante et al. 2012b; Goosens et al. 2013; Miethke et al. 2013).

Cargo may be destined for the Tat pathway for numerous reasons. Many Tat-dependent substrates are known to require complex co-factors for activity and these are incorporated into the protein in the cytoplasm prior to membrane translocation. Certain other proteins that bind divalent metal ions with affinity ranges lower down in the Irving Williams series, such as Mn, may use the Tat pathway to avoid competing ions with higher binding affinities, such as Zn (Tottey et al. 2008; Monteferrante et al. 2012b). Extremophiles and archaea may need to fold the proteins prior to secretion due to the harsh external milieus in which they live (Bolhuis 2002; Rose et al. 2002). Further, some Tat-destined proteins form multi-protein complexes that are translocated in a hitchhiker or piggyback manner (Rodrigue et al. 1999; Friedrich et al. 2000; Wu et al. 2000a).

Tat substrates have been implicated in a wide range of cellular functions and in the case of pathogenic bacteria they have been associated with virulence, antibiotic resistance and antibacterial compounds (McDonough et al. 2005; De Buck et al. 2008; Weatherspoon-Griffin et al. 2011). Notably, certain industrially relevant proteins are difficult to produce due to co-factor—or disulphide-bond requirements and in some organisms, such as *E. coli* and *Corynebacterium glutamicum*, the Tat pathway has been successfully used for export of these types of proteins, including the alkaline phosphatase PhoA, carbohydrate oxidase, antibody fragments and human tissue plasminogen activator (DeLisa et al. 2003; Kim et al. 2005; Bruser 2007; Ribnicky et al. 2007; Panahandeh et al. 2008; Maurer et al. 2009; Matos et al. 2013; Scheele et al. 2013). Also, the Tat system of Gram-positive bacteria has been

used to secrete small enzymes (De Keersmaeker et al. 2006; Kikuchi et al. 2006, 2008; Scheele et al. 2013). However, biotechnological applications have not yet taken full advantage of the potential of Gram-positive bacterial Tat systems. Although the ability to secrete complex cargo directly into the fermentation broth is enticing, production has been hampered by low yields possibly due to yet unidentified quality control requirements.

In addition to a folded state, Tat cargo proteins also have a unique 'twin-arginine' signal peptide signature. In principle, the N-terminal signal peptides of Tat substrates have a similar tripartite structure as the N-terminal signal peptides of Sec substrates; they are made up of a polar N-region, a hydrophobic H-region in the middle and a polar C-region next to the signal peptidase cleavage site (Tjalsma et al. 2004). However, a distinguishing feature of Tat signal peptides is the presence of two-arginine residues in the N-terminal region that form part of the consensus motif SRRxFLK where x is a polar amino acid (Berks et al. 2000; Stanley et al. 2000; Tjalsma et al. 2004). Compared to Sec signal peptides, the Tat signal peptides also tend to be longer, their N-terminal region is more positively charged (Tjalsma et al. 2000), and their H-region is slightly less hydrophobic (Cristobal et al. 1999). While the twin-arginine residues are conserved, mutation studies have shown that changes in the motif result in a variation of phenotypes, ranging from completely blocked to slowed down translocation of the cargo (Chaddock et al. 1995; DeLisa et al. 2002; Mendel et al. 2008).

Several Tat prediction software programs are available, including TatFind and PRED-TAT (Rose et al. 2002; Bendtsen et al. 2005; Bagos et al. 2010). However, although the signal peptide region is important for translocation, amino acid sequence motifs and patterns are not always reliable predictors, especially since Tat cargo has also been associated with piggyback or hitchhiker mechanisms in organisms like *E. coli*. Here, proteins without a signal peptide of their own bind to the Tat substrate possessing the Tat signal peptide and are exported as a complex by the machinery (Rodrigue et al. 1999; Wu et al. 2000a). Further, Sec–Tat substrate overlap has been shown to occur and sequence ambiguity can lead to false-positive identifications (Tjalsma et al. 2000; Jongbloed et al. 2002; Kouwen et al. 2009; Keller et al. 2012; Goosens et al. 2013). Therefore, although bioinformatic tools are invaluable for lead finding, potential Tat substrates need to be confirmed experimentally.

3 Tat Components and Processes

Genes for the Tat system are observed in 77 % of sequenced microbial genomes. The respective organisms typically contain a TatA and TatC pair (Fig. 1), which are often encoded by genes found within a single operon (Wu et al. 2000b; Yen et al. 2002; Simone et al. 2013). Such *tatA-tatC* operons are occasionally located in the vicinity of genes for Tat cargo proteins (Jongbloed et al. 2000, 2004; Biswas et al. 2009).

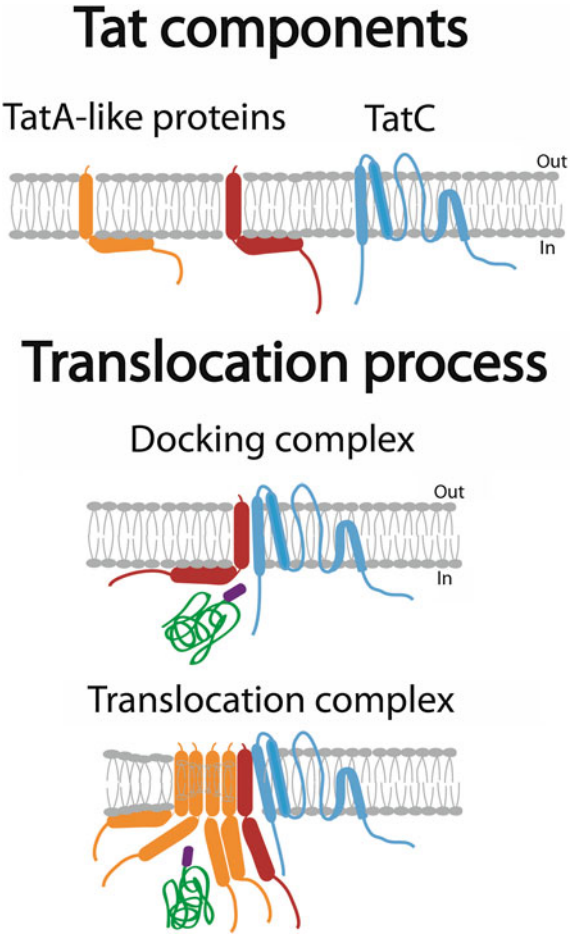


Fig. 1 Tat components and interactions of Tat complexes with their cargo. Tat protein translocases are essentially composed of two types of subunits, namely TatA-like proteins and TatC, which have distinct membrane topologies (*upper panel*). TatA-like proteins (indicated in *orange* and *red*) have an N_{out}–C_{in} topology and consist of two helical domains, one of which spans the membrane while the other one (amphipathic) is exposed to the cytoplasm. TatC is an integral membrane protein with six transmembrane domains and an N_{in}–C_{in} topology. The translocation process is believed to involve two major Tat complexes, namely a docking complex (*middle panel*) and a translocation complex (*lower panel*). The docking complex is composed of TatC and a TatA-like protein (*red*). In some organisms, such as *E. coli*, the latter TatA-like protein has a specialized docking function in which case it is referred to as TatB. Docking of cargo proteins (*green*) involves interaction of the twin-arginine signal peptide (*purple*) with TatC. Once the cargo protein has docked, a large number of TatA-like proteins (*orange*) are recruited to the translocation site, thereby forming the translocation complex

3.1 TatA, TatA-like and TatC Proteins

TatA and TatA-like proteins are small membrane proteins with an N-out C-in topology. They have a small N-terminal extracytoplasmic domain, a single transmembrane helix and an amphipathic helix that lies on the membrane surface or is partially embedded in the membrane on the cytoplasmic side (Fig. 1) (Lange et al. 2007; Hu et al. 2010; Walther et al. 2010). In numerous species, *tatA* genes have undergone multiple duplication events forming the *tatA*-like genes, and these are found in the main *tat* operon or elsewhere in the genome (Wu et al. 2000b; Yen et al. 2002). Some TatA-like proteins can simply be duplicate TatA proteins, such as TatE or TatA2 in *Corynebacterium* species, *Streptomyces coelicolor*, *Salmonella enterica* and *E. coli* (Sargent et al. 1999; Ikeda and Nakagawa 2003; Kalinowski et al. 2003; Nishio et al. 2003; Baglieri et al. 2012). In *B. subtilis* and *C. glutamicum*, the expression of the duplicate proteins TatAc and TatE, respectively, has been shown to assist the activity of the primary TatA protein (Kikuchi et al. 2006; Goosens et al. 2015). However, in some instances, duplicate TatA-like proteins (often referred to as TatB) have further undergone sequence divergence and functional specialization. This seems to have occurred independently numerous times. TatA-like proteins with divergent functions, such as TatB have been studied in *E. coli*, *Streptomyces* species and in the thylakoids of plant chloroplasts (where they are referred to as Hcf106). In these organisms, both TatA and TatB are needed for full translocation activity (Sargent et al. 1999; Mori and Cline 2001; Schaerlaekens et al. 2001; De Keersmaeker et al. 2005a). Intriguingly, certain TatA proteins (e.g. the TatA proteins from *B. subtilis*) are able to functionally replace both *E. coli* TatA and TatB in the TatA-B-C system (Barnett et al. 2008). This is even more remarkable since TatA shares only 20 % sequence similarity with TatB although the *E. coli* TatA and TatB are structurally the same (Hicks et al. 2003; Lange et al. 2007; Hu et al. 2010; Walther et al. 2010). Nevertheless, minor changes to *E. coli* TatA allow it to complement for TatB (Blaudeck et al. 2005; Barrett et al. 2007). Therefore, no clear definition exists that allows one to properly differentiate TatA from TatB, and in many sequence annotations where a second TatA-like protein has been defined as TatB, this annotation may be erroneous. Importantly, over 50 % of the Tat-encoding genomes sequenced to date specify only a TatA-TatC component (Simone et al. 2013). Accordingly, the concept that a core Tat system composed of a core TatA-TatC pair with various assistant TatA-like proteins is gaining more support. This view is backed by studies in *B. subtilis,* where a third TatA protein (TatAc) was shown to assist TatAy in protein translocation, and in *Helicobacter pylori, Campylobacter jejuni,* and *C. glutamicum* where a second TatA-like protein was shown to be essential only under some conditions of Tat-dependent protein translocation but not all (Kikuchi et al. 2006; Benoit and Maier 2014; Liu et al. 2014; Goosens et al. 2015; Oertel et al. 2015). Thus, in many organisms, it is not necessary to have a TatB as is found in *E. coli* and thylakoids, which have TatA, TatB and TatC subunits, each with their own function.

Unlike TatA, TatC proteins are large integral membrane proteins with six membrane-spanning domains (Fig. 1). TatC is central to the Tat pathway as is evident by the multiple protein interactions in which this protein is involved. The TatC transmembrane regions interact with other TatC proteins (Buchanan et al. 2002; Punginelli et al. 2007), but also with the cargo protein (Behrendt et al. 2004; Frobel et al. 2012). Both the cargo protein and TatA(-like) proteins (e.g. *E. coli* TatB) have been shown to interact with the membrane-embedded region of TatC and its conserved cytoplasmic loop (Buchanan et al. 2002; McDevitt et al. 2006; Schreiber et al. 2006; Holzapfel et al. 2007; Strauch and Georgiou 2007; Frobel et al. 2011; Zoufaly et al. 2012; Ma and Cline 2013; Blümmel et al. 2015). Although the extracytoplasmic loops do not share high sequence similarity, random mutagenesis studies showed that the conserved secondary structure is vital (Strauch and Georgiou 2007; Kneuper et al. 2012). Further, the C-terminal tail of TatC has been shown to be essential for successful Tat-dependent protein translocation in *B. subtilis* (Eijlander et al. 2009b).

3.2 Translocation Process

The mechanism of translocating cargo proteins via the Tat pathway has not been concretely defined. However, what is currently agreed upon is that there are at least two Tat complexes and at least three steps to the process (Cline and Mori 2001). The first step, the formation of a docking complex, is initiated when cargo proteins interact with a TatC and TatA(-like) complex at the membrane (Fig. 1) (Bolhuis et al. 2001; Whitaker et al. 2012). In *E. coli* and thylakoids, this is where the (TatA-like) TatB proteins perform their specialized functions. The signal peptide of the cargo protein directly interacts with TatC in the docking complex and is inserted into the membrane after proofreading (Cline and Mori 2001; Alami et al. 2003; Papish et al. 2003; Robinson et al. 2011; Frobel et al. 2012). In the second step, the translocation complex is formed. This occurs once the cargo protein has 'docked', and a large number of TatA proteins are recruited to the translocation site in a manner that is dependent on the proton-motive force, thereby forming the translocation complex (Fig. 1) (Mori and Cline 2002; Alami et al. 2003; Dabney-Smith et al. 2006). The final step is the translocation itself, but how this occurs is not clear. A recent study by Blümmel et al. (2015) indicates that TatBC oligomers can assemble into closed intramembrane substrate-binding cavities, where TatB monomers would form dome-like structures that are surrounded by TatC monomers. These TatBC complexes would bind the N-termini of TatA protomers facilitating contacts with TatB and membrane-inserted cargo proteins.

There are two popular translocation models, which are both speculative: the pore and the membrane-weakening models [reviewed in (Berks 2015; Patel et al. 2014)]. Both models are supported by data and, depending on their interpretation, some results are used to reinforce either. The pore model was conceived based upon single particle electron microscopy studies that showed TatA and TatA-like proteins

self-assemble to form cup-like structures or pores with varying diameters (Gohlke et al. 2005; Oates et al. 2005; Beck et al. 2013). Consistent with this model, the *E. coli* TatA complexes have been shown to form ladders of multiple sizes in native gels (Gohlke et al. 2005; Oates et al. 2005; Beck et al. 2013) giving rise to the hypothesis that a pore made up of TatA proteins adapts its diameter to the globular cargo by varying the amount of TatA components (Fig. 1). The theory goes that TatA-cargo protein interactions within the cup-like TatA structure allow for translocation by the folding-in (trap-door mechanism) or twisting (iris mechanism) of the amphipathic helix of TatA up into the membrane (Berks et al. 2000; Gouffi et al. 2004; Gohlke et al. 2005; Walther et al. 2013). In contrast, the membrane-weakening model predicts that the TatA complexes observed by electron microscopy do not form a pore, but that the aggregates of TatA proteins form destabilised membrane regions that permit cargo passage (Bruser and Sanders 2003). Data that support this include the length of the TatA transmembrane region, which is too short to span the lipid bilayer (Rodriguez et al. 2013). Also, the *B. subtilis* TatAd complexes observed by single particle electron microscopy were structurally too small to represent pores that can accommodate a substrate (Beck et al. 2013). Other evidence not consistent with the pore model is that the large size variation and laddering effect seen in *E. coli* TatA complexes have not been convincingly observed for other TatA and TatA-like proteins (Baglieri et al. 2012; Monteferrante et al. 2012a; Walther et al. 2013). Moreover, NMR studies suggest that, because the TatA amphipathic helix is not flexible (Walther et al. 2010), the movement of the amphipathic helix into the membrane would have to be sudden and, most likely, disruptive. Another piece of evidence that seems to support the membrane-weakening model is the involvement of the phage shock protein PspA in Tat-dependent protein transport. For example, PspA has been implicated in the stabilization of the membrane under stress conditions (Darwin 2005; Vrancken et al. 2008) and in suppressing proton leakage (Kobayashi et al. 2007). PspA binds both *E. coli* TatA (Mehner et al. 2012) and phospholipids (Kobayashi et al. 2007) forming scaffold-like structures in the membrane (Standar et al. 2008). The possible involvement of PspA in Tat-dependent protein transport suggests that the translocation event induces stress. Importantly, expression of PspA has been shown to improve Tat-dependent protein secretion in both *S. lividans* and *E. coli* and, hence, its role in suppressing proton leakage that may occur in the Tat export process may be conserved in both Gram-positive and Gram-negative bacteria (DeLisa et al. 2004; Vrancken et al. 2007).

4 Cargo protein Processing and Quality Control

A defining feature of Tat is its inability to translocate incorrectly folded proteins. Although a small amount of flexibility has been described for small synthetic peptides (Hynds et al. 1998; Richter et al. 2007; Rocco et al. 2012), the system is known to have strict folding requirements regarding its native substrates. If a

protein is not sufficiently folded or does not have its co-factors inserted, the translocation is prevented and the protein is degraded (Jack et al. 2004; Kolkman et al. 2008; Matos et al. 2008). Protein folding, co-factor insertion and quality control prior to Tat complex interactions are therefore considered important for cargo translocation. The quality control step has been shown to occur at the docking complex (Buchanan et al. 2008; Panahandeh et al. 2008; Frobel et al. 2012; Rocco et al. 2012). Further evidence of quality control prior to docking complex formation has been clearly described in *E. coli*, where a number of substrate-specific chaperones have been identified (Oresnik et al. 2001; Jack et al. 2004). However, homologous chaperones have not been characterized in other organisms, and it remains unclear which factors may be involved in pre-translocational protein folding and quality control prior to docking-complex interactions in Gram-positive bacteria. Nonetheless, in *B. subtilis* the Tat-dependent QcrA protein was shown to undergo quality control at two subcellular locations, in the cytoplasm and membrane, and on two different pre-QcrA intermediates. While neither of the pre-QcrA proteins were translocated, pre-QcrA quality control occurred both at the membrane, where the Tat-docking complex is shown to perform proofreading functions, and in the cytoplasm via an as-yet-unknown mechanism (Goosens et al. 2014a). Quantitative proteomic studies have implicated the membrane-targeting chaperone DnaJ and co-factor assembly protein SufS with the Tat pathway in *B. subtilis* (Albrecht et al. 2011; Goosens et al. 2013; Castanie-Cornet et al. 2014). However, functional studies are still required to confirm these links. What has clearly been shown is a direct interaction between the *B. subtilis* TatAd protein and the soluble chemoreceptor HemAT, and between TatAd and the putative pentose transporter CsbC within the membrane. Not only do HemAT and CsbC individually interact with TatAd, but they are also essential for the secretion of the TatAd-specific cargo protein PhoD (Monteferrante et al. 2013). Exactly what the roles of these proteins in the PhoD quality control and Tat-dependent export pathway are remained unclear.

It has been suggested that the Tat-associated quality control is linked to a pool of cytoplasmic TatA (Pop et al. 2003; De Keersmaeker et al. 2005a, b; Schreiber et al. 2006; Westermann et al. 2006; De Keersmaeker et al. 2007; Frielingsdorf et al. 2008). In this model, the cytoplasmic TatA of *B. subtilis*, *Streptomyces lividans* or thylakoids interacts with cargo prior to translocation and guides it to the docking complex in the membrane. Also, overexpressed TatA molecules have been observed to form distinct tubes in the cytoplasm (Berthelmann et al. 2008). However, since the cytoplasmic TatA-cargo protein interaction has only been observed under induced circumstances and the presence of TatA in the cytoplasm has not been shown consistently under all experimental methodologies, future studies will need to verify the possible quality control function of cytoplasmic TatA (Wexler et al. 2000; Barnett et al. 2008; Leake et al. 2008; Barnett et al. 2009; Ridder et al. 2009).

Other steps in the quality control of Tat-dependent cargo proteins occur during or shortly after membrane translocation. In particular, these include the removal of

the twin-arginine signal peptide by signal peptidases, which has been widely observed (Jongbloed et al. 2004; Luke et al. 2009; Widdick et al. 2011; Goosens et al. 2013). In addition to signal peptidases, extracytoplasmic proteases have also been shown to effect the Tat-dependent cargo protein EfeB in *B. subtilis.* EfeB directly interacts with and requires the cell wall-bound protease WprA for processing (Monteferrante et al. 2013), but is degraded by extracellular proteases in the growth medium (Krishnappa et al. 2012). EfeB forms part of a membrane-bound complex with the EfeU and EfeO proteins, and the association with extracellular proteases is an indication of a possible assembly proofreading mechanism (Monteferrante et al. 2013).

5 Flexibility of the Tat System Between Different Organisms

Expression of *B. subtilis* Tat components in *E. coli* leads to Tat-dependent export and functionally replaces the Tat pathway in *E. coli* (Barnett et al. 2008, 2009; Monteferrante et al. 2012a; van der Ploeg et al. 2012). Single *B. subtilis tatA* genes are able to functionally replace both the *E. coli* TatA and TatB proteins (Barnett et al. 2008; Monteferrante et al. 2012a; Beck et al. 2013). However, when similar interspecies experiments were performed in a *B. subtilis* background, complementation was not that simple. Although the Tat systems from *Bacillus cereus* and *Listeria monocytogenes* were functional in *B. subtilis,* the Tat system from *S. aureus* was barely active in *B. subtilis* (Barnett et al. 2008, 2009; van der Ploeg et al. 2011a, 2012). These differences suggest that the Tat pathway alone is not enough to ensure complete translocation and possible chaperone or quality control mechanisms in *E. coli* and *S. aureus* do not match up with those in *B. subtilis*.

Interspecies variations have also been observed with regard to the export of cargo proteins. The addition of a Tat signal peptide to a cargo protein has allowed for heterologous Tat-dependent translocation in many cases, but this does not always equally prove successful in all genetic backgrounds and with all cargo (Thiemann et al. 2006; Meissner et al. 2007; Kikuchi et al. 2008; Widdick et al. 2008; Scheele et al. 2013). Environmental salt (i.e. NaCl) conditions also affected the Tat-dependent export of cargo that was heterologously expressed in *B. subtilis* suggesting external conditions and media may affect Tat-dependent secretion (van der Ploeg et al. 2011a, 2012). However, the environmental salt concentration did not significantly affect the amount of Tat-dependently translocated QcrA in *B. subtilis* (Goosens et al. 2015). The influence of salt on the translocation of other cargo is therefore not necessarily an intrinsic Tat effect.

6 Monoderm Gram-Positive Bacterial Tat Systems

Bacterial phyla are broadly defined by the physical properties of the outer layer of their cell structure. In most cases, bacteria are classified by the outcome of so-called Gram staining. The Gram staining procedure was developed in the late nineteenth century and works by interaction of the stain with the peptidoglycan of the cell wall. The stain is either retained by the peptidoglycan, giving cells a purple colour, or washed out. Accordingly, this gave rise to the common nomenclature of Gram-positive bacteria where the stain is retained, or Gram-negative bacteria where the stain is not retained. The Gram-positive bacteria have a single plasma membrane surrounded by a thick outer cell wall composed of peptidoglycan (i.e. a monoderm cell envelope). In contrast, Gram-negative bacteria have a double membrane with a peptidoglycan layer in between (i.e. a diderm cell envelope). Although the Gram-staining-based nomenclature is generally a good indicator of the physical properties of the outer layer of cells, it can be ambiguous. Some bacteria stain positive, but do in fact have a diderm cell envelope structure. Such bacteria include mycobacteria, corynebacteria, rhodococci and nocardiae. The Tat systems of these diderm Gram-positive bacteria will not be detailed here, as they have been reviewed previously (Goosens et al. 2014b) and not much new information has become available since this review was published.

6.1 Bacillus subtilis

Bacillus subtilis is the major Gram-positive model organism with an extensive array of genetic tools, including in-depth genomic, transcriptomic and proteomic insights (Kunst et al. 1997; Tjalsma et al. 2000; Eymann et al. 2004; Wolff et al. 2007; Buescher et al. 2012; Nicolas et al. 2012). A number of *Bacillus* species are biotechnologically relevant. *B. subtilis, Bacillus licheniformis* and *Bacillus amyloliquefaciens*, for example, have the ability to secrete large titres of proteins, qualify for the Qualified Presumption of Safety (QPS) status of the European Food Safety Authority, and many of their products have a Generally Recognized As Safe (GRAS) status from the US Food and Drug Administration. Furthermore, *B. subtilis* becomes naturally competent, thereby allowing for easy genetic modification (Tjalsma et al. 2000; van Dijl et al. 2002; Graumann 2011).

The *B. subtilis* Tat system is one of the most extensively studied Tat systems within the field, because it has unique characteristics in particular relating to gene duplication. As indicated above, duplication of TatA and TatA-like proteins is a common feature in most Tat systems. However, the duplication of TatC proteins is rare except in *Bacillus* species, where multiple isoforms of TatC have been observed (Jongbloed et al. 2000; Yen et al. 2002; Simone et al. 2013).

The core progenitor operon in *B. subtilis* is *tatAy-tatCy* (Simone et al. 2013). This operon has been duplicated and, consequently, there is a second separate *tat*

operon named *tatAd-tatCd.* Thus, *B. subtilis* Tat is specified by two separate operons, which are expressed at different times and, under normal conditions, do not share substrate specificity (Jongbloed et al. 2004; Eijlander et al. 2009a; Nicolas et al. 2012; Goosens et al. 2013). The predominant Tat pathway is TatAy-TatCy and, although an early bioinformatics analysis has predicted up to 69 potential Tat-dependent substrates, only three substrates have been confirmed to be strictly TatAy-TatCy-dependent, namely EfeB, QcrA and YkuE (Tjalsma et al. 2000; Jongbloed et al. 2002, 2004; Monteferrante et al. 2012b; Goosens et al. 2013). The genes for the second *B. subtilis* Tat pathway, TatAd-TatCd, are found next to the gene for its only known substrate, the phosphate acquisition protein PhoD. Accordingly, the *tatAd-tatCd* operon is only expressed under low-phosphate conditions (Eder et al. 1996; Jongbloed et al. 2000, 2004; Nicolas et al. 2012).

Apart from duplicating the whole *tatAy-tatCy* operon, the Tat system of *B. subtilis* has a further *tatA* duplication, namely *tatAc*. This third *tatA* gene is located elsewhere on the chromosome, and it is expressed constitutively under numerous conditions (Nicolas et al. 2012). Although it was investigated in several studies, a physiological role for TatAc has, until recently, remained enigmatic (Tjalsma et al. 2000; Jongbloed et al. 2002, 2004; Eijlander et al. 2009a; Nicolas et al. 2012). TatAc is unable to form an active translocon when paired with TatC proteins in *B. subtilis* (Eijlander et al. 2009a; Goosens et al. 2015). However, when expressed in *E. coli,* TatAc formed functional translocases with either *E. coli* TatBC, *B. subtilis* TatCd or *B. subtilis* TatCy (Monteferrante et al. 2012a; Beck et al. 2013). This difference illustrates the interpathway flexibility of *E. coli* and further suggests potentially different quality control or chaperone activities in the Tat pathways of *B. subtilis* and *E. coli*. Yeast two-hybrid (Y2H) protein–protein interaction studies have shown that not only does TatAc interact with itself and the *B. subtilis* TatA proteins, but it also directly interacts with HemAT (Monteferrante et al. 2013). HemAT was in turn shown to be essential for the Tat-dependent secretion of PhoD, which therefore suggested a functional role for TatAc in *B. subtilis* (Monteferrante et al. 2013). A functional role for TatAc as a Tat-assistance protein was confirmed when it was shown to permit the translocation of EfeB in cells with an impaired TatAy function, despite the fact that TatAc was unable to replace TatAy (Goosens et al. 2015). It thus seems that TatAc, the third TatA-like protein of *B. subtilis*, reflects an intermediate evolutionary step in TatA-TatB specialization. In this scenario, the presently available data suggest that the defective TatAy protein has a role that is comparable to that of *E. coli* TatB, while TatAc has a role similar to that of *E. coli* TatA. Altogether, it can be concluded that the core Tat translocon in *B. subtilis* is composed of a TatAy-TatCy pair and that the TatAc protein has a non-essential assistant role in translocation. For example, TatAc could allow for more efficient cargo-Tat protein–protein interactions, and it might improve the overall efficiency of the Tat pathway (Goosens et al. 2015).

All confirmed *B. subtilis* Tat-dependent cargo proteins are known to contain co-factors, thereby emphasizing their need for the Tat pathway (Schneider and Schmidt 2005; Monteferrante et al. 2012b; Miethke et al. 2013; Rodriguez et al. 2014). Further, QcrA contains a disulphide bond in addition to its iron-sulphur

cluster (Iwata et al. 1996; Link et al. 1996; Schmidt and Shaw 2001; Hunsicker-Wang et al. 2003). Of note, QcrA has been observed to be a Tat-dependent substrate in a wide-range of organisms (Molik et al. 2001; Bachmann et al. 2006; De Buck et al. 2007; Goosens et al. 2013; Pett and Lavrov 2013b; Oertel et al. 2015). Both QcrA and EfeB form part of larger extracytoplasmic complexes, where the QcrA-B-C proteins form the cytochrome bc_1 complex, whereas the EfeU-O-B proteins form an iron uptake system. This organization of Tat-dependent substrate proteins into larger complexes further suggests that the Tat pathway assists in protein complex assembly (Yu et al. 1995; Schneider and Schmidt 2005; Miethke et al. 2013; Sousa et al. 2013). The extracellular protease WprA also directly affects EfeB and indirectly influences YkuE (Monteferrante et al. 2013). The action of WprA may thus be associated with this complex maturation.

Most of the observed phenotypes associated with Tat-deficiencies have been linked directly to the known substrates, i.e. PhoD is required under phosphate starvation and EfeB is required under conditions of iron deficiency and low salt (Jongbloed et al. 2000; van der Ploeg et al. 2011b). However, quantitative proteomic studies revealed that numerous proteins associated with motility and biofilm formation were decreased in *tatAy-tatCy* deficient strains, leading to the identification of an, as-yet, not-well-understood Tat-associated delayed biofilm formation phenotype (Goosens et al. 2013). Most studies investigating Tat have used Western blotting techniques to validate Tat-dependency of substrates. Although this remains the golden standard and a powerful tool, it does not give an indication of whether the protein is correctly folded and active. In the *B. subtilis* studies, the activity of cargo proteins was determined using the alkaline phosphatase activity of YkuE (Monteferrante et al. 2012b) and the ferric iron uptake to assess EfeB activity (Miethke et al. 2013; Goosens et al. 2015). EfeB is a hemoprotein that oxidizes ferrous iron to ferric iron for uptake via EfeU and EfeO. For this reason, EfeB stimulates growth under microaerobic conditions where ferrous iron is more abundant. In addition, EfeB was shown to have an important role in the protection against cell envelope stress through the elimination of reactive oxygen species that are generated in the presence of ferrous iron (Miethke et al. 2013).

6.2 *Streptomyces*

Streptomyces species are found naturally in the soil where they often form mycelia. These bacteria have become workhorses for industry as they can be used for the high-level production of various antibiotics and secreted proteins (Anne et al. 2012). The Tat system is a major contributor to overall protein secretion in these species with numbers of potential substrates ranging between 100 and 189 (Widdick et al. 2006; Joshi et al. 2010; Palmer and Berks 2012). The Tat system in *Streptomyces* species is composed of at least three Tat components, where the genes for a minimal TatA-TatC system are found clustered and the gene for an extra

TatA-like protein (depending on the species, these are called TatB or TatA2) is located elsewhere on the chromosome (Schaerlaekens et al. 2001; Palmer and Berks 2012). Crosstalk between the Sec and Tat pathways has been observed in the sense that Sec-dependent secretion was enhanced by a mutated Tat system or reduced by an overexpressed Tat system (Schaerlaekens et al. 2004a; De Keersmaeker et al. 2006).

In *S. lividans,* optimal secretion of the Tat-dependent substrates xylanase C and tyrosinase occurred when all three Tat components were present. Secretion did still occur when a single TatA or TatB protein was paired with TatC, albeit at lowered efficiency. It was therefore concluded that the TatA-like proteins in *Streptomyces* were unable to fully functionally replace each other and that each must have a specialized function (De Keersmaeker et al. 2005a). A very interesting observation made in *S. lividans* was that both TatA and TatB proteins were detected in the cytoplasm under native conditions (De Keersmaeker et al. 2005b), and when expression was induced these TatA-like proteins apparently interacted with Tat-dependent cytoplasmic pre-proteins (De Keersmaeker et al. 2005a, 2007).

6.3 Staphylococcus

Not all *Staphylococcus* species have a Tat pathway. However, this pathway has been identified in *Staphylococcus haemolyticus*, *Staphylococcus carnosus, Staphylococcus lugdunensis* and *S. aureus* (Biswas et al. 2009). The staphylococcal Tat pathway is composed of a single TatA–TatC pair. This Tat pathway has been investigated for biotechnological applications in *S. carnosus* and, although shown to secrete heterologous proteins, it was considered inadequate for the required applications (Thiemann et al. 2006; Meissner et al. 2007). To date, only one native Tat-dependent staphylococcal substrate, FepB, has been confirmed, and inactivation of the Tat system did not show any global changes in protein secretion profiles (Yamada et al. 2007; Biswas et al. 2009). FepB is an iron-dependent peroxidase encoded by the *fepABC* operon, and the corresponding complex is very similar to the iron-scavenging EfeUOB complexes in *E. coli, B. subtilis* and *L. monocytogenes* (Biswas et al. 2009; Miethke et al. 2013; Turlin et al. 2013). Interestingly, in a mouse kidney abscess model, the bacterial load of *tat* or *fepB* mutant strains was shown to be decreased, thereby pointing at a physiologically relevant role of Tat-dependent export of FepB in staphylococcal disease (Biswas et al. 2009).

6.4 Listeria monocytogenes

Listeria monocytogenes is a saprophytic bacterium that, once it has entered the food chain, becomes a dangerous food-borne pathogen. The *Listeria* Tat system is composed of a TatA–TatC pair (Desvaux and Hebraud 2006; Machado et al. 2013).

In strains where *tat* genes were deleted, no significant changes in cell viability or virulence have been described (Machado et al. 2013; Halbedel et al. 2014). Bioinformatic analysis predicted two potential Tat-dependent substrates, namely FepB (Lmo0367) and a FabF-like protein (Lmo2201) (Desvaux and Hebraud 2006). However, when these proteins were tagged and expressed neither was Tat-dependent (Halbedel et al. 2014). Nonetheless, the FepB signal peptide was shown to confer Tat-dependent secretion in the *S. lividans* agarase reporter assay, and both a *tatC* and *fepB* mutant strain displayed decreased overall ferric reductase activity (Widdick et al. 2008; Tiwari et al. 2015). Hence, there is evidence of a FepB-Tat association, but none of the currently available data confirms a direct Tat-dependency. The *fepB* gene is co-transcribed in an iron-induced *fepCAB* operon, which is also Fur-regulated (Ledala et al. 2010; Tiwari et al. 2015). Although the *tat* operon is transcribed in the early exponential phase in rich medium, it is Fur-regulated and highly induced under iron starvation conditions (Ledala et al. 2010; Machado et al. 2013; Tiwari et al. 2015). Therefore, it is conceivable that, in order to detect the possible Tat-dependency of FepB, environmental conditions with low iron availability may be required, or as in *B. subtilis,* other environmental conditions such as low salt (van der Ploeg et al. 2011b; Goosens et al. 2015). In fact, the *Listeria fepCAB* operon is highly reminiscent of the Tat-associated *fepABC* and *efeUOB* operons in *S. aureus* and *B. subtilis*, respectively. Accordingly, there appears to be a conservation of the Tat requirement in these iron-scavenging complexes.

6.5 *Streptococcus*

The majority of *Streptococcus* species studied have no identifiable Tat components. However, genes encoding a TatA and TatC protein have been identified in *Streptococcus sanguinis* and *Streptococcus thermophilus*. Intriguingly, in both *S. sanguinis* and *S. thermophilus* these *tat* genes are localized in close genomic proximity to three genes that resemble the *efeUOB*, *fepCAB* and *fepABC* operons of *B. subtilis*, *Listeria* and *Staphylococcus*, respectively. Further, in both *Streptococcus* species, the EfeB-like iron-dependent peroxidase contains a twin-arginine motif in the signal peptide. In the facultative anaerobe *S. thermophilus*, EfeB was shown to be translocated by the Tat system, and mutation of *efeB* or *tatC* resulted in decreased growth under aerobic conditions, suggesting that the respective proteins have a role in protecting the cell against oxidative stress (Xu et al. 2007; Zhang et al. 2015).

7 Conclusion

The investigations on the Tat pathways of Gram-positive bacteria, as described in this review, suggest a general association between the Tat pathway and iron-scavenging complexes, phosphate acquisition and respiratory complexes. Intriguingly, phylogenetic analyses of TatC showed that 89 % of species that have TatC are either facultative aerobes, or facultative or obligate anaerobes, while only 11 % are obligate aerobes (Simone et al. 2013). Thus, the majority of organisms with a Tat system find themselves in anaerobic environments. Of note, anaerobic bacteria and anaerobic growth are, over all, relatively poorly characterized. It thus seems likely that the full potential of the Tat pathway and the spectrum of biological functions that it fulfils are currently substantially underappreciated. With this in mind, an important challenge for future Tat-related research could be the exploration of this pathway in the microbiota of the human gut. Here bacteria, many of which are *Firmicutes*, need to thrive and survive in a challenging anaerobic environment that continuously changes depending on the ingestion of different nutrients by the host, continuous flow-through and influx of oxygen from the gut epithelium (Khan et al. 2012). Indeed, sequence analyses have shown that various dominant gut microbes do contain *tat* genes, and it would be interesting to explore their functions and find out whether they are conditionally essential.

The Tat system is known to be essential in only a few bacteria (Palmer and Berks 2012), including *Mycobacterium tuberculosis* where Tat has been shown to be important for drug resistance and virulence (Raynaud et al. 2002; McDonough et al. 2005). Yet, gene essentiality is often condition-dependent and this also applies to some *tat* genes as exemplified in *B. subtilis*, where the absence of the *tatAy-tatCy* operon leads to severe growth impairment in salt- or iron-depleted environments. This conditional essentiality implies that Tat is a potentially druggable target in notoriously drug-resistant pathogens, such as *M. tuberculosis*.

Today, there are various areas in the Tat field that merit further research, some of which have been touched upon in the present review. For example, this applies to the condition-dependent regulation of *tat* gene expression, including possible roles of antisense RNAs and small non-coding regulatory RNAs. In this respect, it is worth mentioning recent studies, showing that the non-coding RNA Mcr7 of *M. tuberculosis* modulates TatC expression, thereby serving as an intriguing new Tat secretion control mechanism (Solans et al. 2014). Other major knowledge gaps concerning the Tat pathways of Gram-positive bacteria relate to the chaperones that guide cargo folding and cofactor insertion, quality control of cargo prior to membrane translocation, the actual mechanism of Tat-mediated translocation of cargo across the membrane, and post-translocational cargo processing and quality control. Research into these areas will be important, not only to enrich our fundamental understanding of protein translocation mechanisms, but also to open up the enigmatic Tat pathway for the biotechnological secretion of high-value proteins, and to explore the potential of Tat as an antimicrobial drug target.

Acknowledgements VJG and JMvD were in parts supported by the CEU projects PITN-GA-2008-215524 (TranSys), LSHG-CT-2006-037469 and 244093, and the transnational SysMO initiative through projects BACELL SysMO1 and 2 with funding from the Research Council for Earth and Life Sciences of the Netherlands Organization for Scientific Research (NWO-ALW).

References

Alami M, Luke I, Deitermann S, Eisner G, Koch HG, Brunner J, Muller M (2003) Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. Mol Cell 12(4):937–946

Albrecht AG, Peuckert F, Landmann H, Miethke M, Seubert A, Marahiel MA (2011) Mechanistic characterization of sulfur transfer from cysteine desulfurase SufS to the iron-sulfur scaffold SufU in *Bacillus subtilis*. FEBS Lett 585(3):465–470

Anne J, Maldonado B, Van Impe J, Van Mellaert L, Bernaerts K (2012) Recombinant protein production and streptomycetes. J Biotechnol 158(4):159–167

Bachmann J, Bauer B, Zwicker K, Ludwig B, Anderka O (2006) The Rieske protein from Paracoccus denitrificans is inserted into the cytoplasmic membrane by the twin-arginine translocase. FEBS J 273(21):4817–4830

Baglieri J, Beck D, Vasisht N, Smith CJ, Robinson C (2012) Structure of TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter. J Biol Chem 287(10):7335–7344

Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD (2010) Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 26(22):2811–2817

Barnett JP, Eijlander RT, Kuipers OP, Robinson C (2008) A minimal Tat system from a gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes. J Biol Chem 283(5):2534–2542

Barnett JP, van der Ploeg R, Eijlander RT, Nenninger A, Mendel S, Rozeboom R, Kuipers OP, van Dijl JM, Robinson C (2009) The twin-arginine translocation (Tat) systems from *Bacillus subtilis* display a conserved mode of complex organization and similar substrate recognition requirements. FEBS J 276(1):232–243

Barrett CM, Freudl R, Robinson C (2007) Twin arginine translocation (Tat)-dependent export in the apparent absence of TatABC or TatA complexes using modified *Escherichia coli* TatA subunits that substitute for TatB. J Biol Chem 282(50):36206–36213

Beck D, Vasisht N, Baglieri J, Monteferrante CG, van Dijl JM, Robinson C, Smith CJ (2013) Ultrastructural characterisation of *Bacillus subtilis* TatA complexes suggests they are too small to form homooligomeric translocation pores. Biochim Biophys Acta 1833:1811–1819

Behrendt J, Standar K, Lindenstrauss U, Bruser T (2004) Topological studies on the twin-arginine translocase component TatC. FEMS Microbiol Lett 234(2):303–308

Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. BMC Bioinformatics 6:167

Benoit SL, Maier RJ (2014) Twin-arginine translocation system in Helicobacter pylori: TatC, but not TatB, is essential for viability. mBio 5(1):e01016–01013

Berks BC (2015) The twin-arginine protein translocation pathway. Annu Rev Biochem 84: 843–864

Berks BC, Sargent F, Palmer T (2000) The Tat protein export pathway. Mol Microbiol 35(2): 260–274

Berthelmann F, Mehner D, Richter S, Lindenstrauss U, Lunsdorf H, Hause G, Bruser T (2008) Recombinant expression of tatABC and tatAC results in the formation of interacting cytoplasmic TatA tubes in *Escherichia coli*. J Biol Chem 283(37):25281–25289

Biswas L, Biswas R, Nerz C, Ohlsen K, Schlag M, Schafer T, Lamkemeyer T, Ziebandt AK, Hantke K, Rosenstein R, Gotz F (2009) Role of the twin-arginine translocation pathway in Staphylococcus. J Bacteriol 191(19):5921–5929

Blaudeck N, Kreutzenbeck P, Muller M, Sprenger GA, Freudl R (2005) Isolation and characterization of bifunctional *Escherichia coli* TatA mutant proteins that allow efficient tat-dependent protein translocation in the absence of TatB. J Biol Chem 280(5):3426–3432

Blümmel AS, Haag LA, Eimer E, Müller M, Fröbel J (2015) Initial assembly steps of a translocase for folded proteins. Nat Commun 6:7234

Bolhuis A (2002) Protein transport in the halophilic archaeon Halobacterium sp. NRC-1: a major role for the twin-arginine translocation pathway? Microbiology 148(Pt 11):3335–3346

Bolhuis A, Mathers JE, Thomas JD, Barrett CM, Robinson C (2001) TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. J Biol Chem 276(23):20213–20219

Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A (2001) The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. Genome Res 11(5):731–753

Bruser T (2007) The twin-arginine translocation system and its capability for protein secretion in biotechnological protein production. Appl Microbiol Biotechnol 76(1):35–45

Bruser T, Sanders C (2003) An alternative model of the twin arginine translocation system. Microbiol Res 158(1):7–17

Buchanan G, de Leeuw E, Stanley NR, Wexler M, Berks BC, Sargent F, Palmer T (2002) Functional complexity of the twin-arginine translocase TatC component revealed by site-directed mutagenesis. Mol Microbiol 43(6):1457–1470

Buchanan G, Maillard J, Nabuurs SB, Richardson DJ, Palmer T, Sargent F (2008) Features of a twin-arginine signal peptide required for recognition by a Tat proofreading chaperone. FEBS Lett 582(29):3979–3984

Buescher JM, Liebermeister W, Jules M, Uhr M, Muntel J, Botella E, Hessling B, Kleijn RJ, Le Chat L, Lecointe F, Mader U, Nicolas P, Piersma S, Rugheimer F, Becher D, Bessieres P, Bidnenko E, Denham EL, Dervyn E, Devine KM, Doherty G, Drulhe S, Felicori L, Fogg MJ, Goelzer A, Hansen A, Harwood CR, Hecker M, Hubner S, Hultschig C, Jarmer H, Klipp E, Leduc A, Lewis P, Molina F, Noirot P, Peres S, Pigeonneau N, Pohl S, Rasmussen S, Rinn B, Schaffer M, Schnidder J, Schwikowski B, Van Dijl JM, Veiga P, Walsh S, Wilkinson AJ, Stelling J, Aymerich S, Sauer U (2012) Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. Science 335(6072):1099–1103

Castanie-Cornet MP, Bruel N, Genevaux P (2014) Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. Biochim Biophys Acta 1843(8):1442–1456

Chaddock AM, Mant A, Karnauchov I, Brink S, Herrmann RG, Klosgen RB, Robinson C (1995) A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. EMBO J 14(12):2715–2722

Cline K, Mori H (2001) Thylakoid DeltapH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. J Cell Biol 154(4):719–729

Cristobal S, de Gier JW, Nielsen H, von Heijne G (1999) Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. EMBO J 18(11):2982–2990

Dabney-Smith C, Mori H, Cline K (2006) Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport. J Biol Chem 281(9):5476–5483

Darwin AJ (2005) The phage-shock-protein response. Mol Microbiol 57(3):621–628

De Buck E, Vranckx L, Meyen E, Maes L, Vandersmissen L, Anne J, Lammertyn E (2007) The twin-arginine translocation pathway is necessary for correct membrane insertion of the Rieske Fe/S protein in Legionella pneumophila. FEBS Lett 581(2):259–264

De Buck E, Lammertyn E, Anne J (2008) The importance of the twin-arginine translocation pathway for bacterial virulence. Trends Microbiol 16(9):442–453

De Keersmaeker S, Van Mellaert L, Lammertyn E, Vrancken K, Anne J, Geukens N (2005a) Functional analysis of TatA and TatB in Streptomyces lividans. Biochem Biophys Res Commun 335(3):973–982

De Keersmaeker S, Van Mellaert L, Schaerlaekens K, Van Dessel W, Vrancken K, Lammertyn E, Anne J, Geukens N (2005b) Structural organization of the twin-arginine translocation system in *Streptomyces lividans*. FEBS Lett 579(3):797–802

De Keersmaeker S, Vrancken K, Van Mellaert L, Lammertyn E, Anne J, Geukens N (2006) Evaluation of TatABC overproduction on Tat- and Sec-dependent protein secretion in *Streptomyces lividans*. Arch Microbiol 186(6):507–512

De Keersmaeker S, Vrancken K, Van Mellaert L, Anne J, Geukens N (2007) The Tat pathway in *Streptomyces lividans*: interaction of Tat subunits and their role in translocation. Microbiology 153(Pt 4):1087–1094

DeLisa MP, Samuelson P, Palmer T, Georgiou G (2002) Genetic analysis of the twin arginine translocator secretion pathway in bacteria. J Biol Chem 277(33):29825–29831

DeLisa MP, Tullman D, Georgiou G (2003) Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci USA 100(10):6115–6120

DeLisa MP, Lee P, Palmer T, Georgiou G (2004) Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. J Bacteriol 186(2):366–373

Desvaux M, Hebraud M (2006) The protein secretion systems in Listeria: inside out bacterial virulence. FEMS Microbiol Rev 30(5):774–805

Eder S, Shi L, Jensen K, Yamane K, Hulett FM (1996) A *Bacillus subtilis* secreted phosphodiesterase/alkaline phosphatase is the product of a Pho regulon gene, phoD. Microbiology 142:2041–2047

Eijlander RT, Jongbloed JD, Kuipers OP (2009a) Relaxed specificity of the *Bacillus subtilis* TatAdCd translocase in Tat-dependent protein secretion. J Bacteriol 191(1):196–202

Eijlander RT, Kolbusz MA, Berendsen EM, Kuipers OP (2009b) Effects of altered TatC proteins on protein secretion efficiency via the twin-arginine translocation pathway of *Bacillus subtilis*. Microbiology 155(Pt 6):1776–1785

Eymann C, Dreisbach A, Albrecht D, Bernhardt J, Becher D, Gentner S, Le Tam T, Buttner K, Buurman G, Scharf C, Venz S, Volker U, Hecker M (2004) A comprehensive proteome map of growing *Bacillus subtilis* cells. Proteomics 4(10):2849–2876

Friedrich CG, Quentmeier A, Bardischewsky F, Rother D, Kraft R, Kostka S, Prinz H (2000) Novel genes coding for lithotrophic sulfur oxidation of Paracoccus pantotrophus GB17. J Bacteriol 182(17):4677–4687

Frielingsdorf S, Jakob M, Klosgen RB (2008) A stromal pool of TatA promotes Tat-dependent protein transport across the thylakoid membrane. J Biol Chem 283(49):33838–33845

Frobel J, Rose P, Muller M (2011) Early contacts between substrate proteins and TatA translocase component in twin-arginine translocation. J Biol Chem 286(51):43679–43689

Frobel J, Rose P, Lausberg F, Blummel AS, Freudl R, Muller M (2012) Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. Nat Commun 3:1311

Gohlke U, Pullan L, McDevitt CA, Porcelli I, de Leeuw E, Palmer T, Saibil HR, Berks BC (2005) The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. Proc Natl Acad Sci USA 102(30):10482–10486

Goosens VJ, Otto A, Glasner C, Monteferrante CC, van der Ploeg R, Hecker M, Becher D, van Dijl JM (2013) Novel twin-arginine translocation pathway-dependent phenotypes of *Bacillus subtilis* unveiled by quantitative proteomics. J Proteome Res 12(2):796–807

Goosens VJ, Monteferrante CG, van Dijl JM (2014a) Co-factor insertion and disulfide bond requirements for twin-arginine translocase-dependent export of the *Bacillus subtilis* Rieske protein QcrA. J Biol Chem 289(19):13124–13131

Goosens VJ, Monteferrante CG, van Dijl JM (2014b) The Tat system of Gram-positive bacteria. Biochim Biophys Acta 1843(8):1698–1706

Goosens VJ, De-San-Eustaquio-Campillo A, Carballido-Lopez R, van Dijl JM (2015) A Tat menage a trois—the role of *Bacillus subtilis* TatAc in twin-arginine protein translocation. Biochim Biophys Acta 1853(10 Pt A):2745–2753

Gouffi K, Gerard F, Santini CL, Wu LF (2004) Dual topology of the *Escherichia coli* TatA protein. J Biol Chem 279(12):11608–11615

Graumann P (2011) Bacillus: Cellular and Molecular Biology. Caister Academic Press, Wymondham, United Kingdom

Halbedel S, Reiss S, Hahn B, Albrecht D, Mannala GK, Chakraborty T, Hain T, Engelmann S, Flieger A (2014) A systematic proteomic analysis of Listeria monocytogenes house-keeping protein secretion systems. Mol Cell Proteomics 13(11):3063–3081

Hicks MG, de Leeuw E, Porcelli I, Buchanan G, Berks BC, Palmer T (2003) The *Escherichia coli* twin-arginine translocase: conserved residues of TatA and TatB family components involved in protein transport. FEBS Lett 539(1–3):61–67

Holzapfel E, Eisner G, Alami M, Barrett CM, Buchanan G, Luke I, Betton JM, Robinson C, Palmer T, Moser M, Muller M (2007) The entire N-terminal half of TatC is involved in twin-arginine precursor binding. Biochemistry 46(10):2892–2898

Hu Y, Zhao E, Li H, Xia B, Jin C (2010) Solution NMR structure of the TatA component of the twin-arginine protein transport system from gram-positive bacterium *Bacillus subtilis*. J J Am Chem Soc 132(45):15942–15944

Hunsicker-Wang LM, Heine A, Chen Y, Luna EP, Todaro T, Zhang YM, Williams PA, McRee DE, Hirst J, Stout CD, Fee JA (2003) High-resolution structure of the soluble, respiratory-type Rieske protein from Thermus thermophilus: analysis and comparison. Biochemistry 42(24):7303–7317

Hutcheon GW, Bolhuis A (2003) The archaeal twin-arginine translocation pathway. Biochem Soc Trans 31(Pt 3):686–689

Hynds PJ, Robinson D, Robinson C (1998) The sec-independent twin-arginine translocation system can transport both tightly folded and malfolded proteins across the thylakoid membrane. J Biol Chem 273(52):34868–34874

Ikeda M, Nakagawa S (2003) The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. Appl Microbiol Biotechnol 62(2–3):99–109

Iwata S, Saynovits M, Link TA, Michel H (1996) Structure of a water soluble fragment of the 'Rieske' iron-sulfur protein of the bovine heart mitochondrial cytochrome bc1 complex determined by MAD phasing at 1.5 A resolution. Structure (London, England: 1993) 4(5): 567–579

Jack RL, Buchanan G, Dubini A, Hatzixanthis K, Palmer T, Sargent F (2004) Coordinating assembly and export of complex bacterial proteins. EMBO J 23(20):3962–3972

Jongbloed JD, Martin U, Antelmann H, Hecker M, Tjalsma H, Venema G, Bron S, van Dijl JM, Muller J (2000) TatC is a specificity determinant for protein secretion via the twin-arginine translocation pathway. J Biol Chem 275(52):41350–41357

Jongbloed JD, Antelmann H, Hecker M, Nijland R, Bron S, Airaksinen U, Pries F, Quax WJ, van Dijl JM, Braun PG (2002) Selective contribution of the twin-arginine translocation pathway to protein secretion in *Bacillus subtilis*. J Biol Chem 277(46):44068–44078

Jongbloed JD, Grieger U, Antelmann H, Hecker M, Nijland R, Bron S, van Dijl JM (2004) Two minimal Tat translocases in Bacillus. Mol Microbiol 54(5):1319–1325

Joshi MV, Mann SG, Antelmann H, Widdick DA, Fyans JK, Chandra G, Hutchings MI, Toth I, Hecker M, Loria R, Palmer T (2010) The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen Streptomyces scabies. Mol Microbiol 77 (1):252–271

Kalinowski J, Bathe B, Bartels D, Bischoff N, Bott M, Burkovski A, Dusch N, Eggeling L, Eikmanns BJ, Gaigalat L, Goesmann A, Hartmann M, Huthmacher K, Kramer R, Linke B, McHardy AC, Meyer F, Mockel B, Pfefferle W, Puhler A, Rey DA, Ruckert C, Rupp O, Sahm H, Wendisch VF, Wiegrabe I, Tauch A (2003) The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. J Biotechnol 104(1–3):5–25

Keller R, de Keyzer J, Driessen AJ, Palmer T (2012) Co-operation between different targeting pathways during integration of a membrane protein. J Cell Biol 199(2):303–315

Khan MT, Duncan SH, Stams AJ, van Dijl JM, Flint HJ, Harmsen HJ (2012) The gut anaerobe Faecalibacterium prausnitzii uses an extracellular electron shuttle to grow at oxic-anoxic interphases. ISME J 6(8):1578–1585

Kikuchi Y, Date M, Itaya H, Matsui K, Wu LF (2006) Functional analysis of the twin-arginine translocation pathway in *Corynebacterium glutamicum* ATCC 13869. Appl Environ Microbiol 72(11):7183–7192

Kikuchi Y, Itaya H, Date M, Matsui K, Wu LF (2008) Production of Chryseobacterium proteolyticum protein-glutaminase using the twin-arginine translocation pathway in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 78(1):67–74

Kim JY, Fogarty EA, Lu FJ, Zhu H, Wheelock GD, Henderson LA, DeLisa MP (2005) Twin-arginine translocation of active human tissue plasminogen activator in *Escherichia coli*. Appl Environ Microbiol 71(12):8451–8459

Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW, Stiekema W, Lankhorst RM, Bron PA, Hoffer SM, Groot MN, Kerkhoven R, de Vries M, Ursing B, de Vos WM, Siezen RJ (2003) Complete genome sequence of Lactobacillus plantarum WCFS1. Proc Natl Acad Sci U S A 100(4):1990–1995

Kneuper H, Maldonado B, Jager F, Krehenbrink M, Buchanan G, Keller R, Muller M, Berks BC, Palmer T (2012) Molecular dissection of TatC defines critical regions essential for protein transport and a TatB-TatC contact site. Mol Microbiol 85(5):945–961

Kobayashi R, Suzuki T, Yoshida M (2007) *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. Mol Microbiol 66(1):100–109

Kolkman MA, van der Ploeg R, Bertels M, van Dijk M, van der Laan J, van Dijl JM, Ferrari E (2008) The twin-arginine signal peptide of *Bacillus subtilis* YwbN can direct either Tat- or Sec-dependent secretion of different cargo proteins: secretion of active subtilisin via the B. subtilis Tat pathway. Appl Environ Microbiol 74(24):7507–7513

Kouwen TR, van der Ploeg R, Antelmann H, Hecker M, Homuth G, Mader U, van Dijl JM (2009) Overflow of a hyper-produced secretory protein from the Bacillus Sec pathway into the Tat pathway for protein secretion as revealed by proteogenomics. Proteomics 9(4):1018–1032

Krishnappa L, Monteferrante CG, van Dijl JM (2012) Degradation of the Twin-Arginine Translocation Substrate YwbN by Extracytoplasmic Proteases of *Bacillus subtilis*. Appl Environ Microbiol 78(21):7801–7804

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390(6657):249–256

Lange C, Muller SD, Walther TH, Burck J, Ulrich AS (2007) Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach. Biochim Biophys Acta 1768(10):2627–2634

Leake MC, Greene NP, Godun RM, Granjon T, Buchanan G, Chen S, Berry RM, Palmer T, Berks BC (2008) Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. Proc Natl Acad Sci USA 105(40):15376–15381

Ledala N, Sengupta M, Muthaiyan A, Wilkinson BJ, Jayaswal RK (2010) Transcriptomic response of Listeria monocytogenes to iron limitation and Fur mutation. Appl Environ Microbiol 76(2):406–416

Link TA, Saynovits M, Assmann C, Iwata S, Ohnishi T, von Jagow G (1996) Isolation, characterisation and crystallisation of a water-soluble fragment of the Rieske iron-sulfur protein of bovine heart mitochondrial bc1 complex. Eur J Biochem 237(1):71–75

Liu YW, Hitchcock A, Salmon RC, Kelly DJ (2014) It takes two to tango: two TatA paralogues and two redox enzyme-specific chaperones are involved in the localization of twin-arginine translocase substrates in Campylobacter jejuni. Microbiology 160(Pt 9):2053–2066

Luke I, Handford JI, Palmer T, Sargent F (2009) Proteolytic processing of *Escherichia coli* twin-arginine signal peptides by LepB. Arch Microbiol 191(12):919–925

Ma X, Cline K (2013) Mapping the signal peptide binding and oligomer contact sites of the core subunit of the pea twin arginine protein translocase. Plant Cell 25(3):999–1015

Machado H, Lourenco A, Carvalho F, Cabanes D, Kallipolitis BH, Brito L (2013) The Tat pathway is prevalent in Listeria monocytogenes lineage II and is not required for infection and spread in host cells. J Mol Microbiol Biotechnol 23(3):209–218

Matos CF, Robinson C, Di Cola A (2008) The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. EMBO J 27(15):2055–2063

Matos CF, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, Freedman RB, Keshavarz-Moore E (2013) Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. Biotechnol Prog 30(2):281–290

Maurer C, Panahandeh S, Moser M, Muller M (2009) Impairment of twin-arginine-dependent export by seemingly small alterations of substrate conformation. FEBS Lett 583(17): 2849–2853

McDevitt CA, Buchanan G, Sargent F, Palmer T, Berks BC (2006) Subunit composition and in vivo substrate-binding characteristics of *Escherichia coli* Tat protein complexes expressed at native levels. FEBS J 273(24):5656–5668

McDonough JA, Hacker KE, Flores AR, Pavelka MS Jr, Braunstein M (2005) The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial beta-lactamases. J Bacteriol 187(22):7667–7679

Mehner D, Osadnik H, Lunsdorf H, Bruser T (2012) The Tat system for membrane translocation of folded proteins recruits the membrane-stabilizing Psp machinery in *Escherichia coli*. J Biol Chem 287(33):27834–27842

Meissner D, Vollstedt A, van Dijl JM, Freudl R (2007) Comparative analysis of twin-arginine (Tat)-dependent protein secretion of a heterologous model protein (GFP) in three different Gram-positive bacteria. Appl Microbiol Biotechnol 76(3):633–642

Mendel S, McCarthy A, Barnett JP, Eijlander RT, Nenninger A, Kuipers OP, Robinson C (2008) The *Escherichia coli* TatABC system and a *Bacillus subtilis* TatAC-type system recognise three distinct targeting determinants in twin-arginine signal peptides. J Mol Biol 375(3): 661–672

Miethke M, Monteferrante CG, Marahiel MA, van Dijl JM (2013) The *Bacillus subtilis* EfeUOB transporter is essential for high-affinity acquisition of ferrous and ferric iron. Biochim Biophys Acta 1833:2267–2278

Molik S, Karnauchov I, Weidlich C, Herrmann RG, Klosgen RB (2001) The Rieske Fe/S protein of the cytochrome b6/f complex in chloroplasts: missing link in the evolution of protein transport pathways in chloroplasts? J Biol Chem 276(46):42761–42766

Monteferrante CG, Baglieri J, Robinson C, van Dijl JM (2012a) TatAc, the third TatA subunit of *Bacillus subtilis*, can form active twin-arginine translocases with the TatCd and TatCy subunits. Appl Environ Microbiol 78(14):4999–5001

Monteferrante CG, Miethke M, van der Ploeg R, Glasner C, van Dijl JM (2012b) Specific targeting of the metallophosphoesterase YkuE to the bacillus cell wall requires the twin-arginine translocation system. J Biol Chem 287(35):29789–29800

Monteferrante CG, Mackichan C, Marchadier E, Prejean MV, Carballido-Lopez R, van Dijl JM (2013) Mapping the twin-arginine protein translocation network of *Bacillus subtilis*. Proteomics 13(5):800–811

Mori H, Cline K (2001) Post-translational protein translocation into thylakoids by the Sec and DeltapH-dependent pathways. Biochim Biophys Acta 1541(1–2):80–90

Mori H, Cline K (2002) A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid [Delta]pH/Tat translocase. J Cell Biol 157(2):205–210

Nicolas P, Mader U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebeke M, Aymerich S, Becher D, Bisicchia P, Botella E, Delumeau O, Doherty G, Denham EL, Fogg MJ, Fromion V, Goelzer A, Hansen A, Hartig E, Harwood CR, Homuth G, Jarmer H, Jules M, Klipp E, Le Chat L, Lecointe F, Lewis P, Liebermeister W, March A, Mars RA, Nannapaneni P, Noone D, Pohl S, Rinn B, Rugheimer F, Sappa PK, Samson F,

Schaffer M, Schwikowski B, Steil L, Stulke J, Wiegert T, Devine KM, Wilkinson AJ, van Dijl JM, Hecker M, Volker U, Bessieres P, Noirot P (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. Science 335(6072):1103–1106

Nishio Y, Nakamura Y, Kawarabayasi Y, Usuda Y, Kimura E, Sugimoto S, Matsui K, Yamagishi A, Kikuchi H, Ikeo K, Gojobori T (2003) Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens. Genome Res 13(7):1572–1579

Oates J, Barrett CM, Barnett JP, Byrne KG, Bolhuis A, Robinson C (2005) The *Escherichia coli* twin-arginine translocation apparatus incorporates a distinct form of TatABC complex, spectrum of modular TatA complexes and minor TatAB complex. J Mol Biol 346(1):295–305

Oertel D, Schmitz S, Freudl R (2015) A TatABC-Type Tat translocase is required for unimpaired aerobic growth of *Corynebacterium glutamicum* ATCC13032. PLoS ONE 10(4):e0123413

Oresnik IJ, Ladner CL, Turner RJ (2001) Identification of a twin-arginine leader-binding protein. Mol Microbiol 40(2):323–331

Palmer T, Berks BC (2012) The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol 10(7):483–496

Panahandeh S, Maurer C, Moser M, DeLisa MP, Muller M (2008) Following the path of a twin-arginine precursor along the TatABC translocase of *Escherichia coli*. J Biol Chem 283 (48):33267–33275

Papish AL, Ladner CL, Turner RJ (2003) The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the *Escherichia coli* twin-arginine translocase. J Biol Chem 278(35):32501–32506

Patel R, Smith SM, Robinson C (2014) Protein transport by the bacterial Tat pathway. Biochim Biophys Acta 1843(8):1620–1628

Pett W, Lavrov DV (2013a) The twin-arginine subunit C in Oscarella: origin, evolution, and potential functional significance. Integr Comp Biol 53(3):495–502

Pett W, Lavrov DV (2013b) The twin-arginine subunit C in oscarella: origin, evolution, and potential functional significance. Integr Comp Biol 53(3):495–502

Pop OI, Westermann M, Volkmer-Engert R, Schulz D, Lemke C, Schreiber S, Gerlach R, Wetzker R, Muller JP (2003) Sequence-specific binding of prePhoD to soluble TatAd indicates protein-mediated targeting of the Tat export in *Bacillus subtilis*. J Biol Chem 278(40):38428–38436

Punginelli C, Maldonado B, Grahl S, Jack R, Alami M, Schroder J, Berks BC, Palmer T (2007) Cysteine scanning mutagenesis and topological mapping of the *Escherichia coli* twin-arginine translocase TatC Component. J Bacteriol 189(15):5482–5494

Raynaud C, Guilhot C, Rauzier J, Bordat Y, Pelicic V, Manganelli R, Smith I, Gicquel B, Jackson M (2002) Phospholipases C are involved in the virulence of Mycobacterium tuberculosis. Mol Microbiol 45(1):203–217

Ribnicky B, Van Blarcom T, Georgiou G (2007) A scFv antibody mutant isolated in a genetic screen for improved export via the twin arginine transporter pathway exhibits faster folding. J Mol Biol 369(3):631–639

Richter S, Lindenstrauss U, Lucke C, Bayliss R, Bruser T (2007) Functional Tat transport of unstructured, small, hydrophilic proteins. J Biol Chem 282(46):33257–33264

Ridder AN, de Jong EJ, Jongbloed JD, Kuipers OP (2009) Subcellular localization of TatAd of *Bacillus subtilis* depends on the presence of TatCd or TatCy. J Bacteriol 191(13):4410–4418

Robinson C, Matos CF, Beck D, Ren C, Lawrence J, Vasisht N, Mendel S (2011) Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. Biochim Biophys Acta 1808(3):876–884

Rocco MA, Waraho-Zhmayev D, DeLisa MP (2012) Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. Proc Natl Acad Sci USA 109(33):13392–13397

Rodrigue A, Chanal A, Beck K, Muller M, Wu LF (1999) Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. J Biol Chem 274(19):13223–13228

Rodriguez F, Rouse SL, Tait CE, Harmer J, De Riso A, Timmel CR, Sansom MS, Berks BC, Schnell JR (2013) Structural model for the protein-translocating element of the twin-arginine transport system. Proc Natl Acad Sci U S A 110(12):E1092–1101

Rodriguez F, Lillington J, Johnson S, Timmel CR, Lea SM, Berks BC (2014) Crystal structure of the *Bacillus subtilis* phosphodiesterase PhoD reveals an iron and calcium-containing active site. J Biol Chem 289(45):30889–30899

Rose RW, Bruser T, Kissinger JC, Pohlschroder M (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. Mol Microbiol 45(4):943–950

Sargent F, Stanley NR, Berks BC, Palmer T (1999) Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. J Biol Chem 274(51): 36073–36082

Schaerlaekens K, Schierova M, Lammertyn E, Geukens N, Anne J, Van Mellaert L (2001) Twin-arginine translocation pathway in *Streptomyces lividans*. J Bacteriol 183(23):6727–6732

Schaerlaekens K, Lammertyn E, Geukens N, De Keersmaeker S, Anne J, Van Mellaert L (2004a) Comparison of the Sec and Tat secretion pathways for heterologous protein production by Streptomyces lividans. J Biotechnol 112(3):279–288

Schaerlaekens K, Van Mellaert L, Lammertyn E, Geukens N, Anne J (2004b) The importance of the Tat-dependent protein secretion pathway in Streptomyces as revealed by phenotypic changes in tat deletion mutants and genome analysis. Microbiology 150(Pt 1):21–31

Scheele S, Oertel D, Bongaerts J, Evers S, Hellmuth H, Maurer KH, Bott M, Freudl R (2013) Secretory production of an FAD cofactor-containing cytosolic enzyme (sorbitol-xylitol oxidase from Streptomyces coelicolor) using the twin-arginine translocation (Tat) pathway of *Corynebacterium glutamicum*. Microb Biotechnol 6(2):202–206

Schmidt CL, Shaw L (2001) A comprehensive phylogenetic analysis of Rieske and Rieske-type iron-sulfur proteins. J Bioenerg Biomembr 33(1):9–26

Schneider D, Schmidt CL (2005) Multiple Rieske proteins in prokaryotes: where and why? Biochim Biophys Acta 1710(1):1–12

Schreiber S, Stengel R, Westermann M, Volkmer-Engert R, Pop OI, Muller JP (2006) Affinity of TatCd for TatAd elucidates its receptor function in the *Bacillus subtilis* twin arginine translocation (Tat) translocase system. J Biol Chem 281(29):19977–19984

Simone D, Bay DC, Leach T, Turner RJ (2013) Diversity and evolution of bacterial twin arginine translocase protein, TatC, reveals a protein secretion system that is evolving to fit its environmental niche. PLoS ONE 8(11):e78742

Solans L, Gonzalo-Asensio J, Sala C, Benjak A, Uplekar S, Rougemont J, Guilhot C, Malaga W, Martin C, Cole ST (2014) The PhoP-dependent ncRNA Mcr7 modulates the TAT secretion system in Mycobacterium tuberculosis. PLoS Pathog 10(5):e1004183

Sousa PM, Videira MA, Santos FA, Hood BL, Conrads TP, Melo AM (2013) The bc:caa3 supercomplexes from the Gram positive bacterium *Bacillus subtilis* respiratory chain: a megacomplex organization? Arch Biochem Biophys 537(1):153–160

Standar K, Mehner D, Osadnik H, Berthelmann F, Hause G, Lunsdorf H, Bruser T (2008) PspA can form large scaffolds in *Escherichia coli*. FEBS Lett 582(25–26):3585–3589

Stanley NR, Palmer T, Berks BC (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. J Biol Chem 275(16): 11591–11596

Strauch EM, Georgiou G (2007) *Escherichia coli* tatC mutations that suppress defective twin-arginine transporter signal peptides. J Mol Biol 374(2):283–291

Thiemann V, Saake B, Vollstedt A, Schafer T, Puls J, Bertoldo C, Freudl R, Antranikian G (2006) Heterologous expression and characterization of a novel branching enzyme from the thermoalkaliphilic anaerobic bacterium Anaerobranca gottschalkii. Appl Microbiol Biotechnol 72(1):60–71

Tiwari KB, Birlingmair J, Wilkinson BJ, Jayaswal RK (2015) Role of the twin-arginine translocase (tat) system in iron uptake in Listeria monocytogenes. Microbiology 161(Pt 2): 264–271

Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev 64(3):515–547

Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E, Dorenbos R, Dubois JY, Westers H, Zanen G, Quax WJ, Kuipers OP, Bron S, Hecker M, van Dijl JM (2004) Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev 68(2):207–233

Tottey S, Waldron KJ, Firbank SJ, Reale B, Bessant C, Sato K, Cheek TR, Gray J, Banfield MJ, Dennison C, Robinson NJ (2008) Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. Nature 455(7216):1138–1142

Turlin E, Debarbouille M, Augustyniak K, Gilles AM, Wandersman C (2013) Staphylococcus aureus FepA and FepB proteins drive heme iron utilization in *Escherichia coli*. PLoS ONE 8 (2):e56529

van der Ploeg R, Barnett JP, Vasisht N, Goosens VJ, Pother DC, Robinson C, van Dijl JM (2011a) Salt sensitivity of minimal twin arginine translocases. J Biol Chem 286(51):43759–43770

van der Ploeg R, Mader U, Homuth G, Schaffer M, Denham EL, Monteferrante CG, Miethke M, Marahiel MA, Harwood CR, Winter T, Hecker M, Antelmann H, van Dijl JM (2011b) Environmental salinity determines the specificity and need for Tat-dependent secretion of the YwbN protein in *Bacillus subtilis*. PLoS ONE 6(3):e18140

van der Ploeg R, Monteferrante CG, Piersma S, Barnett JP, Kouwen TR, Robinson C, van Dijl JM (2012) High-salinity growth conditions promote Tat-independent secretion of tat substrates in *Bacillus subtilis*. Appl Environ Microbiol 78(21):7733–7744

van Dijl JM, Braun PG, Robinson C, Quax WJ, Antelmann H, Hecker M, Muller J, Tjalsma H, Bron S, Jongbloed JD (2002) Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion. J Biotechnol 98(2–3):243–254

Vrancken K, De Keersmaeker S, Geukens N, Lammertyn E, Anne J, Van Mellaert L (2007) pspA overexpression in Streptomyces lividans improves both Sec- and Tat-dependent protein secretion. Appl Microbiol Biotechnol 73(5):1150–1157

Vrancken K, Van Mellaert L, Anne J (2008) Characterization of the Streptomyces lividans PspA response. J Bacteriol 190(10):3475–3481

Walther TH, Grage SL, Roth N, Ulrich AS (2010) Membrane alignment of the pore-forming component TatA(d) of the twin-arginine translocase from *Bacillus subtilis* resolved by solid-state NMR spectroscopy. J Am Chem Soc 132(45):15945–15956

Walther TH, Gottselig C, Grage SL, Wolf M, Vargiu AV, Klein MJ, Vollmer S, Prock S, Hartmann M, Afonin S, Stockwald E, Heinzmann H, Nolandt OV, Wenzel W, Ruggerone P, Ulrich AS (2013) Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism. Cell 152(1–2):316–326

Weatherspoon-Griffin N, Zhao G, Kong W, Kong Y, Morigen H, Andrews-Polymenis M McClelland, Shi Y (2011) The CpxR/CpxA two-component system up-regulates two Tat-dependent peptidoglycan amidases to confer bacterial resistance to antimicrobial peptide. J Biol Chem 286(7):5529–5539

Westermann M, Pop OI, Gerlach R, Appel TR, Schlormann W, Schreiber S, Muller JP (2006) The TatAd component of the *Bacillus subtilis* twin-arginine protein transport system forms homo-multimeric complexes in its cytosolic and membrane embedded localisation. Biochim Biophys Acta 1758(4):443–451

Wexler M, Sargent F, Jack RL, Stanley NR, Bogsch EG, Robinson C, Berks BC, Palmer T (2000) TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. J Biol Chem 275(22):16717–16722

Whitaker N, Bageshwar UK, Musser SM (2012) Kinetics of precursor interactions with the bacterial Tat translocase detected by real-time FRET. J Biol Chem 287(14):11252–11260

Widdick DA, Dilks K, Chandra G, Bottrill A, Naldrett M, Pohlschroder M, Palmer T (2006) The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor. Proc Natl Acad Sci USA 103(47):17927–17932

Widdick DA, Eijlander RT, van Dijl JM, Kuipers OP, Palmer T (2008) A facile reporter system for the experimental identification of twin-arginine translocation (Tat) signal peptides from all kingdoms of life. J Mol Biol 375(3):595–603

Widdick DA, Hicks MG, Thompson BJ, Tschumi A, Chandra G, Sutcliffe IC, Brulle JK, Sander P, Palmer T, Hutchings MI (2011) Dissecting the complete lipoprotein biogenesis pathway in Streptomyces scabies. Mol Microbiol 80(5):1395–1412

Wolff S, Antelmann H, Albrecht D, Becher D, Bernhardt J, Bron S, Buttner K, van Dijl JM, Eymann C, Otto A, Le Tam T, Hecker M (2007) Towards the entire proteome of the model bacterium *Bacillus subtilis* by gel-based and gel-free approaches. J Chromatogr B Analyt Technol Biomed Life Sci 849(1–2):129–140

Wu LF, Chanal A, Rodrigue A (2000a) Membrane targeting and translocation of bacterial hydrogenases. Arch Microbiol 173(5–6):319–324

Wu LF, Ize B, Chanal A, Quentin Y, Fichant G (2000b) Bacterial twin-arginine signal peptide-dependent protein translocation pathway: evolution and mechanism. J Mol Microbiol Biotechnol 2(2):179–189

Xu P, Alves JM, Kitten T, Brown A, Chen Z, Ozaki LS, Manque P, Ge X, Serrano MG, Puiu D, Hendricks S, Wang Y, Chaplin MD, Akan D, Paik S, Peterson DL, Macrina FL, Buck GA (2007) Genome of the opportunistic pathogen Streptococcus sanguinis. J Bacteriol 189 (8):3166–3175

Yamada K, Sanzen I, Ohkura T, Okamoto A, Torii K, Hasegawa T, Ohta M (2007) Analysis of twin-arginine translocation pathway homologue in *Staphylococcus aureus*. Curr Microbiol 55(1):14–19

Yen MR, Tseng YH, Nguyen EH, Wu LF, Saier MH Jr (2002) Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system. Arch Microbiol 177(6):441–450

Yu J, Hederstedt L, Piggot PJ (1995) The cytochrome bc complex (menaquinone:cytochrome c reductase) in *Bacillus subtilis* has a nontraditional subunit organization. J Bacteriol 177(23):6751–6760

Zhang C, Xin Y, Wang Y, Guo T, Lu S, Kong J (2015) Identification of a novel dye-decolorizing peroxidase, EfeBs, translocated by twin-arginine translocation system in Streptococcus thermophilus CGMCC 7.179. Appl Environ Microbiol 81(18):6108–6119

Zoufaly S, Frobel J, Rose P, Flecken T, Maurer C, Moser M, Muller M (2012) Mapping precursor-binding site on TatC subunit of twin arginine-specific protein translocase by site-specific photo cross-linking. J Biol Chem 287(16):13430–13441

Membrane Translocation and Assembly of Sugar Polymer Precursors

Véronique L. Taylor, Steven M. Huszczynski and Joseph S. Lam

Abstract Bacterial polysaccharides play an essential role in cell viability, virulence, and evasion of host defenses. Although the polysaccharides themselves are highly diverse, the pathways by which bacteria synthesize these essential polymers are conserved in both Gram-negative and Gram-positive organisms. By utilizing a lipid linker, a series of glycosyltransferases and integral membrane proteins act in concert to synthesize capsular polysaccharide, teichoic acid, and teichuronic acid. The pathways used to produce these molecules are the Wzx/Wzy-dependent, the ABC-transporter-dependent, and the synthase-dependent pathways. This chapter will cover the initiation, synthesis of the various polysaccharides on the cytoplasmic face of the membrane using nucleotide sugar precursors, and export of the nascent chain from the cytoplasm to the extracellular milieu. As microbial glycobiology is an emerging field in Gram-positive bacteria research, parallels will be drawn to the more widely studied polysaccharide biosynthesis systems in Gram-negative species in order to provide greater understanding of these biologically significant molecules.

Abbreviations

PG	Peptidoglycan
CPS	Capsular polysaccharide
Und-P	Undecaprenyl phosphate
Und-PP	Undecaprenyl pyrophosphate
GlcNAc	*N*-acetylglucosamine
ManNAc	*N*-acetylmannuronic acid
MurNAc	*N*-acetylmuramic acid
UDP-Glc	UDP-glucose
TMR	Tetramethylrhodamine cadaverine

V.L. Taylor · S.M. Huszczynski · J.S. Lam (✉)
Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON N1G 2W1, Canada
e-mail: jlam@uoguelph.ca

Current Topics in Microbiology and Immunology (2017) 404:95–128
DOI 10.1007/82_2015_5014

Published Online: 16 January 2016

TMS	Transmembrane segment
MOP	Multidrug/oligosaccharidyl-lipid/polysaccharide
MATE	Multidrug and toxin extrusion
MTSES	2-sulfonatoethyl methanethiosulfonate
MVF	Mouse virulence factor
NBD	7-nitro-2,1,3-benzoxadiazol-4-yl
PHPT	Polyprenyl-phosphate hexose-1-phosphate transferase
PST	Polysaccharide transporter
LCP	LytR-CpsA-Psr
Rha	Rhamnose
TA	Teichoic acid
LTA	Lipoteichoic acid
WTA	Wall teichoic acid
GlcA	Uronic acid
TUA	Teichuronic acid
glycerol-3-P	Glycerol-3-phosphate
ribitol-3-P	Ribitol-3-phosphate

Contents

1 Lipid Carrier Involved in Polysaccharide Biosynthesis

The first step in creating cell-wall polysaccharides, such as peptidoglycan (PG) or capsule (CPS), requires the addition of sugars to a lipid carrier to form a precursor compound, which facilitates translocation of sugars across the cytoplasmic membrane. The process of relying on a linear chain polyprenyl-phosphate lipid as the base for translocation of sugar polymers is a highly conserved process across all biological kingdoms. The most commonly used lipid for sugar translocation in bacteria is undecaprenyl-phosphate made of 55-carbon chain length (C_{55}-P or Und-P) (Manat et al. 2014). In brief, the formation of this C_{55}-polyprenyl chain occurs through the sequential condensation of short lipid chains to the appropriate length in a *cis* configuration to a *trans*-linked precursor with the terminal phosphate group in the α-position linked in an unsaturated bond (Hartley and Imperiali 2012). This structure has been shown to be crucial in the translocation of a polar head group through a highly hydrophobic environment by the interaction with a dedicated flippase protein. Though the exact mechanism and the kinetics of interaction between Und-P and the flippase are currently unknown, biophysical evidence demonstrates that the presence of these lipids increases membrane fluidity, thereby promoting a more dynamic state (Valtersson et al. 1985; Wang et al. 2008). Once translocated by a protein such as CpsJ of *Streptococcus pneumoniae*, a putative capsular flippase, the lipid–phosphate–sugar intermediate is believed to act as a donor for the assembly of cellular polysaccharide.

2 Lipid II Translocation

PG synthesis is one of the most intensely studied areas of bacterial research because, with only a few exceptions, it is essential for survival of the bacterial cell. The conserved mechanisms in this pathway have been, and continue to be, targets for developing broad-spectrum antibiotics. The building block of PG is lipid II, an Und-P-linked GlcNAc-MurNAc disaccharide with a pentapeptide side chain that is synthesized on the inner face of the cytoplasmic membrane. Once lipid II is transported across the membrane, it is linked to the protective mesh-like network of PG via penicillin binding proteins (Nikolaidis et al. 2014). Despite the fact that almost each of the steps in the PG biosynthesis pathway has been thoroughly characterized, the protein(s) responsible for the transport of lipid II has remained elusive and the subject of debate; therefore, this action will be the focus of the chapter. It was first proposed that FtsW, an essential cell-division protein, belonging to the shape, elongation, division, and sporulation (SEDS) family, was the sought after flippase (Ghuysen and Hakenbeck 1994; Höltje 1998). Mohammadi et al. (2011) provided rather convincing evidence from in vitro experiments that FtsW was the flippase in *Escherichia coli*. They used a FRET-based assay in which vancomycin and lipid II were labeled with acceptor and donor FRET pairs,

tetramethylrhodamine cadaverine (TMR), and 7-nitro-2,1,3-benzoxadiazol-4-yl (NBD), respectively. Since vancomycin is known to bind to lipid II, this interaction can be monitored by observing the quenching of the NBD signal and concomitant increase in TMR signal. Vancomycin, however, cannot cross the cytoplasmic membrane so it could only bind lipid II that is on the outside of the cytoplasmic membrane. Using this assay, the authors observed that over-expression of FtsW in *E. coli* right-side-out vesicles, or subjecting purified FtsW to be reconstituted in proteoliposomes, decreased the NBD signal and increased the TMR signal, proving that lipid II had been flipped (Mohammadi et al. 2011). Although these data are compelling, a previous study that used bioinformatic approaches has collected evidence to suggest that a protein belonging to the mouse virulence factor family (MVF, part of the multidrug/oligosaccharidyl-lipid/polysaccharide [MOP] exporter superfamily), MviN, herein called MurJ, is the lipid II transporter (Ruiz 2008). Ruiz used a "reductionist approach" by identifying the predicted inner membrane proteins that are present in *E. coli* and in two Gram-negative endosymbiotic bacteria. The author chose endosymbiotic bacteria for comparisons because their proteomes are much smaller, yet they should still contain the machinery for peptidoglycan synthesis. Using this approach, the author identified six genes of unknown function and then searched the genomes of PG- and non-PG-producing bacteria. MurJ was the only one that was present in PG-producing bacteria. Ruiz postulated that since FtsW was present in *Mollicute Eubacterium dolichum* DSM 3991, a non-PG producer, it was likely not the lipid II flippase. However, Mohammadi et al. argued that the presence of FtsW in the absence of PG does not rule out its role in biosynthesis because proteins involved in the cell-wall biosynthesis pathway have been discovered in bacteria that lack PG (Henrichfreise et al. 2009; Mohammadi et al. 2011). Regardless, a line of evidence consistent of naming MurJ as the flippase is that the *murJ* gene is essential for survival (Rudnick et al. 2001; Inoue et al. 2008; Ruiz 2008) as previous attempts to delete the gene proved unsuccessful. Therefore, using an L-arabinose inducible promoter within *murJ*, it was observed that in the absence of added inducer, *E. coli* cells accumulated PG nucleotide precursors and lipid I (MurNAc-pentapeptide) in the cytoplasm (Ruiz 2008). Computational modeling of the MurJ structure revealed a solvent-exposed cavity and the typical "V"-shaped structure observed in crystallized MOP proteins. This model was validated via two topology methods, based on PhoA-LacZα reporter fusions and the substituted cysteine accessibility method (Butler et al. 2013). Further, the authors identified essential positively charged amino acid residues within the solvent-exposed channel and proposed that these residues are involved in the binding of lipid II (Butler et al. 2014). One would anticipate that some of these positive charges might be conserved in other bacteria. Indeed, the MurJ homolog YtgP from *Streptococcus pyogenes* was found to contain 3 analogous essential positive residues in its putative channel as well as several other positive residues that did not align with the MurJ structural model. These differentially located positive charges may reflect the fact that there are differences in the Gram-positive pentapeptide of lipid II (Butler et al. 2014). The results presented for MurJ and YtgP mirror the results of that in the studies of the O-antigen

flippase in *Pseudomonas aeruginosa* explained below. It should be noted that, to date, there is only a handful of crystalized MOP exporter superfamily proteins, and these all belong to the multidrug and toxin extrusion (MATE) class. Therefore, MATE proteins as the top structural hit may be biased and not representative of the true structure of, for instance, a prokaryotic polysaccharide transporter (PST) protein, which is also found in the MOP exporter superfamily. In the aforementioned studies, however, the authors have used an experimental approach to examine the topology of the proteins of interest and have obtained a considerable amount of data to arrive at the structure predictions. Future structure–function studies of MOP exporter proteins should exercise the same degree of scrutiny in the methodology used in order to experimentally validate structure models based on in silico algorithms.

In a subsequent study, the Ruiz group used an in vivo assay to demonstrate MurJ flippase activity in *E. coli.* This assay took advantage of the fact that colicin M cleaves MurNAc-GlcNAc-pentapeptide diphosphate from undecaprenyl in the periplasm and that a cysteine variant of MurJ ($MurJ^{A29C}$) could be inactivated by sodium (2-sulfonatoethyl)methanethiosulfonate (MTSES). *E. coli* cells carrying this $MurJ^{A29C}$ variant were treated with colicin M in the presence or absence of MTSES. In the presence of MTSES, lipid II cleavage by colicin M was blocked, whereas in wild-type cells or untreated $MurJ^{A29C}$ cells, lipid II was produced and made available for processing (Sham et al. 2014).

There was still some doubt that MurJ was the bona fide lipid II flippase because previous experiments that deleted four MurJ homologs in *Bacillus subtilis* were not lethal and did not exhibit any morphological defects. It was suggested that either MurJ is not the flippase or another protein is able to compensate for its absence (Fay and Dworkin 2009). In a recent investigation by Meeske et al. (2015), a protein that has activity for lipid II flipping was identified. In this study, the authors created a *B. subtilis* strain containing the deletion of the four previously investigated proteins and six more MOP exporter superfamily members. Strikingly, this strain grew at a rate that is comparable to the wild type. Using a synthetic lethal screen, *amj* (renamed from *ydaH*) was identified. Several lines of evidence pointed to Amj as an alternate flippase: (i) *amj* could not be deleted in a Δ*murJ* background, (ii) *amj* could be used to complement a *murJ* mutation in *E. coli*, and (iii) Amj or MurJ from *B. subtilis* ($MurJ_{Bs}$) was able to flip lipid II in the colicin M in vivo assay. Expression of Amj was upregulated in the absence of MurJ, which led to the suggestion that Amj represents a "backup" system for lipid II flipping when MurJ becomes non-functional, in the presence of antibiotics (Meeske et al. 2015). Interestingly, it has been difficult to identify Amj as a flippase because this protein does not have any sequence similarity to either the MOP or ABC-transporter protein families. Amj is predicted to contain 6 transmembrane segments (TMS), compared to the MOP family of proteins, which usually contain 12–15 TMS, and it does not have any ATP-binding domains. Meeske et al. (2015) suggested that Amj oligomerizes to form a channel for lipid II flipping. Interestingly, the primordial ancestors of MOP superfamily members are predicted to have six TMS. As MOP

proteins exhibit strong similarity between the first and the second half of their sequence, it is postulated that a duplication event had given rise to the proteins with 12 TMS, and subsequent addition of TMS might have allowed the protein to evolve to a larger size with up to 15 TMS (Hvorup et al. 2003a, b). Whether Amj represents the primordial protein that gave rise to the MOP superfamily or if this is sheer coincidence certainly is an intriguing question for future experiments.

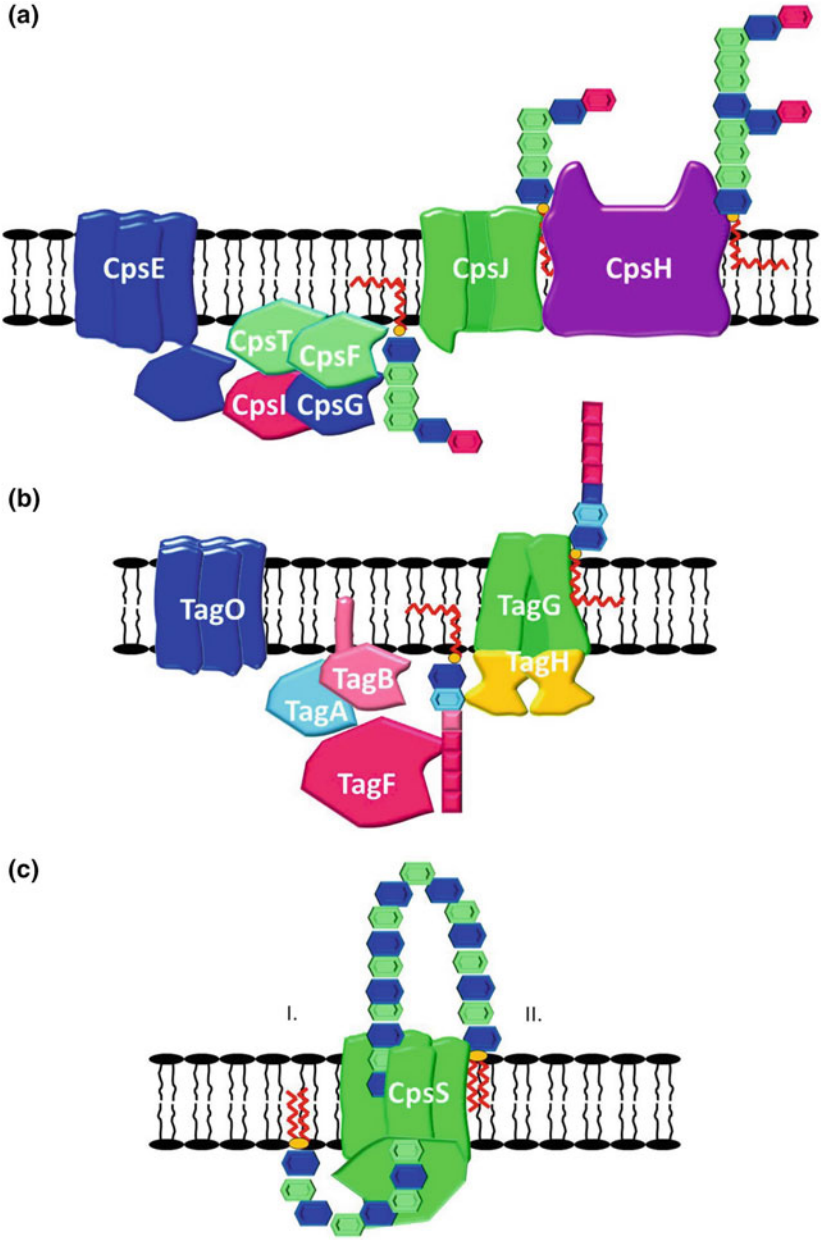
Does FtsW still play a role in lipid II flipping? It has been argued that the in vitro assay of FtsW did not address non-specific transport of lipid II (Butler et al. 2013). Perhaps, in a system where FtsW is the only candidate protein, it can mediate the translocation of lipid-linked sugars non-specifically in a way that is analogous to the flipping of non-native substrates by overexpressing Wzx proteins (see below). Analysis of the substrate specificity of FtsW flipping has been studied by making truncations or point mutations in the protein and subjecting it to a FRET-based in vitro assay (Mohammadi et al. 2014). After deleting 6 of the 10 TMS, the protein is still functional for lipid II flipping. It is intriguing that after removing almost half of this large membrane protein, it could still perform its proposed function. It is somewhat surprising that a protein with only 4 TMS would possess all the necessary structural characteristics to perform the flipping, given that all flippases have been characterized as large proteins. Nonetheless, it cannot be ruled out that the protein monomers could oligomerize to form a sufficiently large structure. No experiments have been performed to assess if these truncations could complement a temperature-sensitive FtsW protein, presumably because the C-terminal domain is important for septal localization and protein–protein interactions (Mohammadi et al. 2014). Evidence that these truncated constructs could retain some function in vivo would certainly substantiate the in vitro results. Alternatively, a detailed structure model based on computational biology methods or obtaining further experimental data may help to reconcile why these deletions did not affect the flippase activity of FtsW. Overall, the current level of understanding that FtsW possesses the activity for flipping lipid II is important; however, as discussed, more questions have been raised, and more experiments are warranted to more clearly define the system of lipid II flipping in Gram-positive bacteria.

In summary, the proteins involved in lipid II flipping remained elusive and a source of controversy. Thus far, the evidence presented in recent research on MurJ strongly supports the notion that it is the bona fide flippase in certain species in both Gram-positive and Gram-negative bacteria. However, backup systems appear to exist, for example, in *B. subtilis* when lipid II flipping can occur in the absence of MurJ. The backup protein Amj, a novel flippase, has very little resemblance to the MOP exporter superfamily proteins and thus it represents a novel family outside of or within this superfamily.

3 Capsular Polysaccharide Biosynthesis

The long-chain exopolysaccharide known as capsular polysaccharide (CPS) is a common structure that extends beyond the cell wall of Gram-negative and Gram-positive bacteria. It forms a surface coat to shield the bacteria against innate host defences and is known to contribute to virulence. In Gram-positive organisms, CPS can be linked to PG. Biosynthesis of this polymer occurs through two separate but conserved pathways: Wzx/Wzy-dependent and synthase dependent (Fig. 1) (Yother 2011). For either of the pathways, capsular biosynthesis is initiated on the inner face of the cytoplasmic membrane where sugar precursors are synthesized on a lipid carrier, i.e., Und-P for Wzx/Wzy-dependent and phosphatidylglycerol for the synthase-dependent one. Once the precursors are formed on the lipid carrier, the pathways differ by the mode of translocation across the cytoplasmic membrane. In the Wzx/Wzy pathway, these Und-P-linked precursors are transported to the outer face of the cytoplasmic membrane by the flippase CpsJ (an equivalent of Wzx, the O-antigen flippase in Gram-negative bacteria) and polymerized by CpsH (an equivalent of Wzy, the O-antigen polymerase). Capsule production, chain length, and attachment to the cell wall are controlled by CpsABCD, which together constitute a phosphoregulatory system (Guidolin et al. 1994; Kolkman et al. 1996; Morona et al. 1999). The synthase pathway, however, begins with phosphatidylglycerol-linked precursors, while subsequent steps, including initiation, polymerization, and export, are apparently performed by a single protein CpsS (synthase).

CPS biosynthesis has been extensively studied in Gram-negative organisms; however, the importance of capsules as virulence factors has inspired the study of this cell surface glycan in Gram-positive organisms with the ultimate goal of developing effective therapeutics (Tarahomjoo 2014). The capsular biosynthesis clusters have been extensively studied in *Staphylococcus aureus*, *Bacillus* spp. and *S. pneumoniae*. Differences in sugar constituents, side groups, and the molecular bonds between the sugar residues distinguish capsules into specific serotypes/serogroups within a particular species; e.g., *S. pneumoniae* alone possesses 93 distinct serotypes (Henrichsen 1995). The genes within the chromosomal locus of *S. pneumoniae* CPS is highly diverse, as reflected by the fact that its cluster ranges from 10 to 30 kb (Bentley et al. 2006). Comparatively, the *S. aureus* CPS clusters range only from 14 to 18 kb depending on the serotype (Moreau et al. 1990; Lin et al. 1994; Sau et al. 1997). The 5′ end of all CPS clusters are highly conserved beginning with the regulatory genes named *cpsABCD*, whose position and order are unchanged. The remainder of the cluster comprises glycosyltransferases and enzymes for synthesizing nucleotide sugar precursors (Bentley et al. 2006). In *S. pneumoniae*, the chromosomal locus of the CPS cluster is flanked by two genes, *dexE* and *ali*.


(a)
CpsE
CpsJ
CpsH
CpsT
CpsF
CpsI
CpsG
(b)
TagO
TagG
TagH
TagB
TagA
TagF
(c)
I.
II.
CpsS

◀ **Fig. 1** Biosynthesis and export of Gram-positive polysaccharides utilize three-distinct pathways: Wzx/Wzy dependent, ABC-transporter dependent, and synthase dependent. The glycosyltransferase enzymes are color coded to distinguish the various components including the respective substrates and the transport proteins (CpsJ, TagG, and CpsT, shown in *Green*). **a** Wzx/Wzy-dependent pathway for capsule biosynthesis. The example used is for *S. pneumoniae* serotype 2. CpsE (shown in *Blue*), the initiating glycosyltransferase, adds a Glc residue to the lipid-carrier undecaprenyl-phosphate (Und-P) allowing the subsequent addition of Rha, Glc, and GlcA to form the capsule-repeat unit. CpsJ (analogous to Wzx), the flippase protein, transports the lipid-linked unit to the outer face of the cytoplasmic membrane where it is polymerized to a desired length by CpsH (shown in *Purple*, analogous to Wzy), the capsule polymerase. **b** ABC-transporter-dependent biosynthesis pathway as seen in teichoic acid biosynthesis in *B. subtilis* 168. TagO (shown in *Blue*) is the initiating glycosyltransferase, which transfers GlcNAc to Und-P. TagA then transfers a ManNAc residue and allows for the subsequent addition of glycerol-3-phosphate (glycerol-3-P) residues by TagB and TagF. Once the glycerol-3-P chain reaches ~40 units, it is exported to the outside of the membrane by TagG and TagH, where TagH hydrolyzes ATP to drive export through the TagG channel. **c** synthase-dependent biosynthesis in *S. pneumoniae* serotype 3. The synthase pathway relies on the action of a single protein CpsS (shown in *Green*) to polymerize and export the polysaccharide composed of a glucuronic acid (GlcA) and glucose (Glc) disaccharide. There are two phases of synthase biosynthesis: (I) The polysaccharide chain is grown on the inner face of the cytoplasmic membrane with phosphatidylglycerol as the lipid linker. (II) Once the growing chain reaches 8 repeat units, a conformational change is proposed to occur to allow for the rapid polymerization of very-long-chain capsule molecules. Although the pathways vary, a lipid carrier is required for the initiation and each requires a designated transporter protein able to assist translocation of the long-chain polysaccharide across the hydrophobic membrane

3.1 Initiating Glycosyltransferase

The first serotype-specific gene identified in the capsule biosynthesis cluster of *S. pneumoniae* was *cpsE* (Guidolin et al. 1994; Kolkman et al. 1996, 1998; Morona et al. 1999; Pelosi et al. 2005), while in *S. aureus*, it was *capM* (Lin et al. 1994; Miyafusa et al. 2013). Both genes encode glycosyltransferases responsible for the reversible addition of a nucleotide-activated sugar to the Und-P lipid carrier (Kolkman et al. 1996; Cartee et al. 2005a, b). CpsE has been identified in 69 of the 98 serotypes of *S. pneumoniae*, and it is a 44-kDa protein that initiates capsular biosynthesis by transferring nucleotide-activated glucose-1-phosphate to Und-P (Bentley et al. 2006). The CpsE family of proteins possesses a high level of sequence identity between serotypes (70–90 %) (Kolkman et al. 1996; van Selm et al. 2002). The mechanism of action and the hydrophobic profile of CpsE proteins characterize them as members of the highly conserved membrane polyprenyl-phosphate hexose-1-phosphate transferase (PHPT) family (Valvano 2011). The topology of CpsE is shown to consist of 4 membrane-spanning domains, a predicted extracellular loop spanning between transmembrane segment (TMS) IV and V that are near the C-terminal cytoplasmic tail (Xayarath and Yother 2007; Saldías et al. 2008; Furlong et al. 2015). However, thus far, no experimental data are available in the literature to biochemically verify the orientation of the

extracellular loop, which is presumably involved in chain length regulation through protein–protein interactions (Xayarath and Yother 2007). Recently, the topology of WcaJ, the CpsE equivalent in *E. coli*, was determined experimentally using a dual-reporter system, whereby 3′ gene truncations are tagged to a reporter whose function is based on the subcellular localization of the residue, and cysteine scanning. With this approach, the authors localized the conserved loop to the cytoplasm; this is in contrast to the previous topology mapping based on in silico predictions which placed the conserved loop to be exposed on the extracellular face of the cytoplasmic membrane (Furlong et al. 2015). A truncated version of CpsE, which only consisted of the C′ terminal domain (residues 260–463) was still able to perform transfer of an activated sugar to the polyprenyl carrier. This region of CpsE maps to the cytoplasmic tail, demonstrating that this domain is the only portion essential for the transfer function (Pelosi et al. 2005).

Investigations into the lipid chain length required for CpsE transferase activity revealed relaxed specificity (35–105 repeats); however, endogenous lipids were utilized throughout the in vitro assays for consistency. It was determined that UDP-Glc is transferred in a *cis* orientation, meaning that the anomeric configuration of the sugar is retained, hence, CpsE performs a retaining glycosyltransferase reaction (Cartee et al. 2005a, b). Although the capsular biosynthesis proteins have been proposed to work in concert for the synthesis and assembly processes, currently no biochemical data have been collected to unequivocally show that the glycosyltransferases interact with one another (Cefalo et al. 2011). In serotypes where a *cpsE* gene is not present, the following proteins have been proposed to perform the initiation, including, WciI, WcjG, or WcjH, as determined by the position of the genes encoding these within their respective biosynthesis clusters (i.e., downstream from *cpsD*), and the strong similarity to the hydropathy profile of the carboxy terminal region of CpsE (Bentley et al. 2006). Genetic evidence has been collected on WcjG, and it showed that this protein has the activity to catalyze the transfer of galactose-1-phosphate which initiates capsular biosynthesis in *Streptococcus oralis* (Yang et al. 2009). The ability to regulate capsular biosynthesis is essential to the physiology of the bacterial cell, as mutations that sequester Und-P and prevent its recycling are generally lethal to both Gram-negative and Gram-positive organisms (Burrows and Lam 1999; Xayarath and Yother 2007). Therefore it is logical that mutations which affect protein function, when identified, would occur at the site of capsule initiation in order to exert an effect on CpsE in order to avoid committing Und-P to a non-functional pathway. As a case in point, deletions in genes that are localized downstream of *cpsE* in *S. pneumoniae* serotype 2, which include *cps2J* (*wzx*) or *cps2H* (*wzy*), cause suppression mutations in *cpsE*. These mutations localize to the extracellular loop and the cytoplasmic domain of CpsE. Further investigation determined that the amino acid residues substituted in the non-functional CpsE mutants were highly conserved among CpsE homologues from other *S. pneumoniae* serotypes and other bacterial genera thereby highlighting the essential nature of these amino acids to CpsE function (James et al. 2013). In addition, it was determined that the bond between the capsular sugar residues formed by the enzymatic activity of CpsE was rather labile, one that could be

hydrolyzed even under mild alkaline conditions (Cartee et al. 2005a, b). This property can be attributed to a response mechanism that would allow the bacterial cell to use Und-P reserves for other purposes in times of stress.

3.2 *Biosynthesis of* S. pneumoniae *Serotype 2 Capsule, Example of Downstream Genes*

The capsular biosynthesis cluster of *S. pneumoniae* serotype 2, regarded as the model organism of such research, has been subjected to extensive scrutiny through genetic and biochemical means. The capsular-repeat structure of serotype 2 has been elucidated and is composed of a tetrasaccharide of glucose → L-rhamnose → L-rhamnose → L-rhamnose with a glucose → glucuronic acid side chain off of the terminal rhamnose (Iannelli et al. 1999). A number of genes within the cluster are highly conserved among the 93 serotypes (Bentley et al. 2006). In the biosynthesis model, after the initial transfer of glucose-1-phosphate to the lipid carrier, the second gene from the clusters in at least 27 serotypes encodes CpsT, which is a β1-4 rhamnosyltransferase that catalyzes the transfer of L-Rha from its nucleotide-activated precursor dTDP-L-Rha to a D-Glc residue (James and Yother 2012). It should be noted that all glycosyltransferase assays that were described for this particular CPS biosynthesis pathway have been performed using membrane fractions of strains expressing CpsT instead of purified protein. The authors opined that using CpsT in situ in membranes is important for stabilizing the enzymes. The addition of L-Rha to the growing polymer was determined to be the "committed" step, i.e., the first irreversible reaction step of the pathway. This notion was substantiated by the lack of any suppressor mutants within a Δ*cpsT* background; this is unlike other genes downstream of *cpsT*, where such mutations have been isolated. Another line of evidence in support of CpsT as an initial glycosyltransferase in the capsule synthesis cascade is that although the majority of the suppressor mutations detected by Xayarath and Yother (2007) were found within *cpsE*, a single mutation was observed in *cpsL*, the first gene in the *cpsLMNO* cluster, which encodes the first enzyme in the multistep pathway conversion of Glc-6-phosphate to Glc-1-phosphate in the biosynthesis of dTDP-L-Rha; hence a mutation in *cpsL* clearly impacts on the function of CpsT by preventing access to the required substrate (James and Yother 2012). The glucose conversion pathway has been thoroughly studied in Gram-negative organisms where it is referred to as the Rml pathway (Rahim et al. 2000). To elongate the polysaccharide, the glycosyltransferases add sugars in a sequential manner in the order matching the organization of the genes in the CPS cluster; as such, *cpsF* encodes the enzyme that adds two L-Rha residues, and *cpsG* encodes a glucosyltransferase that adds the next D-Glc residue, the 4th sugar in the serotype 2 capsule subunit. A glucuronic acid-containing final product of the capsular biosynthesis has not been identified biochemically, and the authors of the study proposed that *cpsL* encodes the last putative

glycosyltransferase based on sequence similarity to other glycosyltransferases carrying similar function and the position of this particular gene within the cluster (James et al. 2013). Gleaning from the knowledge of CPS biosynthesis of *S. pneumonia* serotype 2 has helped to shed light on the interplay among many enzymes associated with synthesis and regulating the level of a certain substrate within the Gram-positive cell. After the biosynthesis steps in the cytoplasmic side of the bacterial membrane, the next logical step is the export of the nascent polysaccharide to the outer face of the membrane based on the activity of CpsJ, the analogous flippase to Wzx.

3.3 Polysaccharide Transporter Proteins: Wzx and CpsJ

Wzx/Wzy-dependent synthesis of oligosaccharides requires the transport of Und-P-linked glycans from the inner to outer face of the Gram-positive cytoplasmic membrane. The candidate enzyme for this action is the "flippase" CpsJ, a Wzx protein. Wzx proteins are classified as PST, a subgroup under the MOP exporter superfamily. There is evidence pointing to PST proteins as the evolutionary founders of this superfamily (Hvorup et al. 2003a, b). In general, Wzx proteins are predominantly hydrophobic and contain 12–14 TMS. This family of proteins may be difficult to identify by homology searches due to the immense sequence variation even between proteins from different serotypes of the same species. The study of Wzx proteins is further impeded by the difficulties in acquiring knockout mutations in the gene. It has been observed that *wzx* mutations are often lethal as Und-P is essential for various cell processes; hence, when Und-P becomes committed to the capsule synthesis but cannot be exported and ultimately recycled, it causes deleterious effects to the bacterial cell (Yother 2011). Therefore suppressor mutations in genes upstream of *wzx* in the biosynthesis cluster have been observed to block different stages of sugar synthesis (Xayarath and Yother 2007). At present, the knowledge concerning the structure and function of Wzx proteins in Gram-positives is lacking.

The recent advance made by He et al. (2010) in solving the crystal structure of NorM from *Vibrio cholera*e has greatly accelerated our efforts toward understanding the structure and mechanisms of activity of Wzx (and PST proteins in general). The 3D structure of NorM is the first to be solved for members of the MATE protein family, which also falls under the MOP exporter superfamily, and is closely related to PST proteins (He et al. 2010). MATE proteins have been shown to couple ion (Na^+ or H^+) binding/influx with drug export (Kuroda and Tsuchiya 2009). The discovery that NorM and Wzx share sequence/structure homology has prompted the testing of the hypothesis that Wzx uses a similar mechanism as NorM (Islam et al. 2012a, b). Islam et al. (2012a, b) used the NorM structure as a template and built a 3D structural model of Wzx from *P. aeruginosa* (Wzx_{Pa}). This model revealed a cationic lumen containing essential amino acid residues, which would presumably help bind the negatively charged O-unit of *P. aeruginosa*

lipopolysaccharide (LPS). In a subsequent report, Islam and colleagues were able to determine that H^+ but not Na^+ was the cation used by Wzx_{Pa} to cause an influx of ions into the proteoliposomes reconstituted with purified Wzx_{Pa} (Islam et al. 2013a, b). Such observations suggested that the protein mediates O-antigen flipping via an antiport mechanism consistent with the activities reported for members of the MATE family (He et al. 2010). Determining the structures of other Wzx proteins by either experimental or computational biology means would be crucial for understanding whether substrate-specific channels exist in other proteins (especially those identified in Gram-positive organisms) and by what mechanisms they might mediate flipping.

Early studies on the substrate specificity of Wzx proteins suggested that they only recognize the first sugar of the Und-P-linked glycan resulting in relaxed specificity (Feldman et al. 1999; Marolda et al. 2004). This seems surprising given the large sequence variation among Wzx proteins; one would expect to find conserved residues and motifs between flippases transporting glycans with the same initiating sugar. Data from more recent studies showed that Wzx proteins are actually highly specific for the substrate of their cognate system. In many cases, it appears that modifying the sugar side branches off the main glycan chain can hinder substrate flipping by Wzx. The system in *Erwinia amylovora* serves as a good example. This species decorates its exopolysaccharide (EPS) glycan unit with pyruvate as a side-branch substituent, yet its genome consists of two putative *wzx* genes for the transport of both pyruvylated and glucosylated capsular units (*amsL1* and *amsL2,* respectively) (Wang et al. 2012). Inactivation of *amsL1* resulted in a loss of EPS, suggesting transport was hindered in this strain. Interestingly, this particular mutant strain could be complemented by the expression of either a plasmid encoded *amsL1* or a gene that adds a terminal Glc, indicating that transport of the exopolysaccharide glycan unit could be facilitated by another flippase, likely AmsL2. Similarly, in another bacterial species, *Pantoea stewartii*, wzx_1 and wzx_2 appear to encode flippase proteins that facilitate the transport of pyruvylated- and glucosylated-EPS molecules, respectively, although this strain apparently only produces the glucosylated form of EPS. In the absence of wzx_1, there is no reduction in EPS production unless the gene for pyruvate capping is also expressed on a plasmid, presumably leading to a build up of glycan units that cannot be flipped by Wzx_2. When the gene for glucose capping was also mutated in a wzx_1-*minus* strain and the pyruvate-capping gene was expressed, EPS synthesis was completely abolished (Wang et al. 2012). Since that report, a study by the Reeves laboratory further substantiated that Wzx proteins possess a high level of substrate specificity with regards to LPS biosynthesis. These authors showed that firstly, in *Salmonella enterica* serogroups D2 and C2, Wzx is specific for the biosynthesis of dideoxyhexose side branch (Hong et al. 2012; Liu et al. 2015). Secondly, in a study on an *E. coli* O16 strain, chromosomal replacement of wzx_{O16} with *wzx* genes from other *E. coli* serotypes, *Salmonella enterica*, or *Shigella flexneri* resulted in either severely reduced or undetectable levels of LPS. However, if some of these *wzx* genes are introduced via a plasmid and overexpressed, they are able to flip O-units of either O16 or O111, suggesting that translocation of non-native substrates is

possible if a system is saturated with protein (Hong and Reeves 2014). Based on these observations, it was suggested that previous reports proposing Wzx as having low specificity was confounded by overexpression. Therefore, it appears that Wzx proteins are specific for their native substrate but can also flip non-native polysaccharide subunits when Wzx is overexpressed (Hong and Reeves 2014). In this case, a higher amount of the flippase protein compensates for the lower levels of activity toward a non-native substrate. Xayarath and Yother (2007) have observed that in *S. pneumoniae* serotype 2, a deletion mutation of *cps2k* (involved in side-chain synthesis) did not affect polymerization of long chains of CPS but could not rule out that Wzx flipping efficiency was affected. Whether Wzx substrate specificity in this Gram-positive system mirrors what has been observed in Gram-negatives remains to be seen. As mentioned earlier, very little is known about Wzx proteins from Gram-positive organisms; hence, one can only surmise how Wzx proteins would function based on the knowledge derived from studies from Gram-negative bacteria.

3.4 Lipid-Linked Polymerase Proteins: Wzy and CpsH

CpsH (the Gram-positive equivalent of Wzy) is the putative polymerase in the Wzx/Wzy-dependent pathway of capsule synthesis in Gram-positive bacteria (Yother 2011). Thus far, biochemical data on its function is lacking. Hence, it is the focus of this section to discuss a possible mechanism of action for polymerization in the better-characterized Gram-negative LPS biosynthesis system in an attempt to stimulate new experiments and development of potential methodologies that could be adapted to studying a protein such as CpsH in Gram-positives.

Similar to the case in *wzx*, characterizing the function of *wzy* has been difficult because there is negligible sequence homology between the encoded polymerases, even among different serotype strains of the same species. Thus far, the approach used by various laboratories to identify *wzy* genes is partly based on their location within a polysaccharide biosynthesis cluster and partly due to their large number of predicted TMS. Because putative *wzy* genes described in numerous publications in the literature are often never characterized, one must be cautious during database searches and not assume that such a protein possesses the polymerization function based on gene annotation by bioinformatics means alone (Islam and Lam 2014).

In a recent study, a landmark experiment was performed to reconstitute the O-polymerization activity of Wzy from *E. coli* O86 in vitro in the presence of purified Wzz and the native O-unit. The authors succeeded in producing LPS polymers, thereby confirming that Wzy is a bona fide O-polymerase (Woodward et al. 2010). Follow-up studies by the same group provided evidence that Wzy acts as a distributive enzyme, i.e., the polymerase is capable of extending the O-polymer through sequential rounds of catalysis and dissociation, as opposed to remaining bound to the polymer (Zhao et al. 2014). Despite these recent findings, knowledge of the exact mechanism of the polymerase reaction is lacking. One of the hurdles

has been the difficulty in identifying which particular domain of the Wzy protein resides on the outer face of the cytoplasmic membrane. Computer-based topology predictions often generate models with critical errors when it comes to studying Wzy proteins; particularly in terms of lacking the precision to localize key amino acid residues of the protein in the periplasm (Islam et al. 2010; Islam and Lam 2013, 2014). In studies of Wzy of *Pseudomonas aeruginosa* (Wzy_{Pa}), the experimentally-derived topology map revealed the presence of two large periplasmic loops called PL3 and PL5. Each of these loops contain an $RX_{10}G$ motif (Islam et al. 2010). Based on site-directed mutagenesis investigations, some of the Arg residues in this motif were deemed essential for polymerization and in some cases, they could not be substituted with another positively charged amino acid, Lys, suggesting that maintaining positive charge alone at these sites is not sufficient (Islam et al. 2011). Interestingly, other than His, Arg is the next most common amino acid implicated in sugar binding (Malik and Ahmad 2007; Elumalai et al. 2010; Islam et al. 2011). In addition, the observation that the predicted pI of PL3 is net-positive while PL5 is predicted to have a net-negative pI has led to the proposed "catch and release" model of O-antigen polymerization in Gram-negative bacteria. This model suggests that PL3 of Wzy_{Pa} is the grabbing arm that binds the negatively-charge O-units, while PL5 retains the polymer loosely (due to the net-negative charge but the presence of the $RX_{10}G$ motif) (Islam et al. 2011). This catch and release mechanism satisfies the requirements for a distributive mechanism suggested by Wang's group who studies Wzy_{EcO86} (Zhao et al. 2014). Exhaustive mutagenesis of wzy_{Pa} revealed a scarcity of critical residues outside of the $RX_{10}G$ motif in the protein, substantiating its importance (Islam et al. 2013a, b). Comparable results were obtained from the mutagenesis of wzy_{Sf} and topology studies of Wzy_{Sf} from *Shigella flexneri* serogroup 2a (Nath and Morona 2015).

Although the catch and release mechanism model still requires biochemical validation, it serves as a platform to stimulate future investigations. For example, a dual-loop topology is logical since the polymerase has to coordinate the addition of the glycan unit to the growing polymer. Therefore, looking for this feature in polymerases for Gram-positive Wzy proteins may provide a starting point for designing future experiments. The approach used to generate the Wzy_{Pa} topology has proven in other cases to closely correlate with the actual structure or the structure of a closely related protein (Alexeyev and Winkler 1999; Islam et al. 2012a, b).

3.5 The Phosphoregulatory System

In Gram-negative bacteria, Wzc and Wzb constitute a tyrosine kinase phospho-regulatory system that controls CPS production. Wzc contains an N-terminal transmembrane activator domain and a C-terminal cytoplasmic domain, which also contains Walker-A and Walker-B motifs essential for ATP binding and the autokinase activity of the protein. The C-terminal domain of Wzc is

dephosphorylated by a separate protein Wzb (Whitfield 2006). Wzc has been shown to associate with Wza and forms a channel between the inner and outer membrane that CPS passes through (Nickerson et al. 2014). In Gram-positives, the N- and C-terminal domains of Wzc are encoded by separate polypeptides, CpsC and CpsD, respectively, while CpsB, which is unrelated to Wzb, is the phosphatase (Morona et al. 2000; Yother 2011; Eberhardt et al. 2012). Based on comparison with the structures of CapAB in *Staphylococcus aureus*, CpsCD likely forms an octameric ring structure, which adopts a more open structure with increased phosphorylation facilitating changes in protein interactions (Olivares-Illana et al. 2008). Evidence has been provided by several studies that dysregulation of the phosphotyrosine system would lead to changes in both the level of capsule production and chain length (Morona et al. 2000; Bender et al. 2003; Morona et al. 2003; Geno et al. 2014).

In *S. pneumoniae*, capsule production is negatively associated with the phosphorylation of CpsD, such that deletion of CpsB or modification of essential tyrosine residues in CpsD causes a decrease in total capsule production (Morona et al. 2002). CpsB is a manganese-dependent enzyme, binding three cations that help stabilize the reaction intermediates, and although the structure of Cps4B from *S. pneumoniae* is quite distinct from Wzb of *E. coli* K30 strain (Wzb_{K30}), it is capable of dephosphorylating Wzc_{K30} (Morona et al. 2002; Hagelueken et al. 2009). CpsB of *Streptococcus, Staphylococcus,* and *Bacillus subtilis* form their own unique family within the polymerase and histidinol phosphatase superfamily (Aravind and Koonin 1998). CpsD is capable of autophosphorylation in the presence of ATP and CpsC. Note that transfer of phosphates from phosphorylated CpsD (CpsD $\sim$ P) to unphosphorylated CpsD can occur in the absence of ATP and CpsC. CpsB has been shown to regulate the phosphorylation of CpsD through its phosphatase activity and blocking of transphosphorylation by a phosphatase-independent mechanism, presumably protein–protein interactions (Bender and Yother 2001).

Geno et al. investigated the role of CpsB and its regulation of CPS biosynthesis in *S. pneumoniae* by examining the effects of low and high oxygen conditions on CPS biosynthesis and the level of CpsD phosphorylation in the presence and absence of CpsB (Geno et al. 2014). They found that under high oxygen conditions, *S. pneumoniae* strain D39 produced less CPS, while under low oxygen conditions more CPS was produced. The authors observed an increase in CpsB activity under high oxygen conditions, a property apparently innate to CpsB and independent of activities from other proteins. However, although CpsB activity appeared to correlate with a decreased level of CpsD phosphorylation (i.e., an increased activity resulted in less CpsD phosphorylation), somehow these two parameters were not connected to the level of CPS production. Rather, the binding of CpsB to CpsC or to other proteins was proposed to be a modulator of CPS biosynthesis under these conditions. This regulation of CPS has a direct implication in the progression of disease due to the pathogenesis of *S. pneumoniae* because in the nasopharynx (high oxygen condition), colonization is dependent on reduced capsule synthesis, whereas in systemic sites (low oxygen conditions), capsule is necessary for evasion of the host defences such as phagocytosis and the complement cascade (Geno et al. 2014).

A study by Toniolo et al. (2015) provides a comprehensive understanding of the functions of *cpsABCD* genes in the capsular biosynthesis in Group B *Streptococcus* (GBS). These authors constructed clean deletion mutants for each of these genes and generated site-specific mutation of functionally important residues/domains and used these to observe their effect on CPS synthesis. Briefly, disruption of individual *cps* genes, the cytoplasmic tail domain or the extracellular domain of CpsC, the extracellular domain of CpsA, or generating a K49A single amino acid change in CpsD, all resulted in reduced levels of CPS production. CpsA appears to play a role in attachment to the cell wall and chain length regulation since mutations in this gene causes an increase in the release of CPS into the medium, and the CPS molecules are much longer than that of the wild type. However, CpsA is not the sole determinant of ligation and other proteins may have the capacity to ligate CPS. The chain length of both peptidoglycan-linked and released CPS was longer in CpsD mutants, suggesting a dysregulation of chain length in these mutants. CpsC was found to interact with both CpsA and CpsD via a bacterial two-hybrid assay. Based on the results described thus far, a model for CPS biosynthesis in GBS was developed: the phosphorylation state of CpsD translates a signal to the extracellular domain of CpsC, which regulates attachment of CPS to the cell wall via CpsA. In a hyper-phosphorylated state, CpsA is allowed to proceed with ligation, whereas in the hypo-phosphorylated state, CpsA-mediated ligation would be blocked (Toniolo et al. 2015). This model is consistent with the one proposed for the phosphoregulatory system in *S. pneumoniae* (Morona et al. 2006).

3.6 Attachment of Sugar Polymers to the Cell Wall

The exact function of CpsA and its role in capsule expression remains elusive. CpsA belongs to the so-called LytR-CpsA-Psr (LCP) protein family. LCP proteins are present in all Gram-positive bacteria except for *Mollicutes* (which lacks a cell wall) and absent in most Gram-negatives (Hübscher et al. 2008). The function of LCP proteins is largely unknown. These proteins seem to be important for transcriptional regulation, cell envelope biogenesis, septum formation and biofilm formation (Hübscher et al. 2008). As described below, several studies have now implicated LCP proteins in the attachment of anionic polymers to the cell wall.

CpsA appears to have a regulatory role in capsule biosynthesis, at least in Group B *Streptococcus*, because it can bind the promoters upstream of *cpsA* and *cpsE*. In the extracellular domain, CpsA contains a DNA processivity factor domain, which, due to its location and actual sequence divergence from the family of DNA-binding sliding clamp proteins, is predicted to be involved in protein–protein interactions with other CPS biosynthesis proteins (Cieslewicz et al. 2001; Hanson et al. 2011; Rowe et al. 2015). A second domain, LytR, is the domain found in the LCP protein family that CpsA belongs to. Although the exact function of LCP proteins is unknown, some of the members of this family have been implicated to play a role in cell division, maintaining a proper cell wall, and cell morphology

(Hübscher et al. 2008). CpsA mutants in GBS lacking the LysR domain exhibit a phenotype of abnormally long bacterial chains and the cells have reduced capsule production. CpsA was found to localize to the division septum (Hanson et al. 2011; Rowe et al. 2015).

There is growing evidence that LCP proteins are involved in attachment of long polymers to the peptidoglycan. In the teichoic acid biosynthesis system, three of the proteins involved were found to belong to the LCP family, including TagT, TagU, and TagV (Kawai et al. 2011). These proteins could not be deleted in concert without causing lethality and defects in the attachment of teichoic acid to peptidoglycan. In the study by Kawai et al., the authors were successful in cocrystallizing the extracellular domain of Cps2A with phosphorylated polyisoprenoid lipids, revealing a binding site for the lipid-linked polymer. Cps2A and TagT undergo a phosphotransferase reaction in a magnesium-dependent manner and TagT and Cps2A exhibit pyrophosphatase activity of lipids during crystal formation (Kawai et al. 2011; Eberhardt et al. 2012). There is evidence that these proteins catalyze the removal of long-chain polymers from Und-P and link it to the cell wall. Further, in *S. aureus* strain Newman, three LCP proteins (A, B, and C) have been implicated in the attachment of wall teichoic acid (WTA) and capsule to peptidoglycan. Knockout mutations of all three genes that encode these proteins abolish retention of the polysaccharides on the surface. LcpA and LcpB apparently ligate WTA, while LcpC presumably functions to ligate capsules, but there is functional overlap between the proteins, such that the TA ligases can partially compensate for an LcpC mutation and vice versa (Chan et al. 2013, 2014).

3.7 Synthase-Dependent Pathway

The synthase-dependent pathway is utilized in the synthesis of *S. pneumoniae* serotype 3 and 37 capsules and in the synthesis of hyaluronic acid (HA), a glycosylaminoglycan, expressed in eukaryotes, *S. pyogenes*, and other less well characterized pathogens including *S. uberus* and *Pasteurella multocida* (Weigel 2002). In both capsule and HA biosynthesis, the enzyme responsible for the addition of the sugars also exports the polymer to the outside of the cytoplasmic membrane (Fig. 1). This system caters to a relatively simpler polysaccharide such as the one from serotype 3, GlcA(β1→4)Glc, as evidenced by the drastically smaller gene cluster containing 3 genes, *cps3D* (encodes UDP-glucose dehydrogenase), *cps3S* (encodes synthase), and *cps3U* (encodes glucose-1 phosphate uridyltransferase) (Bentley et al. 2006). The conserved genes at the 5′ end of the capsule cluster: *cpsABCP* are present but inactive. The tyrosine kinase *cpsD* was renamed *cpsP* within this serotype (Bentley et al. 2006). The protein coded by *cpsS*, the synthase, is a processive β-glycosyltransferase part of the GT-2 family with a conserved 4 TMS region and a cytoplasmic domain. The latter domain, as observed in other membrane bound glycosyltransferases, is the region responsible for the transferase function (Campbell et al. 1998; Bentley et al. 2006). The lipid carrier for

the synthase-dependent pathway in *S. pneumoniae* is phosphatidylglycerol, the most common lipid in the cytoplasmic membrane of the bacterium (Trombe et al. 1979); this is unlike the highly regulated Und-P. The abundance of phosphatidylglycerol lipid helps to explain the lack of lethal phenotypes when any of the three biosynthesis genes are interrupted (Dillard et al. 1995). Isolation of the lipid linker determined that initiation of the polymer begins with the addition of UDP-Glc to the lipid followed by UDP-glucuronic acid (UDP-GlcA) (Cartee et al. 2005a, b). Utilizing radio-labeled substrate and membranes containing the synthase protein became a standard for investigations concerning polymer length and substrate preference. Following initiation, UDP-GlcA and UDP-Glc are added sequentially to create an oligosaccharide of ~ 8 repeats (Cartee et al. 2001; Forsee et al. 2006). Once the key length is achieved, the oligosaccharide precursor becomes more tightly bound to the synthase resulting in a change from casual to processive synthesis, as evidenced by the inability to isolate intermediate polymer lengths by gel filtration chromatography (Forsee et al. 2006; Ventura et al. 2006). Glucose is present in abundance in the cytosol, therefore the chain length regulation and total amount of type 3 capsule synthesized is tightly bound to the cellular availability of UDP-GlcA (Forsee et al. 2006; Ventura et al. 2006; Forsee et al. 2009). This statement is supported by in vivo evidence that mutations which disrupt the active site of CpsD would inhibit or abrogate function and result in the bacteria synthesizing drastically shorter capsule repeats (Ventura et al. 2006). In addition, overexpression of a homologous *cspU* in *E. coli* K5, resulted in overall reduced levels of capsule but no change to the modality (Roman et al. 2003). Biochemical investigations into the synthase kinetics demonstrate that both sugars interact with the same binding site within CpsS as both UDP-Glc and UDP-GlcA possess similar binding and inhibition values; however, the binding of UDP-GlcA increases the affinity for UDP-Glc and alternatively the binding of UDP-Glc decreases affinity of UDP-GlcA demonstrating a processive process (Forsee et al. 2006). The evidence provided for the shift from oligosaccharide to polysaccharide suggests that the synthase may go through a conformational change, allowing faster and more controlled transfer of sugar subunits to the growing chain. The current model is that the lipid linker originates on the inner face of the cytoplasmic membrane and the initiating sugars are added until the octasaccharide is generated. The lipid linker is then flipped to face the outside and the polysaccharide can be synthesized freely into the cytoplasm (Cartee et al. 2005a, b; Forsee et al. 2006). However, thus far, no biochemical evidence has been obtained which would demonstrate the presence of the growing saccharide chain on the outside of the cytoplasmic membrane.

4 Teichoic Acid Biosynthesis

4.1 Wall Teichoic Acid Linkage Unit Initiation

In addition to capsule, another cell surface polymer that is essential for cell viability is teichoic acid (TA). The central role of TA is to assist in cell division and provides temporal and spatial regulation of PG stabilizing proteins such as penicillin binding protein 4 (Atilano et al. 2010). These polymers can be covalently linked to the cell wall: wall teichoic acid (WTA), or to cellular lipids: lipoteichoic acid (LTA). LTAs are differentiated into types based on the complexity of the polymer. There are currently five types of LTA isolated from different organisms. Unlike capsule, both WTA and LTA are often synthesized by an ABC-transporter-dependent pathway; however, the Type IV LTA is flipped to the periplasm by TacF, a member of the MATE protein family (Fig. 2). The basic structure of wall teichoic acid (WTA) is highly conserved among organisms and consists of a disaccharide, *N*-acetylmannosamine (β1→4) *N*-acetylglucosamine 1-phosphate, which acts as a base to allow for polymerization of either glycerol-3-Phosphate (glycerol-3-P) or

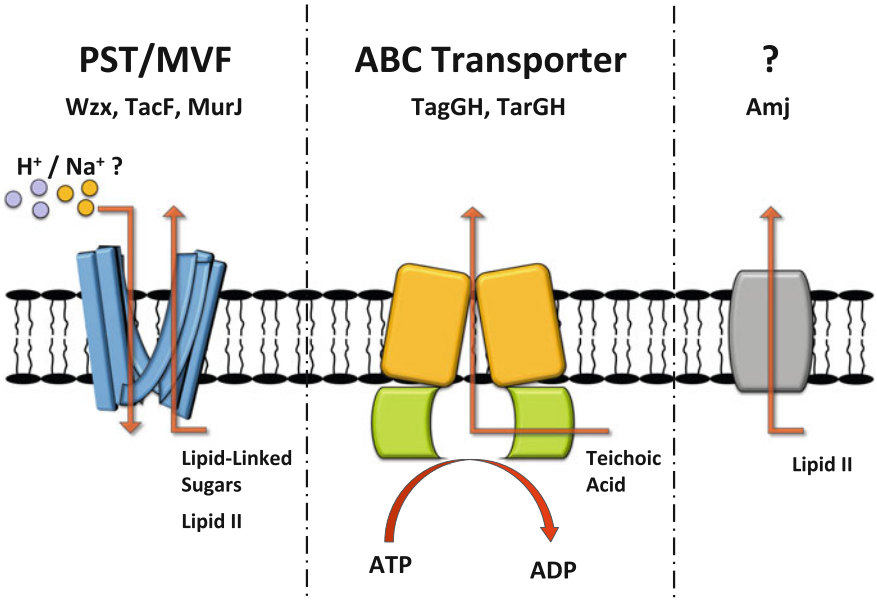


Fig. 2 Variations in mechanisms used by Gram-positive bacteria to transport lipid-linked sugars across the cytoplasmic membrane. *Left panel*, the polysaccharide transporter (PST) and mouse virulence factor (MVF) family proteins may possess a similar structure and H^+/Na^+ antiport mechanism that are consistent with features of multidrug and toxin extrusion (MATE) proteins. Middle panel, ATP-binding cassette (ABC) transporters contain transmembrane domains and nucleotide-binding domains that hydrolyzes ATP to provide energy for transport. *Right panel*, Amj is the sole member of a novel family of proteins that has been recently discovered

ribitol-5-Phosphate (ribitol-5-P), that are anchored to either the PG or the cytoplasmic membrane. The synthesis pathways for both forms of phosphorylated metabolites are highly conserved, although they differ in nomenclature: *tag* (for glycerol-3-P) and *tar* (ribitol-5-P). The model organism for studying WTA biosynthesis has been *Bacillus subtilis* 168a due to its simultaneous expression of two TA phenotypes (Shibaev et al. 1973; Mauël et al. 1989). The genes coding for WTA biosynthesis enzymes are separated into clusters *tagABC* and *tagDEF*, wherein it was determined that *tagABC* possessed hydrophobic residues, likely involved in transfer and export; whereas *tagDEF* code for soluble proteins, which the authors predicted to be associated with the glycerol-3-P precursor synthesis (Mauël et al. 1991). Investigation of the WTA biosynthesis locus is still at an early phase due in part to the intrinsic difficulty in producing high yield and high level of purity of the membrane proteins involved in this system, and in part to the need to synthesize the precise substrates linked to Und-P. Prior investigations into the synthesis of WTA began in 1989 (Honeyman and Stewart 1989) and spanned until 1995 (Lazarevic et al. 1995), by which time some of the genes have been identified and characterized at the biochemical level, while other genes have been annotated with putative functions by virtue of their localization within the cluster and by their similarities with other proteins of known function in GenBank databases. It was not until 2002, that a report on the characterization of the initiating enzyme, *tagO*, was published (Soldo et al. 2002). An ORF located downstream of the teichurionic acid cluster was identified to encode a 385-residue protein with 11 putative TMS, showing similarity to the *E. coli* Rfe, a UDP-*N*-acetylglucosamine: undecaprenyl-P *N*-acetylglucosaminyl 1-P transferase. Intriguingly, the topology of TagO that was presented at the time did not show the presence of a large soluble domain, which would be akin to transferase function. Both Rfe of *E. coli* and *cpsE* of *B. subtilis* share common characteristics, including having membrane domains and catalyzing a reaction that is reversible. This enzyme is apparently conserved across organisms; hence, by using a combination of molecular and biochemical approaches, the role of *tagO* in both teichoic and teichurionic acid biosynthesis was determined. As both teichoic and teichuronic acid biosynthesis requires Und-P, the highly regulated lipid linker, only conditional knockout mutants of *tagO* were made as disruption of constitutive genes would result in observed cell death. In studies that utilized the conditional knockout, a reduced amount of radiolabeled WTA was observed (Soldo et al. 2002). Coincidentally, the target of the antibiotic tunicamycin is undecaprenyl-PP-GlcNAc, the expected product of the reaction catalyzed by TagO. Cells grown in the presence of tunicamycin display a similar phenotype to the *tagO* mutants (Hancock et al. 1976; Lunderberg et al. 2015).

As in capsule biosynthesis, the second gene in the cluster *tagA* (*tarA*) encodes the enzyme that catalyzes the committed step within the TA biosynthesis pathway, i.e., the transfer of ManNAc from UDP-ManNAC to form a ManNAc-(β1 → 4)-GlcNAc disaccharide (Zhang et al. 2006). TagA is a soluble protein; therefore, researchers suggested that it would not recognize Und-P, and enzymatic assays developed for these proteins were done using more soluble lipids (Ginsberg et al. 2006). The WTA linkage unit is primed for polymerization by the action of TagB

(TarB), the third gene within the cluster associated with the three-part process for the biosynthesis of either glycerol-3-P or ribitol-5-P. The function of TagB, a glycerol phosphotransferase, was determined biochemically to be the primase of WTA biosynthesis by transferring a single glycerol-3-P unit to the distal ManNAc residue (Bhavsar et al. 2007). This addition prepares for the sequential downstream polymerization of glycerol-3-P and ribitol-5-P by TagF and TarF, respectively. TagB was first identified due to its similarity to TagF (Mauël et al. 1991), which recognizes the same substrate, and both proteins possess conserved His residues that are essential for function. At present, the mechanism associated with the His residue is unknown (Schertzer et al. 2005). TagB contains an N-terminal amphipathic helix, whose role is localization of TagB to the inner face of the CM in order to bring the protein into the vicinity of the other members of the biosynthesis machinery; and as such, TA biosynthesis is proposed to occur in a complex (Bhavsar et al. 2007). Biochemical assays involving TagB in the absence of detergents determined that the helix is solely involved in structural localization, and not the actual enzymatic reaction (Ginsberg et al. 2006; Bhavsar et al. 2007).

4.2 Glycerol-3-P Versus Ribitol-5-P Biosynthesis

Following the addition of a single glycerol-3-P to the lipid linker by TagB, the two pathways of WTA biosynthesis diverge at the point of long-chain polymerization by the protein TagF (TarF). Data from biochemical assays showed that TagF adds glycerol-3-P to up to 40–60 repeat units on the lipid linker of *B. subtilis* and TarF adds a single glycerol-3-P onto the lipid carrier in *S. aureus* (Brown et al. 2008; Pereira et al. 2008; Sewell et al. 2009). In accordance with the proteins described earlier in the TA pathway, TagF is intimately attached to the membrane through a two-helix domain near the N-terminal. Results from biochemical assays using the product of TagO, TagA, and TagB with a lipid tail, showed that purified TagF adds glycerol-3-P units in a non-processive manner (Sewell et al. 2009). As with TagB, TagF possesses two critical His residues (Schertzer et al. 2005). The crystal structure of TagF from *Staphylococcus epidermidis* positioned these His residues on either side of the substrate binding pocket where His^{584} binds the CDP-glycerol, whereas His^{444} is the active-site base (Lovering et al. 2010). In the ribitol-5-P biosynthesis in *S. aureus*, TarF acts as an additional primer by adding a single glycerol-3-P to the lipid intermediate in the cytoplasm. Reconstitution of the TarA, B, F, and L characterized by stepwise enzymatic assays on synthetic lipids determined that TarL is the ribitol-5-P polymerase which apparently is capable of transferring $\sim$40 repeat units to the lipid anchor (Brown et al. 2008). Alternatively in *B. subtilis* W23, an additional gene from the cluster, *tarK,* codes for a ribitol-5-P primer, which is crucial for allowing the *S. aureus* TarL to act to polymerize ribitol-5-P units (Brown et al. 2008, 2010).

4.3 *TagGH/TarGH is the ABC Transporter in Teichoic Acid Biosynthesis*

TagGH and TarGH are the two-component ABC transporter systems used by Gram-positive bacteria for the export of poly(glycerol-3-P) or poly(ribitol-5-P) WTA from the cytosol to the cell surface, respectively. TagG or TarG of either system constitutes the transmembrane transporter domains, while TagH or TarH represents the ATP-binding domain, which provides the energy for translocation of TA molecules (Brown et al. 2008). To date, little is known about the mechanisms of these transporters. It was suggested that these ABC transporters are able to transport TA with somewhat relaxed specificity and the evidence obtained thus far supports this (Kolkman et al. 1996). For instance, *B. subtilis* contains poly-glycerol phosphate WTA and its cognate ABC transporter is the TagGH system, while *S. aureus* contains a poly-ribitol phosphate WTA and its transporter is the TarGH system. The *tarGH* genes from *S. aureus* can be used to cross-complement a *tagGH*-deletion mutation of *B. subtilis*, suggesting that the Und-P lipid carrier and/or the linkage unit is recognized by these transporters, rather than the glycerol or ribitol moieties (Schirner et al. 2011). Recognition of the Und-P-linked moieties suggests the mechanism of transport involves flipping of the lipid-linked polymers, rather than adopting a simultaneous polymerization and transport process (Schirner et al. 2011; Brown et al. 2013).

4.4 *Teichoic Acid Biosynthesis in* S. pneumoniae *and Type IV Lipoteichoic Acid Biosynthesis*

Unlike WTA, lipoteichoic acid (LTA) is anchored to the membrane by the lipid moiety of the molecule (Fischer 1990). In *S. pneumoniae,* WTA and Type IV LTA share the same polymer structure and are proposed to be synthesized via the same pathway (Fischer and Tomasz 1985; Denapaite et al. 2012). This is an unusual system because most other bacteria synthesize WTA and LTA via different pathways. Further, *S. pneumoniae* LTA contains a complex repeat unit, whereas most other LTA structures are polymers of glycerol-3-P or ribitol-5-P (Denapaite et al. 2012). Current knowledge from the literature proposes that the polymer is made up of repeating units of a pseudopentasaccharide (2-acetamido-4-amino-2,4, 6-trideoxy-D-galactose, Glc, ribitol-5-P, and two residues of GlcNAc) that are decorated with phosphocholine on the terminal GlcNAc residues (Gisch et al. 2013). The model for this pathway is highly speculative because most of the enzyme functions have been inferred based on homology alignment with other proteins and have not been confirmed biochemically (Denapaite and Hakenbeck 2012). It has been proposed that the pseudopentasaccharide is built on Und-P, followed by the addition of phosphocholine and polymerization on the cytoplasmic face of the membrane. This is followed by the translocation of the polymer via TacF, the

proposed flippase for this system (Denapaite and Hakenbeck 2012). *S. pneumoniae* strictly requires exogenous choline for growth in order to add it to the TA chains, and mutations that block the addition of choline can be lethal (Tomasz 1967). In studies whereby *S. pneumoniae* cells were subjected to ethanolamine (a substitute for choline) depletion, several mutants were isolated that no longer required choline for growth and had acquired mutations in the *tacF* gene (Damjanovic et al. 2007; González et al. 2008). This has led to the proposal that these mutations might have changed the substrate specificity from being highly specific to becoming more relaxed. Under normal conditions, this high substrate specificity would ensure that only choline-containing TAs were incorporated. Interestingly, TacF belongs to the same PST family as Wzx proteins; therefore, these data are consistent with the current view that flippases are highly specific for their cognate substrate (Hvorup et al. 2003a, b; Damjanovic et al. 2007). Unfortunately, even less is known about the transport of lipid-linked sugars in the LTA synthesis pathways of other organisms. In *S. aureus,* LtaA is proposed to flip the lipid anchor (Glc_2-DAG) across the membrane so that LtaS can polymerize the GroP chains. LtaA belongs to the major facilitator superfamily. In *L. monocytogenes,* LafC is a predicted large membrane protein which acts downstream of glycolipid synthesis. However, the current evidence is not consistent with the role of flipping the glycolipid anchor (Webb et al. 2009; Reichmann and Gründling 2011). The protein(s) in *B. subtillis, S. agalactiae,* and *E. faecalis* involved in the transport of the glycolipid across the membrane are unidentified (Reichmann and Gründling 2011).

5 Teichuronic Acid Biosynthesis

The alternative cell-wall-linked polysaccharide teichuronic acid (TUA) is formed constitutively in *B. subtilis* strain W23 and conditionally in strain 168 specifically during times of phosphate starvation. In order to maintain cell viability in phosphate-limiting environments, the cellular levels of TA, which is highly phosphorylated, are decreased and synthesis of TUA is activated. (Ellwood and Tempest 1969). TUA is an anionic polymer comprised of negatively charged polysaccharides, which varies from species to species. The *B. subtilis* 168 structure is $[GalNAc\text{-}GlcA]_n$ and that of *Micrococcus luteus* is $[\rightarrow 4)\beta\text{-D-}ManNAcAp\text{-}(1\rightarrow 6)\text{D-}Glc\text{-}\alpha\text{-}(1\rightarrow]_n$, where *n* is the number of disaccharide repeats (Hase and Matsushima 1972). Notice the distinct lack of phosphate groups in comparison with the TA structure. However, the charge of the polymer remains the same, indicating that an overall anionic charge is essential for cell-wall stability (Grant 1979). Although intense studies aiming at elucidating the chemical structure of this polymer began in the 1970s, the synthesis pathway had not been investigated until 1999. A 9.3-kb cluster consisting of 8 genes, *tuaA-H*, was found to be essential for TUA biosynthesis (Soldo et al. 1999). As mentioned above, *tagO* is responsible for producing Und-PP-*N*-acetylglucosamine, the first component of the polymer (Johnson et al. 1984). Currently very little is known about the enzymes responsible for TUA

synthesis. However, molecular manipulations and sequence similarities have painted a picture of the potential role for each of these enzymes within the biosynthesis pathway. The committed step of TUA biosynthesis is proposed to be the reaction catalyzed by an enzyme encoded by *tuaA,* which displays high levels of similarity to other glycosyltransferases that catalyze the linkage between a glycosyl group and undecaprenyl (Soldo et al. 1999). Clearly, more work is warranted for identifying the sugar transferred by TuaA and determining whether this is the committing step of TUA biosynthesis. The regulation of the *tua* cluster occurs through the PhoP-PhoR regulon (Hulett 1996), a two-component regulation system wherein PhoP is the response regulator (Seki et al. 1987) and PhoR is the histidine kinase (Seki et al. 1988). High levels of the TagF-generated TA intermediates suppress phosphorylation of PhoR. The function of TagF requires a large pool of cytoplasmic phosphate as it polymerizes a chain of ~40 glycerol-3-P units. Under conditions where phosphate is limiting, TagF would be unable to produce such long chains, thereby releasing PhoP to be activated. This concept is logical as any previous steps would not be significantly hampered by low phosphate levels (Botella et al. 2014). Once PhoR is activated, it phosphorylates PhoP into PhoP~P. The downstream effect is the decrease of TA synthesis by PhoP~P binding to the *tagA* and *tagD* operons (Qi and Hulett 1998) while also upregulating transcription of the *tua* operon as *tua* possesses a PhoP~P promoter site (Liu nd Hulett 1998). The TA and TUA biosynthesis pathways appear to utilize similar precursors where TA uses TagD (CDP-glycerol) and TUA relies on the product of TuaD and UDP-glucuronate. Cell-free extracts of a wild type and Δ*tuaD* mutant were subjected to a UDP-glucose-6-dehydrogenase assay, wherein oxidation of UDP-Glc resulted in the formation of UDP-GlcA (Pagni et al. 1999). However, so far this is the only protein in the TUA biosynthesis pathway that has been characterized biochemically. Hence, the knowledge of TUA biosynthesis is clearly lacking. The remainder of the genes in the TUA biosynthesis cluster have not been investigated at the molecular or biochemical level; however, based on the sequence similarity to amino sequences uploaded to the GenBank databases, the resulting biosynthesis pathway is predicted to be as follows: TuaC, TuaG, and TuaH show sequence homology to glycosyltransferases from various Gram-negative organisms associated with the metabolic steps of forming lipid-linked precursors (Soldo et al. 1999). The genes associated with the export of TUA have been proposed to be TuaB, TuaE, and TuaF. The membrane topology of each protein was predicted using in silico methods, the results showed each containing multiple TMS, and as such they can be classified as integral membrane proteins. The authors proposed these proteins to be analogous to those found in the Wzx/Wzy-dependent pathways, as each of the proteins display sequence similarity to Wzx (TuaB), Wzy (TuaE), and Wzz(TuaF), respectively (Soldo et al. 1999). However, there is insufficient molecular and biochemical characterization of these genes to determine whether TUA biosynthesis follows the Wzx/Wzy-dependent pathway.

In *M. luteus*, solubilized membranes have been used to demonstrate enzyme activities that catalyze the transfer of a single UDP-GlcNAc onto Und-P and the subsequent addition of UDP-glucose and UDP-ManNAcA to form the anionic

disaccharide (Traxler et al. 1982; Hildebrandt and Anderson 1990). The enzyme responsible for both catalytic reactions was purified from solubilized membranes and observed to run at ~440 kDa on Native PAGE. In subsequent SDS-PAGE experiments, this apparent multimeric protein was resolved into two proteins with apparent molecular masses of 54 and 52.5 kDa, respectively. However, the monomeric form of the proteins did not display any activity associated with the synthesis of a long-chain disaccharide. A model has been proposed that the *M. luteus* teichuronic acid is produced by a synthase pathway wherein four glucosyltransferase and four ManNAcA transferase subunits make an alternating octameric structure which simultaneously polymerizes and exports the polysaccharide (Deng et al. 2010). It is important to point out that this is a highly speculative model that is not substantiated by experimental data. For instance, the authors apparently have purified the complex, but had not made further attempts to identify the proteins within the complex. Recently, the whole-genome sequencing of a *M. luteus* strain has been attained (Accession number: AMYK00000000.2). This should provide future groups with ample information to decipher the genetic components for the TUA biosynthesis pathway in this species (Ghosh et al. 2013).

6 Conclusion

In conclusion, polysaccharides of Gram-positive organisms are synthesized through three separate pathways: Wzx/Wzy-dependent, ABC-transporter-dependent, and synthase-dependent pathways. Although conservation of these pathways is observed between Gram-negative and Gram-positive organisms, there is much to be learned about the key components of each of the pathways in the Gram-positive species. More specifically, one must overcome the obvious challenges including the expression and purification of integral membrane proteins, and the development of biochemical and biophysical methods to characterize the function of this particular group of proteins that are presumably involved in the assembly and transport of the polysaccharides across the Gram-positive cell envelope. Obtaining high yield and high purity of proteins is prerequisite for attempting high-resolution structural studies. The importance of these polysaccharides in virulence and cell viability makes them legitimate targets for novel antimicrobial screens; hence, these challenges/hurdles are worth pursuing.

Acknowledgments Research in the Lam laboratory was supported by funding from the Canadian Institutes of Health Research (CIHR) (grant# MOP-14687) to J.S.L. V.L.T. is a recipient of a Cystic Fibrosis Canada Doctoral Studentship and a Queen Elizabeth II Graduate Student Scholarships in Science and Technology, S.M.H. is a recipient of an Ontario Graduate Scholarship and a CIHR Canada Graduate Scholarship Masters, and J.S.L. holds a Canada Research Chair in Cystic Fibrosis and Microbial Glycobiology.

References

Alexeyev MF, Winkler HH (1999) Membrane topology of the *Rickettsia prowazekii* ATP/ADP translocase revealed by novel dual pho-lac reporters. J Mol Biol 285:1503–1513

Aravind L, Koonin EV (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. Nucleic Acids Res 26:3746–3752

Atilano ML, Pereira PM, Yates J, Reed P, Veiga H, Pinho MG, Filipe SR (2010) Teichoic acids are temporal and spatial regulators of peptidoglycan cross-linking in *Staphylococcus aureus*. PNAS 107:18991–18996

Bender MH, Yother J (2001) CpsB is a modulator of capsule-associated tyrosine kinase activity in *Streptococcus pneumoniae*

Bender MH, Cartee RT, Yother J (2003) Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae* positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae*. J Bacteriol 185:6057–6066

Bentley SD, Aanensen DM, Mavroidi A et al (2006) Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet 2:e31

Bhavsar AP, D'Elia MA, Sahakian TD, Brown ED (2007) The amino terminus of *Bacillus subtilis* TagB possesses separable localization and functional properties. J Bacteriol 189:6816–6823

Botella E, Devine SK, Hubner S et al (2014) PhoR autokinase activity is controlled by an intermediate in wall teichoic acid metabolism that is sensed by the intracellular PAS domain during the PhoPR-mediated phosphate limitation response of *Bacillus subtilis*. Mol Microbiol 94:1242–1259

Brown S, Zhang Y-H, Walker S (2008) A Revised Pathway Proposed for *Staphylococcus aureus* wall teichoic acid biosynthesis based on in vitro reconstitution of the intracellular steps. Chem Biol 15:12–21

Brown S, Meredith T, Swoboda J, Walker S (2010) *Staphylococcus aureus* and *Bacillus subtilis* W23 make polyribitol wall teichoic acids using different enzymatic pathways. Chem Biol 17:1101–1110

Brown S, Santa Maria JP, Walker S (2013) Wall teichoic acids of gram-positive bacteria. Annu Rev Microbiol 67:313–336

Burrows LL, Lam JS (1999) Effect of wzx (rfbX) mutations on A-band and B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* O5. J Bacteriol 181:973–980

Butler EK, Davis RM, Bari V, Nicholson PA, Ruiz N (2013) Structure-function analysis of MurJ reveals a solvent-exposed cavity containing residues essential for peptidoglycan biogenesis in *Escherichia coli*. J Bacteriol 195:4639–4649

Butler EK, Tan WB, Joseph H, Ruiz N (2014) Charge requirements of Lipid II flippase activity in *Escherichia coli*. J Bacteriol 196:4111–4119

Campbell JA, Davies GJ, Bulone V, Henrissat B (1998) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J 329:719

Cartee RT, Forsee WT, Jensen JW, Yother J (2001) Expression of the *Streptococcus pneumoniae* type 3 synthase in *Escherichia coli:* assembly of th type 3 polysaccharide on a lipid primer. J Biol Chem 276:48831–48839

Cartee RT, Forsee WT, Yother J (2005a) Initiation and synthesis of the *Streptococcus pneumoniae* type 3 capsule on a phosphatidylglycerol membrane anchor. J Bacteriol 187:4470–4479

Cartee RT, Forsee WT, Bender MH, Ambrose KD, Yother J (2005b) CpsE from type 2 *Streptococcus pneumoniae* catalyzes the reversible addition of Glucose-1-Phosphate to a polyprenyl phosphate acceptor, initiating type 2 Capsule repeat unit formation. J Bacteriol 187:7425–7433

Cefalo AD, Broadbent JR, Welker DL (2011) Protein–protein interactions among the components of the biosynthetic machinery responsible for exopolysaccharide production in *Streptococcus thermophilus* MR-1C. J Appl Microbiol 110:801–812

Chan YGY, Frankel MB, Dengler V, Schneewind O, Missiakasa D (2013) *Staphylococcus aureus* mutants lacking the lytr-cpsa-Psr family of enzymes release cell wall teichoic acids into the extracellular medium. J Bacteriol 195:4650–4659

Chan YGY, Kim HK, Schneewind O, Missiakas D (2014) The capsular polysaccharide of *Staphylococcus aureus* is attached to peptidoglycan by the LytR-CpsA-Psr (LCP) family of enzymes. J Biol Chem 289:15680–15690

Cieslewicz MJ, Kasper DL, Wang Y, Wessels MR (2001) Functional analysis in type Ia group B Streptococcus of a cluster of genes involved in extracellular polysaccharide production by diverse species of Streptococci. J Biol Chem 276:139–146

Damjanovic M, Kharat AS, Eberhardt A, Tomasz A, Vollmer W (2007) The Essential *tacF* gene is responsible for the choline-dependent growth phenotype of *Streptococcus pneumoniae*. J Bacteriol 189:7105–7111

Denapaite D, Hakenbeck R (2012) Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species. Microb Drug Resist 18:344–358

Denapaite D, Brückner R, Hakenbeck R, Vollmer W (2012) Biosynthesis of teichoic acids in *Streptococcus pneumoniae* and closely related species: lessons from genomes. Microb Drug Resist. 18:344–358

Deng LL, Alexander AA, Lei S, Anderson JS (2010) The cell wall teichuronic acid synthetase (TUAS) Is an enzyme complex located in the cytoplasmic membrane of *Micrococcus luteus*. Biochem Res Int 2010:395758

Dillard JP, Vandersea MW, Yother J (1995) Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. J Exp Med 181:973–983

Eberhardt A, Hoyland CN, Vollmer D et al (2012) Attachment of capsular polysaccharide to the cell wall in *Streptococcus pneumoniae*. Microb Drug Resist (Larchmont, N.Y.) 18:240–255

Ellwood DC, Tempest DW (1969) Control of teichoic acid and teichuronic acid biosynthesis in chemostat cultures of *Bacillus subtilis* var. niger. Biochem J 111:1–5

Elumalai P, Rajasekaran M, Liu HL, Chen C (2010) Investigation of cation-Pi interactions in sugar-binding proteins. Protoplasma 247:13–24

Fay A, Dworkin J (2009) Bacillus subtilis homologs of MviN (MurJ), the putative *Escherichia coli* lipid II flippase, are not essential for growth. J Bacteriol 191:6020–6028

Feldman MF, Marolda CL, Monteiro MA, Perry MB, Parodi AJ, Valvano MA (1999) The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in *Escherichia coli* O antigen assembly is independent of the chemical structure of the O repeat. J Biol Chem 274:35129–35138

Fischer W (1990) Bacterial phosphoglycolipids and lipoteichoic acids. In: Kates M (ed) Glycolipids, phosphoglycolipids, and sulfoglycolipids, vol 6. Springer, US, pp 123–234

Fischer H, Tomasz A (1985) Peptidoglycan cross-linking and teichoic acid attachment in *Streptococcus pneumoniae*. J Bacteriol 163:46–54

Forsee WT, Cartee RT, Yother J (2006) Role of the carbohydrate binding site of the *Streptococcus pneumoniae* capsular polysaccharide type 3 synthase in the transition from oligosaccharide to polysaccharide synthesis. J Biol Chem 281:6283–6289

Forsee WT, Cartee RT, Yother J (2009) Characterization of the lipid linkage region and chain length of the cellubiuronic acid capsule of *Streptococcus pneumoniae*. J Biol Chem 284:11826–11835

Furlong SE, Ford A, Albarnez-Rodriguez L, Valvano MA (2015) Topological analysis of the *Escherichia coli* WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. Sci Rep 5:9178. doi:10.1038/srep09178

Geno KA, Hauser JR, Gupta K, Yother J (2014) *Streptococcus pneumoniae* phosphotyrosine phosphatase CpsB and alterations in capsule production resulting from changes in oxygen availability. J Bacteriol 196:1992–2003

Ghosh A, Chaudhary SA, Apurva SR et al (2013) Whole-genome sequencing of *Micrococcus luteus* strain modasa, of Indian origin. Genome Announc 1:e00076–e00013

Ghuysen JM, Hakenbeck R (1994) Bacterial cell wall. Elsevier Science, Amsterdam
Ginsberg C, Zhang Y-H, Yuan Y, Walker S (2006) In vitro reconstitution of two essential steps in wall teichoic acid biosynthesis. ACS Chem Biol 1:25–28
Gisch N, Kohler T, Ulmer AJ et al (2013) Structural reevaluation of *Streptococcus pneumoniae* lipoteichoic acid and new insights into its immunostimulatory potency. J Biol Chem 288:15654–15667
González A, Llull D, Morales M, García P, García E (2008) Mutations in the *tacF* gene of clinical strains and laboratory transformants of *Streptococcus pneumoniae*: impact on choline auxotrophy and growth rate. J Bacteriol 190:4129–4138
Grant WD (1979) Cell wall teichoic acid as a reserve phosphate source in *Bacillus subtilis*. J Bacteriol 137:35–43
Guidolin A, Morona JK, Morona R, Hansman D, Paton JC (1994) Nucleotide sequence analysis of genes essential for capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* type 19F. Infect Immun 62:5384–5396
Hagelueken G, Huang H, Mainprize IL, Whitfield C, Naismith JH (2009) Crystal structures of Wzb of *Escherichia coli* and CpsB of *Streptococcus pneumoniae*, representatives of two families of tyrosine phosphatases that regulate capsule assembly. J Mol Biol 392:678–688
Hancock IC, Wiseman G, Baddiley J (1976) Biosynthesis of the unit that links teichoic acid to the bacterial wall: inhibition by tunicamycin. FEBS Lett 69:75–80
Hanson BR, Lowe BA, Neely MN (2011) Membrane topology and DNA-binding ability of the Streptococcal CpsA protein. J Bacteriol 193:411–420
Hartley MD, Imperiali B (2012) At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. Arch Biochem Biophys 517:83–97
Hase S, Matsushima Y (1972) Structural studies on a glucose-containing polysaccharide obtained from cell walls of *Micrococcus lysodeikticus*: III. Determination of the structure. J Biochem 72:1117–1128
He X, Szewczyk P, Karyakin A, Evin M, Hong W-X, Zhang Q, Chang G (2010) Structure of a cation-bound multidrug and toxic compound extrusion transporter. Nature 467:991–994
Henrichfreise B, Schiefer A, Schneider T et al (2009) Functional conservation of the lipid II biosynthesis pathway in the cell wall-less bacteria chlamydia and wolbachia: why is lipid II needed? Mol Microbiol 73:913–923
Henrichsen J (1995) Six newly recognized types of *Streptococcus pneumoniae*. J Clin Microbiol 33:2759–2762
Hildebrandt KM, Anderson JS (1990) Biosynthetic elongation of isolated teichuronic acid polymers via glucosyl- and N-acetylmannosaminuronosyl transferases from solubilized cytoplasmic membrane fragments of *Micrococcus luteus*. J Bacteriol 172:5160–5164
Höltje J-V (1998) Growth of the stress-bearing and shape-maintaining murein sacculus of *Escherichia coli*. Microbiol Mol Biol Rev 62:181–203
Honeyman AL, Stewart GC (1989) The nucleotide sequence of the *rodC* operon of *Bacillus subtilis*. Mol Microbiol 3:1257–1268
Hong Y, Reeves PR (2014) Diversity of O-Antigen repeat unit structures can account for the substantial sequence variation of Wzx translocases. J Bacteriol 196:1713–1722
Hong Y, Cunneen MM, Reeves PR (2012) The Wzx translocases for *Salmonella enterica* O-antigen processing have unexpected serotype specificity. Mol Microbiol 1–11
Hübscher J, Lüthy L, Berger-bächi B, Stutzmann P (2008) Phylogenetic distribution and membrane topology of the LytR-CpsA-Psr protein family. BMC Genom 16:1–16
Hulett FM (1996) The signal-transduction network for Pho regulation in *Bacillus subtilis*. Mol Microbiol 19:933–939
Hvorup RN, Winnen B, Chang AB, Jiang Y, Zhou X-F, Saier MH (2003a) The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. Eur J Biochem 270:799–813

Hvorup RN, Winnen B, Chang AB, Jiang Y, Zhou X-F, Saier MH (2003b) The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. Fed Eur Biochem Soc J 270:799–813

Iannelli F, Pearce BJ, Pozzi G (1999) The type 2 capsule locus of *Streptococcus pneumoniae*. J Bacteriol 181:2652–2654

Inoue A, Murata Y, Takahashi H, Tsuji N, Fujisaki S, Kato JI (2008) Involvement of an essential gene, mviN, in murein synthesis in *Escherichia coli*. J Bacteriol 190:7298–7301

Islam ST, Lam JS (2013) Topological mapping methods for α-helical bacterial membrane proteins—an update and a guide. MicrobiologyOpen 2:350–64

Islam ST, Lam JS (2014) Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can J Microbiol 60:697–716

Islam ST, Taylor VL, Qi M, Lam JS (2010) Membrane topology mapping of the O-Antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa* PAO1 reveals novel domain architectures. mBio 1(3):e00189–10. doi: 10.1128/mBio.00189-10

Islam ST, Gold AC, Taylor VL, Anderson EM, Ford RC, Lam JS (2011) Dual conserved periplasmic loops possess essential charge characteristics that support a catch-and-release mechanism of O-antigen polymerization by Wzy in *Pseudomonas aeruginosa* PAO1. J Biol Chem 286:20600–20605

Islam ST, Fieldhouse RJ, Anderson EM, Taylor VL, Keates RAB, Ford RC, Lam JS (2012a) A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in Pseudomonas aeruginosa PAO1. Mol Microbiol 84:1165–1176

Islam ST, Fieldhouse RJ, Anderson EM, Taylor VL, Keates RAB, Ford RC, Lam JS (2012b) A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in *Pseudomonas aeruginosa* PAO1. Mol Microbiol 84:1165–1176

Islam ST, Huszczynski SM, Nugent T, Gold AC, Lam JS (2013) Conserved-residue mutations in Wzy affect O-antigen polymerization and Wzz-mediated chain-length regulation in *Pseudomonas aeruginosa* PAO1. Sci Rep 3:3441. doi: 10.1038/srep03441

Islam ST, Eckford PDW, Jones ML, Nugent T, Bear CE, Vogel C, Lam JS (2013) Proton-dependent gating and proton uptake by Wzx support O-antigen-subunit antiport across the bacterial inner membrane. mBio 4(5): e00678–13. doi: 10.1128/mBio.00678-13

James DBA, Yother J (2012) Genetic and biochemical characterizations of enzymes involved in *Streptococcus pneumoniae* serotype 2 capsule synthesis demonstrate that Cps2T (WchF) catalyzes the committed step by addition of β1-4 rhamnose, the second sugar residue in the repeat unit. J Bacteriol 194:6479–6489

James DBA, Gupta K, Hauser JR, Yother J (2013) Biochemical activities of *Streptococcus pneumoniae* serotype 2 capsular glycosyltransferases and significance of suppressor mutations affecting the initiating glycosyltransferase Cps2E. J Bacteriol 195:5469–5478

Johnson GL, Hoger JH, Ratnayake JH, Anderson JS (1984) Characterization of three intermediates in the biosynthesis of teichuronic acid of *Micrococcus luteus*. Arch Biochem Biophys 235:679–691

Kawai Y, Marles-Wright J, Cleverley RM et al (2011) A widespread family of bacterial cell wall assembly proteins. EMBO J 30:4931–4941

Kolkman MA, Morrison DA, Van Der Zeijst BA, Nuijten PJ (1996) The capsule polysaccharide synthesis locus of *Streptococcus pneumoniae* serotype 14: identification of the glycosyl transferase gene cps14E. J Bacteriol 178:3736–3741

Kolkman MAB, van der Zeijst BAM, Nuijten PJM (1998) Diversity of capsular polysaccharide synthesis gene clusters in *Streptococcus pneumoniae*. J Biochem 123:937–945

Kuroda T, Tsuchiya T (2009) Multidrug efflux transporters in the MATE family. Biochim Biophys Acta 1794:763–768

Lazarevic V, Mauël C, Soldo B, Freymond P-P, Margot P, Karamata D (1995) Sequence analysis of the 308° to 311° segment of the *Bacillus subtilis* 168 chromosome, a region devoted to cell wall metabolism, containing non-coding grey holes which reveal chromosomal rearrangements. Microbiology 141:329–335

Lin WS, Cunneen T, Lee CY (1994) Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. J Bacteriol 176:7005–7016

Liu W, Hulett FM (1998) Comparison of PhoP binding to the *tuaA* promoter with PhoP binding to other Pho-regulon promoters establishes a *Bacillus subtilis* Pho core binding site. Microbiology 144:1443–1450

Liu MA, Stent TL, Hong Y, Reeves PR (2015) Inefficient translocation of a truncated O unit by a Salmonella Wzx affects both O-antigen production and cell growth. FEMS Microbiol Lett 1–7

Lovering AL, Lin LYC, Sewell EW, Spreter T, Brown ED, Strynadka NCJ (2010) Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. Nat Struct Mol Biol 17:582–589

Lunderberg JM, Liszewski Zilla M, Missiakas D, Schneewind O (2015) *Bacillus anthracis* tagO Is required for vegetative growth and secondary cell wall polysaccharide synthesis. J Bacteriol 197:3511–3520

Malik A, Ahmad S (2007) Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network. BMC Struct Biol 7:1

Manat G, Roure S, Auger R, Bouhss A, Barreteau H, Mengin-Lecreulx D, Touzé T (2014) Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. Microb Drug Resist 20:199–214

Marolda CL, Vicarioli J, Valvano MA (2004) Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. Microbiology (Reading, England) 150:4095–4105

Mauël C, Young M, Margot P, Karamata D (1989) The essential nature of teichoic acids in *Bacillus subtilis* as revealed by insertional mutagenesis. Mol Gen Genet 215:388–394

Mauël C, Young M, Karamata D (1991) Genes concerned with synthesis of poly(glycerol phosphate), the essential teichoic acid in *Bacillus subtilis* strain 168, are organized in two divergent transcription units. J Gen Microbiol 137:929–941

Meeske AJ, L-T Sham, Kimsey H, B-M Koo, Gross CA, Bernhardt TG (2015) MurJ and a novel lipid II flippase are required for cell wall biogenesis in *Bacillus subtilis*. PNAS 112:6437–6442

Miyafusa T, Caaveiro JMM, Tanaka Y, Tsumoto K (2013) Dynamic elements govern the catalytic activity of CapE, a capsular polysaccharide-synthesizing enzyme from *Staphylococcus aureus*. FEBS Lett 587:3824–3830

Mohammadi T, van Dam V, Sijbrandi R et al (2011) Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. The EMBO J 30:1425–1432

Mohammadi T, Sijbrandi R, Lutters M et al (2014) Specificity of the transport of lipid II by FtsW in *Escherichia coli*. J Biol Chem 289:14707–14718

Moreau M, Richards JC, Fournier J-M, Byrd RA, Karakawa WW, Vann WF (1990) Structure of the type 5 capsular polysaccharide of *Staphylococcus aureus*. Carbohydr Res 201:285–297

Morona JK, Morona R, Paton JC (1999) Analysis of the 5′ portion of the type 19A capsule locus identifies two classes of *cpsC*, *cpsD*, and *cpsE* Genes in *Streptococcus pneumoniae*. J Bacteriol 181:3599–3605

Morona JK, Paton JC, Miller DC, Morona R (2000) Tyrosine phosphorylation of CpsD negatively regulates capsular polysaccharide biosynthesis in *Streptococcus pneumoniae*. Mol Microbiol 35:1431–1442

Morona JK, Morona R, Miller DC, Paton JC (2002) *Streptococcus pneumoniae* capsule biosynthesis protein CpsB is a novel manganese-dependent phosphotyrosine-protein phosphatase. J Bacteriol 184:577–583

Morona JK, Morona R, Miller DC, Paton JC (2003) Mutational analysis of the carboxy-terminal (YGX)4 repeat domain of CpsD, an autophosphorylating tyrosine kinase required for capsule biosynthesis in *Streptococcus pneumoniae*. J Bacteriol 185:3009–3019

Morona JK, Morona R, Paton JC (2006) Attachment of capsular polysaccharide to the cell wall of *Streptococcus pneumoniae* type 2 is required for invasive disease. Proc Natl Acad Sci USA 103:8505–8510

Nath P, Morona R (2015) Mutational analysis of the major periplasmic loops of *Shigella flexneri* Wzy: identification of the residues affecting O antigen modal chain length control, and Wzz-dependent polymerization activity. Microbiology 161:774–785

Nickerson NN, Mainprize IL, Hampton L, Jones ML, Naismith JH, Whitfield C (2014) Trapped translocation intermediates establish the route for export of capsular polysaccharides across *Escherichia coli* outer membranes. PNAS 111:8203–8208

Nikolaidis I, Favini-Stabile S, Dessen A (2014) Resistance to antibiotics targeted to the bacterial cell wall. Protein Sci 23:243–259

Olivares-Illana V, Meyer P, Bechet E et al (2008) Structural basis for the regulation mechanism of the tyrosine kinase CapB from *Staphylococcus aureus*. PLoS Biol 6:1321–1331

Pagni M, Lazarevic V, Soldo B, Karamata D (1999) Assay for UDPglucose 6-dehydrogenase in phosphate-starved cells: gene tuaD of Bacillus subtilis 168 encodes the UDPglucose 6-dehydrogenase involved in teichuronic acid synthesis. Microbiology 145:1049–1053

Pelosi L, Boumedienne M, Saksouk N, Geiselmann J, Geremia RA (2005) The glucosyl-1-phosphate transferase WchA (Cap8E) primes the capsular polysaccharide repeat unit biosynthesis of *Streptococcus pneumoniae* serotype 8. Biochem Biophys Res Commun 327:857–865

Pereira MP, Schertzer JW, D'Elia MA, Koteva KP, Hughes DW, Wright GD, Brown ED (2008) The wall teichoic acid polymerase TagF efficiently synthesizes poly(glycerol phosphate) on the TagB product lipid III. ChemBioChem 9:1385–1390

Qi Y, Hulett FM (1998) Role of PhoP ~ P in transcriptional regulation of genes involved in cell wall anionic polymer biosynthesis in *Bacillus subtilis*. J Bacteriol 180:4007–4010

Rahim R, Burrows LL, Monteiro MA, Perry MB, Lam JS (2000) Involvement of the rml locus in core oligosaccharide and O polysaccharide assembly in *Pseudomonas aeruginosa*. Microbiology 146:2803–2814

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

Roman E, Roberts I, Lidholt K, Kusche-Gullberg M (2003) Overexpression of UDP-glucose dehydrogenase in *Escherichia coli* results in decreased biosynthesis of K5 polysaccharide. Biochem J 374:767–772

Rowe HM, Hanson BR, Runft DL, Lin Q, Firestine SM, Neely MN (2015) Modification of the CpsA protein reveals a role in alteration of the *Streptococcus agalactiae* cell envelope. Infect Immun 83:1497–1506

Rudnick PA, Arcondéguy T, Kennedy CK, Kahn D (2001) glnD and mviN are genes of an essential operon in *Sinorhizobium meliloti*. J Bacteriol 183:2682–2685

Ruiz N (2008) Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in *Escherichia coli*. PNAS 105:15553–15557

Saldías MS, Patel K, Marolda CL, Bittner M, Contreras I, Valvano MA (2008) Distinct functional domains of the *Salmonella enterica* WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. Microbiology 154:440–453

Sau S, Bhasin N, Wann ER, Lee JC, Foster TJ, Lee CY (1997) The *Staphylococcus aureus* allelic genetic loci for serotype 5 and 8 capsule expression contain the type-specific genes flanked by common genes. Microbiology 143:2395–2405

Schertzer JW, Bhavsar AP, Brown ED (2005) Two conserved histidine residues are critical to the function of the TagF-like family of enzymes. J Biol Chem 280:36683–36690

Schirner K, Stone LK, Walker S (2011) ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers. ACS Chem Biol 407–412

Seki T, Yoshikawa H, Takahashi H, Saito H (1987) Cloning and nucleotide sequence of *phoP*, the regulatory gene for alkaline phosphatase and phosphodiesterase in *Bacillus subtilis*. J Bacteriol 169:2913–2916

Seki T, Yoshikawa H, Takahashi H, Saito H (1988) Nucleotide sequence of the *Bacillus subtilis phoR* gene. J Bacteriol 170:5935–5938

Sewell EWC, Pereira MP, Brown ED (2009) The wall teichoic acid polymerase TagF is non-processive in vitro and amenable to study using steady state kinetic analysis. J Biol Chem 284:21132–21138

Sham LT, Butler EK, Lebar MD, Kahne D, Bernhardt TG, Ruiz N (2014) MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis. Science 345:220–222

Shibaev VN, Duckworth M, Archibald AR, Baddiley J (1973) The structure of a polymer containing galactosamine from walls of *Bacillus subtilis 168*. Biochem J 135:383–384

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

Soldo B, Lazarevic V, Karamata D (2002) tagO is involved in the synthesis of all anionic cell-wall polymers in *Bacillus subtilis* 168. Microbiology 148:2079–2087

Tarahomjoo S (2014) Recent approaches in vaccine development against *Streptococcus pneumoniae*. J Mol Microbiol Biotechnol 24:215–227

Tomasz A (1967) Choline in the cell wall of a bacterium: novel type of polymer-linked choline in Pneumococcus. Science (New York, N.Y.) 157:694–697

Toniolo C, Balducci E, Romano MR et al (2015) *Streptococcus agalactiae* capsule polymer length and attachment is determined by the proteins CpsABCD. J Biol Chem 290:9521–9532

Traxler CI, Goustin AS, Anderson JS (1982) Elongation of teichuronic acid chains by a wall-membrane preparation from *Micrococcus luteus*. J Bacteriol 150:649–656

Trombe M-C, Lanéelle M-A, Lanéelle G (1979) Lipid composition of aminopterin-resistant and sensitive strains of *Streptococcus pneumoniae* effect of aminopterin inhibition. Biochim Biophys Acta 574:290–300

Valtersson C, van Duÿn G, Verkleij AJ, Chojnacki T, de Kruijff B, Dallner G (1985) The influence of dolichol, dolichol esters, and dolichyl phosphate on phospholipid polymorphism and fluidity in model membranes. J Biol Chem 260:2742–2751

Valvano MA (2011) Common themes in glycoconjugate assembly using the biogenesis of O-antigen lipopolysaccharide as a model system. Biochem (Moscow) 76:729–735

van Selm S, Kolkman MAB, van der Zeijst BAM, Zwaagstra KA, Gaastra W, van Putten JPM (2002) Organization and characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* serotype 9V. Microbiology 148:1747–1755

Ventura CL, Cartee RT, Forsee WT, Yother J (2006) Control of capsular polysaccharide chain length by UDP-sugar substrate concentrations in *Streptococcus pneumoniae*. Mol Microbiol 61:723–733

Wang X, Mansourian AR, Quinn PJ (2008) The effect of dolichol on the structure and phase behaviour of phospholipid model membranes. Mol Membr Biol 25:547–556

Wang X, Yang F, von Bodman SB (2012) The genetic and structural basis of two distinct terminal side branch residues in stewartan and amylovoran exopolysaccharides and their potential role in host adaptation. Mol Microbiol 83:195–207

Webb AJ, Karatsa-Dodgson M, Gründling A (2009) Two-enzyme systems for glycolipid and polyglycerolphosphate lipoteichoic acid synthesis in *Listeria monocytogenes*. Mol Microbiol 74:299–314

Weigel PH (2002) Functional characteristics and catalytic mechanisms of the bacterial hyaluronan synthases. IUBMB Life 54:201–211

Whitfield C (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. Annu Rev Biochem 75:39–68

Woodward R, Yi W, Li L et al (2010) In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. Nat Chem Biol 6:418–423

Xayarath B, Yother J (2007) Mutations blocking side chain assembly, polymerization, or transport of a Wzy-dependent *Streptococcus pneumoniae* capsule are lethal in the absence of suppressor mutations and can affect polymer transfer to the cell wall. J Bacteriol 189:3369–3381

Yang J, Ritchey M, Yoshida Y, Bush CA, Cisar JO (2009) Comparative structural and molecular characterization of ribitol-5-phosphate-containing *Streptococcus oralis* coaggregation receptor polysaccharides. J Bacteriol 191:1891–1900

Yother J (2011) Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. Annu Rev Microbiol 65:563–581

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

Zhao G, Wu B, Li L, Wang P (2014) O-antigen polymerase adopts a distributive mechanism for lipopolysaccharide biosynthesis. Appl Microbiol Biotechnol 98:4075–4081

Predicting Subcellular Localization of Proteins by Bioinformatic Algorithms

Henrik Nielsen

Abstract When predicting the subcellular localization of proteins from their amino acid sequences, there are basically three approaches: signal-based, global property-based, and homology-based. Each of these has its advantages and drawbacks, and it is important when comparing methods to know which approach was used. Various statistical and machine learning algorithms are used with all three approaches, and various measures and standards are employed when reporting the performances of the developed methods. This chapter presents a number of available methods for prediction of sorting signals and subcellular localization, but rather than providing a checklist of which predictors to use, it aims to function as a guide for critical assessment of prediction methods.

Abbreviations

ANN	Artificial neural network
BLAST	Basic local alignment search tool
GO	Gene Ontology
HMM	Hidden Markov model
MCC	Matthews correlation coefficient
PWM	Position-weight matrix
SP	Signal peptide
SCL	Subcellular localization
SVM	Support vector machine
TMH	Transmembrane α-helix

H. Nielsen (✉)
Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, Kemitorvet building 208, 2800 Lyngby, Denmark
e-mail: hnielsen@cbs.dtu.dk

Current Topics in Microbiology and Immunology (2017) 404:129–158
DOI 10.1007/82_2015_5006

Published Online: 04 January 2016

Contents

1 Introduction

Prediction of subcellular localization (SCL) of proteins in both eukaryotic and prokaryotic cells has a long history in bioinformatics. The first attempts at predicting the best known sorting signals, the transmembrane α-helix (TMH) and the secretory signal peptide (SP), from the amino acid sequence were published long before bioinformatics was even established as a field (Kyte and Doolittle 1982; von Heijne 1983). Since then, a plethora of methods for predicting sorting signals and SCL has been published, and it can be a daunting task to select the most relevant and reliable methods for analyzing a set of sequences.

Of course, the development of algorithms and the growth in available training data has led to an increase in the predictive performance of the available methods. A decade ago, some of the authors of the PSORTb method for predicting SCL in bacteria (Gardy et al. 2005) even concluded that "on average, recent high-precision computational methods such as PSORTb now have a lower error rate than laboratory methods" (Rey et al. 2005). This conclusion should be taken with a grain of salt; firstly, it applies only to high-throughput laboratory methods; secondly, it should be remembered that computational methods will never be better than the data used to train them. Nevertheless, Rey et al. had a point regarding the experimental sources of error which can easily render a high-throughput experiment less reliable than a well-trained computational method.

It can be difficult, however, to decide what to believe when the authors of every computational method tend to describe their performance as being superior to all others. There are different ways of defining the problem, different ways of measuring the performance, and different prerequisites used for prediction. The aim of this chapter is not to provide a definite answer to which method is best for which problem—such a checklist would quickly become outdated—but instead to install in the reader a toolbox for critically evaluating bioinformatics algorithms. This will involve a number of examples of computational methods selected for their rele-

vance for Gram-positive bacteria. In general, prediction methods are only mentioned if they either provide publicly available Web servers or have strong historical relevance.

2 Three Approaches to Prediction

It is crucial to understand that there are basically three different approaches to predict protein SCL from the amino acid sequence. The first approach is recognition of the actual sorting signals, e.g., SPs, TMHs, or LPXTG-like motifs for cell wall attachment. The above-mentioned early methods for TMH and SP recognition (Kyte and Doolittle 1982; von Heijne 1983) were examples of this. A number of more recent examples is given in Sects. 5, 7, and 8.

The second approach is prediction based on global properties of the proteins, e.g., their amino acid composition. This approach was first used to discriminate between intracellular and extracellular proteins in both prokaryotic end eukaryotic proteins (Nakashima and Nishikawa 1994). It has been shown that the main part of the differences in amino acid composition between intracellular and extracellular proteins resides in the surface-exposed amino acids, which makes sense since the surfaces should be adapted to the varying physicochemical environments of the different SCLs (Andrade et al. 1998). This analysis was done for eukaryotic proteins only, but it would be fair to assume that the observation holds true also for bacterial proteins.

Two early SCL prediction methods which used only the amino acid composition were NNPSL (Reinhardt and Hubbard 1998), and SubLoc[1] (Hua and Sun 2001), based on artificial neural networks (ANN) and support vector machines (SVM), respectively (see Sect. 3). They were limited in their applicability, because their dataset did not include any membrane proteins, and they did not distinguish between Gram-positive and Gram-negative bacteria.

Using only the amino acid composition for prediction of course throws away all sequence information, including possible signatures of actual sorting signals. One way to retain some of this information while still keeping a fixed number of parameters is to count the occurrences of amino acid pairs, either adjacent or separated by a small distance. Nakashima and Nishikawa (1994) thus found that including composition of amino acid pairs with a separation distance of up to five positions improved predictive performance.

The third approach is prediction by sequence homology. When trying to predict functional aspects of an unknown protein, the standard procedure is to do a BLAST search (Altschul et al. 1997) and then infer such aspects from the functional annotations of the found homologues. Therefore, the intuitive expectation is that such a procedure will also work for SCL—in other words, that a protein tends to

[1] http://www.bioinfo.tsinghua.edu.cn/SubLoc/.

stay in the same compartment in the course of evolution. Indeed, a significant part of the "subcellular location" annotations of Swiss-Prot (the manually annotated part of UniProt (The UniProt Consortium 2015)) are found with "sequence similarity" as the evidence (several times as many as the corresponding annotations with experimental evidence).

However, it is not trivial to determine how similar a pair of proteins has to be in order to perform an inference about SCL. Nair and Rost (2002a), working with eukaryotic proteins only, concluded that more than 70 % identical residues in a pairwise BLAST search are needed to correctly infer SCL for 90 % of the query proteins. On the other hand, the authors of the CELLO method for both eukaryotes and bacteria (Yu et al. 2006) found that SCL prediction by a simple BLAST search was better than a machine learning method above a pairwise identity cutoff as low as 30 %.

The simplest possible homology-based prediction is the direct transfer of annotation from the best BLAST hit, i.e., the query protein is used to search a database of proteins with experimentally known SCLs, and then the SCL of the best hit is assigned to the query. However, more advanced approaches to homology-based prediction are also possible, using indirect means to infer SCL from the annotation of homologues which do not necessarily have experimentally known SCLs. This annotation could be derived from keywords or functional descriptions (Nair and Rost 2002b; Lu et al. 2004), or titles and/or abstracts of the literature references (Shatkay et al. 2007; Briesemeister et al. 2009).

In this context, it should be mentioned that many signal-based and global property-based methods use a BLAST search to build a profile of related sequences in order to enhance the prediction. This does not make these methods homology-based, since they do not use the annotations of the found hits.

In addition to the three approaches described here, it is of course possible to construct hybrids of them. Most of the multi-category methods described in Sect. 9 are of the hybrid type.

When comparing methods based on one of the three approaches, it is important to realize that each has its strengths and weaknesses. Homology-based methods, or hybrid methods containing homology-based components, often present the best measured performances, but the performance depends critically on the source of the query protein. Organisms that have been subject to intense research will naturally tend to have more high-quality annotations, so proteins from those and their close relatives will find more close homologues with richer annotations from which to make predictions, while predictions for less well-studied organisms will suffer from lack of annotations of close homologues. This is typically not taken into consideration when reporting the predictive performances of such methods. Signal-based and global property–based methods should be expected to be less sensitive to the source of the query protein, unless the signals in the training data are very organism specific.

There are two advantages to using global property- or homology-based methods. First, they can be used also for those compartments where the actual sorting signals

are not known, or are too poorly characterized to support a prediction method. Second, they may work for sequences that are fragments from which the actual sorting signal may be missing, or for amino acid sequences derived from genomic sequence where the start codon of the protein has not been correctly predicted, thus obscuring any N-terminal sorting signals. On the downside, global property- or homology-based methods do not provide the same degree of insight into the information processing in the cell, since they ignore which parts of the sequence are actually important for sorting. Another drawback is that such methods will not be able to distinguish between very closely related proteins that differ in the presence or absence of a sorting signal, and they will not be able to predict the effects of small mutations that destroy or create a sorting signal.

3 Algorithms for Prediction

A rich variety of computational algorithms have been used in the prediction of SCL from amino acid sequence. Common to all of them is that they take a number of sequence-derived inputs and produce an output which can be the presence or absence of a sorting signal (for signal-based predictors), or an assignment of the protein into one of a number of possible SCL classes (for multi-category predictors). For Gram-positive bacteria, the number of SCL classes is most often defined as four (cytoplasm, membrane, cell wall, and extracellular). It may be discussed whether it makes sense to define a periplasmic compartment in Gram positives; if it does, it is not a rich source of proteins: Presently, UniProt carries only two examples of Gram-positive proteins located in the periplasm with experimental evidence.

Some algorithms (e.g., sequence alignment and Hidden Markov Models (HMM)) are naturally designed to work with sequences, while others (e.g., ANNs and SVMs) take only a fixed number of input values. When working with the latter category, one can either input the sequence as a series of overlapping windows of fixed length (typical for signal-based predictors), or extract a fixed number of features from the sequence (typical for global property-based predictors).

Numerical prediction algorithms can roughly be divided into two groups: statistical and machine learning, although it can sometimes be a matter of definition where to draw the distinction. Both classes of methods have a number of free parameters that must be estimated from the data, but while the parameters in statistical methods can be calculated directly, machine learning methods depend on an iterative optimization process where parameters are gradually changed until the classification error has reached a minimum.

The simplest sequence pattern recognition method is the *consensus sequence*, e.g., “LPXTG” for cell wall localization. It is easy to check whether such a pattern is present in a sequence, but it is also a very crude method, because it defines absolute requirements for certain amino acids at certain positions and only provides “yes” or “no” answers. The pattern “LPXTG,” for example, ignores the fact that the last amino acid in the motif may sometimes be “N” instead of “G.” This could be

accommodated by the use of degenerate positions where one out of a set of amino acids is allowed, turning the pattern into what is known as a *regular expression*. For example, "LPXTG" could be relaxed to "LPXT[GN]," but that expression would not be able to distinguish the typical "G" from the atypical "N." Furthermore, the pattern ignores the fact that not all amino acids are equally probable in the position marked "X."

An alternative to the consensus sequence or regular expression is the *position-weight matrix* (PWM) (Stormo et al. 1982), a statistical window-based method which is very useful for characterizing and predicting short sequence motifs. The procedure when constructing a PWM is to start with a set of examples of the motif of interest—the training set—and count the occurrences of all amino acids at each position. The counts are used for estimating the probability of each amino acid at each position, and this probability is divided by the background probability of that amino acid in proteins in general, and the weight is then calculated as the logarithm of the ratio of the probabilities (it is therefore also known as a log-odds score). The score for a new sequence window can then be calculated by looking up the weights for each amino acid in each position in the window and adding them up. In this way, the weight matrix can give a *quantitative* answer to how well a sequence window fits the pattern.

A graphical counterpart to the PWM is the *sequence logo* (Schneider and Stephens 1990), where each position is summarized by a stack of letters. The height of each stack is equivalent to the information content—a measure of the conservation—of that position, while the height of each letter is proportional to the probability of the corresponding amino acid at that position. Examples of sequence logos are shown in Sects. 5 and 8.

A straightforward extension of the PWM is the *sequence profile*, which allows for insertions and deletions in the sequence and therefore can model motifs of variable length. It is possible to formulate a profile in probabilistic terms—then it becomes an HMM (Krogh et al. 1994). An HMM is basically a generative model where each state can emit an amino acid following a certain probability distribution, and the transitions between the states are also governed by probabilities. A profile HMM contains a number of match states corresponding to the typical length of the domain or motif to be modeled, plus insert states that can emit extra amino acids and delete states that do not emit any amino acids but serve to skip one or more match states. A graphical depiction of a profile HMM is reported in Fig. 1.

An HMM is a machine learning algorithm, since the emission and transition probabilities are found by an iterative optimization process (expectation maximization) from a set of training data. After training, new sequences can be evaluated in terms of the probability that the sequence was generated by the model (a process known as decoding the HMM).

HMMs are part of a broader class of computational models called *graphical models*. Other examples relevant to biological sequence analysis are Bayesian networks (Yu et al. 2010b) and conditional random fields (Chang et al. 2015). Bayesian networks generally specify dependencies between variables in a directed acyclic graph. The simplest possible Bayesian network is the Naïve Bayes

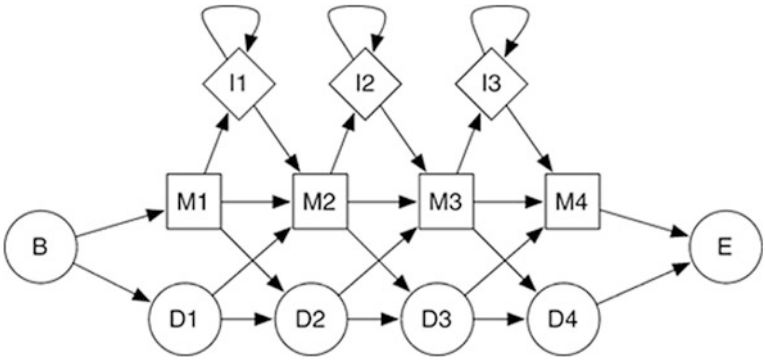


Fig. 1 Schematic drawing of a profile HMM. The model can generate an amino acid sequence by starting in the *begin state* (marked "B") and, following the arrows, concluding in the *end state* (marked "E"). A *match state* (marked "M") or an *insert state* (marked "I") emits an amino acid according to a probability distribution, while a *delete state* (marked "D") does not emit anything but functions as an option for skipping the corresponding match state. If only the match states are used, the generated sequence will be four amino acids long; it can become longer by using the insert states or shorter by using the delete states

classifier, which assumes that all the input variables are independent. It can have surprisingly good performance also in cases where the assumption of independence is known to be violated (Rish 2001), and it is sometimes preferred over more advanced machine learning methods because it offers the opportunity to explain exactly which input variables were important for each prediction (Szafron et al. 2004; Briesemeister et al. 2010).

Several publicly available databases specialize in creating and storing profiles for protein families or domains. Among these are PROSITE[2] (Sigrist et al. 2013), which contains both regular expression patterns and PWM-like profiles, and Pfam[3] (Finn et al. 2014) and TIGRFAMs[4] (Haft et al. 2013), which are both databases of profile HMMs. InterPro[5] (Mitchell et al. 2015) is a special case, since it does not create its own profiles, but collect profiles from a number of contributing databases, including PROSITE, Pfam, and TIGRFAMs. Most profiles in these databases are evolutionarily related families and/or domains, but there are also instances of functional motifs that are similar due to common selection pressure rather than common descent. Among these are a few protein sorting motifs, which can be used as prediction tools—examples is given in Sects. 5 and 8.

To predict the presence of a PROSITE motif in your own sequences, use the ScanProsite service[6] (de Castro et al. 2006), choose "Option 3," and then enter your

[2] http://prosite.expasy.org/prosite.html.

[3] http://pfam.xfam.org/.

[4] http://www.jcvi.org/cgi-bin/tigrfams/index.cgi.

[5] http://www.ebi.ac.uk/interpro/.

[6] http://prosite.expasy.org/scanprosite/.

sequences and the identifier of the PROSITE entry you wish to scan for. Pfam does not offer the ability to search new sequences against a specific profile; you can only scan against the entire database using the "search" function.[7] TIGRFAMs does not have its own search function, but in the hmmscan tool[8] which is part of the HMMER Web server (Finn et al. 2015), you can select TIGRFAMs as (one of) your database(s). Alternatively, you can use InterProScan (Jones et al. 2014) which offers both Pfam and TIGRFAMs among its member databases. The Web server[9] only allows submission of one sequence at a time, but you can download the software and databases to your own computer.

It should be emphasized that not all HMMs are profile HMMs—any grammar that can be described as a diagram of connected states can be modeled as an HMM. For example, a cyclic HMM can describe a repeating pattern, and a branched HMM can describe a choice between alternative patterns.

Another machine learning algorithm which is widely used in biological sequence analysis—and which can also be described in terms of graphical models—is the ANN (Hertz et al. 1991). ANNs are inspired by the way networks of biological neurons are connected; the input patterns are presented to one or more layers of artificial "neurons" that compute a weighted sum of their inputs and apply a non-linear function to the sum. When used on biological sequences, ANNs typically, just like PWMs, treat the sequence as a series of overlapping windows, calculating a score for each window from a number of position-specific weights, but unlike PWMs, the calculation of the score can be nonlinear, allowing correlations between positions to influence the prediction. An example of a correlation could be that a motif needs to contain a proline in one out of two positions, but not both. This is known as an XOR (exclusive or) situation, and if it is depicted in two dimensions (with the occurrence of proline at the two positions along the two axes), it is not possible to separate the positive examples from the negative by a straight line. Therefore, the problem is said not to be *linearly separable*, and it is not possible to solve it by a PWM or a profile HMM. However, an ANN with at least one hidden layer or a branched HMM could be able to do it.

Finally, SVMs are also frequently encountered in biological sequence analysis (Schölkopf and Smola 1998). Like ANNs, they treat each input pattern as a set of numbers which can be represented as a point in space, and an SVM model attempts to map the points so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. Often, the mapping procedure (implemented through the so-called kernel function) is nonlinear and adds many more dimensions than the input patterns originally had, thereby increasing the chance that the problem is linearly separable

[7] http://pfam.xfam.org/search.

[8] http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan/.

[9] http://www.ebi.ac.uk/interpro/search/sequence-search.

in the ensuing high-dimensional space. In this way, the SVM can also solve XOR-type problems. However, SVMs are traditionally rarely used with the window-based approach, but more often with global features.

4 Performance of Prediction Methods

After having trained a statistical or machine learning method, it is crucial to test its predictive performance on another dataset. This is a very important point: It is not enough that a trained method can reproduce its input examples exactly—in fact, it is not even interesting, since a database can do the same. What is interesting is whether a model can *generalize* from the examples in the training set and produce useful output for sequences it has not "seen" before.

There is often a certain degree of trade-off between training set and test set performance. If a model reproduces its training examples in too much detail, it uses its parameters to fit not only the common pattern in the data, but also the individual noise in each data point. When this happens, the performance on the test set goes down, and the model is said to be *overfitted*—colloquially speaking, it cannot see the forest for the trees.

Avoiding overfitting can be tricky; it may involve limiting the number of free parameters in the model, adding some regularizing terms to the parameters, or—especially in the case of ANNs—stopping the training early. In some cases, this is done using the test set performance as a criterion for choosing the optimal number of free parameters or the best point to stop the training, but in fact this is cheating, since the test set in such a procedure has been part of the training process. Instead, three datasets should be used: a training set, a validation set for optimizing the model architecture and training process, and a true test set (also known as evaluation set) for measuring the performance.

Instead of using a fixed part of the data as test set, performance evaluation is often done by *cross-validation*, where the dataset is divided into a number of folds, and each fold is in turn used as a test set, while the others are used as the training set. The final performance is then calculated as an average of the test set performances. The number of folds can vary; most often, five- or ten-fold cross-validation is used, but some authors prefer N-fold cross-validation, where N is the number of data points—in other words, just one example at a time is held out, while the training is performed on all other examples. This is also known as leave-one-out cross-validation or jackknife test.

The necessity for splitting the data into training and test is not special for bioinformatics; it applies to all prediction tasks. However, bioinformatics has an added complication: Sequences are related by descent. If there are sequences in the test set that are closely related to sequences in the training set, the measured performance is arguably not a true test performance. This can be taken into account

by reducing homology in the dataset before splitting it into folds (homology reduction) or by ensuring that no too closely related pair of sequences end up in different folds (homology partitioning). Two widely used algorithms for homology reduction were published early in the history of bioinformatics (Hobohm et al. 1992).

There are diverging views concerning exactly how closely related two sequences should be allowed to be in order to be separated into different folds. Some authors arbitrarily set a rather high cutoff, e.g., 80 or 90 % identity in a pairwise alignment (Reinhardt and Hubbard 1998; Höglund et al. 2006). One approach to a non-arbitrary definition is to identify a cutoff above which the problem could be better solved by alignment than by machine learning (Sander and Schneider 1991; Nielsen et al. 1996). Another approach is to use a cutoff in alignment score above which there is statistical significance of homology (Nielsen and Wernersson 2006). These approaches tend to result in much lower cutoff values, typically corresponding to ≈25 % in long alignments (Sander and Schneider 1991). When comparing reported performances of different methods, it is important to take into account which type of homology reduction or partitioning was used (if any).

However, it is debatable whether homology reduction or partitioning makes sense when constructing homology-based methods. The whole point of such methods is to use the annotations of homologues, the closer the better, and by reducing homology in the dataset, one would be reducing away the very data that the method needs. Still, it is sometimes done, e.g., in the PLoc/mPLoc/iLoc servers (see Sect. 9), where a cutoff of 25 % identity has been used.

When reporting performances of prediction methods, a variety of measures may be used, potentially confusing the untrained reader. The conceptually simplest performance measure, the fraction or percentage of correct answers (also known as *accuracy*), can be misleading if the classes are not the same size. As an example, consider a method for predicting cell wall-binding proteins and a dataset which has 99 negative examples (non-cell wall-binding proteins) for each positive example. If the method consistently answers "non-cell wall binding," it will be correct 99 % of the time, even though the "prediction" is completely non-informative. Instead, a number of alternative measures are often used. When discriminating between two classes, the most important performance measures can be defined in terms of the numbers of true positives (TP), true negatives (TN), false positives (type I errors or overpredictions, FP), and false negatives (type II errors or misses, FN):

- Sensitivity (also known as recall or true positive rate—how many of the positive examples are found?):
  - $\mathrm{Sn} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$

- Specificity (also known as true negative rate—how many of the negative examples are found?):
  - $\mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$
- Precision (also known as positive predictive value—how many of the positive predictions are true?):
  - $\mathrm{Pr} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$
- Matthews correlation coefficient—a measure which takes values between −1 and 1, where 1 is a perfect prediction, 0 is a random guess or non-informative prediction, and −1 is a prediction that is consistently wrong:
  - $\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$

It should be mentioned that the term "specificity" is not unequivocal; it has also been used to denote what is here referred to as precision (e.g., in Gardy et al. 2003).

Whenever a prediction method gives a quantitative output, there is a trade-off between sensitivity and specificity, controlled by the threshold (also known as cutoff) above which a prediction is considered positive. Lowering the threshold reduces the number of false negatives, thereby increasing the sensitivity, but it also increases the number of false positives, thereby reducing the specificity (and the precision). It is possible to plot the sensitivity as a function of the false positive rate (1 minus specificity) for varying threshold values—such a plot is known as a receiver operating characteristic (ROC) curve (see Fig. 3). The area under the ROC curve (usually referred to as AUC or AROC) can be used as a threshold-independent performance measure; it will be 1 for a perfect prediction, 0.5 for random guesses, and 0 for a consistently wrong prediction.

When predicting more than two classes—e.g., a number of SCLs—the maximal information about the prediction is provided by the so-called *confusion matrix*: A table showing, for each observed class, how many examples were predicted to belong to each class. This can be used to see not only how well each class was predicted, but also which classes were particularly difficult to distinguish. From the confusion matrix, sensitivity, specificity, precision, and MCC can be calculated for each class. There are also measures that summarize a whole confusion matrix in one number, such as the Gorodkin correlation coefficient, which is a generalization of the MCC to more than two classes, or the normalized mutual information coefficient (Baldi et al. 2000; Gorodkin 2004). In practice, these are rarely calculated, and the percentage of correct answers is often used instead, despite the shortcomings of this measure.

5 Recognition of Signal Peptides

The secretory SP is among the earliest prediction targets for bioinformatic algorithms. The oldest SP prediction methods used a simple PWM for the SP cleavage site, first with a reduced alphabet (von Heijne 1983) and later with weights for all amino acids (von Heijne 1986). Another very early SP prediction method used two simple sequence-derived features, peak hydrophobicity and length of the uncharged region, to discriminate SPs, but did not predict the cleavage site (McGeoch 1985).

SPs are present in all domains of life, but it was early discovered that there are differences between broadly defined systematic groups (von Heijne and Abrahmsén 1989). SPs of Gram-positive bacteria are longer than those of Gram-negative bacteria, which in turn are longer than those of eukaryotes. A sequence logo of SPs from Gram-positive bacteria is shown in Fig. 2.

In 1997, the SP predictor SignalP[10] was among the first to use ANNs for sorting signal prediction (Nielsen et al. 1997). Later, in versions 2 and 3, an HMM was added to the method (Nielsen and Krogh 1998; Bendtsen et al. 2004), while version 4 is again purely ANN-based (Petersen et al. 2011). SignalP is among the most cited prediction servers in bioinformatics, and it has performed favorably in comparative studies (Menne et al. 2000; Klee and Ellis 2005; Choo et al. 2009), including one specific to Gram-positive bacteria (Zhang et al. 2009). The latter study, however, concluded that a consensus of five methods performed better than any of the individual methods. The included methods besides SignalP were the PWM-based PrediSi[11] (Hiller et al. 2004), the HMM-based Phobius[12] (Käll et al. 2004), SOSUIsignal[13] which is based on amino acid propensities in regions (Gomi et al. 2004), and the unpublished SIG-Pred.

Another SP prediction method worth mentioning is Signal-BLAST[14] (Frank and Sippl 2008) which, quite unusually for a sorting signal prediction method, is homology-based. It uses BLAST (Altschul et al. 1997) with some customized settings to search a reference set of SPs and non-SPs and returns the class of the best hit as its prediction.

The performance of SP prediction in Gram-positive bacteria is fairly high, with SignalP 4.0 reporting an MCC of 0.85 in distinguishing between SPs and non-SPs and a cleavage site precision of 83 %. Interestingly, the cleavage site precision is higher for Gram positives than for the other two organism groups (Gram negatives and eukaryotes). Note that these performances are cross-validation performances on a strictly homology-reduced dataset, so they reflect the performance you would expect if you submitted sequences that were completely unrelated to any in the SignalP 4.0 dataset. The performance measured by applying the finished method

[10] http://www.cbs.dtu.dk/services/SignalP/.

[11] http://www.predisi.de/.

[12] http://phobius.sbc.su.se/.

[13] http://harrier.nagahama-i-bio.ac.jp/sosui/sosuisignal/sosuisignal_submit.html.

[14] http://sigpep.services.came.sbg.ac.at/signalblast.html.

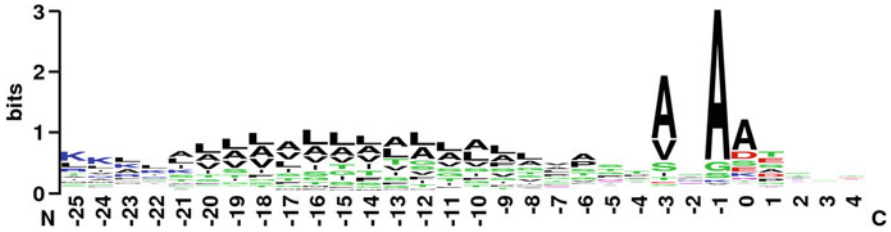


Fig. 2 Sequence logo of signal peptides from Gram-positive bacteria, aligned after their cleavage site (between positions −1 and 0). The height of each stack of letters corresponds to the information (conservation) at that position, while the height of each individual letter is proportional to the fraction of that amino acid at that position. The visible features are the cleavage site specifying residues in −3 and −1 (strong preference for alanine), the hydrophobic region that approximately stretches from −21 to −8, and a preference for the positively charged lysine in the N-terminal region. Picture made with WebLogo (Crooks et al. 2004)

(where the outputs of the different dataset partitions are averaged) to the whole dataset is considerably higher, with an MCC of 0.96. Note that SignalP 4.0 in spite of a higher MCC has a lower sensitivity than SignalP 3.0; the cutoff had simply been placed at a higher value in order to maximize the MCC. In the slightly modified SignalP 4.1, there is an option to select a cutoff that reproduces the sensitivity of SignalP 3.0. Of course, this comes at a price of a higher false positive rate, but it is still lower than that of SignalP 3.0, as can be seen from the ROC curves in Fig. 3.

It should be stressed that the presence of an SP does not necessarily mean that the protein is extracellular or cell wall associated; there may be downstream TMHs keeping the protein integrated in the membrane. It has been reported that cleavable SPs are rarely found in bacterial cytoplasmic membrane proteins (Broome-Smith et al. 1994), but a quick search in UniProt (The UniProt Consortium 2015) reveals that they are not that rare after all, so a prediction of SPs should always be combined with a search for TMHs (see Sect. 7) before drawing conclusions about the SCL.

SignalP and the other SP predictors mentioned so far only predict classical SPs, translocated by the Sec system and cleaved by type I signal peptidases. For lipoproteins cleaved by lipoprotein signal peptidase, there are other prediction methods. LipoP[15] (Juncker et al. 2003) is an HMM-based method (although an ANN was also trained during the development of the method). Even though LipoP has been trained on sequences from Gram-negative bacteria only, both the original paper and a later study (Rahman et al. 2008) report that it has a good performance on sequences from Gram-positive bacteria also. Other methods include the ANN-based SPEPlip[16] (Fariselli et al. 2003), which has separate options for

[15]http://www.cbs.dtu.dk/services/LipoP/.

[16]http://gpcr.biocomp.unibo.it/cgi/predictors/spep/pred_spepcgi.cgi.

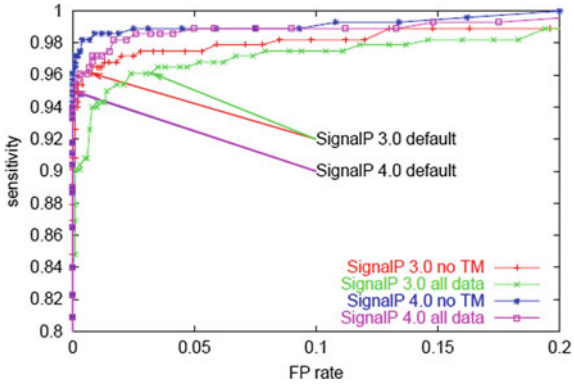


Fig. 3 ROC curve showing performance of SignalP versions 3 and 4 as sensitivity versus false positive rate. "No TM" means performance when the negative set did not contain transmembrane segments; "all data" means that sequences with transmembrane segments were included in the negative data. Observe that although SignalP 4 by default has a lower sensitivity than SignalP 3, this is only a question of cutoff; the curves for SignalP 4 are consistently closer to the *upper left* corner, showing that SignalP 4 is a better method. Note that the sensitivity and false positive rate values depicted here are not cross-validation performances, but measured by applying the finished method to the whole dataset

Gram-negative and Gram-positive bacteria, and the HMM-based PRED-LIPO[17] (Bagos et al. 2008), which is specific to Gram-positive bacteria. In addition to these methods, there is also a profile in PROSITE (Sigrist et al. 2013) dedicated to lipoproteins from both Gram-negative and Gram-positive bacteria, named PROKAR_LIPOPROTEIN.[18] A sequence logo of lipoprotein SPs aligned to this model is shown in Fig. 4.

Rahman et al. (2008) benchmarked seven methods for prediction of lipoprotein SPs in Gram-positive bacteria using a specially selected negative set, where every sequence contained a cysteine that could "fool" the predictors. They found LipoP to be the best, but recommended a consensus approach including the PROSITE profile and a regular expression especially tailored for Gram-positive lipoproteins.

It should be noted that LipoP, SPEPlip, and PRED-LIPO are all able to predict classical SPs as well, differentiating between the two types of SP. Additionally, LipoP differentiates between SPs and N-terminal transmembrane helices (TMHs).

For SPs translocated by the twin-arginine protein translocation (Tat) pathway, there are also a few dedicated prediction methods available. In addition to the twin-arginine motif in the N-terminal region that gave them their name, they also differ from Sec SPs by being on average longer and less hydrophobic (Cristóbal et al. 1999). However, the difference in hydrophobicity may not be significant in

[17] http://bioinformatics.biol.uoa.gr/PRED-LIPO/.

[18] http://prosite.expasy.org/PS51257 and http://prosite.expasy.org/PDOC00013.

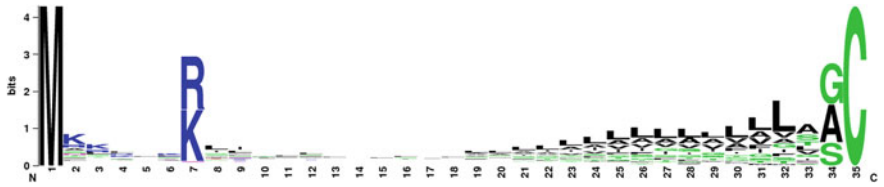


Fig. 4 Sequence logo of lipoprotein signal peptides from both Gram-positive and Gram-negative bacteria, aligned to the PROSITE profile PS51257/PROKAR_LIPOPROTEIN. Lipid attachment occurs at the completely conserved cysteine in position 35. Note that individual sequences may be shorter or longer than 35 amino acids; in the logo, they have been stretched or shortened to fit the model. Picture from PROSITE (Sigrist et al. 2013) made with WebLogo (Crooks et al. 2004)

Gram-positive bacteria (Tjalsma et al. 2000). The available servers are TatFind[19] (Rose et al. 2002), which is based on a regular expression combined with a set of simple rules concerning hydrophobicity and charge, TatP[20] (Bendtsen et al. 2005b), which is based on a regular expression combined with two ANNs, and the newer HMM-based PRED-TAT[21] (Bagos et al. 2010). In addition, there are three motifs available in the family and domain databases: the PROSITE profile TAT,[22] the Pfam profile TAT_signal,[23] and the TIGRFAMs profile TAT_signal_seq.[24] A logo of sequences aligned to the PROSITE profile is shown in Fig. 5. Note that all these methods make no distinction between Gram-positive and Gram-negative bacteria—if the Tat SPs indeed differ between the two bacterial groups, there should be room for improvement of the prediction.

Yet another type of SP is the pseudopilin SP, which directs proteins of pilin-like structures to be secreted via the Com or FPE pathway (Tjalsma et al. 2004; Desvaux et al. 2009). Like the classical SPs, it has a hydrophobic region, but cleavage takes place N-terminally to this region, at a site with the consensus sequence "KGF" (with cleavage between G and F). I am not aware of any predictor that covers this motif.

6 Prediction of Non-classical Secretion

Non-classical secretion has been defined as secretion without an N-terminal, cleaved SP (Bendtsen et al. 2005a). In Gram-negative bacteria, this happens to proteins belonging to secretion systems of type I, III, IV, and VI (Binnewies et al.

[19] http://signalfind.org/tatfind.html.

[20] http://www.cbs.dtu.dk/services/TatP/.

[21] http://www.compgen.org/tools/PRED-TAT/.

[22] http://prosite.expasy.org/PS51318 and http://prosite.expasy.org/PDOC51318.

[23] http://pfam.xfam.org/family/PF10518.

[24] http://www.jcvi.org/cgi-bin/tigrfams/HmmReportPage.cgi?acc=TIGR01409.

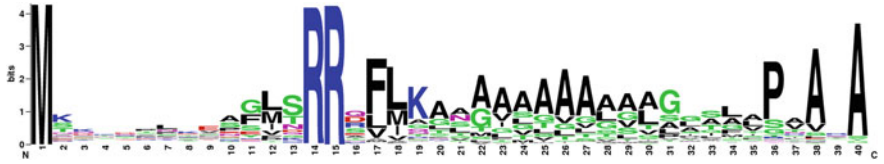


Fig. 5 Sequence logo of Tat signal peptides from both Gram-positive and Gram-negative bacteria, aligned to the PROSITE profile PS51318/TAT. Picture from PROSITE (Sigrist et al. 2013) made with WebLogo (Crooks et al. 2004)

2005; Desvaux et al. 2009). For Gram-positive bacteria, the phenomenon appears to be less important, but there are some examples of proteins exported via, e.g., the Wss, holin, and SecA2 pathways (Bendtsen et al. 2005a; Desvaux et al. 2009).

In addition, some pheromones and bacteriocins (antimicrobial peptides) are exported via an ABC transporter (Tjalsma et al. 2004). This phenomenon does not match the above definition of non-classical secretion, since it does depend on a cleaved N-terminal signal, but since this signal has no hydrophobic region and thus bears no resemblance to the SPs described in the previous section, it cannot be predicted by any of the SP prediction methods; so from a prediction viewpoint, it falls into the category of non-classical secretion.

Since non-classical secretion occurs by a variety of mechanisms, and very few examples are known for each mechanism, there are no signal-based predictors available. It is also difficult to train a global property-based method because of the scarcity of experimentally known data. The prediction method SecretomeP (Bendtsen et al. 2005a) took a different approach, based on the idea that secreted proteins must be expected to share certain features independent of the pathway used to secrete them: The positive training dataset simply consisted of classically secreted proteins with the SP removed. A large number of structural and functional features calculated from the amino acid sequence were then tested for their predictive power for SCL, and the most promising features were used to train an ANN. For Gram-positive bacteria, three simple features—amino acid composition, threonine content, and overall hydrophobicity—and three predicted features—transmembrane helices, secondary structure, and disorder—were selected. These features led to a sensitivity of 89 % (on the truncated examples) and a specificity of 95 %. Among a hand-curated dataset of 14 SP-less Gram-positive proteins with a known extracellular function, 10 were predicted to be non-classically secreted by SecretomeP.

The competing method SecretP (Yu et al. 2010a), which is also based on feature selection, just using SVMs instead of ANNs, used another, more problematic approach to the problem of dataset generation: The positive training set consisted of proteins that were annotated to be secreted but had no annotated SP in UniProt (The UniProt Consortium 2015). The problematic aspect of this is that a lack of annotated SP may simply reflect an incomplete annotation rather than a real absence of

SP. This suspicion is confirmed by the fact that SignalP was found to predict 11 out of 13 supposedly non-classically secreted proteins.

7 Prediction of Transmembrane Topology

Like SPs, prediction of transmembrane α-helices (TMHs) has a long history in bioinformatics. Initially, the basis for the prediction was simply a plot of the hydrophobicity, averaged in a sliding window over the sequence (Kyte and Doolittle 1982; Klein et al. 1985). A slightly more advanced approach was represented by TOP-PRED (von Heijne 1992), which combined hydrophobicity analysis with counting the number of positively charged residues in each loop in order to choose the topological model which best conformed to the "positive-inside rule" (von Heijne and Gavel 1988).

Later, machine learning methods have been used to predict membrane protein topology, i.e., which parts of the sequence are inside, transmembrane, and outside. In particular, the HMM technology has been popular in this area, because it provides the ability to model the "grammar" of the problem: If a TMH follows an inside loop, it must be followed by an outside loop, and vice versa. This is typically modeled by a cyclic HMM, having submodels for helices, inside loops, and outside loops. The best known HMM for TMH prediction is TMHMM[25] (Krogh et al. 2001), but also HMMTOP[26] (Tusnády and Simon 2001) has found a wide usage. A comparative analysis in 2001 found TMHMM to be the best performing TMH predictor (Möller et al. 2001). Newer surveys covering more recently published predictors unfortunately do not provide quantitative performance comparisons (Elofsson and von Heijne 2007; Punta et al. 2007; Tusnády and Simon 2010).

Since hydrophobicity is a feature of both SPs and TMHs, these two are easily confused by prediction methods. TMHMM often falsely predicts an SP as a TMH, and versions 1–3 of SignalP would often predict a TMH close to the N-terminus as an SP. Newer topology prediction methods such as the HMM-based Phobius[27] (Käll et al. 2004), Philius[28] which is based on dynamic Bayesian networks (Reynolds et al. 2008), the ANN-based MEMSAT3 (Jones 2007), the SVM-based MEMSAT-SVM[29] (Nugent and Jones 2009), and the ANN + HMM-based SPOCTOPUS[30] (Viklund et al. 2008) deal with this problem by modeling both

[25] http://www.cbs.dtu.dk/services/TMHMM/.

[26] http://www.enzim.hu/hmmtop/.

[27] http://phobius.sbc.su.se/.

[28] http://www.yeastrc.org/philius/.

[29] Available through the PSIPRED Protein Sequence Analysis Workbench, http://bioinf.cs.ucl.ac.uk/psipred/.

[30] http://octopus.cbr.su.se/.

these signals. However, the paper about version 4 of SignalP (Petersen et al. 2011) reports a better discrimination between SPs and TMHs than all these methods, and it is worth noting that the performance difference is larger for bacterial sequences than for eukaryotic sequences. This probably reflects the fact that the SP models in Phobius, Philius, MEMSAT, and SPOCTOPUS are not divided into organism types, which causes the results to be biased toward the organism group with most data (eukaryotes).

Another confounding factor is the fact that multi-spanning membrane proteins sometimes have so-called reentrant loops—segments of the sequence that dip into the membrane but do not span it, leaving the membrane on the same side from which they entered. Reentrant loops are not very frequent; only five examples from Gram-positive bacteria are currently reported in UniProt. OCTOPUS (Viklund and Elofsson 2008) and SPOCTOPUS make an attempt at predicting reentrant loops.

The use of profiles of homologous sequences generated by BLAST or PSI-BLAST (see Sects. 2 and 3) in the training and prediction of TMH recognition methods has been shown to enhance predictive performance by approximately 10 % units (Viklund and Elofsson 2004). Methods that use profiles include PRODIV-TMHMM (Viklund and Elofsson 2004), PolyPhobius (Käll et al. 2005), MEMSAT3, MEMSAT-SVM, OCTOPUS, and SPOCTOPUS.

An interesting alternative method is SCAMPI[31] (Bernsel et al. 2008) which does not use machine learning nor statistics on a training set to calculate its parameters, instead they are based on a series of experiments where all 20 possible amino acids have been inserted at various positions into a model TMH (Hessa et al. 2007). These experiments have been used to calculate an apparent free energy contribution, ΔG_{app}, which is used as an analogue to a hydrophobicity scale. The overall ΔG_{app} for each sequence window is calculated and used as input to an HMM-like model with only two free parameters to be estimated from the training data. SCAMPI reported a performance comparable to the best machine learning methods.

As is the case for SP prediction, consensus methods for TMH prediction have been shown to perform better than any of the constituent methods. The server TOPCONS[32] (Bernsel et al. 2009; Tsirigos et al. 2015) offers a consensus prediction of both TMHs and SPs based on OCTOPUS, SPOCTOPUS, PolyPhobius, Philius, and SCAMPI. TOPCONS reports 83 % correctly predicted topologies on a benchmark set. The downside of TOPCONS is the running time, increased by the fact that four of the five predictors are based on profiles which first have to be constructed from a database search. An alternative consensus server, only based on methods that do not require profiles, is TOPCONS-single[33] (Hennerdal and Elofsson 2011), which does approximately six percentage units worse than TOPCONS, but 70 times faster.

[31] http://scampi.cbr.su.se/.

[32] http://topcons.cbr.su.se/ or http://topcons.net/.

[33] http://single.topcons.net/.

8 Prediction of Cell Wall-Binding Motifs

There are two ways proteins can be associated with the Gram-positive cell wall: by covalent attachment and non-covalent binding. The first group is characterized by a special sortase-cleaved motif with the consensus sequence "LPXTG" followed by what looks like a reversed SP: a stretch of hydrophobic amino acids followed by a region with positively charged residues. There are two available methods for prediction of this signal: the HMM-based CW-PRED server[34] (Litou et al. 2008; Fimereli et al. 2012) and the PROSITE profile GRAM_POS_ANCHORING.[35] A sequence logo of LPXTG-signals aligned to the PROSITE profile is shown in Fig. 6. CW-PRED seems to be less restrictive than the PROSITE profile; while the latter has an absolute requirement for the proline in the second position, CW-PRED can also detect the variant "LAXTG."

One type of non-covalent binding is described by the approximately 20 aa long PROSITE profile CW[36] and the corresponding Pfam profile CW_binding_1.[37] It occurs mainly in two bacterial Gram-positive protein families: choline-binding proteins and glucosyltransferases (Janeček et al. 2000; López and García 2004). The motif occurs as a repeat, typically many times per protein. The PROSITE profile appears to be more sensitive, with typically more hits per sequence reported in UniProt, than the Pfam profile. In the glucosyltransferases, most of the CW motifs occur in glucan-binding domains (Shah et al. 2004), and TIGRFAMs has an approximately 60 aa long profile associated with glucan binding, glucan_65_rpt,[38] which roughly corresponds to three tandem copies of the CW motif. The TIGRFAMs page states that the 30 aa motif reported by Shah et al. (2004) corresponds to half of glucan_65_rpt or one and a half copies of CW_binding_1. The fact that a cell wall-binding motif forms part of a glucan-binding motif diminishes its value for predicting cell wall localization somewhat, since glucan is not part of the peptidoglycan cell wall, and glucan-binding proteins therefore should be classified as secreted.

Enzymes involved in bacterial cell wall degradation often have a peptidoglycan binding domain of approximately 60 aa (Krogh et al. 1998), which is totally unrelated to the repeats mentioned above. A Pfam profile, PG_binding_1,[39] is available. Another example with a Pfam profile is the WxL domain[40] (Brinster et al. 2007).

[34] http://bioinformatics.biol.uoa.gr/CW-PRED/.

[35] http://prosite.expasy.org/PS50847 and http://prosite.expasy.org/PDOC00373.

[36] http://prosite.expasy.org/PS51170 and http://prosite.expasy.org/PDOC51170.

[37] http://pfam.xfam.org/family/PF01473.

[38] http://www.jcvi.org/cgi-bin/tigrfams/HmmReportPage.cgi?acc=TIGR04035.

[39] http://pfam.xfam.org/family/PF01471.

[40] http://pfam.xfam.org/family/PF13731.

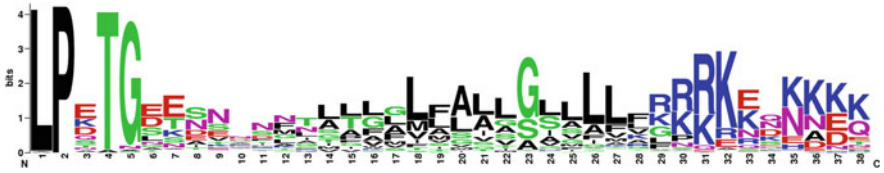


Fig. 6 Sequence logo of cell wall-attached proteins from Gram-positive bacteria, aligned to the PROSITE profile PS50847/GRAM_POS_ANCHORING. Cleavage occurs between positions 4 and 5. Picture from PROSITE (Sigrist et al. 2013) made with WebLogo (Crooks et al. 2004)

9 Multi-category Predictors

The first software to attempt a classification of proteins into multiple SCLs was PSORT (Nakai and Kanehisa 1991). It was basically a signal-based method, incorporating the previously mentioned early methods for prediction of SPs (McGeoch 1985; von Heijne 1986) and TMHs (Klein et al. 1985), but it also used amino acid composition, especially for discriminating outer membrane proteins in Gram-negative bacteria.

For bacteria, PSORT I has been superseded by PSORTb[41] (Gardy et al. 2003, 2005; Yu et al. 2010b), which is now in version 3. Version 1 was for Gram-negative bacteria only, but in version 2, Gram-positive bacteria were included. Version 3 additionally offers predictions for Archaea and the "problematic" bacteria, which either stain Gram positive although they have an outer membrane (such as genus *Deinococcus*) or stain Gram negative although they have no outer membrane (phylum Tenericutes).

PSORTb is a hybrid method, incorporating both signal-based, global property-based, and homology-based predictions. The signal-based component comprises recognition of SPs and TMHs and a database of motifs (regular expressions) derived from PROSITE, which are found to be exclusive to specific SCLs. The global properties component is SVM-based; in version 1, its input consisted of amino acid composition only, but in versions 2 and 3, a collection of overrepresented subsequences is used. The homology-based component is a simple BLAST with direct annotation transfer. Finally, a Bayesian network is used to integrate the outputs from the components and arrive at a final prediction.

The final prediction, however, may be "unknown." PSORTb values precision over recall, so it prefers to deliver no prediction rather than a prediction with weak evidence. It may also arrive at two SCLs, signifying that the protein is predicted to function in both compartments, or belong to the interface between the compartments (e.g., cell membrane/cell wall).

The SCLs predicted by PSORTb 3 extend beyond the standard four categories for Gram-positive bacteria; there are new subcategory SCLs such as "Fimbrial,"

[41] http://www.psort.org/psortb/.

"Flagellar," and "Spore." The reported precision of PSORTb 3 on Gram-positive bacteria (the main categories) is 98 %, with a recall of 93 %. This is tested by five-fold cross-validation with a dataset that was homology reduced, but only down to 80 % identity.

Another predictor that owes its high performance to homology is the SCL predictor built into the prediction workbench Proteome Analyst[42] (Lu et al. 2004; Szafron et al. 2004). It uses a combination of direct and indirect annotation transfer by retrieving up to three hits from the Swiss-Prot part of UniProt by BLAST and then parsing the "subcellular location" field, the keywords, and the cross-referenced InterPro entries. The retrieved words are then processed by a Naïve Bayes classifier. Other machine learning methods (ANN and SVM) were also tried, and although they could enhance the performance by a few percent, the authors decided to stick with Naïve Bayes in order to be able to provide explanations for the individual predictions. The PSORTb 3 paper (Yu et al. 2010b) reports that PSORTb 3.0 and Proteome Analyst 3.0 have comparable precisions, but make complementary predictions, so that a combined analysis with both methods has the highest coverage overall.

Kuo-Chen Chou's group has published a long series of predictors for protein SCL (see, e.g., Chou and Shen 2010). They prefer to publish one Web site per organism group instead of providing one Web site with an option for selecting organism group, and furthermore, they tend to change the name for each new version instead of adding a version number. For Gram-positive bacteria, the relevant predictors are named Gpos-PLoc[43] (Shen and Chou 2007), Gpos-mPLoc[44] (Shen and Chou 2009), and iLoc-Gpos[45] (Wu et al. 2012). The PLoc/mPLoc/iLoc servers are hybrid methods, mostly relying on indirect homology annotation through the Gene Ontology (GO) terms of database hits. GO (Ashburner et al. 2000) is an ordered system (a directed acyclic graph) of controlled terms, which describe the biological process, molecular function, and cellular component of proteins. GO terms are extracted from all database hits with a pairwise identity above a certain cutoff, and then, a *k*-nearest neighbor classifier is applied to the high-dimensional vectors of occurrences of GO terms. If no hits are found, or if the found hits have no GO annotation, a profile-based global property approach is used. The nature of this approach varies between PLoc, mPLoc, and iLoc. However, the corresponding papers contain no information about how often the global property approach is needed, and the performance of this approach has never been reported separately. The overall accuracy, measured by jackknife (leave-one-out cross-validation), is reported to be 83 % for Gpos-PLoc, 82 % for Gpos-mPLoc, and 93 % for iLoc-Gpos, respectively.

[42] http://pa.wishartlab.com/pa/pa/ Note: the website requires login, but registration is free.

[43] http://www.csbio.sjtu.edu.cn/bioinf/Gpos/.

[44] http://www.csbio.sjtu.edu.cn/bioinf/Gpos-multi/.

[45] http://www.jci-bioinfo.cn/iLoc-Gpos.

The new feature of the mPLoc and iLoc servers relative to the PLoc servers is the ability to predict multiple SCLs, i.e., predict whether a protein can exist in more than one cellular compartment. This can be of importance in eukaryotes, where many proteins, for example, may shuttle between the cytoplasm and the nucleus, but it is clearly of minor importance in Gram-positive bacteria. In the dataset for Gpos-mPLoc and iLoc-Gpos, there were only 4 such examples out of 519 proteins.

Another hybrid approach is LocTree3[46] (Goldberg et al. 2014), which has two components: a PSI-BLAST (Altschul et al. 1997) homology search with direct transfer of SCL annotation and a global property approach corresponding to LocTree2 (Goldberg et al. 2012). If no homology hits are found with an E value better than a specified cutoff, the LocTree2 method is applied. It consists of a decision tree of SVMs trained with a "profile kernel," basically using the occurrence of short substrings in profiles made by PSI-BLAST searches as input. When delivering a prediction, LocTree3 reports whether the evidence is based on homology or on LocTree2. For bacteria, LocTree2 has a reported performance (overall accuracy) of 86 %, and LocTree3 has 90 %. Interestingly, in the LocTree3 paper, the measured accuracy for PSORTb 3.0 is only 57 %. This huge difference to PSORTb's own reported performance may reflect different views on the exactly correct way to parse UniProt's SCL annotations.

LocTree2/3 claims to be able to predict SCLs for all domains of life, but it seems less well suited for Gram-positive bacteria, since it offers no opportunity to choose between Gram positives and Gram negatives. Thus, it may predict categories such as periplasm and outer membrane for Gram-positive bacteria, while it totally fails to predict cell wall.

The methods described so far in this section have all been wholly or partly homology-based. However, there is also the global property-based CELLO[47] (Yu et al. 2006), an SVM-based predictor for both eukaryotes, Gram-negative bacteria, and Gram-positive bacteria. The SVMs are organized in a two-level system, where the first level contains a number of SVMs trained on various sequence encodings, and the second layer is a "jury SVM," which decides on the prediction based on the outputs of the first-layer SVMs. The sequences are encoded by total amino acid composition, dipeptide composition, and amino acid composition (in some cases with a reduced alphabet) in a number of partitions of each sequence. The performance for Gram-positive bacteria is unfortunately not reported in the paper, but for Gram-negative bacteria, it is 95 % overall accuracy without homology reduction and 83 % with homology reduction (30 % identity).

Another predictor with special interest in relation to Gram-positive bacteria is TBpred,[48] which is specific to the genus *Mycobacterium* (Rashid et al. 2007). TBpred is a hybrid between a global property-based approach using SVMs and a

[46] https://rostlab.org/services/loctree3/.

[47] http://cello.life.nctu.edu.tw/.

[48] http://www.imtech.res.in/raghava/tbpred/.

signal-based approach using PWMs. The authors state that an organism-specific method performs better than general methods for that organism, but the references they cite for this claim concern human proteins, and unfortunately, they do not show that this is indeed the case by training on a broader class of Gram-positive bacteria for comparison.

Finally, the pipeline LocateP (Zhou et al. 2008) should be mentioned as an example of a signal-based multi-category predictor, even though it does not allow submission of new sequences (instead, a database of LocateP predictions on 427 known genomes of Gram-positive bacteria is available online[49]). LocateP is based on SignalP, PrediSi, TMHMM, Phobius, and TatFind combined with a set of custom-made profile HMMs in a decision tree. LocateP uses an extended ontology of SCLs, counting N-terminally anchored, C-terminally anchored, lipid-anchored, and multi-spanning membrane proteins as separate classes. However, it does not predict cell wall attachment. The authors report a slightly lower accuracy than CELLO when testing on Swiss-Prot data, but a higher accuracy than both CELLO and PSORTb 2 when testing on data from a series of proteomic studies.

10 Discussion

As is apparent from this chapter, the many possible ways of approaching the SCL prediction problem have resulted in a large number of available prediction servers. Comparing their performances can be complicated, and all their authors tend to claim superior performance for their particular method. Add to this that the usability is sometimes limited (some Web servers allow only one or a few sequences in each submission), that response times vary a lot, and that there are almost as many different output formats as there are servers, and you get a rather frustrating situation. Even the definitions of SCLs may vary from server to server—as an example, a peripheral membrane protein may be defined as belonging to the membrane, or to the compartment it protrudes into (inside/outside).

This situation is clearly not ideal for the user, who might prefer a "one-stop shop" to go to for all sequence-based prediction needs, an equivalent of UniProt or InterPro. But this kind of confusion is probably inevitable in a field that is evolving so fast. Scientific competition is basically beneficial, and competing groups should certainly not be discouraged from publishing their predictors independently. That being said, prediction servers ought to follow certain standards concerning usability, definitions, and formats.

Personally, I must admit to having added to the complexity through my involvement in the servers SignalP, LipoP, and TatP (see Sect. 5). In hindsight, we should not have published LipoP and TatP as separate servers, but as functionalities

[49] http://www.cmbi.ru.nl/locatep-db/.

within the SignalP server. Hopefully, the next version of SignalP will be able to predict all these types of SPs in one user interface.

The multi-category prediction methods report quite impressive performances, and as described in Sect. 1, the error rates for prediction may now be lower than the error rates for high-throughput experiments. However, it is important to keep in mind that these performances are achieved by analyzing the annotations of homologues found by sequence similarity searches. I see three problems with this. First, predictions for novel organisms and metagenomics samples with few known homologues will necessarily be harder than for the organisms the training and test sets were built from, so coverage and precision for such organisms will be considerably lower than the reported performances. Second, the annotations used for prediction are themselves error-prone and not necessarily derived from experiments. In particular, when relying on keywords and GO terms, there is a real danger of circular reasoning, where annotations based on predictions are used as a basis for new predictions, which then may enter the databases as annotations. Third, homology-based predictions do not reflect a real biological knowledge about the protein sorting process in the way a successful signal-based predictor does.

But there are quite successful signal-based predictors available for SPs, TMHs, and some of the cell wall-binding motifs. A combination of such predictors into a homology-independent multi-category prediction method is an important task waiting to be done. LocateP (see previous section) is an attempt in this direction, but is not implemented as a prediction server.

As machine learning algorithms continue to evolve, new classes of algorithms should also be expected to be applied to prediction of SCL. In fields such as image processing and speech recognition, novel types of ANNs—deep and recurrent neural networks—have been extensively used in recent years (Krizhevsky et al. 2012; Dahl et al. 2012), and they are beginning to be employed also in bioinformatics, e.g., for predicting protein secondary structure (Magnan and Baldi 2014) or alternative splicing (Xiong et al. 2015). The advantage of recurrent neural networks is that they are naturally designed to handle sequential data, so the sequence is not chopped up into apparently unrelated windows, and they can potentially learn long-range correlations. A first attempt at using a recurrent neural network in SCL prediction has been published recently (Sønderby et al. 2015), so far, only for eukaryotic data, but the results seem promising. Coupled with a so-called convolutional layer—basically a series of PWMs of varying width though which the input sequences were presented—the network was able to learn from the data where in each sequence to focus its attention. Performance was much better than other methods working only on sequence and on the same level as advanced homology-based methods. This technology represents a new kind of compromise between signal-based and global property-based methods, since it is apparently able to find sorting signals in sequences even though it has only been given the sequences and their SCL categories during training. It will be very interesting to see where this and other novel technologies will take SCL prediction in the coming years.

References

Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389. doi:10.1093/nar/25.17.3389

Andrade MA, O'Donoghue SI, Rost B (1998) Adaptation of protein surfaces to subcellular location. J Mol Biol 276:517–525. doi:10.1006/jmbi.1997.1498

Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. Nat Genet 25:25–29. doi:10.1038/75556

Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD (2010) Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 26:2811–2817. doi:10.1093/bioinformatics/btq530

Bagos PG, Tsirigos KD, Liakopoulos TD, Hamodrakas SJ (2008) Prediction of lipoprotein signal peptides in Gram-positive bacteria with a hidden Markov model. J Proteome Res 7:5082–5093. doi:10.1021/pr800162c

Baldi P, Brunak S, Chauvin Y et al (2000) Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16:412–424. doi:10.1093/bioinformatics/16.5.412

Bendtsen JD, Kiemer L, Fausbøll A, Brunak S (2005a) Non-classical protein secretion in bacteria. BMC Microbiol 5:58. doi:10.1186/1471-2180-5-58

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795. doi:10.1016/j.jmb.2004.05.028

Bendtsen JD, Nielsen H, Widdick D et al (2005b) Prediction of twin-arginine signal peptides. BMC Bioinform 6:167. doi:10.1186/1471-2105-6-167

Bernsel A, Viklund H, Falk J et al (2008) Prediction of membrane-protein topology from first principles. Proc Natl Acad Sci 105:7177–7181. doi:10.1073/pnas.0711151105

Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res 37:W465–W468. doi:10.1093/nar/gkp363

Binnewies TT, Bendtsen JD, Hallin PF et al (2005) Genome update: protein secretion systems in 225 bacterial genomes. Microbiology 151:1013–1016. doi:10.1099/mic.0.27966-0

Briesemeister S, Blum T, Brady S et al (2009) SherLoc2: a high-accuracy hybrid method for predicting subcellular localization of proteins. J Proteome Res 8:5363–5366. doi:10.1021/pr900665y

Briesemeister S, Rahnenführer J, Kohlbacher O (2010) Going from where to why—interpretable prediction of protein subcellular localization. Bioinformatics 26:1232–1238. doi:10.1093/bioinformatics/btq115

Brinster S, Furlan S, Serror P (2007) C-Terminal WxL domain mediates cell wall binding in *Enterococcus faecalis* and other Gram-positive bacteria. J Bacteriol 189:1244–1253. doi:10.1128/JB.00773-06

Broome-Smith JK, Gnaneshan S, Hunt LA et al (1994) Cleavable signal peptides are rarely found in bacterial cytoplasmic membrane proteins. Mol Membr Biol 11:3–8. doi:10.3109/09687689409161023

Chang KY, Lin T, Shih L-Y, Wang C-K (2015) Analysis and prediction of the critical regions of antimicrobial peptides based on conditional random fields. PLoS ONE 10:e0119490. doi:10.1371/journal.pone.0119490

Choo K, Tan T, Ranganathan S (2009) A comprehensive assessment of N-terminal signal peptides prediction methods. BMC Bioinform 10:S2. doi:10.1186/1471-2105-10-S15-S2

Chou K-C, Shen H-B (2010) Cell-PLoc 2.0: an improved package of web-servers for predicting subcellular localization of proteins in various organisms. Nat Sci 02:1090–1103. doi:10.4236/ns.2010.210136

Cristóbal S, de Gier J-W, Nielsen H, von Heijne G (1999) Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*. EMBO J 18:2982–2990. doi:10.1093/emboj/18.11.2982

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190. doi:10.1101/gr.849004

Dahl GE, Yu D, Deng L, Acero A (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process 20:30–42. doi:10.1109/TASL.2011.2134090

de Castro E, Sigrist CJA, Gattiker A et al (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res 34:W362–W365. doi:10.1093/nar/gkl124

Desvaux M, Hébraud M, Talon R, Henderson IR (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends Microbiol 17:139–145. doi:10.1016/j.tim.2009.01.004

Elofsson A, von Heijne G (2007) Membrane protein structure: prediction versus reality. Annu Rev Biochem 76:125–140. doi:10.1146/annurev.biochem.76.052705.163539

Fariselli P, Finocchiaro G, Casadio R (2003) SPEPlip: the detection of signal peptide and lipoprotein cleavage sites. Bioinformatics 19:2498–2499. doi:10.1093/bioinformatics/btg360

Fimereli DK, Tsirigos KD, Litou ZI et al (2012) CW-PRED: a HMM-based method for the classification of cell wall-anchored proteins of Gram-positive bacteria. In: Maglogiannis I, Plagianakos V, Vlahavas I (eds) Artificial intelligence: theories and applications. Springer, Berlin, pp 285–290. doi:10.1007/978-3-642-30448-4_36

Finn RD, Bateman A, Clements J et al (2014) Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. doi:10.1093/nar/gkt1223

Finn RD, Clements J, Arndt W et al (2015) HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. doi:10.1093/nar/gkv397

Frank K, Sippl MJ (2008) High-performance signal peptide prediction based on sequence alignment techniques. Bioinformatics 24:2172–2176. doi:10.1093/bioinformatics/btn422

Gardy JL, Laird MR, Chen F et al (2005) PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21:617–623. doi:10.1093/bioinformatics/bti057

Gardy JL, Spencer C, Wang K et al (2003) PSORT-B: improving protein subcellular localization prediction for Gram-negative bacteria. Nucleic Acids Res 31:3613–3617. doi:10.1093/nar/gkg602

Goldberg T, Hamp T, Rost B (2012) LocTree2 predicts localization for all domains of life. Bioinformatics 28:i458–i465. doi:10.1093/bioinformatics/bts390

Goldberg T, Hecht M, Hamp T et al (2014) LocTree3 prediction of localization. Nucleic Acids Res 42:W350–W355. doi:10.1093/nar/gku396

Gomi M, Sonoyama M, Mitaku S (2004) High performance system for signal peptide prediction: SOSUIsignal. Chem-Bio Inform J 4:142–147. doi:10.1273/cbij.4.142

Gorodkin J (2004) Comparing two K-category assignments by a K-category correlation coefficient. Comput Biol Chem 28:367–374. doi:10.1016/j.compbiolchem.2004.09.006

Haft DH, Selengut JD, Richter RA et al (2013) TIGRFAMs and genome properties in 2013. Nucleic Acids Res 41:D387–D395. doi:10.1093/nar/gks1234

Hennerdal A, Elofsson A (2011) Rapid membrane protein topology prediction. Bioinformatics 27:1322–1323. doi:10.1093/bioinformatics/btr119

Hertz JA, Krogh AS, Palmer RG (1991) Introduction to the theory of neural computation. Westview Press, Redwood City, Calif

Hessa T, Meindl-Beinker NM, Bernsel A et al (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450:1026–1030. doi:10.1038/nature06387

Hiller K, Grote A, Scheer M et al (2004) PrediSi: prediction of signal peptides and their cleavage positions. Nucl Acids Res 32:W375–W379. doi:10.1093/nar/gkh378

Hobohm U, Scharf M, Schneider R, Sander C (1992) Selection of representative protein data sets. Protein Sci 1:409–417. doi:10.1002/pro.5560010313

Höglund A, Dönnes P, Blum T et al (2006) MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. Bioinformatics 22:1158–1165. doi:10.1093/bioinformatics/btl002

Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721–728. doi:10.1093/bioinformatics/17.8.721

Janeček Š, Svensson B, Russell RRB (2000) Location of repeat elements in glucansucrases of Leuconostoc and Streptococcus species. FEMS Microbiol Lett 192:53–57. doi:10.1111/j.1574-6968.2000.tb09358.x

Jones DT (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 23:538–544. doi:10.1093/bioinformatics/btl677

Jones P, Binns D, Chang H-Y et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. doi:10.1093/bioinformatics/btu031

Juncker AS, Willenbrock H, von Heijne G et al (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662. doi:10.1110/ps.0303703

Käll L, Krogh A, Sonnhammer EL (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics 21:i251–i257. doi:10.1093/bioinformatics/bti1014

Käll L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338:1027–1036. doi:10.1016/j.jmb.2004.03.016

Klee E, Ellis L (2005) Evaluating eukaryotic secreted protein prediction. BMC Bioinform 6:1–7. doi:10.1186/1471-2105-6-256

Klein P, Kanehisa M, DeLisi C (1985) The detection and classification of membrane-spanning proteins. Biochim Biophys Acta BBA Biomembr 815:468–476. doi:10.1016/0005-2736(85)90375-X

Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems 25. Curran Associates, Inc., pp 1097–1105

Krogh A, Brown M, Mian IS et al (1994) Hidden Markov models in computational biology: applications to protein modeling. J Mol Biol 235:1501–1531. doi:10.1006/jmbi.1994.1104

Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. doi:10.1006/jmbi.2000.4315

Krogh S, Jørgensen ST, Devine KM (1998) Lysis genes of the *Bacillus subtilis* defective prophage PBSX. J Bacteriol 180:2110–2117

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132. doi:10.1016/0022-2836(82)90515-0

Litou ZI, Bagos PG, Tsirigos KD et al (2008) Prediction of cell wall sorting signals in gram-positive bacteria with a hidden Markov model: application to complete genomes. J Bioinform Comput Biol 6:387–401. doi:10.1142/S0219720008003382

López R, García E (2004) Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. FEMS Microbiol Rev 28:553–580. doi:10.1016/j.femsre.2004.05.002

Lu Z, Szafron D, Greiner R et al (2004) Predicting subcellular localization of proteins using machine-learned classifiers. Bioinformatics 20:547–556. doi:10.1093/bioinformatics/btg447

Magnan CN, Baldi P (2014) SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. Bioinformatics 30:2592–2597. doi:10.1093/bioinformatics/btu352

McGeoch DJ (1985) On the predictive recognition of signal peptide sequences. Virus Res 3:271–286. doi:10.1016/0168-1702(85)90051-6

Menne KML, Hermjakob H, Apweiler R (2000) A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics 16:741–742. doi:10.1093/bioinformatics/16.8.741

Mitchell A, Chang H-Y, Daugherty L et al (2015) The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res 43:D213–D221. doi:10.1093/nar/gku1243

Möller S, Croning MDR, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646–653. doi:10.1093/bioinformatics/17.7.646

Nair R, Rost B (2002a) Sequence conserved for subcellular localization. Protein Sci 11:2836–2847. doi:10.1110/ps.0207402

Nair R, Rost B (2002b) Inferring sub-cellular localization through automated lexical analysis. Bioinformatics 18(Suppl 1):S78–S86. doi:10.1093/bioinformatics/18.suppl_1.S78

Nakai K, Kanehisa M (1991) Expert system for predicting protein localization sites in gram-negative bacteria. Proteins Struct Funct Bioinform 11:95–110. doi:10.1002/prot.340110203

Nakashima H, Nishikawa K (1994) Discrimination of intracellular an extracellular proteins using amino acid composition and residue-pair frequencies. J Mol Biol 238:54–61. doi:10.1006/jmbi.1994.1267

Nielsen H, Brunak S, Engelbrecht J, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6. doi:10.1093/protein/10.1.1

Nielsen H, Engelbrecht J, von Heijne G, Brunak S (1996) Defining a similarity threshold for a functional protein sequence pattern: The signal peptide cleavage site. Proteins Struct Funct Bioinform 24:165–177. doi:10.1002/(SICI)1097-0134(199602)24:2<165::AID-PROT4>3.0.CO;2-I

Nielsen H, Krogh A (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol 6:122–130

Nielsen H, Wernersson R (2006) An overabundance of phase 0 introns immediately after the start codon in eukaryotic genes. BMC Genom 7:256. doi:10.1186/1471-2164-7-256

Nugent T, Jones D (2009) Transmembrane protein topology prediction using support vector machines. BMC Bioinform 10:159. doi:10.1186/1471-2105-10-159

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8:785–786. doi:10.1038/nmeth.1701

Punta M, Forrest LR, Bigelow H et al (2007) Membrane protein prediction methods. Methods 41:460–474. doi:10.1016/j.ymeth.2006.07.026

Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC (2008) Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World J Microbiol Biotechnol 24:2377–2382. doi:10.1007/s11274-008-9795-2

Rashid M, Saha S, Raghava G (2007) Support vector machine-based method for predicting subcellular localization of mycobacterial proteins using evolutionary information and motifs. BMC Bioinform 8:337. doi:10.1186/1471-2105-8-337

Reinhardt A, Hubbard T (1998) Using neural networks for prediction of the subcellular location of proteins. Nucleic Acids Res 26:2230–2236. doi:10.1093/nar/26.9.2230

Reynolds SM, Käll L, Riffle ME et al (2008) Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. PLoS Comput Biol 4:e1000213. doi:10.1371/journal.pcbi.1000213

Rey S, Gardy J, Brinkman F (2005) Assessing the precision of high-throughput computational and laboratory approaches for the genome-wide identification of protein subcellular localization in bacteria. BMC Genom 6:162. doi:10.1186/1471-2164-6-162

Rish I (2001) An empirical study of the naive Bayes classifier. In: IJCAI 2001 workshop on empirical methods in artificial intelligence. IBM New York, pp 41–46

Rose RW, Brüser T, Kissinger JC, Pohlschröder M (2002) Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. Mol Microbiol 45:943–950. doi:10.1046/j.1365-2958.2002.03090.x

Sander C, Schneider R (1991) Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins Struct Funct Bioinform 9:56–68. doi:10.1002/prot.340090107

Schneider TD, Stephens RM (1990) Sequence logos: a new way to display consensus sequences. Nucleic Acids Res 18:6097–6100. doi:10.1093/nar/18.20.6097

Schölkopf B, Smola A (1998) Support vector machines. In: Encyclopedia of Biostatistics. Wiley, New York. doi:10.1002/0470011815.b2a14038

Shah DSH, Joucla G, Remaud-Simeon M, Russell RRB (2004) Conserved repeat motifs and glucan binding by Glucansucrases of oral Streptococci and *Leuconostoc mesenteroides*. J Bacteriol 186:8301–8308. doi:10.1128/JB.186.24.8301-8308.2004

Shatkay H, Höglund A, Brady S et al (2007) SherLoc: high-accuracy prediction of protein subcellular localization by integrating text and protein sequence data. Bioinformatics 23:1410–1417. doi:10.1093/bioinformatics/btm115

Shen H-B, Chou K-C (2007) Gpos-PLoc: an ensemble classifier for predicting subcellular localization of Gram-positive bacterial proteins. Protein Eng Des Sel 20:39–46. doi:10.1093/protein/gzl053

Shen H-B, Chou K-C (2009) Gpos-mPLoc: a top-down approach to improve the quality of predicting subcellular localization of Gram-positive bacterial proteins. Protein Pept Lett 16:1478–1484. doi:10.2174/092986609789839322

Sigrist CJA, de Castro E, Cerutti L et al (2013) New and continuing developments at PROSITE. Nucleic Acids Res 41:D344–D347. doi:10.1093/nar/gks1067

Sønderby SK, Sønderby CK, Nielsen H, Winther O (2015) Convolutional LSTM networks for subcellular localization of proteins. In: Dediu A-H, Hernández-Quiroz F, Martin-Vide C, Rosenblueth DA (eds) Algorithms for computational biology, pp 68–80. Springer, Berlin. doi:10.1007/978-3-319-21233-3_6

Stormo GD, Schneider TD, Gold L, Ehrenfeucht A (1982) Use of the "perceptron" algorithm to distinguish translational initiation sites in *E. coli*. Nucleic Acids Res 10:2997–3011. doi:10.1093/nar/10.9.2997

Szafron D, Lu P, Greiner R et al (2004) Proteome analyst: custom predictions with explanations in a web-based tool for high-throughput proteome annotations. Nucleic Acids Res 32:W365–W371. doi:10.1093/nar/gkh485

The UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. doi:10.1093/nar/gku989

Tjalsma H, Antelmann H, Jongbloed JDH et al (2004) Proteomics of Protein Secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev 68:207–233. doi:10.1128/MMBR.68.2.207-233.2004

Tjalsma H, Bolhuis A, Jongbloed JDH et al (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev 64:515–547. doi:10.1128/MMBR.64.3.515-547.2000

Tsirigos KD, Peters C, Shu N et al (2015) The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res 43:W401–W407. doi:10.1093/nar/gkv485

Tusnády GE, Simon I (2001) The HMMTOP transmembrane topology prediction server. Bioinformatics 17:849–850. doi:10.1093/bioinformatics/17.9.849

Tusnády GE, Simon I (2010) Topology prediction of helical transmembrane proteins: how far have we reached? Curr Protein Pept Sci 11:550–561. doi:10.2174/138920310794109184

Viklund H, Bernsel A, Skwark M, Elofsson A (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 24:2928–2929. doi:10.1093/bioinformatics/btn550

Viklund H, Elofsson A (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24:1662–1668. doi:10.1093/bioinformatics/btn221

Viklund H, Elofsson A (2004) Best α-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. Protein Sci 13:1908–1917. doi:10.1110/ps.04625404

von Heijne G (1983) Patterns of amino acids near signal-sequence cleavage sites. Eur J Biochem 133:17–21. doi:10.1111/j.1432-1033.1983.tb07424.x

von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683–4690. doi:10.1093/nar/14.11.4683

von Heijne G (1992) Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule. J Mol Biol 225:487–494. doi:10.1016/0022-2836(92)90934-C

von Heijne G, Abrahmsén L (1989) Species-specific variation in signal peptide design implications for protein secretion in foreign hosts. FEBS Lett 244:439–446. doi:10.1016/0014-5793(89)80579-4

von Heijne G, Gavel Y (1988) Topogenic signals in integral membrane proteins. Eur J Biochem 174:671–678. doi:10.1111/j.1432-1033.1988.tb14150.x

Wu Z-C, Xiao X, Chou K-C (2012) iLoc-Gpos: a multi-layer classifier for predicting the subcellular localization of singleplex and multiplex Gram-positive bacterial proteins. Protein Pept Lett 19:4–14. doi:10.2174/092986612798472839

Xiong HY, Alipanahi B, Lee LJ et al (2015) The human splicing code reveals new insights into the genetic determinants of disease. Science 347:1254806. doi:10.1126/science.1254806

Yu C-S, Chen Y-C, Lu C-H, Hwang J-K (2006) Prediction of protein subcellular localization. Proteins 64:643–651. doi:10.1002/prot.21018

Yu L, Guo Y, Li Y et al (2010a) SecretP: identifying bacterial secreted proteins by fusing new features into Chou's pseudo-amino acid composition. J Theor Biol 267:1–6. doi:10.1016/j.jtbi.2010.08.001

Yu NY, Wagner JR, Laird MR et al (2010b) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. doi:10.1093/bioinformatics/btq249

Zhang X, Li Y, Li Y (2009) Evaluating signal peptide prediction methods for Gram-positive bacteria. Biologia (Bratisl) 64:655–659. doi:10.2478/s11756-009-0118-3

Zhou M, Boekhorst J, Francke C, Siezen R (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. BMC Bioinform 9:173. doi:10.1186/1471-2105-9-173

Anchoring of LPXTG-Like Proteins to the Gram-Positive Cell Wall Envelope

Sara D. Siegel, Melissa E. Reardon and Hung Ton-That

Abstract In Gram-positive bacteria, protein precursors with a signal peptide and a cell wall sorting signal (CWSS)—which begins with an LPXTG motif, followed by a hydrophobic domain and a tail of positively charged residues—are targeted to the cell envelope by a transpeptidase enzyme call sortase. Evolution and selective pressure gave rise to six classes of sortase, i.e., SrtA-F. Only class C sortases are capable of polymerizing substrates harboring the pilin motif and CWSS into protein polymers known as pili or fimbriae, whereas the others perform cell wall anchoring functions. Regardless of the products generated from these sortases, the basic principle of sortase-catalyzed transpeptidation is the same. It begins with the cleavage of the LPXTG motif, followed by the cross-linking of this cleaved product at the threonine residue to a nucleophile, i.e., an active amino group of the peptidoglycan stem peptide or the lysine residue of the pilin motif. This chapter will summarize the efforts to identify and characterize sortases and their associated pathways with emphasis on the cell wall anchoring function.

Contents

S.D. Siegel · M.E. Reardon · H. Ton-That
Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, TX, USA

H. Ton-That (✉)
Department of Microbiology and Molecular Genetics, The University of Texas Medical School at Houston, 6431 Fannin Street, MSE R213, Houston, TX 77030, USA
e-mail: ton-that.hung@uth.tmc.edu

Current Topics in Microbiology and Immunology (2017) 404:159–175
DOI 10.1007/82_2016_8

Published Online: 21 April 2016

1 Introduction

More than 42 years ago, Sjöquist et al. (1972) might have provided the first physical evidence of cell wall anchoring of surface proteins in Gram-positive bacteria. By treating *Staphylococcus aureus* with lysostaphin, a cell wall hydrolase that cleaves the pentaglycine cross-bridge of staphylococcal peptidoglycan, they observed that protein A was released from the bacterium. Two decades later, Schneewind et al. (1992) reported that protein A is anchored to the bacterial cell wall via a cell wall sorting signal (CWSS) located at its C-terminus. Comprised of an LPXTG motif, a hydrophobic domain, and a positively charged tail, this CWSS is a common feature of cell wall-anchored surface proteins in Gram-positive bacteria (Schneewind et al. 1993). Because the LPXTG motif was found to be cleaved between threonine (T) and glycine (G), a processing pathway of cell wall-anchored proteins was postulated (Navarre and Schneewind 1994). A search for a possible proteolytic enzyme by the Schneewind laboratory led to the discovery of sortase SrtA, the transpeptidase that catalyzes cell wall anchoring of protein A (Mazmanian et al. 1999). To date, SrtA homologs have been identified in nearly all Gram-positive bacteria, with many coding multiple sortase proteins. Classification —based on primary sequences, substrate association, and degrees of homology— has divided sortase homologs into 6 classes, i.e., class A, B, C, D, E, and F (Comfort and Clubb 2004; Dramsi et al. 2005; Spirig et al. 2011). The class A sortases, with *S. aureus* SrtA as the founding member, are thought to perform a "housekeeping function" by anchoring a large number of LPXTG-harboring substrates to the cell wall (Ton-That et al. 2004). In contrast, surface proteins with a NPQTN motif are anchored to the bacterial peptidoglycan by class B sortases (Mazmanian et al. 2002). Class C sortases, which constitute the largest groups, function as polymerase enzymes that link monomeric LPXTG-containing pilin subunits into pilus polymers, as first reported in *Corynebacterium diphtheriae* (Ton-That and Schneewind 2003). This chapter will mainly focus on the mechanism of sortase-catalyzed cell wall anchoring of LPXTG-containing proteins, in addition to a brief discussion on other processes mediated by different classes of sortase enzymes.

2 The Cell Wall Sorting Signal

The first description of the LPXTG motif was reported by Schneewind et al. (1990), when they characterized T6, a trypsin-resistant surface protein of *Streptococcus pyogenes*. A sequence alignment of the C-terminal region of T6, protein A, and some other Gram-positive proteins revealed a conservation of the LPXTG motif, which is followed by a long stretch of hydrophobic and positively charged residues. Together, the three elements constitute the C-terminal CWSS, which was termed a few years later in a seminal report using protein A as a model substrate (Schneewind et al. 1992). Removal of any element of this tripartite signal caused protein A to be secreted into the extracellular milieu, whereas deletion of the LPXTG motif led to mislocalization of protein A; the protein was found in all subcellular compartments. When the signal peptide and the CWSS of protein A were fused to PhoA, a periplasmic alkaline phosphatase enzyme of *Escherichia coli*, the fusion protein was "sorted" to the staphylococcal cell wall in the same manner as that of protein A. To exclude the possibility that the signal peptide might influence cell wall anchoring, the CWSS of protein A was then fused to the staphylococcus enterotoxin B (SEB). Although SEB is normally secreted, the SEB fusion protein (SEB-$Spa_{490\text{-}524}$) was found to be exclusively associated within the cell wall compartment (Schneewind et al. 1993). Furthermore, while the LPXTG motif was later shown to be the site for sortase cleavage and cell wall attachment (see below), the charged tail appears to be a retention signal to keep protein A within the secretory pathway, and its spacing from the LPXTG motif determined by the hydrophobic domain is also a contributing factor (Schneewind et al. 1993).

Interestingly, the fusion protein experiments above suggested that the CWSS contains an essential element conferring sortase specificity; within a CWSS, the conserved LPXTG motif is seemingly obvious for sortase specificity. In fact, all *S. aureus* surface proteins predicted to be SrtA substrates contain the canonical LPXTG motif (Mazmanian et al. 2001), but substrates for SrtB (class B sortase) harbor the NPQTN motif (Mazmanian et al. 2002). However, it remains unclear whether the rest of the CWSS sequence would contribute to sortase specificity.

3 Sortase SrtA and the Cell Wall Anchoring Pathway

3.1 *Identification of* S. aureus *SrtA*

Analyses of the proteolytic cleavage of the LPXTG motif led to the initial deduction that a transpeptidase enzyme was responsible for the processing and anchoring of surface proteins containing a CWSS (Navarre and Schneewind 1994). To identify this factor, Mazmanian et al. constructed a library of chemically mutagenized temperature-sensitive *S. aureus* mutants expressing the SEB fusion protein

mentioned above. The library was used to screen for clones that are defective in cell wall anchoring of this reporter protein and found one that was mapped to a gene termed *srtA* (surface protein sorting A) (Mazmanian et al. 1999). An insertion mutant of *srtA* was generated, and the *srtA* mutant failed to display protein A on the bacterial surface. Significantly, this mutant was also attenuated in a mouse model of infection (Mazmanian et al. 2000). The fact that a mutant of *srtA* was obtained by allelic exchange refutes the initial conjecture of *srtA* essentiality in *S. aureus* with the intent usage of the temperature-sensitive mutants.

3.2 *Transpeptidase Activities of* S. aureus *SrtA*

S. aureus SrtA was cloned in and purified from *E. coli*; this recombinant protein displayed transpeptidase activities by cleaving a quencher away from a fluorescent peptide *Dabcyl*-QA**LPETG**EE-*Edans* between T and G and linking the carboxyl group of cleaved threonine to the amino group of triglycines (Ton-That et al. 1999, 2000). This is consistent with the cell wall-linked structure of the SEB-Spa fusion protein isolated from *S. aureus* after lysostaphin treatment (Ton-That et al. 1997). The transpeptidase activity of SrtA requires the conserved residues Cys^{184} and His^{120} (Ton-That et al. 2002) and is enhanced by the addition of calcium ions (Ilangovan et al. 2001). Furthermore, Arg^{197}, part of the enzyme's structural triad, was shown to aid optimal catalysis by SrtA, possibly by facilitating formation of the Cys^{184} thiolate during transpeptidation reactions (Marraffini et al. 2004). Mechanistically, the Cys^{184} thiolate of the sortase enzyme initiates a nucleophilic attack of the carbonyl carbon of the substrate threonine residue, shifting electrons to the oxygen and creating an oxyanion at this position, and a sortase–substrate thioester intermediate. The active site arginine has been proposed to stabilize this tetrahedral intermediate by forming an oxyanion hole, which was shown for to exist in the *S. aureus* SrtB structure (Frankel et al. 2007; Jacobitz et al. 2014).

The three-dimensional structure of sortase was first solved with *S. aureus* SrtA by nuclear magnetic resonance (NMR) and revealed that the enzyme forms a β-barrel structure comprised of eight β-strands and two peripheral α-helices (Ilangovan et al. 2001). The β6-β7 strands form a highly flexible β-loop structure and a calcium coordination site. The catalytic residues His120, Cys184, and Arg197 lie within the hydrophobic groove of the β-loop (Ilangovan et al. 2001; Bentley et al. 2008). It was proposed that the binding of a calcium ion to sortase stabilizes the active site β-loop structure, hence increasing sortase activity (Ilangovan et al. 2001; Naik et al. 2006). It was shown that an LPXTG substrate peptide binds to the hydrophobic groove and interacts with the reactive site (Suree et al. 2009; Zong et al. 2004). However, these two studies differed in the conformation of the bound substrate. Further evidence that this β-loop also plays a critical role in specificity was shown by mutagenesis (Piotukh et al. 2011; Bentley et al. 2007). Using a directed evolution approach to identify sortase A molecules that have an altered specificity, the residues involving expanded specificity were mapped to changes in

the β6-β7 loop (Piotukh et al. 2011). As previously mentioned, *S. aureus* SrtB recognizes the NPQTN motif, and its structure determined by X-ray crystallography also displays the similar β-barrel conformation (Zong et al. 2004; Zhang et al. 2004). Consistent with the findings above, it was shown that swapping the β6-β7 loop from SrtB into SrtA, the hybrid SrtA was able to recognize the NPQTG substrate, but unable to catalyze transpeptidation reactions (Bentley et al. 2007).

3.3 *The Cell Wall Anchoring Pathway Mediated by Sortase SrtA*

In the current model (Schneewind and Missiakas 2014; Reardon-Robinson et al. 2015a, b) (see Fig. 1), it was proposed that protein precursors destined for the sortase machine are transported across the cytoplasmic membrane by the Sec apparatus in unfolded states. In high GC-content Gram-positive bacteria or actinobacteria, such as *C. diphtheriae* and *Actinomyces oris*, posttranslocational folding of these precursors is catalyzed by a membrane-bound thiol–disulfide oxidoreductase named MdbA (Reardon-Robinson et al. 2015a, b; Reardon-Robinson and Ton-That 2016). The folded precursors are embedded into the membrane via the hydrophobic domain of their CWSS. There, sortase cleaves the LPXTG motif of a substrate between T and G, forming an acyl–enzyme intermediate with the substrate via a thioester bond. Subsequently, the amino group of the stem peptide within a Lipid II precursor

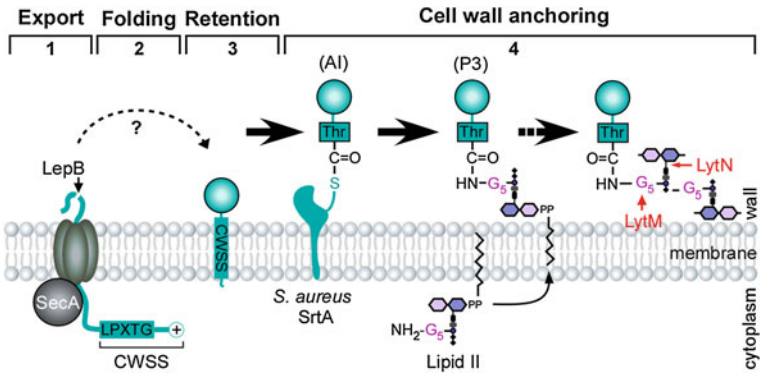


Fig. 1 Cell wall anchoring of surface proteins catalyzed by sortase SrtA. The pathway is modeled with *S. aureus* SrtA (see text for details). It begins with the transport of unfolded protein precursors with a signal peptide and the cell wall sorting signal (*CWSS*), after their synthesis in the cytoplasm, by the SecA-mediated translocation, followed by cleavage of the signal peptide by a signal peptidase (*LepB*) (Step 1). After folding (Step 2) by uncharacterized factors and insertion into the membrane (Step 3), the precursors are cleaved by SrtA, forming an acyl–enzyme intermediate (*AI*), which is then resolved by a nucleophilic attack by Lipid II via the amino group of pentaglycine (G_5), resulting in the intermediate P3. This product is incorporated into the cell wall peptidoglycan (Step 4)

resolves this intermediate (AI), hence linking the cleaved substrate via the carboxyl group of Thr to the Lipid II molecule (P3). As the bacterial cell wall is built from Lipid II molecules, surface proteins are subsequently anchored to the cell envelope. This model has been supported by several lines of evidence. (1) Treatment of *S. aureus* cells with sodium azide, an inhibitor of SecA, resulted in accumulation of protein precursors (Ton-That and Schneewind 1999). (2) In *C. diphtheriae* and *A. oris*, genetic disruption of *mdbA* led to secretion of degraded surface proteins, concomitant of abrogation of their surface assembly. (3) As mentioned above, cell wall anchoring of reporter proteins was shown to require sortase as demonstrated by in vitro and in vivo studies. (4) Treatment of *S. aureus* cells with sortase inhibitors also prevents cell wall anchoring (Ton-That and Schneewind 1999; Zhang et al. 2014). (5) Evidence of Lipid II as a sortase substrate comes from [^{32}P]-labeling experiments, whereby a [^{32}P]-labeled cell wall-anchored molecules were shown to bind to nisin, an inhibitor forming a complex with Lipid II (Perry et al. 2002). This finding was also supported by in vitro studies that showed sortase-catalyzed transpeptidation reactions that link the cleaved product of a LPXTG substrate peptide to pentaglycine or Lipid II molecules (Ton-That et al. 2000; Ruzin et al. 2002) (Fig. 1).

While many major steps of the above pathway are inarguably supported by experimental data, the presence of AI and P3 species in vivo has not been isolated. It is conceivable that their formation in the cell is so transient that the currently employed experimental conditions failed to detect them. The usage of fluorescent LPXTG substrates recently made available (Hansenova Manaskova et al. 2014) in combination with inhibitors, which potentially decelerate sortase-catalyzed transpeptidation reactions, and [^{32}P]-labeling may facilitate isolation of these intermediates.

3.4 *Modulation of Cell Wall Anchoring and Extracellular Release of LPXTG-Containing Proteins*

It has been observed in Gram-positive bacteria that cell wall-anchored proteins including pili are released into the extracellular milieu during bacterial growth, presumably as by-products of cell wall turnover (Ton-That and Schneewind 2003; Gaspar and Ton-That 2006; Mishra et al. 2007; Becker et al. 2014). This argument may be valid, because the covalent linkage of surface proteins to the peptidoglycan is irreversible. However, Becker and colleagues showed that the immunomodulatory protein A of *S. aureus* is actively released, peaked at early logarithmic growth (Becker et al. 2014). It was then shown that LytM, a glycyl-glycine endopeptidase, is involved in this release by cleaving the penta-glycine cross-bridges, the attachment site of protein A to the staphylococcal peptidoglycan. In addition, LytN, a cell wall hydrolase, contributes to protein A release by cleaving amino sugars (Becker et al. 2014). Given that LytM is regulated by the two-component system WalKR

(Dubrac et al. 2007), it is likely that the LytM-mediated cleavage of protein A is regulated. Since protein A is a key factor for *S. aureus* immune evasion, it is proposed that protein A is released early during infection to help the pathogen elude the host immune responses when bacterial numbers are low (Becker et al. 2014).

Remarkably, a similar proteolytic cleavage of a cell wall-anchored protein has been observed in *Clostridium difficile*. This Gram-positive spore-forming bacterium expresses a cell wall-anchored adhesin named CD2831. Expression of CD2831 is under the control of a c-diGMP-dependent type II riboswitch (Peltier et al. 2015); it was shown that CD2831 is proteolytically cleaved at the C-terminus by the zinc metalloprotease ZmpI in cells with low levels of c-diGMP, hence released extracellularly. Intriguingly, transcription of CD2831 is upregulated at high levels of c-diGMP, while transcription of *zmpI* is repressed (Soutourina et al. 2013). Considering that high levels of c-diGMP are typically associated with biofilm formation, hence sessile, whereas low levels of c-diGMP promote motility (Sondermann et al. 2012; Hengge 2009; Romling et al. 2013), Peltier and colleagues propose that ZmpI-mediated release of cell wall-anchored proteins may be involved in or promote transitioning from a sessile to a mobile state or vice versa (Peltier et al. 2015). Thus, Gram-positive bacteria might have solved an "irreversible" issue of covalent linkages produced by sortase-mediated transpeptidation.

4 Essentiality of Sortase SrtA in Cell Wall Anchoring of Surface Proteins in *Actinomyces oris*

Actinomyces oris is an oral biofilm-forming actinobacterium that possesses two class C sortases SrtC1 and SrtC2 and a class A sortase, i.e., SrtA. While the former are involved in pilus assembly (Mishra et al. 2007; Wu et al. 2011), the latter is predicted to catalyze cell wall anchoring of pilus polymers and fourteen LPXTG-containing surface proteins (Reardon-Robinson et al. 2014). In contrast to *srtA* in other Gram-positive bacteria studied to date, *A. oris srtA* is an essential gene. To investigate its role in cell wall anchoring of surface proteins, Wu and colleagues made several failed attempts to generate a *srtA* deletion mutant. Essentiality of *A. oris srtA* was demonstrated by a conditional gene deletion method, whereby chromosomal *srtA* can be removed when SrtA is ectopically provided under the control of a *tetR*-inducible promoter (Wu et al. 2014). When grown in the absence of inducers for SrtA expression, the *srtA* mutant ceased to grow. Strikingly, *srtA* depleted cells exhibited aberrant cell morphology with multiple septa and expansion of the cell envelope. Interestingly, envelope expansion was only observed at one end of the cells.

Using transposon mutagenesis, Wu and coworkers found five sets of mutants that suppress the lethal phenotypes of *srtA* deletion (Wu et al. 2014). The first set of suppressors was mapped to genes coding for a LytR-CpsA-Psr (LCP)-like protein and a LPXTG-containing protein named GspA. It has been proposed that LCP

attaches glycan strands to a membrane-bound form of GspA, which is then anchored to the cell wall by sortase SrtA. In the absence of SrtA, the glycosylated GspA polymers accumulated in the membrane, which was proposed to be toxic. This could be possibly due to membrane jamming of the SecA-dependent transport system that causes cell stress, growth arrest, and ultimately cell death (Wu et al. 2014). This conjecture is supported by the fact that *A. oris* cells expressing a GspA mutant devoid of the CWSS are able to survive without *srtA*. It still remains unclear how other suppressor mutants are involved in this lethality; many of the targeted genes encode transporters and ATPases. Perhaps, they may be required for transport of glycan precursors that are substrates for LCP-catalyzed glycosylation of GspA. Without glycan strands, a stalled membrane-bound form of GspA may not so be detrimental.

The unexpected finding that *A. oris srtA* is essential in *A. oris* provided a convenient cell-based assay to identify inhibitors that can penetrate the Gram-positive cell envelope and inactivate sortase activities, hence bacterial virulence. This class of inhibitors—anti-virulence inhibitors—may not impose a selective pressure on other Gram-positive pathogens like inhibitors targeting essential genes or pathway.

5 Cell Wall Anchoring of Iron-Regulated Surface-Determinant (Isd) Proteins Catalyzed by Sortase SrtB

Many Gram-positive Firmicutes, including *S. aureus*, *Bacillus anthracis*, *Listeria monocytogenes*, and *S. pyogenes*, harbor the *srtB* gene, which encodes a class B sortase enzyme (Spirig et al. 2011; Mandlik et al. 2008). Except for *S. pyogenes*, SrtB is assigned to iron acquisition that was first reported with the Isd system in *S. aureus* (Isd for iron-regulated surface determinants) (Mazmanian et al. 2003). In *S. aureus*, *srtB* is part of an eight-gene cluster consisting of *isdA-G* and *srtB*, while *isdH* and *isdI* are located elsewhere in the chromosome (Mazmanian et al. 2003; Skaar et al. 2004). The gene products of this cluster and the other two constitute the major heme utilization system in *S. aureus*, which is comprised of a hemoglobin/haptoglobin and heme-scrounging platform (IsdH, IsdB, and IsdA), a heme transport apparatus (IsdC and IsdDEF), and a heme-degrading machine (IsdG and IsdI) (Hammer and Skaar 2011). IsdA, IsdB, and IsdH are cell wall-anchored proteins containing the conserved LPXTG motif that is recognized by SrtA, whereas IsdC with the NQPTN motif is the only known cell wall-anchored substrate of SrtB (Mazmanian et al. 2003). Intriguingly, these proteins have distinct topologies in the cell envelope; unlike IsdA, IsdB, or IsdH, IsdC appears to be buried within the bacterial peptidoglycan (Marraffini and Schneewind 2005), supporting the model that IsdC is part of a relay system that transfers heme from IsdA,

IsdB, and IsdH to the membrane-bound lipoprotein IsdE (Hammer and Skaar 2011). Similar to *S. aureus*, *isdC* is the first gene of the 8-gene *isd* cluster in *B. anthracis* (Skaar et al. 2006). It was shown that IsdC and its cognate sortase SrtB are required for heme–iron scavenging (Maresso et al. 2006).

6 Cell Wall Anchoring of Pilus Polymers by Non-polymerizing Sortase Enzymes

Many Gram-positive bacteria, including *C. diphtheriae*, *A. oris*, *B. cereus*, enterococci, several streptococcal species, lactococci and lactobacilli, produce covalently linked protein polymers known as pili or fimbriae that are assembled by class C sortase enzymes (Ton-That and Schneewind 2003; Mishra et al. 2007; Budzik et al. 2007; Nallapareddy et al. 2006; Lauer et al. 2005; Dramsi et al. 2006; Abbot et al. 2007; Manetti et al. 2007; Falker et al. 2008; Hilleringmann et al. 2008; Meyrand et al. 2013; Rintahaka et al. 2014). The mechanism of sortase-catalyzed pilus assembly was first described with the prototype SpaA pili in *C. diphtheriae* (Ton-That and Schneewind 2003). The SpaA pilus, expressed from the gene cluster *spaA–srtA–spaB–spaC*, is comprised of the pilus shaft SpaA, the tip pilin SpaC, and the pilus base SpaB. Like protein A of *S. aureus*, each Spa pilin harbors a CWSS, in addition to a signal peptide sequence targeted by SecA-mediated translocation. Additionally, SpaA contains a pilin motif with the conserved lysine residue essential for pilin cross-linking (Ton-That and Schneewind 2003; Ton-That et al. 2004). Like most of Gram-positive pilus systems studied to date, polymerization of the SpaA pilus requires a class C sortase, initially termed SrtA expressed by the same gene cluster (Ton-That and Schneewind 2003). The CWSS and the pilin motif appear to be the only two elements critical for pilus polymerization. A fusion protein of staphylococcus enterotoxin B (SEB), N-terminally fused to the SpaA fragment, which encompasses the signal peptide sequence and the pilin motif, and C-terminally linked to the SpaA CWSS, can be polymerized by sortase SrtA (Ton-That et al. 2004). Reminiscent of sortase-catalyzed transpeptidation that links surface proteins to the cell wall, pilus polymerization catalyzed by corynebacterial SrtA occurs in two steps; first is the cleavage of the SpaA LPXTG motif by SrtA, leading to the formation of an acyl–enzyme intermediate. A nucleophilic attack by the amino group of the pilin motif lysine residue from an incoming SpaA pilin subunit resolves this intermediate, resulting in extension of pilus polymers (Ton-That and Schneewind 2004). Pilus polymerization is switched to cell wall anchoring when the pilus base SpaB is incorporated (Mandlik et al. 2008).

The notion that SpaA pili are covalently linked to the bacterial peptidoglycan came from the experiment that utilized mutanolysin, a cell wall hydrolase that cleaves the *N*-acetyl-muramyl-(1 → 4)-*N*-acetylglucosamine linkage of the cell wall polysaccharides. Treatment of corynebacterial cells with mutanolysin released pilus polymers from the cell envelope, and these isolated polymers were resistant to

hot SDS (Ton-That and Schneewind 2003). To biochemically address this issue, Budzik and colleagues isolated cell wall-linked BcpA pili, which are comprised of the pilus shaft BcpA and the tip pilin BcpB (Budzik et al. 2009), after muralytic digestion of the *B. cereus* cell wall with mutanolysin and purification of the released pili by affinity chromatography. Using mass spectrometry analysis, they demonstrated that the BcpA pilus is covalently linked to the side chain amino group of diaminopimelic acid within the peptidoglycan stem peptide (Budzik et al. 2008).

Unlike the *B. cereus* BcpA pilus, in which the last BcpA pilus shaft is linked to the peptidoglycan (see above) (Budzik et al. 2008), the *C. diphtheriae* SpaA pilus harbors the pilus base SpaB that is covalently attached to the cell wall (Mandlik et al. 2008). The linkage joining the last SpaA subunit and SpaB is formed between the threonine residue of the SpaA LPETG motif and the lysine residue K139 near the C-terminus of SpaB (Mandlik et al. 2008). In turn, SpaB is linked to the peptidoglycan by the housekeeping sortase SrtF (Mandlik et al. 2008; Swaminathan et al. 2007), a class E sortase (Spirig et al. 2011). The biphasic mode of pilus assembly, i.e., the cooperation between a class C sortase and a non-polymerizing sortase for pilus polymerization and cell wall anchoring of pilus polymers (Mandlik et al. 2008), appears to a universal feature in the assembly process of many Gram-positive pili (Budzik et al. 2007; Nobbs et al. 2008; Nielsen et al. 2013; Sillanpaa et al. 2013).

Based on recent findings of posttranslocational folding of unfolded pilin precursors transported by the Sec apparatus (Reardon-Robinson et al. 2015a, b; Reardon-Robinson and Ton-That 2016), a revised biphasic model of sortase-catalyzed pilus polymerization and cell wall anchoring is proposed with the

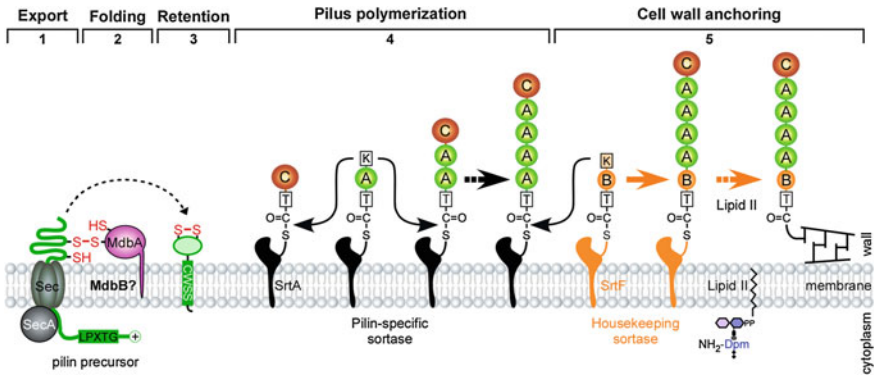


Fig. 2 Pilus biogenesis in *Corynebacterium diphtheriae.* The pathway is modeled with the SpaABC-type pilus with SpaA forming the shaft, SpaC at the tip, and SpaB at the base. Pilus assembly begins with the SecA-mediated transport of unfolded Spa prepilins (Step 1). The thiol–disulfide oxidoreductase MdbA catalyzes oxidative protein folding (Step 2), leading to insertion of the precursors (Step 3). The pilin-specific sortase SrtA catalyzes pilus polymerization, linking pilin subunits via lysine-mediated transpeptidation reactions (Step 4). Pilus polymerization is terminated when SpaB enters the pilus base, which is in turn anchored to the cell wall by the housekeeping sortase SrtF (Step 5). It is proposed that another oxidoreductase termed MdbB is required for reoxidation of MdbA. *Dashed arrows* indicate multi-steps; adapted from Reardon-Robinson et al. (2015a, b; Reardon-Robinson and Ton-That 2016; Mandlik et al. 2008)

C. diphtheriae SpaA pilus as an example (Fig. 2). Accordingly, after their synthesis in the cytoplasm, pilin precursors with a CWSS are transported across the cytoplasmic membrane by SecA-mediated translocation in an unfolded state (Step 1). The membrane-bound oxidoreductase MdbA catalyzes oxidative folding of the precursors emerging from the Sec translocon (Step 2). The folded precursors, embedded in the membrane (Step 3), are subject to sortase-mediated polymerization (Step 4). Pilus polymerization is terminated when SpaB enters the pilus base, and the resulting pilus polymers are then attached to the cell wall via Lipid II by the housekeeping sortase SrtF, a class E sortase (Step 5). Of note, the oxidative protein folding pathway that is coupled with protein translocation and pilus assembly appears to be conserved in actinobacteria (Reardon-Robinson and Ton-That 2016). However, it is not clear that a similar pathway would be found for pilus assembly and cell wall anchoring in the Gram-positive Firmicutes, given the cysteine exclusion of exported proteins in these organisms (Daniels et al. 2010).

7 Perspectives

The basic principle of sortase-catalyzed transpeptidation is the cleavage of the LPXTG motif, followed by the cross-linking of the cleaved product at the threonine residue to a nucleophile (e.g., the amino group of the peptidoglycan stem peptide or the pilin motif lysine residue). The products of sortase-catalyzed transpeptidation reactions in vivo can be in the forms of monomers (e.g., protein A) or polymers (e.g., pili). Regardless, the final products are ultimately anchored to the peptidoglycan. Thus, it is abundantly clear that sortase-catalyzed cell wall anchoring of surface proteins is a conserved feature of cell envelope assembly in Gram-positive bacteria. Since many Gram-positive cell wall-anchored proteins are virulence factors (Navarre and Schneewind 1999), inhibition of sortase activity would render the pathogenic potential of bacterial pathogens. Consistent with the important role of sortase in surface assembly of virulence factors, genetic disruption of sortase in vivo results in attenuation of bacterial virulence (Mazmanian et al. 2000; Garandeau et al. 2002; Paterson and Mitchell 2006; Kemp et al. 2007; Guiton et al. 2010). Efforts have been made to find sortase inhibitors that can be developed as an anti-infective therapy (Maresso and Schneewind 2008; Cascioferro et al. 2015). Notably, using a virtual screening approach, Zhang et al. identified thiadiazole derivatives that inhibit sortase activity in vitro and in vivo. Administration of mice with a lead compound prior to *S. aureus* infection improved the animal survival in a bacteremia model (Zhang et al. 2014). More recently, it was shown that a natural compound named chlorogenic acid (CHA) exhibits inhibitory sortase activity in vitro and can prevent mice from infection in a *S. aureus*-induced renal abscess model (Wang et al. 2015). None of the sortase inhibitors identified to date, however, have been shown to be effective to treat infection. Furthermore, it is unclear these inhibitors could have a broad range of sortase inhibition, i.e., capable of inhibiting different classes of sortase enyzmes.

Apart from sortase inhibition studies, the last several years have seen a shift in focus on sortase-related research. Sortase-mediated ligation or sortagging has become a great application of sortase-catalyzed transpeptidation that allows protein ligation, labeling, and modification (Tsukiji and Nagamune 2009; Popp et al. 2007; Schmohl and Schwarzer 2014). Sortase even finds its way to a therapeutic and imaging system for B cell lymphoma (Fang et al. 2016). Sortase-like mechanisms now have been reported in Gram-negative bacteria (Gorasia et al. 2015; Craig et al. 2011). On the other hand, many aspects of sortase-mediated mechanisms in Gram-positive bacteria have not been well understood, for example, sortase specificity and pilus hijacking of surface proteins. It is more than often that multiple sortases are expressed by one organism. In some bacteria, sortase-encoding genes are located in different locations in the chromosome; in others, they reside on the same gene clusters (Mandlik et al. 2008). In the case of *C. diphtheriae* SpaD pili, either SrtB or SrtC, class C sortases, is capable of polymerizing the pilus shaft pilin SpaD, but only SrtB is specific for incorporating SpaE into the pilus base (Gaspar and Ton-That 2006). Interestingly, corynebacterial sortase SrtD is able to polymerize the fimbrillin FimA of *A. oris* when this fimbrillin is expressed in *C. diphtheriae* (Ton-That et al. 2004). What determines sortase specificity is not well understood; perhaps, molecular mimicry is one possibility. In *A. oris*, CafA, a predicted cell wall-anchored protein not genetically linked to the type 2 fimbriae, is found at the tip of the FimA fimbrial structures, which also contain the canonical tip fimbrillin FimB. It has been postulated that the similarity of the CafA CWSS with that of FimB allows its recognition by SrtC2, a class C sortase (Reardon-Robinson et al. 2014).

Clearly, the LPXTG motif is not the only determinant of sortase specificity, since the *C. diphtheriae* SpaE CWSS contains the LALTG motif, predicted to be a substrate of class E sortase (Spirig et al. 2011), whereas the CWSS of SpaA and SpaD is similar, i.e., LPLTG and LPMTG, respectively (Gaspar and Ton-That 2006). However, *C. diphtheriae* SrtA is only able to recognize SpaA (Ton-That and Schneewind 2003). It is possible that the close proximity of substrates and their cognate sortase enzymes in the assembly center may impose some specificity. Nonetheless, what brings them there together is another intriguing question.

Acknowledgments We thank our laboratory members for their comments. Work in the laboratory is supported by the National Institute of Dental and Craniofacial Research of the NIH under Awards F31DE024004 (to M.E.R.-R.) and DE017382 and DE025015 (to H.T.-T.).

References

Abbot EL, Smith WD, Siou GP, Chiriboga C, Smith RJ, Wilson JA, Hirst BH, Kehoe MA (2007) Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. Cell Microbiol 9(7):1822–1833

Becker S, Frankel MB, Schneewind O, Missiakas D (2014) Release of protein A from the cell wall of *Staphylococcus aureus*. Proc Natl Acad Sci USA 111(4):1574–1579. doi:10.1073/pnas.1317181111 (1317181111 [pii])

Bentley ML, Gaweska H, Kielec JM, McCafferty DG (2007) Engineering the substrate specificity of *Staphylococcus aureus* Sortase A. The beta6/beta7 loop from SrtB confers NPQTN recognition to SrtA. J Biol Chem 282(9):6571–6581. doi:10.1074/jbc.M610519200 (M610519200 [pii])

Bentley ML, Lamb EC, McCafferty DG (2008) Mutagenesis studies of substrate recognition and catalysis in the sortase A transpeptidase from *Staphylococcus aureus*. J Biol Chem 283 (21):14762-14771. doi:10.1074/jbc.M800974200 (M800974200 [pii])

Budzik JM, Marraffini LA, Schneewind O (2007) Assembly of pili on the surface of *Bacillus cereus* vegetative cells. Mol Microbiol 66(2):495–510

Budzik JM, Oh SY, Schneewind O (2008) Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. J Biol Chem 283(52):36676–36686. doi:10.1074/jbc.M806796200 (M806796200 [pii])

Budzik JM, Oh SY, Schneewind O (2009) Sortase D forms the covalent bond that links bcpb to the tip of *Bacillus cereus* Pili. J Biol Chem 284(19):12989–12997. doi:10.1074/jbc.M900927200 (M900927200 [pii])

Cascioferro S, Raffa D, Maggio B, Raimondi MV, Schillaci D, Daidone G (2015) Sortase A Inhibitors: recent advances and future perspectives. J Med Chem 58(23):9108–9123. doi:10.1021/acs.jmedchem.5b00779

Comfort D, Clubb RT (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. Infect Immun 72(5):2710–2722

Craig JW, Cherry MA, Brady SF (2011) Long-chain N-acyl amino acid synthases are linked to the putative PEP-CTERM/exosortase protein-sorting system in gram-negative bacteria. J Bacteriol 193(20):5707–5715. doi:10.1128/JB.05426-11

Daniels R, Mellroth P, Bernsel A, Neiers F, Normark S, von Heijne G, Henriques-Normark B (2010) Disulfide bond formation and cysteine exclusion in gram-positive bacteria. J Biol Chem 285(5):3300–3309. doi:10.1074/jbc.M109.081398 (M109.081398 [pii])

Dramsi S, Trieu-Cuot P, Bierne H (2005) Sorting sortases: a nomenclature proposal for the various sortases of gram-positive bacteria. Res Microbiol 156(3):289–297

Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, Kojadinovic M, Lalioui L, Poyart C, Trieu-Cuot P (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60 (6):1401–1413

Dubrac S, Boneca IG, Poupel O, Msadek T (2007) New insights into the WalK/WalR (YycG/YycF) essential signal transduction pathway reveal a major role in controlling cell wall metabolism and biofilm formation in *Staphylococcus aureus*. J Bacteriol 189(22):8257–8269. doi:10.1128/JB.00645-07 (JB.00645-07 [pii])

Falker S, Nelson AL, Morfeldt E, Jonas K, Hultenby K, Ries J, Melefors O, Normark S, Henriques-Normark B (2008) Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Mol Microbiol 70(3):595–607. doi:10.1111/j.1365-2958.2008.06396.x (MMI6396 [pii])

Fang T, Duarte JN, Ling J, Li Z, Guzman JS, Ploegh HL (2016) Structurally defined alphaMHC-II nanobody-drug conjugates: a therapeutic and imaging system for B-cell lymphoma. Angew Chem Int Ed Engl 55(7):2416–2420. doi:10.1002/anie.201509432

Frankel BA, Tong Y, Bentley ML, Fitzgerald MC, McCafferty DG (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. Biochemistry 46 (24):7269–7278. doi:10.1021/bi700448e

Garandeau C, Reglier-Poupet H, Dubail I, Beretti JL, Berche P, Charbit A (2002) The sortase SrtA of Listeria monocytogenes is involved in processing of internalin and in virulence. Infect Immun 70(3):1382–1390

Gaspar AH, Ton-That H (2006) Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. J Bacteriol 188(4):1526–1533

Gorasia DG, Veith PD, Chen D, Seers CA, Mitchell HA, Chen YY, Glew MD, Dashper SG, Reynolds EC (2015) *Porphyromonas gingivalis* Type IX secretion substrates are cleaved and modified by a sortase-like mechanism. PLoS Pathog 11(9):e1005152. doi:10.1371/journal.ppat.1005152

Guiton PS, Hung CS, Hancock LE, Caparon MG, Hultgren SJ (2010) Enterococcal biofilm formation and virulence in an optimized murine model of foreign body-associated urinary tract infections. Infect Immun 78(10):4166–4175. doi:10.1128/IAI.00711-10

Hammer ND, Skaar EP (2011) Molecular mechanisms of *Staphylococcus aureus* iron acquisition. Annu Rev Microbiol 65:129–147. doi:10.1146/annurev-micro-090110-102851

Hansenova Manaskova S, Nazmi K, van Belkum A, Bikker FJ, van Wamel WJ, Veerman EC (2014) Synthetic LPETG-containing peptide incorporation in the *Staphylococcus aureus* cell-wall in a sortase A- and growth phase-dependent manner. PLoS ONE 9(2):e89260. doi:10.1371/journal.pone.0089260 (PONE-D-13-47535 [pii])

Hengge R (2009) Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol 7(4):263–273. doi:10.1038/nrmicro2109

Hilleringmann M, Giusti F, Baudner BC, Masignani V, Covacci A, Rappuoli R, Barocchi MA, Ferlenghi I (2008) Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. PLoS Pathog 4(3):e1000026. doi:10.1371/journal.ppat.1000026

Ilangovan U, Ton-That H, Iwahara J, Schneewind O, Clubb RT (2001) Structure of sortase, the transpeptidase that anchors proteins to the cell wall of *Staphylococcus aureus*. Proc Natl Acad Sci USA 98(11):6056–6061

Jacobitz AW, Wereszczynski J, Yi SW, Amer BR, Huang GL, Nguyen AV, Sawaya MR, Jung ME, McCammon JA, Clubb RT (2014) Structural and computational studies of the *Staphylococcus aureus* sortase B-substrate complex reveal a substrate-stabilized oxyanion hole. J Biol Chem 289(13):8891–8902. doi:10.1074/jbc (M113.509273 [pii])

Kemp KD, Singh KV, Nallapareddy SR, Murray BE (2007) Relative contributions of *Enterococcus faecalis* OG1RF sortase-encoding genes, srtA and bps (srtC), to biofilm formation and a murine model of urinary tract infection. Infect Immun 75(11):5399–5404. doi:10.1128/IAI.00663-07

Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R, Taddei AR, Mora M, Rappuoli R, Grandi G, Telford JL (2005) Genome analysis reveals pili in Group B Streptococcus. Science 309(5731):105

Mandlik A, Swierczynski A, Das A, Ton-That H (2008a) Pili in gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends Microbiol 16(1):33–40

Mandlik A, Das A, Ton-That H (2008b) The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. Proc Natl Acad Sci USA 105(37):14147–14152. doi:10.1073/pnas.0806350105 (0806350105 [pii])

Manetti AG, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F, Gambellini G, Bensi G, Mora M, Edwards AM, Musser JM, Graviss EA, Telford JL, Grandi G, Margarit I (2007) *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. Mol Microbiol 64(4):968–983

Maresso AW, Schneewind O (2008) Sortase as a target of anti-infective therapy. Pharmacol Rev 60(1):128–141. doi:10.1124/pr.107.07110 (pr.107.07110 [pii])

Maresso AW, Chapa TJ, Schneewind O (2006) Surface protein IsdC and Sortase B are required for heme-iron scavenging of *Bacillus anthracis*. J Bacteriol 188(23):8145–8152. doi:10.1128/JB.01011-06 (JB.01011-06 [pii])

Marraffini LA, Schneewind O (2005) Anchor structure of staphylococcal surface proteins. V. Anchor structure of the sortase B substrate IsdC. J Biol Chem 280(16):16263–16271. doi:10.1074/jbc.M500071200

Marraffini LA, Ton-That H, Zong Y, Narayana SV, Schneewind O (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):37763–37770

Mazmanian SK, Liu G, Ton-That H, Schneewind O (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. Science 285(5428):760–763

Mazmanian SK, Liu G, Jensen ER, Lenoy E, Schneewind O (2000) *Staphylococcus aureus* sortase mutants defective in the display of surface proteins and in the pathogenesis of animal infections. Proc Natl Acad Sci USA 97(10):5510–5515

Mazmanian SK, Ton-That H, Schneewind O (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Mol Microbiol 40(5):1049–1057

Mazmanian SK, Ton-That H, Su K, Schneewind O (2002) An iron-regulated sortase anchors a class of surface protein during *Staphylococcus aureus* pathogenesis. Proc Natl Acad Sci USA 99(4):2293–2298

Mazmanian SK, Skaar EP, Gaspar AH, Humayun M, Gornicki P, Jelenska J, Joachmiak A, Missiakas DM, Schneewind O (2003) Passage of heme-iron across the envelope of *Staphylococcus aureus*. Science 299(5608):906–909

Meyrand M, Guillot A, Goin M, Furlan S, Armalyte J, Kulakauskas S, Cortes-Perez NG, Thomas G, Chat S, Pechoux C, Dupres V, Hols P, Dufrene YF, Trugnan G, Chapot-Chartier MP (2013) Surface proteome analysis of a natural isolate of *Lactococcus lactis* reveals the presence of pili able to bind human intestinal epithelial cells. Mol Cell Proteomics 12 (12):3935–3947. doi:10.1074/mcp.M113.029066

Mishra A, Das A, Cisar JO, Ton-That H (2007) Sortase-catalyzed assembly of distinct heteromeric fimbriae in *Actinomyces naeslundii*. J Bacteriol 189(8):3156–3165

Naik MT, Suree N, Ilangovan U, Liew CK, Thieu W, Campbell DO, Clemens JJ, Jung ME, Clubb RT (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):1817–1826. doi:10.1074/jbc.M506123200 (M506123200 [pii])

Nallapareddy SR, Singh KV, Sillanpaa J, Garsin DA, Hook M, Erlandsen SL, Murray BE (2006) Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. J Clin Invest 116(10):2799–2807

Navarre WW, Schneewind O (1994) Proteolytic cleavage and cell wall anchoring at the LPXTG motif of surface proteins in gram-positive bacteria. Mol Microbiol 14(1):115–121

Navarre WW, Schneewind O (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev 63(1):174–229

Nielsen HV, Flores-Mireles AL, Kau AL, Kline KA, Pinkner JS, Neiers F, Normark S, Henriques-Normark B, Caparon MG, Hultgren SJ (2013) Pilin and sortase residues critical for endocarditis- and biofilm-associated pilus biogenesis in *Enterococcus faecalis*. J Bacteriol 195 (19):4484–4495. doi:10.1128/JB.00451-13 (JB.00451-13 [pii])

Nobbs AH, Rosini R, Rinaudo CD, Maione D, Grandi G, Telford JL (2008) Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in *Streptococcus agalactiae*. Infect Immun 76(8):3550–3560. doi:10.1128/IAI.01613-07 (IAI.01613-07 [pii])

Paterson GK, Mitchell TJ (2006) The role of *Streptococcus pneumoniae* sortase A in colonisation and pathogenesis. Microbes Infect 8(1):145–153

Peltier J, Shaw HA, Couchman EC, Dawson LF, Yu L, Choudhary JS, Kaever V, Wren BW, Fairweather NF (2015) Cyclic diGMP regulates production of sortase substrates of *Clostridium difficile* and their surface exposure through ZmpI protease-mediated cleavage. J Biol Chem 290 (40):24453–24469. doi:10.1074/jbc.M115.665091

Perry AM, Ton-That H, Mazmanian SK, Schneewind O (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. J Biol Chem 277(18):16241–16248

Piotukh K, Geltinger B, Heinrich N, Gerth F, Beyermann M, Freund C, Schwarzer D (2011) Directed evolution of sortase A mutants with altered substrate selectivity profiles. J Am Chem Soc 133(44):17536–17539. doi:10.1021/ja205630g

Popp MW, Antos JM, Grotenbreg GM, Spooner E, Ploegh HL (2007) Sortagging: a versatile method for protein labeling. Nat Chem Biol 3(11):707–708. doi:10.1038/nchembio.2007.31

Reardon-Robinson ME, Ton-That H (2016) Disulfide-bond-forming pathways in gram-positive bacteria. J Bacteriol 198(5):746–754. doi:10.1128/JB.00769-15

Reardon-Robinson ME, Wu C, Mishra A, Chang C, Bier N, Das A, Ton-That H (2014) Pilus hijacking by a bacterial coaggregation factor critical for oral biofilm development. Proc Natl Acad Sci U S A 111(10):3835–3840. doi:10.1073/pnas.1321417111 (1321417111 [pii])

Reardon-Robinson ME, Osipiuk J, Chang C, Wu C, Jooya N, Joachimiak A, Das A, Ton-That H (2015a) A disulfide bond-forming machine is linked to the sortase-mediated pilus assembly

pathway in the gram-positive bacterium *Actinomyces oris*. J Biol Chem 290(35):21393–21405. doi:10.1074/jbc.M115.672253

Reardon-Robinson ME, Osipiuk J, Jooya N, Chang C, Joachimiak A, Das A, Ton-That H (2015b) A thiol-disulfide oxidoreductase of the gram-positive pathogen *Corynebacterium diphtheriae* is essential for viability, pilus assembly, toxin production and virulence. Mol Microbiol. doi:10.1111/mmi.13172

Rintahaka J, Yu X, Kant R, Palva A, von Ossowski I (2014) Phenotypical analysis of the Lactobacillus rhamnosus GG fimbrial spaFED operon: surface expression and functional characterization of recombinant SpaFED pili in *Lactococcus lactis*. PLoS ONE 9(11):e113922. doi:10.1371/journal.pone.0113922

Romling U, Galperin MY, Gomelsky M (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev 77(1):1–52. doi:10.1128/MMBR.00043-12

Ruzin A, Severin A, Ritacco F, Tabei K, Singh G, Bradford PA, Siegel MM, Projan SJ, Shlaes DM (2002) Further evidence that a cell wall precursor [C(55)-MurNAc-(peptide)-GlcNAc] serves as an acceptor in a sorting reaction. J Bacteriol 184(8):2141–2147

Schmohl L, Schwarzer D (2014) Sortase-mediated ligations for the site-specific modification of proteins. Curr Opin Chem Biol 22:122–128. doi:10.1016/j.cbpa.2014.09.020

Schneewind O (1843) Missiakas D (2014) Sec-secretion and sortase-mediated anchoring of proteins in gram-positive bacteria. Biochim Biophys Acta 8:1687–1697. doi:10.1016/j.bbamcr.2013.11.009 (S0167-4889(13)00400-X [pii])

Schneewind O, Jones KF, Fischetti VA (1990) Sequence and structural characteristics of the trypsin-resistant T6 surface protein of group A streptococci. J Bacteriol 172(6):3310–3317

Schneewind O, Model P, Fischetti VA (1992) Sorting of protein A to the staphylococcal cell wall. Cell 70(2):267–281

Schneewind O, Mihaylova-Petkov D, Model P (1993) Cell wall sorting signals in surface proteins of gram-positive bacteria. EMBO J 12(12):4803–4811

Sillanpaa J, Chang C, Singh KV, Montealegre MC, Nallapareddy SR, Harvey BR, Ton-That H, Murray BE (2013) Contribution of individual Ebp pilus subunits of *Enterococcus faecalis* OG1RF to pilus biogenesis, biofilm formation and urinary tract infection. PLoS ONE 8(7):e68813. doi:10.1371/journal.pone.0068813 (PONE-D-13-17347 [pii])

Sjöquist J, Meloun B, Hjelm H (1972) Protein A isolated from Staphylococcus aureus after digestion with lysostaphin. Eur J Biochem 29(3):572–578

Skaar EP, Gaspar AH, Schneewind O (2004) IsdG and IsdI, heme-degrading enzymes in the cytoplasm of *Staphylococcus aureus*. J Biol Chem 279(1):436–443

Skaar EP, Gaspar AH, Schneewind O (2006) Bacillus anthracis IsdG, a heme-degrading monooxygenase. J Bacteriol 188(3):1071–1080. doi:10.1128/JB.188.3.1071-1080.2006

Sondermann H, Shikuma NJ, Yildiz FH (2012) You've come a long way: c-di-GMP signaling. Curr Opin Microbiol 15(2):140–146. doi:10.1016/j.mib.2011.12.008

Soutourina OA, Monot M, Boudry P, Saujet L, Pichon C, Sismeiro O, Semenova E, Severinov K, Le Bouguenec C, Coppee JY, Dupuy B, Martin-Verstraete I (2013) Genome-wide identification of regulatory RNAs in the human pathogen *Clostridium difficile*. PLoS Genet 9(5):e1003493. doi:10.1371/journal.pgen.1003493

Spirig T, Weiner EM, Clubb RT (2011) Sortase enzymes in gram-positive bacteria. Mol Microbiol 82(5):1044–1059. doi:10.1111/j.1365-2958.2011.07887.x

Suree N, Liew CK, Villareal VA, Thieu W, Fadeev EA, Clemens JJ, Jung ME, Clubb RT (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):24465–24477. doi:10.1074/jbc.M109.022624 M109.022624 [pii]

Swaminathan A, Mandlik A, Swierczynski A, Gaspar A, Das A, Ton-That H (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. Mol Microbiol 66(4):961–974. doi:10.1111/j.1365-2958.2007.05968.x (MMI5968 [pii])

Ton-That H, Schneewind O (1999) Anchor structure of staphylococcal surface proteins. IV. Inhibitors of the cell wall sorting reaction. J Biol Chem 274(34):24316–24320

Ton-That H, Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol 50(4):1429–1438

Ton-That H, Schneewind O (2004) Assembly of pili in gram-positive bacteria. Trends Microbiol 12(5):228–234

Ton-That H, Faull KF, Schneewind O (1997) Anchor structure of staphylococcal surface proteins. A branched peptide that links the carboxyl terminus of proteins to the cell wall. J Biol Chem 272(35):22285–22292

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 U S A 96(22):12424–12429

Ton-That H, Mazmanian SK, Faull KF, Schneewind O (2000) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. J Biol Chem 275(13):9876–9881

Ton-That H, Mazmanian SK, Alksne L, Schneewind O (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):7447–7452

Ton-That H, Marraffini LA, Schneewind O (2004a) Protein sorting to the cell wall envelope of gram-positive bacteria. Biochim Biophys Acta 1694(1–3):269–278

Ton-That H, Marraffini LA, Schneewind O (2004b) Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. Mol Microbiol 53(1):251–261

Tsukiji S, Nagamune T (2009) Sortase-mediated ligation: a gift from gram-positive bacteria to protein engineering. ChemBioChem 10(5):787–798. doi:10.1002/cbic.200800724

Wang L, Bi C, Cai H, Liu B, Zhong X, Deng X, Wang T, Xiang H, Niu X, Wang D (2015) The therapeutic effect of chlorogenic acid against *Staphylococcus aureus* infection through sortase A inhibition. Front Microbiol 6:1031. doi:10.3389/fmicb.2015.01031

Wu C, Mishra A, Yang J, Cisar JO, Das A, Ton-That H (2011) Dual function of a tip fimbrillin of Actinomyces in fimbrial assembly and receptor binding. J Bacteriol 193(13):3197–3206. doi:10.1128/JB.00173-11 (JB.00173-11 [pii])

Wu C, Huang IH, Chang C, Reardon-Robinson ME, Das A, Ton-That H (2014) Lethality of sortase depletion in Actinomyces oris caused by excessive membrane accumulation of a surface glycoprotein. Mol Microbiol 94(6):1227–1241. doi:10.1111/mmi.12780

Zhang R, Wu R, Joachimiak G, Mazmanian SK, Missiakas DM, Gornicki P, Schneewind O, Joachimiak A (2004) Structures of sortase B from *Staphylococcus aureus* and *Bacillus anthracis* reveal catalytic amino acid triad in the active site. Structure (Camb) 12(7):1147–1156

Zhang J, Liu H, Zhu K, Gong S, Dramsi S, Wang YT, Li J, Chen F, Zhang R, Zhou L, Lan L, Jiang H, Schneewind O, Luo C, Yang CG (2014) Antiinfective therapy with a small molecule inhibitor of *Staphylococcus aureus* sortase. Proc Natl Acad Sci USA 111(37):13517–13522. doi:10.1073/pnas.1408601111

Zong Y, Bice TW, Ton-That H, Schneewind O, Narayana SV (2004a) Crystal structures of *Staphylococcus aureus* sortase A and its substrate complex. J Biol Chem 279(30):31383–31389

Zong Y, Mazmanian SK, Schneewind O, Narayana SV (2004b) The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. Structure (Camb) 12(1):105–112

Spatial Organization of Cell Wall-Anchored Proteins at the Surface of Gram-Positive Bacteria

Shaynoor Dramsi and Hélène Bierne

Abstract Bacterial surface proteins constitute an amazing repertoire of molecules with important functions such as adherence, invasion, signalling and interaction with the host immune system or environment. In Gram-positive bacteria, many surface proteins of the "LPxTG" family are anchored to the peptidoglycan (PG) by an enzyme named sortase. While this anchoring mechanism has been clearly deciphered, less is known about the spatial organization of cell wall-anchored proteins in the bacterial envelope. Here, we review the question of the precise spatial and temporal positioning of LPxTG proteins in subcellular domains in spherical and ellipsoid bacteria (*Staphylococcus aureus*, *Streptococcus pyogenes*, *Streptococcus agalactiae* and *Enterococcus faecalis*) and in the rod-shaped bacterium *Listeria monocytogenes*. Deposition at specific sites of the cell wall is a dynamic process tightly connected to cell division, secretion, cell morphogenesis and levels of gene expression. Studying spatial occupancy of these cell wall-anchored proteins not only provides information on PG dynamics in responses to environmental changes, but also suggests that pathogenic bacteria control the distribution of virulence factors at specific sites of the surface, including pole, septa or lateral sites, during the infectious process.

S. Dramsi (✉)
Unité de Biologie des Bactéries Pathogènes à Gram-positif, Institut Pasteur, Paris 75015, France
e-mail: shaynoor.dramsi@pasteur.fr

S. Dramsi
CNRS ERL 3526, Paris 75015, France

H. Bierne
Micalis Institute, INRA, AgroParis Tech Université Paris-Saclay, Jouy-en-Josas 78350, France
e-mail: helene.bierne@jouy.inra.fr

Current Topics in Microbiology and Immunology (2017) 404:177–201
DOI 10.1007/82_2016_4

Published Online: 05 March 2016

Contents

1 Introduction

Gram-positive bacteria of the Firmicute phylum are responsible for the most common infections in humans [e.g. staphylococci, streptococci and enterococci (Willems et al. 2011)]. They also constitute a major component of the gut microbiota [e.g. clostridiales, lactobacilli and ruminococci (Jandhyala et al. 2015)], whose imbalance can be deleterious to health. These bacteria produce a wide repertoire of surface proteins that play key roles in the interaction with their host, such as adhesion to eukaryotic cells, binding to extracellular matrix, mucus or plasma proteins, transport of nutrients and evasion from the host innate immune defences. Surface proteins are thus long-standing targets for the development of antibacterial therapeutics.

Knowing the precise location of bacterial proteins is often critical to understand their function. The distinct subcellular distribution of proteins in Firmicutes has greatly benefited from pioneering studies that revealed the role of MreB (Jones et al. 2001) and FtsZ (Bi and Lutkenhaus 1991; Lowe and Amos 1998), the bacterial counterparts of eukaryotic actin and tubulin. These bacterial cytoskeletal elements determine PG sidewall or septal synthesis, which is required for cell elongation or division, respectively. Following initial discoveries made by immunoelectron and immunofluorescence microscopy on fixed cells, a major breakthrough came from the use of green fluorescent protein (GFP) as a fluorescent reporter that could retain the subcellular localization of the protein fused to it, opening the way to a dynamic view of protein localization in living bacteria (Phillips 2001). The development of new fluorescence proteins and the advance of single molecule and super-resolution imaging techniques over the last decade (Chiu and Leake 2011; Gahlmann and Moerner 2014) also radically changed the knowledge of bacterial cell organization that was for a long time limited due to the small size of bacteria. The asymmetric

distribution of ActA, the actin polymerization factor of *Listeria monocytogenes*, is one of the first and most striking evidence of the importance of precise subcellular distribution of a surface protein for its function (Smith et al. 1995; Rafelski and Theriot 2006). However, in Gram-positive bacteria, the fluorescence of GFP seems to be lost after its translocation outside of the plasma membrane (Yu and Gotz 2012), and a dynamic vision of proteins at the cell surface remains a challenge.

The majority of surface proteins are secreted via the general Sec secretory pathway and can be classified in five different categories according to their means of association with the cell surface: (i) membrane-spanning proteins; (ii) lipoproteins; (iii) cell wall-anchored proteins (CWA), which are covalently linked to the PG by sortases (detailed below); (iv) non-covalently associated proteins through interaction with cell wall components (e.g. PG, lipoteichoic and/or teichoic acids, other glycopolymers); and (v) non-conventional ("moonlighting") surface proteins present at the surface by unknown mechanisms. Several protein motifs constituting cell wall targeting domains have been characterized: the LysM motif, first identified as 44 amino acid repeats in *Bacillus subtilis* ϕ29 lysozyme, is found in many cell wall hydrolases responsible for bacterial cell separation. It was shown that the LysM domains bind to the PG disaccharide β-*N*-acetylmuramic acid (1 → 4)-β-*N*-acetylglucosamine but not to the PG decorated with teichoic acid (Frankel and Schneewind 2012). The WxL motif first identified in *E. faecalis* surface protein also binds to PG (Brinster et al. 2007); the SLH domain present in the S-layer proteins of *Bacillus anthracis* interacts with pyruvylated cell wall polysaccharide (Mesnage et al. 2000); the CBD domain first described in *Streptococcus pneumoniae* interacts with the choline residues decorating both lipoteichoic and teichoic acids, whereas the GW modules of *L. monocytogenes* Internalin B and Ami autolysin bind to lipoteichoic acids (Braun et al. 1997; Jonquieres et al. 1999; Bierne and Cossart 2007).

Sortase substrates (i.e. CWA proteins) of Gram-positive bacteria, such as LPxTG proteins of *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Listeria* and other Firmicutes, constitute an amazing repertoire of colonizing factors, toxins, proteases and enzymes that enable these opportunistic pathogens to mount successful infections (Navarre and Schneewind 1999; Lindahl et al. 2005; Bierne and Cossart 2007; Hendrickx et al. 2009; Nobbs et al. 2009; Speziale et al. 2009). Following the pioneering work of Schneewind and co-workers showing that LPxTG proteins are attached to the PG by an enzyme called "sortase", the mode of action of sortases has been intensively studied and a plethora of sortase genes has been identified in almost all Gram-positive bacteria, with often more than one sortase per genome (Comfort and Clubb 2004; Dramsi et al. 2005). Sortases are now classified into six classes named A–F: the housekeeping sortase A (SrtA), which anchors a wide range of LPxTG proteins, sortase B (SrtB) recognizes NP(Q/K)TN-like motifs in proteins involved in iron uptake systems, sortase C dedicated to pilus assembly and other less characterized sortases D, E and F (Comfort and Clubb 2004; Dramsi et al. 2005; Spirig et al. 2011; Bradshaw et al. 2015). The prototypical sortase A (SrtA) of *Staphylococcus aureus* is a membrane-bound transpeptidase, which recognizes a conserved carboxylic sorting motif referred as LPxTG motif where x represents any amino acid, and cleaves between the threonyl and the glycyl residues of this motif,

and links the carboxyl group of the threonine to the PG precursor lipid II (Perry et al. 2002). The protein–lipid II-linked product is then incorporated into the mature PG via the penicillin-binding proteins (PBPs) constituting the cell wall biosynthesis machinery. Thus, surface protein precursors are first processed by the secretory pathway via N-terminal signal peptides, then scanned by SrtA for a C-terminal sorting signal tagged with an LPxTG motif and finally anchored to the nascent PG (Marraffini et al. 2006; Schneewind and Missiakas 2012). This sequence of events summarized in Fig. 1 supports the idea that the final localization of CWA proteins in the cell wall depends on protein synthesis and secretion, cell wall biosynthetic and turnover activities, as well as sortase-mediated anchoring.

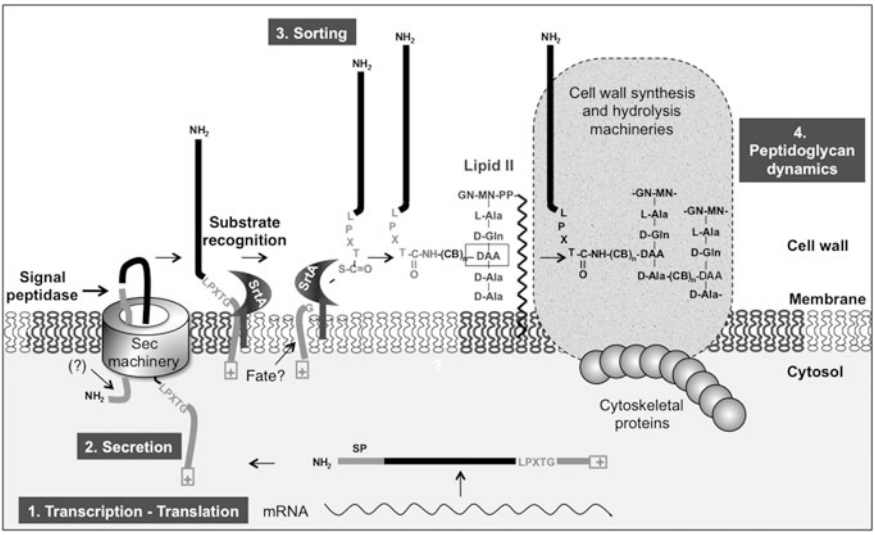


Fig. 1 Anchoring of LPxTG proteins by SrtA at the surface of Gram-positive bacteria. Following transcription and translation, polypeptide precursors bearing a N-terminal signal peptide (SP, *dark grey*) and a C-terminal sorting signal [in *light grey* LPxTG motif, hydrophobic domain and positively charged tail (boxed +)] are secreted across the membrane through the Sec pathway. Following cleavage of the signal peptide, the exported protein is transiently retained in the membrane and then processed by the membrane-bound transpeptidase sortase SrtA, which recognizes the LPxTG sequence and cleaves its substrate between the threonine and glycine residues of the motif. An acyl-enzyme intermediate is formed between the active site cysteine of SrtA and the carboxyl group of threonine. The enzyme then recognizes 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 (CB) results in covalent attachment of the protein to lipid II. The surface protein is next incorporated into the mature PG via the cell wall synthesis machinery (i.e. a multi-protein complex containing penicillin-binding proteins and other PG assembly factors, coupled to the cytoskeletal proteins). It then follows expansion and reshaping of the PG during cell growth and division, upon the action of hydrolytic enzymes (autolysins). MN, *N*-acetylmuramic acid; GN, *N*-acetylglucosamine; DAA (*D*-amino acid residue). Secretion and cell wall synthesis machineries might be associated with lipid domains [adapted from Bierne and Dramsi (2012), Schneewind and Missiakas (2012)]

The PG is the main constituent of the bacterial cell wall. It is a large polymer consisting of linear chains of two glycans (*N*-acetylglucosamine and *N*-acetylmuramic acid), which are further reticulated by linear and branched peptide links. Indeed, each molecule of *N*-acetylmuramic acid is attached to a specific short amino acid chain (Fig. 1). The other constituents of the cell wall in Gram-positive bacteria are extracellular polysaccharides, such as capsule and/or group-specific carbohydrate, teichoic acids and/or lipoteichoic acids. The bacterial surface is constantly remodelled during cell growth to achieve a particular shape and is subjected to deep changes upon switches between different physiological states, such as transition between actively dividing, biofilm-like or quiescent states. The distribution of sortase substrates, whose anchoring process is linked to PG metabolism, thus depends on specific characteristics of PG architecture and dynamics, which is different in spherical and rod-shaped bacteria.

Here, we will update our previous contribution on the dynamics of subcellular localization of cell wall-anchored proteins within the three-dimensional bacterial envelope (Bierne and Dramsi 2012). Our goal is not to be exhaustive but rather to focus on a few examples that are illustrative of general principles of protein localization in spherical, ovococcal and rod-shaped Firmicutes. These examples were chosen from our own model bacteria or from closely related organisms.

2 Localization of CWA Proteins in Gram-Positive Cocci: The Staphylococcal and Streptococcal Paradigms

2.1 Influence of Signal Peptides

A priori, there are several possible ways that protein localization could be specified. Each protein could be localized through a unique set of interactions with other localized proteins. This mechanism also known as "diffusion and capture" is observed for the assembly of multi-proteins complexes, e.g. division proteins associated with the FtsZ ring at the cell division plane forming the "divisome". Alternatively, localized proteins may contain sequences or structural information that directs them to particular sites via general sorting pathways, e.g. nuclear localization in eukaryotic cells by recognition of nuclear localization signal (NLS). One simple mechanism to direct protein localization in bacteria is through peptide signal sequences. The N-terminal signal sequences target proteins to the Sec or Tat (twin-arginine translocation) pathway for secretion or membrane insertion.

In 2006, an additional localization cue was discovered in the signal peptides of two proteins in *Streptococcus pyogenes* (Carlsson et al. 2006). In this pioneer study, Carlsson et al. demonstrated that the signal peptide sequence of M protein, which contains an YSIRK-G/S motif (SP_{+YSIRK}), directs the protein to the equatorial division site, while the signal peptide of protein F (PrtF, also known as SfbI) lacking this motif (SP_{-YSIRK}) directs PrtF to the old pole. These findings were

confirmed by an independent study, which also provided a dynamic 3D view of these proteins during bacterial cell cycle (Raz and Fischetti 2008). The SP_{+YSIRK} M protein was rapidly detected at the septum, simultaneously at the mother and daughter septa, whereas the SP_{-YSIRK} SfbI protein accumulated gradually on peripheral PG resulting in a polar distribution (Fig. 2). Localization of these two proteins (M and SfbI) was not altered in a SrtA mutant, suggesting that signal sequences direct localized secretion before anchoring to the cell wall (Raz et al. 2012). In contrast, alteration of PG synthesis using methicillin, a beta-lactam antibiotic inhibiting PG cross-linking, resulted in the significant reduction of the amount of M protein on the bacterial surface, but not of Sfbl. Thus, PG synthesis initiated at the septum during bacterial division is logically connected to the anchoring of M protein at the septum and may also control their expression/stability. It is important to note that the majority of proteins carrying the YSIRK sequence in their signal peptide are highly expressed proteins.

The YSIRK signal peptide localization rule was generalized in *S. aureus* (DeDent et al. 2008). Five SP_{+YSIRK} CWA proteins (protein A, ClfA, SdrC, SdrD and FnbpB) were shown to localize in a ring-like manner close to the equatorial division sites, whereas four SP_{-YSIRK} proteins (SasA, SasD, SasF and SasK) were found at discrete loci on the bacterial envelope that do not overlap with the division site. It is worth noting that mutations of the conserved YSIRK motif did not change protein localization indicating that the YSIRK per se is not the septum localization signal (DeDent et al. 2008). In an independent study, generation of a mCherry reporter system was developed in *S. aureus* to investigate the spatial distribution of surface proteins in live bacteria; Yu and Gotz (2012) showed that in contrast to GFP variants (e.g. GFM-mut3; super-folder GFP), mCherry can be secreted and anchored to staphylococcal cell wall while maintaining stable fluorescence. To our knowledge, this tool constitutes the first fluorescent probe for localizing CWA proteins in living bacteria (Yu and Gotz 2012). This study demonstrated that the signal peptides SP_{+YSIRK} and SP_{-YSIRK} are necessary and sufficient to drive the localization of CWA proteins either to the septum or pole, respectively (Yu and Gotz 2012).

As shown for protein M, treatment with sublethal concentrations of penicillin, a beta-lactam antibiotic, shifts the localization of the non-YSIRK mCherry hybrids from the peripheral cell wall to the division site. This shift is probably due to the increased amounts of lipid II precursors at the septum in these conditions. However, a reduction in the rate of protein synthesis and bacterial division may also contribute to this shift. From the results gathered in *S. pyogenes* and *S. aureus*, it appears that highly expressed proteins, such as M protein of *S. pyogenes* or protein A of *S. aureus*, are incorporated at the most active cell wall synthesis site (the septum), whereas lowly expressed proteins, for example PrtF of *S. pyogenes*, are incorporated elsewhere because of lack of free lipid II available at the septum.

The YSIRK signal peptide rule has also been investigated in *Streptococcus agalactiae*, a close relative of *S. pyogenes.* Localization of the secreted protein Bsp, whose signal peptide contains a canonical YSIRKG/S motif, has been studied by immunofluorescence (Brega et al. 2013). As predicted by the model, Bsp localized

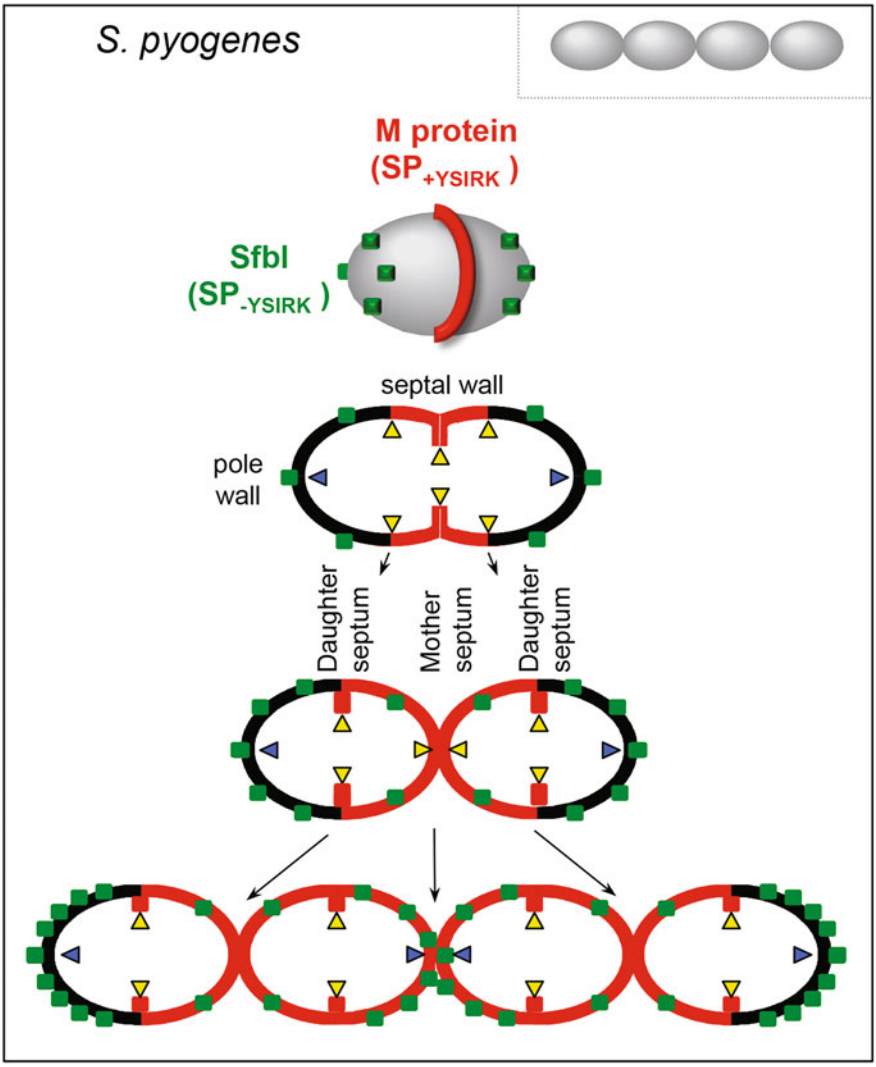


Fig. 2 A model representation of the spatial distribution of CWA proteins with SP_{+YSIRK} or SP_{-YSIRK} motifs at the surface of *S. pyogenes* (adapted from Carlsson et al. 2006; Raz and Fischetti 2008 and Bierne and Dramsi 2012). The surface of a bacterium is shown on the *top*. Spatial organization of daughter cells after successive division cycles is squared. 2-D longitudinal cross sections through the centre of the bacterium are shown in the *bottom* following protease removal of pre-existing surface proteins. SP_{+YSIRK} M protein (*in red*) rapidly localizes as a ring in the septal wall, while SP_{-YSIRK} SfbI (*in green*) localizes as patches in the polar regions. Active secretion and anchoring of the M protein (indicated by *yellow triangles*) are coupled to peptidoglycan synthesis at the septum. As daughter septum assembly starts before the mother septum is completely closed, anchoring of M protein occurs simultaneously at both mother and daughter septa. After two cell divisions, M protein covers areas of the surface with newly synthesized peptidoglycan. In contrast, as SfbI is preferentially anchored to peripheral peptidoglycan, it accumulates at sites of older peptidoglycan, particularly at poles (secretion and anchoring of SfbI are indicated by *blue triangles*)

at the septum. However, replacing the signal peptide of Bsp with four other signal peptides, containing or not the YSIRK motif, did not altered the localization of Bsp at the equatorial ring. Cell wall-anchored proteins displayed polar, punctate or uniform distribution on the surface of *S. agalactiae*. These results indicate that the YSIRK rule does not apply to all Gram-positive cocci (Brega et al. 2013).

The molecular bases, underlying the differential targeting of SP_{+YSIRK}/SP_{-YSIRK} proteins, remain largely unknown. As mentioned before, most proteins carrying the YSIRK sequence in their signal peptide are highly expressed proteins. Interestingly, it was observed that the cellular amounts of SP_{+YSIRK} mCherry hybrids are higher than those without the YSIRK motif, suggesting that signal peptides might affect expression at a transcriptional or post-transcriptional level. In addition, mutations of the YSIRK sequence significantly reduced the rate of signal peptide processing of protein A of *S. aureus* (Bae and Schneewind 2003). Thus, one cannot rule out the possibility that the differential targeting is due to different expression levels rather than to the intrinsic properties of signal peptides. In *S. aureus*, three genes (*spdA*, *spdB*, *spdC*) encoding transmembrane proteins with abortive infectivity (ABI) domains, elements first described in lactococci for their role in phage exclusion, were shown to be involved in the transcription of genes encoding YSIRK-containing proteins. Despite the reduced amount of protein A in these *spd* mutants, the spatial distribution of protein A was not significantly altered (Frankel et al. 2010). Finally, recent studies indicate biased codon usage within the signal peptide of *E. coli* proteins with a high frequency of non-optimal codons, which allow efficient folding and export (Zalucki et al. 2009). It is thus highly likely that a difference in the codon usage between YSIRK-containing signal peptides and non-YSIRK signal peptides may explain a better coupling between translation and secretion, resulting in higher surface levels of YSIRK-containing proteins.

2.2 *Influence of Secretion, Sorting and Cell Wall Synthesis Machineries*

Since LPxTG proteins are exported through the general Sec secretory pathway and anchored to the PG precursor lipid II by the sortase A, the distribution of secretion, sorting and cell wall synthesis machineries is likely to play an important role in the final localization of CWA proteins. So far, a clear rule for the localization of these different components in cocci does not emerge. A pioneer study in *S. pyogenes* reported that secretion of proteins occurs at a single microdomain in the membrane, through a macromolecular structure termed the ExPortal, enriched in anionic lipids that cluster Sec translocons and accessory factors (Rosch and Caparon 2004). More recently, the same group demonstrated that the integrity of the ExPortal is dependent on the cell wall biogenesis in *S. pyogenes* (Vega et al. 2013). However, two other independent studies challenged the ExPortal view and showed that SecA is randomly distributed in the membrane of *S. pyogenes* using either immunoelectron

microscopy (Carlsson et al. 2006) or deconvolution immunofluorescence microscopy (Raz and Fischetti 2008).

In *S. pneumoniae*, SecA and SecY proteins were found both at the septum and periphery but not at a focal domain (Tsui et al. 2011). Regarding the subcellular distribution of the housekeeping sortase, SrtA was distributed in an apparently random punctate pattern over the surface of cells without any preferential distribution to the septum or pole (Tsui et al. 2011). SrtA molecules of *S. pyogenes* were detected at the division septa and, to a lesser extent, at the poles where sortase foci tend to be smaller and less frequent than the septal ones (Raz and Fischetti 2008). In contrast, in *S. agalactiae,* both SecA and SrtA were shown to be colocalized at the septum (Brega et al. 2013). Similarly, SrtA has been shown to colocalize with SecA in *Enterococcus faecalis* (Kline et al. 2009) and *Streptococcus mutans* (Hu et al. 2008) at discrete foci using immunoelectron microscopy. More recently, immunofluorescence analyses indicate that SrtA and SecA are localized at nascent septal sites in *E. faecalis* but do not always colocalize through the bacterial cell cycle (Kandaswamy et al. 2013).

Direct time-lapse visualization of secretion machineries (secretons) and sortases is still limited by the constraints of conventional optical microscopy, which could explain the differences in the aforementioned studies. These technical issues also impair a detailed analysis of the incorporation of SrtA substrate–lipid II-linked products into the PG by the cell wall biosynthesis machinery, a multi-protein complex containing synthases (PBPs) and hydrolases (autolysins), coupled to cytoskeletal proteins (den Blaauwen et al. 2008; Typas et al. 2012; Wang et al. 2012). It has been proposed that in cocci and ovococci, the new cell wall is synthesized only at mid-cell, also referred to as the "equator" by opposition to the perpendicular or "meridian" pole, leading to spherical or ovoid shapes (Zapun et al. 2008; Perez-Nunez et al. 2011; Jiang et al. 2015). Ellipsoid-shaped bacteria, like streptococci, rely on the activity of two finely coordinated PG machineries, which are dedicated either to cell elongation or division (Massidda et al. 2013; Philippe et al. 2014). In contrast, spherical bacteria, like staphylococci, were thought to possess a unique septal PG synthesis system, generating new PG synthesis only at the septum and promoting cell division in orthogonal planes. However, the application of novel techniques to visualize *S. aureus* division cycle including super-resolution microscopy techniques and fluorescent D-alanine derivatives for labelling nascent PG recently challenged this dogma (Gautam et al. 2015; Zhou et al. 2015; Monteiro et al. 2015). These new studies indicate that staphylococci synthesize a peripheral PG before cell division (Fig. 3), and among the four staphylococcal PBPs, PBP4 was identified as being critical player for this process. According to this new model, *S. aureus* cells do not divide in equal planes. Although daughter cells appear spherical and perfectly similar, the bacterial surface is made of about ¾ of the old PG and ¼ of new PG (Zhou et al. 2015; Monteiro et al. 2015).

Of note, the class B sortase (SrtB) is thought to use non-cross-linked, mature PG as a substrate for IsdC attachment and this internal localization contributes to the function of IsdC as haem transfer from the cell wall to the Isd membrane transporter

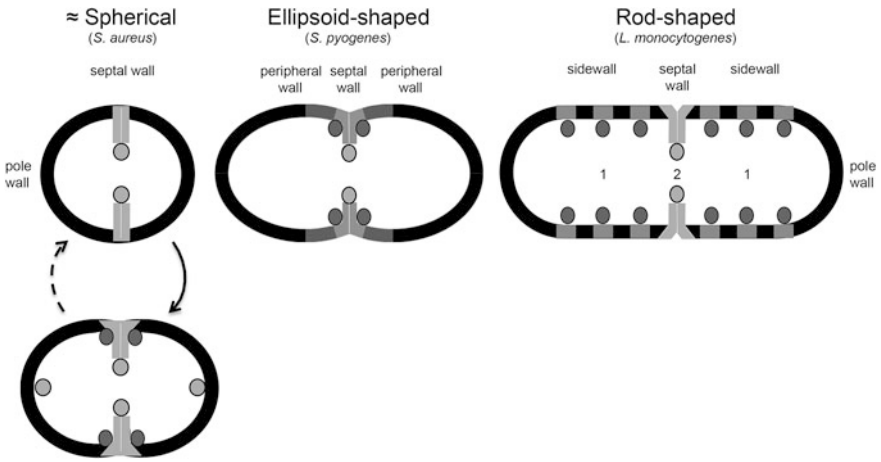


Fig. 3 Models of cell wall growth patterns in Gram-positive (Firmicutes) bacteria. Here are depicted simplified models for cell wall assembly and localization of PG factories in spherical cocci (e.g *S. aureus*), ellipsoid-shaped ovococci (e.g *S. pyogenes*) and rod-shaped bacteria (e.g *L. monocytogenes*). *Light and dark grey lines* show regions where new PG material is inserted, at peripheral or septal sites, respectively, and associated PG synthesis machineries are represented by ovals. Ovococci (ellipsoid) use two machineries localized at the mid-cell to synthesize the peripheral and septal wall, respectively. In *S. aureus* (spherical), new data suggest the existence of a peripheral machinery in addition to the primary septal machinery. Rod-shaped bacteria first have an elongation stage characterized by insertion of PG in the form of helices along the sidewall of the cell cylinder (*1*) and then synthesize septal wall to form two young poles (*2*). For simplicity, the diagram does not represent the dynamic localization of PG factories in time and space during cell division [Adapted from Bierne and Dramsi (2012)]

(Mazmanian et al. 2003). It would be interesting to investigate whether SrtA can use a different substrate than lipid II, including mature PG, to anchor its substrates at non-growing regions of the PG. Another important question to be solved concerns the "active" or "passive" localization of sortase A at poles.

We cannot close this section without mentioning the sortase-assembled polymers called pili, discovered a decade ago in most streptococci and enterococci (Telford et al. 2006). These long surface appendages have been involved in adherence, biofilm formation, modulation of the immune responses and virulence (Danne and Dramsi 2012). Pili are generally composed of two or three different LPXTG proteins, which are covalently assembled into a pilus fibre by one to three sortase C enzymes (Kang and Baker 2012). In *S. pneumoniae*, a discrete non-homogenous distribution of pilus subunits was observed near the division site (Reis et al. 2010). In *E. faecalis*, the sortase C enzyme responsible for Ebp pilus assembly co-localized with SecA and SrtA (Kline et al. 2009). Similarly, in *S. agalactiae*, pili are found at mother and daughter division sites and following cell separation at poles where they accumulate due to their high stability (SD, unpublished data). Collectively, these data argue for the existence of a multi-protein complex responsible for the secretion

and anchoring of sortase substrates. Interestingly, human cationic β-defensins were shown to interact with *E. faecalis* at septal sites and disrupt sites of localized secretion and sorting (Kandaswamy et al. 2013).

3 Dynamics of CWA Protein Localization in the Rod-Shaped Bacterium *Listeria monocytogenes*

3.1 The "Cylinder" Challenge

The subcellular distribution of surface proteins cannot be dissociated from bacterial morphogenesis (Fig. 3). To elongate with a constant diameter, several rod-like bacterial species, including the Gram-positive *B. subtilis* and Gram-negative *Escherichia coli*, use lateral sidewall PG synthesis to elongate the bacterial body prior to septal wall synthesis (Zapun et al. 2008; Margolin 2009; Jiang et al. 2015). In the past ten years, many reports have investigated this mechanism (Amir and van Teeffelen 2014), and even if differences exist, in line with the issue of building a flat (*E. coli*) or a thick (*B. subtilis*) cell wall, a common scenario can be defined. The biosynthesis of the sidewall PG in these organisms is achieved by the dispersed incorporation of PG precursors at discrete sites of the cylindrical bacterial body. This depends on an elongation multi-protein complex, named "elongase", containing synthases (PBPs) and hydrolases (autolysins) coupled to actin-like MreB proteins (den Blaauwen et al. 2008; Typas et al. 2012; Wang et al. 2012). Visualization of MreB by TIRF microscopy revealed that in Gram-positive (Dominguez-Escobar et al. 2011; Garner et al. 2011; Olshausen et al. 2013; Reimold et al. 2013) as in Gram-negative bacteria (van Teeffelen et al. 2011), MreB rotates in a processive manner along tracks organized as rings, perpendicular to the long axis of the cells. MreB filaments are proposed to act as a platform for PG-assembling factories and/or to control the direction and speeds of these factories, enabling insertion of new PG strands in helicoidal bands. Studies performed in *E. coli* suggested that MreB trajectories are indeed slightly helical (van Teeffelen et al. 2011). This model of a helical orientation of the PG meshwork caused by the orientation of MreB filaments was further supported by combining microscopy and optical trapping to follow fluorescent beads at the bacterial surface (Wang et al. 2012). The curvature and twist of the MreB filaments, but also the curved surface of the cylindrical cell envelope, might cause this orientation. The architecture of the mature PG in *B. subtilis* analysed by electron cryotomography and surface atomic force microscopy (AFM) experiments also suggests the existence of a higher-order three-dimensional helicoidal structure (Hayhurst et al. 2008; Andre et al. 2010; Beeby et al. 2013). In this context, CWA proteins could be nice tools to further explore the PG dynamics in Gram-positive bacteria.

3.2 *Helix-Like Localization Patterns of CWA Proteins in* Listeria monocytogenes

With an exceptional high number of surface proteins, the rod-shaped pathogen *L. monocytogenes* constitutes a good model to study CWA proteins (Bierne and Cossart 2007). We identified its sortase A and sortase B more than ten years ago (Bierne et al. 2002; Bierne et al. 2004). In silico genome analysis, microscopy and biochemical analyses further revealed about forty SrtA substrates and two SrtB substrates (Calvo et al. 2005; Pucciarelli et al. 2005; Bierne and Cossart 2007). The *L. monocytogenes* LPxTG protein family encompasses many proteins involved in the pathogenic process, such as internalins InlA, InlH, InlJ and InlK (Gaillard et al. 1991; Sabet et al. 2008; Personnic et al. 2010; Dortet et al. 2012), the virulence protein VIP (Cabanes et al. 2005), the adhesin LapB (Reis et al. 2010) and the mucin-binding protein LmiA (Mariscotti et al. 2014). The haem-binding proteins Hbp1 (formely Lmo2186) and Hbp2 (formely Lmo2185 or SvpA) are the SrtB substrates, which are involved in iron metabolism (Bierne and Cossart 2002; Klebba et al. 2012). By using 3D imaging, we studied spatial organization of a set of these CWA proteins, namely InlA, InlH, InlJ and Hbp2/SvpA. We found that at steady-state expression levels, all of them were arranged along non-overlapping helices in the sidewall (Bruck et al. 2011) (Fig. 4a). Such distribution is in agreement with the helicoidal architecture of the PG in the related species *B. subtilis*. Based on these observations, we propose that listerial LPxTG proteins may be sequentially incorporated in a same PG helical band and/or incorporated in different bands (Fig. 4b).

The listerial LPxTG proteins decorate the cylindrical sidewall as helices in close proximity of the nascent cell wall synthesis sites labelled with fluorescent vancomycin (Van-FL) (Bruck et al. 2011). This organization is compatible with the helical cell wall incorporation model described in *B. subtilis* (Daniel and Errington 2003; Tiyanont et al. 2006). Since LPxTG proteins are proposed to be anchored to the cell wall precursor lipid II, prior to being incorporated into the mature PG (Perry et al. 2002), one can expect a spatial proximity of the secretion, sortase and PG factories, as discussed above for cocci. Remarkably, Sec translocons exhibit helix-like patterns in *B. subtilis* and *E. coli* (Campo et al. 2004; Shiomi et al. 2006), but there are no data on their mobility in the bacterial membrane. Of interest, two putative sortases (YhcS and YwpE) and their substrates (YhcR and YfkN) have been identified by in silico analyses in *B. subtilis* and a recent study confirmed YhcS and YhcR as sortase and a sortase substrate, respectively (Liew et al. 2012). In agreement with the model in *Listeria*, microscopy assessment of a GFP–sortase fusion indicated that the *B. subtilis* sortase localized in helical arcs or tracks (Liew et al. 2012). Application of TIRF microscopy to Sec and sortase–GFP fusions in both species would be extremely meaningful to get a more dynamic view of the coupling between secretion and sorting.

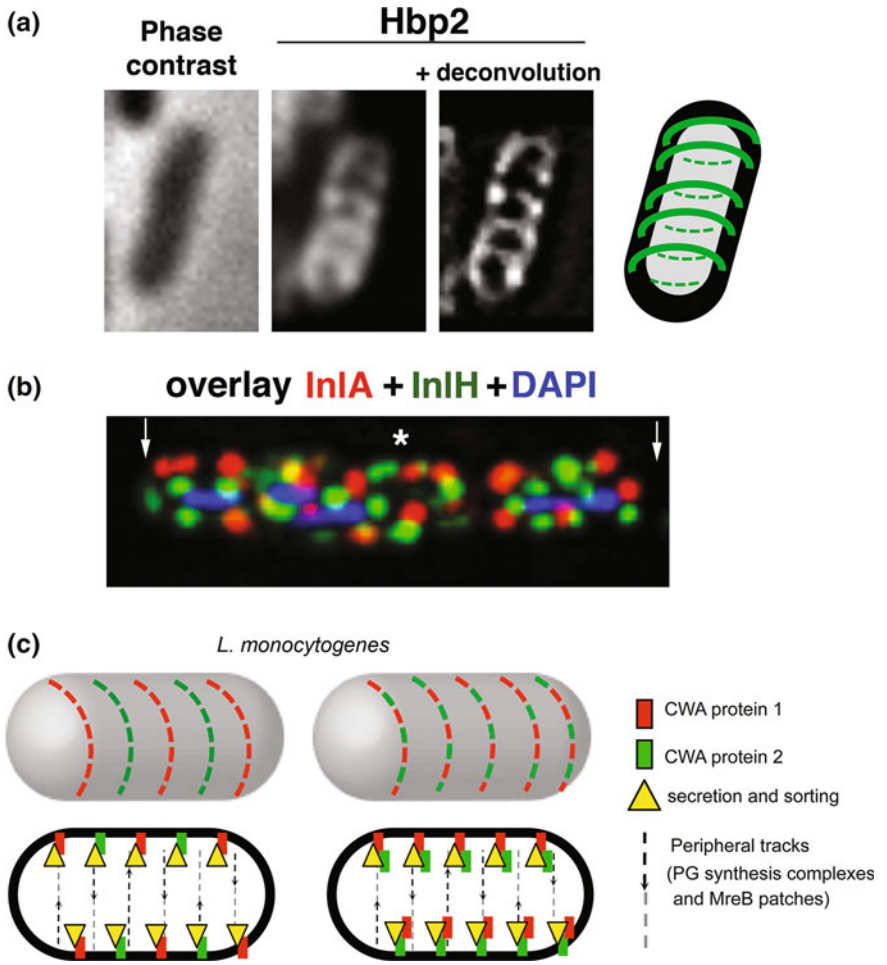


Fig. 4 Helical distribution of sortase substrates at the surface of *L. monocytogenes* during the exponential growth phase (A–B) Immunofluorescence micrographs of *L. monocytogenes* grown at mid-log phase in standard BHI medium. **a** Helix-like patterns of the sortase B substrate Hbp2/SvpA in the sidewall. The staining of the bacterial surface with Hbp2 antibodies highlights helical patterns, as illustrated in the diagram (from Bruck et al. 2011). **b** Bacteria were labelled with InlA and InlH antibodies and DAPI to label the nucleoid. The septum (*asterisk*) and poles (*arrow*) are also indicated. Z-stacks of deconvoluted images were merged. InlA (*red dots*) and InlH (*green dots*) have mostly non-overlapping distribution. **c** Models of CWA protein anchoring in exponentially growing *L. monocytogenes*. Two model LPxTG proteins (represented as *dotted red* and *green lines*) are distributed as helix-like patterns at the bacterial surface. CWA proteins may be secreted and anchored at different sites and arranged as distinct helices (*left*) or sequentially secreted and anchored at a same site and part of the same helix (*right*). A 2-D longitudinal cross section through the centre of the bacterium is shown *below*. After secretion and sortase-mediated anchoring to the PG (indicated by *yellow triangles*), CWA proteins follow the helical expansion of the murein sacculus orchestrated by the PG elongation complex in association with MreB cytoskeletal filaments [Adapted from Bierne and Dramsi (2012)]

3.3 *Shifts to Polar Localization of* L. monocytogenes *CWA Proteins in Response to Environmental Changes*

PG is also synthesized at the cell division site, creating the septal wall required for the formation of new poles. As for elongation, a specialized multi-component machinery known as the divisome, containing PG synthases and hydrolases coupled to the tubulin-like protein FtsZ, is involved in this complex and essential process (den Blaauwen et al. 2008; Typas et al. 2012). As previously shown in *B. subtilis* (Tiyanont et al. 2006), the PG precursors of *L. monocytogenes*, labelled with Van-FL, markedly stained the septal wall but were also detected as dots on the old poles (Bruck et al. 2011). This observation supports the idea that PG synthesis may continue for some period at the youngest poles following cell division (Mobley et al. 1984). With such a distribution of lipid II, one would expect sortase substrates to localize to septal and polar cell walls, in addition to sidewall in *Listeria* and other rod-shaped bacteria. However, these sites seem to be used only in specific conditions, in response to environmental changes and/or altered division rates, enabling fluidity in the distribution of surface proteins in time and space.

Genome-wide transcriptomic studies (Camejo et al. 2009; Toledo-Arana et al. 2009) have shown that CWA protein-encoding genes are differentially expressed upon transition from life in the environment to colonization of host. High-resolution mass spectrometry applied to purified cell walls confirmed that the repertoire of CWA proteins in *L. monocytogenes* evolves qualitatively and quantitatively in the envelope, with the different phase of growth, in different media (nutrient-rich to minimal medium), and with the entry of the bacterium into the cytosol of infected epithelial cells (Pucciarelli et al. 2005; Garcia-del Portillo et al. 2011). As discussed above for cocci, it seems that variations in expression levels also affect spatial organization of CWA proteins of *L. monocytogenes*. In particular, there is a striking enrichment of SrtA substrates, InlA and InlJ, at both poles of exponentially growing bacteria, in response to specific switches in gene expression. InlA localization to the poles was proposed to involve passive helical cell wall growth (Rafelski and Theriot 2006), but our data rather support an active process of secretion and anchoring at the pole, when a critical protein concentration is achieved (Bruck et al. 2011), as depicted in the model Fig. 5a. Thus, under high expression levels induced by the transcription regulator PrfA, helical anchoring of InlA might be saturated allowing polar positions to become utilized. Reaching a critical protein concentration would trigger accumulation of LPxTG proteins at the septum and, following cell division, further secretion and anchoring by SrtA to lipid II at newly formed pole. As the synthesis of PG continues at the pole for a while, proteins progressively accumulate at the ageing pole until it becomes metabolically inert. In a similar way, the *E. coli* enzyme MurG localizes at the poles when synthesized at high levels (Michaelis and Gitai 2010).

Another switch in localization patterns is observed with σ^B-dependent LPxTG proteins, as exemplified with InlA and InlH. When bacteria enter into the stationary phase or are submitted to oxidative stress in exponential phase, these proteins

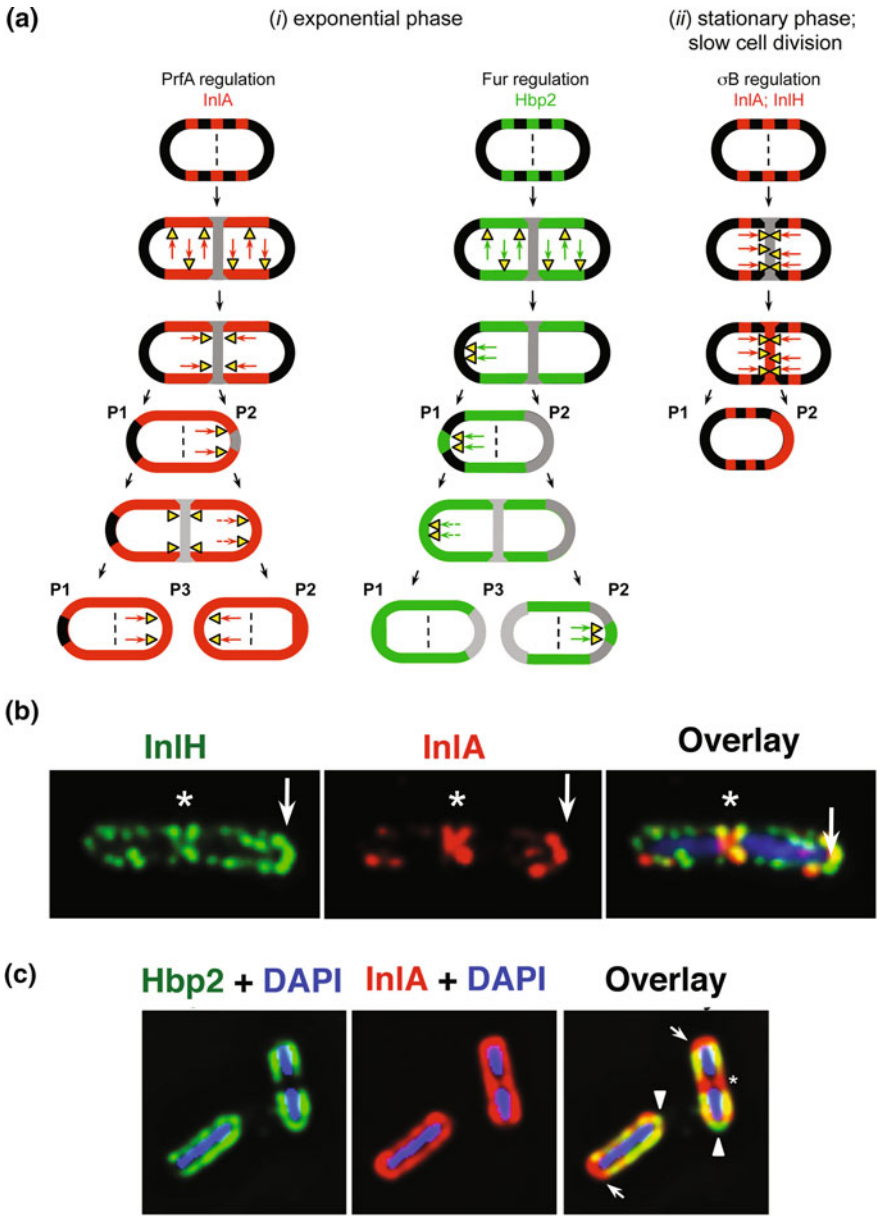
shuttle from the sidewall to septal wall (Bruck et al. 2011). Anchoring to the septal PG precursors enables these proteins to be rapidly exposed at the surface of cells exposed to a sudden stress, when cell elongation is impaired (Fig. 5a, b). Conversely, heat shock, which disrupts septum formation and induces cell elongation, prevents σ^B-dependent LPxTG proteins to be localized at the mid-cell. These dynamic changes suggest interconnected regulatory circuits that coordinate cell wall dynamics, cell wall anchoring and σ^B-dependent stress responses.

What determines the relocalization of CWA proteins to the septal region is unknown. There is so far no data supporting a role for spatial information encoded in the signal peptides in *Listeria* or other rod-shaped bacteria, as described for cocci above. Polar localization information may be carried by messenger RNAs (Blaylock et al. 2008; Nevo-Dinur et al. 2011). Alternatively, σ^B, InlA, InlH and/or SrtA may shuttle from one place to another by interacting with components of elongasome or divisome complexes, or with factors that sense membrane curvature and play role in the secretion at septal regions, such as the membrane protein DivIV (Halbedel et al. 2012). σ^B-dependent PG proteins may also be more resistant to proteolysis at the poles. These are important issues that need to be addressed in future studies.

The subcellular localization of the SrtB substrate, Hbp2/SvpA, whose production is repressed by Fur and induced by iron limitation, contrasts with that of classical LPxTG proteins such as InlA (Bruck et al. 2011). When HpB2 expression is activated, it localizes to the old pole, in addition to the sidewall, with a complete exclusion from the septal region (Bruck et al. 2011). SrtB has been proposed to anchor its substrate directly to the mature PG, suggesting that SrtA and SrtB substrates localize to distinct sites of the bacterial surface (Marraffini and Schneewind 2005). Consistent with this hypothesis, InlA and Hbp2 are deposited to distinct sites of the surface (Bruck et al. 2011) (Fig. 5a, c). This differential distribution of surface proteins displaying unrelated functions may favour the separation of functional activities: InlA-dependent entry into epithelial cells versus Hpb2-mediated iron transport pathway (Klebba et al. 2012).

3.4 Importance of Other Domains Than the C-terminal Sortase-Anchoring Motif

Listerial CWA proteins have a modular architecture with distinct domains (or regions) upstream of the C-terminal sorting signal, such as leucine-rich repeats (LRR), PKD repeats, mucin-binding (MucBP), BIG-3 and collagen-binding domains. Some of these domains are involved in interaction with host factors, the best example being the N-terminal LRR domain found in CWA proteins of the internalin family (Bierne et al. 2007). The LRR domain of InlA, the paradigm of this family, binds to the mammalian adhesion molecule E-cadherin (Ecad) (Mengaud et al. 1996). InlA–Ecad interactions are sufficient for bacterial uptake



into epithelial cells, a key step for the crossing of intestinal, placental and blood–brain barriers by *L. monocytogenes* (Bonazzi et al. 2009; Disson and Lecuit 2013). It is worth mentioning that while the eukaryotic signalling cascades promoting

◂**Fig. 5** Changes from helical to polar distribution of *L. monocytogenes* CWA proteins in response to environmental changes. **a** Models of spatial and temporal localization of CWA virulence proteins in the rod-shape bacterium *L. monocytogenes*. (i) During the exponential phase, CWA proteins localize as helices in the cylindrical part of the cell (*rectangles represent* helical foci). In response to environmental changes and increased expression levels, they may reach a critical concentration that saturates helical positions (*vertical arrows*). Following activation of its regulator PrfA, the SrtA substrate InlA (*in red*) accumulates in the septal region (*horizontal arrows*) and is anchored to the new pole wall (P2) only after cell division is achieved. At the next cell division, InlA is anchored to the younger pole P3, while still incorporated at P2 until this pole becomes inert, leading to distribution of InlA at the two poles. When overexpressed under iron-limited conditions, the Fur-regulated SrtB substrate Hbp2 does not accumulate at the mid-cell but instead is anchored to the mature PG at the old pole (P1). Hbp2 accumulates at P1 and starts to be anchored at P2 at the second generation, while remaining excluded from the newly synthesized PG at P3, leading to an asymmetric distribution at the surface. (ii) σ^B-dependent CWA proteins, such as InlA or InlH, are actively produced in response to stress. Upon entry into the stationary phase and inhibition of cell elongation, InlA and InlH become anchored to the septal cell wall precursors before cell division is achieved, thus decorate the septal wall and are fastly exposed at the newly formed pole (P2). **b–c** Immunofluorescence micrographs illustrating the relocalization of CWA proteins at the surface of *L. monocytogenes* (adapted from Bierne and Dramsi 2012). **b** InlH and InlA relocalize to overlapping sites at septa and poles of *L. monocytogenes* (strain EGDe) entering into the stationary phase. An immunofluorescence micrograph of a *Listeria* labelled with InlA and InlH antibodies and DAPI (*blue*, in the overlay) is shown. InlA colocalizes with InlH at the septum (*asterisk*) and pole (*arrow*). **c** When induced for expression, InlA and Hbp2 relocalize to distinct sites at poles. Bacteria grown with charcoal (to activate InlA expression) and dipyridyl (to activate Hbp2 expression) were labelled with InlA and Hbp2 antibodies and DAPI. *Arrows* and *arrowheads* point to non-overlapping InlA and Hbp2 signals at poles and *asterisk* marks the septum [From Bruck et al. (2011)]

entry are now quite well known, there is no data addressing the role of the topological organization of InlA molecules at the bacterial surface in the receptor-guided mechanism of phagocytic cup assembly.

Recent studies on Lmo0171, another member of the internal family, open the possibility that some LPXTG proteins can interact directly with PG assembly and/or turnover machineries. Indeed, disruption of the *lmo0171* gene results in notable changes in the bacterial cell morphology, with asymmetry in cell wall growth and curvature of the bacterial body. This shape defect coincides with a decrease in the ability of bacteria to invade human epithelial cells (Stachowiak et al. 2015). How Lmo0171 is involved in the regulation and/or stabilization of the cell wall structure is unknown.

4 Importance of Sortase Substrates Localization in the Infectious Process

The covalent anchoring of LPXTG virulence factors to the PG appears to play a key role in their pathogenic properties. This has been demonstrated for InlA, for which truncation of the cell wall sorting motif abolishes the entry of *L. monocytogenes*

into human intestinal cells (Lebrun et al. 1996; Jonquieres et al. 1998). In addition, replacement of the C-terminal sorting motif of InlA with the transmembrane domain of ActA results in decreased surface accessibility of InlA and consequently decreased invasiveness (Lebrun et al. 1996). Similarly, it was previously shown deletion of the LPETGE sequence of protein A, a major *S. aureus* virulence protein, abolished its cell wall anchoring and thus its immunoglobulin binding function (Stachowiak et al. 2015). Finally, proteomic studies have highlighted an important remodelling of the *L. monocytogenes* cell wall inside eukaryotic cells changing the repertoire of surface proteins. For example, high amounts of internalins InlA and InlB have been detected, suggesting that these proteins may interact not only with host cell surface receptors, but also with other host intracellular molecules (Garcia-del Portillo et al. 2011).

If surface display is critical, do spiral, ring-like, focal or polar organizations have any consequence on the interaction of PG-bound proteins with the host molecules? As discussed above, information is lacking on this subject. Yet, such repartitions might potentiate interaction with host ligands and compartmentalize specific functions. For example, spatial exclusion of listerial InlA and Hbp2 likely favours separation of unrelated functions such as entry into eukaryotic cells and iron transport, respectively. Conversely, spatial proximity should increase physical interactions required for common functions, as exemplified by staphylococcal haem iron transport components IsdA and IsdB (Pishchany et al. 2009). A recent study showed that IsdP, a novel iron-regulated autolysin, remodels the cell wall to facilitate surface localization of IsdC, another srtB substrate acting with IsdA and IsdB. In *Staphylococcus lugdunensis*, IsdP inactivation alters IsdC surface display, which in turn leads to a defect in haemoglobin utilization (Farrand et al. 2015). To our knowledge, this is the first study demonstrating the role of a specialized autolysin in the remodelling of the cell wall for the correct surface display of a cell wall-anchored protein, which in turn is key for protein function.

5 Conclusions

The toolbox for bacterial cell biology has expanded considerably within recent years (for a review, (Yao and Carballido-Lopez 2014)) with the development of optimized fluorescent proteins for Gram-positive bacteria. In addition, recent advances in super-resolution microscopy techniques applied to prokaryotic cells have provided insights into the process of bacterial wall elongation and division (Fleurie et al. 2014; Zhou et al. 2015). Fluorescent probes based on D-Ala, a key component of the stem peptide of PG (Fig. 1), are tools that will certainly provide many novel information of the cell wall dynamics in complex bacterial communities. For instance, such probes recently revealed the diversity of PG assembly in

the hundreds of bacteria found in a salivary sample (Kuru et al. 2015). Importantly, not all bacteria divide in a binary fission mode described for *E. coli*, *B. subtilis* or *S. aureus*. For example, *Corynebacterium glutamicum*, a Gram-positive bacterium belonging to the group of Actinobacteria, grows bipolar and divides from the tip (or poles). How cell wall-anchored proteins are organized in these bacteria may reveal novel connections between cell division and protein localization. The composition and/or structure of the PG can be heterogeneous along the bacterial surface and may constitute an important localization clue: indeed, it was recently shown that MapZ, which binds to PG, has the dual role of marking the cell division site and positioning the FtsZ ring in *S. pneumoniae* (Fleurie et al. 2014). It is proposed that MapZ recognizes a PG structure specific to mid-cell such as the equatorial mark, which is reminiscent of the "piecrust" structure previously reported in *S. pneumoniae*, *E. faecalis* and *S. aureus* (Tomasz et al. 1964; Higgins and Shockman 1970; Wheeler et al. 2011).

Another important concern is the experimental conditions in which CWA and PG dynamics are analysed. So far, most of the studies have been performed on model bacteria growing in standard broth medium. However, the bacterial surface profoundly changes in different physiological states, and one can predict that protein secretion and anchoring machineries would be deeply influenced by environmental and developmental conditions. A recent study showed that in *B. subtilis* the cytoskeletal protein MreB is affected by the development of competence during stationary phase. The competence regulator ComK activates the expression of *mreB* and MreB is targeted to the pole, where it colocalizes with the late competence protein ComGA. It is proposed that ComGA sequesters MreB to prevent cell elongation and escape from competence (Mirouze et al. 2015). This adaptation to competence might prove to be true for other physiological processes and is likely to affect the spatio-temporal organization of cell wall proteins. More generally, it will be important to examine the dynamics of surface proteins during biofilm development or bacterial interaction with the host environment, as well as in complex ecosystems harbouring thousand of different bacteria.

Many questions remain unsolved on the underlying mechanism driving protein subcellular localization at the bacterial surface. Protein targeting via specific mRNA localization may exist as shown in *E. coli* (Govindarajan et al. 2012). We anticipate that the fluidity of the cell surface proteins within the cell wall is tightly controlled and this has functional implications to the pathogenic or symbiotic property of prokaryotes. How bacterial replication, transcription and translation processes are organized and coordinated to quickly adapt to environmental signals or developmental changes in the prokaryotic cell represents the biggest challenge for the future.

Acknowledgments We thank Nienke Buddelmeijer for critical reading of the manuscript. H. Bierne acknowledges supports from the French National Research Agency (ANR 11-BSV3-0003 EPILIS) and the iXcore Foundation for Research.

References

Amir A, van Teeffelen S (2014) Getting into shape: How do rod-like bacteria control their geometry? Syst Synth Biol 8(3):227–235. doi:10.1007/s11693-014-9143-9

Andre G, Kulakauskas S, Chapot-Chartier MP, Navet B, Deghorain M, Bernard E, Hols P, Dufrene YF (2010) Imaging the nanoscale organization of peptidoglycan in living *Lactococcus lactis* cells. Nat Commun 1:27. doi:10.1038/ncomms1027

Bae T, Schneewind O (2003) The YSIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing. J Bacteriol 185(9):2910–2919

Beeby M, Gumbart JC, Roux B, Jensen GJ (2013) Architecture and assembly of the Gram-positive cell wall. Mol Microbiol 88(4):664–672. doi:10.1111/mmi.12203

Bi EF, Lutkenhaus J (1991) FtsZ ring structure associated with division in *Escherichia coli*. Nature 354(6349):161–164. doi:10.1038/354161a0

Bierne H, Cossart P (2002) InlB, a surface protein of *Listeria monocytogenes* that behaves as an invasin and a growth factor. J Cell Sci 115(Pt 17):3357–3367

Bierne H, Cossart P (2007) *Listeria monocytogenes* surface proteins: from genome predictions to function. Microbiol Mol Biol Rev 71(2):377–397

Bierne H, Dramsi S (2012) Spatial positioning of cell wall-anchored virulence factors in Gram-positive bacteria. Curr Opin Microbiol 15(6):715–723. doi:10.1016/j.mib.2012.10.010

Bierne H, Garandeau C, Pucciarelli MG, Sabet C, Newton S, Garcia-del Portillo F, Cossart P, Charbit A (2004) Sortase B, a new class of sortase in *Listeria monocytogenes*. J Bacteriol 186 (7):1972–1982

Bierne H, Mazmanian SK, Trost M, Pucciarelli MG, Liu G, Dehoux P, Jansch L, Garcia-del Portillo F, Schneewind O, Cossart P, European Listeria Genome C (2002) Inactivation of the srtA gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. Mol Microbiol 43(4):869–881

Bierne H, Sabet C, Personnic N, Cossart P (2007) Internalins: a complex family of leucine-rich repeat-containing proteins in *Listeria monocytogenes*. Microbes and infection/Institut Pasteur 9 (10):1156–1166. doi:10.1016/j.micinf.2007.05.003

Blaylock B, Sorg JA, Schneewind O (2008) *Yersinia enterocolitica* type III secretion of YopR requires a structure in its mRNA. Mol Microbiol 70(5):1210–1222. doi:10.1111/j.1365-2958.2008.06474.x (MMI6474 [pii])

Bonazzi M, Lecuit M, Cossart P (2009) *Listeria monocytogenes* internalin and E-cadherin: from structure to pathogenesis. Cell Microbiol 11(5):693–702. doi:10.1111/j.1462-5822.2009.01293.x

Bradshaw WJ, Davies AH, Chambers CJ, Roberts AK, Shone CC, Acharya KR (2015) Molecular features of the sortase enzyme family. FEBS J 282(11):2097–2114. doi:10.1111/febs.13288

Braun L, Dramsi S, Dehoux P, Bierne H, Lindahl G, Cossart P (1997) InlB: an invasion protein of *Listeria monocytogenes* with a novel type of surface association. Mol Microbiol 25(2):285–294

Brega S, Caliot E, Trieu-Cuot P, Dramsi S (2013) SecA localization and SecA-dependent secretion occurs at new division septa in group B Streptococcus. PLoS ONE 8(6):e65832. doi:10.1371/journal.pone.0065832

Brinster S, Furlan S, Serror P (2007) C-terminal WxL domain mediates cell wall binding in Enterococcus faecalis and other gram-positive bacteria. J Bacteriol 189:1244–1253. doi:10.1128/jb.00773-06

Bruck S, Personnic N, Prevost MC, Cossart P, Bierne H (2011) Regulated shift from helical to polar localization of *Listeria monocytogenes* cell wall-anchored proteins. J Bacteriol 193 (17):4425–4437. doi:10.1128/JB.01154-10

Cabanes D, Sousa S, Cebria A, Lecuit M, Garcia-del Portillo F, Cossart P (2005) Gp96 is a receptor for a novel *Listeria monocytogenes* virulence factor, Vip, a surface protein. EMBO J 24(15):2827–2838. doi:10.1038/sj.emboj.7600750

Calvo E, Pucciarelli MG, Bierne H, Cossart P, Albar JP, Garcia-Del Portillo F (2005) Analysis of the Listeria cell wall proteome by two-dimensional nanoliquid chromatography coupled to mass spectrometry. Proteomics 5(2):433–443. doi:10.1002/pmic.200400936

Camejo A, Buchrieser C, Couve E, Carvalho F, Reis O, Ferreira P, Sousa S, Cossart P, Cabanes D (2009) In vivo transcriptional profiling of *Listeria monocytogenes* and mutagenesis identify new virulence factors involved in infection. PLoS Pathog 5(5):e1000449

Campo N, Tjalsma H, Buist G, Stepniak D, Meijer M, Veenhuis M, Westermann M, Muller JP, Bron S, Kok J, Kuipers OP, Jongbloed JD (2004) Subcellular sites for bacterial protein export. Mol Microbiol 53(6):1583–1599

Carlsson F, Stalhammar-Carlemalm M, Flardh K, Sandin C, Carlemalm E, Lindahl G (2006) Signal sequence directs localized secretion of bacterial surface proteins. Nature 442 (7105):943–946

Chiu SW, Leake MC (2011) Functioning nanomachines seen in real-time in living bacteria using single-molecule and super-resolution fluorescence imaging. Int J Mol Sci 12(4):2518–2542. doi:10.3390/ijms12042518

Comfort D, Clubb RT (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. Infect Immun 72(5):2710–2722

Daniel RA, Errington J (2003) Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell. Cell 113(6):767–776

Danne C, Dramsi S (2012) Pili of gram-positive bacteria: roles in host colonization. Res Microbiol 163(9–10):645–658. doi:10.1016/j.resmic.2012.10.012

DeDent A, Bae T, Missiakas DM, Schneewind O (2008) Signal peptides direct surface proteins to two distinct envelope locations of *Staphylococcus aureus*. EMBO J 27(20):2656–2668

den Blaauwen T, de Pedro MA, Nguyen-Disteche M, Ayala JA (2008) Morphogenesis of rod-shaped sacculi. FEMS Microbiol Rev 32(2):321–344

Disson O, Lecuit M (2013) In vitro and in vivo models to study human listeriosis: mind the gap. Microbes and Infection 15(14–15):971–980. doi:10.1016/j.micinf.2013.09.012

Dominguez-Escobar J, Chastanet A, Crevenna AH, Fromion V, Wedlich-Soldner R, Carballido-Lopez R (2011) Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria. Science 333(6039):225–228. doi:10.1126/science.1203466

Dortet L, Mostowy S, Cossart P (2012) *Listeria* and autophagy escape: involvement of InlK, an internalin-like protein. Autophagy 8(1):132–134. doi:10.4161/auto.8.1.18218

Dramsi S, Trieu-Cuot P, Bierne H (2005) Sorting sortases: a nomenclature proposal for the various sortases of Gram-positive bacteria. Res Microbiol 156(3):289–297

Farrand AJ, Haley KP, Lareau NM, Heilbronner S, McLean JA, Foster T, Skaar EP (2015) An iron-regulated autolysin remodels the cell wall to facilitate heme acquisition in *Staphylococcus lugdunensis*. Infect Immun 83(9):3578–3589. doi:10.1128/IAI.00397-15

Fleurie A, Lesterlin C, Manuse S, Zhao C, Cluzel C, Lavergne JP, Franz-Wachtel M, Macek B, Combet C, Kuru E, VanNieuwenhze MS, Brun YV, Sherratt D, Grangeasse C (2014) MapZ marks the division sites and positions FtsZ rings in *Streptococcus pneumoniae*. Nature 516 (7530):259–262. doi:10.1038/nature13966

Frankel MB, Schneewind O (2012) Determinants of murein hydrolase targeting to cross-wall of *Staphylococcus aureus* peptidoglycan. J Biol Chem 287(13):10460–10471. doi:10.1074/jbc.M111.336404

Frankel MB, Wojcik BM, DeDent AC, Missiakas DM, Schneewind O (2010) ABI domain-containing proteins contribute to surface protein display and cell division in *Staphylococcus aureus*. Mol Microbiol 78(1):238–252. doi:10.1111/j.1365-2958.2010.07334.x

Gahlmann A, Moerner WE (2014) Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. Nat Rev Microbiol 12(1):9–22. doi:10.1038/nrmicro3154

Gaillard JL, Berche P, Frehel C, Gouin E, Cossart P (1991) Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci. Cell 65(7):1127–1141

Garcia-del Portillo F, Calvo E, D'Orazio V, Pucciarelli MG (2011) Association of ActA to peptidoglycan revealed by cell wall proteomics of intracellular *Listeria monocytogenes*. J Biol Chem 286(40):34675–34689. doi:10.1074/jbc.M111.230441

Garner EC, Bernard R, Wang W, Zhuang X, Rudner DZ, Mitchison T (2011) Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in *B. subtilis*. Science 333(6039):222–225. doi:10.1126/science.1203285

Gautam S, Kim T, Spiegel DA (2015) Chemical probes reveal an extraseptal mode of cross-linking in *Staphylococcus aureus*. J Am Chem Soc 137(23):7441–7447. doi:10.1021/jacs.5b02972

Govindarajan S, Nevo-Dinur K, Amster-Choder O (2012) Compartmentalization and spatiotemporal organization of macromolecules in bacteria. FEMS Microbiol Rev 36(5):1005–1022. doi:10.1111/j.1574-6976.2012.00348.x

Halbedel S, Hahn B, Daniel RA, Flieger A (2012) DivIVA affects secretion of virulence-related autolysins in *Listeria monocytogenes*. Mol Microbiol 83(4):821–839. doi:10.1111/j.1365-2958.2012.07969.x

Hayhurst EJ, Kailas L, Hobbs JK, Foster SJ (2008) Cell wall peptidoglycan architecture in *Bacillus subtilis*. Proc Natl Acad Sci USA 105(38):14603–14608

Hendrickx AP, Willems RJ, Bonten MJ, van Schaik W (2009) LPxTG surface proteins of enterococci. Trends Microbiol 17(9):423–430. doi:10.1016/j.tim.2009.06.004

Higgins ML, Shockman GD (1970) Model for cell wall growth of *Streptococcus faecalis*. J Bacteriol 101(2):643–648

Hu P, Bian Z, Fan M, Huang M, Zhang P (2008) Sec translocase and sortase A are colocalised in a locus in the cytoplasmic membrane of *Streptococcus mutans*. Arch Oral Biol 53(2):150–154. doi:10.1016/j.archoralbio.2007.08.008

Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, Nageshwar Reddy D (2015) Role of the normal gut microbiota. World J Gastroenterol 21(29):8787–8803. doi:10.3748/wjg.v21.i29.8787

Jiang C, Caccamo PD, Brun YV (2015) Mechanisms of bacterial morphogenesis: evolutionary cell biology approaches provide new insights. BioEssays 37(4):413–425. doi:10.1002/bies.201400098

Jones LJ, Carballido-Lopez R, Errington J (2001) Control of cell shape in bacteria: helical, actin-like filaments in *Bacillus subtilis*. Cell 104(6):913–922

Jonquieres R, Bierne H, Fiedler F, Gounon P, Cossart P (1999) Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: a novel mechanism of protein association at the surface of gram-positive bacteria. Mol Microbiol 34(5):902–914

Jonquieres R, Bierne H, Mengaud J, Cossart P (1998) The inlA gene of *Listeria monocytogenes* LO28 harbors a nonsense mutation resulting in release of internalin. Infect Immun 66(7):3420–3422

Kandaswamy K, Liew TH, Wang CY, Huston-Warren E, Meyer-Hoffert U, Hultenby K, Schroder JM, Caparon MG, Normark S, Henriques-Normark B, Hultgren SJ, Kline KA (2013) Focal targeting by human beta-defensin 2 disrupts localized virulence factor assembly sites in *Enterococcus faecalis*. Proc Natl Acad Sci USA 110(50):20230–20235. doi:10.1073/pnas.1319066110

Kang HJ, Baker EN (2012) Structure and assembly of Gram-positive bacterial pili: unique covalent polymers. Curr Opin Struct Biol 22(2):200–207. doi:10.1016/j.sbi.2012.01.009

Klebba PE, Charbit A, Xiao Q, Jiang X, Newton SM (2012) Mechanisms of iron and haem transport by *Listeria monocytogenes*. Mol Membr Biol 29(3–4):69–86. doi:10.3109/09687688.2012.694485

Kline KA, Kau AL, Chen SL, Lim A, Pinkner JS, Rosch J, Nallapareddy SR, Murray BE, Henriques-Normark B, Beatty W, Caparon MG, Hultgren SJ (2009) Mechanism for sortase localization and the role of sortase localization in efficient pilus assembly in *Enterococcus faecalis*. J Bacteriol 191(10):3237–3247

Kuru E, Tekkam S, Hall E, Brun YV, Van Nieuwenhze MS (2015) Synthesis of fluorescent D-amino acids and their use for probing peptidoglycan synthesis and bacterial growth in situ. Nat Protoc 10(1):33–52. doi:10.1038/nprot.2014.197

Lebrun M, Mengaud J, Ohayon H, Nato F, Cossart P (1996) Internalin must be on the bacterial surface to mediate entry of *Listeria monocytogenes* into epithelial cells. Mol Microbiol 21 (3):579–592

Liew PX, Wang CL, Wong SL (2012) Functional characterization and localization of a *Bacillus subtilis* sortase and its substrate and use of this sortase system to covalently anchor a heterologous protein to the *B. subtilis* cell wall for surface display. J Bacteriol 194(1):161–175. doi:10.1128/JB.05711-11

Lindahl G, Stalhammar-Carlemalm M, Areschoug T (2005) Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. Clin Microbiol Rev 18(1):102–127. doi:10.1128/CMR.18.1.102-127.2005

Lowe J, Amos LA (1998) Crystal structure of the bacterial cell-division protein FtsZ. Nature 391 (6663):203–206. doi:10.1038/34472

Margolin W (2009) Sculpting the bacterial cell. Curr Biol 19(17):R812–R822

Mariscotti JF, Quereda JJ, Garcia-Del Portillo F, Pucciarelli MG (2014) The *Listeria monocytogenes* LPXTG surface protein Lmo1413 is an invasin with capacity to bind mucin. Int J Med Microbiol 304(3–4):393–404. doi:10.1016/j.ijmm.2014.01.003

Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. Microbiol Mol Biol Rev 70(1):192–221

Marraffini LA, Schneewind O (2005) Anchor structure of staphylococcal surface proteins. V. Anchor structure of the sortase B substrate IsdC. J Biol Chem 280(16):16263–16271

Massidda O, Novakova L, Vollmer W (2013) From models to pathogens: how much have we learned about *Streptococcus pneumoniae* cell division? Environ Microbiol 15(12):3133–3157. doi:10.1111/1462-2920.12189

Mazmanian SK, Skaar EP, Gaspar AH, Humayun M, Gornicki P, Jelenska J, Joachmiak A, Missiakas DM, Schneewind O (2003) Passage of heme-iron across the envelope of *Staphylococcus aureus*. Science 299(5608):906–909. doi:10.1126/science.1081147

Mengaud J, Ohayon H, Gounon P, Mege RM, Cossart P (1996) E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells. Cell 84 (6):923–932

Mesnage S, Fontaine T, Mignot T, Delepierre M, Mock M, Fouet A (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J Sep 1;19(17):4473–4484

Michaelis AM, Gitai Z (2010) Dynamic polar sequestration of excess MurG may regulate enzymatic function. J Bacteriol 192(18):4597–4605. doi:10.1128/JB.00676-10

Mirouze N, Ferret C, Yao Z, Chastanet A, Carballido-Lopez R (2015) MreB-Dependent Inhibition of Cell Elongation during the Escape from Competence in *Bacillus subtilis*. PLoS Genet 11(6): e1005299. doi:10.1371/journal.pgen.1005299

Mobley HL, Koch AL, Doyle RJ, Streips UN (1984) Insertion and fate of the cell wall in *Bacillus subtilis*. J Bacteriol 158(1):169–179

Monteiro JM, Fernandes PB, Vaz F, Pereira AR, Tavares AC, Ferreira MT, Pereira PM, Veiga H, Kuru E, VanNieuwenhze MS, Brun YV, Filipe SR, Pinho MG (2015) Cell shape dynamics during the staphylococcal cell cycle. Nat Commun 6:8055. doi:10.1038/ncomms9055

Navarre WW, Schneewind O (1999) Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev: MMBR 63(1):174–229

Nevo-Dinur K, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O (2011) Translation-independent localization of mRNA in *E. coli*. Science 331 (6020):1081–1084. doi:10.1126/science.1195691

Nobbs AH, Lamont RJ, Jenkinson HF (2009) Streptococcus adherence and colonization. Microbiol Mol Biol Rev: MMBR 73(3):407–450. doi:10.1128/MMBR.00014-09

Olshausen PV, Defeu Soufo HJ, Wicker K, Heintzmann R, Graumann PL, Rohrbach A (2013) Superresolution imaging of dynamic MreB filaments in *B. subtilis*–a multiple-motor-driven transport? Biophys J 105(5):1171–1181. doi:10.1016/j.bpj.2013.07.038

Perez-Nunez D, Briandet R, David B, Gautier C, Renault P, Hallet B, Hols P, Carballido-Lopez R, Guedon E (2011) A new morphogenesis pathway in bacteria: unbalanced activity of cell wall

synthesis machineries leads to coccus-to-rod transition and filamentation in ovococci. Mol Microbiol 79(3):759–771. doi:10.1111/j.1365-2958.2010.07483.x

Perry AM, Ton-That H, Mazmanian SK, Schneewind O (2002) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. III. Lipid II is an in vivo peptidoglycan substrate for sortase-catalyzed surface protein anchoring. J Biol Chem 277(18):16241–16248

Personnic N, Bruck S, Nahori MA, Toledo-Arana A, Nikitas G, Lecuit M, Dussurget O, Cossart P, Bierne H (2010) The stress-induced virulence protein InlH controls interleukin-6 production during murine listeriosis. Infect Immun 78(5):1979–1989. doi:10.1128/IAI.01096-09

Philippe J, Vernet T, Zapun A (2014) The elongation of ovococci. Microb Drug Resist 20(3):215–221. doi:10.1089/mdr.2014.0032

Phillips GJ (2001) Green fluorescent protein–a bright idea for the study of bacterial protein localization. FEMS Microbiol Lett 204(1):9–18

Pishchany G, Dickey SE, Skaar EP (2009) Subcellular localization of the *Staphylococcus aureus* heme iron transport components IsdA and IsdB. Infect Immun 77(7):2624–2634. doi:10.1128/IAI.01531-08

Pucciarelli MG, Calvo E, Sabet C, Bierne H, Cossart P, Garcia-del Portillo F (2005) Identification of substrates of the *Listeria monocytogenes* sortases A and B by a non-gel proteomic analysis. Proteomics 5(18):4808–4817

Rafelski SM, Theriot JA (2006) Mechanism of polarization of *Listeria monocytogenes* surface protein ActA. Mol Microbiol 59(4):1262–1279

Raz A, Fischetti VA (2008) Sortase A localizes to distinct foci on the *Streptococcus pyogenes* membrane. Proc Natl Acad Sci USA 105(47):18549–18554. doi:10.1073/pnas.0808301105 (0808301105 [pii])

Raz A, Talay SR, Fischetti VA (2012) Cellular aspects of the distinct M protein and SfbI anchoring pathways in *Streptococcus pyogenes*. Mol Microbiol 84(4):631–647. doi:10.1111/j.1365-2958.2012.08047.x

Reimold C, Defeu Soufo HJ, Dempwolff F, Graumann PL (2013) Motion of variable-length MreB filaments at the bacterial cell membrane influences cell morphology. Mol Biol Cell 24(15):2340–2349. doi:10.1091/mbc.E12-10-0728

Reis O, Sousa S, Camejo A, Villiers V, Gouin E, Cossart P, Cabanes D (2010) LapB, a novel *Listeria monocytogenes* LPXTG surface adhesin, required for entry into eukaryotic cells and virulence. J Infect Dis 202(4):551–562. doi:10.1086/654880

Rosch J, Caparon M (2004) A microdomain for protein secretion in Gram-positive bacteria. Science 304(5676):1513–1515. doi:10.1126/science.1097404

Sabet C, Toledo-Arana A, Personnic N, Lecuit M, Dubrac S, Poupel O, Gouin E, Nahori MA, Cossart P, Bierne H (2008) The *Listeria monocytogenes* virulence factor InlJ is specifically expressed in vivo and behaves as an adhesin. Infect Immun 76(4):1368–1378. doi:10.1128/IAI.01519-07

Schneewind O, Missiakas DM (2012) Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc Lond B Biol Sci 367(1592):1123–1139. doi:10.1098/rstb.2011.0210

Shiomi D, Yoshimoto M, Homma M, Kawagishi I (2006) Helical distribution of the bacterial chemoreceptor via colocalization with the Sec protein translocation machinery. Mol Microbiol 60(4):894–906

Smith GA, Portnoy DA, Theriot JA (1995) Asymmetric distribution of the *Listeria monocytogenes* ActA protein is required and sufficient to direct actin-based motility. Mol Microbiol 17(5):945–951

Speziale P, Pietrocola G, Rindi S, Provenzano M, Provenza G, Di Poto A, Visai L, Arciola CR (2009) Structural and functional role of *Staphylococcus aureus* surface components recognizing adhesive matrix molecules of the host. Future microbiology 4(10):1337–1352. doi:10.2217/fmb.09.102

Spirig T, Weiner EM, Clubb RT (2011) Sortase enzymes in Gram-positive bacteria. Mol Microbiol 82(5):1044–1059. doi:10.1111/j.1365-2958.2011.07887.x

Stachowiak R, Jagielski T, Roeske K, Osinska O, Gunerka P, Wisniewski J, Bielecki J (2015) Lmo0171, a novel internalin-like protein, determines cell morphology of *Listeria monocytogenes* and its ability to invade human cell lines. Curr Microbiol 70(2):267–274. doi:10.1007/s00284-014-0715-4

Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in gram-positive pathogens. Nat Rev Microbiol 4(7):509–519. doi:10.1038/nrmicro1443

Tiyanont K, Doan T, Lazarus MB, Fang X, Rudner DZ, Walker S (2006) Imaging peptidoglycan biosynthesis in *Bacillus subtilis* with fluorescent antibiotics. Proc Natl Acad Sci USA 103 (29):11033–11038. doi:10.1073/pnas.0600829103

Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 459(7249):950–956

Tomasz A, Jamieson JD, Ottolenghi E (1964) The fine structure of *Diplococcus pneumoniae*. J Cell Biol 22:453–467

Tsui HC, Keen SK, Sham LT, Wayne KJ, Winkler ME (2011) Dynamic distribution of the SecA and SecY translocase subunits and septal localization of the HtrA surface chaperone/protease during *Streptococcus pneumoniae* D39 cell division. mBio 2(5):e00202-11. doi:10.1128/mBio.00202-11

Typas A, Banzhaf M, Gross CA, Vollmer W (2012) From the regulation of peptidoglycan synthesis to bacterial growth and morphology. Nat Rev Microbiol 10(2):123–136. doi:10.1038/nrmicro2677

van Teeffelen S, Wang S, Furchtgott L, Huang KC, Wingreen NS, Shaevitz JW, Gitai Z (2011) The bacterial actin MreB rotates, and rotation depends on cell-wall assembly. Proc Natl Acad Sci USA 108(38):15822–15827. doi:10.1073/pnas.1108999108

Vega LA, Port GC, Caparon MG (2013) An association between peptidoglycan synthesis and organization of the *Streptococcus pyogenes* ExPortal. mBio 4(5):e00485–00413. doi:10.1128/mBio.00485-13

Wang S, Furchtgott L, Huang KC, Shaevitz JW (2012) Helical insertion of peptidoglycan produces chiral ordering of the bacterial cell wall. Proc Natl Acad Sci USA 109(10):E595–E604. doi:10.1073/pnas.1117132109

Wheeler R, Mesnage S, Boneca IG, Hobbs JK, Foster SJ (2011) Super-resolution microscopy reveals cell wall dynamics and peptidoglycan architecture in ovococcal bacteria. Mol Microbiol 82(5):1096–1109. doi:10.1111/j.1365-2958.2011.07871.x

Willems RJ, Hanage WP, Bessen DE, Feil EJ (2011) Population biology of Gram-positive pathogens: high-risk clones for dissemination of antibiotic resistance. FEMS Microbiol Rev 35 (5):872–900. doi:10.1111/j.1574-6976.2011.00284.x

Yao Z, Carballido-Lopez R (2014) Fluorescence imaging for bacterial cell biology: from localization to dynamics, from ensembles to single molecules. Annu Rev Microbiol 68:459–476. doi:10.1146/annurev-micro-091213-113034

Yu W, Gotz F (2012) Cell wall antibiotics provoke accumulation of anchored mCherry in the cross wall of *Staphylococcus aureus*. PLoS ONE 7(1):e30076. doi:10.1371/journal.pone.0030076

Zalucki YM, Beacham IR, Jennings MP (2009) Biased codon usage in signal peptides: a role in protein export. Trends Microbiol 17(4):146–150. doi:10.1016/j.tim.2009.01.005

Zapun A, Vernet T, Pinho MG (2008) The different shapes of cocci. FEMS Microbiol Rev 32 (2):345–360. doi:10.1111/j.1574-6976.2007.00098.x

Zhou X, Halladin DK, Rojas ER, Koslover EF, Lee TK, Huang KC, Theriot JA (2015) Bacterial division. Mechanical crack propagation drives millisecond daughter cell separation in *Staphylococcus aureus*. Science 348(6234):574–578. doi:10.1126/science.aaa1511

Pilus Assembly in Gram-Positive Bacteria

Werner Pansegrau and Fabio Bagnoli

Abstract Pili of Gram-positive bacteria are unique structures on the bacterial surface, assembled from covalently linked polypeptide subunits. Pilus assembly proceeds by transpeptidation reactions catalyzed by sortases, followed by covalent anchoring of the filament in the peptidoglycan layer. Another distinctive property is the presence of intramolecular isopeptide bonds, conferring extraordinary chemical and mechanical stability to these elongated structures. Besides their function in cell adhesion and biofilm formation, this section discusses possible application of pilus constituents as vaccine components against Gram-positive pathogens.

Abbreviations

BP	Backbone protein
AP	Ancillary protein
CWSS	Cell wall sorting signal
m-Dap	Meso-2,6-diaminopimelic acid
GAS	Group A streptococcus (*S. pyogenes*)
GBS	Group B streptococcus (*S. agalactiae*)
MIDAS	Metal ion-dependent adhesion site
MSCRAMMs	Microbial surface components recognizing adhesive matrix molecules
PI	Pilus islet
vWFA	von Willebrand factor type A domain

W. Pansegrau (✉) · F. Bagnoli
GSK Vaccines, Siena, Italy
e-mail: werner.x.pansegrau@gsk.com

Current Topics in Microbiology and Immunology (2017) 404:203–233
DOI 10.1007/82_2015_5016

Published Online: 05 January 2016

Contents

1 Introduction

Pili are bacterial surface organelles implied in a variety of important processes regarding host–pathogen interactions, such as cell adhesion, colonization and biofilm formation, but also transformation and motility (Thanassi et al. 2012). It is therefore surprising that, although pili in Gram-positive bacteria had been observed for many years (Kumazawa and Yanagawa 1972; Yanagawa and Otsuki 1970; Yanagawa et al. 1968), their rigorous characterization by molecular biology and biochemistry has started only relatively recently with the dissection of the pili found on the surface of corynebacteria (Ton-That and Schneewind 2004; Ton-That et al. 2004; Ton-That and Schneewind 2003) and streptococci (Barocchi et al. 2006; Lauer et al. 2005; Bagnoli et al. 2008; Kang et al. 2007; Mora et al. 2005). Following these seminal studies, pili have been described in a great variety of pathogenic, but also in commensal and probiotic Gram-positive bacteria.

The surface of Gram-positive bacteria is characterized by the presence of a thick peptidoglycan layer, while an outer membrane, characteristic for the Gram negatives, is absent. It is obvious therefore that the pilus architecture in Gram-positive

bacteria must be fundamentally different from that of pili in Gram-negative species. Indeed, typical single pili found in Gram-positive bacteria are much thinner, ca. 2–3 nm, (Kang et al. 2007) than any known type of pilus in Gram negatives. Another exclusive property is their formation from covalently linked subunits, a process that is catalyzed by sortases transpeptidase enzymes, that not only assemble the pilus filament but also mediate the reaction that leads to anchoring the pilus base covalently in the peptidoglycan.

With the ongoing analysis of these surface structures, a wealth of structural information on Gram-positive pilus components has become available recently (Kang and Baker 2012), revealing another surprising feature: the presence of autocatalytically formed intramolecular isopeptide bonds (Kang et al. 2007, 2009). While in pili from Gram-negative bacteria stabilizing disulfide bonds are widespread, Gram-positive bacteria are lacking machinery that could catalyze the formation of disulfide bonds on the cell surface (Dutton et al. 2008; Heras et al. 2009). Thus, it is tempting to speculate that isopeptide formation in Gram-positive pili has evolved as an alternative means to introduce stabilizing cross-links (Kang and Baker 2011). Another unusual finding in pili of Gram positives consists in the recent discovery of thioester bonds, forming a covalent linkage between pilus tip adhesins and host proteins (Pointon et al. 2010; Linke-Winnebeck et al. 2014) demonstrating the presence of an uninterrupted chain of covalent interactions between the bacterial cell wall and host cell surface proteins.

This review will also discuss the use of pilus components in vaccines against Gram-positive pathogens. Since pili extend from the bacterial surface, evolved to interact with the environment and traversing even the bacterial capsule, they are often the major proteinaceous antigens of a given pathogen that is visible to the immune system. Therefore, pilus components are highly immunogenic, a property that makes them promising candidates for use in vaccines.

2 Pili Are Encoded by Distinct Genetic Islands

In virtually all Gram-positive bacteria in which pili have been described, the genetic information necessary for their formation is specified by distinct genetic elements, designated as pilus islands. Pilus islands encode at least one sortase enzyme and the genes expressing the precursors for the pilus backbone and minor ("ancillary") pilus components (Fig. 1). Although not present in all pilus islands, genes specifying signal peptidase-like polypeptides (SipA) have been found, and in these systems, they are required for pilus formation (Zähner and Scott 2008).

In some cases, regulatory gene products have been identified, suggesting that pilus expression might be transcriptionally regulated (Kreikemeyer et al. 2011). In the case of *Streptococcus pneumoniae* strains harboring pilus island 1 (PI-1), bistability in pilus expression has been demonstrated by the finding that in vitro

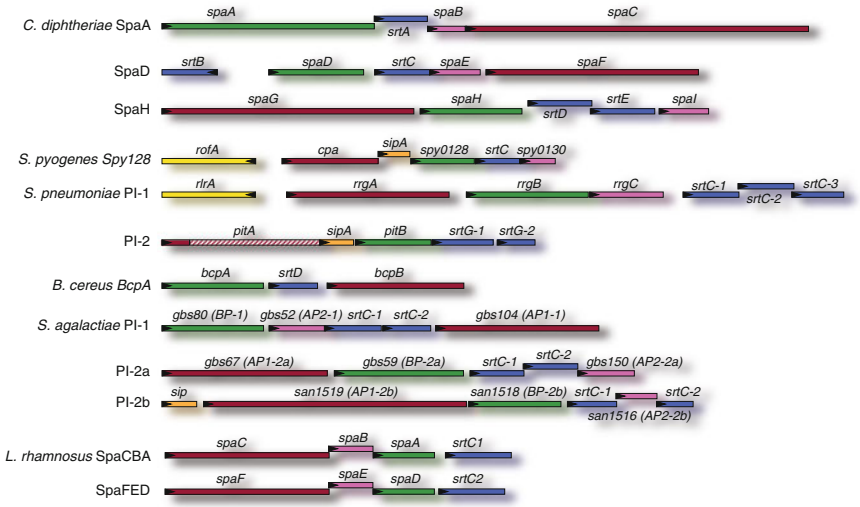


Fig. 1 Operon structure of selected pilus islands. Gene products are colored according to their functions: *red* pilus tip (AP1); *green* pilus shaft/backbone (BP); *magenta* pilus base (AP2); *blue* sortase; *orange* signal peptidase-like protein; *yellow* transcriptional regulator. A *black arrowhead* indicates the direction of transcription. *Boxes* that represent genes in which stop and start codons of adjacent reading frame overlap are depicted in a raised position. The region of the putative *pitA* pseudogene located downstream of a stop codon interrupting the *pitA* reading frame (Bagnoli et al. 2008) is shown as a *hatched bar*

stable bacterial populations can be isolated by colony selection in which either a high (H) or low (L) proportion of individual bacterial cells is expressing pili or not (Pancotto et al. 2013; De Angelis et al. 2011). Expression of PI-1 pili is linked to the positive regulator RlrA, the expression of which seems to be sufficient to switch pilus biogenesis from an "off" to an "on" state (De Angelis et al. 2011) with the biphasic expression pattern resulting from RlrA controlling its own expression via a positive feedback loop (Hava et al. 2003). Infecting mice with either H, L, or unselected (medium, M) populations shows that pilus-1 is preferentially expressed at early colonization stages, consistent with its role in adhesion, while at later stages the expression is partially switched off, probably to avoid exposure of pilus antigens to the host immune system (Pancotto et al. 2013).

Similar characteristics of pilus expression were found in *Streptococcus gallolyticus* where pilus expression depends on a phase variation mechanism involving addition/deletion of GCAGA repeats that modifies the length of an upstream leader peptide. Synthesis of longer leader peptides potentiates the transcription of the pil1 genes through ribosome-induced destabilization of a premature stem-loop transcription terminator (Danne et al. 2014).

3 Sortases Assemble Pili of Gram-Positive Bacteria

3.1 Classes of Sortases and Their Substrates

Based on sequence similarity, sortases have been classified in six main classes designated with the letters A–F (Spirig et al. 2011).

In most Gram-positive species, class A sortases function as the housekeeping sortases, catalyzing the covalent linkage of a variety of surface proteins to the surface peptidoglycan layer (for recent reviews, see (Bradshaw et al. 2015; Spirig et al. 2011). Class B sortases are mostly associated with anchoring gene products implied in iron uptake such as hemoproteins (Mazmanian et al. 2003), but in some cases, e.g., *Streptococcus pyogenes,* they have been also found to catalyze pilus assembly (Kang et al. 2011). Class C sortases are the typical pilus-related sortases specified by pilus islands and required for initiation and elongation in pilus assembly to form extended filaments. Class D sortases have been found in bacilli, but so far, only the one of *Bacillus anthracis* (Ba-SrtC) has been characterized to some extent (Marraffini and Schneewind 2006, 2007), where it seems to be implied in sporulation under certain conditions (Marraffini and Schneewind 2006). Class E sortases in some species, such as *Corynebacterium diphtheriae*, act as housekeeping sortases (Ton-That and Schneewind 2003). Interestingly, the presence of class A and E sortases seems to be mutually exclusive, suggesting that they are indeed functional homologues (Comfort and Clubb 2004). Finally, the class F sortases are found in a wide variety of *Actinomyces* species; however, their role is largely unknown (Spirig et al. 2011).

Proteins that are substrates for sortases in general are exported in a Sec-dependent manner across the cellular membrane, remaining anchored to the latter by a C-terminal hydrophobic sequence patch (Schneewind and Missiakas 2014). A pentapeptide sorting motif, adjacent to the C-terminus and specific for a certain sortase class, is recognized by its cognate sortase, likewise linked to the cellular membrane by a hydrophobic anchor. The sortase cleaves the pentapeptide motif at a certain position and then catalyzes the linkage of a nucleophile to the new C-terminus of the substrate protein. Both the sorting motif and the nucleophile are specific to the sortase implied in the reaction (Fig. 2; Table 1).

3.2 Transpeptidation Mechanism

The catalytic site of sortases consists of a cleft large enough to accommodate a protein substrate with the appropriate LPXTG pentapeptide sorting signal and exposing the catalytic residues constituted by a triad of a cysteine, a histidine, and an arginine residues (Marraffini et al. 2006). The cleavage reaction proceeds via nucleophilic attack by the SH group of the active site cysteine in the conserved catalytic TLXTC motif on a specific peptide bond in the LPXTG pentapeptide

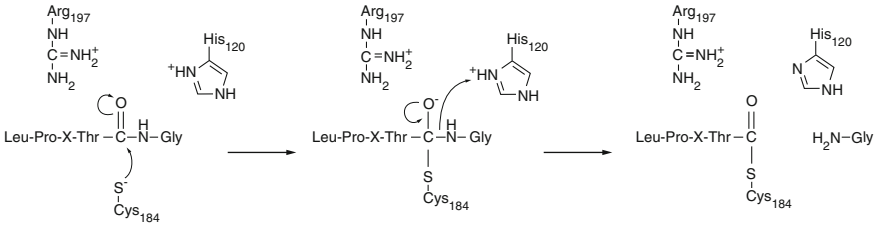


Fig. 2 Reverse protonation model of the sortase cleavage step (Frankel et al. 2007)

Table 1 Sortase classes and their specific sorting motifs and nucleophiles

Sortase class	Pentapeptide motif[a]	Nucleophile	Function
A	LPXT-G	Lipid II (Ala, Gly,), Peptidoglycan cross-bridges (m-Dap)	Housekeeping sortase, pilus anchoring
B	N(P/A)(Q/K)(T/S)-(N/S)	Peptidoglycan cross-bridges (Lys), Pilin (Lys)	Iron deprivation response, pilus polymerization
C	(Y/I/L)(P/A)XT-G	YPKN pilin motif (Lys)	Pilus polymerization and anchoring
D	LPNT-A	Peptidoglycan cross-bridges (m-Dap)	Sporulation, anchoring of endospore envelop proteins
E	LAXT-G	Peptidoglycan cross-bridges (m-Dap)	Housekeeping in some high G + C bacteria

[a]The sortase cleavage site is indicated by a dash (-)

motif, usually after the threonine. This results in the formation of an acyl (thioester) intermediate with the N-terminal portion of the substrate protein and a C-terminal cleavage product, a peptide with a free N-terminus (Fig. 2).

The mechanism of this reaction has been subject to intense debate (Frankel et al. 2005; Ton-That et al. 2002), favoring a reverse protonation model (Frankel et al. 2007) where the LPXTG substrate binds to active sortase in which the cysteine and histidine are reverse protonated; only a minor fraction of total sortase population is present in this configuration. The nucleophilic cysteine thiolate attacks the carbonyl of the scissile Thr–Gly bond to result in the formation of a short-lived tetrahedral oxyanion intermediate, which is stabilized by interaction with the positively charged side chain of the arginine. Protonation of the leaving group facilitates collapse of the intermediate and formation of the acyl-enzyme (Fig. 2; Frankel et al. 2007).

The thioester intermediate that is formed by this mechanism is resolved by a consecutive nucleophilic substitution reaction with the sortase determining the specificity of the nucleophile: either by an N-terminal amino acid in the peptide moiety of a lipid II molecule (class A), meso-2,6-diaminopimelic acid (mDap) in a peptidoglycan cross-bridge (class A, D and E Sortases), a lysine residue in a peptidoglycan cross-bridge (class B sortases), or a lysine residue that is part of the "YPKN" motif of a pilin precursor (class C sortases and certain class B sortases,

(Ton-That and Schneewind 2004; Kang et al. 2011)). The nucleophilic substitution completes the transpeptidation reaction, resulting in anchoring of substrate molecules to peptidoglycan (class A, B and D and E sortases, eventually after incorporation of the attached lipid II into the peptidoglycan) or in incorporation of a pilin precursor into a growing pilus filament (class C and some class B sortases).

A particular characteristic of the class C sortases is the presence of an N-terminal extension not present in the housekeeping sortases (Neiers et al. 2009; Manzano et al. 2009). This region forms a highly flexible loop that acts as a pseudosubstrate and blocks access to the active site (Cozzi et al. 2012). This so-called lid apparently is implied in the regulation of activity and substrate specificity of the class C sortases. Mutagenesis of the *Streptococcus agalactiae* PI-2a sortase SrtC-1 demonstrates that abrogating a particular interaction between a tyrosine residue in the lid and the active site cysteine results in a mutant enzyme that, other than the wild-type enzyme, efficiently polymerizes recombinant BP-2a backbone pilin (GBS59) in vitro (Cozzi et al. 2013). However, while the C-terminal IPQTG sorting motif in BP-2a is essential for the polymerization reaction to occur, the presence of the lysine residue that should provide the nucleophilic ε-amino group to resolve the acyl intermediate is not, suggesting that either the unmutated lid or other unknown factors provide additional specificity determinants allowing only the lysine of the YPKN motif to act as nucleophile in the reaction (Cozzi et al. 2013). In line with these observations, (Khare et al. 2011) presented the crystal structure of a SrtC-1 apo-enzyme dimer in which the lid was shown to be displaced by a loop from a neighboring SrtC-1 molecule containing an LPXTG-like sequence, thus mimicking a substrate intruding into the active site. A part of the displaced lid and the C-terminal hinge that connects the lid to the rest of the molecule in this structure forms a five-turn α-helix that is arranged in a way that it allows access to the active site (Khare et al. 2011).

4 Pili Are Composed of Subunits Linked by Intermolecular Isopeptide Bonds

4.1 Pilus Assembly

Following Sec-dependent transport of pilus precursors across the cell membrane, the pilus building blocks remain attached to the membrane via a C-terminal hydrophobic region acting as a membrane anchor (Fig. 3; Schneewind and Missiakas 2014). According to the accepted model, assembly of pili starts when a C class sortase, likewise anchored in the membrane, recognizes and cleaves the specific sortase motif near the C-terminus of a pilus tip precursor. As detailed before, the initial cleavage reaction results in the formation of thioester intermediate between precursor and sortase that is resolved through nucleophilic attack by the ε-amino group of a specific lysine residue of a backbone precursor–sortase intermediate. Usually, this lysine is a part of the conserved "YPKN" motif (Ton-That

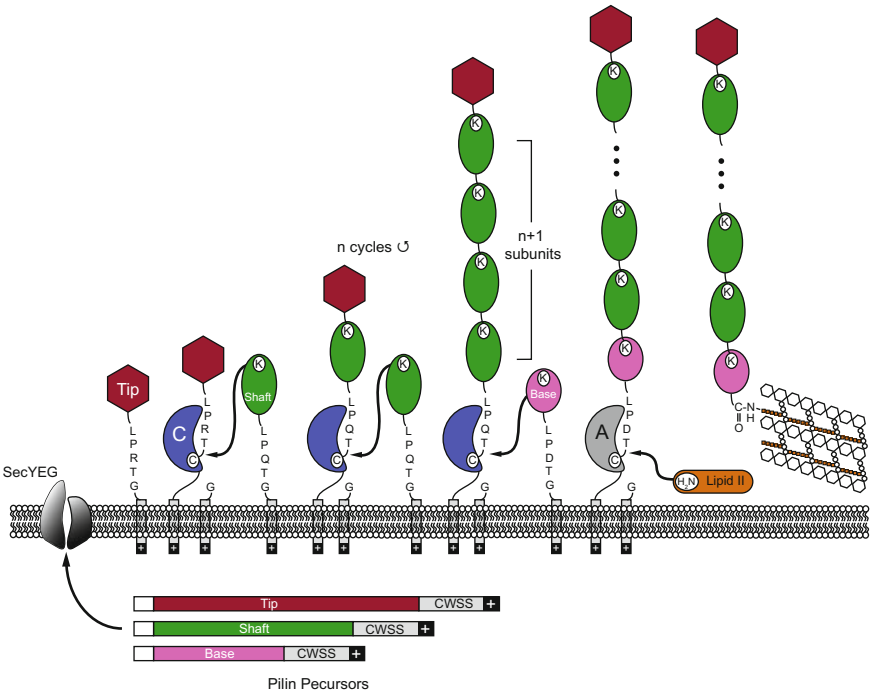


Fig. 3 Current model of pilus assembly. Following secretion of pilin precursors in a sec-dependent manner, the folded intermediates are retained at the surface of the bacterial membrane by a hydrophobic anchor (*CWSS*). In the initiating step, a C-type sortase (*blue*) recognizes the cognate sortase motif and forms a thioacyl intermediate with a pilus tip precursor (*red*). This intermediate is resolved by nucleophilic attack by the ε-amino group of the YPKN lysine of a pilus shaft precursor (*green*) which itself is subject to recognition and thioester formation by a C class sortase. Resolution of the latter intermediate by another shaft precursor and repetition of the process results in elongation of the pilus fiber. Thus, a number of *n* cycles of elongation result in a fiber containing *n* + 1 backbone subunits. Termination of pilus assembly occurs after incorporation of a base precursor (*magenta*), formation of a thioester intermediate with a housekeeping sortase (*gray*), and resolution of the intermediate by lipid II (*orange*). Alternatively, in systems that do not comprise a dedicated base protein, the housekeeping sortase recognizes a shaft precursor and links it to lipid II (not shown). In both cases, the lipid II moiety is finally incorporated into the peptidoglycan layer, covalently anchoring the newly formed pilus and completing the process. Three dots are indicating that pilus fibers may contain a large number of backbone subunits

and Schneewind 2004). The result is a covalent intermolecular isopeptide linkage between pilus tip and the first backbone unit. The latter reaction repeats several times forming the extended pilus filament by adding more and more pilus backbone units (Fig. 3).

How pilus polymerization terminates is species-dependent: In some species such as *Bacillus cereus*, the housekeeping sortase SrtA is able to recognize and form intermediates with backbone precursors, albeit with lower frequency, so that after

incorporation of a certain number of backbone units, a backbone intermediate formed with the housekeeping sortase may be incorporated into the growing pilus (Budzik et al. 2008). This intermediate can only be resolved by nucleophilic attack of the N-terminal glycine of the peptide side chain of a lipid II molecule, so that incorporation of a SrtA-backbone intermediate into the growing pilus acts as a switch, terminating pilus elongation. The lipid II with the attached pilus is then incorporated into the peptidoglycan layer covalently anchoring the pilus base (Budzik et al. 2008).

In many other species, such as *C. diphtheriae* (Swaminathan et al. 2007), *S. agalactiae* (Necchi et al. 2011), or *S. pneumoniae* (Shaik et al. 2014), pilus polymerization terminates when a dedicated precursor (pilus base/anchor), which is only be recognized by the housekeeping sortase, is cleaved and incorporated at the base of the growing pilus via a specific lysine–isopeptide linkage. Since the housekeeping sortase determines the nature of the nucleophile, this intermediate can only be resolved by lipid II, followed by incorporation into the peptidoglycan and anchoring the pilus (Fig. 3).

5 Intramolecular Isopeptide Bonds Stabilize Pili

In their natural environment, the components that form pilus structures may be exposed to very harsh conditions, such as, extreme pH, chaotropes, or proteases, usually resulting in unfolding or degradation of proteins. Moreover, as mediators of bacterial adhesion, pili might also be exposed to extensive mechanical stress by shear forces. To deal with these challenges, pilus components from Gram-negative bacterial species have evolved strategies to withstand, including cross-linking of the polypeptide chain by disulfide bridges (Piatek et al. 2010), presentation of extremely hydrophobic surfaces (Li et al. 2012), and protection of the protein termini by cyclization (Kalkum et al. 2002). Pili of Gram-positive bacteria have developed an alternative strategy, consisting in the formation of intramolecular isopeptide bonds.

5.1 CnaA and B Domains Are Central Elements of Pilus Components

Adhesins in Gram-positive bacteria comprise a class of surface-anchored multidomain proteins, known as MSCRAMMs (microbial surface components recognizing adhesive matrix molecules) (Patti et al. 1994), the prototype of which is the collagen-binding adhesin Cna of *Staphylococcus aureus*. Cna contains a series of IgG-like domains classified as types CnaA (DEv-IgG) and CnaB (IgG-rev) (Deivanayagam et al. 2002; Vengadesan et al. 2011). While the function of CnaA consists in binding of collagen, CnaB-type domains are supposed to form a stalk exposing the CnaA domains at the bacterial surface.

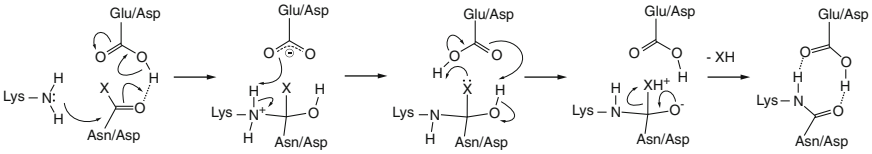


Fig. 4 Autocatalytic formation of intramolecular isopeptide bonds (Kang and Baker 2011). X is either NH_2 (in Asn) or OH (in Asp)

Interestingly, CnaB-like domains were also recognized as the basic modular element in the components that form the polymeric pilus assemblies of Gram-positive bacteria (Kang et al. 2007). A common structural characteristic of CnaB domains is the presence of autocatalytically formed isopeptide bonds. Virtually all pilus proteins in fact harbor one or more CnaB domains with the majority bearing intramolecular isopeptide bonds formed between the ε-amino group of a lysine and the carbonyl of a glutamic/aspartic acid, or the amide group of an asparagine residue.

A canonical arrangement of amino acid side chains has been identified, implied in forming the intramolecular isopeptide linkages (Fig. 4; Kang and Baker 2011). Isopeptide bond formation involves an essential glutamate (or aspartate) residue that by structural analysis has been found to form hydrogen bonds to the isopeptide C = O and NH groups. A plausible mechanism for isopeptide bond formation was first suggested for the bacteriophage capsid protein HK97 (Helgstrand et al. 2003). It consists in polarization of the C = O bond of the Asp/Asn side chain by the protonated glutamate/aspartate, inducing positive charge on Cγ. A subsequent nucleophilic attack on Cγ by the unprotonated Lys ε-amino group results in formation of the isopeptide bond (Fig. 4; Kang et al. 2007).

The catalytic residues are situated in a conserved hydrophobic pocket, and the isopeptide bonds are positioned in most CnaB domains in such a way that they are tying together the first and last β strand, located just before the interdomain connections. In this way, the rest of the domain structure is isolated from external mechanical stress, explaining the extreme resistance of pilus fibers against mechanical unfolding (Alegre-Cebollada et al. 2010).

6 Ancillary Proteins and Pilus Specialization—Examples for Pilus Composition and Architecture in Various Species

6.1 Corynebacterium diphtheriae

C. diphtheriae was one of the first Gram-positive organisms in which pili were extensively characterized and its pili assembly systems have become a paradigm.

Genome sequencing revealed the presence of three pilus islands in *C. diphtheriae* NCTC13129, encoding the SpaA, SpaD, and SpaH pilus types, named after the gene products representing the respective backbone proteins (Ton-That and Schneewind 2003). Each pilus island encodes two ancillary proteins (pilus tip/pilus base) and at least one sortase (Fig. 1).

While most of the work that lead to dissection of the pilus assembly mechanism in *Corynebacterium* has been done using the SpaA pilus as a model system, it was demonstrated that assembly of morphological distinct SpaD pili occurs autonomously and dependent on its cognate sortases, SrtB and C (Gaspar and Ton-That 2006). The prototype pilus contains the backbone pilin SpaA which is forming the shaft, the tip pilin SpaC, and another ancillary pilin, SpaB. SpaB functions as the base protein, demonstrated by the finding that cells lacking SpaB, although forming pilus fibers, mostly shed pili into the medium, a phenotype also observed when cells lack the housekeeping sortase SrtF (Mandlik et al. 2008).

The structure of the backbone pilin SpaA has been solved (Kang et al. 2009), showing that the protein consists of three tandem IgG-like domains, two of which, the N-terminal D1 and the C-terminal D3 domains, with the CnaB-type fold, while the middle D2 domain is of the CnaA-type. The D2 and D3 domains are stabilized by intramolecular isopeptide bonds, absent from the D1 domain. Remarkably, the D3 domain of SpaA contains also a stabilizing disulfide bond (Kang et al. 2009). SpaD and SpaA share several stabilizing features, including the isopeptide bonds in domains D2 and D3 and the disulfide bond in D3. SpaD has an additional isopeptide bond in its D1 domain, while in SpaA, the residues corresponding to the isopeptide bond-forming Lys, Asn, and Asp residues of SpaD are replaced by Ala, His, and Gln, and consequently, in SpaA no D1 isopeptide bond exists (Kang et al. 2014).

6.2 *Bacillus cereus*

B. cereus encodes one of the simpler pilus systems, consisting of a C class sortase, SrtD, required for pilus polymerization, the pilus backbone BcpA, and the pilus tip protein BcpB (Budzik et al. 2007) (Fig. 1). The IPNTG sorting motif of BcpB is recognized exclusively by the C class sortase SrtD, the latter accepting only the ε-amino group of the lysine residue of the YPKN motif of BcpA as a nucleophile, thus ensuring incorporation of at least one BcpA unit into the growing pilus (Budzik et al. 2009a). The LPVTG sorting motif of BcpA on the other hand is also accepted by the housekeeping sortase SrtA, which, upon cleavage, terminates pilus polymerization and anchors the respective subunit to a peptide cross-bridge of the peptidoglycan (Budzik et al. 2008). Recognition of the BcpA pilus subunits by either sortase appears to be stochastic; thus, pilus length is regulated by the local abundances of housekeeping and pilus-related sortases. The involvement of SrtA in both anchoring and determining the average length of pili is demonstrated by the finding that SrtA mutants produce unusually elongated pili that are released into the surrounding medium (Budzik et al. 2007).

The major pilin, BcpA, consist of three CnaB-like domains, CNA_1, CNA_2, and CNA_3, and a jelly-roll domain, XNA, each of which, when part of the assembled pilus, contain one lysine–asparagine isopeptide bond (Budzik et al. 2009b). Interestingly, the isopeptide bond of the N-terminal CNA_1 domain is absent in recombinant BcpA protein, apparently forming only when the precursor becomes part of the pilus structure (Budzik et al. 2009b). It has been speculated that the close vicinity of the Lys162, which is part of the BcpA YPKN motif and forms the intermolecular amide bond with the adjacent pilin, and Asn163, implied in the intramolecular isopeptide bond in CNA1, necessitates that the intramolecular amide bond is formed only in the assembled pilus. This would permit Lys162 to exert its nucleophilic attack on the sortase thioester intermediate, a step that requires major flexibility in this region (Kang et al. 2014). A similar situation was found in *S. pneumoniae* RrgB where the isopeptide of the D1 domain forms only in the assembled pilus or when the protein is arranged in a fiber-like crystal structure (see below) (El Mortaji et al. 2012a).

6.3 Streptococcus pneumoniae

6.3.1 Pilus Islet-1

The more common type of pili in *S. pneumoniae*, also designated as *rlrA* pili, is encoded by a 14-kb region called pilus islet-1 (PI-1) (Barocchi et al. 2006; LeMieux et al. 2006) and present in ca. 30 % of a global collection of clinical isolates of *S. pneumoniae* (Moschioni et al. 2008). The associated pilus operon encodes seven gene products: the positive transcriptional regulator RlrA, three sortases, designated as SrtC-1, SrtC-2, and SrtC-3, the pilus backbone protein RrgB, the pilus tip RrgA, and the pilus base RrgC (LeMieux et al. 2006; Hilleringmann et al. 2008), Fig. 1.

Understanding the significance of the presence of multiple sortases is still fragmentary. While SrtC-1 seems to be catalytically most active, elimination of single sortases does not result in a defective phenotype, indicating overlapping specificities (LeMieux et al. 2008; Fälker et al. 2008). Coexpressing the pilus components with single sortases in *Escherichia coli* (El Mortaji et al. 2012b) suggests that SrtC-1 indeed is most active in polymerizing the backbone RrgB protein, while SrtC-2 apparently is implied in association with RrgA to RrgB and multimerization of RrgA. Finally, SrtC-3 appears to catalyze the occasional integration of RrgA into the pilus shaft and the association between RrgB and the base protein RrgC (El Mortaji et al. 2012b). One should keep in mind, however, that these results were obtained in a heterologous system that may not mimic the high specificities of sortase activities in *S. pneumoniae*. In fact, neither RrgA multimerization nor incorporation of RrgA in the pilus shaft was actually demonstrated in pili isolated from *S. pneumoniae* (Hilleringmann et al. 2009).

The pilus backbone protein RrgB consists of four CnaB-like domains, two of which are arranged in a side by side rather than in an end-to-end arrangement resulting from the fact that D3 appears to be inserted into a loop of D2 (El Mortaji et al. 2012a; Spraggon et al. 2010). Interestingly, RrgB molecules crystallize in a fiber-like arrangement, in which the IPQTG sorting motif of a given RrgB molecule appears to be docked to the D1 domain of an adjacent RrgB. This crystal structure revealed the presence of an isopeptide in the RrgB D1 domain, while an analogous amide bond was not detected in the isolated D1 domain (Gentile et al. 2011; El Mortaji et al. 2012a). This result is in line with observations showing that the intramolecular isopeptide bond of the backbone domain that presents the lysine residue implied in the intermolecular isopeptide bond forms only in the context of the assembled fiber, even if, as in the crystal, the association is non-covalent. Similar observations were made in *B. cereus* BcpA, where an isopeptide bond was detected in the assembled pilus but not in the recombinant protein (see above) (Budzik et al. 2009b).

The ancillary tip protein RrgA mediates adherence to respiratory epithelia and extracellular matrix components such as fibronectin, collagen I, and laminin (Nelson et al. 2007; Hilleringmann et al. 2008). More recent findings show that RrgA interacts also with factors of the innate immune system such as Toll-like receptor (TLR) 2 (Basset et al. 2013) and complement receptor (CR) 3 (Orrskog et al. 2012). The RrgA protein consists of four domains, D1–D4, linked by flexible linker regions that, similarly as in RrgB, are not arranged in an end-to-end arrangement but apparently have evolved by insertion of the D3 domain into a loop of the D2 domain, however leaving the overall structure of D2 intact. The structure is even more complicate since D2/D3 itself is inserted into a loop of D1, so that only the C-terminal D4 domain is uninterrupted. Isopeptide bonds stabilize D2 and D4. The D3 domain contains a MIDAS motif and a region resembling the A3 domain of human von Willebrand factor (vWFA) (Izoré et al. 2010), a molecule that interacts with collagens I and III (Bienkowska et al. 1997; Cruz et al. 1995).

Electron microscopy studies could show for the first time that PI-1 pili have a defined directionality, with the majority of pili exposing one copy of RrgA at the tip and one of RrgC at the pilus base (Hilleringmann et al. 2009). Moreover, direct immune electron microscopic analysis of isolated pili demonstrated the presence of not more than one copy of RrgA per pilus (Hilleringmann et al. 2009). Indeed, RrgC functions as anchor protein, tethering the pilus structure to the peptidoglycan layer (Shaik et al. 2014). The sorting motif of RrgC, VPDTG, is cleaved by the housekeeping sortase linking the protein to the peptide bridge of lipid II (Shaik et al. 2014). Structural analysis of RrgC shows the presence of three CnaB-like domains with reverse IgG fold, designated D1–D3. Isopeptide bonds stabilize D2 and D3, while D1 shows an exposed lysine residue that is part of an IYPK-like motif. In the assembled pilus, this residue is probably linked to the C-terminus of an adjacent RrgB backbone subunit (Shaik et al. 2014).

6.3.2 Pilus Islet-2

Pilus islet-2 is present in 16 % of strains of the above-mentioned global collection of clinical isolates. It is 6575 bp in length and encodes the sortases SrtG1, a pseudogene encoding an incomplete sortase (*srtG2*), and a signal peptidase-like SipA protein (Bagnoli et al. 2008) (Fig. 1). There seems to be only a pilus backbone protein, PitB. A putative ancillary protein could be specified by another pseudogene, *pitA*; however, the reading frame of *pitA* appears to be interrupted by a stop codon. Although the possibility of alternative translation cannot be excluded, a corresponding gene product was not detected and mutants in *pitA* do not show any detectable phenotype (Bagnoli et al. 2008). As in *S. pyogenes*, pilus assembly is dependent on the presence of an active sortase, here SrtG1, and the signal peptidase-like protein SipA (Bagnoli et al. 2008).

The sortase SrtG1 is somewhat particular: Although showing extended similarity to C class sortases, it contains three insertions otherwise present only in B class sortases. The reading frame of the *srtG2* pseudogene terminates just upstream of the position where in SrtG1, a B class sortase insertion is located. Moreover, the putative sorting signal present near the C-terminus of PitB, VTPTG, is different from any known consensus sequence. For this reason, a separate class, G, was proposed for these enzymes (Bagnoli et al. 2008).

The piliation phenotype associated with PI-2 differs considerably from that of PI-1: Whereas in the latter huge numbers of pili are present on the surface of expressing cells, PI-2-type pili usually are present as a single filament that extends further from the cell surface than PI-1 pili (Bagnoli et al. 2008). Despite the apparent absence of a tip protein, usually fulfilling the adhesin function, PI-2 pili have been demonstrated to mediate adhesion of *S. pneumoniae* to various respiratory cell lines, suggesting that the PitB backbone protein has also adhesin function (Bagnoli et al. 2008).

6.4 Streptococcus agalactiae

Three related pilus islands (PI-1, PI-2a, and Pi-2b) have been identified in circulating S. *agalactiae* (GBS) strains (Rosini et al. 2006; Dramsi et al. 2006). All GBS isolates contain at least one pilus island, with PI-2a and PI-2b mutually exclusive since they are located at the same chromosomal position. Each pilus island encodes three pilus components, a backbone protein designated BP, and two ancillary proteins, designated AP1 and AP2. In addition, each PI specifies two C class sortases required for pilus assembly. All three pilus islands show a similar gene organization (Fig. 1). The pilus proteins of PI-1 and PI-2b differ by very few amino acids, while PI-2a is more variable with seven alleles described for both the BP and the AP1 presenting sequence identities between 48–98 % for BP and 87–98 % for AP2 (Rosini and Margarit 2015).

The structures of backbone proteins of representatives of PI-1, PI-2a, and PI-2b have been partially determined, showing that they are composed of classical IgG-like folds, with BP-1 composed of three (Vengadesan et al. 2011), BP-2a of four (Nuccitelli et al. 2011), and BP-2b (Cozzi et al. 2015) of three independently folded domains. All three structures were obtained only after removal of the N-terminal D1 domain; BP-1 consists of three IgG-like domains, with D2 of the Dev-IgG/CnaA-type and D3 of the IgG-rev/CnaB-type fold. Molecular modeling using the *C. diphtheriae* SpaA structure as a template suggests that the D1 domain protrudes from the pilus shaft, probably contributing to the rigidity of the pilus fiber (Vengadesan et al. 2011). The 3D structure of BP-2a resembles that of *S. pneumoniae* RrgB, with an organization into four independently folded IgG-like domains (one CnaA and three CnaB domains). Analogous to *S. pneumoniae* RrgB, the third (D3) domain is positioned as a lateral insertion into the CnaA-type D2 domain (Nuccitelli et al. 2011).The spatial arrangement of D2 and D4 domains overlaps the structural architecture of D2 and D3 domains of BP-2b or BP-1, confirming that the overall fold of the core of these proteins is very well conserved.

Gbs104 is the pilus tip protein encoded by PI-1(AP1-1). Its structure is very similar to that of *S. pneumoniae* RrgA, constituted by four independently folded domains (N1-4), with N1 and N4 belonging to the CnaB-type, while N2 is a CnaA-type domain (Krishnan et al. 2013). As in RrgA, the domain structure is nonlinear with N3 inserted into N2 and N2/N3 inserted into the N-terminal N1 domain. The N3 domain, as in RrgA, is a vWFA-like domain containing a metal ion-dependent adhesion site (MIDAS). The C-terminal 7 helix and the MIDAS of vWFA-like domains (also known as I domains in integrins) can adopt two conformations: a closed low-affinity state and an open high-affinity state (Zhang et al. 2009; Luo et al. 2004; Krishnan et al. 2013). After mutating Thr564 and Lys571 to Cys residues in the isolated GBS104 N3 domain, the MIDAS motif was shown to be locked in the open conformation, resulting in considerably higher affinity to fibronectin as compared to the corresponding wild-type construct (Krishnan et al. 2013). Likewise, for the corresponding PI-2a tip protein AP1-2a, an essential role of the vWFA domain for adhesion but not for biofilm formation was demonstrated (Konto-Ghiorghi et al. 2009).

AP2 is the pilus base protein found close to the bacterial surface. The essential role of the GBS housekeeping sortase SrtA in anchoring the GBS PI-2a pilus was shown by specific cleavage of quenched fluorescent peptides containing either the pentapeptide motifs of BP-2a, AP1-2a, or AP2-2a (Necchi et al. 2011). Only the AP2-2a-related peptide was cleaved by GBS SrtA, as demonstrated by an increase of fluorescence intensity. The same exclusive specificity was observed during incubation of recombinant AP2-2a protein with GBS SrtA, resulting in the formation of transpeptidation and cleavage products, whereas AP1-2a and BP-2a under the same conditions remained inert (Necchi et al. 2011). The structure of PI-1 AP2-1 (GBS52) revealed the presence of two CnaB/IgG-rev folds, N1 and N2, and a YPKN-like motif near the region that links the two domains possibly providing the lysine residue forming the intermolecular isopeptide bond that links AP2-1 to the pilus shaft (Vengadesan et al. 2011).

6.5 *Streptococcus pyogenes*

More than 70 years ago, Rebecca Lancefield and coworkers (Lancefield 1940) described a method to classify hemolytic group A streptococci into serotypes, based on the serological properties of two antigen classes: the M and the T antigens. While the M proteins had been studied extensively (Smeesters et al. 2010), until recently much less was known about the nature of the trypsin-resistant T antigens of which 21 serotypes have been identified. With the first description of pili in GAS, it became clear that the different T antigens correspond to different pilus types that are encoded by the fibronectin-binding collagen-binding T antigen (FCT) regions of the *S. pyogenes* chromosome (Falugi et al. 2008; Mora et al. 2005). So far, nine different FCT regions have been identified, encoding 15 different variants of pilus backbone proteins and accounting for 17 of the 21 Lancefield T serotypes (Falugi et al. 2008).

The backbone protein of GAS strain M1 (Spy128, Fig. 1) was characterized extensively and has become one of the paradigms in pilus structure and assembly (Kang et al. 2007). Spy128 consists of two IgG-like domains each of which contains an intramolecular isopeptide bond. Crystal packing of the protein resembles apparently the arrangement of the monomers in the pilus fiber, placing an invariant lysine residue near the C-terminus of an adjacent monomer (Kang et al. 2007). Interestingly, this lysine residue is not part of a canonic YPKN motif found in other pilins, and the sortase involved in the assembly of the pilus fiber is a B-type sortase (Kang et al. 2011). Spy128 became one of the first proteins in which the thermodynamic (Kang and Baker 2009) and mechanic properties (Alegre-Cebollada et al. 2010) of pilus backbone proteins were studied, showing the extraordinary resistance of pilus subunits stabilized by isopeptide bonds.

Several FCA regions, i.e., types 2, 3, and 4, encode signal peptidase-like proteins. For the T3 pilus, it was demonstrated that the corresponding SipA2 protein is essential for T3 pilin polymerization and linkage of the Cpa tip protein (Zähner and Scott 2008). Although the protein sequence of SipA2 shows clear similarity to signal peptidases, in T3 pilus assembly, SipA2 apparently functions more like a chaperone, interacting with pilus components but lacking catalytic residues required for protease function (Zähner and Scott 2008).

Spy125/Cpa and Spy130 are the ancillary components of the *S. pyogenes* M1 pilus with Spy125/Cpa functioning as the tip adhesin (AP1) (Abbot et al. 2007; Smith et al. 2010) and Spy130 as the base protein (AP2) recognized by the housekeeping sortase and linked to the peptidoglycan (Smith et al. 2010). The structures of both proteins have been characterized (Linke et al. 2010; Solovyova et al. 2010). The crystal structure of a three-domain C-terminal fragment of the AP1 protein Cpa from the T1/M1 strain SF370 unexpectedly revealed a thioester bond joining the side chains of a cysteine residue (Cys426) and a glutamine residue (Gln575) in its CpaT domain (Pointon et al. 2010). The thioester bond was found in a solvent-accessible groove on the protein surface. While not contributing to the stability of Cpa (Walden et al. 2014; Pointon et al. 2010), the occurrence of a

thioester bond in a pilus tip protein known to function as an adhesin (Quigley et al. 2009; Smith et al. 2010) is intriguing.

Cys–Gln thioester bonds are present in the human complement proteins C3 and C4 (Law and Dodds 1997), and proteolytic activation of C3 and C4 results in a conformational rearrangement exposing these thioester bonds. Nucleophilic moieties such as amino or hydroxyl groups present in proteins and bacterial cell wall components may attack the exposed thioester, leading to the formation of a covalent linkage between the glutamine residue and the attacking amine or hydroxyl group (Law and Dodds 1997).

The thioester group in Cpa apparently reacts in an analogous way since it was shown that Cpa forms a covalent complex with the polyamine spermidine (Linke-Winnebeck et al. 2014). Spermidine most probably is not the natural receptor for Cpa; however, as a small and abundant molecule, it may preferentially reach the partially exposed thioester group. It is likely that binding of Cpa to its actual receptor molecule results in a conformational change allowing close contact between a nucleophile on the receptor surface and the thioester group. Cpa consists of four domains: the CnaB domains, CnaB1 and CnaB2, and the CpaN and CpaT domains which show extended sequence homology among each other, including the thioester motif. It has been demonstrated that also CpaN does indeed contain a thioester bond so that the Cpa present in most *S. pyogenes* strains has two reactive thioester bonds, hence representing two-headed adhesins (Linke-Winnebeck et al. 2014). The thioester sequence motif is conserved in equivalent positions in all AP1 proteins of *S. pyogenes* FCT types 2, 3, and 4; moreover, it has been identified in a variety of surface-exposed proteins from various Gram-positive species, suggesting that thioester-mediated covalent adhesion is actually a widespread mechanism (Linke-Winnebeck et al. 2014).

6.6 *Lactobacillus rhamnosus*

Lactobacillus rhamnosus GG is a well-known Gram-positive probiotic strain, whose health-benefiting properties correlate with prolonged residence in the gastrointestinal tract, a property probably caused by adherence to the intestinal mucosa (Reunanen et al. 2012). Two pilus gene clusters (*spaCBA* and *spaFED*; Fig. 1) were identified in the genome of *L. rhamnosus*, each of which contained the predicted genes for three pilin subunits and a single sortase (Kankainen et al. 2009). While SpaCBA pili were demonstrated to be present on the surface of *L. rhamnosus*, the significance of the SpaFED pilus island is less clear (Reunanen et al. 2012). Expression of SpaFED pili in wild-type *L. rhamnosus* has not been demonstrated yet, however putting the operon under transcriptional control of a nisin-inducible promoter and expressing it in *Lactococcus lactis* resulted in assembly of SpaFED pili on the *L. lactis* surface (Rintahaka et al. 2014). This result demonstrates that the spaFED pilus island has the potential for independent expression of a second pilus type in *L. rhamnosus*, although the conditions that could lead to its expression have

yet to be defined. Heterologous expression of the SpaFED pilus, however, allowed for a comparison of its effects host cellular responses, showing that SpaFED pili, in contrast to SpaCBA pili (von Ossowski et al. 2013) exert a dampening effect on immunogenic-related reactions (Rintahaka et al. 2014).

The intestinal mucus-binding capacity of *L. rhamnosus* has been attributed to the presence of the SpaCBA pilus and in particular to one of the pilin subunits, SpaC (von Ossowski et al. 2010). In fact, also SpaC displays similarity with von Willebrand lectin-binding domains (Kankainen et al. 2009). As a major exception to the canonical Gram-positive pilus model, the position of the SpaC subunit is not restricted to the pilus tip. Instead, SpaC can be found all along the pilus fiber and, furthermore, has a 1:2 ratio in molecular numbers with the backbone protein SpaA (Reunanen et al. 2012). Using single-molecule atomic force microscopy to unravel the binding mechanism of the pili confirmed that SpaC is a multifunctional adhesin with broad specificity (Tripathi et al. 2013). SpaC mediates homophilic transinteractions engaged in bacterial aggregation and specifically binds mucin and collagen, two major extracellular components of host epithelial layers. Homophilic and heterophilic interactions display similar binding strengths and dissociation rates (Tripathi et al. 2013). Performing pulling experiments on living bacteria, the authors demonstrated that SpaCBA pili exhibit two unique mechanical responses, the zipper-like adhesion, involving multiple SpaC molecules distributed along the pilus length, and a nanospring property, enabling the pili to resist high forces (Tripathi et al. 2013).

7 Pilus Proteins Are Promising Vaccine Components

The polypeptides that form pilus structures on the surface of pathogens are considered as very promising candidates for vaccine components. Being elongated structures, pili are well exposed to the host immune system and as their role usually consist in mediating attachment, thus promoting host cell adhesion or biolayer formation, inhibition of these functions by specific antibodies should result in disruption of processes important for the virulence of a respective pathogen. Thus, protectivity of pilus-specific antibodies would not only rely on antigen binding followed by opsonization but also on direct disruption of bacterial virulence functions and blocking infection already in the colonization phase (Soriani and Telford 2010).

Finally, in most human pathogens, genes encoding pilus components are sufficiently conserved, so that including only one or a few components in a vaccine would cover a broad variety of strains of a given pathogen. In some cases, not all strains of a given pathogen are piliated (e.g., *S. pneumoniae*, see below); in these cases, successful vaccination would require additional component to cover the non-piliated variants. The following paragraphs, together with Table 2, will provide some examples of pilus components used in experimental vaccines or in vaccines currently in clinical studies.

Table 2 Selection of pilus components in vaccines against Gram-positive pathogens

Pathogen	Indication	Available vaccines on the market	Medical need	Antigens	Coverage	References
S. agalactiae	Maternal vaccination to prevent GBS infection of newborns	None	Protection against neonatal meningitis and sepsis caused by GBS infections	GBS80 (BP-1), GBS 104 (AP-1)	87 % against a panel of 12 strains	(Maione et al. 2005)
S. agalactiae	Maternal vaccination to prevent GBS infection of newborns	None	Protection against neonatal meningitis and sepsis caused by GBS infections	BP-1 BP-2b AP1-2a	94 % predicted	(Margarit et al. 2009)
S. agalactiae	Maternal vaccination to prevent GBS infection of newborns	None	Protection against neonatal meningitis and sepsis caused by GBS infections	6xD3 (fusion protein consisting of the D3 domains from six BP-2a variants)	GBS carrying PI-2a	(Nuccitelli et al. 2011)
S. pneumoniae	Otitis media, pneumonia	Conjugate vaccines (PCV7, PCV13); Polysaccharide vaccine (PPV-23)	Inexpensive vaccine with broad coverage against various *S. pneumoniae* infections	RrgA, RrgB, RrgC	30 % (100 % for piliated strains)	(Gianfaldoni et al. 2007)
S. pneumoniae	Otitis media, pneumonia	Conjugate vaccines (PCV7, PCV13); Polysaccharide vaccine (PPV-23)	Inexpensive vaccine with broad coverage against various *S. pneumoniae* infections	RrgB231	30 % (100 % for piliated strains)	(Harfouche et al. 2012)
S. pyogenes	*S. pyogenes* infections	None	Broad coverage against infections by *S. pyogenes*	Cpa, Spy128, Spy130	Not known	(Mora et al. 2005)

(continued)

Table 2 (continued)

Pathogen	Indication	Available vaccines on the market	Medical need	Antigens	Coverage	References
E. faecalis	CAUTI	None	Control of catheter-associated urinary tract infections in hospital environment	EbpA, EbpA^{NTD} (N-terminal domain)	Not known	(Flores-Mireles et al. 2014)
L. lactis	N.a. (Technology platform)	N.a.	An inexpensive platform for expressing and presenting heterologous pilus-based vaccine antigens	Examples: Gene products of GBS PI-2a, *S. pyogenes* FCT-2	N.a.	(Buccato et al. 2006; Manetti et al. 2007)
L. lactis	N.a. (Technology Platform)	N.a.	An inexpensive platform for expressing and presenting various vaccine antigens as tip protein hybrids (UPTOP)	Example: *E. coli* MBP-Cpa hybrid	N.a.	(Quigley et al. 2010)

N.a. Not applicable

7.1 *Streptococcus agalactiae*

S. agalactiae (Group B Streptococcus: GBS) is a widespread microorganism that colonizes 15–35 % of healthy women (Schuchat 1998), but it can also cause overwhelming neonatal infections as a result of direct transmission of the bacteria from colonized pregnant women to their babies. The risk of neonatal infection is inversely proportional to the maternal amount of specific antibodies to the capsular polysaccharide, which are transferred from the mothers to the babies through the placenta (Lin et al. 2004). GBS vaccines are expected to induce not only protective antibodies that can be transplacentally transferred to the fetus but also mucosal antibodies preventing colonization of the mothers, thus further contributing to neonate protection.

A first attempt to create a universal vaccine against GBS consisted in screening a panel of GBS isolates of diverse serotype for protective antigens (Maione et al. 2005). Four proteins were identified that elicited protective responses when the offspring of immunized female mice was challenged with the same panel of strains. In combination, these four antigens turned out to be broadly protective against a wide variety of GBS strains (Maione et al. 2005). It turned out that two of the best protective antigens were the pilus backbone protein BP-1 and the tip ancillary protein AP-1 encoded by PI-1 (Lauer et al. 2005).

Based on the observation that the best protective antigens in GBS are pilus components, another approach consisted in combining pilus components from the three PIs that are found in circulating strains of GBS (Margarit et al. 2009). All GBS isolates bear at least one of the three pilus islands mentioned before. While gene products of homologous PI are well conserved, there is much less sequence similarity between functional homologs from different PIs. Therefore, a combination of components from the three PIs that would cover virtually every circulating GBS strain was conceived by combining those gene products that showed the lowest variability among alleles in different strains. Indeed, the backbone proteins of PI-1 and PI-2b combined with the tip ancillary protein 1 of PI-2a turned out to be a highly protective combination, eliciting significant protection against all tested strains (Margarit et al. 2009).

A third attempt involved the creation of a broadly protective antigen based on the GBS BP-2a protein and, for the first time, applied a structural vaccinology approach (Nuccitelli et al. 2011). Following structural analysis of one of the six main variants of BP-2a (BP-2a-515), single domains of BP-2a-515 were expressed in *E. coli* and tested for their protectivity in the maternal mouse immunization/neonatal pup challenge model. As a result, the D3 domain turned out to be the most important for protection. In the next step, a fusion protein containing a head-to-tail arrangement of the D3 domains of the six main variants of BP-2a, in which the domains were linked by pentapeptide spacers, was constructed and expressed in *E. coli* (Nuccitelli et al. 2011). The resulting fusion protein, designated 6xD3, elicited broad protection against neonatal challenge with a panel of GBS strains expressing different BP-2a allelic variants (Nuccitelli et al. 2011).

7.2 *Streptococcus pneumoniae*

S. pneumoniae, is a frequent colonizer of the human upper respiratory tract. On the other hand, this commensal pathogen can cause serious diseases such as pneumonia and meningitis and is known to be the most important vaccine-preventable cause of death in children under 5, affecting up to 1 million children each year. Pneumococci are characterized by the great variety of capsule types, making the generation of a polysaccharide-based vaccine a costly and difficult undertaking. Current vaccines against *S. pneumoniae* are 13 (PCV13 or Prevnar 13®) or even 23-valent (PPSV23 or Pneumovax 23®), so a protein-based vaccine with high coverage is highly desirable (Barocchi et al. 2007). Following the discovery of pilus island 1 in pneumococci (Barocchi et al. 2006; LeMieux et al. 2006), the respective pilus components were tested for their ability to elicit protective antibodies (Gianfaldoni et al. 2007). Using a mouse model of infection, the authors could demonstrate that intraperitoneal immunization with the three subunits, RrgA, RrgB, and RrgC, alone or in combination conferred active and passive protection against strains expressing PI-1 (Gianfaldoni et al. 2007). In order to broaden the coverage after the discovery that three main clades of PI-1 with considerable sequence diversion exist in *S. pneumoniae* (Moschioni et al. 2008), a fusion protein, consisting of a tandem arrangement of the backbone proteins encoded by the three alleles, was constructed and tested in animal models (Harfouche et al. 2012). RrgB231, as the fusion protein was called, elicited antibodies against proteins from all three clades and protected mice against challenge with pneumococcal strains expressing PI-1. RrgB321 antisera mediated complement-dependent opsonophagocytosis of piliated strains at levels comparable to those achieved with the PCV7 glycoconjugate vaccine (Harfouche et al. 2012). Interestingly, protective immunity was observed even with *S. pneumoniae* that were low pilus-expressing populations as mentioned in Sect. 2 (Moschioni et al. 2012). Since only 30 % of circulating clinical *S. pneumoniae* strains contain PI-1 (Moschioni et al. 2008), RrgB321 was conceptualized as a component to be used in combination with other antigens that would cover the non-piliated strains.

7.3 *Enterococcus faecalis*

Catheter-associated urinary tract infections (CAUTIs) are the most common cause of hospital-acquired infections. A frequent cause for these infections are enterococci, which are especially difficult to control in the hospital environment due to their intrinsic resistance to antibiotics along with tolerance to heat and aseptic solutions. Furthermore, the advent of multidrug-resistant enterococci withstanding treatment with all commonly used antibiotics, including vancomycin (Arias and Murray 2008), requires new prophylactic strategies that could most notably include

vaccination. Animal models of CAUTI have shown that endocarditis- and biofilm-associated pili (Ebp) are contributing both to biofilm formation and disease (Nallapareddy et al. 2006).

Ebp pili are typical members of the three component Gram-positive pilus family with the major shaft protein EbpC and two ancillary proteins, EbpA and EbpB, which represent pilus tip and base proteins, respectively (Sillanpää et al. 2013). Finally, a C class sortase, SrtC, is implicated in pilus assembly, while the housekeeping sortase SrtA is required for covalently attaching the mature pilus fiber to the enterococcus peptidoglycan layer (Nielsen et al. 2013; Sillanpää et al. 2013). Of particular interest is the EbpA protein, a fibrinogen-binding protein that, similarly to the *S. pneumoniae* RrgA and GBS AP1 tip proteins, contains a van Willebrand factor A (vWA) domain with a MIDAS motif (Nielsen et al. 2012). The fibrinogen-binding property of EbpA plays a pivotal role in CAUTIs since fibrinogen is deposited on the catheter surface as part of the host inflammatory response, allowing piliated enterococci to colonize the urinary tract (Flores-Mireles et al. 2014). In addition, fibrinogen is used by *E. faecalis* for growth and thus further promotes biofilm formation on the catheter. Immunizing mice either with the full-length EbpA or with the N-terminal domain of EbpA, containing the fibrinogen-binding domain with the vWA motif and the MIDAS, inhibited biofilm formation in vivo and protected the mice against CAUTI (Flores-Mireles et al. 2014). In contrast, immunizing with the C-terminal part of EbpA or with the other pilus components had no significant protective effect. Analyses in vitro demonstrated that protection was associated with a serum antibody response that blocked EbpA binding to fibrinogen and the formation of a fibrinogen-dependent biofilm on catheters (Flores-Mireles et al. 2014).

7.4 *Streptococcus pyogenes*

The human-specific Gram-positive pathogen *S. pyogenes* (Group A streptococcus, GAS) causes a great variety of human diseases, such as pharyngitis, impetigo, invasive disease, necrotizing fasciitis, and autoimmune sequels, and therefore, the availability of a vaccine would be of great benefit. Following the discovery that the *S. pyogenes* (GAS) Lancefield T antigens in fact correspond to pilus backbone proteins, Mora et al. (2005) confirmed that a combination of three recombinant pilus proteins, Cpa, Spy128, and Spy130 originating from the M1 strain SF370, protected mice against nasal infection with a mouse-adapted M1 strain (Mora et al. 2005). Under conditions where ca. 90 % of unimmunized mice were killed, more than 70 % of the immunized mice survived, a result comparable to that obtained by immunization with M protein-based vaccines (Mora et al. 2005). However, while there are more than 100 serotypes of M protein, the number of T types is considerably lower, ca. 20, so that with a pilus-based GAS vaccine, broader coverage may be achieved combining fewer antigens in a vaccine. Considering that adhesion to human pharyngeal cells in *S. pyogenes* is mediated by pili, while in pilus-negative mutants,

aggregation, biofilm formation, and adhesion are severely impaired (Manetti et al. 2007), a vaccine that elicited a specific immune response against GAS pili is expected to interrupt GAS infections already in the colonization phase.

7.5 *Lactococcus lactis*

L. lactis is a nonpathogenic organism, employing pili for mucosal attachment and biofilm formation (Oxaran et al. 2012; Meyrand et al. 2013; Le et al. 2013). Several attempts have been undertaken to use *L. lactis* to expose heterologous pili on its surface. One promising approach consists in expressing complete pilus islands in *L. lactis*. This approach was exploited for GBS (Buccato et al. 2006) and GAS (Manetti et al. 2007) showing that in this way, the pili were indeed assembled on the *L. lactis* surface and among the most immunogenic and broadly recognized antigens.

Another way to expose antigens on the *L. lactis* surface is based on a principle for which the term UPTOP (unhindered presentation on tips of pili) was coined (Quigley et al. 2010). It is based on the creation of hybrid pilus tip proteins containing portions of pilus components originating in pathogenic bacteria, but retaining the potential to be incorporated into the pilus (Quigley et al. 2010).

Both approaches may provide an inexpensive and stable alternative to current mechanisms of immunization for many serious human pathogens.

8 Conclusions

Pili of Gram-positive bacteria are fascinating and unique macromolecular structures implied in host-bacterium and interbacterial interactions. While the basic transpeptidation reactions leading to their assembly by specialized sortase enzymes are known, details of the process are still missing. For example, what are the in vivo events leading to opening of the lid that covers the active site in C class sortases, so that pilus polymerization can proceed? Abundant structural information has contributed a lot to understanding the function of pilus components. In particular, the tip proteins reveal some unexpected features, such as thioester groups promoting covalent adhesion and the van Willebrand factor domains, both until recently only known as functional elements of eukaryotic proteins.

Another highly interesting field opened with the discovery of intramolecular isopeptide bonds. While their important role in hardening pilus filaments against chemical and mechanical stress seems obvious, again many details remain to be elucidated. Where and when do they form? It seems unlikely that Sec-dependent protein transport could accomplish secretion of fully formed CnaA or CnaB domains, so it remains to be understood what keeps isopeptide bonds from forming prematurely in the cytoplasm.

Further open questions concern the regulation of pilus assembly. Is pilus formation regulated in response to external stimuli and if yes: What are these stimuli and how does pilus formation result in a benefit for the bacteria when responding to a certain signal? Other poorly understand aspects of regulation concern spatial distribution of pili on the cell surface (density, polar or equatorial preference, foci or equal distribution) and the dynamics of pilus formation.

Apart from the scientific interest, pilus components have important applications as vaccine components as shown by the successful formulation of backbone and tip proteins of GBS in vaccines currently undergoing clinical trials. Another hall mark in vaccine development is the recent application of structural vaccinology in the construction of a hybrid fusion protein that assembles the most protective epitopes from major GBS variants in a chain of independently folded domains originating from different pilus backbone variants and providing a broad protective response. Such an approach could have an extremely wide range of applicability in vaccine development, whenever numerous variants of an infectious agent exist. Finally, the pilus assembly machinery itself may be employed as an inexpensive delivery system by exposing vaccine antigens on the surface of nonpathogenic bacteria such as *L. lactis*.

Acknowledgments The authors are grateful to Markus Hilleringmann for critical reading of the manuscript and numerous highly appreciated suggestions.

References

Abbot EL, Smith WD, Siou GPS, Chiriboga C, Smith RJ, Wilson JA et al (2007) Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. Cell Microbiol 9:1822–1833

Alegre-Cebollada J, Badilla CL, Fernández JM (2010) Isopeptide bonds block the mechanical extension of pili in pathogenic *Streptococcus pyogenes*. J Biol Chem 285:11235–11242

Arias CA, Murray BE (2008) Emergence and management of drug-resistant enterococcal infections. Expert Rev Anti Infect Ther 6:637–655

Bagnoli F, Moschioni M, Donati C, Dimitrovska V, Ferlenghi I, Facciotti C et al (2008) A second pilus type in *Streptococcus pneumoniae* is prevalent in emerging serotypes and mediates adhesion to host cells. J Bacteriol 190:5480–5492

Barocchi MA, Ries J, Zogaj X, Hemsley C, Albiger B, Kanth A et al (2006) A pneumococcal pilus influences virulence and host inflammatory responses. Proc Natl Acad Sci USA 103:2857–2862

Barocchi MA, Censini S, Rappuoli R (2007) Vaccines in the era of genomics: the pneumococcal challenge. Vaccine 25:2963–2973

Basset A, Zhang F, Benes C, Sayeed S, Herd M, Thompson C et al (2013) Toll-like receptor (TLR) 2 mediates inflammatory responses to oligomerized RrgA pneumococcal pilus type 1 protein. J Biol Chem 288:2665–2675

Bienkowska J, Cruz M, Atiemo A, Handin R, Liddington R (1997) The von Willebrand factor A3 domain does not contain a metal ion-dependent adhesion site motif. J Biol Chem 272:25162–25167

Bradshaw WJ, Davies AH, Chambers CJ, Roberts AK, Shone CC, Acharya KR (2015) Molecular features of the sortase enzyme family. FEBS J 282:2097–2114

Buccato S, Maione D, Rinaudo CD, Volpini G, Taddei AR, Rosini R et al (2006) Use of Lactococcus lactis expressing pili from group B Streptococcus as a broad-coverage vaccine against streptococcal disease. J Infect Dis 194:331–340

Budzik JM, Marraffini LA, Schneewind O (2007) Assembly of pili on the surface of *Bacillus cereus* vegetative cells. Mol Microbiol 66:495–510

Budzik JM, Oh SY, Schneewind O (2008) Cell wall anchor structure of BcpA pili in *Bacillus anthracis*. J Biol Chem 283:36676–36686

Budzik JM, Oh SY, Schneewind O (2009a) Sortase D forms the covalent bond that links BcpB to the tip of *Bacillus cereus* pili. J Biol Chem 284:12989–12997

Budzik JM, Poor CB, Faull KF, Whitelegge JP, He C, Schneewind O (2009b) Intramolecular amide bonds stabilize pili on the surface of bacilli. Proc Natl Acad Sci USA 106:19992–19997

Comfort D, Clubb RT (2004) A comparative genome analysis identifies distinct sorting pathways in gram-positive bacteria. Infect Immun 72:2710–2722

Cozzi R, Prigozhin D, Rosini R, Abate F, Bottomley MJ, Grandi G et al (2012) Structural basis for group B streptococcus pilus 1 sortases C regulation and specificity. PLoS ONE 7:e49048

Cozzi R, Zerbini F, Assfalg M, D'Onofrio M, Biagini M, Martinelli M et al (2013) Group B streptococcus pilus sortase regulation: a single mutation in the lid region induces pilin protein polymerization in vitro. FASEB J 27:3144–3154

Cozzi R, Malito E, Lazzarin M, Nuccitelli A, Castagnetti A, Bottomley MJ et al (2015) Structure and assembly of group B streptococcus pilus 2b backbone protein. PLoS ONE 10:e0125875

Cruz MA, Yuan H, Lee JR, Wise RJ, Handin RI (1995) Interaction of the von Willebrand factor (vWF) with collagen. Localization of the primary collagen-binding site by analysis of recombinant vWF a domain polypeptides. J Biol Chem 270:10822–10827

Danne C, Dubrac S, Trieu-Cuot P, Dramsi S (2014) Single cell stochastic regulation of pilus phase variation by an attenuation-like mechanism. PLoS Pathog 10:e1003860

De Angelis G, Moschioni M, Muzzi A, Pezzicoli A, Censini S, Delany I et al (2011) The *Streptococcus pneumoniae* pilus-1 displays a biphasic expression pattern. PLoS ONE 6:e21269

Deivanayagam CC, Wann ER, Chen W, Carson M, Rajashankar KR, Hook M, Narayana SV (2002) A novel variant of the immunoglobulin fold in surface adhesins of *Staphylococcus aureus*: crystal structure of the fibrinogen-binding MSCRAMM, clumping factor A. EMBO J 21:6660–6672

Dramsi S, Caliot E, Bonne I, Guadagnini SP, Prévost MC, Kojadinovic M. et al (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60:1401–1413

Dutton RJ, Boyd D, Berkmen M, Beckwith J (2008) Bacterial species exhibit diversity in their mechanisms and capacity for protein disulfide bond formation. Proc Natl Acad Sci USA 105:11933–11938

El Mortaji L, Contreras-Martel C, Moschioni M, Ferlenghi I, Manzano C, Vernet T et al (2012a) The full-length *Streptococcus pneumoniae* major pilin RrgB crystallizes in a fibre-like structure, which presents the D1 isopeptide bond and provides details on the mechanism of pilus polymerization. Biochem J 441:833–841

El Mortaji L, Fenel D, Vernet T, Di G (2012b) Association of RrgA and RrgC into the *Streptococcus pneumoniae* pilus by sortases C-2 and C-3. Biochemistry 51:342–352

Fälker S, Nelson AL, Morfeldt E, Jonas K, Hultenby K, Ries J et al (2008) Sortase-mediated assembly and surface topology of adhesive pneumococcal pili. Mol Microbiol 70:595–607

Falugi F, Zingaretti C, Pinto V, Mariani M, Amodeo L, Manetti AGO et al (2008) Sequence variation in group A streptococcus pili and association of pilus backbone types with lancefield T serotypes. J Infect Dis 198:1834–1841

Flores-Mireles AL, Pinkner JS, Caparon MG, Hultgren SJ (2014) EbpA vaccine antibodies block binding of *Enterococcus faecalis* to fibrinogen to prevent catheter-associated bladder infection in mice. Sci Transl Med 6:254ra127

Frankel BA, Kruger RG, Robinson DE, Kelleher NL, McCafferty DG (2005) *Staphylococcus aureus* sortase transpeptidase SrtA: insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. Biochemistry 44:11188–11200

Frankel BA, Tong Y, Bentley ML, Fitzgerald MC, McCafferty DG (2007) Mutational analysis of active site residues in the *Staphylococcus aureus* transpeptidase SrtA. Biochemistry 46:7269–7278

Gaspar AH, Ton-That H (2006) Assembly of distinct pilus structures on the surface of *Corynebacterium diphtheriae*. J Bacteriol 188:1526–1533

Gentile MA, Melchiorre S, Emolo C, Moschioni M, Gianfaldoni C, Pancotto L et al (2011) Structural and functional characterization of the *Streptococcus pneumoniae* RrgB pilus backbone D1 domain. J Biol Chem 286:14588–14597

Gianfaldoni C, Censini S, Hilleringmann M, Moschioni M, Facciotti C, Pansegrau W et al (2007) *Streptococcus pneumoniae* pilus subunits protect mice against lethal challenge. Infect Immun 75:1059–1062

Harfouche C, Filippini S, Gianfaldoni C, Ruggiero P, Moschioni M, Maccari S et al (2012) RrgB321, a fusion protein of the three variants of the pneumococcal pilus backbone RrgB, is protective *in vivo* and elicits opsonic antibodies. Infect Immun 80:451–460

Hava DL, Hemsley CJ, Camilli A (2003) Transcriptional regulation in the *Streptococcus pneumoniae rlrA* pathogenicity islet by RlrA. J Bacteriol 185:413–421

Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L (2003) The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. J Mol Biol 334:885–899

Heras B, Shouldice SR, Totsika M, Scanlon MJ, Schembri MA, Martin JL (2009) DSB proteins and bacterial pathogenicity. Nat Rev Microbiol 7:215–225

Hilleringmann M, Giusti F, Baudner BC, Masignani V, Covacci A, Rappuoli R et al (2008) Pneumococcal pili are composed of protofilaments exposing adhesive clusters of Rrg A. PLoS Pathog 4:e1000026

Hilleringmann M, Ringler P, Muller SA, De AG, Rappuoli R, Ferlenghi I, Engel A (2009) Molecular architecture of *Streptococcus pneumoniae* TIGR4 pili. EMBO J 28:3921–3930

Izoré T, Contreras-Martel C, El ML, Manzano C, Terrasse R, Vernet T et al (2010) Structural basis of host cell recognition by the pilus adhesin from *Streptococcus pneumoniae*. Structure 18:106–115

Kalkum M, Eisenbrandt R, Lurz R, Lanka E (2002) Tying rings for sex. Trends Microbiol 10:382–387

Kang HJ, Baker EN (2009) Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of *Streptococcus pyogenes*. J Biol Chem 284:20729–20737

Kang HJ, Baker EN (2011) Intramolecular isopeptide bonds: protein crosslinks built for stress? Trends Biochem Sci 36:229–237

Kang HJ, Baker EN (2012) Structure and assembly of Gram-positive bacterial pili: unique covalent polymers. Curr Opin Struct Biol 22:200–207

Kang HJ, Coulibaly Fl, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. Science 318:1625–1628

Kang HJ, Paterson NG, Gaspar AH, Ton-That H, Baker EN (2009) The *Corynebacterium diphtheriae* shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. Proc Natl Acad Sci USA 106:16967–16971

Kang HJ, Coulibaly F, Proft T, Baker EN (2011) Crystal structure of Spy0129, a *Streptococcus pyogenes* class B sortase involved in pilus assembly. PLoS ONE 6:e15969

Kang HJ, Paterson NG, Kim CU, Middleditch M, Chang C, Ton-That H, Baker EN (2014) A slow-forming isopeptide bond in the structure of the major pilin SpaD from *Corynebacterium diphtheriae* has implications for pilus assembly. Acta Crystallogr D Biol Crystallogr 70:1190–1201

Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P et al (2009) Comparative genomic analysis of Lactobacillus rhamnosus GG reveals pili containing a human- mucus binding protein. Proc Natl Acad Sci USA 106:17193–17198

Khare B, Fu ZQ, Huang IH, Ton-That H, Narayana SV (2011) The crystal structure analysis of group B Streptococcus sortase C1: a model for the "lid" movement upon substrate binding. J Mol Biol 414:563–577

Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, Caliot E, Trieu-Cuot P, Dramsi S (2009) Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. PLoS Pathog 5:e1000422

Kreikemeyer B, Gamez G, Margarit I, Giard JC, Hammerschmidt S, Hartke A, Podbielski A (2011) Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic Streptococci and Enterococci. Int J Med Microbiol 301:240–251

Krishnan V, Dwivedi P, Kim BJ, Samal A, Macon K, Ma X et al (2013) Structure of *Streptococcus agalactiae* tip pilin GBS104: a model for GBS pili assembly and host interactions. Acta Crystallogr D Biol Crystallogr 69:1073–1089

Kumazawa N, Yanagawa R (1972) Chemical properties of the pili of *Corynebacterium renale*. Infect Immun 5:27–30

Lancefield RC (1940) Type-specific antigens, M and T, of matt and glossy variants of group a hemolytic streptococci. J Exp Med 71:521–537

Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, Rosini R et al (2005) Genome analysis reveals pili in group B Streptococcus. Science 309:105

Law SK, Dodds AW (1997) The internal thioester and the covalent binding properties of the complement proteins C3 and C4. Protein Sci 6:263–274

Le DT, Tran TL, Duviau MP, Meyrand M, Guerardel Y, Castelain M et al (2013) Unraveling the role of surface mucus-binding protein and pili in muco-adhesion of *Lactococcus lactis*. PLoS ONE 8:e79850

LeMieux J, Hava DL, Basset A, Camilli A (2006) RrgA and RrgB are components of a multisubunit pilus encoded by the *Streptococcus pneumoniae rlrA* pathogenicity islet. Infect Immun 74:2453–2456

LeMieux J, Woody S, Camilli A (2008) Roles of the sortases of *Streptococcus pneumoniae* in assembly of the RlrA pilus. J Bacteriol 190:6002–6013

Li J, Egelman EH, Craig L (2012) Structure of the *Vibrio cholerae* type IVb Pilus and stability comparison with the *Neisseria gonorrhoeae* type IVa pilus. J Mol Biol 418:47–64

Lin FY, Weisman LE, Azimi PH, Philips JB III, Clark P, Regan J et al (2004) Level of maternal IgG anti-group B streptococcus type III antibody correlated with protection of neonates against early-onset disease caused by this pathogen. J Infect Dis 190:928–934

Linke C, Young PG, Kang HJ, Bunker RD, Middleditch MJ, Caradoc-Davies TT et al (2010) Crystal structure of the minor pilin FctB reveals determinants of Group A streptococcal pilus anchoring. J Biol Chem 285:20381–20389

Linke-Winnebeck C, Paterson NG, Young PG, Middleditch MJ, Greenwood DR, Witte G, Baker EN (2014) Structural model for covalent adhesion of the *Streptococcus pyogenes* pilus through a thioester bond. J Biol Chem 289:177–189

Luo BH, Takagi J, Springer TA (2004) Locking the beta3 integrin I-like domain into high and low affinity conformations with disulfides. J Biol Chem 279:10215–10221

Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M et al (2005) Identification of a universal group B streptococcus vaccine by multiple genome screen. Science 309:148–150

Mandlik A, Das A, Ton-That H (2008) The molecular switch that activates the cell wall anchoring step of pilus assembly in gram-positive bacteria. Proc Natl Acad Sci USA 105:14147–14152

Manetti AGO, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F et al (2007) *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. Mol Microbiol 64:968–983

Manzano C, Izore T, Job V, Di Guilmi AM, Dessen AA, Izoré T (2009) Sortase activity is controlled by a flexible lid in the pilus biogenesis mechanism of gram-positive pathogens. Biochemistry 48:10549–10557

Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezzo C, Buttazzoni E et al (2009) Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis 199:108–115

Marraffini LA, Schneewind O (2006) Targeting proteins to the cell wall of sporulating *Bacillus anthracis*. Mol Microbiol 62:1402–1417

Marraffini LA, Schneewind O (2007) Sortase C-mediated anchoring of BasI to the cell wall envelope of *Bacillus anthracis*. J Bacteriol 189:6425–6436

Marraffini LA, Dedent AC, Schneewind O (2006) Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. Microbiol Mol Biol Rev 70:192–221

Mazmanian SK, Skaar EP, Gaspar AH, Humayun M, Gornicki P, Jelenska J et al (2003) Passage of heme-iron across the envelope of *Staphylococcus aureus*. Science 299:906–909

Meyrand M, Guillot A, Goin M, Furlan S, Armalyte J, Kulakauskas S et al (2013) Surface proteome analysis of a natural isolate of *Lactococcus lactis* reveals the presence of pili able to bind human intestinal epithelial cells. Mol Cell Proteomics 12:3935–3947

Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG et al (2005) Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci USA 102:15641–15646

Moschioni M, Donati C, Muzzi A, Masignani V, Censini S, Hanage WP et al (2008) *Streptococcus pneumoniae* contains 3 rlrA pilus variants that are clonally related. J Infect Dis 197:888–896

Moschioni M, De Angelis G, Harfouche C, Bizzarri E, Filippini S, Mori E et al (2012) Immunization with the RrgB321 fusion protein protects mice against both high and low pilus-expressing *Streptococcus pneumoniae* populations. Vaccine 30:1349–1356

Nallapareddy SR, Singh KV, Sillanpaa J, Garsin DA, Hook M, Erlandsen SL, Murray BE (2006) Endocarditis and biofilm-associated pili of *Enterococcus faecalis*. J Clin Invest 116:2799–2807

Necchi F, Nardi-Dei V, Biagini M, Assfalg M, Nuccitelli A, Cozzi R et al (2011) Sortase a substrate specificity in GBS pilus 2a cell wall anchoring. PLoS ONE 6:e25300

Neiers F, Madhurantakam C, Fälker S, Manzano C, Dessen A, Normark S et al (2009) Two crystal structures of pneumococcal pilus sortase C provide novel insights into catalysis and substrate specificity. J Mol Biol 393:704–716

Nelson AL, Ries J, Bagnoli F, Dahlberg S, Fälker S, Rounioja S et al (2007) RrgA is a pilus-associated adhesin in *Streptococcus pneumoniae*. Mol Microbiol 66:329–340

Nielsen HV, Guiton PS, Kline KA, Port GC, Pinkner JS, Neiers F et al (2012) The metal ion-dependent adhesion site motif of the *Enterococcus faecalis* EbpA pilin mediates pilus function in catheter-associated urinary tract infection. MBio 3:e00177–12

Nielsen HV, Flores-Mireles AL, Kau AL, Kline K, Pinkner JS, Neiers F et al (2013) Pilin and sortase residues critical for endocarditis-and biofilm-associated pilus biogenesis in *Enterococcus faecalis*. J Bacteriol 195:4484–4495

Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, Necchi F, Norais N 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 USA 108:10278–10283

Orrskog S, Rounioja S, Spadafina T, Gallotta M, Norman M, Hentrich K et al (2012) Pilus adhesin RrgA interacts with complement receptor 3, thereby affecting macrophage function and systemic pneumococcal disease. MBio 4:e00535–12

Oxaran V, Ledue-Clier F, Dieye Y, Herry JM, Pechoux C, Meylheuc T et al (2012) Pilus biogenesis in *Lactococcus lactis*: molecular characterization and role in aggregation and biofilm formation. PLoS ONE 7:e50989

Pancotto L, De AG, Bizzarri E, Barocchi MA, Del GG, Moschioni M, Ruggiero P (2013) Expression of the *Streptococcus pneumoniae* pilus-1 undergoes on and off switching during colonization in mice. Sci Rep 3:2040

Patti JM, Allen BL, McGavin MJ, Hook M (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. Annu Rev Microbiol 48:585–617

Piatek R, Bruzdziak P, Wojciechowski M, Zalewska-Piatek B, Kur J (2010) The noncanonical disulfide bond as the important stabilizing element of the immunoglobulin fold of the Dr fimbrial DraE subunit. Biochemistry 49:1460–1468

Pointon JA, Smith WD, Saalbach G, Crow A, Kehoe MA, Banfield MJ (2010) A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. J Biol Chem 285:33858–33866

Quigley BR, Zähner D, Hatkoff M, Thanassi DG, Scott JR, Zähner D (2009) Linkage of T3 and Cpa pilins in the *Streptococcus pyogenes* M3 pilus. Mol Microbiol 72:1379–1394

Quigley BR, Hatkoff M, Thanassi DG, Ouattara M, Eichenbaum Z, Scott JR (2010) A foreign protein incorporated on the Tip of T3 pili in *Lactococcus lactis* elicits systemic and mucosal immunity. Infect Immun 78:1294–1303

Reunanen J, von Ossowski I, Hendrickx AP, Palva A, de Vos WM (2012) Characterization of the SpaCBA pilus fibers in the probiotic Lactobacillus rhamnosus GG. Appl Environ Microbiol 78:2337–2344

Rintahaka J, Yu X, Kant R, Palva A, von Ossowski I (2014) Phenotypical analysis of the *Lactobacillus rhamnosus* GG fimbrial spaFED operon: surface expression and functional characterization of recombinant SpaFED pili in *Lactococcus lactis*. PLoS ONE 9:e113922

Rosini R, Margarit I (2015) Biofilm formation by *Streptococcus agalactiae*: influence of environmental conditions and implicated virulence factors. Front Cell Infect Microbiol 5:6

Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, Maione D et al (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in *Streptococcus agalactiae*. Mol Microbiol 61:126–141

Schneewind O, Missiakas D (2014) Sec-secretion and sortase-mediated anchoring of proteins in gram-positive bacteria. Biochim Biophys Acta 1843:1687–1697

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

Shaik MM, Maccagni A, Tourcier G, Di Guilmi AM, Dessen AA (2014) Structural basis of pilus anchoring by the ancillary pilin RrgC of *Streptococcus pneumoniae*. J Biol Chem 289:16988–16997

Sillanpää J, Chang C, Singh KV, Montealegre MC, Nallapareddy SR, Harvey BR et al (2013) Contribution of individual Ebp Pilus subunits of *Enterococcus faecalis* OG1RF to pilus biogenesis, biofilm formation and urinary tract infection. PLoS ONE 8:e68813

Smeesters PR, McMillan DJ, Sriprakash KS (2010) The streptococcal M protein: a highly versatile molecule. Trends Microbiol 18:275–282

Smith WD, Pointon J, Abbot E, Kang HJ, Baker EN, Hirst BH et al (2010) Roles of minor pilin subunits Spy0125 and Spy0130 in the serotype M1 *Streptococcus pyogenes* strain SF370. J Bacteriol 192:4651–4659

Solovyova AS, Pointon JA, Race PR, Smith WD, Kehoe MA, Banfield MJ (2010) Solution structure of the major (Spy0128) and minor (Spy0125 and Spy0130) pili subunits from *Streptococcus pyogenes*. Eur Biophys J EBJ 39:469–480

Soriani M, Telford JL (2010) Relevance of pili in pathogenic streptococci pathogenesis and vaccine development. Future Microbiol 5:735–747

Spirig T, Weiner EM, Clubb RT (2011) Sortase enzymes in gram-positive bacteria. Mol Microbiol 82:1044–1059

Spraggon G, Koesema E, Scarselli M, Malito E, Biagini M, Norais N et al (2010) Supramolecular organization of the repetitive backbone unit of the *Streptococcus pneumoniae* pilus. PLoS ONE 5:e10919

Swaminathan A, Mandlik A, Swierczynski A, Gaspar A, Das A, Ton-That H (2007) Housekeeping sortase facilitates the cell wall anchoring of pilus polymers in *Corynebacterium diphtheriae*. Mol Microbiol 66:961–974

Thanassi DG, Bliska JB, Christie PJ (2012) Surface organelles assembled by secretion systems of gram-negative bacteria: diversity in structure and function. FEMS Microbiol Rev 36:1046–1082

Ton-That H, Schneewind O (2003) Assembly of pili on the surface of *Corynebacterium diphtheriae*. Mol Microbiol 50:1429–1438

Ton-That H, Schneewind O (2004) Assembly of pili in gram-positive bacteria. Trends Microbiol 12:228–234

Ton-That H, Mazmanian SK, Alksne L, Schneewind O (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 277:7447–7452

Ton-That H, Marraffini LA, Schneewind O (2004) Sortases and pilin elements involved in pilus assembly of *Corynebacterium diphtheriae*. Mol Microbiol 53:251–261

Tripathi P, Beaussart A, Alsteens D, Dupres V, Claes I, von Ossowski I et al (2013) Adhesion and nanomechanics of pili from the probiotic *Lactobacillus rhamnosus* GG. ACS nano 7:3685–3697

Vengadesan K, Ma X, Dwivedi P, Ton-That H, Narayana SV (2011) A model for group B Streptococcus pilus type 1: the structure of a 35-kDa C-terminal fragment of the major pilin GBS80. J Mol Biol 407:731–743

von Ossowski I, Reunanen J, Satokari R, Vesterlund S, Kankainen M, Huhtinen H et al (2010) Mucosal adhesion properties of the probiotic *Lactobacillus rhamnosus* GG SpaCBA and SpaFED pilin subunits. Appl Environ Microbiol 76:2049–2057

von Ossowski I, Pietila TE, Rintahaka J, Nummenmaa E, Makinen VM, Reunanen J et al (2013) Using recombinant Lactococci as an approach to dissect the immunomodulating capacity of surface piliation in probiotic *Lactobacillus rhamnosus* GG. PloS one 8:e64416

Walden M, Crow A, Nelson MD, Banfield MJ (2014) Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the *S. pyogenes* pilus adhesin Spy0125. Proteins 82:517–527

Yanagawa R, Otsuki K (1970) Some properties of the pili of *Corynebacterium renale*. J Bacteriol 101:1063–1069

Yanagawa R, Otsuki K, Tokui T (1968) Electron microscopy of fine structure of *Corynebacterium renale* with special reference to pili. Jpn J Vet Res 16:31–37

Zähner D, Scott JR (2008) SipA is required for pilus formation in *Streptococcus pyogenes* serotype M3. J Bacteriol 190:527–535

Zhang H, Liu JH, Yang W, Springer T, Shimaoka M, Wang JH (2009) Structural basis of activation-dependent binding of ligand-mimetic antibody AL-57 to integrin LFA-1. Proc Natl Acad Sci USA 106:18345–18350

Type VII Secretion Systems in Gram-Positive Bacteria

Daria Bottai, Matthias I. Gröschel and Roland Brosch

Abstract Bacterial secretion systems are sophisticated molecular machines that fulfil a wide range of important functions, which reach from export/secretion of essential proteins or virulence factors to the implication in conjugation processes. In contrast to the widely distributed Sec and Twin Arginine Translocation (TAT) systems, the recently identified ESX/type VII systems show a more restricted distribution and are typical for mycobacteria and other high-GC *Actinobacteria*. Similarly, type VII-like secretion systems have been described in low-GC Gram-positive bacteria belonging to the phylum *Firmicutes*. While the most complex organization of type VII secretion systems currently known is found in slow-growing mycobacteria, which harbour up to 5 chromosomal-encoded systems (ESX-1 to ESX-5), much simpler organization is reported for type VII-like systems in *Firmicutes*. In this chapter, we describe common and divergent features of type VII- and type VII-like secretion pathways and also comment on their biological key roles, many of which are related to species-/genus-specific host–pathogen interactions and/or virulence mechanisms.

Contents

D. Bottai
Dipartimento di Ricerca Traslazionale e delle Nuove Tecnologie in Medicina e Chirurgia, University of Pisa, Pisa, Italy

M.I. Gröschel · R. Brosch (✉)
Institut Pasteur, Unit for Integrated Mycobacterial Pathogenomics, Paris, France
e-mail: roland.brosch@pasteur.fr

M.I. Gröschel
Department of Pulmonary Diseases and Tuberculosis, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Current Topics in Microbiology and Immunology (2017) 404:235–265
DOI 10.1007/82_2015_5015

Published Online: 05 January 2016

1 Introduction

Bacteria have evolved a multitude of secretion pathways for the export of a wide range of substrates (Costa et al. 2015), some of which may also help the bacteria to shape and/or adapt to an intracellular environment (Stanley and Cox 2013; Majlessi et al. 2015). The effector molecules transported by the secretion systems often play crucial roles in different physiologic processes, in horizontal gene transfer and/or in host–pathogen interactions (adhesion, pathogenicity and survival). In addition to the widely distributed Sec and Twin Arginine Translocation (TAT) systems, which are responsible for substrate translocation across the cytoplasmic membrane, both in Gram-positive and Gram-negative bacteria, other less broadly conserved secretion systems fulfil more specialized functions. These latter systems include the different specialized secretion systems of Gram-negative bacteria, known as type I, type II, type III, type IV, type V and type VI secretion pathways, which are responsible for secretion of substrates across the typical cell envelop of Gram-negative bacteria that is constituted by a plasma membrane and an outer membrane (Costa et al. 2015). In contrast, few specialized secretion systems are known in Gram-positive bacteria. One exception constitute the so-called type VII secretion systems, which have been first described in *Mycobacteria*, where they serve as specialized secretion machineries devoted to the export of peculiar subsets of protein substrates across the complex and highly hydrophobic mycobacterial cell envelope (Mahairas et al. 1996; Pym et al. 2002; Abdallah et al. 2007; Brodin et al. 2004b). However, type VII-like systems have also been identified by in silico analyses in other closely related high-C+G bacterial species (*Actinobacteria*), as well as in more distantly related Gram-positive bacteria belonging to the low-C+G group (*Firmicutes*) of Gram-positive bacteria (Gey Van Pittius et al. 2001; Pallen 2002; Houben et al. 2014).

In this book chapter, we focus on type VII secretion systems in *Actinobacteria* and also describe the type VII-like secretion systems in other Gram-positive

bacteria, with focus on the most relevant features of the Ess system of *Staphylococcus aureus* and the Yuk/Yue system of *Bacillus subtilis*, for which a functional secretion activity has recently been demonstrated.

2 The Discovery of Type VII Secretion Systems

The first type VII substrate to be identified was the 6-kDa early-secreted antigenic target ESAT-6 (EsxA) of *Mycobacterium tuberculosis*. This small protein, which is lacking a classical N-terminal signal sequence, is present in large amounts in short-term culture filtrate of *M. tuberculosis* and was shown to act as an immunodominant T-cell antigen (Sorensen et al. 1995). During comparative and functional genomic analyses, it was found that certain attenuated strains (e.g. *Mycobacterium bovis* BCG or *Mycobacterium microti*) lacked the gene *esxA* encoding ESAT-6, as well as the genomic region up- and downstream of *esxA* (Mahairas et al. 1996; Pym et al. 2002; Brodin et al. 2002), suggesting a key role of this region in mycobacterial virulence, which was later confirmed by complementation (Pym et al. 2002, 2003; Brodin et al. 2004a) and gene knockout studies (Lewis et al. 2003; Stanley et al. 2003; Hsu et al. 2003). These studies also provided evidence of an ESAT-6 specialized secretion system, independently referred to as ESAT-6 system (ESX-1) (Brodin et al. 2004b), Snm system (Converse and Cox 2005; MacGurn et al. 2005; Stanley et al. 2003), or type VII secretion system (Abdallah et al. 2007; Bitter et al. 2009a, b). This classification is in line with the diderm structure of the mycobacterial cell envelope (Zuber et al. 2008; Hoffmann et al. 2008), which in addition to the cytosolic inner membrane includes an outer membrane (the mycomembrane). It was speculated that the mycomembrane might be functionally equivalent to the classical Gram-negative outer membrane despite a different chemical composition (Hoffmann et al. 2008).

In *M. tuberculosis*, ESAT-6 (EsxA) and its protein partner CFP-10 (10-kDa culture filtrate protein, EsxB) are representative members of the large Esx protein family that is constituted of 23 small (size of ~100 amino acids (aa)), highly immunogenic secreted proteins, sharing a conserved Trp-Xaa-Gly (WXG) motif and a characteristic hairpin structure with the WXG domain at the helix-turn-helix bend (Pallen 2002; Cole et al. 1998; Gey Van Pittius et al. 2001). Genes encoding for Esx family members usually lie in tandem pairs and, in five cases, are flanked by blocks of conserved gene clusters coding for components of the ESX secretory apparatus responsible for secretion of the corresponding ESAT-6-like proteins (Cole et al. 1998; Tekaia et al. 1999; Gey Van Pittius et al. 2001). Bioinformatic analyses identified genes encoding ESAT-6-like proteins both in the genomes of other *Actinobacteria* (*Nocardia*, *Corynebacteriae*, *Streptomyces*) (Gey Van Pittius et al. 2001) and in a number of bacterial species belonging to the phylum *Firmicutes* (Pallen 2002). Although the homology in the primary sequence is low relative to ESAT-6 from *M. tuberculosis*, certain features allow to classify these proteins into the WXG-100 superfamily (Pallen 2002). Bacteria with WXG proteins

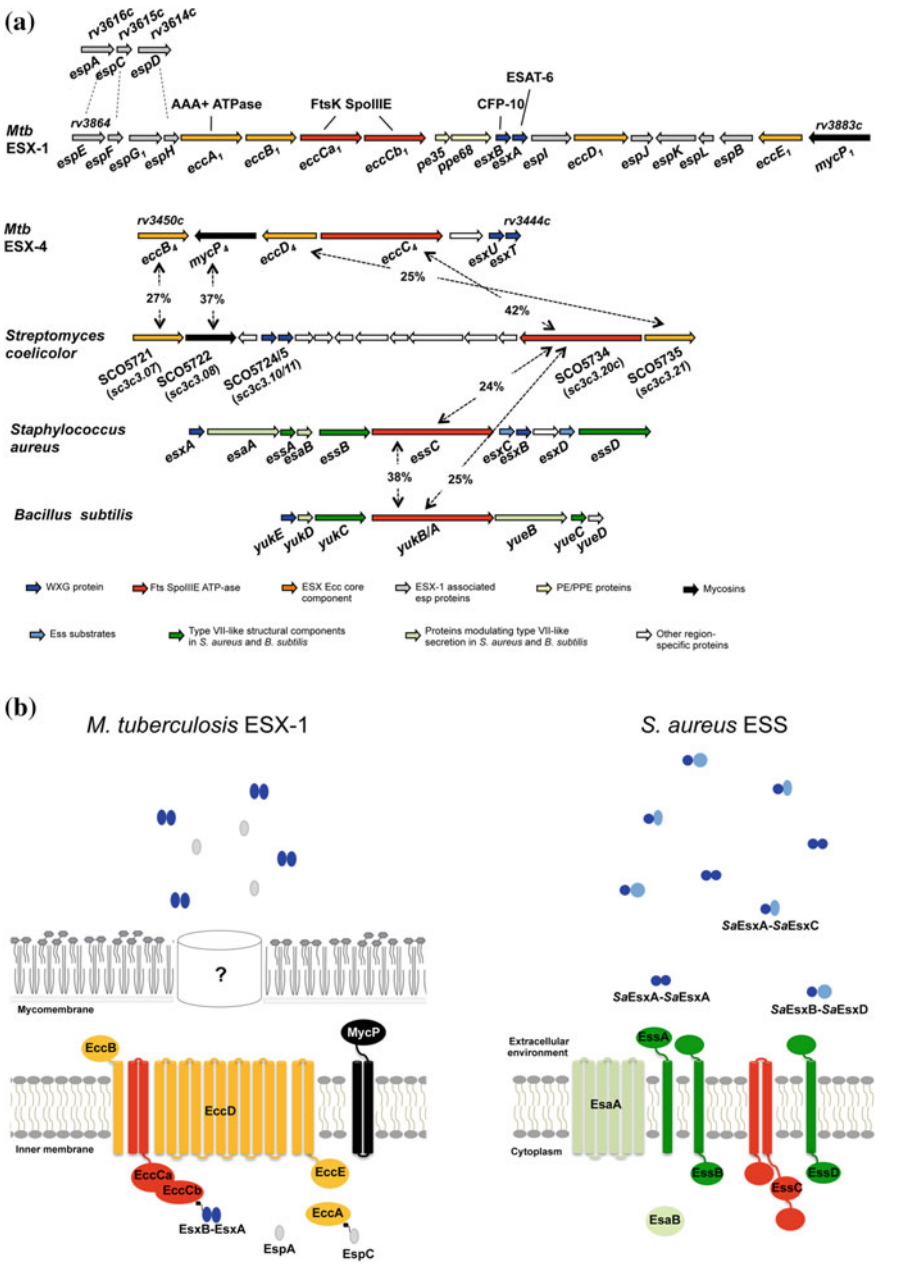


Fig. 1 **a** Genetic organization of type VII system genetic loci in *Mycobacterium tuberculosis* (Mtb) showing ESX-1 and ESX-4 clusters in comparison with type VII-like loci in *Streptomyces coelicolor*, *Staphylococcus aureus* and *Bacillus subtilis*. **b** Working model of the type VII- and type VII-like secretion machineries of *M. tuberculosis* (ESX-1) and *S. aureus* (ESS), respectively. Note that the colours of the represented proteins refer to the same colour code as used in panel A

also often harbour ATP-binding proteins belonging to the FtsK/SpoIIIE family, whose encoding genes are encoded next to those encoding WXG-100 proteins (Fig. 1) (Bitter et al. 2009a). FtsK/SpoIIIE ATPases are involved in translocation of macromolecules (proteins and DNA) across a membrane-bound channel in a wide range of cellular processes, including protein secretion, DNA conjugation, cell division and sporulation (Iyer et al. 2004). In the Gram-negative type IV secretion pathway, for example, FtsK/SpoIIIE family proteins play a key role in substrate recognition and ATP-mediated protein transport (Christie 2001). Consistently, they have been supposed to be involved in generating energy for translocation of WXG proteins/ESX substrates (Pallen 2002). Functionally active type VII-like secretion systems have been described in *Streptomyces* (Fyans et al. 2013; Akpe San Roman et al. 2010), *Staphylococcus aureus* (Burts et al. 2005), *Streptococcus agalactiae* (Shukla et al. 2010), *Listeria monocytogenes* (Way and Wilson 2005), *B. subtilis* (Baptista et al. 2013; Huppert et al. 2014) and *Bacillus anthracis* (Garufi et al. 2008).

3 Gene Clusters Encoding Type VII Secretion Systems

Typical features of the gene clusters encoding type VII secretion machineries both in *Actinobacteria* and *Firmicutes* are the presence of one or two genes encoding WXG-100 proteins, and a gene encoding an FtsK/SpoIIIE ATPase family member (Fig. 1) (Bitter et al. 2009a, b; Pallen 2002). Apart from these conserved genes, each locus includes a variable number of genus-specific genes encoding for structural or accessory system-specific components, essential for type VII- or type VII-like secretion in mycobacteria or other Gram-positive bacterial species, respectively. These differences in gene content may result in varying complexity and functionality of the corresponding secretion machineries and may account for their different roles in various biological processes and/or adaptation to specific host/environments.

3.1 The Mycobacterial ESX Loci

The *M. tuberculosis* genome harbours five highly conserved ESX gene clusters encoding type VII secretion systems, designated ESX-1 to ESX-5 (Tekaia et al. 1999; Gey Van Pittius et al. 2001). Phylogenetically, the ESX-4 locus is considered the most ancestral ESX cluster in mycobacteria, showing a relatively small number of genes and a simple gene organization relative to the other ESX clusters (Fig. 1) (Gey Van Pittius et al. 2001; Gey van Pittius et al. 2006). ESX-4 also displays similarity to the ESX-related loci identified in the genome of a wide range of *Actinobacteria* (Gey Van Pittius et al. 2001; Gey van Pittius et al. 2006). It was hypothesized that other ESX clusters might have evolved from ESX-4 through gene

duplication/diversification events and insertions of additional genes. Conversely, ESX-5 is considered the most recent ESX locus, whose evolution correlates with the differentiation of slow-growing mycobacterial species. ESX-5 orthologous loci are exclusively conserved in the genomes of mycobacterial species belonging to the slow-growing group including several human pathogens (*M. tuberculosis*, *Mycobacterium leprae* and *Mycobacterium ulcerans*) or the fish pathogen *Mycobacterium marinum*. In contrast, ESX-5-like loci are not present in the genomes of phylogenetically more distant, fast-growing mycobacterial species, which are mainly constituted by saprophytic mycobacteria.

Although the gene content may vary in different ESX loci (in terms of genes encoding chaperones, ATPases, transcription factors and substrates), a set of genes is highly conserved in all ESX clusters. Each ESX locus consists of the following: (i) a pair of *esx* genes, coding for WXG-100 proteins (ESAT-6 and CFP-10 in the case of ESX-1 system/locus, the paradigm of ESX loci); (ii) *ecc* (*esx conserved components*) genes, which encode proteins with one or more transmembrane domain(s) and ATP-binding proteins, representing core components of the secretion machinery responsible for the ATP-dependent translocation of ESX substrates; (iii) a *mycP* gene, coding for mycobacteria-specific membrane-anchored proteins belonging to the subtilisin-like serine protease family; *iv) pe* and *ppe* genes, which encode for two large classes of mycobacteria-specific proteins, named after their conserved proline–glutamic acid (PE) or proline–proline–glutamic acid (PPE) motifs at their N-termini, respectively (Bitter et al. 2009a; Tekaia et al. 1999; Cole et al. 1998). The ESX-associated *pe* and *ppe* genes are included in all ESX loci with the exception of ESX-4 and might represent the most ancestral *pe* and *ppe* genes of mycobacteria from which other members of these large gene families seem to have evolved by duplication and diversification (Gey van Pittius et al. 2006). Interestingly, selected *pe* and or *ppe* gene family members carrying characteristic GC-rich repetitive sequences (Cole et al. 1998) have also been acquired by horizontal gene transfer. This is the case of the *pe_pgrs33* (*rv1818*) gene that is present in all members of the *M. tuberculosis* complex but is missing from the supposed ancestral gene pool of tubercle bacilli, represented by the different *Mycobacterium canettii* strains (Supply et al. 2013; Boritsch et al. 2014). In addition to *ecc* genes, ESX loci may also contain genes encoding for ESX secretion-associated proteins (Esp) (Fortune et al. 2005; Bitter et al. 2009a). Several *esp* gene products are secreted via the type VII pathway and/or are involved in modulation of the ESX activity. In other cases, Esp proteins may act as chaperones in assisting the secretion of ESX substrates (as detailed below). Some Esp proteins, e.g. EspB or EspI, are part of the core ESX-1 cluster (Chen et al. 2013; Zhang et al. 2014) whereas others, such as EspA, EspC, EspD or EspR, exert fundamental roles in the regulation and secretion of ESX-1 proteins, but are not encoded in the core region (MacGurn et al. 2005; Frigui et al. 2008; Blasco et al. 2012; Pang et al. 2013; Majlessi et al. 2015). This is particularly relevant for the *espACD* locus which in contrast to the core ESX-1 region is exclusively present in the genomes of pathogenic mycobacteria, such as of *M. tuberculosis*, *M. marinum* or *M. leprae*. It is noteworthy that in these different mycobacterial species, the *espACD* region is located in non-syntenic

genomic sections (Simeone et al. 2012), which are reminiscent of genomic islands and suggest an independent acquisition of the *espACD* region in the different mycobacterial species (Majlessi et al. 2015). Future work will have to elucidate how these *espACD* loci have been interconnected with the ESX-1 secretion activity in the concerned mycobacterial species.

A new type VII locus has recently been identified in the conjugative pRAW plasmid isolated from *M. marinum* (Ummels et al. 2014). This locus, which was referred to as ESX-P1, harbours genes highly homologous to the genome-associated ESX-5 cluster, although the gene order and the gene content closely resembles to the ESX-2 locus. The ESX-P1 locus differs from all known ESX loci for the presence at the 5' terminus of two genes coding for members of the NLP60 family of peptidoglycan-associated glycosides hydrolases, which might be predictive of a novel and different function for ESX-P1. In addition to ESX-P1, the pRAW plasmid also contains genes homologous to type IV secretion system components and proteins with predicted relaxase activity. The ESX-P1-harbouring plasmid seems to be efficiently transferred among various slow-growing mycobacterial species and can also be transferred into *M. tuberculosis* under certain experimental conditions (Ummels et al. 2014). The conjugative process linked to the pRAW plasmid requires both, type VII and type IV secretion machinery components, with the ESXP-1-encoded $EccC_{P1}$ and VirB4 homologs being involved in the gene transfer process. The implication of a secretion system in conjugal gene transfer is reminiscent of some biological properties of type IV secretion systems in Gram-negative bacteria. For mycobacteria, conjugative gene transfer has also been reported. Apart from the aforementioned implication of the pRAW plasmid in conjugation within the group of selected slow-growing mycobacteria, conjugative processes have been demonstrated for *M. smegmatis.* In this fast-growing mycobacterial model organism, an involvement of the ESX-1 type VII secretion machinery in distributive conjugal transfer has been described that results in genome-wide mosaicism (Flint et al. 2004; Gray et al. 2013). ESX-1 seems to play a dual role in the process: while ESX-1 components are negative regulators of DNA transfer in donor cells (Flint et al. 2004), the ESX-1 activity is required for DNA acquisition in recipient cells (Coros et al. 2008; Gray et al. 2013). Future studies will show whether distributive conjugal transfer may also play a role in the evolution of other (myco)bacterial species.

3.2 *Type VII-Like Gene Clusters in Actinobacteria and Firmicutes*

Apart from *M. tuberculosis* and other mycobacterial species, type VII-like secretion systems are found in the genomes of other mycolic acid-containing genera (mycolata), such as *Nocardia* or *Corynebacterium*, as well as in *Streptomyces* species (Gey Van Pittius et al. 2001). The genomic loci encoding type VII-like secretion

systems in mycolata show the highest degree of homology with the ESX-4 system of *M. tuberculosis*. In the more closely related *Corynebacterium diphteriae*, an ESX-4-encoding gene cluster is located in a genomic region that shares a syntenic gene organization with *M. tuberculosis*, which suggests that ESX-4 like type VII systems might have played an important role in the evolution of selected branches of *Actinobacteria*. The genes in this region are predicted to be functional, as no deletions, frameshifts or stop codons were detected (Gey Van Pittius et al. 2001). The genome of the more distantly related *Streptomyces coelicolor* harbours an ESX locus including four out of six orthologous genes conserved in the *M. tuberculosis* ESX-4 cluster (*eccB*$_4$, *eccC*$_4$, *eccD*$_4$ and *mycP*$_4$), and other species-specific genes (Fig. 1) (Gey Van Pittius et al. 2001).

However, the simplest gene organization compared to the mycobacterial ESX loci was found in gene clusters encoding type VII-like secretion systems in *Firmicutes*. The *Ess* locus of *S. aureus* consists of 11 type VII system–associated genes (Burts et al. 2005; Anderson et al. 2013), which include *esxA*, *esxB* and *essC* genes encoding WXG-100 proteins (the *S. aureus* EsxA and EsxB variants) and a membrane-anchored FtsK/SpoIIIE-like ATPase (EssC) (Fig. 1). The *Ess* locus also contains *essA*, *essB* and *essD* genes coding for membrane-embedded proteins required for secretion of Ess substrates, as well as *esaC/esxC* and *esaD/esxD* genes, which encode small proteins that were recently identified as specific Ess substrates of staphylococci (Burts et al. 2008; Anderson et al. 2011, 2013, Chen et al. 2012). Finally, among the genes at this locus *esaB* encodes a negative regulator of Ess secretion activity (Burts et al. 2008).

WXG-100-encoding genes have also been identified in different species of the genus *Bacillus*, both in the non-pathogenic species *B. subtilis* and in the virulent species *Bacillus cereus*, *Bacillus thuringiensis* and *B. anthracis* (Baptista et al. 2013; Huppert et al. 2014). The genome of *B. subtilis* harbours two genes encoding WXG-100 proteins, *yukE* and *yfiA*, situated in distant genomic regions. *YukE* is the first gene of a cluster consisting of five annotated genes (*yukEABCD*), which is conserved in the genomes of *B. cereus*, *B. thuringiensis*, but not in *B. anthracis* (Baptista et al. 2013; Huppert et al. 2014). For some *Bacillus* strains, the split *yukA/B* gene encodes for a predicted FtsK/SpoIIIE ATPase, homologous to EccCa$_1$ and EccCb$_1$ from *M. tuberculosis* or EssC from *S. aureus*. The *yukC* and *yukD* genes encode two additional structural proteins, thought to be required for YukE secretion (Baptista et al. 2013; Huppert et al. 2014). Also, the *yueB/C* cluster, lying adjacent to the *yuk* locus in the genome of *B. subtilis*, has recently been demonstrated to encode components of the secretion apparatus involved in YukE secretion (Huppert et al. 2014). Both YueB and YueC are homologous to Ess proteins of *S. aureus* and are conserved in putative type VII-like secretion systems of other species within the *Firmicutes*, such as *Streptococcus agalactiae* or *L. monocytogenes* (Huppert et al. 2014).

Inspection of the genome of *B. anthracis* identified six genes encoding for WXG proteins. One of them, the *Ba-esxB* (BAS2036), shows similarity with the mycobacterial *esxB* and staphylococcal *esxA* and lies in a gene cluster immediately downstream the *Ba-essC* gene (BAS2035), encoding an FtsK/SpoIIIE-like ATPase

(Garufi et al. 2008). Four other WXG encoding genes, namely *Ba-esxL*, *Ba-esxQ*, *Ba-esxV* and *Ba-esxW*, are located elsewhere on the *B. anthracis* chromosome, whereas the fifth gene, *Ba-esxP,* is encoded on the virulence plasmid pOX1. All these five genes encode for proteins harbouring a large C-terminal domain that is not found in mycobacterial and staphylococcal Esx proteins and appears to be a specific signature of the *B. cereus* group (Garufi et al. 2008).

4 Type VII Secretion Machineries

Because of their relatively recent discovery, almost no structural data on type VII secretion machineries are available (Rosenberg et al. 2015; Korotkova et al. 2014, 2015; van der Woude et al. 2013). However, in silico predictions and homology studies, in combination with comparative secretome analyses from different type VII mutant strains suggest that each type VII secretion apparatus secretes multiple substrates and constitutes a complex, multi-protein machinery, consisting of several components (structural and accessory factors). Mycobacterial ESX/type VII secretion machineries are the most extensively characterized, although some data on the composition and regulation of type VII-like secretion machineries in *Firmicutes* became recently available.

4.1 Mycobacterial ESX Secretion Machineries: The Paradigm of Type VII Secretion Systems

Most of the current knowledge on ESX secretion machineries comes from the characterization of ESX-1, ESX-3 and ESX-5 systems from *M. marinum, M. tuberculosis* or *M. smegmatis* (van der Woude et al. 2013).

The different ESX secretion machines are predicted to contain an inner-membrane-bound protein complex, thought to drive ATP-dependent translocation of ESX substrates across the cytoplasmic membrane in association with a yet un-identified outer-membrane-embedded channel, which seem to manage substrate export across the mycomembrane to the extracellular environment (see below). Structural components of the membrane-anchored apparatus are three transmembrane-domain-containing proteins (EccB, EccD, EccE), a membrane-anchored ATP-binding protein (EccC) and a cytosolic ATP-binding protein belonging to the AAA+ATPase family (EccA). Based on mass spectrometry analyses and the evaluation of the apparent molecular weight of the ESX-5 complex from *M. marinum*, each ESX membrane complex seems to be composed of 6 copies of EccB, EccC and EccD components and three copies of EccE (Houben et al. 2012b). EccB, EccC and EccD proteins have been demonstrated to be required for secretion of the corresponding ESX substrates in all ESX systems characterized so far (ESX-1, ESX-3

and ESX-5) (Siegrist et al. 2014; Di Luca et al. 2012; Brodin et al. 2006; MacGurn et al. 2005; Guinn et al. 2004; Hsu et al. 2003). Among these components, EccD is a highly hydrophobic protein, containing 11 predicted transmembrane domains. This protein is predicted to build the channel for transport of substrates across the cytoplasmic membrane (Bitter et al. 2009a). EccC proteins are thus key components of the ESX core complex and are encoded either by a single gene, or, as in the case for ESX-1, by two adjacent genes (*eccCa*$_1$ and *eccCb*$_1$), whose products are predicted to form a single functional unit (Cole et al. 1998; Bitter et al. 2009a). EccC proteins have functional ATP-binding domains, which are homologous to the FtsK/SpoIIIE family of ATPases. In contrast to the FtsK/SpoIIIE ATPases characterized so far, which contain only one ATPase domain, the ESX-encoded EccC component has a unique multi-domain structure consisting of three functionally distinct FtsK/SpoIIIE-like ATP-binding domains, supposed to play distinct roles in substrate translocation and complex formation (Ramsdell et al. 2015; van der Woude et al. 2013). While the most N-terminally situated ATPase domain seems to be involved in ATP hydrolysis, the second and third ATPase/nucleotide binding domains seem to be required for the assembly of a functional ESX secretion machinery (Ramsdell et al. 2015). The ESX-1-encoded EccC component interacts with the ESX-1 substrates EsxB, via the last seven C-terminal residues of EsxB (Stanley et al. 2003; Champion et al. 2006). This C-terminal short domain represents a signal sequence that is linked to secretion via ESX-1 (Champion et al. 2006). The involvement of an ATPase in the contact between ESX substrates and their corresponding secretion apparatus resembles type IV secretion systems, where a FtsK/SpoIIIE ATPase recognizes an unstructured C-terminal sequence and hence directs substrates across the cytoplasmic membrane. It has recently been observed that the binding of EsxB via its signal sequence at the C-terminal domain to EccC$_1$ induces the activation of EccC$_1$ by stimulating its multimerization, thus resulting in increased ATPase activity (Rosenberg et al. 2015). These findings suggest a model in which ESX substrates (e.g. EsxB) might modulate the coordinated release of substrates from the bacterium (Rosenberg et al. 2015). Apart from the FtsK/SpoIIIE ATPase, ESX secretion machineries also include a cytosolic ATP-binding protein (EccA) belonging to the AAA+ family (ATPase associated with various cellular activities). Proteins belonging to this family play a role in various cellular processes: assembly and disassembly of protein complexes, protein degradation and signal transduction. Although some EccA protein family members have been demonstrated to be involved in ESX activity (EccA$_1$ and EccA$_5$ are implicated ESX-1 and ESX-5 secretion, respectively, while the ESX-3-encoded EccA$_3$ is required for viability of *M. tuberculosis*) (Siegrist et al. 2014; Bottai et al. 2012; Brodin et al. 2006), the specific function of AAA+ ATPases in ESX-mediated secretion remains unknown. Structural data on the ESX-1-encoded EccA$_1$ revealed that the C-terminal part of EccA$_1$ contains an oligomerization domain, which might induce the formation of a hexamer. It was also proposed that EsxA$_1$ possesses ATPase activity, possibly regulated by the N-terminal part of the protein (Wagner et al. 2013). The finding that EccA$_1$ interacts with EspC and EspF$_1$ in two-hybrid studies in vitro (Champion et al. 2009) suggests that these proteins might be involved in the

assembly/disassembly of secreted substrate molecules, or alternatively, secreted substrates and their chaperones.

In addition to the ESX secretion components mentioned above, ESX activity also involves mycosins ($MycP_1$–$MycP_5$). These proteins correspond to specific subtilisin-like serine proteases (Brown et al. 2000; Dave et al. 2002) and are predicted to be anchored in the cytoplasmic membrane via the C-terminal transmembrane domain, with the active site exposed to the extracytoplasmic space (Sonnhammer et al. 1998; Gardy et al. 2005). This protein topology suggests a role for these proteins in proteolytic digestion of ESX substrates. To date, proteolytic activity has been demonstrated for $MycP_1$, the ESX-1 associated mycosin. $MycP_1$ was shown to digest EspB at its C-terminal domain, thereby modulating ESX-1 activity (Ohol et al. 2010). More recently, the ESX-3-encoded mycosin $MycP_3$ has been demonstrated to be involved in stabilization or/and optimal secretion of the EsxG–EsxH complex in *M. smegmatis* (see below) (Siegrist et al. 2014).

The actual way of how ESX substrates are transported across the mycobacterial outer membrane to the extracellular environment remains unknown. In Gram-negative bacteria, protein secretion may occur by a "one-step" mechanism, in which a protein channel spanning both the inner and the outer membrane mediates the substrate translocation across the diderm cell envelope (type I, type III and type VI secretion pathways). Alternatively, this transport can also occur by a "two-step" mechanism, in which the Sec or TAT machineries are responsible for substrate transport across the cytoplasmic membrane, and a different, more specialized secretion apparatus is involved in subsequent substrate translocation across the outer membrane (type II and type IV secretion pathways). In mycobacteria, the absence of a Sec or TAT signal sequence in ESX substrates and the presence of transmembrane domains in all ESX core components suggest that type VII protein transport across—at least the inner membrane—occurs by the ESX systems. It has been speculated that some of the Ecc-membrane-bound components, such as EccE or EccC, might span across the mycomembrane (Houben et al. 2012b), favouring a one-step secretion mechanisms of ESX substrates. However, as detailed structural data of the type VII apparatus are missing, it is also possible that proteins with yet undefined roles might mediate the ESX protein translocation to the extracellular environment.

4.2 ESX Substrates

In addition to the above-described Esx/WXG-100 proteins representing the first known substrates of mycobacterial ESX/type VII secretion systems, PE/PPE protein family members and a number of Esp proteins are also transported by ESX machineries.

Although Esx, PE/PPE and Esp proteins display very low similarity in their primary amino acid sequence, all these classes of ESX substrates share several typical features: (a) the ability to be secreted as heterodimers (for Esx and PE/PPE

proteins); (b) the codependent secretion where deletion/absence of one substrate affects the secretion of the cognate ESX substrate; (c) the presence of a C-terminal signal domain responsible for protein targeting to the type VII secretion pathway. The resolved structure of the ESAT-6-CFP-10 heterodimer, representative of the Esx protein complexes, revealed the presence of two helices on each Esx protein connected by a turn formed by the WXG motif. In this complex, the two proteins are in antiparallel orientation and have flexible N- and C-terminal tails (Renshaw et al. 2005). A similar structure has been reported for the PE25-PPE41 heterodimer, representative of the PE/PPE protein complexes: the domains of the proteins that are responsible for interaction form a four-helix bundle in an antiparallel fashion, and a WXG motif is located in the turn between the two helices of the PPE protein (Strong et al. 2006). Both, EsxB and PE25 proteins in the ESAT-6-CFP-10 and PE25-PPE41 heterodimers, respectively, have an identical C-terminal short motif, which is crucial for secretion (Champion et al. 2006; Renshaw et al. 2005; Strong et al. 2006). Although ESX substrates lack a classical N-terminal signal sequence, they share a conserved C-terminal secretion domain which interacts with a conserved ESX structural component (e.g. the short C-terminal domain of CFP-10 is recognized by the FtsK/SpoIIIE ATPase EccC), thus targeting the substrate and its protein partner to the corresponding secretion machinery. More recently, a conserved C-terminal YxxxD/E motif was identified as the general sequence required for targeting proteins to type VII secretion pathway (Daleke et al. 2012). This YxxxD/E domain seems also to be located adjacent to the helix-turn-helix motif of several PE proteins and EspB (Daleke et al. 2012). Moreover, an additional hydrophobic residue located seven positions downstream the YxxxD/E motif (the Y subdomain) was proposed to be involved in ESX substrate recognition (Poulsen et al. 2014). However, the signal sequences that specifically target each ESX substrate to the corresponding ESX secretion apparatus are still unknown. In this respect, it has been hypothesized that selected EspG proteins (e.g. the *M. marinum* ESX-1- and ESX-5-encoded EspG) might act as chaperons in directing type VII substrates to the corresponding ESX machineries (see below).

Apart from Esx and PE/PPE proteins, several ESX-1-related Esp proteins such as EspA, EspC, EspE, EspF, EspJ, EspK, EspB have been proposed as ESX-1 substrates (Champion et al. 2009; McLaughlin et al. 2007; Sani et al. 2010; Carlsson et al. 2009). Several of these Esp proteins carry the C-terminal YxxxD/E motif, conserved in type VII substrates or substrate complexes (Daleke et al. 2012). Secretion of some Esp proteins (e.g. EspA and EspC) or Esp protein activity (e.g. EspB, EspD) impact on ESAT-6 secretion and in turn, on mycobacterial virulence. The mutually codependent secretion of EspA and EspC with ESAT-6 and CFP-10 represents a key element of the regulation of the ESX-1 activity (see below). Differently, EspB secretion via ESX-1 requires neither EspA/EspC/EspD expression nor ESAT-6/CFP-10 secretion (Chen et al. 2013). However, the proteolytic digestion of EspB by $MycP_1$ modulates the amount of proteins secreted by ESX-1. EspB is able to bind bioactive host phospholipids such as phosphatidic acid and phosphatidylserine, thus interfering with eukaryotic cell signalling (Chen et al. 2013). This feature reveals an additional, ESAT-6-independent impact of EspB on

mycobacterial virulence. Recent data on the resolved crystal structure of EspB demonstrated that the protein can adopt a PE/PPE-like fold and oligomerize to form a barrel-shaped structure with heptameric symmetry, thus suggesting the possibility that EspB might be part of a structural subunit of a cell-wall-associated complex (Solomonson et al. 2015; Korotkova et al. 2015).

Other ESX-1-associated Esp proteins (e.g. EspD) do not seem to be ESX-1 substrates, but are involved in modulation of the ESX-1 activity as they seem to influence EspA and EspC stabilization and secretion (Chen et al. 2013). Moreover, a species-specific impact on ESX-1 secretion has been reported for some Esp proteins (EspF and $EspG_1$) (Bottai et al. 2011; Gao et al. 2004; Converse and Cox 2005), suggesting that the function of similar proteins in related bacterial species might have changed during evolution due to adaptation processes. EspF and $EspG_1$ have been reported to be implicated in ESAT-6 and CFP-10 secretion in *M. smegmatis* and *M. marinum*, whereas they are dispensable for secretion, post-transcriptional modification, and immunogenicity of ESAT-6 in *M. tuberculosis*. However, both, EspF and $EspG_1$, are required for full virulence of *M. tuberculosis* (Bottai et al. 2011). Although the function of $EspG_1$ is still unknown, the impact of the protein on virulence of *M. tuberculosis* might be related to specific interactions with other ESX components, such as PE/PPE proteins (Teutschbein et al. 2009). Deletion of *espG*$_1$ from *M. tuberculosis* results indeed in lower amounts of PPE68 in cell lysates, suggesting a possible role for $EspG_1$ in folding or stability of PPE68 (Bottai et al. 2011), in accordance with the proposed chaperone activity of EspG-like proteins encoded by other ESX systems (Ekiert and Cox 2014; Korotkova et al. 2014).

EspI is an ESX-1-associated protein for which a putative function as negative regulator of ESX-1 activity has been postulated. Although dispensable for ESX-1-mediated secretion and full virulence of *M. tuberculosis*, EspI seems to be involved in the shutdown of ESX-1 secretion activity in case of ATP depletion (Zhang et al. 2014).

4.3 Regulation of ESX Secretion in Mycobacteria

Because of the relatively recent discovery of type VII systems, the regulation of ESX activity is a rather unexplored field of research on mycobacterial type VII secretion. As opposed to type III and type IV secretion in Gram-negative bacteria, where substrate secretion occurs after host–cell contact, ESX/type VII secretion occurs also in in vitro growing mycobacteria. This feature suggests that mycobacterial ESX secretion systems might play more general roles in the physiology of the bacteria, whereby their implication in virulence could have evolved as an adaptation to the hostile intracellular environment. Recent experimental data provide evidence that ESX-mediated secretion and/or ESX gene cluster expression is controlled by different global transcriptional regulators. These include the putative transcriptional repressors Lrs2 and CRP (Rickman et al. 2005; Gordon et al.

2010) and the activators PhoP and EspR (Frigui et al. 2008; Blasco et al. 2012), which regulate ESX-1 secretion; the iron-dependent regulator IdeR (Rodriguez et al. 2002) and the zinc uptake regulator Zur (Maciag et al. 2007), which control the expression of the ESX-3 gene cluster; the global transcriptional regulator WhiB5 and the alternative sigma factor SigM which regulate genes at ESX-2 and ESX-4 loci (Casonato et al. 2012; Raman et al. 2006; Agarwal et al. 2007).

The regulation of ESX-1-mediated secretion of ESAT-6 and CFP-10 mainly involves the *espACD* locus due to the codependent secretion of these proteins with EspA and EspC. According to the currently proposed model, PhoP/R, a major two-component regulatory system of *M. tuberculosis* (Walters et al. 2006), influences the expression of the nucleoid-associated protein/regulator EspR (Blasco et al. 2012; Raghavan et al. 2008), which controls the *espACD* operon (Blasco et al. 2012; Hunt et al. 2012). The link between a functional PhoP protein and ESAT-6 secretion was first demonstrated in *M. tuberculosis* H37Ra (Frigui et al. 2008), the paradigm attenuated *M. tuberculosis* strain, which carries a point mutation in its *phoP* gene that interferes with the DNA binding capacities of PhoP (Wang et al. 2007). Complementation of H37Ra with a wild-type copy of the *phoP* locus restored ESAT-6 secretion and partially increased the virulence of the recombinant H37Ra::*phoP* strain (Frigui et al. 2008). Moreover, a second two-component signal transduction system named MprA/B (for mycobacterial persistence regulator) was also found to modulate ESX-1 functions via regulation of the *espACD* locus (Pang et al. 2013). Interestingly, regulation of the *espACD* locus by two-component regulators may also be bypassed, as observed in a lineage of tubercle bacilli that have the region of difference 8 (RD8) deleted, leading to ESAT-6 secretion despite PhoP/R mutations (Gonzalo-Asensio et al. 2014). The RD8 deletion is located just upstream of the *espACD* operon in *Mycobacterium africanum* lineage 6 strains, as well as in animal-adapted members of the *M. tuberculosis* complex, such as *M. microti* or *M. bovis* (Boritsch et al. 2014).

However, apart from regulation of ESAT-6 secretion via EspA/C expression, regulation of ESX-1 genes may also occur directly at the ESX-1 locus by WhiB6, a member of large WhiB protein regulatory family (Solans et al. 2014). A point mutation in the putative *whiB6* promoter region present in the reference strain *M. tuberculosis* H37Rv determines the formation of a stem-loop structure in the binding region of the regulator PhoP, resulting in lower expression of ESAT-6 in *M. tuberculosis* H37Rv relative to *M. tuberculosis* clinical isolates and other reference strains (e.g. CDC1551) (Solans et al. 2014).

4.4 *Type VII-like Secretion Machineries in Non-mycobacterial Species*

The Ess secretion system of *S. aureus* is one of the best-characterized type VII secretion systems in *Firmicutes*. It consists of four structural components (EssA, EssB, EssC and EssD), all required for transport of Ess substrates across the

staphylococcal cell envelope and two accessory proteins (EsaA and EsaB) (Burts et al. 2005, 2008, Anderson et al. 2013, 2011, Chen et al. 2012, Kneuper et al. 2014). Among the Ess core components, EssC is a membrane-bound FtsK/SpoIIIE ATPase, which represents a hallmark for type VII secretion machineries. Moreover, the locus also encodes phylum-specific EssA, EssB and EssD, which are membrane-anchored proteins that are not conserved in the mycobacterial ESX systems (Fig. 1) (Burts et al. 2005, 2008, Anderson et al. 2011, 2013). The Ess secretion machineries also include EsaB, a small cytosolic protein, for which a role in regulation of the expression and production of a specific subset of Ess substrates (e.g. EsxC/EsaC, see below) has been demonstrated (Burts et al. 2008, Kneuper et al. 2014). Secretome analysis of a panel of *S. aureus* Ess mutant strains in two different genetic backgrounds (*S. aureus* USA300 and *S. aureus* Newman clinical isolates) identified four Ess substrates: the canonical *Sa*EsxA and *Sa*EsxB (Burts et al. 2005), as well as the non-canonical type VII substrates *Sa*EsxC and *Sa*EsxD (Anderson et al. 2011, 2013). *S. aureus* EsxA and EsxB share some features with the *M. tuberculosis* ESAT-6-CFP-10, including the presence of a WXG motif (located in the middle of a 100-aa-long protein) and the codependent secretion (genetic deletion of *Sa*EsxA or *Sa*EsxB impairs the secretion of the related WXG protein). However, unlike ESAT-6 and CFP-10 from *M. tuberculosis*, *Sa*EsxA and *Sa*EsxB from *S. aureus* do not seem to interact (Anderson et al. 2013). Instead, *Sa*EsxA dimerizes with itself or associates with *Sa*EsxC, while *Sa*EsxB interacts with *Sa*EsxD (Anderson et al. 2013). The recently resolved crystal structure of the EsxA-EsxA homodimer reveals that each EsxA subunit folds into an elongated cylinder of two helices bent by a hairpin, carrying the WXG motif (Sundaramoorthy et al. 2008), in a structure that is reminiscent of the *M. tuberculosis* ESAT-6-CFP-10 or EsxG-EsxH heterodimers secreted by ESX-1 and ESX-3 systems, respectively.

In contrast to *Sa*EsxA and *Sa*EsxB, staphylococcal EsxC and EsxD do not share obvious sequence features with WXG proteins, nor with other reported ESX substrates, such as PE/PPE or Esp protein family members (Anderson et al. 2013). However, *Sa*EsxD contains a C-terminal YxxxD/E motif (Anderson et al. 2013) identical to that involved in targeting substrates to the type VII secretion pathways in mycobacteria. It was found that *Sa*EsxD can play a role in the interplay among Ess substrates and Ess core components. Deletion of the entire *Sa*EsxD, as well as deletion of its C-terminal domain, abolished the production of *Sa*EsxB and affects the secretion of *Sa*EsxA and *Sa*EsxB. These effects on Ess secretion seem to be related to a direct effect of the absence of *Sa*EsxD on EssD stability (Anderson et al. 2013). This suggests a model in which *Sa*EsxD might contribute to the stability of EssD via the last 6 amino acids at the C-terminal domain, thus influencing the secretion of other Ess substrates. Recently, some data on regulation of Ess activity became available: similarly to the mycobacterial ESX loci, whose expression is under the control of different global transcriptional regulators, the Ess locus is negatively regulated by the response regulator SaeR (Anderson et al. 2013). SaeR is the transcriptional regulator of the two-component regulation system SaeR/SaeS, which modulates the secretion of a plethora of virulence factors in *S. aureus*

(Giraudo et al. 1997; Cheung et al. 2004; Novick 2003). Transposon insertion in SaeR or SaeS results in an increased production of both Ess core components (EssD) and Ess-dependent substrates (*Sa*EsxA and *Sa*EsxB) (Anderson et al. 2013). The SaeR-/SaeS-dependent Ess expression might account for the increased Ess secretion activity observed in the *S. aureus* USA300 strain relative to that observed for the Newman strain (Anderson et al. 2013), where *Sa*EsxA and *Sa*EsxB are produced and secreted at low levels, and *Sa*EsxC is produced only after *Sa*EsxB deletion (Burts et al. 2005, 2008). A point mutation in the SaeS encoding gene in the Newman strain results indeed in a constitutively active variant of the signalling kinase SaeS (Adhikari and Novick 2008).

The simplest protein composition was predicted for the Yuk/Yue type VII-like system in *B. subtilis*. In addition to the FtsK/SpoIIIE-like ATPase YukA/B, other components of the secretion machinery are the ubiquitin-like YukC and YukD proteins, and the membrane-bound YueB/YueC proteins, homologous to the Ess components in *S. aureus* and conserved in type VII-like secretion systems of other *Firmicutes* (*Streptococcus agalactiae* and *L. monocytogenes*) (Baptista et al. 2013; Huppert et al. 2014). Interestingly, YueB is a membrane receptor essential for phage infection (Huppert et al. 2014). The WXG-100 protein member YukE is the only type VII-like substrate identified so far (Huppert et al. 2014). YukE accumulates in stationary growth phase culture supernatants as homodimer (Sysoeva et al. 2014), whose predicted structure consists in a putative helix-loop-helix fold, with the WXG domain lying in the loop and the two helices in an antiparallel configuration (Huppert et al. 2014). Mutagenesis studies combined with cross-linking experiments revealed that the C-terminal residues of YukE are important for secretion, thus providing further evidence for a general mode of recognition of type VII-like and ESX substrates in *Firmicutes* and *Actinobacteria,* respectively (Sysoeva et al. 2014). However, the tryptophan and glycine residues of the WXG motif of YukE seem also to be required for an efficient translocation of the protein homodimer towards the outside of the cell (Sysoeva et al. 2014). Based on structural data, YukE homodimers contain two sites composed by the C terminus and the WXG turn, but only one intact bipartite site seems to be required for protein export. These features suggest that substrate secretion via the type VII-like pathway in *B. subtilis* requires a composite, bipartite signal formed by two folded YukE polypeptides (Sysoeva et al. 2014).

Yuk/Yue expression in *B. subtilis* is regulated by the two-component system DegU–DegS (Ogura et al. 2001; Mader et al. 2002; Kobayashi 2007), which controls several post-exponential processes (genetic competence, biofilm formation and cell motility) (Murray et al. 2009; Hsueh et al. 2011; Lopez and Kolter 2010). Stable production of YukE requires high levels of phosphorylated transcriptional regulator DegU (Baptista et al. 2013), whose levels increase in cells upon transition to the stationary growth phase. This could explain the reason why no Ess activity could be demonstrated in the *B. subtilis* strain 168, which is impaired in DegU-P-dependent processes (Baptista et al. 2013).

5 Type VII Secretion Systems in Bacterial Virulence and Pathogenicity

The biological function of several type VII-secreted substrates and effector molecules is still unknown. Likewise, it remains unclear whether type VII secretion machineries might share conserved functions. It is well established that mycobacterial ESX secretion systems (ESX-1, ESX-3 and ESX-5) are virulence determinants in *M. tuberculosis* and other pathogenic mycobacteria (ESX-1 and ESX-5), or play crucial roles in essential metabolic pathways (ESX-3) (Bottai et al. 2015; Bottai et al. 2014; Majlessi et al. 2015). Similarly, type VII-like secretion systems have a strong impact on bacterial virulence and pathogenicity in *S. aureus* (Korea et al. 2014; Burts et al. 2005) and *B. anthracis* (Garufi et al. 2008). However, some other bacterial pathogens are known, whose type VII-like secretion systems do not seem to impact on virulence, as this is, for example, the case for *L. monocytogenes* (Way and Wilson 2005). Likewise, type VII-like secretion is also not required for the virulence of the plant pathogen *Streptomyces scabies* (Fyans et al. 2013), or of *S. coelicolor*, where the type VII system plays a role in modulation of sporulation and development (Akpe San Roman et al. 2010).

5.1 The Mycobacterial ESX-1, ESX-3 and ESX-5 Secretion Systems

The ESX-1 system impacts on the ability of *M. tuberculosis* and other pathogenic mycobacteria to establish infection due to its role as modulator of the mycobacterial trafficking in phagocytic host cells (macrophages and dendritic cells). Several cellular events have been reported to be related to the activity of a functional ESX-1 system: inhibition of phagosomal maturation and acidification, cytosolic access, autophagy, host cell-death, and modulation of the inflammatory response. ESX-1-proficient mycobacterial species *M. tuberculosis*, *M. leprae* and *M marinum*, as well as recombinant BCG::ESX-1 variants were shown to access the host cytosol of infected phagocytic cells, while strains with an interrupted ESX-1 system remained enclosed in the phagovacuole (Stamm et al. 2003; van der Wel et al. 2007; Simeone et al. 2012; Houben et al. 2012a; Simeone et al. 2015). Translocation into the cytosol plays a key role not only in cultured phagocytic cells, but was also demonstrated for a murine infection model (Simeone et al. 2015). Such ESX-1-mediated access to the cytosolic compartment seems to be related to the ability of ESAT-6 and potential, other ESX-1-secreted effector molecules to interact with biomembranes under specific conditions and induce membrane lysis (de Jonge et al. 2007; De Leon et al. 2012). The finding that recombinant ESX-1 strains expressing mutated variants of ESAT-6 are unable to induce phagosomal rupture despite secretion of the protein further supports a role for ESAT-6 in this process (Simeone et al. 2012; Houben et al. 2012a). The ability to lyse vacuolar

membranes, thereby allowing pathogenic mycobacteria and/or bacterial components to gain access to the cytosolic compartment of the host cell, might account for several features which characterize the pathogenic potential of *M. tuberculosis* and other ESX-1-proficient mycobacterial species: cell-to-cell spread (Guinn et al. 2004; Hsu et al. 2003), autophagy induction and/or impairment (Watson et al. 2012; Romagnoli et al. 2012), and access of mycobacterial proteins to the class I-processing machinery contained in the proteasome, with impact on NLRP3 inflammasome activation (Mishra et al. 2010; Wong and Jacobs 2011; Dorhoi et al. 2012), type I interferon (type I-IFN) responses (Stanley et al. 2007) and induction of increased CD8+ T-cell responses (Ryan et al. 2009). Recent studies demonstrated a role of cytosolic *M. tuberculosis* DNA in the induction of the synthesis of type I-IFN, via the induction of the cGAMP signalling cascade, both in human and murine macrophages (Wassermann et al. 2015; Watson et al. 2015; Collins et al. 2015). This DNA-dependent induction was suggested to be linked to ESX-1-mediated access of the pathogen to the host cytosol (Majlessi and Brosch 2015).

In addition to ESX-1, ESX-5 is a key virulence determinant of pathogenic mycobacteria, due to its role as specialized secretion system devoted to the export of PE and PPE proteins. Apart from EsxM (the ESAT-6 homolog encoded at the ESX-5 locus), a number of PE/PPE proteins have been identified as ESX-5 substrates, both in *M. tuberculosis* and in *M. marinum*. At first, the representative PE/PPE proteins PE25-PPE41 (Abdallah et al. 2006, 2009; Bottai et al. 2012) were found to be ESX-5 substrates, similar to LipY (Daleke et al. 2011), which represents a mycobacterial lipase that is involved in degradation of long chain triacylglycerols during late phases of infection (Deb et al. 2006). Other putative ESX-5 substrates are the members of PE-PGRS and PPE-MPTR subfamilies (Abdallah et al. 2009; Houben et al. 2012b). These phylogenetically most recent subclasses of the PE and PPE proteins seem to have emerged from ancestral PE and PPE proteins encoded at the ESX-5 locus (Gey van Pittius et al. 2006).

Although the biological function of most of the PE/PPE proteins remains unknown, some of them have been suggested to play a role in mycobacterial virulence. They are involved in mycobacterial growth in macrophages, efficiency of phagocytosis, inhibition of phagosome maturation, and virulence in the mouse infection model (Ramakrishnan et al. 2000; Brennan et al. 2001; Sampson et al. 2001; Li et al. 2005; Goldstone et al. 2009; Brodin et al. 2010; Dong et al. 2012; Iantomasi et al. 2012; Bottai et al. 2012). PE/PPE proteins have high immunogenic potential and represent a rich source of B- and T-cell epitopes (Copin et al. 2014). Their polymorphic nature, in combination with their predicted localization at the mycobacterial surface (Sampson et al. 2001; Banu et al. 2002; Cascioferro et al. 2007; Song et al. 2008; Chaturvedi et al. 2010) and the large variability in expression at different stages of the infection (Voskuil et al. 2004) suggested a role for PE/PPE proteins as a source of antigenic variation (Delogu and Brennan 2001; Bottai and Brosch 2009; Cole et al. 1998). However, it was recently reported that the T-cell epitopes were located within the conserved N-terminal parts of the PE/PPE proteins and not in the variable PGRS and MPTR sections of the proteins,

which questions the previous hypothesis (Copin et al. 2014) and suggests that the variable parts might have different, possibly structural, roles for the mycobacteria. Indeed, when ESX-5-associated PE and PPE proteins were investigated, which harbour mainly the conserved N-terminal part of PE/PPE proteins, strong, specific T-cell responses were recorded (Sayes et al. 2012). The responses were directed against the ESX-5-associated PE/PPE proteins (PPE25, PE18, PPE26, PPE27, PE19), but also against a number of non-ESX-5-encoded homologs (Sayes et al. 2012). These responses were completely abolished in mutant strains where the ESX-5 secretion system was non-functional (Sayes et al. 2012). These findings further confirm a role for ESX-5 in the release of PE and PPE proteins from the bacterial cell during infection and emphasize the impact of ESX-5 as key modulator of the adaptive antimycobacterial host immune response against PE/PPE proteins.

A functionally active ESX-5 system is essential for full virulence of pathogenic mycobacteria. In *M. tuberculosis*, ESX-5 inactivation via disruption of single components of the ESX-5 secretion machineries (e.g. the predicted transmembrane channel $EccD_5$) caused strong attenuation of the corresponding mutant, which was neither able to replicate in murine macrophages nor in severely combined immunodeficient (SCID) mice (Bottai et al. 2012). Furthermore, an intact ESX-5 system is required for in vitro growth of several mycobacterial species (*M. tuberculosis*, BCG and *M. marinum*) (Di Luca et al. 2012; Ates et al. 2015). Deletion of large portions of the *M. tuberculosis* ESX-5 locus (e.g. the $eccB_5$-$eccC_5$ operon that encodes key building blocks of the ESX-5-membrane-bound protein complex) results in the loss of viability of tubercle bacilli (Di Luca et al. 2012). Moreover, a putative role for certain duplicated regions of the ESX-5 locus was recently suggested, whereby knockout mutants of the so-called ESX-5a region, comprising *esxI*, *esxJ*, *ppe15* and *pe8, of M. marinum* and/or *M. tuberculosis* showed differences in secretion of selected PE/PPE proteins and some other proteins that were not members of these protein families (Shah et al. 2015). The impact of ESX-5 on *M. tuberculosis* viability and virulence seems to be related to its role in transport of PE/PPE proteins as well as its involvement in maintaining the cell envelope stability and functionality. ESX-5 disruption results in extensive damage of the mycobacterial cell envelope, as revealed by increased sensitivity of ESX-5 mutants to detergents and hydrophilic antibiotics to which mycobacteria are naturally resistant (Bottai et al. 2012). The detrimental effect of ESX-5 disruption on cell envelope functionality and mycobacterial viability can be restored by increasing the permeability of the mycomembrane, by altering its lipid composition or by introducing the heterologous outer-membrane-associated mycobacterial porin MspA (Ates et al. 2015). The existence of a functional link between ESX-mediated secretion and cell-wall biogenesis is supported by results of ChIP-on-chip analyses, which demonstrated that genes at the ESX-2 and ESX-5 loci, as well as a number of genes encoding enzymes involved in cell-wall biosynthesis were regulated by EspR (Blasco et al. 2012).

Some data are now available on the functional role of ESX-3. The finding that the expression of the ESX-3 locus is regulated by the iron-dependent transcriptional repressor IdeR (Rodriguez et al. 2002) and the zinc uptake repressor Zur (Maciag

et al. 2007) and is up-regulated in iron/zinc limiting conditions provides evidence for the implication of this secretion system in homeostasis of ferric and zinc metal ions (Siegrist et al. 2009, 2014, Serafini et al. 2009). Consistent with its involvement in fundamental metabolic pathways, ESX-3 is highly conserved in the genome of all mycobacterial species. In *M. tuberculosis*, where the uptake of ferric ions exclusively occurs via the mycobactin-mediated siderophore pathway, ESX-3 is essential for viability (Serafini et al. 2009). In contrast, in *M. smegmatis*, where the iron uptake also occurs by the alternate exochelin pathway, ESX-3 is not essential for in vitro growth (Siegrist et al. 2014). Selective inactivation of different *M. smegmatis* ESX-3 genes (*eccC*$_3$, *eccD*$_3$ and *espG*$_3$) confirmed the implication of the ESX-3 secretion system in mycobactine-mediated iron acquisition. EsxG and EsxH form a characteristic heterodimer resembling the ESAT-6-CFP-10 complex (Ilghari et al. 2011), whose secretion is enhanced in iron-limiting conditions (Siegrist et al. 2014). NMR spectroscopy of the *M. tuberculosis* EsxG–EsxH complex revealed the presence of a specific Zn^{2+} binding site in EsxH (Ilghari et al. 2011). The Zn^{2+} binding site is specific for EsxH proteins of mycobacterial species belonging to the *M. tuberculosis* complex, and it is not conserved in EsxH paralogs from other mycobacterial species (*M. smegmatis*, *M. marinum* and *M. ulcerans*), nor in other Esx proteins. Although the functional role of this site is still unknown, the Zn^{2+} binding site might regulate the stability of the EsxG–EsxH complex, its interactions with other protein partners, or might be implicated in zinc uptake.

Apart from the functional role in iron uptake, a role of ESX-3 in virulence of mycobacteria has been recently reported. The EsxG–EsxH complex was found as being involved in inhibition of phagosome maturation. This process involves the disruption of the host endosomal sorting complex required for transport (ESCRT) responsible to the delivery of *M. tuberculosis*-loaded phagosomes to the lysosomes (Mehra et al. 2013).

5.2 *Ess Secretion System in Virulence of* S. Aureus

Several observations indicate that Ess plays an important role in virulence of *S. aureus*. *Sa*EsxA and *Sa*EsxB are required for *S. aureus* replication in organs and tissues of infected mice (Burts et al. 2005). Moreover, Ess secretion activity is also required for the establishment of staphylococcal abscesses wherein the pathogen can persist and evade the host immune response (Burts et al. 2005, 2008; Anderson et al. 2011). Inactivation of the Ess pathway by the disruption of key components of the Ess secretion machinery (EssC) as well as the deletion of *esxA* and *esxB* causes a significant reduction in the ability of *S. aureus* to establish kidney or liver abscesses in a murine model of staphylococcal blood-borne dissemination and abscess formation (Burts et al. 2005, 2008; Anderson et al. 2011). Other Ess substrates, such as the staphylococcal-specific EsaC protein, although being dispensable for the establishment of acute infections, were required for the formation of persistent infection in animal models (Burts et al. 2008). Moreover, the Ess

secretion system has recently been demonstrated to be required for nasal colonization and virulence in a murine lung pneumonia model, using two different *S. aureus* strains (RN6390 and COL) for infection (Kneuper et al. 2014). More insights on the cellular mechanisms responsible for the impact of EsxA in virulence of *S. aureus* have recently been provided (Korea et al. 2014). Although *S. aureus* is primarily an extracellular pathogen, whose presence in intracellular environment during in vivo infections remains unclear, *S. aureus* was recently reported to be able to invade and replicate in several non-phagocytic cells (Garzoni and Kelley 2009; Clement et al. 2005; Sachse et al. 2010). It has also been proposed that a transient intracellular lifestyle might potentially provide protection against exposure to antibiotics and the host immune response and might represent a favourable environment for the formation of resistant variants (Tuchscherr et al. 2011; Fraunholz and Sinha 2012). *Sa*EsxA is able to interfere with the host cell apoptotic pathways in human epithelial cells, thus affecting bacterial survival and mediating the release of *S. aureus* from the host cells (Korea et al. 2014).

5.3 *The ESAT-like Secretion System in Virulence of Bacillus Anthracis*

B. anthracis encodes six WXG proteins, referred to as EsxB, EsxL, EsxP, EsxQ, EsxV and EsxW. EsxB and EsxW are secreted into the extracellular environment during growth in liquid medium, although the secretion process seems to not require the *Ba*EssC ATPase (Garufi et al. 2008). In contrast, no expression was found for other Esx-like proteins under laboratory growth conditions (Garufi et al. 2008). To date, the impact of ESX-like secreted substrates in virulence of *B. anthracis* is still unknown. However, the finding that specific humoral immune responses against EsxB-, EsxP- and EsxW-like proteins are detectable in *B. anthracis*-infected guinea pigs suggests that the type VII-like secretion pathway is activated during *B. anthracis* in vivo infections.

6 Concluding Comments

Taken together, the different aspects of type VII- and type VII-like secretion in Gram-positive bacteria, reviewed in this book chapter, emphasize the actuality of this topic for bacterial physiology and host–pathogen interaction. Despite considerable advances in understanding the mechanisms and functions of the different type VII- and type VII-like systems in high- and low-C + G-content Gram-positive bacteria, namely *Actinobacteria* and *Firmicutes*, many questions remain to be answered (Schneewind and Missiakas 2012). These challenges for future work have relevance for pathogenicity research (Le Chevalier et al. 2014; Majlessi and Brosch 2015),

novel treatment strategies (Rybniker et al. 2014; Christophe et al. 2009), or vaccine design (Bottai et al. 2015; Pym et al. 2003) and shall thus receive the continued attention of a broad spectrum of fundamental and applied research.

Acknowledgements Support by the European Community (Grant H2020-PHC-643381) and the Fondation pour la Recherche Médicale FRM (DEQ 20130326471) is gratefully acknowledged. M.I.G. is supported by an MD–PhD grant of the University of Groningen, The Netherlands.

References

Abdallah AM, Verboom T, Hannes F, Safi M, Strong M, Eisenberg D, Musters RJ, Vandenbroucke-Grauls CM, Appelmelk BJ, Luirink J, Bitter W (2006) A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol Microbiol 62:667–679

Abdallah AM, Gey van Pittius NC, Champion PA, Cox J, Luirink J, Vandenbroucke-Grauls CM, Appelmelk BJ, Bitter W (2007) Type VII secretion system of mycobacteria show the way. Nat Rev Microbiol 5:883–891

Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jimenez C, Parra M, Cadieux N, Brennan MJ, Appelmelk BJ, Bitter W (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol Microbiol 73:329–340

Adhikari RP, Novick RP (2008) Regulatory organization of the staphylococcal *sae* locus. Microbiology 154:949–959

Agarwal N, Woolwine SC, Tyagi S, Bishai WR (2007) Characterization of the *Mycobacterium tuberculosis* sigma factor SigM by assessment of virulence and identification of SigM-dependent genes. Infect Immun 75:452–461

Akpe San Roman S, Facey PD, Fernandez-Martinez L, Rodriguez C, Vallin C, Del Sol R, Dyson P (2010) A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. Microbiology 156:1719–1729

Anderson M, Chen YH, Butler EK, Missiakas DM (2011) EsaD, a secretion factor for the Ess pathway in *Staphylococcus aureus*. J Bacteriol 193:1583–1589

Anderson M, Aly KA, Chen YH, Missiakas D (2013) Secretion of atypical protein substrates by the ESAT-6 secretion system of *Staphylococcus aureus*. Mol Microbiol 90:734–743

Ates LS, Ummels R, Commandeur S, van der Weerd R, Sparrius M, Weerdenburg E, Alber M, Kalscheuer R, Piersma SR, Abdallah AM, Abd El Ghany M, Abdel-Haleem AM, Pain A, Jimenez CR, Bitter W, Houben EN (2015) Essential role of the ESX-5 secretion system in outer membrane permeability of pathogenic mycobacteria. PLoS Genet 11:e1005190

Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol Microbiol 44:9–19

Baptista C, Barreto HC, Sao-Jose C (2013) High levels of DegU-P activate an Esat-6-like secretion system in *Bacillus subtilis*. PLoS ONE 8:e67840

Bitter W, Houben EN, Bottai D, Brodin P, Brown EJ, Cox JS, Derbyshire K, Fortune SM, Gao LY, Liu J, Gey van Pittius NC, Pym AS, Rubin EJ, Sherman DR, Cole ST, Brosch R (2009a) Systematic genetic nomenclature for type VII secretion systems. PLoS Pathog 5: e1000507

Bitter W, Houben EN, Luirink J, Appelmelk BJ (2009b) Type VII secretion in mycobacteria: classification in line with cell envelope structure. Trends Microbiol 17:337–338

Blasco B, Chen JM, Hartkoorn R, Sala C, Uplekar S, Rougemont J, Pojer F, Cole ST (2012) Virulence regulator EspR of *Mycobacterium tuberculosis* is a nucleoid-associated protein. PLoS Pathog 8:e1002621

Boritsch EC, Supply P, Honore N, Seemann T, Stinear TP, Brosch R (2014) A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. Mol Microbiol 93:835–852

Bottai D, Brosch R (2009) Mycobacterial PE, PPE and ESX clusters: novel insights into the secretion of these most unusual protein families. Mol Microbiol 73:325–328

Bottai D, Majlessi L, Simeone R, Frigui W, Laurent C, Lenormand P, Chen J, Rosenkrands I, Huerre M, Leclerc C, Cole ST, Brosch R (2011) ESAT-6 secretion-independent impact of ESX-1 genes *espF* and *espG1* on virulence of *Mycobacterium tuberculosis*. J Infect Dis 203:1155–1164

Bottai D, Di Luca M, Majlessi L, Frigui W, Simeone R, Sayes F, Bitter W, Brennan MJ, Leclerc C, Batoni G, Campa M, Brosch R, Esin S (2012) Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation. Mol Microbiol 83:1195–1209

Bottai D, Stinear TP, Supply P, Brosch R (2014) Mycobacterial Pathogenomics and Evolution. Microbiol Spectr 2:MGM2-0025-2013

Bottai D, Frigui W, Clark S, Rayner E, Zelmer A, Andreu N, de Jonge MI, Bancroft GJ, Williams A, Brodin P, Brosch R (2015) Increased protective efficacy of recombinant BCG strains expressing virulence-neutral proteins of the ESX-1 secretion system. Vaccine 33:2710–2718

Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, Jacobs WR Jr (2001) Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immun 69:7326–7333

Brodin P, Eiglmeier K, Marmiesse M, Billault A, Garnier T, Niemann S, Cole ST, Brosch R (2002) Bacterial artificial chromosome-based comparative genomic analysis identifies *Mycobacterium microti* as a natural ESAT-6 deletion mutant. Infect Immun 70:5568–5578

Brodin P, Majlessi L, Brosch R, Smith D, Bancroft G, Clark S, Williams A, Leclerc C, Cole ST (2004a) Enhanced protection against tuberculosis by vaccination with recombinant *Mycobacterium microti* vaccine that induces T cell immunity against region of difference 1 antigens. J Infect Dis 190:115–122

Brodin P, Rosenkrands I, Andersen P, Cole ST, Brosch R (2004b) ESAT-6 proteins: protective antigens and virulence factors? Trends Microbiol 12:500–508

Brodin P, Majlessi L, Marsollier L, de Jonge MI, Bottai D, Demangel C, Hinds J, Neyrolles O, Butcher PD, Leclerc C, Cole ST, Brosch R (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. Infect Immun 74:88–98

Brodin P, Poquet Y, Levillain F, Peguillet I, Larrouy-Maumus G, Gilleron M, Ewann F, Christophe T, Fenistein D, Jang J, Jang MS, Park SJ, Rauzier J, Carralot JP, Shrimpton R, Genovesio A, Gonzalo-Asensio JA, Puzo G, Martin C, Brosch R, Stewart GR, Gicquel B, Neyrolles O (2010) High content phenotypic cell-based visual screen identifies *Mycobacterium tuberculosis* acyltrehalose-containing glycolipids involved in phagosome remodeling. PLoS Pathog 6:e1001100

Brown GD, Dave JA, Gey van Pittius NC, Stevens L, Ehlers MR, Beyers AD (2000) The mycosins of *Mycobacterium tuberculosis* H37Rv: a family of subtilisin-like serine proteases. Gene 254:147–155

Burts ML, Williams WA, Debord K, Missiakas DM (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. Proc Natl Acad Sci U S A 102:1169–1174

Burts ML, DeDent AC, Missiakas DM (2008) EsaC substrate for the ESAT-6 secretion pathway and its role in persistent infections of *Staphylococcus aureus*. Mol Microbiol 69:736–746

Carlsson F, Joshi SA, Rangell L, Brown EJ (2009) Polar localization of virulence-related Esx-1 secretion in mycobacteria. PLoS Pathog 5:e1000285

Cascioferro A, Delogu G, Colone M, Sali M, Stringaro A, Arancia G, Fadda G, Palu G, Manganelli R (2007) PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall. Mol Microbiol 66:1536–1547

Casonato S, Cervantes Sanchez A, Haruki H, Rengifo Gonzalez M, Provvedi R, Dainese E, Jaouen T, Gola S, Bini E, Vicente M, Johnsson K, Ghisotti D, Palu G, Hernandez-Pando R, Manganelli R (2012) WhiB5, a transcriptional regulator that contributes to *Mycobacterium tuberculosis* virulence and reactivation. Infect Immun 80:3132–3144

Champion PA, Stanley SA, Champion MM, Brown EJ, Cox JS (2006) C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*. Science 313:1632–1636

Champion PA, Champion MM, Manzanillo P, Cox JS (2009) ESX-1 secreted virulence factors are recognized by multiple cytosolic AAA ATPases in pathogenic mycobacteria. Mol Microbiol 73:950–962

Chaturvedi R, Bansal K, Narayana Y, Kapoor N, Sukumar N, Togarsimalemath SK, Chandra N, Mishra S, Ajitkumar P, Joshi B, Katoch VM, Patil SA, Balaji KN (2010) The multifunctional PE_PGRS11 protein from *Mycobacterium tuberculosis* plays a role in regulating resistance to oxidative stress. J Biol Chem 285:30389–30403

Chen YH, Anderson M, Hendrickx AP, Missiakas D (2012) Characterization of EssB, a protein required for secretion of ESAT-6 like proteins in *Staphylococcus aureus*. BMC Microbiol 12:219

Chen JM, Zhang M, Rybniker J, Boy-Rottger S, Dhar N, Pojer F, Cole ST (2013) *Mycobacterium tuberculosis* EspB binds phospholipids and mediates EsxA-independent virulence. Mol Microbiol 89:1154–1166

Cheung AL, Bayer AS, Zhang G, Gresham H, Xiong YQ (2004) Regulation of virulence determinants in vitro and in vivo in *Staphylococcus aureus*. FEMS Immunol Med Microbiol 40:1–9

Christie PJ (2001) Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. Mol Microbiol 40:294–305

Christophe T, Jackson M, Jeon HK, Fenistein D, Contreras-Dominguez M, Kim J, Genovesio A, Carralot JP, Ewann F, Kim EH, Lee SY, Kang S, Seo MJ, Park EJ, Skovierova H, Pham H, Riccardi G, Nam JY, Marsollier L, Kempf M, Joly-Guillou ML, Oh T, Shin WK, No Z, Nehrbass U, Brosch R, Cole ST, Brodin P (2009) High content screening identifies decaprenyl-phosphoribose 2' epimerase as a target for intracellular antimycobacterial inhibitors. PLoS Pathog 5:e1000645

Clement S, Vaudaux P, Francois P, Schrenzel J, Huggler E, Kampf S, Chaponnier C, Lew D, Lacroix JS (2005) Evidence of an intracellular reservoir in the nasal mucosa of patients with recurrent *Staphylococcus aureus* rhinosinusitis. J Infect Dis 192:1023–1028

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544

Collins AC, Cai H, Li T, Franco LH, Li X-D, Nair VR, Scharn CR, Stamm CE, Levine B, Chen ZJ, Shiloh MU (2015) Cyclic GMP-AMP Synthase (cGAS) is an innate immune DNA sensor for *Mycobacterium tuberculosis*. Cell Host Microbe 17:820–828

Converse SE, Cox JS (2005) A protein secretion pathway critical for Mycobacterium tuberculosis virulence is conserved and functional in *Mycobacterium smegmatis*. J Bacteriol 187:1238–1245

Copin R, Coscolla M, Seiffert SN, Bothamley G, Sutherland J, Mbayo G, Gagneux S, Ernst JD (2014) Sequence diversity in the pe_pgrs genes of *Mycobacterium tuberculosis* is independent of human T cell recognition. MBio 5:e00960–13

Coros A, Callahan B, Battaglioli E, Derbyshire KM (2008) The specialized secretory apparatus ESX-1 is essential for DNA transfer in *Mycobacterium smegmatis*. Mol Microbiol 69:794–808

Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. Nat Rev Microbiol 13:343–359

Daleke MH, Cascioferro A, de Punder K, Ummels R, Abdallah AM, van der Wel N, Peters PJ, Luirink J, Manganelli R, Bitter W (2011) Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway. J Biol Chem 286:19024–19034

Daleke MH, Ummels R, Bawono P, Heringa J, Vandenbroucke-Grauls CM, Luirink J, Bitter W (2012) General secretion signal for the mycobacterial type VII secretion pathway. Proc Natl Acad Sci U S A 109:11342–11347

Dave JA, Gey van Pittius NC, Beyers AD, Ehlers MR, Brown GD (2002) Mycosin-1, a subtilisin-like serine protease of *Mycobacterium tuberculosis*, is cell wall-associated and expressed during infection of macrophages. BMC Microbiol 2:30

de Jonge MI, Pehau-Arnaudet G, Fretz MM, Romain F, Bottai D, Brodin P, Honore N, Marchal G, Jiskoot W, England P, Cole ST, Brosch R (2007) ESAT-6 from *Mycobacterium tuberculosis* dissociates from its putative chaperone CFP-10 under acidic conditions and exhibits membrane-lysing activity. J Bacteriol 189:6028–6034

De Leon J, Jiang G, Ma Y, Rubin E, Fortune S, Sun J (2012) *Mycobacterium tuberculosis* ESAT-6 exhibits a unique membrane-interacting activity that is not found in its ortholog from non-pathogenic *Mycobacterium smegmatis*. J Biol Chem 287:44184–44191

Deb C, Daniel J, Sirakova TD, Abomoelak B, Dubey VS, Kolattukudy PE (2006) A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *Mycobacterium tuberculosis*. J Biol Chem 281:3866–3875

Delogu G, Brennan MJ (2001) Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. Infect Immun 69:5606–5611

Di Luca M, Bottai D, Batoni G, Orgeur M, Aulicino A, Counoupas C, Campa M, Brosch R, Esin S (2012) The ESX-5 associated *eccB-eccC* locus Is essential for *Mycobacterium tuberculosis* viability. PLoS ONE 7:e52059

Dong D, Wang D, Li M, Wang H, Yu J, Wang C, Liu J, Gao Q (2012) PPE38 modulates the innate immune response and is required for *Mycobacterium marinum* virulence. Infect Immun 80:43–54

Dorhoi A, Nouailles G, Jorg S, Hagens K, Heinemann E, Pradl L, Oberbeck-Muller D, Duque-Correa MA, Reece ST, Ruland J, Brosch R, Tschopp J, Gross O, Kaufmann SHE (2012) Activation of the NLRP3 inflammasome by *Mycobacterium tuberculosis* is uncoupled from susceptibility to active tuberculosis. Eur J Immunol 42:374–384

Ekiert DC, Cox JS (2014) Structure of a PE-PPE-EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion. Proc Natl Acad Sci U S A 111:14758–14763

Flint JL, Kowalski JC, Karnati PK, Derbyshire KM (2004) The RD1 virulence locus of *Mycobacterium tuberculosis* regulates DNA transfer in *Mycobacterium smegmatis*. Proc Natl Acad Sci U S A 101:12598–12603

Fortune SM, Jaeger A, Sarracino DA, Chase MR, Sassetti CM, Sherman DR, Bloom BR, Rubin EJ (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. Proc Natl Acad Sci U S A 102:10676–10681

Fraunholz M, Sinha B (2012) Intracellular *Staphylococcus aureus*: live-in and let die. Front Cell Infect Microbiol 2:43

Frigui W, Bottai D, Majlessi L, Monot M, Josselin E, Brodin P, Garnier T, Gicquel B, Martin C, Leclerc C, Cole S, Brosch R (2008) Control of *M. tuberculosis* ESAT-6 secretion and specific T cell recognition by PhoP. PLoS Pathog 4:e33

Fyans JK, Bignell D, Loria R, Toth I, Palmer T (2013) The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen *Streptomyces scabies*. Mol Plant Pathol 14:119–130

Gao LY, Guo S, McLaughlin B, Morisaki H, Engel JN, Brown EJ (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. Mol Microbiol 53:1677–1693

Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FS (2005) PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21:617–623

Garufi G, Butler E, Missiakas D (2008) ESAT-6-like protein secretion in Bacillus anthracis. J Bacteriol 190:7004–7011

Garzoni C, Kelley WL (2009) *Staphylococcus aureus*: new evidence for intracellular persistence. Trends Microbiol 17:59–65

Gey Van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ, Beyers AD (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G + C Gram-positive bacteria. Genome Biol 2:RESEARCH0044

Gey van Pittius NC, Sampson SL, Lee H, Kim Y, van Helden PD, Warren RM (2006) Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol Biol 6:95

Giraudo AT, Cheung AL, Nagel R (1997) The sae locus of *Staphylococcus aureus* controls exoprotein synthesis at the transcriptional level. Arch Microbiol 168:53–58

Goldstone RM, Goonesekera SD, Bloom BR, Sampson SL (2009) The transcriptional regulator Rv0485 modulates the expression of a *pe* and *ppe* gene pair and is required for *Mycobacterium tuberculosis* virulence. Infect Immun 77:4654–4667

Gonzalo-Asensio J, Malaga W, Pawlik A, Astarie-Dequeker C, Passemar C, Moreau F, Laval F, Daffe M, Martin C, Brosch R, Guilhot C (2014) Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc Natl Acad Sci U S A 111:11491–11496

Gordon BR, Li Y, Wang L, Sintsova A, van Bakel H, Tian S, Navarre WW, Xia B, Liu J (2010) Lsr2 is a nucleoid-associated protein that targets AT-rich sequences and virulence genes in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 107:5154–5159

Gray TA, Krywy JA, Harold J, Palumbo MJ, Derbyshire KM (2013) Distributive conjugal transfer in mycobacteria generates progeny with meiotic-like genome-wide mosaicism, allowing mapping of a mating identity locus. PLoS Biol 11:e1001602

Guinn KM, Hickey MJ, Mathur SK, Zakel KL, Grotzke JE, Lewinsohn DM, Smith S, Sherman DR (2004) Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*. Mol Microbiol 51:359–370

Hoffmann C, Leis A, Niederweis M, Plitzko JM, Engelhardt H (2008) Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. Proc Natl Acad Sci U S A 105:3963–3967

Houben D, Demangel C, van Ingen J, Perez J, Baldeon L, Abdallah AM, Caleechurn L, Bottai D, van Zon M, de Punder K, van der Laan T, Kant A, Bossers-de Vries R, Willemsen P, Bitter W, van Soolingen D, Brosch R, van der Wel N, Peters PJ (2012a) ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. Cell Microbiol 14:1287–1298

Houben EN, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jimenez CR, Ottenhoff TH, Luirink J, Bitter W (2012b) Composition of the type VII secretion system membrane complex. Mol Microbiol 86:472–484

Houben EN, Korotkov KV, Bitter W (2014) Take five - Type VII secretion systems of Mycobacteria. Biochim Biophys Acta 1844:1707–1716

Hsu T, Hingley-Wilson SM, Chen B, Chen M, Dai AZ, Morin PM, Marks CB, Padiyar J, Goulding C, Gingery M, Eisenberg D, Russell RG, Derrick SC, Collins FM, Morris SL, King CH, Jacobs WR Jr (2003) The primary mechanism of attenuation of bacillus Calmette-Guerin is a loss of secreted lytic function required for invasion of lung interstitial tissue. Proc Natl Acad Sci U S A 100:12420–12425

Hsueh YH, Cozy LM, Sham LT, Calvo RA, Gutu AD, Winkler ME, Kearns DB (2011) DegU-phosphate activates expression of the anti-sigma factor FlgM in *Bacillus subtilis*. Mol Microbiol 81:1092–1108

Hunt DM, Sweeney NP, Mori L, Whalan RH, Comas I, Norman L, Cortes T, Arnvig KB, Davis EO, Stapleton MR, Green J, Buxton RS (2012) Long-range transcriptional control of an

operon necessary for virulence-critical ESX-1 secretion in *Mycobacterium tuberculosis*. J Bacteriol 194:2307–2320

Huppert LA, Ramsdell TL, Chase MR, Sarracino DA, Fortune SM, Burton BM (2014) The ESX system in *Bacillus subtilis* mediates protein secretion. PLoS ONE 9:e96267

Iantomasi R, Sali M, Cascioferro A, Palucci I, Zumbo A, Soldini S, Rocca S, Greco E, Maulucci G, De Spirito M, Fraziano M, Fadda G, Manganelli R, Delogu G (2012) PE_PGRS30 is required for the full virulence of *Mycobacterium tuberculosis*. Cell Microbiol 14:356–367

Ilghari D, Lightbody KL, Veverka V, Waters LC, Muskett FW, Renshaw PS, Carr MD (2011) Solution structure of the Mycobacterium tuberculosis EsxG.EsxH complex: functional implications and comparisons with other *M. tuberculosis* Esx family complexes. J Biol Chem 286:29993–30002

Iyer LM, Makarova KS, Koonin EV, Aravind L (2004) Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. Nucleic Acids Res 32:5260–5279

Kneuper H, Cao ZP, Twomey KB, Zoltner M, Jager F, Cargill JS, Chalmers J, van der Kooi-Pol MM, van Dijl JM, Ryan RP, Hunter WN, Palmer T (2014) Heterogeneity in *ess* transcriptional organization and variable contribution of the Ess/Type VII protein secretion system to virulence across closely related *Staphylocccus aureus* strains. Mol Microbiol 93:928–943

Kobayashi K (2007) Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in Bacillus subtilis. Mol Microbiol 66:395–409

Korea CG, Balsamo G, Pezzicoli A, Merakou C, Tavarini S, Bagnoli F, Serruto D, Unnikrishnan M (2014) Staphylococcal Esx proteins modulate apoptosis and release of intracellular *Staphylococcus aureus* during infection in epithelial cells. Infect Immun 82:4144–4153

Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, Wilmanns M, Bitter W, Parret AH, Houben EN, Korotkov KV (2014) Structure of the *Mycobacterium tuberculosis* type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. Mol Microbiol 94:367–382

Korotkova N, Piton J, Wagner JM, Boy-Rottger S, Japaridze A, Evans TJ, Cole ST, Pojer F, Korotkov KV (2015) Structure of EspB, a secreted substrate of the ESX-1 secretion system of *Mycobacterium tuberculosis*. J Struct Biol 191:236–244

Le Chevalier F, Cascioferro A, Majlessi L, Herrmann JL, Brosch R (2014) *Mycobacterium tuberculosis* evolutionary pathogenesis and its putative impact on drug development. Future Microbiol 9:969–985

Lewis KN, Liao R, Guinn KM, Hickey MJ, Smith S, Behr MA, Sherman DR (2003) Deletion of RD1 from *Mycobacterium tuberculosis* mimics bacille Calmette-Guerin attenuation. J Infect Dis 187:117–123

Li Y, Miltner E, Wu M, Petrofsky M, Bermudez LE (2005) A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol 7:539–548

Lopez D, Kolter R (2010) Extracellular signals that define distinct and coexisting cell fates in *Bacillus subtilis*. FEMS Microbiol Rev 34:134–149

MacGurn JA, Raghavan S, Stanley SA, Cox JS (2005) A non-RD1 gene cluster is required for Snm secretion in *Mycobacterium tuberculosis*. Mol Microbiol 57:1653–1663

Maciag A, Dainese E, Rodriguez GM, Milano A, Provvedi R, Pasca MR, Smith I, Palu G, Riccardi G, Manganelli R (2007) Global analysis of the *Mycobacterium tuberculosis* Zur (FurB) regulon. J Bacteriol 189:730–740

Mader U, Antelmann H, Buder T, Dahl MK, Hecker M, Homuth G (2002) *Bacillus subtilis* functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. Mol Genet Genomics 268:455–467

Mahairas GG, Sabo PJ, Hickey MJ, Singh DC, Stover CK (1996) Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol 178:1274–1282

Majlessi L, Brosch R (2015) *Mycobacterium tuberculosis* meets the cytosol: The role of cGAS in anti-mycobacterial immunity. Cell Host Microbe 17:733–735

Majlessi L, Prados-Rosales R, Casadevall A, Brosch R (2015) Release of mycobacterial antigens. Immunol Rev 264:25–45

McLaughlin B, Chon JS, MacGurn JA, Carlsson F, Cheng TL, Cox JS, Brown EJ (2007) A mycobacterium ESX-1-secreted virulence factor with unique requirements for export. PLoS Pathog 3:e105

Mehra A, Zahra A, Thompson V, Sirisaengtaksin N, Wells A, Porto M, Koster S, Penberthy K, Kubota Y, Dricot A, Rogan D, Vidal M, Hill DE, Bean AJ, Philips JA (2013) *Mycobacterium tuberculosis* type VII secreted effector EsxH targets host ESCRT to impair trafficking. PLoS Pathog 9:e1003734

Mishra BB, Moura-Alves P, Sonawane A, Hacohen N, Griffiths G, Moita LF, Anes E (2010) *Mycobacterium tuberculosis* protein ESAT-6 is a potent activator of the NLRP3/ASC inflammasome. Cell Microbiol 12:1046–1063

Murray EJ, Kiley TB, Stanley-Wall NR (2009) A pivotal role for the response regulator DegU in controlling multicellular behaviour. Microbiology 155:1–8

Novick RP (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. Mol Microbiol 48:1429–1449

Ogura M, Yamaguchi H, Yoshida K, Fujita Y, Tanaka T (2001) DNA microarray analysis of *Bacillus subtilis* DegU, ComA and PhoP regulons: an approach to comprehensive analysis of *B. subtilis* two-component regulatory systems. Nucleic Acids Res 29:3804–3813

Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS (2010) *Mycobacterium tuberculosis* MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell Host Microbe 7:210–220

Pallen MJ (2002) The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? Trends Microbiol 10:209–212

Pang X, Samten B, Cao G, Wang X, Tvinnereim AR, Chen XL, Howard ST (2013) MprAB regulates the espA operon in *Mycobacterium tuberculosis* and modulates ESX-1 function and host cytokine response. J Bacteriol 195:66–75

Poulsen C, Panjikar S, Holton SJ, Wilmanns M, Song YH (2014) WXG100 protein superfamily consists of three subfamilies and exhibits an alpha-helical C-terminal conserved residue pattern. PLoS ONE 9:e89313

Pym AS, Brodin P, Brosch R, Huerre M, Cole ST (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. Mol Microbiol 46:709–717

Pym AS, Brodin P, Majlessi L, Brosch R, Demangel C, Williams A, Griffiths KE, Marchal G, Leclerc C, Cole ST (2003) Recombinant BCG exporting ESAT-6 confers enhanced protection against tuberculosis. Nat Med 9:533–539

Raghavan S, Manzanillo P, Chan K, Dovey C, Cox J (2008) Secreted transcription factor controls *Mycobacterium tuberculosis* virulence. Nature 454:717–721

Ramakrishnan L, Federspiel NA, Falkow S (2000) Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family. Science 288:1436–1439

Raman S, Puyang X, Cheng TY, Young DC, Moody DB, Husson RN (2006) *Mycobacterium tuberculosis* SigM positively regulates Esx secreted protein and nonribosomal peptide synthetase genes and down regulates virulence-associated surface lipid synthesis. J Bacteriol 188:8460–8468

Ramsdell TL, Huppert LA, Sysoeva TA, Fortune SM, Burton BM (2015) Linked domain architectures allow for specialization of function in the FtsK/SpoIIIE ATPases of ESX secretion systems. J Mol Biol 427:1119–1132

Renshaw PS, Lightbody KL, Veverka V, Muskett FW, Kelly G, Frenkiel TA, Gordon SV, Hewinson RG, Burke B, Norman J, Williamson RA, Carr MD (2005) Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. EMBO J 24:2491–2498

Rickman L, Scott C, Hunt DM, Hutchinson T, Menendez MC, Whalan R, Hinds J, Colston MJ, Green J, Buxton RS (2005) A member of the cAMP receptor protein family of transcription regulators in *Mycobacterium tuberculosis* is required for virulence in mice and controls transcription of the *rpfA* gene coding for a resuscitation promoting factor. Mol Microbiol 56:1274–1286

Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I (2002) *ideR*, an essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect Immun 70:3371–3381

Romagnoli A, Etna MP, Giacomini E, Pardini M, Remoli ME, Corazzari M, Falasca L, Goletti D, Gafa V, Simeone R, Delogu G, Piacentini M, Brosch R, Fimia GM, Coccia EM (2012) ESX-1 dependent impairment of autophagic flux by *Mycobacterium tuberculosis* in human dendritic cells. Autophagy 8:1357–1370

Rosenberg OS, Dovala D, Li X, Connolly L, Bendebury A, Finer-Moore J, Holton J, Cheng Y, Stroud RM, Cox JS (2015) Substrates control multimerization and activation of the multi-domain ATPase motor of type VII secretion. Cell 161:501–512

Ryan AA, Nambiar JK, Wozniak TM, Roediger B, Shklovskaya E, Britton WJ, Fazekas de St Groth B, Triccas JA (2009) Antigen load governs the differential priming of CD8 T cells in response to the bacille Calmette Guerin vaccine or *Mycobacterium tuberculosis* infection. J Immunol 182:7172–7177

Rybniker J, Chen JM, Sala C, Hartkoorn RC, Vocat A, Benjak A, Boy-Rottger S, Zhang M, Szekely R, Greff Z, Orfi L, Szabadkai I, Pato J, Keri G, Cole ST (2014) Anticytolytic screen identifies inhibitors of mycobacterial virulence protein secretion. Cell Host Microbe 16:538–548

Sachse F, Becker K, von Eiff C, Metze D, Rudack C (2010) *Staphylococcus aureus* invades the epithelium in nasal polyposis and induces IL-6 in nasal epithelial cells in vitro. Allergy 65:1430–1437

Sampson SL, Lukey P, Warren RM, van Helden PD, Richardson M, Everett MJ (2001) Expression, characterization and subcellular localization of the *Mycobacterium tuberculosis* PPE gene Rv1917c. Tuberculosis (Edinb) 81:305–317

Sani M, Houben EN, Geurtsen J, Pierson J, de Punder K, van Zon M, Wever B, Piersma SR, Jimenez CR, Daffe M, Appelmelk BJ, Bitter W, van der Wel N, Peters PJ (2010) Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. PLoS Pathog 6:e1000794

Sayes F, Sun L, Di Luca M, Simeone R, Degaiffier N, Fiette L, Esin S, Brosch R, Bottai D, Leclerc C, Majlessi L (2012) Strong immunogenicity and cross-reactivity of *Mycobacterium tuberculosis* ESX-5 Type VII secretion- encoded PE-PPE proteins predicts vaccine potential. Cell Host Microbe 11:352–363

Schneewind O, Missiakas DM (2012) Protein secretion and surface display in Gram-positive bacteria. Philos Trans R Soc Lond B Biol Sci 367:1123–1139

Serafini A, Boldrin F, Palu G, Manganelli R (2009) Characterization of a *Mycobacterium tuberculosis* ESX-3 conditional mutant: essentiality and rescue by iron and zinc. J Bacteriol 191:6340–6344

Shah S, Cannon JR, Fenselau C, Briken V (2015) A duplicated ESAT-6 region of ESX-5 is involved in protein export and virulence of mycobacteria. Infect Immun 83:4349–4361

Shukla A, Pallen M, Anthony M, White SA (2010) The homodimeric GBS1074 from *Streptococcus agalactiae*. Acta Crystallogr, Sect F: Struct Biol Cryst Commun 66:1421–1425

Siegrist MS, Unnikrishnan M, McConnell MJ, Borowsky M, Cheng TY, Siddiqi N, Fortune SM, Moody DB, Rubin EJ (2009) Mycobacterial Esx-3 is required for mycobactin-mediated iron acquisition. Proc Natl Acad Sci U S A 106:18792–18797

Siegrist MS, Steigedal M, Ahmad R, Mehra A, Dragset MS, Schuster BM, Philips JA, Carr SA, Rubin EJ (2014) Mycobacterial Esx-3 requires multiple components for iron acquisition. MBio 5:e01073–14

Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, Enninga J (2012) Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. PLoS Pathog 8: e1002507

Simeone R, Sayes F, Song O, Groschel MI, Brodin P, Brosch R, Majlessi L (2015) Cytosolic access of *Mycobacterium tuberculosis*: Critical impact of phagosomal acidification control and demonstration of occurrence in vivo. PLoS Pathog 11:e1004650

Solans L, Aguilo N, Samper S, Pawlik A, Frigui W, Martin C, Brosch R, Gonzalo-Asensio J (2014) A specific polymorphism in *Mycobacterium tuberculosis* H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1 component. Infect Immun 82:3446–3456

Solomonson M, Setiaputra D, Makepeace KA, Lameignere E, Petrotchenko EV, Conrady DG, Bergeron JR, Vuckovic M, DiMaio F, Borchers CH, Yip CK, Strynadka NC (2015) Structure of EspB from the ESX-1 type VII secretion system and insights into its export mechanism. Structure 23:571–583

Song H, Sandie R, Wang Y, Andrade-Navarro MA, Niederweis M (2008) Identification of outer membrane proteins of *Mycobacterium tuberculosis*. Tuberculosis (Edinb) 88:526–544

Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175–182

Sorensen AL, Nagai S, Houen G, Andersen P, Andersen AB (1995) Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infect Immun 63:1710–1717

Stamm LM, Morisaki JH, Gao LY, Jeng RL, McDonald KL, Roth R, Takeshita S, Heuser J, Welch MD, Brown EJ (2003) Mycobacterium marinum escapes from phagosomes and is propelled by actin-based motility. J Exp Med 198:1361–1368

Stanley SA, Cox JS (2013) Host-pathogen interactions during Mycobacterium tuberculosis infections. Curr Top Microbiol Immunol 374:211–241

Stanley SA, Raghavan S, Hwang WW, Cox JS (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. Proc Natl Acad Sci U S A 100:13001–13006

Stanley SA, Johndrow JE, Manzanillo P, Cox JS (2007) The Type I IFN Response to Infection with *Mycobacterium tuberculosis* Requires ESX-1-Mediated Secretion and Contributes to Pathogenesis. J Immunol 178:3143–3152

Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 103:8060–8065

Sundaramoorthy R, Fyfe PK, Hunter WN (2008) Structure of *Staphylococcus aureus* EsxA suggests a contribution to virulence by action as a transport chaperone and/or adaptor protein. J Mol Biol 383:603–614

Supply P, Marceau M, Mangenot S, Roche D, Rouanet C, Khanna V, Majlessi L, Criscuolo A, Tap J, Pawlik A, Fiette L, Orgeur M, Fabre M, Parmentier C, Frigui W, Simeone R, Boritsch EC, Debrie AS, Willery E, Walker D, Quail MA, Ma L, Bouchier C, Salvignol G, Sayes F, Cascioferro A, Seemann T, Barbe V, Locht C, Gutierrez MC, Leclerc C, Bentley SD, Stinear TP, Brisse S, Médigue C, Parkhill J, Cruveiller S, Brosch R (2013) Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat Genet 45:172–179

Sysoeva TA, Zepeda-Rivera MA, Huppert LA, Burton BM (2014) Dimer recognition and secretion by the ESX secretion system in *Bacillus subtilis*. Proc Natl Acad Sci U S A 111:7653–7658

Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tuber Lung Dis 79:329–342

Teutschbein J, Schumann G, Mollmann U, Grabley S, Cole ST, Munder T (2009) A protein linkage map of the ESAT-6 secretion system 1 (ESX-1) of *Mycobacterium tuberculosis*. Microbiol Res 164:253–259

Tuchscherr L, Medina E, Hussain M, Volker W, Heitmann V, Niemann S, Holzinger D, Roth J, Proctor RA, Becker K, Peters G, Loffler B (2011) *Staphylococcus aureus* phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. EMBO Mol Med 3:129–141

Ummels R, Abdallah AM, Kuiper V, Aajoud A, Sparrius M, Naeem R, Spaink HP, van Soolingen D, Pain A, Bitter W (2014) Identification of a novel conjugative plasmid in mycobacteria that requires both type IV and type VII secretion. MBio 5:e01744–14

van der Wel N, Hava D, Houben D, Fluitsma D, van Zon M, Pierson J, Brenner M, Peters PJ (2007) *M. tuberculosis* and *M. leprae* translocate from the phagolysosome to the cytosol in myeloid cells. Cell 129:1287–1298

van der Woude AD, Luirink J, Bitter W (2013) Getting across the cell envelope: mycobacterial protein secretion. Curr Top Microbiol Immunol 374:109–134

Voskuil MI, Schnappinger D, Rutherford R, Liu Y, Schoolnik GK (2004) Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. Tuberculosis (Edinb) 84:256–262

Wagner JM, Evans TJ, Chen J, Zhu H, Houben EN, Bitter W, Korotkov KV (2013) Understanding specificity of the mycosin proteases in ESX/type VII secretion by structural and functional analysis. J Struct Biol 184:115–128

Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I (2006) The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. Mol Microbiol 60:312–330

Wang S, Engohang-Ndong J, Smith I (2007) Structure of the DNA-binding domain of the response regulator PhoP from *Mycobacterium tuberculosis*. Biochemistry 46:14751–14761

Wassermann R, Gulen MF, Sala C, Perin SG, Lou Y, Rybniker J, Schmid-Burgk JL, Schmidt T, Hornung V, Cole ST, Ablasser A (2015) *Mycobacterium tuberculosis* differentially aActivates cGAS- and inflammasome-dependent intracellular immune responses through ESX-1. Cell Host Microbe 17:799–810

Watson RO, Manzanillo PS, Cox JS (2012) Extracellular *M. tuberculosis* DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. Cell 150:803–815

Watson RO, Bell SL, MacDuff DA, Kimmey JM, Diner EJ, Olivas J, Vance RE, Stallings CL, Virgin HW, Cox JS (2015) The cytosolic sensor cGAS detects *Mycobacterium tuberculosis* DNA to Induce type I Interferons and activate autophagy. Cell Host Microbe 17:811–819

Way SS, Wilson CB (2005) The *Mycobacterium tuberculosis* ESAT-6 homologue in *Listeria monocytogenes* is dispensable for growth in vitro and in vivo. Infect Immun 73:6151–6153

Wong KW, Jacobs WR Jr (2011) Critical role for NLRP3 in necrotic death triggered by *Mycobacterium tuberculosis*. Cell Microbiol 13:1371–1384

Zhang M, Chen JM, Sala C, Rybniker J, Dhar N, Cole ST (2014) EspI regulates the ESX-1 secretion system in response to ATP levels in *Mycobacterium tuberculosis*. Mol Microbiol 93:1057–1065

Zuber B, Chami M, Houssin C, Dubochet J, Griffiths G, Daffe M (2008) Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. J Bacteriol 190:5672–5680

Protein Secretion in Gram-Positive Bacteria: From Multiple Pathways to Biotechnology

Jozef Anné, Anastassios Economou and Kristel Bernaerts

Abstract A number of Gram-positive bacteria are important players in industry as producers of a diverse array of economically interesting metabolites and proteins. As discussed in this overview, several Gram-positive bacteria are valuable hosts for the production of heterologous proteins. In contrast to Gram-negative bacteria, proteins secreted by Gram-positive bacteria are released into the culture medium where conditions for correct folding are more appropriate, thus facilitating the isolation and purification of active proteins. Although seven different protein secretion pathways have been identified in Gram-positive bacteria, the majority of heterologous proteins are produced via the general secretion or Sec pathway. Not all proteins are equally well secreted, because heterologous protein production often faces bottlenecks including hampered secretion, susceptibility to proteases, secretion stress, and metabolic burden. These bottlenecks are associated with reduced yields leading to non-marketable products. In this chapter, besides a general overview of the different protein secretion pathways, possible hurdles that may hinder efficient protein secretion are described and attempts to improve yield are discussed including modification of components of the Sec pathway. Attention is also paid to omics-based approaches that may offer a more rational approach to optimize production of heterologous proteins.

J. Anné (✉) · A. Economou
Laboratory of Molecular Bacteriology, Department of Microbiology, and Immunology, KU Leuven, Rega Institute for Medical Research, 3000 Leuven, Belgium
e-mail: jozef.anne@kuleuven.be

A. Economou
e-mail: tassos.economou@kuleuven.be

K. Bernaerts
Bio- & Chemical Systems Technology, Reactor Engineering and Safety Section, Department of Chemical Engineering, KU Leuven, Celestijnenlaan 200F, 3001 Leuven, Belgium
e-mail: kristel.bernaerts@kuleuven.be

Current Topics in Microbiology and Immunology (2017) 404:267–308
DOI 10.1007/82_2016_49

Published Online: 22 November 2016

Abbreviations

FEA	Flagella export apparatus
GFP	Green fluorescent protein
GRAS	Generally recognized as safe
LAB	Lactic acid bacteria
mTNFα	Mouse tumor necrosis factor α
PMF	Proton motive force
PSPa	Phage shock protein A
Sec pathway	General secretory pathway
SPase I	Signal peptidase type I
T4SS	Type 4 secretion system
T7SS	Type VII secretion system
Tat	Twin-arginine translocation

Contents

1 Introduction

Biotechnology involves the use of living organisms and their products for the benefit of mankind in different areas. The biotechnological manufacturing of products of biomedical interest such as antibiotics, vaccines, antibodies, and other

biopharmaceuticals are termed red biotechnology. On the other hand, the production of industrial enzymes for the sustainable processing and production of chemicals, materials, and fuels is classified as white biotechnology, while green biotechnology serves agriculture and involves, for example, the development of pest-resistant plants and the control of food and feed. Recently, blue biotechnology gained interest, whereby marine organisms and their products are used for the making of valuable compounds including drugs and bioenergy. Molecules of interest can be synthetized by the original producing organism, but in many cases they are heterologously produced. This means that the necessary genes derived from other organisms are cloned into the host cell of choice with the intention to express the molecules encoded by the cloned DNA so that the host produces the molecules that it normally does not make. The global biotechnology market size is estimated to have a value of more than US$398 billion in 2015 and a growth expectation at a compound annual growth rate of 12.3 %. To be profitable in this vast market, not only the value of the molecules produced counts, but the cost-effectiveness and environmental-friendliness of the production process are at least as important. The main challenge therefore is to have the most effective production process for lowering the cost of the final product to obtain a commercially viable process. Optimization of microbes in production processes has been done over the past decades using, for example, random mutagenesis and selection, identification of metabolic reactions whose activities should be modified to achieve the desired cellular objective, genome-scale modeling of metabolism, and by fermentation optimization. More recently, these approaches are enhanced by synthetic biology tools.

To produce heterologous molecules, a variety of different expression systems has been described, each with its own advantages and disadvantages. The most popular host is *Escherichia coli*, as proven by the fact that nearly half of the approved recombinant biopharmaceuticals are synthesized using this host. Reasons why *E. coli* is so popular in recombinant protein production and as a workhorse in the laboratories of academia and the biopharma industry are obvious: its genetics are far better understood than those of other microorganisms; there are many genetic tools available, and proteins, if expressed, can be obtained in high production yields. As other bacteria, *E. coli* grows quickly on cheap media to high cell density. A drawback, however, is that *E. coli* are Gram-negative bacteria inherently having an outer membrane bilayer (OM) composed of lipopolysaccharides (LPS). The OM acts as an effective permeability barrier hindering secreted proteins from being released into the extracellular medium. LPS contains the endotoxin lipid A, that if released in the blood can cause septic shock, a systemic inflammatory response syndrome. When overexpressed in *E. coli*, many proteins become misfolded and accumulate in the cytoplasm as inclusion bodies. To become active, these inclusion bodies need to be solubilized, and proteins refolded into bioactive molecules, an often cumbersome process, with poor recovery and accounting for the major cost in the production process of recombinant proteins. Therefore, new solubilization techniques have been proposed, as well as genetic approaches to make *E. coli* a better host, for example, engineering *E. coli* strains that possess an

oxidative cytoplasmic environment that favors disulfide bond formation, overexpression of different chaperones or combinations thereof, the fusion of appropriate tags to the N- or C-terminus of the overexpressed protein, secretion in the medium using the alpha-hemolysin secretion system. These improvements can be helpful, but are not used at an industrial scale, and depending on the proteins to be obtained other host cells derived from Gram-positive bacteria might be a better option.

An important asset is that Gram-positive bacteria are monoderm with a cell envelop that surrounds the cytoplasmic membrane with a thick peptidoglycan layer and associated teichoic acids. This structure protects the cell from mechanical or osmolytic lysis and is an anchor place for proteins, glycopolymers, and cations. Notwithstanding its complex structure, the cell envelop of Gram-positive bacteria is permeable to proteins as it does not contain an outer membrane. Consequently, secreted proteins will be released into the culture medium, where they can obtain their native conformation simplifying downstream processing. This is an advantage for the industrial production of heterologous proteins.

In this review, a survey will be presented about the protein secretion pathways in Gram-positive bacteria together with possible applications for specific species. Complementary to the beneficial properties of heterologous expression in Gram-positive bacteria, strategies for enhancing heterologous protein production are developed to acquire commercially acceptable production and yield. This review will further give a brief compilation of approaches tackling the bottlenecks at the level of expression up to metabolic fluxes.

2 Protein Secretion Pathways in Gram-Positive Bacteria

Protein secretion is a vital process for all organisms, since about 35 % of all proteins made in a cell are either membrane-embedded or secreted (Orfanoudaki and Economou 2014). To do so bacteria have different secretion pathways at their disposal. Whereas at least 7 diverse secretion systems (type I–VII) have been identified for Gram-negative bacteria, 6 protein translocation systems were reported for Gram-positive bacteria [for an overview, see Forster and Marquis (2012)] as explained below for Gram-positive bacteria. Only two of them are used for biotechnology purposes (Fig. 1).

2.1 General Secretion (Sec) System

The most important protein secretion pathway is the general secretion (Sec) system, which directs proteins to the cytoplasmic membrane for their insertion into or translocation across the membrane. Proteins destined for secretion are in general synthesized as preproteins with an N-terminally extended sequence, named the signal peptide. The primary sequences of signal peptides are not homologous, although they

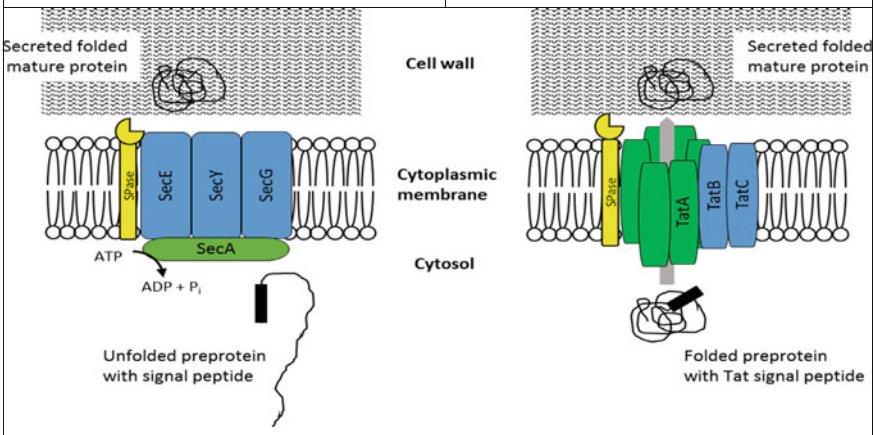
Sec pathway	Twin-arginine transport pathway
	
Used with species such as *B. subtilis, B. licheniformis, B. amyloliquefaciens S. lividans, S. coelicolor, C. sporogenes, C. glutamicum, L. lactis, L. plantarum, C. acetobutylicum, C. sporogenes, C. novyi*-NT, Bacterial enzymes and several eukaryotic proteins	Used with species: *B. subtilis, S. lividans, C. glutamicum* Bacterial enzymes, GFP and other eukaryotic proteins
Non-classically secreted protein pathway	Tested at lab scale with *B. subtilis* for secretion of heterologous bacterial enzymes
Type IV secretion	Transfer of heterologous effector proteins by *Escherichia coli* pKM101-encoded conjugation system (Whitaker et al. 2016)
Type VII secretion system	Not applied
Flagella export apparatus	Secretion of recombinant proteins in *Salmonella* (Vondervisztet al. 2012)
Holins	Not applied

Fig. 1 Overview of protein secretion pathways of Gram-positive bacteria and their possible use in biotechnology

do have 3 common structural features: a net positively charged N-terminus, a hydrophobic core region (H-region), and a polar C-terminal end containing the signal peptidase recognition site. In bacteria, the signal peptide is between 20 and 30 residues long, but can contain even more than 50 amino acids residues depending on the species and the protein to be secreted. The role of the signal peptide is to guide the protein to the secretion channel following binding to soluble targeting factors. The involvement of chaperones to target newly synthesized proteins to the translocation

pathways, either co- or posttranslationally, is best studied for *E. coli*, as summarized hereafter. The chaperones trigger factor (Tf), DnaK/DnaJ/GrpE (DnaKJE) and GroEL do not only play a major role in the folding of newly synthesized cytosolic proteins, but are also important for posttranslational protein targeting (Grady et al. 2012; Castanie-Cornet et al. 2014). Besides Tf, DnaKJE and GroEL, of prime importance for secretion in proteobacteria, is SecB, which keeps proteins in an unfolded secretion-competent state and delivers them to the Sec translocon SecYEG via its interaction with SecA, the dimeric ATPase subunit of bacterial protein translocase (Karamanou et al. 1999). The peripherally associated motor protein SecA drives then the protein translocation step by repeated cycles of ATP-binding and hydrolysis resulting in SecA membrane insertion/deinsertion and stepwise exportation of the preprotein through the channel (Chatzi et al. 2013). Additional energy promoting translocation, when the preprotein is detached from SecA, comes from the proton motive force [PMF; (Schiebel et al. 1991)]. During or following translocation, the signal peptide is cleaved off by the membrane-embedded signal peptidase I or II (the latter specializing on secreted lipoprotein signal peptides) at the signal peptidase recognition site located in its C-terminal end. This recognition site is often a canonical A-X-A motif, but other residues are permitted as well. It was recently shown by computational analyses of $\sim$1500 genomes that numerous major evolutionary clades have replaced the canonical signal peptide sequence with novel motifs (Payne et al. 2012).

Besides the Sec pathway, there is also the signal recognition pathway (SRP) for cotranslational secretion that in bacteria mainly deals with membrane protein insertion and to a lesser extent with protein secretion. SRP binds to particularly hydrophobic N-terminal signal sequences or hydrophobic transmembrane segments as they emerge from the ribosome. The SRP/RNC (*r*ibosome *n*ascent *c*hain) complex interacts with the membrane-bound *S*RP *R*eceptor (SR) and the delivery of the RNC to the translocation channel SecYEG in the membrane finally leads to the dissociation of the SRP/SR complex, whereupon the preprotein is driven across the translocation channel with the help of continuing translation and/or SecA.

It is generally assumed that the secretion of proteins in Gram-positive bacteria follows similar steps as they occur in *E. coli* since genes involved in the Sec-dependent protein secretion pathway are identified in the genome of Gram-positive bacteria. Nonetheless SecB is absent from the genomes of Gram-positive bacteria, although some SecB-like genes are present. For example, SecB-like protein (Rv1957) is present in the *Mycobacterium tuberculosis* genome where it specifically controls a stress-responsive toxin–antitoxin system. Experiments suggest that Rv1957 could play a role in protein export of *M. tuberculosis* (Sala et al. 2014), but its role is completely different from that of SecB in Gram-negative bacteria. In *Bacillus subtilis*, the SecB chaperone function has been attributed to CsaA (Shapova and Paetzel 2007). How Sec-dependent secretory proteins are kept in a Sec-secretion-competent way, and which chaperones are involved is not clear yet.

2.2 Twin-Arginine Transport Pathway (Tat)

A special characteristic of the Tat pathway is that proteins are transported across the cytoplasmic membrane in a folded state, and the energy for translocation comes from the proton motive force (PMF). Many Tat substrates receive cofactors and fold prior to translocation. Based on genome sequence analysis of prokaryotes, the Tat pathway is present in nearly 80 % of all bacteria, also in *Archaea* (Simone et al. 2013). Results indicate that the Tat pathway is utilized to highly varying extents. It operates in parallel with the Sec pathway. Signal peptides that target proteins to the Tat pathway resemble Sec signal peptides, but with a conserved S/T-R-R-x-FLK consensus motif at the end of the N-region, where the twin-arginines are invariant and normally essential for efficient export by the Tat pathway (Stanley et al. 2000). However, the Tat-specific signal sequence with two arginine residues may not be an absolute prerequisite for the Tat pathway (Watanabe et al. 2009). The Tat translocon comprises two kinds of small membrane proteins: TatC, a highly hydrophobic protein with 6 predicted transmembrane helices, and with its N- and C-termini at the cytoplasmic face of the membrane. The Tat translocon contains also one or two members of the TatA protein family, named TatA and TatB, sequence-related proteins with a common structure, each predicted to comprise a membrane-spanning α-helix at the N-terminus, immediately followed by an amphipathic helix located at the cytoplasmic side of the membrane and a C-terminal region of variable length. TatB and TatC form an oligomeric, multivalent receptor complex that binds Tat substrates, while multiple protomers of TatA assemble at substrate-bound TatBC receptors to facilitate substrate transport (Cleon et al. 2015). Minimal Tat systems contain only one type of TatA and one type of TatC. When the signal peptide of a Tat-dependent protein is recognized, it will be bound by a multi-subunit TatBC complex located in the membrane and Tat secretion is initiated. This binding event triggers the PMF-dependent recruitment and oligomerization of TatA protomers from a pool in the membrane to form the active TatABC-containing translocation site (Berks 2015). Possible cross talk between the Tat- and Sec-dependent protein secretion pathways has been reported (Goosens et al. 2014). This assumed interaction, however, needs to be further investigated.

In actinomycetes including *Streptomyces lividans* and other streptomycetes, *Mycobacterium* (McDonough et al. 2005) and *Corynebacterium* (Oertel et al. 2015), TatA, TatB, and TatC are the components for this pathway, similarly to Gram-negative bacteria. Of the Tat components in *Streptomyces*, TatC is essential whereas TatA and TatB are individually dispensable and are, next to the membrane-embedded localization, also found as active soluble complexes in the cytoplasm (De Keersmaeker et al. 2007). In contrast to the majority of Tat-containing organisms in which the Sec pathway is the major route for protein transport, the Tat pathway seems to be an important protein secretion route in *Streptomyces*. It is estimated that nearly 20 % of proteins of the extracellular proteome is secreted via the Tat pathway (Widdick et al. 2006). Using enhanced green fluorescent protein (eGFP) and mCherry fusions of the proteins of the Tat

machinery, Willemse et al. (2012) tried to determine their subcellular localization in *Streptomyces coelicolor* throughout the complex life cycle of this organism. They showed that TatA, TatB, and TatC dynamically co-localize in the vegetative hyphae, with a strong preference for apical sites in growing hyphae. For *Corynebacterium,* secretion is mainly studied in *Corynebacterium glutamicum* (Kikuchi et al. 2006). A functional *C. glutamicum* Tat system requires TatA and TatC, while the TatB protein seems to be dispensable, but it is important for maximal efficiency, and it was also found to be essential for the secretion of a heterologous Tat-dependent model protein into the *C. glutamicum* culture supernatant (Oertel et al. 2015). It was further shown that TatB (in combination with TatA and TatC) is strictly required for unimpaired aerobic growth (Oertel et al. 2015).

Both *M. tuberculosis* and *Mycobacterium smegmatis* have a functional Tat pathway. As for other actinomycetes, the Tat translocon consists of TatA, TatB, and TatC (McDonough et al. 2005). In *M. tuberculosis,* the Tat pathway is essential for growth (Saint-Joanis et al. 2006) as concluded from the inability to obtain viable deletion mutants. This contrasts with *M. smegmatis* for which viable Tat mutants could be obtained, although these mutants showed growth defects. In addition, increased sensitivity to β-lactam antibiotics was also noticed, this as a consequence of reduced export of β-lactamase BlaS, a protein with a predicted Tat signal peptide (Saint-Joanis et al. 2006).

For staphylococci, only some species contain a functional Tat system, including *Staphylococcus carnosus* (Meissner et al. 2007), *Staphylococcus haemolyticus* (Yamada et al. 2007), and *Staphylococcus aureus* (Biswas et al. 2009). The Tat system is composed of TatA and TatC and was proven to translocate iron-dependent peroxide FepB in *S. aureus* (Biswas et al. 2009). *B. subtilis* secretion system has been studied extensively (Goosens et al. 2014). Genes for two TatC (TatCd and TatCy) and three TatA components (TatAd, TatAy, and TatAc) have been identified in the *B. subtilis* genome. The core Tat complex consists of TatA and TatC, namely TatAy–TatCy, of which the latter is constitutively expressed and exports more substrates, including the Dyp-type peroxidase EfeB (YwbN), the Rieske iron-sulfur protein QcrA, and the alkaline phosphatase YkuE, while TatAdCd is only expressed under phosphate limitation (Pop et al. 2002). A third TatA-like protein TatAc can be combined as TatAcCd and TatAcCy. It is supposed to be an intermediate evolutionary step in TatA–TatB specialization (Goosens et al. 2015).

2.3 Type IV Secretion Systems (T4SSs)

T4SSs transport a diverse array of substrates from DNA to nucleoprotein complexes and effector proteins. They are multi-subunit, membrane-spanning translocation systems found in Gram-positive as well as Gram-negative bacteria and in some archaea (Chandran Darbari and Waksman 2015). They have evolved from a self-transmissible, single-stranded DNA conjugation system with VirB4-like AAA + ATPase to systems with an enormous diversity in their overall structure

and the types of substrates secreted. T4SSs can be divided into 3 groups (Bhatty et al. 2013): (1) a conjugation system to deliver ssDNA and one or more proteins across the membrane to the bacterial or eukaryotic target cell whereby direct cell contact is required; (2) the effector translocation system by a contact-dependent mechanism to deliver proteins to the cytosol of eukaryotic target cells, and (3) a release/uptake system to export/import molecules from/to the extracellular milieu. As a consequence, T4SSs are involved in a variety of functions including type 4 pilus formation, toxin and other protein secretion, gene transfer, and biofilm formation. Secreted substrates are involved in pathogenesis and adaptation to the cellular host environment. T4SSs translocate also proteins that form pilin-like structures (Chen and Dubnau 2004). A typical characteristic of these proteins is the presence of a specialized leader peptide that is cleaved off by a cognate membrane-bound type 4 prepilin peptidase during the process of secretion. Only T4SS conjugation systems are known in Gram-positive bacteria and Archaea to date.

2.4 Type VII Secretion System (T7SS)

Recent studies have uncovered a T7SS or early secretory antigen 6-KDa (ESX) secretion system. Originally, it was detected via an in silico analysis of the *M. tuberculosis* virulence effectors ESAT-6 (early-secreted antigenic target, 6 kDa) and the associated 10-kDa culture filtrate protein (CFP-10, EsxB) encoded by the *esxA* and *esxB* genes, respectively. They were known to be secreted despite the lack of a recognizable secretion signal (Tekaia et al. 1999). Esx proteins are characterized by their small size ($\sim$100 residues) and a WXG motif in the middle of the protein that forms a hairpin bend (Pallen 2002). Therefore, an alternative name was proposed for T7SS, the WXG100 secretion system (Wss) (Sutcliffe 2011), because distant homologues of ESAT-6/CFP-10 identified in Gram-positive bacteria all share a central WXG motif. T7SSs are widespread in actinomycetes and Gram-positive bacteria and affect a range of bacterial processes including sporulation, conjugation, and cell wall stability (Sysoeva et al. 2014). The T7SS is a complex system with many components and substrates, at least in mycobacteria. *M. tuberculosis* has five T7S systems, designated ESX-1 through ESX-5 (Stoop et al. 2012), which show similarity in gene content and gene order. Of these T7S systems, 3 are important for survival in the host, namely ESX-3, responsible for the uptake of iron and zinc, and ESX-5, responsible for the secretion of immunomodulatory effector proteins, and ESX-1 is most crucial for virulence. First detected in *M. tuberculosis* (Stanley et al. 2003), it was shown afterward also to be present in the non-pathogenic species *M. smegmatis* (Converse and Cox 2005) and the fish pathogen *Mycobacterium marinum* (Abdallah et al. 2009). On the other hand, ESX-1 is absent in *Mycobacterium bovis* BCG, the attenuated vaccine strain. PE/PPE is also secreted by the T7SS. The PE/PPE protein family, which has a conserved signature motif proline–glutamate and proline–proline–glutamate residues near the start of their encoded proteins, affects mycobacterial interactions with

the innate immune system, specifically inhibiting macrophage function (Ahmed et al. 2015). ESX-1 and ESX-5 have been implicated in major roles in the secretion of PE/PPE proteins (Abdallah et al. 2009). ESX secretion seems to be crucial for establishing and maintaining the infection for *M. tuberculosis.* A PSI-BLAST search on sequences retrieved from the NCBI or the ViruloGenome databases further evidenced the presence of ESAT-6 homologues in a number of low-GC Gram-positive bacteria, and also in several actinobacteria other than *Mycobacterium* (Pallen 2002), including all sequenced *Streptomyces* genomes such as *S. coelicolor*, *S. lividans,* and *S. scabies*. The biological importance of this pathway for streptomycetes is, so far, less well-known and begins only just to be revealed. For *S. scabies* no role in virulence for any of the T7SS components in any of the plant infection models tested could be detected, but it was demonstrated that components encoded by the T7SS gene cluster are required for the normal growth and development of *S. scabies* (Fyans et al. 2013). By mutagenesis analysis, it was shown that also proteins encoded by the *esxBA* operon and belonging to the WXG-100 superfamily play a role in morphogenesis in *S. coelicolor* (San Roman et al. 2010). In the sequenced genomes of other Gram-positive bacteria including *B. subtilis, Bacillus anthracis, Clostridium acetobutylicum, Listeria monocytogenes*, and *S. aureus,* ESAT-6 homologues were also discovered (Pallen 2002) and confirmed experimentally, for example, for *S. aureus* (Burts et al. 2005), *B. subtilis* (Sysoeva et al. 2014), *B. anthracis* (Fan et al. 2015).

2.5 Flagella Export Apparatus (FEA)

This specific protein export apparatus serves to secrete proteins that form the flagella hook, filament, and cap (Erhardt et al. 2010). Flagellar T3SS are present both in Gram-positive and Gram-negative bacteria, and it has been proposed that the type III secretion required for pathogenesis evolved from flagellar-specific T3SS (Hueck 1998). To transport proteins that form the flagella hook, filament, and cap to the distal growing end, the FEA utilizes ATP and PMF as the energy source (Paul et al. 2008). The flagellar export apparatus is thought to be the ancestor of all T3SS functions in the export of several components of the flagellum across the cytoplasmic membrane into the channel of the flagellum for assembly. Not much is known, however, about the FEA for Gram-positive bacteria. One report mentions that in *B. subtilis* FlgM is secreted by the flagellar export apparatus, consistent with the model of morphogenetic coupling proposed in *Salmonella enterica* (Calvo and Kearns 2015).

2.6 Holins

Originally holins were used to describe a group of phage-encoded pore forming membrane proteins that control access of phage-encoded endolysins to the

peptidoglycan layer. During the phage lytic cycle, holins insert into the bacterial cell membrane to translocate phage-encoded cell wall hydrolases (Wang et al. 2000). Holins may also be important for a variety of other functions in Gram-positive phage-free bacteria (Saier and Reddy 2015) such as (i) spore morphogenesis and germination in *B. subtilis* (Real et al. 2005); (ii) biofilm formation and DNA release for *S. aureus* (Fischer et al. 2014) (iii) programmed cell death and acetate metabolism in *S. aureus* (Ahn et al. 2012); and (iv) biofilm formation and oxidative stress adaptation in *Streptococcus mutans* (Westbye et al. 2013). A number of practical applications have been described for the holin/lysin systems, for example, aiming to control bacterial or viral infections (Yan et al. 2013; Shi et al. 2012) or to deliver drugs, nucleic acids, and proteins to animal cells (Kuo et al. 2009).

2.7 Non-classically Secreted Proteins

Extracellular proteomic studies revealed that a number of proteins are found in the extracellular medium without any secretion signal (Tjalsma et al. 2004). As their secretion route is not known, they are indicated as "non-classically secreted proteins" (Wang et al. 2016). Although there is a debate if these proteins in the extracellular medium are not a consequence of cell lysis, evidence was given by Yang et al. (2011), who experimentally showed that the *B. subtilis* carboxylesterase Est55, and several other cytoplasmic proteins are secreted through a process in which the protein domain structure plays a contributing role. Furthermore, using enolase to which the heterologous protein GFP was fused it was shown that the intact long N-terminus including the hydrophobic helix domain is required to serve as a non-cleavable signal for the secretion of enolase (Yang et al. 2014). Moreover, signals of "non-classically secreted proteins" could be more generally used for the secretion of heterologous proteins (Chen et al. 2016).

Despite the presence in Gram-positive bacteria of a variety of different export systems, the industrial production of (heterologous) proteins has relied primarily on the Sec-dependent pathway and to a far lesser extent the Tat pathway. Below an overview is given for using these systems in a number of different Gram-positive host cells.

3 Gram-Positive Bacteria as Hosts for Heterologous Protein Production

Gram-positive bacteria are considered interesting hosts for the production of heterologous proteins. An important advantage they have is that secreted proteins are released into the culture medium in which the conditions are favorable for the correct

folding of heterologous proteins. This contrasts to the reducing environment of the cytoplasm, in which secretory proteins that undergo oxidative folding, cannot fold. In addition, secreted proteins have the advantage that the mature protein has no methionine extension, but the authentic N-terminal amino acid sequence because of the cleavage by the signal peptidase. Therefore, several Gram-positive bacteria have been evaluated as hosts for the secretory production of heterologous proteins. Reasons why specific species have been tested are, for example, their industrial importance (or that of their relatives) and known fermentation technology, their proven secretion capacity, the absence of pathogenicity and toxicity and available tools for genetic manipulation. In Table 1, a number of possible advantages/disadvantages are compared for Gram-positive bacteria versus Gram-negative bacteria (*E. coli*).

3.1 Streptomyces

Streptomycetes belong to the phylum *Actinobacteria,* filamentous or rod-shaped bacteria, of which the filamentous forms tend to produce branching filaments. These Gram-positive soil bacteria are widespread in nature, they have a high guanine and cytosine content in their DNA (70–73 % GC) and a remarkably large genome size of up to 11.9 Mbps (*S. bingchenggensis* BCW-1; Accession ID: CP002047) (Wang et al. 2010) and with gene clusters from just a few to more than 30 pathways for the biosynthesis of a diverse range of secondary metabolites (Nett et al. 2009). Various *Actinomycetales* species are the richest source of natural products, they account for about 45 % of all microbial bioactive secondary metabolites with about 80 % of the 7600 compounds being produced by streptomycetes (Berdy 2005). Many of these secondary metabolites are of industrial and pharmaceutical value, including clinically important antibiotics for human and veterinary medicine or applied in agriculture, anticancer, and immunosuppressive agents, other pharmacologically active compounds, antiparasitic agents, and herbicides. Streptomycetes also play an important role in nature. Thanks to the large variety of enzymes they produce, such as cellulases and chitinases, they help to break down decaying vegetation as such playing an important role in the C- and N-cycle and replenishing the soil with nutrients. Typical for streptomycetes is their complex life cycle: Under suitable growth conditions, exospores germinate and subsequently develop into hyphae, which frequently become branched forming the vegetative mycelia that subsequently differentiate to aerial mycelia. Finally, aerial mycelia become divided into long chains of prespore compartments, which eventually mature to thick-walled exospores. In this phase of the life cycle, a large array of secondary metabolites is produced (van Wezel and McDowall 2011).

Strains of the class Actinobacteria include some of the most common soil, freshwater, and marine life. Other Actinobacteria inhabit plants and animals, including some pathogens such as *M. tuberculosis*, several strains of *Corynebacterium, Nocardia, Rhodococcus* spp., and a few *Streptomyce*s species (Goodfellow 2012).

Table 1 Strengths and weaknesses of Gram-positive host cells compared to *E. coli* for heterologous protein production

	Gram-positive bacteria	Gram-negative bacteria (*E. coli*)
Strengths		
Growth and physiology	Fast unicellular growth for a number of strains (e.g., *Bacillus, Lactobacilli*, *Corynebacterium*), facilitating easy fermentation standardization	Fast growth, unicellular facilitating easy fermentation standardization
		Many different expression hosts available
Media	Cheap, no complex production media	Cheap, no complex production media
Genetic tools	DNA sequences of most important strains are available	Extensive knowledge on genetics
		Many genetic tools available, hereby facilitating easy cloning and recombination
		Large variety of promoters (strong and regulated)
Yield		Very high expression levels; high biomass yield in fed-batch fermentation
Safety	GRAS status (Lactobacilli, *B. subtilis, C. glutamicum, S. lividans*)	
Product quality and downstream processing	Correct preprotein processing during secretion creating a native mature protein	
	Extracellular secretion allows easy downstream processing and product recovery	
	Extracellular secretion in the medium promotes correct folding	
	Extracellular secretion minimizes contamination from host protein	
	Hosts with different GC content	
	Extensive knowledge of industrial processes for several hosts	
Weaknesses		
Growth and physiology		Outer membrane with lipopolysaccharides requiring careful downstream processing
	Streptomyces: slower growth, mycelium morphology and clump formation for *Streptomyces* making fermentation more challenging	
Genetic tools	Limited number of genetic tools	
	Less variety of promoters	

(continued)

Table 1 (continued)

	Gram-positive bacteria	Gram-negative bacteria (*E. coli*)
Product quality and downstream processing	Product degradation likely for several *Bacillus* strains due to an excess of proteases	N-terminal methionine of non-secreted proteins; extracellular protein secretion not easily achieved
		Proteins are not secreted and inclusion bodies can be formed
		Inclusion bodies make downstream processing more expensive and environmentally unfriendly
	Minimal posttranslational modification of proteins	Minimal posttranslational modification of proteins
	Multi-domain eukaryotic proteins expressed are difficult to express as functional proteins	Multi-domain eukaryotic proteins expressed are difficult to express as functional proteins

Over the last 20 years, some of these high-GC Gram-positive bacteria have been studied extensively as an alternative expression system [reviewed in Anné et al. (2014)]. *Streptomyces* is extremely well suited for the expression of DNA from other actinomycetes and genomes of high-GC content. Furthermore, due to its high innate secretion capacity, *Streptomyces* can be a better system than *E. coli* for the production of many extracellular proteins. The host of choice for secretory protein production of heterologous proteins using Streptomyces is *S. lividans*. The main reasons are its limited restriction–modification system as such avoiding the requirement to use non-methylated DNA for transformation or conjugation, and its low endogenous protease activity, when compared to many other streptomycetes (Butler et al. 1996). Based on whole-genome sequence analysis of *S. lividans* TK24 (Ruckert et al. 2015) and RNAseq analysis, *S. lividans* transcribes only a limited number of genes encoding proteases under standard growth conditions in minimal media. Only one-third of the genes encoding secreted proteases are transcribed at medium to high level (Tobias Busche and Jörn Kalinowski, personal communication). In case of cytoplasmic or membrane-bound proteases, 75 % of the encoding genes are transcribed at medium to high level.

A wide variety of host–vector systems have been developed, many of which are based on plasmid pIJ101, such as pIJ702 and pIJ486 (Kieser et al. 2000), but in addition, a large array of new vectors has been developed including replicative plasmid vectors, integrative plasmid and phage vectors, and special vectors for integrating DNA into the *Streptomyces* chromosome [for an overview see Rebets et al. (2016)]. As protoplast transformation with *Streptomyces* is time-consuming, conjugative plasmids are most often used for cloning purposes.

A number of heterologous prokaryotic and eukaryotic proteins have been successfully produced to economically interesting yields. For example, the L-Lysine α-oxidase (LysOX) gene from *Trichoderma viride,* a homodimeric 112-kDa flavoenzyme LysOX, was cloned and heterologously expressed in *S. lividans* TK24

with an enzyme activity up to 9.8 U/mL. Cel6A-(His)6 was secreted in *S. lividans* supernatant after 84 h of cultivation amounted to 5.56 U/mL. The maximum expression level of Cel6A-(His)6 in *S. lividans* supernatant reached up to 173 mg/L after 84 h of cultivation (Li et al. 2013). Using the promoter and signal sequence of subtilisin inhibitor of *S. venezuelae* CBS762.70 (Van Mellaert et al. 1998) yields of up to 300 mg/mL biologically active mouse TNFα could be obtained and monomeric red fluorescent protein yielded up to 500 mg/mL. In some cases, proteins which could hardly or not be produced in *B. subtilis* or *E. coli* such as e.g., xyloglucanase from *Jonesia* sp. (Sianidis et al. 2006) and CelA from *Rhodothermus marinus* (Halldórsdóttir et al. 1998) were successfully produced as secreted proteins with *S. lividans*. *Mycobacterium* Ag85A produced in *S. lividans* used in combination with rCFP-10, rESAT-6, rAPA, rPstS-1 obtained via *E. coli* heterologous production in an ELISA multi-antigen was shown to be an efficient, complementary tool for the diagnosis of active pulmonary tuberculosis (Ayala et al. 2015). In other cases, however, only low yields could be obtained, a phenomenon also experienced with other expression systems. For the examples mentioned above, the Sec-dependent secretion pathway was used. For a more complete overview of heterologous proteins secreted using recombinant *S. lividans*, see Anné et al. (2012). After the detection of the Tat-dependent pathway in bacteria, one was convinced that the latter pathway could be a solution for the production of heterologous proteins not or hardly produced via the Sec pathway. However, so far this hope has not been materialized. This does not mean, of course, that the Tat pathway cannot be used for the production of heterologous proteins, but it still has to be investigated at a larger scale. Surprisingly, Sec-dependent translocation in *tat* deletions mutants and especially in Δ*tatB* mutants showed an increase (Schaerlaekens et al. 2004). No real explanation for this phenomenon could be given up to now.

3.2 Corynebacterium

Other Gram-positive bacteria with high-GC content are corynebacteria. Some species such as *C. diphtheria* are important pathogens, while the majority are not pathogenic and some are industrially very important as major producers of amino acids including glutamic acid, lysine, threonine and valine (Mitsuhashi 2014), nucleotides, and vitamins. In particular, *C. glutamicum* is of major industrial importance. Corynebacteria are also able to produce large amounts of extracellular proteins despite their diderm-mycolate cell wall. *C. glutamicum* has several attractive features, making it a potentially interesting host for the production of heterologous proteins at an industrial scale: it secretes few endogenous proteins, and no proteases in the culture filtrate are detected, although a proteome analysis revealed the presence of more than 40 proteins in the culture supernatant (Hermann et al. 2001). As a consequence, *C. glutamicum* has been shown to be a valuable host for the production of heterologous proteins including functionally active human

epidermal growth factor (Date et al. 2006), thus demonstrating its potential for industrial-scale production of human proteins. Expressed proteins can be secreted through the Sec or the Tat pathway. For example, isomaltodextranase (IMD) of *Arthrobacter globiformis* and *Streptomyces mobaraensis* pro-transglutaminase (MTG) was produced via the *C. glutamicum* Tat pathway and yields could reach approximately 100 mg/L in flask cultures. This achievement implies a great potential for the industrial-scale production of proteins that are not efficiently secreted via other systems (Kikuchi et al. 2006).

3.3 Bacillus

Bacillus species are aerobic, endospore-forming, rod-shaped cells that are ubiquitously present in nature. Various *Bacillus* species including *B. subtilis, Bacillus licheniformis,* and *Bacillus amyloliquefaciens* can produce various enzymes including proteases, amylases, and lipases in amounts up to 25 g/L. These proteins are used in different industrial and household applications such as in the cleaning, paper, textile, food, and feed industry and also for bioremediation. Because of the efficient protein secretion of these bacilli, their secretion process is intensively investigated, in particular for *B. subtilis* which for this purpose is considered the model organism among Gram-positive bacteria. More insight in the fundamentals of the secretion process is meaningful to develop strains with superior secretion capacity. Whereas improvement of protein secretion was in general quite successful for homologous proteins, production of heterologous proteins was more cumbersome. The most important reasons therefore are a combination of the properties of the secretion pathway, the bacterial cell envelope, and the presence of a number of membrane-bound, cell-wall-bound, and secreted proteases (Westers et al. 2008). For example, the quality control proteases, WprA, HtrA, and HtrB, and feeding proteases, NprB, AprE, Epr, Bpr, NprE, Mpr, and VprA, quickly degrade slow-folding, or wrongly folded proteins (Pohl and Harwood 2010). Nevertheless, some heterologous prokaryotic and eukaryotic proteins could be well expressed, as quantities ranging from less than 10 μg up to more than 200 mg/L could be obtained [for an overview, see Schumann (2007), Kang et al. (2014)] Other reports mention the overproduction of α-amylase from *B. licheniformis* in a recombinant *B. subtilis* strain (Chen et al. 2015a); and the accumulation of biologically active hIL-3 in the growth medium in amounts of up to 100 mg/L (Westers et al. 2006).

3.4 Lactobacilli

Lactic acid bacteria (LAB) are a phylogenetically diverse group of Gram-positive, aerotolerant, non-spore-forming rods or cocci with a low-GC genome. They ferment carbohydrates with lactic acid as the major end-product. They are commonly

used to ferment food and as probiotics. *Lactobacilli* are part of the normal flora of humans and animals. The reason why strains of LAB, mainly lactococci and lactobacilli, are chosen as cell factories are plentiful: (1) many species are generally recognized as safe (GRAS) organisms because they are traditionally used in food products; *Lactobacillus* infections occur very rarely; if so, they are opportunistic infections, especially in immunocompromised individuals (Schlegel et al. 1998); (2) genetic tools for manipulation of LAB are well-developed; (3) strains of *Lactococcus lactis* secrete relatively few proteins and express very few membrane-bound or secreted proteases. In such strains, HtrA is the only protease that has been characterized on the extracellular surface (Poquet et al. 2000). (4) A variety of constitutive and inducible vector systems have been developed including the well-known 2-component NIsin-Controlled gene Expression system (NICE). This system derives from the auto-induced expression of nisin, an antibacterial polycyclic peptide produced by some strains of *L. lactis* (Kuipers et al. 1998). When nisin binds to the receptor NisK, a membrane-associated protein kinase, NisR becomes phosphorylated. The activated NisR then induces the nisin promoter (Mierau and Kleerebezem 2005). Small amounts of nisin are sufficient to activate the promoter. The NICE system is widely used for the expression of heterologous proteins in *L. lactis* (Mierau and Kleerebezem 2005).

Several proteins could be efficiently secreted using *L. lactis* and *Lactobacillus plantarum* as hosts as illustrated hereafter with a few examples. *S. aureus* nuclease NucA was secreted in amounts of more than 200 mg/L culture medium (Tremillon et al. 2010; Karlskas et al. 2014); the C-terminal region of staphylococcal HtrA transmembrane proteins could efficiently be produced and secreted in *L. lactis* as correctly folded proteins (Samazan et al. 2015); *L. lactis* was shown to be a suitable host to express a variety of structurally different glycoside hydrolases of LAB in their native, multi-meric form (Schwab et al. 2010); *B. subtilis* oxalate decarboxylase (Anbazhagan et al. 2013); *Thermobifida fusca* cellulases and xylanases to convert biomass to biofuels using *Lactobacillus plantarum* as a host (Morais et al. 2013); recombinant *L. lactis* was able to secrete biologically active human interferon-γ inducible protein-10 (Villatoro-Hernandez et al. 2008). Chitosanase (CsnA) and a β-mannanase (ManB) from *B. licheniformis* and *B. subtilis*, respectively, were efficiently produced in *L. plantarum* (Sak-Ubol et al. 2016). More examples can be found in an overview given by Le Loir et al. (2005). For the expression and secretion of the heterologous proteins, mentioned in these examples, different plasmids (inducible), promoters, and signal peptides have been used.

An additionally interesting aspect of recombinant lactococci is that they can be used as live vectors for the delivery of antigenic or therapeutic proteins to mucosal surfaces in the framework of the treatment of allergic, infectious, and gastrointestinal diseases. This use has the potential to elicit antigen-specific secretory immunoglobin A responses at mucosal surfaces (Pontes et al. 2011; Bermudez-Humaran et al. 2011). *L. lactis* engineered to secrete bioactive molecules such as Interleukin-10 (IL-10), an anti-inflammatory cytokine, was shown to be beneficial in the treatment of inflammatory bowel disease (IBD). *L. lactis* producing IL-10 markedly reduced the pathology of colitis in several mouse models. Another

strain expressing a Fab against TNF-α was also effective in the treatment of IBD (Vandenbroucke et al. 2010). A truncated version of the A2 antigen from *Leishmania donovani* expressed in *L. lactis* as cell wall anchored protein effectively gave induced high levels of antigen-specific serum antibodies (Yam et al. 2011). Subcutaneous immunization with live *L. lactis* expressing the LACK antigen anchored to the cell wall and *L. lactis* secreting IL-12 significantly delayed footpad swelling in *Leishmania major* infected BALB/c mice (Hugentobler et al. 2012). For a more extensive overview of protection studies with LAB vaccines, see among others in Wells and Mercenier (2008).

In addition, lactobacilli can be used for the delivery of DNA at the mucosal membrane. To improve the delivery, so-called invasive *L. lactis* strains were developed. These recombinant strains expressed *S. aureus* fibronectin-binding protein A or internalin A of *Listeria monocytogenes* (de Azevedo et al. 2015) or a mutated form thereof (Pontes et al. 2014) to increase the invasiveness of the strain and subsequent DNA delivery. Several examples showed the feasibility of this approach to elicit an immune response using DNA vaccination with *L. lactis* as a vector.

3.5 Clostridium *and* Bifidobacterium

These genera have in common that they are both anaerobic and Gram-positive. Clostridia are rod-shaped, endospore-forming bacteria with a low-GC content. *Clostridium* is mainly known for its pathogens like *Clostridium tetani, Clostridium botulinum,* and *Clostridium perfringens*, which secrete potent toxins leading to the life-threatening diseases tetanus, botulinum, and gangrene, respectively. *Clostridium difficile* is mainly a nosocomial pathogen and the causative agent of antibiotic-associated pseudomembranous enterocolitis. From the biotechnology point of view, *C. acetobutylicum* is an important producer of butanol and acetone. Pasteur was the first to report the fermentation process of butanol already in 1861 (Jones and Woods 1986). Stimulated by the First World War, the acetone and butanol fermentation gradually became a most important industrial fermentation processes until the 1950s. Then, the interest for the fermentative production of butanol and acetone wasted away because of cheap crude oil prices as raw material for their chemical synthesis. However, the acetone and butanol fermentation recently regained importance in the framework of renewable resources for biobutanol production. For this reason, several individual cellulosomal components and mini-cellulosomes from *C. thermocellum* and *Clostridium cellulolyticum* have been cloned and expressed in *C. acetobutylicum* and their gene products such as Cel5A, Cel8C, and Cel9M were successfully secreted into the medium. On the other hand, other cellulosomal component proteins such as Cel48F, Cel9G, and Cel9E could

not be recombinantly obtained (Mingardon et al. 2011). The development of allele-coupled exchange (ACE) (Heap et al. 2012) for *Clostridium* allowed the generation of stable and iterative integrations within a relatively short period of time. As such three genes of *C. thermocellum*-derived cellulosome components inserted into the genome of *C. acetobutylicum* could be efficiently expressed, with subsequent secretion and complex formation (Kovacs et al. 2013).

Clostridium spp. came also in the focus of research for a totally different application, notably in the framework of anticancer therapy. As anaerobic bacteria survive and multiply only under anaerobic conditions, after intravenous administration they selectively colonize, if present, in the hypoxic/necrotic areas of solid tumor tissue, a consequence of inconsistent and insufficient blood flow within regions of the tumor. When administered to a tumor-bearing body, the hypoxic/necrotic zones in solid tumors are ideal niches for the growth of anaerobic bacteria, as other tissues in a body are well oxygenized. This selectivity is repeatedly demonstrated with experimental animals (Umer et al. 2012; Roberts et al. 2014). Strains tested during these experiments belong to the following species: *C. acetobutylicum*, *C. sporogenes*, an attenuated *C. novyi*-NT or *C. beijrinckii* and more recently also *C. ghonii* (Wei 2013). When multiplying in these tumor tissues, they destroy (part of) the tumor by the hydrolytic enzymes they produce. Besides, by combinatorial treatment antitumor activity can be increased by using recombinant strains in which genes for prodrug converting enzymes are cloned. Examples of such genes are nitroreductase that converts the CB1954 prodrug to an active antitumor drug (Theys et al. 2006; Heap et al. 2014) or cytosine deaminase (CDase) which converts 5-fluorocytosine to the cytotoxic drug 5-fluorouracil. Using appropriate signal peptides, CDase can be secreted in sufficient amounts to be of biological relevance as is also the case for cloned TNF-α or IL-10, cytokines with an antitumoral but also with an immune stimulating activity to combat the tumors [for an overview see Umer et al. (2012)].

Bifidobacteria are non-spore-forming, non-motile, often branched rod-shaped bacteria with a % GC value of circa 60. They are ubiquitously found in the intestines, and because they have a probiotic function they are often used in yogurt. Engineered *Bifidobacterium adolescentis* expressing endostatin (specifically inhibiting the proliferation of vessel endothelial cells stimulated by basic fibroblast growth factor, and hence also inhibiting tumor growth), when intravenously administered to tumor-bearing mice were found only in the tumor. They inhibited angiogenesis and tumor growth (Li et al. 2003). It must be mentioned that in this case endostatin was not secreted into the medium, but was expressed intracellularly. The above-mentioned results show the potential of using (recombinant) anaerobic bacteria as tumor-specific vectors to transport anticancer genes/proteins to tumor tissues.

4 Bottlenecks in Protein Secretion and Possible Remediation

To be economically interesting, recombinant strains should produce sufficient amounts of the protein of interest. However, in many cases concentrations are low or too low. Various reasons could be at the root of the problem (see Table 2). As a consequence, several approaches can be followed in an attempt to increase the yield: from strain engineering at several levels up to fermentation optimization and this using either rather empirical approaches up to more sophisticated ones, applying state-of-the-art technologies (Fig. 2). In the following paragraphs, several examples will be used to illustrate these possibilities.

4.1 Modulation of Components of the Protein Secretion Pathway

It is evident that the promoter is of utmost importance for high expression levels. Looking into the literature, a wide variety of promoters are available for different bacteria, either constitutive or inducible. Sources of promoters vary from native to

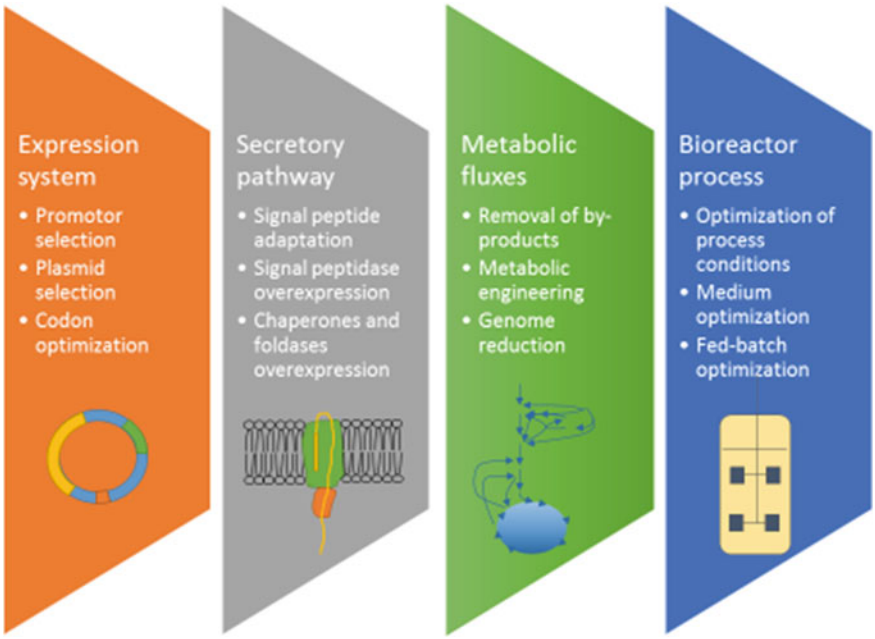


Fig. 2 Schematic overview of possible strategies for increased protein production

Table 2 Potential bottlenecks for the secretory production of recombinant proteins by Gram-positive bacteria and possible solutions

Bottleneck	Possible solutions
Unsatisfactory expression	Change promoter
	Change host strain
	Change expression vector
Translation	Optimize Ribosome Binding Site
	Use codon-optimized genes
Secretion	Modify signal peptide
	Use Tat instead of Sec secretion (or vice versa)
	Overexpress chaperones/foldases
	Overexpress signal peptidase(s)
	Overexpress Sec or Tat components
	Mutate or make hybrid SecA
	Use fusion proteins
Incorrect folding	Overexpress foldases and chaperones
Breakdown of protein of interest	Delete protease(s)
Metabolic burden	Identify problem via—omics analysis
	Bypass or solve problem by strain modification (overexpression or deletion of target gene)
	Reduce genome
Yield	Optimize fermentation
	Change carbon source and/or medium

synthetic promoter libraries. Promoter strength can be compared using reporter proteins such as β-glucuronidase (GusA) (Siegl et al. 2013), mCherry (Heiss et al. 2016) and others. To have a recent overview of different promoters used for Streptomycetes, we refer to Rebets et al. (2016).

The availability of strong promoters does not guarantee that the protein will be produced at sufficient levels, because bottlenecks are mainly at the secretion level and more downstream. Therefore, several approaches have been attempted related to the protein secretion pathway itself (Fig. 1).

4.1.1 Signal Peptide Adaptation

It is not clear what determines the sequence of an "efficient" signal peptide; therefore, several approaches are being investigated including single amino acid replacements in the N-terminal region of the signal peptide (Lammertyn and Anné 1998) or testing large Sec-type signal peptide libraries (Mathiesen et al. 2009; Degering et al. 2010). Considerable differences exist between different signal peptides but also for different mature proteins for the same set of signal peptides and this independently of the host tested. Amino acid extension by which the amino

acids in the neighborhood of the signal peptidase cleavage site are conserved could also be helpful (Sevillano et al. 2016).

4.1.2 Signal Peptidase Overexpression

As explained above, several proteins constitute the protein secretion pathway with proteins different for the Sec- and Tat-dependent pathway. What they have in common is the signal peptidase type I, an enzyme needed to release the signal peptide from the mature protein upon translocation. Most Gram-negative bacteria have only 1 chromosomally encoded signal peptidase I, but some have more. For example, *Pseudomonas aeruginosa* has two (LepB and PA1303), each with a different role in virulence and physiology (Waite et al. 2012). On the other hand, with the exception of i.a. *Streptococcus pneumoniae*, *M. tuberculosis,* and *S. aureus,* many Gram-positive bacteria have more than 1 SPase I with a maximum for *B. subtilis* which contains 5 chromosomally encoded SPase I genes (*sip*); namely *sipT*, *sipS*, *sipU*, *sipV,* and *sipW*. In addition, various *B. subtilis* strains contain in addition 2 plasmid encoded (*sipP*) SPase I genes (van Roosmalen et al. 2004). SipS and SipT are key to preprotein processing, while SipU, SipV, and SipW appear to play minor roles in protein secretion (Tjalsma et al. 1998). Overexpression of signal peptidases resulted in an increased level of secretion (Pummi et al. 2002). Similarly, in *Bacillus megaterium* MS941, co-overexpression of its unique signal peptidase SipM increased the heterologously expressed *Leuconostoc mesenteroides* dextransucrase (Malten et al. 2005).

S. lividans has 4 chromosomally encoded SPases I (SipW, SipX, SipY, and SipZ) (Parro et al. 1999). None of the individual SPases I was found to be essential for cell viability, indicating they have an overlapping substrate specificity. Nevertheless, SipY was shown to be the major SPase as the secretome of a SipY-deficient strain is severely affected (Palacin et al. 2002) on growth as well as on morphology (Gullón et al. 2012). Moreover, in particular cases the SipY mutant was shown to have some interesting advantages compared to the wild-type *S. lividans* for the overproduction of extracellular agarose probably as a consequence of the diminished extracellular proteolytic activity (Gabarró et al. 2016). Alternatively, co-overexpression of all 4 *sip* genes led to the highest increase in total preprotein processing capacity of the cell, and also to a higher amount of extracellular human CC16. It can thus be concluded that for *S. lividans* both overexpression and inactivation of individual Sip proteins can be advantageous for yield improvement of secretory proteins (Geukens 2002).

4.1.3 Overexpression of Chaperones and Foldases

When heterologous proteins are expressed, they ought to obtain their correct conformation, both for activity and stability, as incorrectly folded proteins are more prone to proteases and aggregation. Correct conformation is obtained with the help

of chaperones. Folding facilitators are, for example, the DnaK chaperone (DnaK, DnaJ, and the nucleotide-exchange factor GrpE) and GroEL/ES (mainly studied in *E. coli*) which assist the folding of newly synthesized proteins and prevent protein aggregation. Following secretion, peptidyl-prolyl cis/trans isomerases (PPIases) and disulfide bond formation proteins (Dsb) are needed for formation and rearrangement of disulfide bonds. Six Dsb (A-G) have been identified in *E. coli*. In Gram-positive bacteria, this folding process is hardly investigated except for *Bacillus*. The main components responsible for secretory protein folding and quality control in *B. subtilis* are summarized in Sarvas et al. (2004). The lipoprotein PrsA, a putative peptidyl-prolyl cis/trans isomerase, plays a major role in protein secretion by helping the posttranslocational extracellular folding of several secreted proteins. The presence of the extracytoplasmic enzymes thiol-disulfide oxidoreductases (TDOR) in *B. subtilis* were identified based on data searches. They were named BdbA (YolI), BdbB (YolK), BdbC (YvgU), and BdbD (YvgV) (Kouwen and van Dijl 2009). It was shown that BdbB and BdbC are involved in the folding of tested proteins including PhoA and A13i-Bla (Bolhuis et al. 1999). Overexpression of chaperones is considered an attractive approach to increase yield of heterologous proteins (Mogk et al. 2002). Overexpression of *prsA* in *Bacillus,* for example, increased the secretion of α-amylases, recombinant protective antigen, and a protease (Williams et al. 2003; Vitikainen et al. 2005; Chen et al. 2015b). Overexpression of the *B. subtilis* TDOR genes, however, did not improve the folding of the secreted heterologous proteins as investigated with PhoA. On the other hand, overexpression of the DsbA from *S. aureus* or the *S. carnosus* DsbA allowed the secretion of active PhoA at elevated levels (Kouwen and van Dijl 2009). Folding modulators in other Gram-positive bacteria and their impact on heterologous protein production have not yet been investigated. Based on homology searches in the genome of the *S. lividans* TK24, chaperones and peptidyl-prolyl cis/trans isomerases have been identified (Tobias Busche and Jörn Kalinowski, personal communication), but their effect on (heterologous) protein secretion has still to be investigated.

4.1.4 Sec Components

The Sec translocase consists of the integral membrane complex SecYEG, the ATPase SecA, and two additional membrane proteins that promote the release of the mature peptide across the cytoplasmic membrane (SecD and SecF). In *E. coli,* SecD and SecF are two separate membrane proteins, whereas in *B. subtilis* they are present as one polypeptide, named SecDF (Bolhuis et al. 1998). It is required to maintain a high capacity for secretion. It is not essential, but its deletion results in low-temperature sensitivity, aberrant cell division, and impaired protein secretion. The *secDF* deletion mutant exhibits a reduced level of secreted proteins (Vorös et al. 2014). Few attempts have been made to modulate specific Sec proteins to improve protein secretion. One example is the co-expression in *B. subtilis* of the *E. coli* SecB and a hybrid SecA of *B. subtilis* in which the 32 C-terminal amino

acids end was replaced by the corresponding SecA fragment of *E. coli* (Diao et al. 2012). This artificial protein targeting pathway led to a significant increase in the secretion of 2 model proteins tested, mutant maltose binding protein (MalE11) and alkaline phosphatase (PhoA), which *B. subtilis* could hardly export using the native secretion pathway. Kakeshita et al. (2010) deleted 61 amino acids of the C-terminus of SecA, a region, known to bind SecB in *E. coli.* In Gram-positive bacteria; however, SecB is absent, and the C-terminal region is not essential for protein secretion nor for growth. Moreover, the 61 amino acid deletion dramatically increased the extracellular production of the heterologous proteins alkaliphilic *Bacillus* sp. thermostable alkaline cellulase (Egl-237) and human interferon a (hIFN-a2b) in *B. subtilis*. Differential expression of SecA demonstrates that various precursors may exhibit major differences in their dependency on the amount of functional SecA in the cell (Leloup et al. 1999). In some cases, therefore, SecA overexpression or mutation might be beneficial for improved protein secretion as shown for cutinase in *B. subtilis* (Brockmeier 2006).

4.1.5 Tat Translocon Overexpression

The Tat pathway represents an alternative pathway for the production of secreted recombinant proteins, in particular for proteins prefolded in the cytoplasm. Notwithstanding this particular property, this pathway is so far not much explored for the industrial production of (heterologous) proteins. Several reasons account for this: Protein yield of Tat-exported proteins is in general substantially lower than of Sec-secreted proteins, and much of the synthesized proteins is retained in the cytoplasm (DeLisa et al. 2004). This might be a consequence of the fact that the export machinery becomes easily saturated not only by overexpressed target proteins, but even for native Tat-exported proteins (Barrett et al. 2003). The saturation of the export machinery can partially be relieved by co-expression of proteins of the Tat translocon. The stoichiometry of the TatABC components seems, however, critical for export function. For example, in *E. coli* overexpression of *tatB* resulted in complete loss of Tat transport, overexpression of *tatA* has a less severe but nonetheless significant effect on translocation (Sargent et al. 1999), while high expression of *tatC* can relieve saturation of the Tat pathway (DeLisa et al. 2004). Therefore, most attempts to relieve the saturation problem of the Tat translocon have been done by the coordinated overexpression of TatABC. This can certainly have a positive effect on the secretion of Tat-dependent proteins as illustrated for different organisms both Gram-positive and Gram-negative bacteria. When TatABC were overproduced in *S. lividans*, a fivefold increased xylanase C secretion was noticed. Surprisingly, the overproduction of TatABC in *S. lividans* caused a strong reduction in the secretion of the monitored Sec-dependent substrates (De Keersmaeker et al. 2006), suggesting a possible cross talk between the Tat- and Sec-dependent protein secretion pathway. Also for *C. glutamicum* overexpression of Tat components dramatically increased the secretion of *Chryseobacterium proteolyticum* pro-protein glutaminase (pro-PG) and *Streptomyces mobaraensis*

pro-transglutaminase (pro-TG). The amounts of secreted pro-PG were more than threefold higher when TatC or TatAC was overexpressed, and there was a further threefold increase when TatABC were overexpressed (Kikuchi et al. 2009). More recently, Albiniak et al. (2013) investigated the ability of an *E. coli tat* null mutant containing *B. subtilis* TatAdCd system to export the Tat-dependent model protein GFP. These cells do indeed export GFP to the periplasm with high efficiency; moreover, the protein was subsequently released into the extracellular medium during batch fermentation. The latter property was a consequence of the fact that the *E. coli tat* null mutant strain has impaired outer membrane integrity (Ize et al. 2003). Such an example shows that within Gram-positive bacteria, similar approaches can be tested to optimize secretion yield of (heterologous) proteins and to broaden the spectrum of proteins that can be produced by Gram-positive bacteria.

Tat-dependent secretion could be further increased if the phage shock protein (PspA) was overexpressed as shown for *S. lividans* (Vrancken et al. 2007), which was also true for the Gram-negative *E. coli* (DeLisa et al. 2004). The beneficial effect of PspA overproduction could be a consequence of its effector role in the maintenance of the integrity of the cytoplasmic membrane and proton motive force (Darwin 2005), the latter providing the energy in Tat-dependent protein translocation.

4.2 Omics Approaches for Enhanced Protein Secretion

Thanks to the new and fast techniques of DNA sequencing for genome analysis and RNAseq for transcriptome analysis, the availability of new methods for proteome analyses and the massive amounts of data and intelligent bioinformatics tools, it has now become more easy to have an insight at the systems-level burden caused by the overproduction of proteins, by the presence of plasmids and the biosynthesis and secretion of heterologous proteins. Stepping closer to the observed phenotype, metabolites and metabolic fluxes matter most and can be investigated using metabolomics and fluxomics techniques. Omics approaches are new drivers for rational engineering of host strains for improved fitness and increased productivity. Despite the availability of these new resources and their potentialities, so far not much research has been done in this field.

4.2.1 Transcriptomics and Proteomics

One of the first studies in which transcriptomics and proteomics studies were combined to understand the physiological and metabolic changes that occurred in high cell density cultivation (in order to obtain higher yield) was done with *E. coli* (Yoon et al. 2003). A recent study with *B. licheniformis* investigated the early responses to physical stress and nutrient starvation using integrated transcriptomics and proteomics (Voigt et al. 2014). With this approach, they were able to identify

general and specific marker proteins for different stress and starvation conditions including high protein secretion. Such markers might be interesting to follow the production process, and when needed to adapt it accordingly.

Because of its improved Tat-dependent protein secretion, the transcriptional profile of the *S. lividans pspA* mutant (see Sect. 4.1.5) was compared with the wild-type strain to see whether genes were differentially expressed in the *pspA* mutant. A number of genes were shown to be up- or downregulated in the mutant strain using a microarray screen containing all genes of *S. lividans* (Anné et al. 2014). Sixty-seven genes were twofold or more upregulated in the *pspA* mutant, while 117 genes were down-regulated. Among the genes encoding proteins for which a function is known or predicted, there are several which are linked to stress regulation (cold shock proteins, sigma factors), while others are involved in metabolic processes such as energy production and conversion and general metabolism. Among others, an increased expression of *sco6996* in the *pspA* mutant was identified. The corresponding protein SCO6996 shows some homology to the RNA polymerase sigma factor RpoE and experiments in *Salmonella* Typhimurium previously showed that RpoE can (at least partially) compensate for the lack of PspA (Becker et al. 2005). Loss of either *pspA* or *rpoE* leads to a depolarization of the membrane potential, indicating that both can affect the PMF. Moreover, PspA overproduction could partially compensate for the loss of RpoE in a *Salmonella* Typhimurium ΔrpoE strain. Furthermore, Gordon et al. (2008) recently showed that overexpression of one particular sigma factor (SigU) in *S. coelicolor* could lead to a significant alteration in the secretome. The SigU-overproducing strain secreted a much greater quantity and diversity of proteins than the wild-type strain, revealing that modification of the sigma factor expression in *S. lividans* might also affect protein secretion. Overexpression of *sco6996* led to an increased secretion of the tested proteins (XylC, eGFP) through the Tat pathway. This increase was far less pronounced than in the case of PspA overexpression, but still yielded a 20 % increase in final protein yield, which is still highly interesting. In another study, transcriptomics expression profiles of *S. lividans* TK24 strains producing the heterologous proteins human/mouse tumor necrosis factor alpha (hTNFα/mTNFα), monomeric red fluorescent protein, and xyloglucanase were compared to the corresponding control strain containing the empty vector only. Based on these analyses, a number of genes showed a significant twofold change in the recombinant strains overproducing the heterologous proteins. One gene, encoding a phosphoenolpyruvate carboxykinase (PEP carboxykinase) involved in the tricarboxylic acid (TCA) cycle and gluconeogenesis, was selected for further investigation. Overexpression of this gene in *S. lividans* TK24 hTNFα and xyloglucanase C production strains increased almost twofold the yield of recombinant hTNFα (Lule et al. 2012) and XylC in comparison with the initial production strains. Overall, these results show that a transcriptomics-based approach represents a useful tool for a rational optimization of heterologous protein secretion in *S. lividans*.

4.2.2 Metabolomics and Fluxomics

Metabolomics refers to the comprehensive analysis of small molecules produced by cellular metabolism. Metabolites inside as well as outside the cell (referred to as the endo- and exometabolome) can help to understand phenotypic behavior of recombinant strains, can be used for metabolic flux estimation, and can assist in strain development when combined with other omics data. Analysis is mostly done with mass spectrometry (MS) preceded by chromatographic separation, for which the technique of choice depends on the depth of analysis, the targeted metabolites, and the type of application [e.g., Garcia-Ochoa and Gomez (2009)]. Metabolomics, however, does not reach the same high resolution as RNAseq- or MS-based proteome analysis [e.g., Goodacre et al. (2004)]. From the vast pool of small molecules, some hundreds of metabolites can be detected in untargeted analysis but less than a hundred metabolites can usually be identified and quantified in a targeted metabolome analysis. When analyzing for intracellular metabolites, rapid quenching is required (e.g., in cold methanol, in liquid nitrogen) since their metabolite levels can quickly change (e.g., order of seconds in the central carbon metabolism) upon sampling. Quenched cells are then separated from the culture broth (e.g., centrifugation), and metabolites are extracted from the cell pellet (e.g., freeze-thawing cycles, ethanol boiling). Final derivatization follows when using GC-MS analysis. All stages need to be carefully evaluated and optimized to avoid leakage during quenching, to ensure complete extraction of metabolites, and to minimize loss of metabolites. Protocols can be found in literature but require validation prior to their application. Some examples of exo- and endometabolome analysis for Gram-positive hosts for heterologous protein expression are given in the next paragraph.

D'Huys et al. (2011) performed a comprehensive exometabolome profiling of wild-type, empty plasmid-containing and mTNFα-producing *S. lividans*. Metabolite profiles revealed that glutamate and aspartate are two important growth-determining amino acids. Cometabolization with glucose results in a high growth rate, although this fast biomass accumulation did not coincide with the highest mTNFα to biomass yield. Overflow of alanine and organic acids was typical for the fast growth phase and pointed out the imbalance in carbon and nitrogen metabolism. After depletion of aspartate and glutamate, growth slows down and mTNFα yield increases. Entering the stationary growth phase after glucose depletion, a diauxic shift toward consumption of overflow metabolites can be observed and mTNFα yield was maximal. Fed-batch processing is proposed as a strategy for tackling overflow metabolism. Based on the protocol for endometabolome analysis developed in Kassama et al. (2010), Muhamadali et al. (2015) performed a complementary endometabolome analysis which confirmed the intracellular metabolic shifts and observed organic acids and sugar overflow inside mTNFα-producing *S. lividans*.

A first example of using metabolome profiling for debottlenecking heterologous protein production is described in Korneli et al. (2012). Green fluorescent protein (GFP) production by *B. megaterium* was investigated. Large-scale bioreactor conditions are mimicked in small-scale bioreactors by intermitted feeding of

substrate, hereby inducing periods of feast and famine and resulting in a reduced process performance and product yield. Detailed time course of intracellular metabolites uncovered limitations in particular amino acids which could be resolved by supplementation of these amino acids during fermentation.

To fully understand the nature of metabolic bottlenecks and associated metabolome profiles, however, one needs to investigate metabolic fluxes in metabolic reaction networks. Fluxomics refers here to any technique applicable for this metabolic flux analysis. A genome-wide analysis of metabolic fluxes uses constrained-based metabolic modeling approaches [e.g., Lewis et al. (2012) and cited references therein], in which flux balance analysis (FBA) forms a central methodology. This FBA method is based on measured exchange rates of substrates and products, a genome-scale stoichiometric network model, steady-state assumption for intracellular metabolites, reaction flux constraints, and the optimization of a cellular objective function such as biomass growth or redox potential. This technique is tractable because of its genome-wide scope and commonly used for testing metabolic capacity of strains and for development of in silico strain engineering programs [e.g., Kim and Reed (2010), Schellenberger et al. (2011) and Wiechert (2001)]. However, exact knowledge of parallel reactions, bidirectional reactions, cycles, and flux split ratios requires ^{13}C-based metabolic flux analysis [e.g., Wiechert (2001); Zamboni et al. (2009)]. ^{13}C-based fluxomics is particularly suited for accurate flux calculations in the central carbon metabolism. Fluxes are estimated from intracellular mass isotopomer distributions in free intracellular metabolites or proteinogenic amino acids observed after feeding a ^{13}C-labeled carbon source. ^{13}C-labeling distributions will be determined by the actual reaction rates. Published flux maps are usually snapshots taken during a specific growth phase adopting a pseudo steady-state condition, but transient profiles of metabolic fluxes can also be modeled using dynamic FBA [e.g., Hjersted and Henson (2009)] or in stationary ^{13}C-based flux analysis [e.g., Wiechert and Nöh (2013)].

Genome-scale FBA was applied by D'Huys et al. (2012) to get understanding in the metabolome profiles and growth of *S. lividans* in a complex medium. In contrast to the maximum biomass formation capacity predicted from the complex medium, *S. lividans* shows suboptimal growth illustrating that rich media do not necessarily support maximum biomass growth. Overflow metabolism could not be predicted but needed to be imposed by constraints. Uptake of amino acids clearly contributed to biomass growth by augmenting the pool of available amino acids and by increasing the fluxes in the tricarboxylic (TCA) cycle. Genome-scale analysis of metabolic fluxes during human growth hormone production with *B subtilis* also showed metabolic shifts during batch fermentation on a minimal medium as well as shifts in the number of reactions that carried fluxes (Özdamar et al. 2010). A ^{13}C-based fluxomics was performed by Umakoshi et al. (2011) on batch cultures of *C. glutamicum* secreting heterologous transglutaminase (TGase). An increased flux through the pentose phosphate pathway for NADPH generation and also an increased flux through the TCA cycle augmenting the NADH/NAD and ATP/ADP ratios could be observed. This inspired the authors to increase the NADH/NAD

ratio by promoting lactate production. Elevation of the pH from 6.2 to 7.0 gave a small yet notably 1.4-fold increase in product yield.

Fluxomics can also form the foundation for rational strain engineering, i.e., for the identification of interesting gene knockouts redirecting fluxes and leading to higher yields of the desired product. Advantage of fluxomics-based strain design is that the interconnected nature of the cellular metabolic reactions is taken into account. In the broader context of LAB-based vaccine production (Oddone et al. 2009), for example, applied dynamic genome-scale FBA to identify targets for enhanced heterologous protein production in *L. lactis*. Green fluorescent protein (GFP) expression (as a model protein) could be increased with 15 % by implementing predicted gene targets.

A recent trend to increase production performance of microbial host cells is genome reduction where large segments of the genomic DNA are removed with the intension of removing metabolic ballast and increasing resources of product formation. Genome reduction efforts often focus on production of secondary metabolites [e.g., Gomez-Escribano and Bibb (2011)], but their application to heterologous protein production has also been reported by Toya et al. (2014) and Lieder et al. (2015). The genome-reduced *Pseudomonas putida* strain created exhibits a 40 % increased GFP production (Lieder et al. 2015). Toya et al. (2014) transformed a genome-reduced *B. subtilis* (Morimoto et al. 2008) to produce heterologous cellulase and investigated fluxes in the central carbon metabolism using ^{13}C fluxomics. A 1.7-fold increase in specific cellulase production rate as compared to the parental strain with empty plasmid was attained and flux maps reflects higher pentose phosphate pathway flux and thus NADPH generation, which seems to be a general requirement for enhanced recombinant protein production.

5 Fermentation

Development of a recombinant protein production process starts under laboratory conditions in small volume shake flasks. Many screening experiments for strain selection, testing of vectors, promoters and signal peptidases, and medium optimization are required. Laboratory-scale bioreactor experiments are performed for defining optimal culture conditions such as pH, dissolved oxygen, stirrer speed, and process operation (mostly batch or fed-batch). Screening can be greatly speed up by using high-throughput microbioreactor platforms. A first proof of principle for filamentous bacteria is reported for *S. coelicolor* by Sohoni et al. (2012). Rohe et al. (2012) developed a milliliter bioreactor screening platform and validated this setup for heterologous protein secretion of *Fusarium solani pisi* cutinase by *C. glutamicum*. Multiple cultures are run in parallel in a microtiter plate cultivation system (Biolector®) in which each well is stirred and dissolved oxygen, pH and biomass are monitored online. A liquid-handling robot enables swift media preparation, online dosing (e.g., for optimization of inducer concentration and time) and sampling. Auxiliary devices can be added for sample handling and online assaying

(Unthan et al. 2015). Scalability for *C. glutamicum* was proven excellent to 1-L bioreactors and even up to 20-L bioreactors. This advanced automated platform of Rohe et al. (2012) proved also be applicable for *Streptomyces*, a species with a more complex growth physiology including clump formation (J. Koepff and M. Oldiges, Forschungszentrum Jülich GmbH, personal communication).

Yield of secreted heterologous proteins and productivity are affected by the medium constitution, growth phase and fermentation time. Optimization of the medium composition can be done randomly but a more rational approach is guided by design of experiments techniques [e.g., Mandenius and Brundin (2008)]. A satisfactory recombinant protein production typically requires nutrient-rich media containing amino acids. Pozidis et al. (2001), for example, illustrate fermentation upscaling and medium selection for murine tumor necrosis factors alfa (mTNF-α) production with *S. lividans* TK24 and tested different amino acid rich media. Final biomass and heterologous protein concentration show no consistency and protein yields are better in a less efficient growth medium, even with the use of a constitutive promotor. D'Huys et al. (2011) further demonstrated that the yield of mTNF-α increases after glutamate and aspartate depletion from the nutrient-rich medium and when the biomass growth rate slows down. Secreted protein yield becomes highest in the stationary phase but the fermentation must be stopped when degradation by extracellular protease activity is observed. Although complex media are commonly used in *Streptomyces* cultivations, cells do not exploit their nutritional resources optimally toward biomass formation and by-product formation is usually observed (D'Huys et al. 2012). Media can also be defined and identification of the most essential amino acids can be a tedious job. Nowruzi et al. (2008), for example, screened for the impact of different amino acids and defined mixtures of them on the heterologous expression of recombinant human interleukin 3 (rHuIL-3) in *S. lividans* 66.

Batch and fed-batch cultivation are both industrially relevant *modi operandi* in heterologous protein production. Batch operation is simple and flexible, but controlled substrate addition in fed-batch fermentations enables metabolic control and high density growth by avoiding by-product formation related to overflow metabolism. Minimization of acetate overflow by restricting the specific biomass growth rate is a common practice in *E. coli* recombinant protein production [e.g., Eiteman and Altman (2006)]. After reaching the high cell density at the end of the substrate feeding phase, heterologous protein expression is started by inducer addition. Heterologous protein production by Gram-positive bacteria can also be favored by fed-batch operation, although the fed-batch control strategy depends on the expressed protein, the promotor used and other factors. No standardized approaches as those established for *E. coli* seem to exist. A good illustration of the diversity and complexity in fed-batch operation strategies is given in Oztürk et al. (2016). Fed-batch processes for homologous and heterologous expression by *Bacillus* are reviewed and associated fed-batch operation strategies are derived. No consensus fed-batch operation strategy could be found and feeding strategies largely depended on promoter choice.

From a practical and financial point of view, microorganisms growing as single cells are more favorable for easy fermentation. A filamentous growth morphology increases the power input requirements in aerated bioreactors, increases cooling requirements and results into clump formation which introduces diffusion limitations and biomass heterogeneity. Not all cells in pellets and clumps of *Streptomyces* are biologically active and metabolically equal (Manteca et al. 2008; Rioseras et al. 2014). Large and dense clump formation can be partially counteracted by addition of hydrophilic polymers like polyethylene glycol [e.g., Kieser et al. (2000)]. Clump formation, mycelial differentiation and programmed cell death have been linked to the production of antibiotics and are important factors for secondary metabolite production by *Streptomyces* [Rioseras et al. (2014) and cited references therein]. Studies linking heterologous protein secretion and cellular morphology are limited, but van Wezel and coworkers found that overexpression of *ssgA* in *Streptomyces* leads to more fragmented growth without substantial clump formation and increases heterologous protein yields in *S. lividans* 1326 (van Wezel et al. 2006; Sevillano et al. 2016). Morphology engineering could be a strategy for enhanced protein secretion.

Upscaling from laboratory-scale to industrial-scale bioreactors generally reduces the final product yields due to spatial gradients, oxygen transfer limitations, shear stress, medium differences, etc. Guidelines for upscaling of bioreactor processes are described in many handbooks on bioreactor process engineering or fermentation technology [e.g., Doran (2013) and McNeil (2008)]. Many empirical relations have been established to estimate important parameters like oxygen transfer coefficients in large-scale fermenters, but upscaling generally remains a trial and error process based on some generally accepted rules of thumbs [e.g., Garcia-Ochoa and Gomez (2009)]. Keeping a constant dissolved oxygen concentration for aerobic processes with non-filamentous organisms or keeping a constant impeller tip speed for filamentous organisms are good starting points for upscaling. Oxygen supply to the production biomass is of key importance, and spatial differences are common in large bioreactors. Oscillations in oxygen availability can lead to temporary metabolic shifts, by-product formation, and eventually multi-substrate growth. Kass et al. (2014) characterize these effects for *C. glutamicum* and strain robustness can be guaranteed when temporary gradients are limited to the scale of a few minutes. Metabolic robustness toward spatial gradients in large bioreactors is a desired property of a robust production strain, but effects on heterologous protein secretion are not yet characterized.

6 Conclusion

Bacteria are fast-growing organisms able to be replicate in cheap culture media. This property makes them potentially attractive cell factories in biotechnological processes such as for the production of heterologous proteins. For well-known reasons, *E. coli* remains the host of choice but cannot meet all expectations as not

all proteins are equally well produced in this Gram-negative host. Therefore, several Gram-positive bacteria are being explored and used as alternative bacterial hosts for the production of heterologous proteins. An important motivation therefore is that Gram-positive bacteria secrete proteins in the extracellular medium allowing correct folding, a problem encountered with *E. coli* in which case proteins often are precipitated in inclusion bodies impeding the purification process of the proteins to correctly folded active compounds.

The use of Gram-positive bacteria for heterologous proteins in secreted form shows mixed successes. While some proteins are produced in industrially viable amounts as secreted proteins, others give only small or disappointingly low yields. The new techniques that in recent years became available, including next-generation sequencing (NGS), RNA-seq, proteomics, metabolomics, and fluxomics combined with more advanced bioinformatics, and the improved understanding of the protein secretion pathways help to understand the cellular background that underlies production yield. Using this understanding allows rational strain engineering, possibly in combination with synthetic biology tools, and will undoubtedly broaden the applicability of Gram-positive bacteria for efficient use in protein secretion biotechnology.

Acknowledgments Part of the research leading to the results described for Streptomyces has received funding from the European Commission's Seventh Framework Program (FP7/2007–2013) under the grant agreement STREPSYNTH (Project No. 613877). Tobias Busche and Jörn Kalinowski (CeBiTec, Bielefeld, Germany), Joachim Koepff and Marco Oldiges (Forschungszentrum Jülich GmbH, Germany) are acknowledged for communication of results before publication).

References

Abdallah AM, Verboom T, Weerdenburg EM, Gey van Pittius NC, Mahasha PW, Jimenez C, Parra M, Cadieux N, Brennan MJ, Appelmelk BJ, Bitter W (2009) PPE and PE_PGRS proteins of *Mycobacterium marinum* are transported via the type VII secretion system ESX-5. Mol Microbiol 73:329–340

Ahmed A, Das A, Mukhopadhyay S (2015) Immunoregulatory functions and expression patterns of PE/PPE family members: roles in pathogenicity and impact on anti-tuberculosis vaccine and drug design. IUBMB Life 67:414–427

Ahn SJ, Qu MD, Roberts E, Burne RA, Rice KC (2012) Identification of the *Streptococcus mutans* LytST two-component regulon reveals its contribution to oxidative stress tolerance. BMC Microbiol 12:187

Albiniak AM, Matos CFRO, Branston SD, Freedman RB, Keshavarz-Moore E, Robinson C (2013) High-level secretion of a recombinant protein to the culture medium with a *Bacillus subtilis* twin-arginine translocation system in *Escherichia coli*. FEBS J 280:3810–3821

Anbazhagan K, Sasikumar P, Gomathi S, Priya HP, Selvam GS (2013) In vitro degradation of oxalate by recombinant *Lactobacillus plantarum* expressing heterologous oxalate decarboxylase. J Appl Microbiol 115:880–887

Anné J, Maldonado B, Van Impe J, Van Mellaert L, Bernaerts K (2012) Recombinant protein production and streptomycetes. J Biotechnol 158:159–167

Anné J, Vrancken K, Van Mellaert L, Van Impe J, Bernaerts K (2014) Protein secretion biotechnology in Gram-positive bacteria with special emphasis on *Streptomyces lividans*. Biochim Biophys Acta 1843:1750–1761

Ayala JC, Pimienta E, Rodríguez C, Sarzo M, Jones J, Vallín C, Guerrero A, Milanés MT, Anné J, Mellaert LV, Huygen K (2015) Assessment of an ELISA for serodiagnosis of active pulmonary tuberculosis in a Cuban population. Asian Pac J Trop Dis 5:772–778

Barrett CM, Ray N, Thomas JD, Robinson C, Bolhuis A (2003) Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in *Escherichia coli*. Biochem Biophys Res Commun 304:279–284

Becker LA, Bang IS, Crouch ML, Fang FC (2005) Compensatory role of PspA, a member of the phage shock protein operon, in rpoE mutant *Salmonella enterica* serovar Typhimurium. Mol Microbiol 56:1004–1016

Berdy J (2005) Bioactive microbial metabolites. J Antibiot 58:1–26

Berks BC (2015) The twin-arginine protein translocation pathway. Annu Rev Biochem 84:843–864

Bermudez-Humaran LG, Kharrat P, Chatel JM, Langella P (2011) Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. Microb Cell Fact 10 (Suppl 1):S4

Bhatty M, Laverde Gomez JA, Christie PJ (2013) The expanding bacterial type IV secretion lexicon. Res Microbiol 164:620–639

Biswas L, Biswas R, Nerz C, Ohlsen K, Schlag M, Schafer T, Lamkemeyer T, Ziebandt AK, Hantke K, Rosenstein R, Gotz F (2009) Role of the twin-arginine translocation pathway in Staphylococcus. J Bacteriol 191:5921–5929

Bolhuis A, Broekhuizen CP, Sorokin A, van Roosmalen ML, Venema G, Bron S, Quax WJ, van Dijl JM (1998) SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins. J Biol Chem 273:21217–21224

Bolhuis A, Venema G, Quax WJ, Bron S, van Dijl JM (1999) Functional analysis of paralogous thiol-disulfide oxidoreductases in *Bacillus subtilis*. J Biol Chem 274:24531–24538

Brockmeier U (2006) New strategies to optimize the secretion capacity for heterologous proteins in *Bacillus subtilis*. Ruhr-Universität Bochum Bochum, Germany

Burts ML, Williams WA, DeBord K, Missiakas DM (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. Proc Natl Acad Sci USA 102:1169–1174

Butler MJ, Aphale JS, Binnie C, DiZonno MA, Krygsman P, Soltes G, Walczyk E, Malek LT (1996) Cloning and analysis of a gene from *Streptomyces lividans* 66 encoding a novel secreted protease exhibiting homology to subtilisin BPN'. Appl Microbiol Biotechnol 45:141–147

Calvo RA, Kearns DB (2015) FlgM is secreted by the flagellar export apparatus in *Bacillus subtilis*. J Bacteriol 197:81–91

Castanie-Cornet MP, Bruel N, Genevaux P (2014) Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. Biochim Biophys Acta 1843:1442–1456

Chandran Darbari V, Waksman G (2015) Structural biology of bacterial type IV secretion systems. Annu Rev Biochem 84:603–629

Chatzi KE, Sardis MF, Karamanou S, Economou A (2013) Breaking on through to the other side: protein export through the bacterial Sec system. Biochem J 449:25–37

Chen I, Dubnau D (2004) DNA uptake during bacterial transformation. Nat Rev Microbiol 2:241–249

Chen J, Fu G, Gai Y, Zheng P, Zhang D, Wen J (2015a) Combinatorial Sec pathway analysis for improved heterologous protein secretion in *Bacillus subtilis*: identification of bottlenecks by systematic gene overexpression. Microb Cell Fact 14:92

Chen J, Gai Y, Fu G, Zhou W, Zhang D, Wen J (2015b) Enhanced extracellular production of alpha-amylase in *Bacillus subtilis* by optimization of regulatory elements and over-expression of PrsA lipoprotein. Biotechnol Lett 37:899–906

Chen J, Zhao L, Fu G, Zhou W, Sun Y, Zheng P, Sun J, Zhang D (2016) A novel strategy for protein production using non-classical secretion pathway in *Bacillus subtilis*. Microb Cell Fact 15:69

Cleon F, Habersetzer J, Alcock F, Kneuper H, Stansfeld PJ, Basit H, Wallace MI, Berks BC, Palmer T (2015) The TatC component of the twin-arginine protein translocase functions as an obligate oligomer. Mol Microbiol 98:111–129

Converse SE, Cox JS (2005) A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*. J Bacteriol 187:1238–1245

D'Huys PJ, Lule I, Van Hove S, Vercammen D, Wouters C, Bernaerts K, Anné J, Van Impe JF (2011) Amino acid uptake profiling of wild type and recombinant *Streptomyces lividans* TK24 batch fermentations. J Biotechnol 152:132–143

D'Huys PJ, Lule I, Vercammen D, Anné J, Van Impe JF, Bernaerts K (2012) Genome-scale metabolic flux analysis of *Streptomyces lividans* growing on a complex medium. J Biotechnol 161:1–13

Darwin AJ (2005) The phage-shock-protein response. Mol Microbiol 57:621–628

Date M, Itaya H, Matsui H, Kikuchi Y (2006) Secretion of human epidermal growth factor by *Corynebacterium glutamicum*. Lett Appl Microbiol 42:66–70

de Azevedo M, Meijerink M, Taverne N, Pereira VB, LeBlanc JG, Azevedo V, Miyoshi A, Langella P, Wells JM, Chatel JM (2015) Recombinant invasive *Lactococcus lactis* can transfer DNA vaccines either directly to dendritic cells or across an epithelial cell monolayer. Vaccine 33:4807–4812

De Keersmaeker S, Vrancken K, Van Mellaert L, Anné J, Geukens N (2007) The Tat pathway in *Streptomyces lividans*: interaction of Tat subunits and their role in translocation. Microbiology (Reading, England) 153:1087–1094

De Keersmaeker S, Vrancken K, Van Mellaert L, Lammertyn E, Anné J, Geukens N (2006) Evaluation of TatABC overproduction on Tat- and Sec-dependent protein secretion in *Streptomyces lividans*. Arch Microbiol 186:507–512

Degering C, Eggert T, Puls M, Bongaerts J, Evers S, Maurer KH, Jaeger KE (2010) Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. Appl Environ Microb 76:6370–6376

DeLisa MP, Lee P, Palmer T, Georgiou G (2004) Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. J Bacteriol 186:366–373

Diao L, Dong Q, Xu Z, Yang S, Zhou J, Freudl R (2012) Functional implementation of the posttranslational SecB-SecA protein-targeting pathway in *Bacillus subtilis*. Appl Environ Microb 78:651–659

Doran PM (2013) Bioprocess engineering principles, 2nd edn. Academic Press, London

Eiteman MA, Altman E (2006) Overcoming acetate in *Escherichia coli* recombinant protein fermentations. Trends Biotech 24:530–536

Erhardt M, Namba K, Hughes KT (2010) Bacterial nanomachines: the flagellum and type III injectisome. Cold Spring Harb Perspect Biol 2:a000299

Fan Y, Tan K, Chhor G, Butler EK, Jedrzejczak RP, Missiakas D, Joachimiak A (2015) EsxB, a secreted protein from *Bacillus anthracis* forms two distinct helical bundles. Protein Sci Publ Protein Soc 24:1389–1400

Fischer A, Kambara K, Meyer H, Stenz L, Bonetti EJ, Girard M, Lalk M, Francois P, Schrenzel J (2014) GdpS contributes to *Staphylococcus aureus* biofilm formation by regulation of eDNA release. Int J Med Microbiol 304:284–299

Forster BM, Marquis H (2012) Protein transport across the cell wall of monoderm Gram-positive bacteria. Mol Microbiol 84:405–413

Fyans JK, Bignell D, Loria R, Toth I, Palmer T (2013) The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen *Streptomyces scabies*. Mol Plant Pathol 14:119–130

Gabarró MlV, Gullón S, Vicente RL, Caminal G, Mellado RP, López-Santín J (2016) A *Streptomyces lividans* SipY deficient strain as a host for protein production: standardization of operational alternatives for model proteins. J Chem Technol Biotechnol doi:10.1002/jctb.4933

Garcia-Ochoa F, Gomez E (2009) Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview. Biotechnol Adv 27:153–176

Geukens N (2002) Characterisation of *Streptomyces lividans* signal peptidases. KU Leuven, Leuven, Belgium

Gomez-Escribano JP, Bibb MJ (2011) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4:207–215

Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB (2004) Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotech 22:245–252

Goodfellow M (2012) Phylum XXVI. Actinobacteria phyl. nov. In: Whitman WG, Goodfellow M. Kämpfer P, Busse H-J, Trujillo ME, Ludwig W, Suzuki K-i, Parte A (eds) Bergey's manual® of systematic bacteriology, vol 5. The Actinobacteria, Part A and B, p 2028

Goosens VJ, De-San-Eustaquio-Campillo A, Carballido-Lopez R, van Dijl JM (2015) A Tat menage a trois—the role of *Bacillus subtilis* TatAc in twin-arginine protein translocation. Biochim Biophys Acta 1853:2745–2753

Goosens VJ, Monteferrante CG, van Dijl JM (2014) The Tat system of Gram-positive bacteria. Biochim Biophys Acta 1843:1698–1706

Gordon ND, Ottaviano GL, Connell SE, Tobkin GV, Son CH, Shterental S, Gehring AM (2008) Secreted-protein response to sigmaU activity in *Streptomyces coelicolor*. J Bacteriol 190: 894–904

Grady LM, Michtavy J, Oliver DB (2012) Characterization of the *Escherichia coli* SecA signal peptide-binding site. J Bacteriol 194:307–316

Gullón S, Palomino C, Navajas R, Paradela A, Mellado RP (2012) Translocase and major signal peptidase malfunctions affect aerial mycelium formation in *Streptomyces lividans*. J Biotechnol 160:112–122

Halldórsdóttir S, Thórólfsdóttir ET, Spilliaert R, Johansson M, Thorbjarnardóttir SH, Palsdottir A, Hreggvidsson GO, Kristjánsson JK, Holst O, Eggertsson G (1998) Cloning, sequencing and overexpression of a *Rhodothermus marinus* gene encoding a thermostable cellulase of glycosyl hydrolase family 12. Appl Microbiol Biotechnol 49:277–284

Heap JT, Ehsaan M, Cooksley CM, Ng YK, Cartman ST, Winzer K, Minton NP (2012) Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. Nucleic Acids Res 40:e59

Heap JT, Theys J, Ehsaan M, Kubiak AM, Dubois L, Paesmans K, Mellaert LV, Knox R, Kuehne SA, Lambin P, Minton NP (2014) Spores of Clostridium engineered for clinical efficacy and safety cause regression and cure of tumors in vivo. Oncotarget 5:1761–1769

Heiss S, Hormann A, Tauer C, Sonnleitner M, Egger E, Grabherr R, Heinl S (2016) Evaluation of novel inducible promoter/repressor systems for recombinant protein expression in *Lactobacillus plantarum*. Microb Cell Fact 15:50

Hermann T, Pfefferle W, Baumann C, Busker E, Schaffer S, Bott M, Sahm H, Dusch N, Kalinowski J, Puhler A, Bendt AK, Kramer R, Burkovski A (2001) Proteome analysis of *Corynebacterium glutamicum*. Electrophoresis 22:1712–1723

Hjersted JL, Henson MA (2009) Steady-state and dynamic flux balance analysis of ethanol production by *Saccharomyces cerevisiae*. IET Syst Biol 3:167–179

Hueck CJ (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. Microbiol Mol Biol Rev 62:379–433

Hugentobler F, Yam KK, Gillard J, Mahbuba R, Olivier M, Cousineau B (2012) Immunization against *Leishmania major* infection using LACK- and IL-12-expressing *Lactococcus lactis* induces delay in footpad swelling. PLoS ONE 7:e30945

Ize B, Stanley NR, Buchanan G, Palmer T (2003) Role of the *Escherichia coli* Tat pathway in outer membrane integrity. Mol Microbiol 48:1183–1193

Jones DT, Woods DR (1986) Acetone-butanol fermentation revisited. Microbiol Rev 50:484–524

Kakeshita H, Kageyama Y, Ara K, Ozaki K, Nakamura K (2010) Enhanced extracellular production of heterologous proteins in *Bacillus subtilis* by deleting the C-terminal region of the SecA secretory machinery. Mol Biotechnol 46:250–257

Kang Z, Yang S, Du G, Chen J (2014) Molecular engineering of secretory machinery components for high-level secretion of proteins in Bacillus species. J Ind Microbiol Biotechnol 41: 1599–1607

Karamanou S, Vrontou E, Sianidis G, Baud C, Roos T, Kuhn A, Politou AS, Economou A (1999) A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. Mol Microbiol 34:1133–1145

Karlskas IL, Maudal K, Axelsson L, Rud I, Eijsink VG, Mathiesen G (2014) Heterologous protein secretion in Lactobacilli with modified pSIP vectors. PLoS ONE 9:e91125

Kass F, Junne S, Neubauer P, Wiechert W, Oldiges M (2014) Process inhomogeneity leads to rapid side product turnover in cultivation of *Corynebacterium glutamicum*. Microb Cell Fact 13:6

Kassama Y, Xu Y, Dunn WB, Geukens N, Anné J, Goodacre R (2010) Assessment of adaptive focused acoustics versus manual vortex/freeze-thaw for intracellular metabolite extraction from *Streptomyces lividans* producing recombinant proteins using GC-MS and multi-block principal component analysis. Analyst 135:934–942

Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical streptomyces genetics. The John Innes Foundation, Norwich

Kikuchi Y, Date M, Itaya H, Matsui K, Wu LF (2006) Functional analysis of the twin-arginine translocation pathway in *Corynebacterium glutamicum* ATCC 13869. Appl Environ Microb 72:7183–7192

Kikuchi Y, Itaya H, Date M, Matsui K, Wu LF (2009) TatABC overexpression improves *Corynebacterium glutamicum* Tat-dependent protein secretion. Appl Environ Microb 75: 603–607

Kim J, Reed JL (2010) OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol 4:53

Korneli C, Bolten CJ, Godard T, Franco-Lara E, Wittmann C (2012) Debottlenecking recombinant protein production in *Bacillus megaterium* under large-scale conditions—targeted precursor feeding designed from metabolomics. Biotechnol Bioeng 109:1538–1550

Kouwen TR, van Dijl JM (2009) Applications of thiol-disulfide oxidoreductases for optimized in vivo production of functionally active proteins in Bacillus. Appl Microbiol Biotechnol 85:45–52

Kovacs K, Willson BJ, Schwarz K, Heap JT, Jackson A, Bolam DN, Winzer K, Minton NP (2013) Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. Biotechnol Biofuels 6:117

Kuipers OP, de Ruyter PGGA, Kleerebezem M, de Vos WM (1998) Quorum sensing-controlled gene expression in lactic acid bacteria. J Biotechnol 64:15–21

Kuo CY, Sinha S, Jazayeri JA, Pouton CW (2009) A stably engineered, suicidal strain of listeria monocytogenes delivers protein and/or DNA to fully differentiated intestinal epithelial monolayers. Mol Pharm 6:1052–1061

Lammertyn E, Anné J (1998) Modifications of Streptomyces signal peptides and their effects on protein production and secretion. FEMS Microbiol Lett 160:1–10

Le Loir Y, Azevedo V, Oliveira SC, Freitas DA, Miyoshi A, Bermudez-Humaran LG, Nouaille S, Ribeiro LA, Leclercq S, Gabriel JE, Guimaraes VD, Oliveira MN, Charlier C, Gautier M, Langella P (2005) Protein secretion in *Lactococcus lactis*: an efficient way to increase the overall heterologous protein production. Microb Cell Fact 4:1

Leloup L, Driessen AJM, Freudl R, Chambert R, Petit-Glatron MF (1999) Differential dependence of levansucrase and alpha-amylase secretion on SecA (Div) during the exponential phase of growth of *Bacillus subtilis*. J Bacteriol 181:1820–1826

Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10:291–305

Li J-X, Zhao L-M, Wu R-J, Zheng Z-J, Zhang R-J (2013) High-Level overproduction of Thermobifida Enzyme in *Streptomyces lividans* using a novel expression vector. Int J Mol Sci 14:18629–18639

Li X, Fu GF, Fan YR, Liu WH, Liu XJ, Wang JJ, Xu GX (2003) Bifidobacterium adolescentis as a delivery system of endostatin for cancer gene therapy: selective inhibitor of angiogenesis and hypoxic tumor growth. Cancer Gene Ther 10:105–111

Lieder S, Nikel PI, de Lorenzo V, Takors R (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. Microb Cell Fact 14:23

Lule I, Maldonado B, D'Huys PJ, Van Mellaert L, Van Impe J, Bernaerts K, Anné J (2012) On the influence of overexpression of phosphoenolpyruvate carboxykinase in *Streptomyces lividans* on growth and production of human tumour necrosis factor-alpha. Appl Microbiol Biotechnol 96:367–372

Malten M, Nahrstedt H, Meinhardt F, Jahn D (2005) Coexpression of the type I signal peptidase gene sipM increases recombinant protein production and export in *Bacillus megaterium* MS941. Biotechnol Bioeng 91:616–621

Mandenius CF, Brundin A (2008) Bioprocess optimization using design-of-experiments methodology. Biotechnol Prog 24:1191–1203

Manteca A, Alvarez R, Salazar N, Yague P, Sanchez J (2008) Mycelium differentiation and antibiotic production in submerged cultures of *Streptomyces coelicolor*. Appl Environ Microbiol 74:3877–3886

Mathiesen G, Sveen A, Brurberg MB, Fredriksen L, Axelsson L, Eijsink VG (2009) Genome-wide analysis of signal peptide functionality in *Lactobacillus plantarum* WCFS1. BMC Genom 10:425

McDonough JA, Hacker KE, Flores AR, Pavelka MS Jr, Braunstein M (2005) The twin-arginine translocation pathway of *Mycobacterium smegmatis* is functional and required for the export of mycobacterial beta-lactamases. J Bacteriol 187:7667–7679

McNeil BH, Harvey LM (2008) Practical fermentation technology. Wiley, Chichester

Meissner D, Vollstedt A, van Dijl JM, Freudl R (2007) Comparative analysis of twin-arginine (Tat)-dependent protein secretion of a heterologous model protein (GFP) in three different Gram-positive bacteria. Appl Microbiol Biotechnol 76:633–642

Mierau I, Kleerebezem M (2005) 10 years of the nisin-controlled gene expression system (NICE) in *Lactococcus lactis*. Appl Microbiol Biotechnol 68:705–717

Mingardon F, Chanal A, Tardif C, Fierobe HP (2011) The issue of secretion in heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in *Clostridium aceto-butylicum* ATCC 824. Appl Environ Microb 77:2831–2838

Mitsuhashi S (2014) Current topics in the biotechnological production of essential amino acids, functional amino acids, and dipeptides. Curr Opin Biotechnol 26:38–44

Mogk A, Mayer MP, Deuerling E (2002) Mechanisms of protein folding: molecular chaperones and their application in biotechnology. ChemBioChem 3:807–814

Morais S, Shterzer N, Grinberg IR, Mathiesen G, Eijsink VG, Axelsson L, Lamed R, Bayer EA, Mizrahi I (2013) Establishment of a simple *Lactobacillus plantarum* cell consortium for cellulase-xylanase synergistic interactions. Appl Environ Microb 79:5242–5249

Morimoto T, Kadoya R, Endo K, Tohata M, Sawada K, Liu S, Ozawa T, Kodama T, Kakeshita H, Kageyama Y, Manabe K, Kanaya S, Ara K, Ozaki K, Ogasawara N (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. DNA Res 15:73–81

Muhamadali H, Xu Y, Ellis DI, Trivedi DK, Rattray NJ, Bernaerts K, Goodacre R (2015) Metabolomics investigation of recombinant mTNFalpha production in *Streptomyces lividans*. Microb Cell Fact 14:157

Nett M, Ikeda H, Moore BS (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. Nat Prod Rep 26:1362–1384

Nowruzi K, Elkamel A, Scharer JM, Cossar D, Moo-Young M (2008) Development of a minimal defined medium for recombinant human interleukin-3 production by *Streptomyces lividans* 66. Biotechnol Bioeng 99:214–222

Oddone GM, Mills DA, Block DE (2009) A dynamic, genome-scale flux model of *Lactococcus lactis* to increase specific recombinant protein expression. Metab Eng 11:367–381

Oertel D, Schmitz S, Freudl R (2015) A TatABC-type Tat translocase is required for unimpaired aerobic growth of *Corynebacterium glutamicum* ATCC13032. PLoS ONE 10:e0123413

Orfanoudaki G, Economou A (2014) Proteome-wide subcellular topologies of *E. coli* polypeptides database (STEPdb). Mol Cell Proteomics 13:3674–3687

Özdamar TH, Şentürk B, Yılmaz ÖD, Kocabaş P, Çalık G, Çalık P (2010) Bioreaction network flux analysis for human protein producing *Bacillus subtilis* based on genome-scale model. Chem Eng Sci 65:574–580

Oztürk S, Calik P, Ozdamar TH (2016) Fed-batch biomolecule production by *Bacillus subtilis*: a state of the art review. Trends Biotech 34:329–345

Palacin A, Parro V, Geukens N, Anné J, Mellado RP (2002) SipY Is the *Streptomyces lividans* type I signal peptidase exerting a major effect on protein secretion. J Bacteriol 184:4875–4880

Pallen MJ (2002) The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? Trends Microbiol 10:209–212

Parro V, Schacht S, Anné J, Mellado RP (1999) Four genes encoding different type I signal peptidases are organized in a cluster in *Streptomyces lividans* TK21. Microbiology (Reading, England) 145(Pt 9):2255–2263

Paul K, Erhardt M, Hirano T, Blair DF, Hughes KT (2008) Energy source of flagellar type III secretion. Nature 451:489–492

Payne SH, Bonissone S, Wu S, Brown RN, Ivankov DN, Frishman D, Pasa-Tolic L, Smith RD, Pevzner PA (2012) Unexpected diversity of signal peptides in prokaryotes. MBio 3:e00339-12

Pohl S, Harwood CR (2010) Heterologous protein secretion by bacillus species from the cradle to the grave. Adv Appl Microbiol 73:1–25

Pontes D, Azevedo M, Innocentin S, Blugeon S, Lefevre F, Azevedo V, Miyoshi A, Courtin P, Chapot-Chartier MP, Langella P, Chatel JM (2014) Immune response elicited by DNA vaccination using *Lactococcus lactis* is modified by the production of surface exposed pathogenic protein. PLoS ONE 9:e84509

Pontes DS, de Azevedo MS, Chatel JM, Langella P, Azevedo V, Miyoshi A (2011) *Lactococcus lactis* as a live vector: heterologous protein production and DNA delivery systems. Protein Expr Purif 79:165–175

Pop O, Martin U, Abel C, Muller JP (2002) The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous Tat translocation system. J Biol Chem 277:3268–3273

Poquet I, Saint V, Seznec E, Simoes N, Bolotin A, Gruss A (2000) HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. Mol Microbiol 35:1042–1051

Pozidis C, Lammertyn E, Politou AS, Anné J, Tsiftsoglou AS, Sianidis G, Economou A (2001) Protein secretion biotechnology using *Streptomyces lividans*: large-scale production of functional trimeric tumor necrosis factor alpha. Biotechnol Bioeng 72:611–619

Pummi T, Leskela S, Wahlstrom E, Gerth U, Tjalsma H, Hecker M, Sarvas M, Kontinen VP (2002) ClpXP protease regulates the signal peptide cleavage of secretory preproteins in *Bacillus subtilis* with a mechanism distinct from that of the Ecs ABC transporter. J Bacteriol 184:1010–1018

Real G, Pinto SM, Schyns G, Costa T, Henriques AO, Moran CP Jr (2005) A gene encoding a holin-like protein involved in spore morphogenesis and spore germination in *Bacillus subtilis*. J Bacteriol 187:6443–6453

Rebets Y, Kormanec J, Lutzhetskyy A, Bernaerts K, Anné J (2016) Cloning and expression of metagenomic DNA in *Streptomyces lividans* and its subsequent fermentation for optimized production. In: Streit W, Daniel R (eds) Metagenomics—methods and protocols, vol 1539. Springer, Berlin

Rioseras B, Lopez-Garcia MT, Yague P, Sanchez J, Manteca A (2014) Mycelium differentiation and development of *Streptomyces coelicolor* in lab-scale bioreactors: programmed cell death,

differentiation, and lysis are closely linked to undecylprodigiosin and actinorhodin production. Bioresour Technol 151:191–198

Roberts NJ, Zhang L, Janku F, Collins A, Bai RY, Staedtke V, Rusk AW, Tung D, Miller M, Roix J, Khanna KV, Murthy R, Benjamin RS, Helgason T, Szvalb AD, Bird JE, Roy-Chowdhuri S, Zhang HH, Qiao Y, Karim B, McDaniel J, Elpiner A, Sahora A, Lachowicz J, Phillips B, Turner A, Klein MK, Post G, Diaz LA Jr, Riggins GJ, Papadopoulos N, Kinzler KW, Vogelstein B, Bettegowda C, Huso DL, Varterasian M, Saha S, Zhou S (2014) Intratumoral injection of *Clostridium novyi*-NT spores induces antitumor responses. Sci Transl Med 6:249ra111

Rohe P, Venkanna D, Kleine B, Freudl R, Oldiges M (2012) An automated workflow for enhancing microbial bioprocess optimization on a novel microbioreactor platform. Microb Cell Fact 11:144

Ruckert C, Albersmeier A, Busche T, Jaenicke S, Winkler A, Friethjonsson OH, Hreggviethsson GO, Lambert C, Badcock D, Bernaerts K, Anné J, Economou A, Kalinowski J (2015) Complete genome sequence of *Streptomyces lividans* TK24. J Biotechnol 199:21–22

Saier MH Jr, Reddy BL (2015) Holins in bacteria, eukaryotes, and archaea: multifunctional xenologues with potential biotechnological and biomedical applications. J Bacteriol 197:7–17

Saint-Joanis B, Demangel C, Jackson M, Brodin P, Marsollier L, Boshoff H, Cole ST (2006) Inactivation of Rv2525c, a substrate of the twin arginine translocation (Tat) system of *Mycobacterium tuberculosis*, increases beta-lactam susceptibility and virulence. J Bacteriol 188:6669–6679

Sak-Ubol S, Namvijitr P, Pechsrichuang P, Haltrich D, Nguyen TH, Mathiesen G, Eijsink VG, Yamabhai M (2016) Secretory production of a beta-mannanase and a chitosanase using a *Lactobacillus plantarum* expression system. Microb Cell Fact 15:81

Sala A, Bordes P, Genevaux P (2014) Multitasking SecB chaperones in bacteria. Front Microbiol 5:666

Samazan F, Rokbi B, Seguin D, Telles F, Gautier V, Richarme G, Chevret D, Varela PF, Velours C, Poquet I (2015) Production, secretion and purification of a correctly folded staphylococcal antigen in *Lactococcus lactis*. Microb Cell Fact 14:104

San Roman SA, Facey PD, Fernandez-Martinez L, Rodriguez C, Vallin C, Del Sol R, Dyson P (2010) A heterodimer of EsxA and EsxB is involved in sporulation and is secreted by a type VII secretion system in *Streptomyces coelicolor*. Microbiol-Sgm 156:1719–1729

Sargent F, Stanley NR, Berks BC, Palmer T (1999) Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. J Biol Chem 274:36073–36082

Sarvas M, Harwood CR, Bron S, van Dijl JM (2004) Post-translocational folding of secretory proteins in Gram-positive bacteria. Biochim Biophys Acta 1694:311–327

Schaerlaekens K, Van Mellaert L, Lammertyn E, Geukens N, Anné J (2004) The importance of the Tat-dependent protein secretion pathway in Streptomyces as revealed by phenotypic changes in tat deletion mutants and genome analysis. Microbiology (Reading, England) 150:21–31

Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BO (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc 6:1290–1307

Schiebel E, Driessen AJ, Hartl FU, Wickner W (1991) Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. Cell 64:927–939

Schlegel L, Lemerle S, Geslin P (1998) Lactobacillus species as opportunistic pathogens in immunocompromised patients. Eur J Clin Microbiol Infect Dis 17:887–888

Schumann W (2007) Production of recombinant proteins in *Bacillus subtilis*. Adv Appl Microbiol 62:137–189

Schwab C, Sorensen KI, Ganzle MG (2010) Heterologous expression of glycoside hydrolase family 2 and 42 beta-galactosidases of lactic acid bacteria in *Lactococcus lactis*. Syst Appl Microbiol 33:300–307

Sevillano L, Vijgenboom E, van Wezel GP, Diaz M, Santamaria RI (2016) New approaches to achieve high level enzyme production in *Streptomyces lividans*. Microb Cell Fact 15:28

Shapova YA, Paetzel M (2007) Crystallographic analysis of *Bacillus subtilis* CsaA. Acta Crystallogr Sect D Biol Crystallogr 63:478–485

Shi Y, Li N, Yan Y, Wang H, Li Y, Lu C, Sun J (2012) Combined antibacterial activity of phage lytic proteins holin and lysin from *Streptococcus suis* bacteriophage SMP. Curr Microbiol 65:28–34

Sianidis G, Pozidis C, Becker F, Vrancken K, Sjoeholm C, Karamanou S, Takamiya-Wik M, van Mellaert L, Schaefer T, Anné J, Economou A (2006) Functional large-scale production of a novel Jonesia sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*. J Biotechnol 121:498–507

Siegl T, Tokovenko B, Myronovskyi M, Luzhetskyy A (2013) Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab Eng 19:98–106

Simone D, Bay DC, Leach T, Turner RJ (2013) Diversity and evolution of bacterial twin arginine translocase protein, TatC, reveals a protein secretion system that is evolving to fit its environmental niche. PLoS ONE 8:e78742

Sohoni SV, Bapat PM, Lantz AE (2012) Robust, small-scale cultivation platform for *Streptomyces coelicolor*. Microb Cell Fact 11:9

Stanley NR, Palmer T, Berks BC (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. J Biol Chem 275:11591–11596

Stanley SA, Raghavan S, Hwang WW, Cox JS (2003) Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system. Proc Natl Acad Sci USA 100:13001–13006

Stoop EJ, Bitter W, van der Sar AM (2012) Tubercle bacilli rely on a type VII army for pathogenicity. Trends Microbiol 20:477–484

Sutcliffe IC (2011) New insights into the distribution of WXG100 protein secretion systems. Antonie Van Leeuwenhoek 99:127–131

Sysoeva TA, Zepeda-Rivera MA, Huppert LA, Burton BM (2014) Dimer recognition and secretion by the ESX secretion system in *Bacillus subtilis*. Proc Natl Acad Sci USA 111: 7653–7658

Tekaia F, Gordon SV, Garnier T, Brosch R, Barrell BG, Cole ST (1999) Analysis of the proteome of *Mycobacterium tuberculosis* in silico. Tubercle 79:329–342

Theys J, Pennington O, Dubois L, Anlezark G, Vaughan T, Mengesha A, Landuyt W, Anné J, Burke PJ, Durre P, Wouters BG, Minton NP, Lambin P (2006) Repeated cycles of Clostridium-directed enzyme prodrug therapy result in sustained antitumour effects in vivo. Br J Cancer 95:1212–1219

Tjalsma H, Antelmann H, Jongbloed JD, Braun PG, Darmon E, Dorenbos R, Dubois JY, Westers H, Zanen G, Quax WJ, Kuipers OP, Bron S, Hecker M, van Dijl JM (2004) Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol Mol Biol Rev 68:207–233

Tjalsma H, Bolhuis A, van Roosmalen ML, Wiegert T, Schumann W, Broekhuizen CP, Quax WJ, Venema G, Bron S, van Dijl JM (1998) Functional analysis of the secretory precursor processing machinery of *Bacillus subtilis*: identification of a eubacterial homolog of archaeal and eukaryotic signal peptidases. Genes Dev 12:2318–2331

Toya Y, Hirasawa T, Morimoto T, Masuda K, Kageyama Y, Ozaki K, Ogasawara N, Shimizu H (2014) 13 C-metabolic flux analysis in heterologous cellulase production by *Bacillus subtilis* genome-reduced strain. J Biotechnol 179:42–49

Tremillon N, Issaly N, Mozo J, Duvignau T, Ginisty H, Devic E, Poquet I (2010) Production and purification of staphylococcal nuclease in *Lactococcus lactis* using a new expression-secretion system and a pH-regulated mini-reactor. Microb Cell Fact 9:37

Umakoshi M, Hirasawa T, Furusawa C, Takenaka Y, Kikuchi Y, Shimizu H (2011) Improving protein secretion of a transglutaminase-secreting *Corynebacterium glutamicum* recombinant strain on the basis of 13C metabolic flux analysis. J Biosci Bioeng 112:595–601

Umer B, Good D, Anné J, Duan W, Wei MQ (2012) Clostridial spores for cancer therapy: targeting solid tumour microenvironment. J Toxicol 2012:862764

Unthan S, Radek A, Wiechert W, Oldiges M, Noack S (2015) Bioprocess automation on a Mini Pilot Plant enables fast quantitative microbial phenotyping. Microb Cell Fact 14:32

Van Mellaert L, Lammertyn E, Schacht S, Proost P, Van Damme J, Wroblowski B, Anné J, Scarcez T, Sablon E, Raeymaeckers J, Van Broekhoven A (1998) Molecular characterization of a novel subtilisin inhibitor protein produced by *Streptomyces venezuelae* CBS762.70. DNA Seq 9:19–30

van Roosmalen ML, Geukens N, Jongbloed JD, Tjalsma H, Dubois JY, Bron S, van Dijl JM, Anné J (2004) Type I signal peptidases of Gram-positive bacteria. Biochim Biophys Acta 1694: 279–297

van Wezel GP, Krabben P, Traag BA, Keijser BJ, Kerste R, Vijgenboom E, Heijnen JJ, Kraal B (2006) Unlocking Streptomyces spp. for use as sustainable industrial production platforms by morphological engineering. Appl Environ Microbiol 72:5283–5288

van Wezel GP, McDowall KJ (2011) The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. Nat Prod Rep 28:1311–1333

Vandenbroucke K, de Haard H, Beirnaert E, Dreier T, Lauwereys M, Huyck L, Van Huysse J, Demetter P, Steidler L, Remaut E, Cuvelier C, Rottiers P (2010) Orally administered *L. lactis* secreting an anti-TNF Nanobody demonstrate efficacy in chronic colitis. Mucosal Immunol 3:49–56

Villatoro-Hernandez J, Loera-Arias MJ, Gamez-Escobedo A, Franco-Molina M, Gomez-Gutierrez JG, Rodriguez-Rocha H, Gutierrez-Puente Y, Saucedo-Cardenas O, Valdes-Flores J, Montes-de-Oca-Luna R (2008) Secretion of biologically active interferon-gamma inducible protein-10 (IP-10) by *Lactococcus lactis*. Microb Cell Fact 7:22

Vitikainen M, Hyyrylainen HL, Kivimaki A, Kontinen VP, Sarvas M (2005) Secretion of heterologous proteins in *Bacillus subtilis* can be improved by engineering cell components affecting post-translocational protein folding and degradation. J Appl Microbiol 99:363–375

Voigt B, Schroeter R, Schweder T, Jurgen B, Albrecht D, van Dijl JM, Maurer KH, Hecker M (2014) A proteomic view of cell physiology of the industrial workhorse *Bacillus licheniformis*. J Biotechnol 191:139–149

Vonderviszt F, Sajo R, Dobo J, Zavodszky P (2012) The use of a flagellar export signal for the secretion of recombinant proteins in Salmonella. Methods Mol Biol 824:131–143

Vorös A, Simm R, Slamti L, McKay MJ, Hegna IK, Nielsen-LeRoux C, Hassan KA, Paulsen IT, Lereclus D, Okstad OA, Molloy MP, Kolsto AB (2014) SecDF as part of the Sec-translocase facilitates efficient secretion of *Bacillus cereus* toxins and cell wall-associated proteins. PLoS ONE 9:e103326

Vrancken K, De Keersmaeker S, Geukens N, Lammertyn E, Anné J, Van Mellaert L (2007) pspA overexpression in *Streptomyces lividans* improves both Sec- and Tat-dependent protein secretion. Appl Microbiol Biotechnol 73:1150–1157

Waite RD, Rose RS, Rangarajan M, Aduse-Opoku J, Hashim A, Curtis MA (2012) Pseudomonas aeruginosa possesses two putative type I signal peptidases, LepB and PA1303, each with distinct roles in physiology and virulence. J Bacteriol 194:4521–4536

Wang G, Xia Y, Song X, Ai L (2016) Common non-classically secreted bacterial proteins with experimental evidence. Curr Microbiol 72:102–111

Wang IN, Smith DL, Young R (2000) Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol 54:799–825

Wang XJ, Yan YJ, Zhang B, An J, Wang JJ, Tian J, Jiang L, Chen YH, Huang SX, Yin M, Zhang J, Gao AL, Liu CX, Zhu ZX, Xiang WS (2010) Genome sequence of the milbemycin-producing bacterium *Streptomyces bingchenggensis*. J Bacteriol 192:4526–4527

Watanabe K, Tsuchida Y, Okibe N, Teramoto H, Suzuki N, Inui M, Yukawa H (2009) Scanning the *Corynebacterium glutamicum* R genome for high-efficiency secretion signal sequences. Microbiology (Reading, England) 155:741–750

Wei M (2013) Oncolytic clostridium ghonii strains, and methods of production and use. Google Patents

Wells JM, Mercenier A (2008) Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. Nat Rev Microbiol 6:349–362

Westbye AB, Leung MM, Florizone SM, Taylor TA, Johnson JA, Fogg PC, Beatty JT (2013) Phosphate concentration and the putative sensor kinase protein CckA modulate cell lysis and release of the *Rhodobacter capsulatus* gene transfer agent. J Bacteriol 195:5025–5040

Westers L, Dijkstra DS, Westers H, van Dijl JM, Quax WJ (2006) Secretion of functional human interleukin-3 from *Bacillus subtilis*. J Biotechnol 123:211–224

Westers L, Westers H, Zanen G, Antelmann H, Hecker M, Noone D, Devine KM, van Dijl JM, Quax WJ (2008) Genetic or chemical protease inhibition causes significant changes in the *Bacillus subtilis* exoproteome. Proteomics 8:2704–2713

Whitaker N, Berry TM, Rosenthal N, Gordon JE, Gonzalez-Rivera C, Sheehan KB, Truchan HK, VieBrock L, Newton IL, Carlyon JA, Christie PJ (2016) Chimeric coupling proteins mediate transfer of heterologous type IV effectors through the *Escherichia coli* pKM101-encoded conjugation machine. J Bacteriol 198:2701–2718

Widdick DA, Dilks K, Chandra G, Bottrill A, Naldrett M, Pohlschroder M, Palmer T (2006) The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*. Proc Natl Acad Sci USA 103:17927–17932

Wiechert W (2001) 13C metabolic flux analysis. Metab Eng 3:195–206

Wiechert W, Nöh K (2013) Isotopically non-stationary metabolic flux analysis: complex yet highly informative. Curr Opin Biotechnol 24:979–986

Willemse J, Ruban-Osmialowska B, Widdick D, Celler K, Hutchings MI, van Wezel GP, Palmer T (2012) Dynamic localization of Tat protein transport machinery components in *Streptomyces coelicolor*. J Bacteriol 194:6272–6281

Williams RC, Rees ML, Jacobs MF, Pragai Z, Thwaite JE, Baillie LW, Emmerson PT, Harwood CR (2003) Production of *Bacillus anthracis* protective antigen is dependent on the extracellular chaperone, PrsA. J Biol Chem 278:18056–18062

Yam KK, Hugentobler F, Pouliot P, Stern AM, Lalande JD, Matlashewski G, Olivier M, Cousineau B (2011) Generation and evaluation of A2-expressing *Lactococcus lactis* live vaccines against *Leishmania donovani* in BALB/c mice. J Med Microbiol 60:1248–1260

Yamada K, Sanzen I, Ohkura T, Okamoto A, Torii K, Hasegawa T, Ohta M (2007) Analysis of twin-arginine translocation pathway homologue in *Staphylococcus aureus*. Curr Microbiol 55:14–19

Yan J, Fan X, Xie J (2013) Emerging biomedicines based on bacteriophages. Crit Rev Eukaryot Gene Expr 23:299–308

Yang CK, Ewis HE, Zhang X, Lu CD, Hu HJ, Pan Y, Abdelal AT, Tai PC (2011) Nonclassical protein secretion by *Bacillus subtilis* in the stationary phase is not due to cell lysis. J Bacteriol 193:5607–5615

Yang CK, Zhang XZ, Lu CD, Tai PC (2014) An internal hydrophobic helical domain of *Bacillus subtilis* enolase is essential but not sufficient as a non-cleavable signal for its secretion. Biochem Biophys Res Commun 446:901–905

Yoon SH, Han MJ, Lee SY, Jeong KJ, Yoo JS (2003) Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. Biotechnol Bioeng 81:753–767

Zamboni N, Fendt S-M, Ruhl M, Sauer U (2009) 13C-based metabolic flux analysis. Nat Protoc 4:878–892

Surface and Exoproteomes of Gram-Positive Pathogens for Vaccine Discovery

Massimiliano Biagini, Fabio Bagnoli and Nathalie Norais

Abstract Reverse vaccinology has been very successful in the discovery of vaccine candidates against many pathogenic bacteria by integrating genome and proteome mining. This great achievement was facilitated by the complementarity of the in silico prediction of antigens and the empirical data on protein localization, expression, and immunogenicity obtained through different techniques based on electrophoresis, immunoblotting and mass spectrometry. An iterative process between information provided by DNA sequence analysis and proteomic data has been established leading to precise antigen identification. In this review, we report how the identification of surface and exoproteomes of Gram-positive pathogens have contributed to the selection of vaccine candidates. Moreover, we show how quantitative mass spectrometry is now paving the way for identifying protective antigens that play key roles during infection and represent the most promising vaccine targets.

Abbreviations

2DE	Bi-dimensional electrophoresis
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
LC	Liquid chromatography
MS	Mass spectrometry
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
ORF	Open reading frame
SRM	Selected reaction monitoring
SSTI	Skin and soft tissue infection

M. Biagini · F. Bagnoli · N. Norais (✉)
GSK Vaccines, Via Fiorentina 1, 53100 Siena, Italy
e-mail: nathalie.x.norais@gsk.com

Current Topics in Microbiology and Immunology (2017) 404:309–337
DOI 10.1007/82_2016_50

Published Online: 14 February 2017

Contents

1 Introduction

The licensing of the Bexsero® vaccine (CHMP, EMA2013; Vesikari et al. 2013), almost twenty years after the publication of the first bacterial genome (Fleischmann et al. 1995) probably represents one of the major achievements of what the genomic era applied to microbiology has brought to the medical field. The sequencing of the *Neisseria meningitidis* serogroup B genome (Tettelin et al. 2000) and the following in silico screening for the cellular localization prediction of encoded proteins allowed the identification of the three protective surface-exposed antigens (Giuliani et al. 2006; Pizza et al. 2000) that were finally included in the Bexsero® vaccine. This revolutionary approach for the identification of new vaccine candidates was referred as reverse vaccinology (Rappuoli 2001). While bioinformatic tools seem to have been extensively applied for vaccine discovery, proteomic studies leading to the identification of new vaccine candidates seem to remain marginal. A research on PubMed (http://www.ncbi.nlm.nih.gov/pubmed) with the key words "*bacteria and proteomics*" and "*bacteria, proteomics and vaccine*" gives the impression that less than 10% of the proteomic studies carried out on bacteria were dedicated to vaccine discovery and development (Fig. 1a). This observation does not reflect the real contribution of proteomics to vaccine discovery. From the first application of the reverse vaccinology, genomics and proteomics have driven the vaccine research in a very tight synergy, with each science "boosted" from the progresses of the other one. In one hand, proteomics depends on genomics to build protein sequence databases from DNA sequence analysis. On the other hand, while the sequencing of bacterial genomes has become routine, resulting in the availability of an incredible amount of sequence information, the correct annotation of protein coding regions remains challenging. Shotgun proteomics contributed to supplement genomic data by adding a new level of information for the interpretation of genomic sequences, such as the identification of protein regions that are absent from or incorrectly

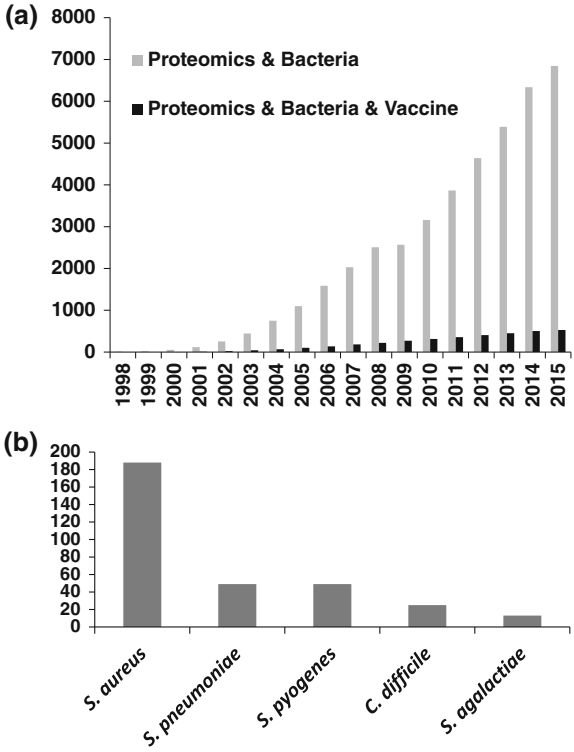


Fig. 1 Number of publications downloaded from PubMed using the key words "bacteria and proteomics" or "bacteria, proteomics and vaccine" **a** and "proteomics and different Gram-positive bacteria" **b**. PubMed: http://www.ncbi.nlm.nih.gov/pubmed, search done on June 2015

represented in current gene annotations or wrongly annotated in term of translation initiation sites from annotated ORFs. In the first period of the genomics-proteomics interplay, proteomics was also fundamental to assign cellular compartment of proteins and trained the algorithms for protein localization prediction. From this back and forth, the number of vaccine candidates selected by reverse vaccinology was considerably reduced and significant examples which have allowed the identification of vaccine candidates against pathogenic Gram-positive bacteria have been reported, such as for *Streptococcus agalactiae* (Maione et al. 2005), *Streptococcus pneumoniae* (Argondizzo et al. 2015; Donati et al. 2010; Wizemann et al. 2001), *Streptococcus pyogenes* (Mora et al. 2005; Rodriguez-Ortega et al. 2006), *Clostridium difficile* (Lawley et al. 2009; Stabler et al. 2006, 2009), *Staphylococcus aureus* (Bagnoli et al. 2015). While nowadays the algorithms for the predictions are more reliable, the main interrogation became how to assess the level of antigen expression especially in model of infection. The development of targeted proteomic tools, allowing the quantification in very complex matrices of hundreds of peptides (or proteins) selected from the genomic analysis, is paving the way to rapidly identify potentially highly expressed antigens which represent promising vaccine candidates in clinical trials. In this review, we report how the surface and the exoproteomes of Gram-positive pathogens through proteomics, and

in particular of the most studied human pathogen *S. aureus* (Fig. 1b), have contributed to vaccine discovery and what are the perspectives for the future.

2 The Secretome and Associated Bioinformatic Tools

The term "*secretome*" has been defined by Tjalsma et al. (2000) from the bioinformatic analysis of the *Bacillus subtilis* genome as the proteins predicted to belong to the secretory machinery and their "substrates". These substrates are proteins that present a cleavable amino-terminal signal peptide that drives the protein export. In Gram-positive bacteria, they are: (i) the proteins exported to the extracellular environment, (ii) those proteins, substrates of sortases, that share a determinant at their carboxyl termini for covalent attachment to the cell wall and (iii) the lipoproteins (Lpps), which remain anchored to the membrane through the attachment of a lipid moiety at their N-terminal after the removal of the amino-terminal signal peptide. Their empirical identification requires proteomic analysis of bacterial culture supernatant, cell wall, and Lpps-enriched fractions. Moreover, their accessibility on the bacteria surface could also be monitored by proteomic analysis. These analyses have generated a broad terminology which is often misused leading to a quite high confusion in the literature. Several authors proposed a harmonization of the terminology and definitions (Armengaud et al. 2012; Chagnot et al. 2013; Desvaux et al. 2009). In this review, we adopted the nomenclature that we schematized in Fig. 2. Briefly, we distinguished the surface proteome, which includes proteins accessible on the bacterial surface from the exoproteome, which includes proteins identified in the growth culture supernatant independently of their localization prediction.

The first step of a genomic analysis aimed at predicting secreted proteins consists in a screening carried out on DNA segments or contigs using databases and computer programs for defining open reading frames (ORFs). Then, the predicted ORFs are screened with dedicated algorithms to identify specific features to assign a cellular localization to the encoded putative proteins (Mora et al. 2003). The most popular software programs for cellular localization prediction are publically available and have been recently reviewed (Armengaud et al. 2012; Romine 2011). Software programs, such as SignalP (Petersen et al. 2011), are designed to identify possible peptide signals at the N-termini of proteins or transmembrane helices, stretches of particular amino acid composition, secondary structures, and disordered regions necessary for nonclassical secretion methods (Armengaud et al. 2012). PSORTb is the most commonly used software to predict localization of a protein in prokaryotes (Yu et al. 2010). It combines six analytical modules and takes into consideration sequence homology with proteins, integrating experimental datasets, and knowledge about secretion pathways (Yu et al. 2010). It assigns four localization predictions for Gram-positive bacterial proteins (cytoplasm, cytoplasmic membrane, cell wall and extracellular). Furthermore LocateP distinguishes seven subcellular localizations of proteins within Gram-positive bacteria: intracellular,

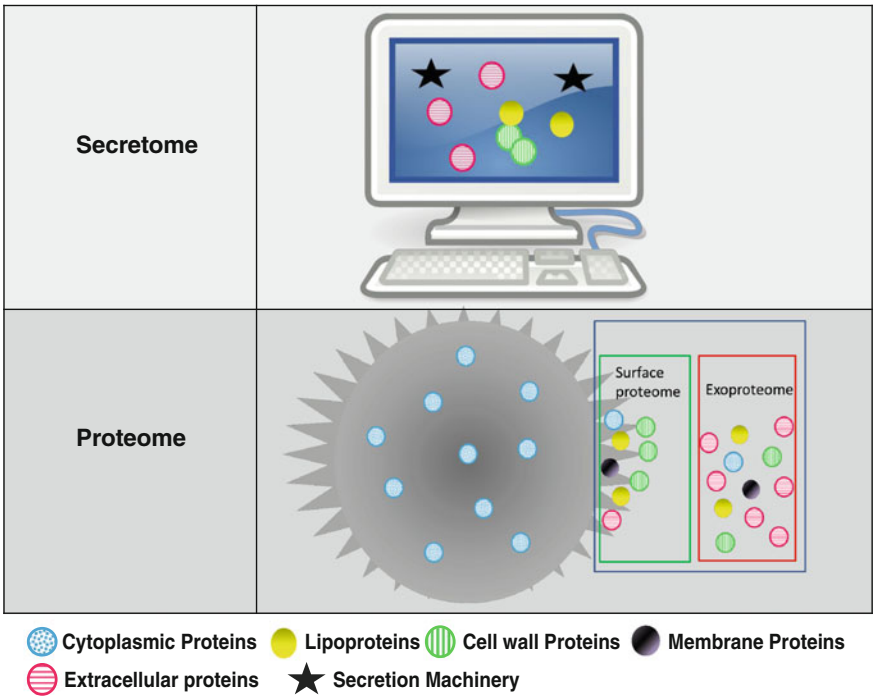


Fig. 2 Schematic representation of the terminology used in this review for predictive **a** and empirical proteins localizations **b** for Gram-positive bacteria

multi-transmembrane, N-terminally membrane anchored, C-terminally membrane anchored, lipid anchored, LPxTG-type cell wall anchored, and secreted/released proteins (Zhou et al. 2008). With the need of more and more accurate predictions, meta-analytical approaches have been recently proposed by merging the results of different algorithms and proposing a meta-scoring for the consensus prediction (Magnus et al. 2012; Sato et al. 2011). However, there is a body of experimental evidences that several typical cytoplasmic proteins could appear on the bacterial surface or extracellular medium and have a role in pathogenesis, such as adhesion, plasma protein binding, and modulation of host immune response (Henderson and Martin 2013). Some of them have been already shown to be highly immunogenic and protective in several animal models of disease (Alves et al. 2015; Feng et al. 2009). Software programs such as SecretomeP (Bendtsen et al. 2005) or VirulentPred (Garg and Gupta 2008) predict non-classical secreted virulence factors. The recent Protectome, a localization-unbiased clustering approach based on functional domain homology, has also been proposed for the identification of new vaccine candidates (Altindis et al. 2015).

3 Analysis of Surface and Exoproteomes Through the Main Proteomic Tools

The most recent and advanced proteomic analysis of a Gram-positive pathogen allowed covering about the 80% of the predicted ORFs in a single study (Karlsson et al. 2012). For long time, only a small percentage of the predicted proteins could be empirically identified. Improvement of mass spectrometers was a motor to deepen proteome analysis to a level that was thought impossible to reach at the beginning of the proteomic era. The evolution of proteomics as well as some major milestones in method development and the outcome that derived from these technological achievements was reported in Fig. 3. The three main strategies employed (bi-dimensional electrophoresis (2DE) coupled to mass spectrometry, shotgun proteomics, and targeted proteomics) and their contribution to the analysis of Gram-positive surface proteome and exoproteome are discussed below.

3.1 Bi-dimensional Electrophoresis and MALDI-TOF Mass Spectrometry

Proteomics has emerged from the combination of three technical innovations (Rabilloud et al. 2010): 2DE for the separation of complex protein samples

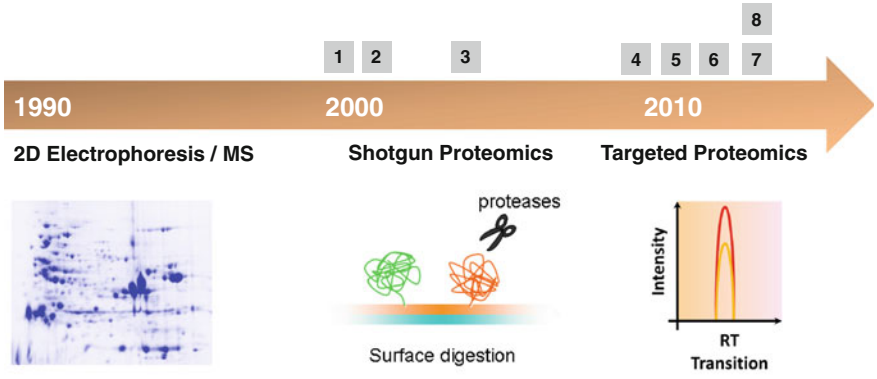


Fig. 3 Key milestones of the application of the mass spectrometry for vaccine discovery of Gram-positive bacteria. The list of the achievements reported is not exhaustive. *1* Proteome analysis of extracellular proteins through 2DE (Hirose et al. 2000). *2* Serological proteome analysis (*SERPA*) (Vytvytska et al. 2002). *3* Bacterial surface shaving (Rodriguez-Ortega et al. 2006). *4* Biotinylation of surface-exposed proteins (Becher et al. 2009). *5* Investigation of pathogen responses to the host cell environment by host cell-bacteria co-cultivation (Schmidt et al. 2010). *6* Proteome-wide selected reaction monitoring assays (Karlsson et al. 2012). *7* In vivo proteomic analysis (Diep et al. 2014). *8* Targeted analysis of data-independent acquisition MS data (Rost et al. 2014)

(Gorg et al. 1988a, b; O'Farrell 1975), the introduction of mass spectrometry-based methods (in particular MALDI-TOF MS) for protein identification through peptide mass fingerprinting (Henzel et al. 1993; James et al. 1993; Pappin et al. 1993; Yates et al. 1993), and the creation of databases and software programs that allowed the interpretation of the mass spectra by peptide mass fingerprint for protein identifications (Clauser et al. 1999; Eng et al. 1994; Perkins et al. 1999). The exoproteome of *B. subtilis* was the first to be studied through this method (Hirose et al. 2000). Soon after, analyses of the exoproteome of *Streptococcus* species (Lei et al. 2000) and *S. aureus* have been reported (Ziebandt et al. 2001). Although this method is not anymore the strategy of choice, it presents the advantage to provide a global visualization of the whole protein content of a cellular compartment. Ziebandt et al. (2010) have performed a comprehensive survey of the composition and variability of the *S. aureus* exoproteome of 25 clinical isolates revealing that only 7 out of the 63 identified exoproteins were produced by all isolates. Combined with complementary approaches, such as multilocus sequence typing and prophage profiling by multiplex sequencing, it was deduced that the observed variations were caused not only by genome plasticity, but also by an unprecedented variation in secretory protein production due to differences in transcriptional and posttranscriptional regulation (Ziebandt et al. 2010). The different secretion pattern implies that genomic studies on antigen conservation need to be complemented by the analyses of antigen expression in order ensure the selection of vaccine candidates able to induce a broad cross protection. Kush and Engelmann, by combining this information with their own results and other published data, reported that a "*pan*" exoproteome of 68 *S. aureus* clinical isolates comprises a core of 26 proteins constitutively released while 40 proteins are variably released (Kusch and Engelmann 2014; Wolf et al. 2011; Ziebandt et al. 2010).

Another approach, known as "*serological proteome analysis*" (SERPA), couples the 2DE analysis with immunoblotting and presents the advantage to "easily" monitor the immune response of infected subjects allowing the identification of antigens produced by the pathogens during the infection and presented to the host immune system (Couto et al. 2016; Hendrickson et al. 2000; Klade 2002; Klade et al. 2001). One of the first attempts was performed on *S. aureus* and led to the identification of 15 proteins including known and novel vaccine candidates (Vytvytska et al. 2002). A recent study analyzed the presence of immunogenic proteins in *S. aureus* strains isolated from patients presenting bacteraemia, skin, and soft tissue infections (SSTI), or from healthy carriers. Immunogenic proteins were identified using the corresponding patient sera. The study revealed that 12 proteins with cytoplasmic features were found to be consistently present in more than 50% of the bacteremia isolates, while none of the SSTI or healthy-carrier isolates showed any of these proteins (Liew et al. 2015).

The "first-generation" of proteomic strategies designed to identify surface proteins relied on biochemical fractionation and/or enrichment of cell wall proteins and/or Lpps. Common approaches to isolate cell wall proteins are based on their extraction by different cell wall lysis procedures that include enzymatic methods, heat treatment in presence of strong detergents (e.g., sodium dodecylsulfate) or use of LiCl

(Nandakumar et al. 2005). The main drawback of these protocols is the high level of contamination with proteins from other cellular compartments. Enrichment protocols for Lpps include their extraction from the membrane using Tween detergents. However, these procedures lead to low recovery of material not suitable for 2DE analysis (Thompson et al. 2010). The emergency of gel-free methods has contributed to the development of new methodologies that did not requested fractionations steps and offered a more accurate determination of the topology of Gram-positive proteins.

3.2 Shotgun Proteomics

After its introduction, liquid chromatography tandem mass spectrometry (LC-MS/MS)-based proteomics quickly became the method of choice for large-scale protein identification and quantification (Aebersold and Mann 2003, 2016). The approach, also called "*Shotgun Proteomics*", has been and is still an extremely powerful tool for discovery-based proteomics. It allowed the advent of two new strategies to study the surface proteome of Gram-positive bacteria, initially referred with new terminologies: "*surfome*" (mainly related to Streptococcus genus) (Doro et al. 2009; Olaya-Abril et al. 2013, 2014a; Rodriguez-Ortega et al. 2006), "*surfacome*" (mainly related to *S. aureus*) (Dreisbach et al. 2010), or "*surfaceome*" (Dalla Vecchia et al. 2014; Dreisbach et al. 2010). The term "*surface proteome*" is now preferred since it stresses that the analysis refers to the protein content and not to all the components of the bacterial cell surface (Chagnot et al. 2013). The first approach consists in the treatment of live, intact Gram-positive bacterial cells with proteases (generally trypsin or proteinase K) so that surface-exposed proteins are "shaved" and the released peptides are analyzed by LC-MS/MS for protein identification (Rodriguez-Ortega et al. 2006). This approach is particularly informative for investigating the native topology of proteins present on the bacterial surface, providing important insights on surface-exposed domains (Doro et al. 2009; Rodriguez-Ortega et al. 2006). The second method, which relays on biotinylation approaches with subsequent enrichment of labeled proteins, was mainly applied to dissect the staphylococcal surface proteome (Becher et al. 2009; Foulston et al. 2014; Moche et al. 2015). Quantitative strategy combining ^{14}N–^{15}N metabolic labeling, biotinylation, and mass spectrometry has also been reported (Hempel et al. 2010). Biotinylation reagents are usually membrane impermeable but can easily penetrate the peptidoglycan and can also label proteins that are buried within the cell wall. To avoid unspecific biotin labeling or trypsin digestion of non-surface-associated proteins, autolysis has to be prevented. The percentage of the cytoplasmic proteins identified from these studies is quite variable from laboratory to laboratory, probably reflecting both different sample preparation methodologies and sensitivity of the mass spectrometry instruments used. In shaving experiments, several conditions have been tested to counteract the osmotic stress due to membrane and cell wall impairment, such as the addition of sucrose or arabinose in the digestion buffer (Dreisbach et al. 2010). These sugars had no

stabilizing effect to reduce the lysis of pneumococci since autolysis mediated by autolysin is a natural process in this bacterial species. In this case, cytoplasmic protein contamination was significantly reduced when pneumococci were incubated in a phosphate buffer supplemented with choline (Pribyl et al. 2014), which is known to cause the release from the pneumococcal surface of choline binding protein family including autolysin (Briese and Hakenbeck 1985).

Otto et al. (2014) noted that up to date shotgun proteomic analysis of *S. aureus* allowed to confirm the expression of more than 90% of predicted Lpps and cell wall proteins as well as 80% of the other signal peptide containing proteins. As highlighted by the authors, from the first analysis performed in 2009, the increase of identification was about 10% for Lpps and cell wall proteins and more than 40% for other signal peptide containing proteins (Otto et al. 2014). Gel-free proteomic approach has been helpful for the development of prediction software and allowed proteomics to supplement genomic data for the interpretation of not annotated sequences. Referred as proteogenomics, one of the main approaches consists in identifying peptides in biological samples by searching the six-frame translations of the genome sequence, as opposed to database matching strategy usually used for protein sequence identification. This enables identification of protein regions that are absent from or incorrectly represented in current gene annotations, and thus allows improvement of the gene annotations. The identification of a novel *Bacillus cereus* exoprotein named EntD, involved in the regulation of a high number of proteins either in cellular proteome and the exoproteome could be cited as a very recent example of the proteogenomic contribution (Omer et al. 2015). In this perspective, N-terminomics is a specific proteogenomic approach that confidently predicts the N-terminus sequence of mature proteins (Armengaud 2009, 2010). Incorrect predictions of protein subcellular location are often due to the wrong initial methionine assignment in amino acid sequence of the protein. Surprisingly, from a recent large N-terminome study of the Gram-negative bacterium *Deinococcus deserti*, the 18% of the genes were wrongly annotated in terms of translation initiation sites from annotated ORFs (Baudet et al. 2010). Comparative proteogenomics is a complementary discipline that compares proteomic dataset from several related bacterial species and exploits the homology between their proteins to improve genome annotations. Combined with a quantitative approach Malmstrom et al. (2015) were able to identify several proteins differentially expressed between invasive and noninvasive *S. pyogenes* clinical isolates.

Up to date, proteomic studies on Gram-positive bacteria have almost exclusively been performed in vitro. In vivo proteomics represents a step forward for the elucidation of host-pathogen interactions and for the identification of those antigens highly expressed during bacterial infection, leading to the successfully selection of new vaccine candidates. With the attempt to identify those proteins, culture conditions mimicking stresses and challenges that the pathogens may encounter during infection have been tested. They included iron or metal deprivation (Stentzel et al. 2014), temperature variations (Alreshidi et al. 2015; Sanchez et al. 2010), osmotic or pH shocks (Karlsson et al. 2012), presence of serum (Johansson et al. 2005; Lange et al. 2008; Malmstrom et al. 2012; Sjoholm et al. 2014) or nutrient

starvation (Liebeke et al. 2011; Liu et al. 2014; Michalik et al. 2009, 2012), exposure to antibiotics (Diep et al. 2014; Fischer et al. 2011; Hessling et al. 2013), and cultivation in biofilm or planktonic status (Atshan et al. 2015). The most comprehensive survey of in vitro growth conditions explored the staphylococcal proteome variations testing nine different infection-related stresses and starvation stimuli (H_2O_2, diamide, paraquat, nitrogen monoxide, fermentative metabolism, nitrate respiration, heat shock, puromycin, mupirocin) (Fuchs et al. 2013). These studies avoided the limitations imposed by the minute amounts of bacterial sample that can be obtained from in vivo models, as well as the interference caused by host materials like proteins and/or nucleic acids (Bumann 2010).

A step forward was the innovative work that Schmidt and Völker developed to assess the proteome of bacteria internalized in eukaryotic cells, contributing to a better understanding of *S. aureus* interactions with host tissues. This procedure employs three essential steps: (i) metabolic pulse-chase labeling and infection assay; (ii) isolation of bacteria by fluorescence-assisted cell sorting; (iii) mass spectrometry-based proteome analysis (Schmidt and Volker 2011). The approach has been successively applied in three proteomic analyses of *S. aureus* co-cultivated with two different human alveolar epithelial cell lines (Pfortner et al. 2014; Surmann et al. 2014, 2015) and a macrophage cell line (Miller et al. 2011). Of interest, proteins identified from total extract of bacteria recovered from cell cultures included almost all the antigens that have been recently proposed as vaccine candidates against *S. aureus* (Bagnoli et al. 2015; Hawkins et al. 2012) (Table 1). These are three Lpps, the ferric hydroxamate-binding Lpp FhuD2 which is involved in iron uptake in early stages of invasive *S. aureus* infection (Mariotti et al. 2013; Mishra et al. 2012; Sebulsky and Heinrichs 2001), the conserved staphylococcal antigen 1A (Csa1A) which belongs to a family of proteins encoded in at least four distinct loci sharing from 54 to 91% sequence identity and immunological cross reactive (Schluepen et al. 2013), and the manganese transporter C (MntC) (Anderson et al. 2012). FhuD2, MntC and at least one of the Csa1A paralogs were identified in almost all experiments of co-cultivation of *S. aureus* with host cells. The other four selected antigens are the extracellular proteins α-hemolysin (Hla), Ess extracellular A (EsxA), and Ess extracellular B (EsxB) (Bagnoli et al. 2015) and the cell wall protein clumping factor A (ClfA) (Hawkins et al. 2012). Hla is one of the most studied staphylococcal toxins and has been recognized to be involved in the first stages of invasive and skin infections in animal models (Bubeck Wardenburg and Schneewind 2008; Kennedy et al. 2010). EsxA and EsxB are associated with abscess formation and may contribute to the persistence and diffusion of the staphylococcus in the infected host (Burts et al. 2005; Korea et al. 2014). These antigens, predicted to be released and empirically identified from growth cultures, were found tightly associated to bacteria recovered from infected host cells (Table 1). Moreover to the best of our knowledge EsxB, that has been identified from human S9 bronchial epithelial cells infected with *S. aureus* (Pfortner et al. 2014), has never been reported in any other proteomic studies. The ClfA cell wall protein, a highly conserved fibrinogen (Fg)-binding protein and virulence factor that contributes to host tissue adhesion (Hawkins et al. 2012), was also

Table 1 *S. aureus* vaccine candidates identified in bacteria recovered from co-cultivation with host cells or from bacteria grown in culture media

		Host cell-bacteria co-cultivation					Culture media growth		
		2010 Schmidt F et al.	2011 Miller M et al.	2014 Pförtner H et al.	2014 Surmann K et al.	2015 Surmann K et al.	2011 Miller M et al.	2013 Liu X et al.	2014 Pförtner H et al.
Antigen	Gene ID	Human S9 bronchial epithelial cells	Human macrohage (THB-1)	Human S9 bronchail epithelial cells	Human alveolar epithelial cells A549 and S9 kidney cell line HEK293	Human alveolar epithelial cells A549 and S9 kidney cell line HEK293	TSB	MH until mid-exp phase and treated with e or w/o oxacillin	pMEM until mid-exp phase
FhuD2	SAOUHSC_02554	•	•	•	•	•	•		•
Csa1A[a]	SAOUHSC_00052	•		•	•	•			•
MntC	SAOUHSC_00634	•		•	•	•		•	•
Hla	SAOUHSC_01121			•		•			
EsxA	SAOUHSC_00257	•	•	•	•	•		•	•
EsxB	SAOUHSC_00265			•					
CflA	SAOUHSC_00812			•	•	•			•

[a]Csa1A or one of its paralogs

identified from the total extracts of bacteria recovered from co-cultivation with host cells while it seems challenging to be identified from total extracts of bacteria grown in culture media (Table 1). Therefore, these observations suggest that proteomic studies conducted using bacteria recovered from infected tissues or cells can provide valuable information for vaccine candidate identifications.

Although they are very informative, in vitro studies might not truly represent the conditions encountered by the pathogen during infections. Thanks to the introduction of the latest generation of mass spectrometers, a new chapter of bacterial proteomics has been opened. The instruments can now overcome some of the limitations due to the low number of bacterial cells recovered from infected cell lines or organs and the high noise due to host proteins. Diep et al. (2014) reported the first characterization of the *S. aureus* proteome derived directly from infected mice. The experimental approach designed by the authors consisted in achieving sufficient enrichment of bacteria over a vast excess of kidney debris by bacterial cell sorting through *S. aureus*-specific antibodies and by enriching the sample of surface-associated proteins through the treatment of recovered bacteria with lysostaphin. Mass spectrometry analysis identified 342 *S. aureus* proteins, among which 57 were predicted to contain signal sequences for secretion. The authors noticed that while proteins involved in host cell adhesion made up 20–30% of the surface proteome in vitro, they only represented about 1% of the surface proteome identified in bacteria recovered from infected kidneys and that the majority of the in vivo-expressed surface proteins were Lpps involved in nutrient acquisition (Diep et al. 2014).

In spite of these great achievements, there is still a portion of the proteome that can be only revealed by immunoassays such as Western blot or by specific enrichment steps (Schluepen et al. 2013), and the identification of low abundance peptides remains a challenge. To reach this "unseen" proteome, Muntel et al. employed exclusion lists during MS in combination with optimized MS parameter to quantify more than 990 *S. aureus* proteins without labeling techniques (Muntel et al. 2012). Until now, most of the shotgun proteomic analyses were performed using the traditional data-dependent acquisition (DDA), for which only peptides with the strongest signal in a full-scan mass spectrum are selected for fragmentation, producing tandem mass spectra (MS/MS) that can be matched to spectra available in databases. Alternatively to the DDA, in data-independent acquisition (DIA), all peptides that can give a MS signal are analyzed. Although the DIA strategy has been known to the mass spectrometry community for more than a decade, the development of adequate software allowed its application to complex biological systems to deepen the breath of proteome providing many novel insights into many aspects of biology (Distler et al. 2014; Rost et al. 2014).

3.3 Targeted Proteomics

The development of selected reaction monitoring (SRM)-MS analysis (also called multiple reaction monitoring, MRM) has become a viable complement to shotgun

proteomics analysis. SRM is a targeted MS technology where preselected pairs of peptide precursor and derived fragment ion masses, also known as transitions, are explicitly monitored over time in a triple quadrupole MS instrument. SRM represents an attractive option for studying bacterial proteomes given that the dynamic range of their protein amount is estimated to be 4–5 orders of magnitude (Malmstrom et al. 2009), which is smaller than the linear dynamic range of SRM (Stahl-Zeng et al. 2007). The execution of SRM experiments is dependent on a priori knowledge regarding which peptides and transitions to target. This knowledge is typically obtained by creating deep proteome maps using multidimensional peptide fractionation strategies followed by LC–MS/MS analysis. From such proteome maps, proteotypic peptides (PTP's) uniquely identifying proteins of interest, and presenting strong MS signal are selected and suitable transitions are optimized (Kuster et al. 2005). The transitions are subsequently used by the mass spectrometer to isolate peptide ions in the first quadrupole, fragment them in the second quadrupole and select the fragment ions in the third one, providing a high-degree of selectivity and sensitivity for the detection of the targeted peptides. Many software programs for such complex analysis have been developed and are publically available (MacLean et al. 2010; Reiter et al. 2011). Synthetic peptides containing heavy stable isotope amino acid residues can be spiked in the analyzed samples as standards and the comparison of labeled with unlabeled peak area provides precise quantification of the endogenous analyte (Picotti and Aebersold 2012). The limited availability of SRM assays or spectral libraries is still a bottleneck for carrying out targeted analyses since their creations are time consuming and expensive. In a joined effort the teams led by Aebersold and Malmström provided to the scientific community a SRM spectral library targeting 1332 proteins of *S. pyogenes* out of 1905 ORFs by mapping more than 20,027 peptides (Karlsson et al. 2012; Rost et al. 2014). Using the spectral library to study proteomic changes that occur upon exposure of *S. pyogenes* up to 10% human plasma, the authors identified and quantified 927 proteins out of 1322 targeted proteins. Out of these, 767 proteins were quantified by more than one peptide per protein (Rost et al. 2014). A spectral library to address *S. aureus* proteome has recently been made accessible by combining 144 high precision proteomic data sets, 19,109 peptides from 2088 proteins of the *S. aureus* strain HG001 (accounting for 72% of the predicted ORFs) were identified by Depke et al. (2015). Although these data represent a considerable coverage of the ORFs of the two pathogens, a significant proportion of the predicted proteins were not identified. Specific peptides for these proteins unidentified from a shotgun proteomic approach could be predicted from the genomic sequence and their behavior in the mass spectrometer tested with a synthetic version of the peptides. A powerful alternative consists in expressing as recombinant proteins those proteins in order to identify specific peptides that can be added to the spectral library. This newest implementation give the hope to identify and quantify Gram-positive proteins in highly complex host-pathogen systems.

4 Proteins Identified on the Surface Proteome and Exoproteome of Gram-Positive Bacteria

4.1 Cell Wall Proteins

The cell wall peptidoglycan of Gram-positive bacteria functions as a surface organelle for the assembly of proteins that interact with the environment. The cell wall proteins have been mainly identified through surface proteomics and in particular by surface digestion. Hempel et al. (2011) reported in a comparative study that, from the surface of *S. aureus,* only two sortase substrate proteins were identified through the biotinylation approach, and 15 from the bacterial surface digestion, representing then 80% of the predicted sortase substrate proteins. Similar percentages were identified using the shaving approach with *S. pyogenes* (Rodriguez-Ortega et al. 2006; Severin et al. 2007), although very structured proteins such as the pilus subunits, which are stabilized by internal isopeptide bonds (Kang et al. 2007), were particularly resistant to proteolytic digestion under non denaturing conditions (Walden et al. 2014) and were underrepresented with this approach. The accessibility of the cell wall proteins varies between different bacterial species. In spite of several attempts of our group, the surface digestion of *C. difficile* failed, probably due to the presence of the crystalline bacterial cell surface layer proteins (S-layers), which represents the outermost cell envelope in these bacteria (Biagini M., unpublished data). Nevertheless, cell wall proteins represented the major fraction of the proteins identified in the culture supernatant of this Gram-positive bacterium (Cafardi et al. 2013). The mechanism by which cell wall anchored proteins are released in the culture medium is still unclear. An ultracentrifugation of the growth medium reveals that the identified cell wall proteins are not released as macrostructures leading to the hypothesis that they might be shaved or released from the bacterium (Cafardi et al. 2013), in agreement with the identification of a novel secreted metalloprotease (CD2830) that cleaves specific proline-containing sequences in *C. difficile* cell surface proteins (Cafardi et al. 2013; Hensbergen et al. 2014). Streptococcal surface proteins are also released from the bacterial surface through the action of a cysteine protease that cleaves and releases biologically active fragments of surface proteins in post-exponential phases of growth (Loughman and Caparon 2006; Rasmussen and Bjorck 2002). Nevertheless, some proteins have multiple release mechanisms. For example, staphylococcal protein A (SpA), a cell wall envelope anchored protein was shown to be released from the bacterial surface with the murein tetrapeptide-tetraglycyl-linked to its C-terminal threonyl residue (Becker et al. 2014), or with an unprocessed sorting signal (O'Halloran et al. 2015).

4.2 Lipoproteins

Lpps represent approximately 2–3% of the encoded bacterial proteome. They play a variety of biological functions, some important for cell physiology and others essential for virulence, such as adhesion, colonization, and persistence. Lpps have been described as virulence factors because they play critical roles in membrane stabilization, nutrient uptake, antibiotic resistance, bacterial adhesion to host cells, protein maturation and secretion and many of them have still unknown functions (Kovacs-Simon et al. 2011). After the translation in the cytoplasm, the prepro-lipoprotein precursor translocates across the membrane via the Sec-pathway by the recognition of the specific N-terminal signal peptide. Then in Gram-positive bacteria Lpps are processed by two key enzymes: the prolipoprotein diacylglyceryl transferase (Lgt) and the Lpp signal peptidase (Lsp). The Lgt enzyme recognizes a so-called lipobox motif in the C-terminal region of the signal peptide of a premature Lpp and transfers a diacylglyceryl moiety to the cysteine residue of the lipobox. Subsequently, the Lsp enzyme cleaves the signal peptide resulting in a mature Lpp (Hussain et al. 1982; Pfortner et al. 2014). Nevertheless, recent reports have suggested that Gram-positives lack the Gram-negative homologous apolipoprotein N-acyltransferase (Lnt) gene responsible for N-acylation (Kurokawa et al. 2009; Vidal-Ingigliardi et al. 2007). Despite that it seems that the Lpp N-terminal of some Gram-positive bacteria can also be modified with an acetyl group (Asanuma et al. 2011; Tawaratsumida et al. 2009). Chemical phase partitioning of membrane components with the detergent Triton X-114 allowed the isolation of Gram-positive Lpps (Cockayne et al. 1998). Seven *S. aureus* native Lpps were isolated and then structurally characterized using this protocol of extraction, but it required the impressive volume of 10 liters of bacterial cell culture to obtain 4 mg of Lpp fraction (Tawaratsumida et al. 2009). Interestingly, we have recently reported that the treatment of streptococci with sublethal concentration of penicillin allowed the release in the extracellular medium of membrane vesicles constituted almost exclusively of Lpps (Biagini et al. 2015) enabling a fast enrichment tool for characterization of Gram-positive native Lpps. Another interesting approach for Lpps study is the bacterial surface biotinylation, in which surface-exposed proteins were covalently labeled with biotin-conjugated sulfo-N-hydroxy-sulfosuccinimide ester cross-linker, a membrane impermeable reagent that preferentially reacts with the ε-amino group of lysine residues of surface-associated proteins. This approach was already applied in Gram-negative bacteria for the isolation of Lpps (Cullen et al. 2003) and has been successfully applied to *S. aureus* (Hempel et al. 2010). The power of this approach in the identification of Lpps has been further confirmed in a comparative study, where different approaches have been evaluated with different *S. aureus* subcellular fractions (Hempel et al. 2011). An optimized biotinylation approach and a highly specific shaving approach were applied to study the pneumococcal surface proteome, and preventing autolysis of pneumococci during incubation, nearly the 95% of the predicted Lpps were identified (Pribyl et al. 2014).

4.3 Proteins Released in the Extracellular Milieu

Most of the virulence factors are found in the extracellular milieu, independently from the way they reach out the bacterial cell. Many adhesins, that are covalently or non-covalently bound to the cell surface and mediate bacterial adhesion to a variety of host molecules (Foster et al. 2014), are often detected in the extracellular milieu due to cleavage exerted by bacterial proteases (Hensbergen et al. 2014). However the major components (not necessarily in terms of abundance) are toxins that are represented by several protein families and classified according to their different function or effect on the host (Otto 2014; Vandenesch et al. 2012): generally they are grouped in membrane-damaging factors, or molecules that interfere with host receptor functionality, degrade host molecules, and interfere with host immune defence mechanisms (Otto 2014). Among exotoxins, the staphylococcal superantigens represent a peculiar class of molecules able to activate in a nonconventional way the T-cell immune response and they are associated with food poisoning and toxic shock syndromes (Xu and McCormick 2012). Exfoliative toxins well represent the last class of toxins, acting as serine proteases in the degradation of the skin (in particular hydrolysing desmoglein and causing the Ritter's disease, also known as staphylococcal scalded skin syndrome) (Bukowski et al. 2010). Other small proteins and peptides have been identified to be released by Gram-positive bacteria as virulence factors and they are represented by Panton-Valentine leukocidins (LukS and LukF) (Shallcross et al. 2013) and phenol-soluble modulins (Li et al. 2014; Peschel and Otto 2013), recently proven to be effective vaccines in animal models of infection (Brown et al. 2009; Karauzum et al. 2013; van den Berg et al. 2015). Noteworthy other small peptides with multiple roles as virulence factors, such as lantibiotics (Alkhatib et al. 2012; Dischinger et al. 2014) and circular bacteriocins (Gabrielsen et al. 2014) have also been identified in the *S. aureus* proteome.

4.4 Cytoplasmic Predicted Proteins

The proteomic analysis of cell surface and/or medium revealed the presence of cytoplasmic proteins which do not carry pre-sequence for secretion and/or with function associated to the cytoplasm. The number of cytoplasmic proteins identified from these compartments is often proportional to the extent of bacterial lysis in the culture (Rice and Bayles 2008). The protective immunity observed with cytoplasmic proteins such as the GAPDH, a key glycolytic enzyme (Perez-Casal and Potter 2016), or the recognition of cytoplasmic proteins by sera obtained from infected subjected (Zysk et al. 2000), indicate that they are exposed to the immune system through presentation on the bacterial surface or by bacterial lysis occurring during the infection. Although a non-classical protein secretion pathway (e.g., SecA2) has been described as responsible for secretion of some proteins lacking the N-term

signal peptide (Bendtsen et al. 2005), it is not excluded that other molecular machineries are responsible for the translocation of cytoplasmic proteins. It remains extremely challenging to evaluate what is the part of potential translocation and lysis. Interestingly some of the predicted cytoplasmic proteins identified on the surface of bacteria have been demonstrated to function as adhesins and their protective function after active immunization was proved (Alves et al. 2015; Elhaik Goldman et al. 2016). For this reason, these proteins, which are also involved in energetic metabolism or genetic apparatus (enolase, phosphoglycerate kinase, DNA-directed RNA polymerase β), were described also as "*moonlighting*" to reinforce their double and diverse roles on the surface and in the cytoplasm (Henderson and Martin 2011, 2013; Kainulainen and Korhonen 2014).

4.5 Membrane Vesicles in Gram-Positive Bacteria

The release from live cells of nano-sized membrane vesicles into the extracellular environment is a common phenomenon observed in all the three domains of life: Bacteria, Archaea, and Eukarya (Kim et al. 2015). In Gram-negative bacteria, they were first observed by electron microscopy in *E. coli* since the 1960s (Knox et al. 1966) and then generally termed as outer membrane vesicles (OMV). A plethora of functions have been associated to OMV, ranging from genetic transformation, virulence factor release, decoy from antimicrobial attack and protection from cellular stress as well as biofilm formation (Manning and Kuehn 2013). Nevertheless the experimental evidence of the release of membrane vesicles (MV) in Gram-positive bacteria was found five decades later, with two studies performed on *S. aureus* (Lee et al. 2009) and *Bacillus anthracis* (Rivera et al. 2010). In *S. aureus* MV more than 90 proteins were identified by proteomic studies, including surface-associated and extracellular virulent proteins such as L-lactamase, coagulase, hemolysin, IgG-binding protein Sbi, and N-acetylmuramoyl-l-alanine amidase (Lee et al. 2009). Intriguing functional roles of MV in bacteria—bacteria and bacteria—host interactions have been suggested, in particular the association of *S. aureus* MV with the development of atopic dermatitis (Hong et al. 2011, 2014). *S. aureus* MV displayed a cytotoxic effect on various host cells and caused cell death by delivering bacterial virulent factors to host cells in a cholesterol-dependent manner (Gurung et al. 2011; Thay et al. 2013). Rivera and colleagues demonstrated that *B. anthracis* also produces MV with "*cargo*" activity acting as a means to release biologically active toxins including lethal factor, edema toxin, and anthrolysin (Rivera et al. 2010). Notably mice immunized with MV survived significantly longer than controls after *B. anthracis* challenge (Rivera et al. 2010). After these first studies MV release was discovered in other Gram-positive species such as *B. subtilis* (Brown et al. 2014), *Clostridium perfringens* (Jiang et al. 2014), *Listeria monocytogenes* (Lee et al. 2013), *Streptococcus mutans* (Liao et al. 2014), *S. pneumoniae* (Olaya-Abril et al. 2014b) and *Streptococcus suis* (Haas and Grenier 2015), *Streptomices coelicolor* (Schrempf et al. 2011), *Streptomices lividans*

(Schrempf and Merling 2015). Lipoprotein-rich Membrane Vesicles (LMV) from streptococcal species have been reported very recently (Biagini et al. 2015). Proteomic analysis of LMV revealed that they are almost exclusively constituted of Lpps. In addition, the very low abundance of proteins with multiple transmembrane domains and the presence of proteins associated to ExPortal led to the hypothesis that the LMVs are not representative of the bacteria membrane and may derive from micro-domains involved in the protein maturation and lipoprotein secretion (Biagini et al. 2015).

5 Conclusions

Several preclinical and clinical studies conducted with protein subunit-antigens from Gram-positive bacteria have evidenced that the identification of protective antigens remains a challenging task (Bagnoli et al. 2012). Poor understanding of mechanisms of immune evasion and adaptation of Gram-positive bacteria to different host niches is delaying vaccine discovery. It is likely that pathogens express and secrete proteins of unknown function that contribute to those processes supporting the concept that the successful development of efficacious vaccines may also depend on the identification of virulence factors expressed in the surface proteome and/or exoproteome in conditions mimicking host-pathogen interactions. The identification of those vaccine candidates requires to: (i) identify all ORFs of a pathogen; (ii) assign cellular localization to the encoded proteins and predict vaccine candidates; (iii) assess the level of expression of the selected antigens in animal models of infection or in models mimicking infection conditions. In this process, genomics and proteomics have driven the vaccine research in a very tight synergy and this was facilitated by the simultaneous development of tools as well as new techniques and experimental approaches. Until recently, proteomics complements genomics by validating the identification of the ORFs, and the assignment of cellular compartments of the antigens encoded by these ORFs allowing the development and release of predictive software programs of higher reliability. We are now assisting to a new phase of this synergy where the main contribution of proteomics is the quantification of predicted antigens in model of infection to prioritize vaccine candidates with high potential to be effective in clinical trials.

Funding
GSK funded all costs associated with the development and the publishing of the present manuscript.

Declaration of interest
Fabio Bagnoli, Massimiliano Biagini and Nathalie Norais are employees of GSK Vaccines. Fabio Bagnoli, Massimiliano Biagini, and Nathalie Norais own patents on *S. aureus* vaccine candidates. Fabio Bagnoli declared stock ownership in the GSK group of companies. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or

financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Authorship

Massimiliano Biagini, Nathalie Norais and Fabio Bagnoli were involved in the conception and design of the manuscript. All authors were involved in writing the manuscript or revising it critically for important intellectual content. All authors approved the manuscript before it was submitted.

References

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422:198–207

Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. Nature 537:347–355

Alkhatib Z, Abts A, Mavaro A, Schmitt L, Smits SH (2012) Lantibiotics: how do producers become self-protected? J Biotechnol 159:145–154

Alreshidi MM, Dunstan RH, Macdonald MM, Smith ND, Gottfries J, Roberts TK (2015) Metabolomic and proteomic responses of *Staphylococcus aureus* to prolonged cold stress. J Proteomics 121:44–55

Altindis E, Cozzi R, Di Palo B, Necchi F, Mishra RP, Fontana MR, Soriani M, Bagnoli F, Maione D, Grandi G, Liberatori S (2015) Protectome analysis: a new selective bioinformatics tool for bacterial vaccine candidate discovery. Mol Cell Proteomics 14:418–429

Alves J, Madureira P, Baltazar MT, Barros L, Oliveira L, Dinis-Oliveira RJ, Andrade EB, Ribeiro A, Vieira LM, Trieu-Cuot P, Duarte JA, Carvalho F, Ferreira P (2015) A safe and stable neonatal vaccine targeting GAPDH confers protection against group B Streptococcus infections in adult susceptible mice. PLoS ONE 10:e0144196

Anderson AS, Scully IL, Timofeyeva Y, Murphy E, McNeil LK, Mininni T, Nunez L, Carriere M, Singer C, Dilts DA, Jansen KU (2012) *Staphylococcus aureus* manganese transport protein C is a highly conserved cell surface protein that elicits protective immunity against *S. aureus* and *Staphylococcus epidermidis*. J Infect Dis 205:1688–1696

Argondizzo AP, da Mota FF, Pestana CP, Reis JN, de Miranda AB, Galler R, Medeiros MA (2015) Identification of proteins in *Streptococcus pneumoniae* by reverse vaccinology and genetic diversity of these proteins in clinical isolates. Appl Biochem Biotechnol 175: 2124–2165

Armengaud J (2009) A perfect genome annotation is within reach with the proteomics and genomics alliance. Curr Opin Microbiol 12:292–300

Armengaud J (2010) Proteogenomics and systems biology: quest for the ultimate missing parts. Expert Rev Proteomics 7:65–77

Armengaud J, Christie-Oleza JA, Clair G, Malard V, Duport C (2012) Exoproteomics: exploring the world around biological systems. Expert Rev Proteomics 9:561–575

Asanuma M, Kurokawa K, Ichikawa R, Ryu KH, Chae JH, Dohmae N, Lee BL, Nakayama H (2011) Structural evidence of alpha-aminoacylated lipoproteins of *Staphylococcus aureus*. FEBS J 278:716–728

Atshan SS, Shamsudin MN, Sekawi Z, Thian Lung LT, Barantalab F, Liew YK, Alreshidi MA, Abduljaleel SA, Hamat RA (2015) Comparative proteomic analysis of extracellular proteins expressed by various clonal types of *Staphylococcus aureus* and during planktonic growth and biofilm development. Front Microbiol 6:524

Bagnoli F, Bertholet S, Grandi G (2012) Inferring reasons for the failure of *Staphylococcus aureus* vaccines in clinical trials. Front Cell Infect Microbiol 2:16

Bagnoli F, Fontana MR, Soldaini E, Mishra RP, Fiaschi L, Cartocci E, Nardi-Dei V, Ruggiero P, Nosari S, De Falco MG, Lofano G, Marchi S, Galletti B, Mariotti P, Bacconi M, Torre A, Maccari S, Scarselli M, Rinaudo CD, Inoshima N, Savino S, Mori E, Rossi-Paccani S, Baudner B, Pallaoro M, Swennen E, Petracca R, Brettoni C, Liberatori S, Norais N, Monaci E, Bubeck Wardenburg J, Schneewind O, O'Hagan DT, Valiante NM, Bensi G, Bertholet S, De Gregorio E, Rappuoli R, Grandi G (2015) Vaccine composition formulated with a novel TLR7-dependent adjuvant induces high and broad protection against *Staphylococcus aureus*. Proc Natl Acad Sci USA 112:3680–3685

Baudet M, Ortet P, Gaillard JC, Fernandez B, Guerin P, Enjalbal C, Subra G, de Groot A, Barakat M, Dedieu A, Armengaud J (2010) Proteomics-based refinement of *Deinococcus deserti* genome annotation reveals an unwonted use of non-canonical translation initiation codons. Mol Cell Proteomics MCP 9:415–426

Becher D, Hempel K, Sievers S, Zuhlke D, Pane-Farre J, Otto A, Fuchs S, Albrecht D, Bernhardt J, Engelmann S, Volker U, van Dijl JM, Hecker M (2009) A proteomic view of an important human pathogen—towards the quantification of the entire *Staphylococcus aureus* proteome. PLoS ONE 4:e8176

Becker S, Frankel MB, Schneewind O, Missiakas D (2014) Release of protein A from the cell wall of *Staphylococcus aureus*. Proc Natl Acad Sci USA 111:1574–1579

Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005) Non-classical protein secretion in bacteria. BMC Microbiol 5:58

Biagini M, Garibaldi M, Aprea S, Pezzicoli A, Doro F, Becherelli M, Taddei AR, Tani C, Tavarini S, Mora M, Teti G, D'Oro U, Nuti S, Soriani M, Margarit I, Rappuoli R, Grandi G, Norais N (2015) The human pathogen Streptococcus pyogenes releases lipoproteins as lipoprotein-rich membrane vesicles. Mol Cell Proteomics MCP

Briese T, Hakenbeck R (1985) Interaction of the pneumococcal amidase with lipoteichoic acid and choline. Eur J Biochem/FEBS 146:417–427

Brown EL, Dumitrescu O, Thomas D, Badiou C, Koers EM, Choudhury P, Vazquez V, Etienne J, Lina G, Vandenesch F, Bowden MG (2009) The Panton-Valentine leukocidin vaccine protects mice against lung and skin infections caused by *Staphylococcus aureus* USA300. Clin Microbiol Infect 15:156–164

Brown L, Kessler A, Cabezas-Sanchez P, Luque-Garcia JL, Casadevall A (2014) Extracellular vesicles produced by the Gram-positive bacterium *Bacillus subtilis* are disrupted by the lipopeptide surfactin. Mol Microbiol 93:183–198

Bubeck Wardenburg J, Schneewind O (2008) Vaccine protection against *Staphylococcus aureus* pneumonia. J Exp Med 205:287–294

Bukowski M, Wladyka B, Dubin G (2010) Exfoliative toxins of *Staphylococcus aureus*. Toxins (Basel) 2:1148–1165

Bumann D (2010) Pathogen proteomes during infection: a basis for infection research and novel control strategies. J Proteomics 73:2267–2276

Burts ML, Williams WA, DeBord K, Missiakas DM (2005) EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections. Proc Natl Acad Sci USA 102:1169–1174

Cafardi V, Biagini M, Martinelli M, Leuzzi R, Rubino JT, Cantini F, Norais N, Scarselli M, Serruto D, Unnikrishnan M (2013) Identification of a novel zinc metalloprotease through a global analysis of Clostridium difficile extracellular proteins. PLoS ONE 8:e81306

Chagnot C, Zorgani MA, Astruc T, Desvaux M (2013) Proteinaceous determinants of surface colonization in bacteria: bacterial adhesion and biofilm formation from a protein secretion perspective. Front Microbiol 4:303

Clauser KR, Baker P, Burlingame AL (1999) Role of accurate mass measurement (±10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal Chem 71:2871–2882

Cockayne A, Hill PJ, Powell NB, Bishop K, Sims C, Williams P (1998) Molecular cloning of a 32-kilodalton lipoprotein component of a novel iron-regulated *Staphylococcus epidermidis* ABC transporter. Infect Immun 66:3767–3774

Couto N, Martins J, Lourenco AM, Pomba C, Varela Coelho A (2016) Identification of vaccine candidate antigens of *Staphylococcus pseudintermedius* by whole proteome characterization and serological proteomic analyses. J Proteomics 133:113–124

Cullen PA, Haake DA, Bulach DM, Zuerner RL, Adler B (2003) LipL21 is a novel surface-exposed lipoprotein of pathogenic Leptospira species. Infect Immun 71:2414–2421

Dalla Vecchia E, Shao PP, Suvorova E, Chiappe D, Hamelin R, Bernier-Latmani R (2014) Characterization of the surfaceome of the metal-reducing bacterium Desulfotomaculum reducens. Front Microbiol 5:432

Depke M, Michalik S, Rabe A, Surmann K, Brinkmann L, Jehmlich N, Bernhardt J, Hecker M, Wollscheid B, Sun Z, Moritz RL, Volker U, Schmidt F (2015) A peptide resource for the analysis of *Staphylococcus aureus* in host-pathogen interaction studies. *Proteomics*

Desvaux M, Hebraud M, Talon R, Henderson IR (2009) Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends Microbiol 17:139–145

Diep BA, Phung Q, Date S, Arnott D, Bakalarski C, Xu M, Nakamura G, Swem DL, Alexander MK, Le HN, Mai TT, Tan MW, Brown EJ, Nishiyama M (2014) Identifying potential therapeutic targets of methicillin-resistant *Staphylococcus aureus* through in vivo proteomic analysis. J Infect Dis 209:1533–1541

Dischinger J, Basi Chipalu S, Bierbaum G (2014) Lantibiotics: promising candidates for future applications in health care. Int J Med Microbiol IJMM 304:51–62

Distler U, Kuharev J, Navarro P, Levin Y, Schild H, Tenzer S (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. Nat Meth 11:167–170

Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V (2010) Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol 11:R107

Doro F, Liberatori S, Rodriguez-Ortega MJ, Rinaudo CD, Rosini R, Mora M, Scarselli M, Altindis E, D'Aurizio R, Stella M, Margarit I, Maione D, Telford JL, Norais N, Grandi G (2009) Surfome analysis as a fast track to vaccine discovery: identification of a novel protective antigen for group B Streptococcus hypervirulent strain COH1. Mol Cell Proteomics MCP 8:1728–1737

Dreisbach A, Hempel K, Buist G, Hecker M, Becher D, van Dijl JM (2010) Profiling the surfacome of *Staphylococcus aureus*. Proteomics 10:3082–3096

Elhaik Goldman S, Dotan S, Talias A, Lilo A, Azriel S, Malka I, Portnoi M, Ohayon A, Kafka D, Ellis R, Elkabets M, Porgador A, Levin D, Azhari R, Swiatlo E, Ling E, Feldman G, Tal M, Dagan R, Mizrachi Nebenzahl Y (2016) *Streptococcus pneumoniae* fructose-1,6-bisphosphate aldolase, a protein vaccine candidate, elicits Th1/Th2/Th17-type cytokine responses in mice. Int J Mol Med 37:1127–1138

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5:976–989

Feng Y, Pan X, Sun W, Wang C, Zhang H, Li X, Ma Y, Shao Z, Ge J, Zheng F, Gao GF, Tang J (2009) *Streptococcus suis* enolase functions as a protective antigen displayed on the bacterial cell surface. J Infect Dis 200:1583–1592

Fischer A, Yang SJ, Bayer AS, Vaezzadeh AR, Herzig S, Stenz L, Girard M, Sakoulas G, Scherl A, Yeaman MR, Proctor RA, Schrenzel J, Francois P (2011) Daptomycin resistance mechanisms in clinically derived *Staphylococcus aureus* strains assessed by a combined transcriptomics and proteomics approach. J Antimicrob Chemother 66:1696–1711

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM et al (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512

Foster TJ, Geoghegan JA, Ganesh VK, Hook M (2014) Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. Nat Rev Microbiol 12:49–62

Foulston L, Elsholz AK, DeFrancesco AS, Losick R (2014) The extracellular matrix of *Staphylococcus aureus* biofilms comprises cytoplasmic proteins that associate with the cell surface in response to decreasing pH. *mBio* 5:e01667-01614

Fuchs S, Zuhlke D, Pane-Farre J, Kusch H, Wolf C, Reiss S, Binh le TN, Albrecht D, Riedel K, Hecker M, Engelmann S (2013) Aureolib—a proteome signature library: towards an understanding of *Staphylococcus aureus* pathophysiology. *PloS one* 8:e70669

Gabrielsen C, Brede DA, Nes IF, Diep DB (2014) Circular bacteriocins: biosynthesis and mode of action. Appl Environ Microbiol 80:6854–6862

Garg A, Gupta D (2008) VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. BMC Bioinf 9:62

Giuliani MM, Adu-Bobie J, Comanducci M, Arico B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B, Cartocci E, Ciucchi L, Di Marcello F, Ferlicca F, Galli B, Luzzi E, Masignani V, Serruto D, Veggi D, Contorni M, Morandi M, Bartalesi A, Cinotti V, Mannucci D, Titta F, Ovidi E, Welsch JA, Granoff D, Rappuoli R, Pizza M (2006) A universal vaccine for serogroup B meningococcus. Proc Natl Acad Sci USA 103:10834–10839

Gorg A, Postel W, Domscheit A, Gunther S (1988a) Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (Hordeum vulgare): method, reproducibility and genetic aspects. Electrophoresis 9:681–692

Gorg A, Postel W, Gunther S (1988b) The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis 9:531–546

Gurung M, Moon DC, Choi CW, Lee JH, Bae YC, Kim J, Lee YC, Seol SY, Cho DT, Kim SI, Lee JC (2011) *Staphylococcus aureus* produces membrane-derived vesicles that induce host cell death. PLoS ONE 6:e27958

Haas B, Grenier D (2015) Isolation, characterization and biological properties of membrane vesicles produced by the swine pathogen *Streptococcus suis*. PLoS ONE 10:e0130528

Hawkins J, Kodali S, Matsuka YV, McNeil LK, Mininni T, Scully IL, Vernachio JH, Severina E, Girgenti D, Jansen KU, Anderson AS, Donald RG (2012) A recombinant clumping factor A-containing vaccine induces functional antibodies to *Staphylococcus aureus* that are not observed after natural exposure. Clin Vaccine Immunol CVI 19:1641–1650

Hempel K, Pane-Farre J, Otto A, Sievers S, Hecker M, Becher D (2010) Quantitative cell surface proteome profiling for SigB-dependent protein expression in the human pathogen *Staphylococcus aureus* via biotinylation approach. J Proteome Res 9:1579–1590

Hempel K, Herbst FA, Moche M, Hecker M, Becher D (2011) Quantitative proteomic view on secreted, cell surface-associated, and cytoplasmic proteins of the methicillin-resistant human pathogen *Staphylococcus aureus* under iron-limited conditions. J Proteome Res 10:1657–1666

Henderson B, Martin A (2011) Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease. Infect Immun 79: 3476–3491

Henderson B, Martin A (2013) Bacterial moonlighting proteins and bacterial virulence. Curr Top Microbiol Immunol 358:155–213

Hendrickson RC, Douglass JF, Reynolds LD, McNeill PD, Carter D, Reed SG, Houghton RL (2000) Mass spectrometric identification of mtb81, a novel serological marker for tuberculosis. J Clin Microbiol 38:2354–2361

Hensbergen PJ, Klychnikov OI, Bakker D, van Winden VJ, Ras N, Kemp AC, Cordfunke RA, Dragan I, Deelder AM, Kuijper EJ, Corver J, Drijfhout JW, van Leeuwen HC (2014) A novel secreted metalloprotease (CD2830) from Clostridium difficile cleaves specific proline sequences in LPXTG cell surface proteins. Mol Cell Proteomics MCP 13:1231–1244

Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. Proc Natl Acad Sci USA 90:5011–5015

Hessling B, Bonn F, Otto A, Herbst FA, Rappen GM, Bernhardt J, Hecker M, Becher D (2013) Global proteome analysis of vancomycin stress in *Staphylococcus aureus*. Int J Med Microbiol IJMM 303:624–634

Hirose I, Sano K, Shioda I, Kumano M, Nakamura K, Yamane K (2000) Proteome analysis of *Bacillus subtilis* extracellular proteins: a two-dimensional protein electrophoretic study. Microbiology 146(Pt 1):65–75

Hong SW, Kim MR, Lee EY, Kim JH, Kim YS, Jeon SG, Yang JM, Lee BJ, Pyun BY, Gho YS, Kim YK (2011) Extracellular vesicles derived from *Staphylococcus aureus* induce atopic dermatitis-like skin inflammation. Allergy 66:351–359

Hong SW, Choi EB, Min TK, Kim JH, Kim MH, Jeon SG, Lee BJ, Gho YS, Jee YK, Pyun BY, Kim YK (2014) An important role of alpha-hemolysin in extracellular vesicles on the development of atopic dermatitis induced by *Staphylococcus aureus*. PLoS ONE 9:e100499

Hussain M, Ozawa Y, Ichihara S, Mizushima S (1982) Signal peptide digestion in *Escherichia coli.* Effect of protease inhibitors on hydrolysis of the cleaved signal peptide of the major outer-membrane lipoprotein. Eur J biochem/FEBS 129:233–239

James P, Quadroni M, Carafoli E, Gonnet G (1993) Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun 195:58–64

Jiang Y, Kong Q, Roland KL, Curtiss R 3rd (2014) Membrane vesicles of Clostridium perfringens type A strains induce innate and adaptive immunity. Int J Med Microbiol IJMM 304:431–443

Johansson BP, Levander F, von Pawel-Rammingen U, Berggard T, Bjorck L, James P (2005) The protein expression of Streptococcus pyogenes is significantly influenced by human plasma. J Proteome Res 4:2302–2311

Kainulainen V, Korhonen TK (2014) Dancing to another tune-adhesive moonlighting proteins in bacteria. Biology 3:178–204

Kang HJ, Coulibaly F, Clow F, Proft T, Baker EN (2007) Stabilizing isopeptide bonds revealed in Gram-positive bacterial pilus structure. Science 318:1625–1628

Karauzum H, Adhikari RP, Sarwar J, Devi VS, Abaandou L, Haudenschild C, Mahmoudieh M, Boroun AR, Vu H, Nguyen T, Warfield KL, Shulenin S, Aman MJ (2013) Structurally designed attenuated subunit vaccines for *S. aureus* LukS-PV and LukF-PV confer protection in a mouse bacteremia model. PLoS ONE 8:e65384

Karlsson C, Malmstrom L, Aebersold R, Malmstrom J (2012) Proteome-wide selected reaction monitoring assays for the human pathogen Streptococcus pyogenes. Nat Commun 3:1301

Kennedy AD, Bubeck Wardenburg J, Gardner DJ, Long D, Whitney AR, Braughton KR, Schneewind O, DeLeo FR (2010) Targeting of alpha-hemolysin by active or passive immunization decreases severity of USA300 skin infection in a mouse model. J Infect Dis 202:1050–1058

Kim JH, Lee J, Park J, Gho YS (2015) Gram-negative and Gram-positive bacterial extracellular vesicles. Semin Cell Dev Biol 40:97–104

Klade CS (2002) Proteomics approaches towards antigen discovery and vaccine development. Curr Opin Mol Ther 4:216–223

Klade CS, Voss T, Krystek E, Ahorn H, Zatloukal K, Pummer K, Adolf GR (2001) Identification of tumor antigens in renal cell carcinoma by serological proteome analysis. Proteomics 1: 890–898

Knox KW, Vesk M, Work E (1966) Relation between excreted lipopolysaccharide complexes and surface structures of a lysine-limited culture of *Escherichia coli*. J Bacteriol 92:1206–1217

Korea CG, Balsamo G, Pezzicoli A, Merakou C, Tavarini S, Bagnoli F, Serruto D, Unnikrishnan M (2014) Staphylococcal Esx proteins modulate apoptosis and release of intracellular *Staphylococcus aureus* during infection in epithelial cells. Infect Immun 82: 4144–4153

Kovacs-Simon A, Titball RW, Michell SL (2011) Lipoproteins of bacterial pathogens. Infect Immun 79:548–561

Kurokawa K, Lee H, Roh KB, Asanuma M, Kim YS, Nakayama H, Shiratsuchi A, Choi Y, Takeuchi O, Kang HJ, Dohmae N, Nakanishi Y, Akira S, Sekimizu K, Lee BL (2009) The triacylated ATP binding cluster transporter substrate-binding lipoprotein of *Staphylococcus aureus* functions as a native ligand for Toll-like receptor 2. J Biol Chem 284:8406–8411

Kusch H, Engelmann S (2014) Secrets of the secretome in *Staphylococcus aureus*. Int J Med Microbiol IJMM 304:133–141

Kuster B, Schirle M, Mallick P, Aebersold R (2005) Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol 6:577–583

Lange V, Malmstrom JA, Didion J, King NL, Johansson BP, Schafer J, Rameseder J, Wong CH, Deutsch EW, Brusniak MY, Buhlmann P, Bjorck L, Domon B, Aebersold R (2008) Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. Mol Cell Proteomics MCP 7:1489–1500

Lawley TD, Croucher NJ, Yu L, Clare S, Sebaihia M, Goulding D, Pickard DJ, Parkhill J, Choudhary J, Dougan G (2009) Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores. J Bacteriol 191:5377–5386

Lee EY, Choi DY, Kim DK, Kim JW, Park JO, Kim S, Kim SH, Desiderio DM, Kim YK, Kim KP, Gho YS (2009) Gram-positive bacteria produce membrane vesicles: proteomics-based characterization of *Staphylococcus aureus*-derived membrane vesicles. Proteomics 9:5425–5436

Lee JH, Choi CW, Lee T, Kim SI, Lee JC, Shin JH (2013) Transcription factor sigmaB plays an important role in the production of extracellular membrane-derived vesicles in Listeria monocytogenes. PLoS ONE 8:e73196

Lei B, Mackie S, Lukomski S, Musser JM (2000) Identification and immunogenicity of group A Streptococcus culture supernatant proteins. Infect Immun 68:6807–6818

Li S, Huang H, Rao X, Chen W, Wang Z, Hu X (2014) Phenol-soluble modulins: novel virulence-associated peptides of staphylococci. Future Microbiol 9:203–216

Liao S, Klein MI, Heim KP, Fan Y, Bitoun JP, Ahn SJ, Burne RA, Koo H, Brady LJ, Wen ZT (2014) Streptococcus mutans extracellular DNA is upregulated during growth in biofilms, actively released via membrane vesicles, and influenced by components of the protein secretion machinery. J Bacteriol 196:2355–2366

Liebeke M, Dorries K, Zuhlke D, Bernhardt J, Fuchs S, Pane-Farre J, Engelmann S, Volker U, Bode R, Dandekar T, Lindequist U, Hecker M, Lalk M (2011) A metabolomics and proteomics study of the adaptation of *Staphylococcus aureus* to glucose starvation. Mol BioSyst 7: 1241–1253

Liew YK, Awang Hamat R, van Belkum A, Chong PP, Neela V (2015) Comparative exoproteomics and host inflammatory response in *Staphylococcus aureus* skin and soft tissue infections, bacteremia, and subclinical colonization. Clin Vaccine Immunol CVI 22:593–603

Liu X, Hu Y, Pai PJ, Chen D, Lam H (2014) Label-free quantitative proteomics analysis of antibiotic response in *Staphylococcus aureus* to oxacillin. J Proteome Res 13:1223–1233

Loughman JA, Caparon M (2006) Regulation of SpeB in Streptococcus pyogenes by pH and NaCl: a model for in vivo gene expression. J Bacteriol 188:399–408

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26:966–968

Magnus M, Pawlowski M, Bujnicki JM (2012) MetaLocGramN: a meta-predictor of protein subcellular localization for Gram-negative bacteria. Biochim Biophys Acta 1824:1425–1433

Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi-Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G (2005) Identification of a universal group B Streptococcus vaccine by multiple genome screen. Science 309:148–150

Malmstrom J, Beck M, Schmidt A, Lange V, Deutsch EW, Aebersold R (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. Nature 460:762–765

Malmstrom J, Karlsson C, Nordenfelt P, Ossola R, Weisser H, Quandt A, Hansson K, Aebersold R, Malmstrom L, Bjorck L (2012) Streptococcus pyogenes in human plasma: adaptive mechanisms analyzed by mass spectrometry-based proteomics. J Biol Chem 287:1415–1425

Malmstrom L, Bakochi A, Svensson G, Kilsgard O, Lantz H, Petersson AC, Hauri S, Karlsson C, Malmstrom J (2015) Quantitative proteogenomics of human pathogens using DIA-MS. J Proteomics 129:98–107

Manning AJ, Kuehn MJ (2013) Functional advantages conferred by extracellular prokaryotic membrane vesicles. J Mol Microbiol Biotechnol 23:131–141

Mariotti P, Malito E, Biancucci M, Lo Surdo P, Mishra RP, Nardi-Dei V, Savino S, Nissum M, Spraggon G, Grandi G, Bagnoli F, Bottomley MJ (2013) Structural and functional characterization of the *Staphylococcus aureus* virulence factor and vaccine candidate FhuD2. Biochem J 449:683–693

Michalik S, Liebeke M, Zuhlke D, Lalk M, Bernhardt J, Gerth U, Hecker M (2009) Proteolysis during long-term glucose starvation in *Staphylococcus aureus* COL. Proteomics 9:4468–4477

Michalik S, Bernhardt J, Otto A, Moche M, Becher D, Meyer H, Lalk M, Schurmann C, Schluter R, Kock H, Gerth U, Hecker M (2012) Life and death of proteins: a case study of glucose-starved *Staphylococcus aureus*. Mol Cell Proteomics 11:558–570

Miller M, Dreisbach A, Otto A, Becher D, Bernhardt J, Hecker M, Peppelenbosch MP, van Dijl JM (2011) Mapping of interactions between human macrophages and *Staphylococcus aureus* reveals an involvement of MAP kinase signaling in the host defense. J Proteome Res 10:4018–4032

Mishra RP, Mariotti P, Fiaschi L, Nosari S, Maccari S, Liberatori S, Fontana MR, Pezzicoli A, De Falco MG, Falugi F, Altindis E, Serruto D, Grandi G, Bagnoli F (2012) *Staphylococcus aureus* FhuD2 is involved in the early phase of staphylococcal dissemination and generates protective immunity in mice. J Infect Dis 206:1041–1049

Moche M, Schluter R, Bernhardt J, Plate K, Riedel K, Hecker M, Becher D (2015) Time-resolved analysis of cytosolic and surface-associated proteins of *Staphylococcus aureus* HG001 under planktonic and biofilm conditions. J Proteome Res 14:3804–3822

Mora M, Veggi D, Santini L, Pizza M, Rappuoli R (2003) Reverse vaccinology. Drug Discov Today 8:459–464

Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AG, Maggi T, Taddei AR, Grandi G, Telford JL (2005) Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci USA 102:15641–15646

Muntel J, Hecker M, Becher D (2012) An exclusion list based label-free proteome quantification approach using an LTQ Orbitrap. Rapid Commun Mass Spectrom 26:701–709

Nandakumar R, Nandakumar MP, Marten MR, Ross JM (2005) Proteome analysis of membrane and cell wall associated proteins from *Staphylococcus aureus*. J Proteome Res 4:250–257

Novartis receives positive CHMP opinion for MenB vaccine Bexsero (2013) Human Vaccines Immunother 9:8–9

O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. J Biol Chem 250:4007–4021

O'Halloran DP, Wynne K, Geoghegan JA (2015) Protein A is released into the *Staphylococcus aureus* culture supernatant with an unprocessed sorting signal. Infect Immun 83:1598–1609

Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, Obando I, Rodriguez-Ortega MJ (2013) Identification of potential new protein vaccine candidates through pan-surfomic analysis of pneumococcal clinical isolates from adults. PLoS ONE 8:e70365

Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, Rodriguez-Ortega MJ (2014a) Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. J Proteomics 97:164–176

Olaya-Abril A, Prados-Rosales R, McConnell MJ, Martin-Pena R, Gonzalez-Reyes JA, Jimenez-Munguia I, Gomez-Gascon L, Fernandez J, Luque-Garcia JL, Garcia-Lidon C, Estevez H, Pachon J, Obando I, Casadevall A, Pirofski LA, Rodriguez-Ortega MJ (2014b) Characterization of protective extracellular membrane-derived vesicles produced by Streptococcus pneumoniae. J Proteomics 106:46–60

Omer H, Alpha-Bazin B, Brunet JL, Armengaud J, Duport C (2015) Proteomics identifies *Bacillus cereus* EntD as a pivotal protein for the production of numerous virulence factors. Front Microbiol 6:1004

Otto M (2014) *Staphylococcus epidermidis* pathogenesis. Meth Mol Biol 1106:17–31

Otto A, van Dijl JM, Hecker M, Becher D (2014) The *Staphylococcus aureus* proteome. Int J Med Microbiol IJMM 304:110–120

Pappin DJ, Hojrup P, Bleasby AJ (1993) Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3:327–332

Perez-Casal J, Potter AA (2016) Glyceradehyde-3-phosphate dehydrogenase as a suitable vaccine candidate for protection against bacterial and parasitic diseases. Vaccine 34:1012–1017

Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567

Peschel A, Otto M (2013) Phenol-soluble modulins and staphylococcal infection. Nat Rev Microbiol 11:667–673

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8:785–786

Pfortner H, Burian MS, Michalik S, Depke M, Hildebrandt P, Dhople VM, Pane-Farre J, Hecker M, Schmidt F, Volker U (2014) Activation of the alternative sigma factor SigB of *Staphylococcus aureus* following internalization by epithelial cells—an in vivo proteomics perspective. Int J Med Microbiol IJMM 304:177–187

Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Meth 9:555–566

Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Galeotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R (2000) Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287:1816–1820

Pribyl T, Moche M, Dreisbach A, Bijlsma JJ, Saleh M, Abdullah MR, Hecker M, van Dijl JM, Becher D, Hammerschmidt S (2014) Influence of impaired lipoprotein biogenesis on surface and exoproteome of *Streptococcus pneumoniae*. J Proteome Res 13:650–667

Rabilloud T, Chevallet M, Luche S, Lelong C (2010) Two-dimensional gel electrophoresis in proteomics: past, present and future. J Proteomics 73:2064–2077

Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 19:2688–2691

Rasmussen M, Bjorck L (2002) Proteolysis and its regulation at the surface of *Streptococcus pyogenes*. Mol Microbiol 43:537–544

Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, Hengartner MO, Aebersold R (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nat Meth 8:430–435

Rice KC, Bayles KW (2008) Molecular control of bacterial death and lysis. Microbiol Mol Biol Rev MMBR 72:85–109, table of contents

Rivera J, Cordero RJ, Nakouzi AS, Frases S, Nicola A, Casadevall A (2010) *Bacillus anthracis* produces membrane-derived vesicles containing biologically active toxins. Proc Natl Acad Sci USA 107:19002–19007

Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, Scarselli M, Doro F, Ferrari G, Garaguso I, Maggi T, Neumann A, Covre A, Telford JL, Grandi G (2006) Characterization and identification of vaccine candidate proteins through analysis of the group A Streptococcus surface proteome. Nat Biotechnol 24:191–197

Romine MF (2011) Genome-wide protein localization prediction strategies for gram negative bacteria. BMC Genom 12(Suppl 1):S1

Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, Wolski W, Collins BC, Malmstrom J, Malmstrom L, Aebersold R (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol 32:219–223

Sanchez B, Cabo ML, Margolles A, Herrera JJ (2010) A proteomic approach to cold acclimation of *Staphylococcus aureus* CECT 976 grown at room and human body temperatures. Int J Food Microbiol 144:160–168

Sato Y, Takaya A, Yamamoto T (2011) Meta-analytic approach to the accurate prediction of secreted virulence effectors in gram-negative bacteria. BMC Bioinf 12:442

Schluepen C, Malito E, Marongiu A, Schirle M, McWhinnie E, Lo Surdo P, Biancucci M, Falugi F, Nardi-Dei V, Marchi S, Fontana MR, Lombardi B, De Falco MG, Rinaudo CD, Spraggon G, Nissum M, Bagnoli F, Grandi G, Bottomley MJ, Liberatori S (2013) Mining the bacterial unknown proteome: identification and characterization of a novel family of highly conserved protective antigens in *Staphylococcus aureus*. Biochem J 455:273–284

Schmidt F, Volker U (2011) Proteome analysis of host-pathogen interactions: investigation of pathogen responses to the host cell environment. Proteomics 11:3203–3211

Schmidt F, Scharf SS, Hildebrandt P, Burian M, Bernhardt J, Dhople V, Kalinka J, Gutjahr M, Hammer E, Volker U (2010) Time-resolved quantitative proteome profiling of host-pathogen interactions: the response of Staphylococcus aureus RN1HG to internalisation by human airway epithelial cells. Proteomics 10:2801–2811

Schrempf H, Merling P (2015) Extracellular Streptomyces lividans vesicles: composition, biogenesis and antimicrobial activity. Microb Biotechnol 8:644–658

Schrempf H, Koebsch I, Walter S, Engelhardt H, Meschke H (2011) Extracellular Streptomyces vesicles: amphorae for survival and defence. Microb Biotechnol 4:286–299

Sebulsky MT, Heinrichs DE (2001) Identification and characterization of fhuD1 and fhuD2, two genes involved in iron-hydroxamate uptake in *Staphylococcus aureus*. J Bacteriol 183:4994–5000

Severin A, Nickbarg E, Wooters J, Quazi SA, Matsuka YV, Murphy E, Moutsatsos IK, Zagursky RJ, Olmsted SB (2007) Proteomic analysis and identification of Streptococcus pyogenes surface-associated proteins. J Bacteriol 189:1514–1522

Shallcross LJ, Fragaszy E, Johnson AM, Hayward AC (2013) The role of the Panton-Valentine leucocidin toxin in staphylococcal disease: a systematic review and meta-analysis. Lancet Infect Dis 13:43–54

Sjoholm K, Karlsson C, Linder A, Malmstrom J (2014) A comprehensive analysis of the Streptococcus pyogenes and human plasma protein interaction network. Mol BioSyst 10:1698–1708

Stabler RA, Gerding DN, Songer JG, Drudy D, Brazier JS, Trinh HT, Witney AA, Hinds J, Wren BW (2006) Comparative phylogenomics of Clostridium difficile reveals clade specificity and microevolution of hypervirulent strains. J Bacteriol 188:7297–7305

Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW (2009) Comparative genome and phenotypic analysis of Clostridium difficile 027 strains provides insight into the evolution of a hypervirulent bacterium. Genome Biol 10:R102

Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, Aebersold R, Domon B (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. Mol Cell Proteomics 6:1809–1817

Stentzel S, Vu HC, Weyrich AM, Jehmlich N, Schmidt F, Salazar MG, Steil L, Volker U, Broker BM (2014) Altered immune proteome of *Staphylococcus aureus* under iron-restricted growth conditions. Proteomics 14:1857–1867

Surmann K, Michalik S, Hildebrandt P, Gierok P, Depke M, Brinkmann L, Bernhardt J, Salazar MG, Sun Z, Shteynberg D, Kusebauch U, Moritz RL, Wollscheid B, Lalk M, Volker U, Schmidt F (2014) Comparative proteome analysis reveals conserved and specific adaptation patterns of *Staphylococcus aureus* after internalization by different types of human non-professional phagocytic host cells. Front Microbiol 5:392

Surmann K, Simon M, Hildebrandt P, Pfortner H, Michalik S, Stentzel S, Steil L, Dhople VM, Bernhardt J, Schluter R, Depke M, Gierok P, Lalk M, Broker BM, Schmidt F, Volker U (2015) A proteomic perspective of the interplay of *Staphylococcus aureus* and human alveolar epithelial cells during infection. J Proteomics 128:203–217

Tawaratsumida K, Furuyashiki M, Katsumoto M, Fujimoto Y, Fukase K, Suda Y, Hashimoto M (2009) Characterization of N-terminal structure of TLR2-activating lipoprotein in *Staphylococcus aureus*. J Biol Chem 284:9147–9152

Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK,

Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC (2000) Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 287:1809–1815

Thay B, Wai SN, Oscarsson J (2013) *Staphylococcus aureus* alpha-toxin-dependent induction of host cell death by membrane-derived vesicles. PLoS ONE 8:e54661

Thompson BJ, Widdick DA, Hicks MG, Chandra G, Sutcliffe IC, Palmer T, Hutchings MI (2010) Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium Streptomyces coelicolor. Mol Microbiol 77:943–957

Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev MMBR 64:515–547

van den Berg S, Koedijk DG, Back JW, Neef J, Dreisbach A, van Dijl JM, Bakker-Woudenberg IA, Buist G (2015) Active immunization with an octa-valent *Staphylococcus aureus* antigen mixture in models of *S. aureus* bacteremia and skin infection in mice. PLoS ONE 10:e0116847

Vandenesch F, Lina G, Henry T (2012) *Staphylococcus aureus* hemolysins, bi-component leukocidins, and cytolytic peptides: a redundant arsenal of membrane-damaging virulence factors? Front Cell Infect Microbiol 2:12

Vesikari T, Esposito S, Prymula R, Ypma E, Kohl I, Toneatto D, Dull P, Kimura A, E.U.M.B.I.V. S. group (2013) Immunogenicity and safety of an investigational multicomponent, recombinant, meningococcal serogroup B vaccine (4CMenB) administered concomitantly with routine infant and child vaccinations: results of two randomised trials. Lancet 381:825–835

Vidal-Ingigliardi D, Lewenza S, Buddelmeijer N (2007) Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family. J Bacteriol 189:4456–4464

Vytvytska O, Nagy E, Bluggel M, Meyer HE, Kurzbauer R, Huber LA, Klade CS (2002) Identification of vaccine candidate antigens of *Staphylococcus aureus* by serological proteome analysis. Proteomics 2:580–590

Walden M, Crow A, Nelson MD, Banfield MJ (2014) Intramolecular isopeptide but not internal thioester bonds confer proteolytic and significant thermal stability to the *S. pyogenes* pilus adhesin Spy0125. Proteins 82:517–527

Wizemann TM, Heinrichs JH, Adamou JE, Erwin AL, Kunsch C, Choi GH, Barash SC, Rosen CA, Masure HR, Tuomanen E, Gayle A, Brewah YA, Walsh W, Barren P, Lathigra R, Hanson M, Langermann S, Johnson S, Koenig S (2001) Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. Infect Immun 69:1593–1598

Wolf C, Kusch H, Monecke S, Albrecht D, Holtfreter S, von Eiff C, Petzl W, Rainard P, Broker BM, Engelmann S (2011) Genomic and proteomic characterization of *Staphylococcus aureus* mastitis isolates of bovine origin. Proteomics 11:2491–2502

Xu SX, McCormick JK (2012) Staphylococcal superantigens in colonization and disease. Front Cell Infect Microbiol 2:52

Yates JR 3rd, Speicher S, Griffin PR, Hunkapiller T (1993) Peptide mass maps: a highly informative approach to protein identification. Anal Biochem 214:397–408

Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615

Zhou M, Boekhorst J, Francke C, Siezen RJ (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. BMC Bioinf 9:173

Ziebandt AK, Weber H, Rudolph J, Schmid R, Hoper D, Engelmann S, Hecker M (2001) Extracellular proteins of *Staphylococcus aureus* and the role of SarA and sigma B. Proteomics 1:480–493

Ziebandt AK, Kusch H, Degner M, Jaglitz S, Sibbald MJ, Arends JP, Chlebowicz MA, Albrecht D, Pantucek R, Doskar J, Ziebuhr W, Broker BM, Hecker M, van Dijl JM, Engelmann S (2010) Proteomics uncovers extreme heterogeneity in the *Staphylococcus aureus* exoproteome due to genomic plasticity and variant gene regulation. Proteomics 10:1634–1644

Zysk G, Bongaerts RJ, ten Thoren E, Bethe G, Hakenbeck R, Heinz HP (2000) Detection of 23 immunogenic pneumococcal proteins using convalescent-phase serum. Infect Immun 68:3740–3743