university of groningen

University Medical Center Groningen

University of Groningen

The Staphylococcus aureus secretome

Sibbald, M.J.J.B.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Sibbald, M. J. J. B. (2010). *The Staphylococcus aureus secretome*. s.n.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-12-2022

Chapter 2

Mapping the pathways to staphylococcal pathogenesis by comparative secretomics

M.J.J.B. Sibbald, *A.K. Ziebandt, S. Engelmann, M. Hecker, A. de Jong, H.J.M. Harmsen, G.C. Raangs, I. Stokroos, J.P. Arends, J.Y.F. Dubois, and J.M. van Dijl*

Published in *Microbiology and Molecular Biology Reviews* (2006) **70**, 755-788

Summary

The Gram-positive bacterium *Staphylococcus aureus* is a frequent component of the human microbial flora that can turn into a dangerous pathogen. As such, this organism is capable of infecting almost every tissue and organ system in the human body. It does so by actively exporting a variety of virulence factors to the cell surface and extracellular milieu. Upon reaching their respective destinations, these virulence factors have pivotal roles in the colonization and subversion of the human host. It is therefore of major importance to obtain a clear understanding of the protein transport pathways that are active in *S. aureus*. The present review aims to provide a state-of-the-art roadmap of staphylococcal secretomes, which include both protein transport pathways and the extracytoplasmic proteins of these organisms. Specifically, an overview is presented of the exported virulence factors, pathways for protein transport, signals for cellular protein retention or secretion, and the exoproteomes of different *S. aureus* isolates. The focus is on *S. aureus*, but comparisons with *Staphylococcus epidermidis* and other Gram-positive bacteria like *Bacillus subtilis* are included where appropriate. Importantly, the results of genomic and proteomic studies on *S. aureus* secretomes are integrated through a comparative "secretomics" approach, resulting in a first definition of the core and variant secretomes of this bacterium. While the core secretome seems to be largely employed for general house-keeping functions, necessary to thrive in particular niches provided by the human host, the variant secretome seems to contain the “gadgets” that *S. aureus* needs to conquer these well-protected niches.

General introduction and scope of this review

The Gram-positive bacterium *Staphylococcus aureus* is a frequent component of the human microbial flora that can turn into a dangerous pathogen. As such, this organism is capable of infecting almost every tissue and organ system in the human body. It does so by exporting a variety of virulence factors to the cell surface and extracellular milieu of the human host. As in all living organisms (Wickner and Schekman, 2005), *S. aureus* contains several protein transport pathways, of which the general secretory (Sec) pathway is the most well known and best described. Proteins that need to be transported to an extracytoplasmic location contain, in general, an N-terminal signal peptide that is needed to target the newly synthesized protein from the ribosome to the translocation machinery in the cytoplasmic membrane. Next, the protein is threaded through the Sec translocon in an unfolded state. During this translocation step, or shortly thereafter, the signal peptide is removed by a so-called signal peptidase. Upon complete membrane translocation, the protein has to fold into its correct conformation and will then be retained in an extracytoplasmic compartment of the cell, or secreted into the extracellular milieu. In the case of Gram-positive cocci, such as *S. aureus* (Figure 1), we distinguish three extracytoplasmic subcellular compartments: the membrane, the membrane-cell wall interface and the cell wall. Since surface-exposed and secreted proteins of *S. aureus* play pivotal roles in the colonization and subversion of the human host, it is of major importance to obtain a clear understanding of the protein transport pathways that are active in this organism (Lee and Schneewind, 2001). Knowledge about the protein sorting mechanism has become all the more relevant with the upcoming of staphylococcal resistance against last-defence antibiotics, such as vancomycin. The scope of this review is to provide a state-of-the-art roadmap of staphylococcal secretomes, which include both protein transport pathways and the extracytoplasmic proteins of these organisms. The focus is on *S. aureus*, but comparisons with *Staphylococcus epidermidis* and the best characterized Gram-positive bacterium *Bacillus subtilis* are included where appropriate. Importantly, the present review aims to integrate the results of genomic and proteomic studies on *S. aureus* secretomes, representing the first documented "comparative secretomics" study. Specifically, this review deals with known and predicted exported virulence factors, pathways for protein transport, signals for subcellular protein sorting or secretion, and the exoproteomes of different *S. aureus* isolates as defined by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) and mass spectrometry (Figures 2 and 3). The exoproteome is defined by all *S. aureus* proteins that can be identified in the extracellular milieu of this organism and thus includes proteins actively secreted by living cells and the remains of dead cells. For a clear appreciation of the present review, it is important to bear in mind that the proteins exported from the cytoplasm could be directly involved in staphylococcal virulence, whereas the respective protein export systems represent the "pathways to pathogenesis".

Figure 1. Imaging of* S. aureus *RN6390. *(**A**) For scanning electron microscopy a drop of washed culture of bacteria was fixated for 30 min with 2% glutaraldehyde in 0.1 M Cacodylate buffer, pH 7.38. Next, the fixated bacteria were placed on a piece (1 cm^2) of cleaved 0.1% Poly-L Lysine coated mica sheet and washed in 0.1 M Cacodylate buffer. This specimen was dehydrated in ethanol series consisting of 30%, 50%, 70%, 96% and anhydrous 100% solution (3X) 10 min each, then Critical point dried with CO_2 and sputter-coated with 2-3 nm Au/Pd (Balzers coater). The specimen was fixed on a SEM-stub-holder and observed in a JEOL FE-SEM 6301F. (**B**) Micrograph of a cluster of* S. aureus *cells grown in blood culture medium. The cells were fixed with ethanol and hybridized with the fluorescein labelled Peptide Nucleid Acid (PNA) probe PNA-Stau. The image was generated by merging an epi-fluorescence image with the negative of a phase-contrast image.*

***S. aureus* MRSA 693331** | ***S. aureus* 035699y/bm** | ***S. aureus* RN6390**

***S. aureus* 044057g/rmo** | ***S. aureus* CA-MRSA 021708m/rmo** | ***S. aureus* COL**

Figure 2. The extracellular proteomes of different* S. aureus *strains
Proteins of the growth medium fractions of different staphylococcal isolates, grown in TSB medium (37 °C) to an optical density at 540 nm (OD_{540}) of 10, were separated by 2D-PAGE using immobilized pH gradient (IPG) strips in the pH range of 3 to 10 (Amersham Pharmacia Biotech, Piscataway, N. J.). Each gel was loaded with 350 µg protein extracts and, after electrophoresis, stained with Colloidal Coomassie. Proteins were identified by MALDI-TOF mass spectrometry. The corresponding protein spots are labelled with protein names according to the S. aureus *N315 database or NCBI entries for proteins not present in N315. The* S. aureus *strains that were used in these experiments are RN6390 and COL, and four clinical isolates from the University Medical Center Groningen named MRSA693331, 035699y/bm, 0440579/rmo, CA-MRSA021708m/rmo.*

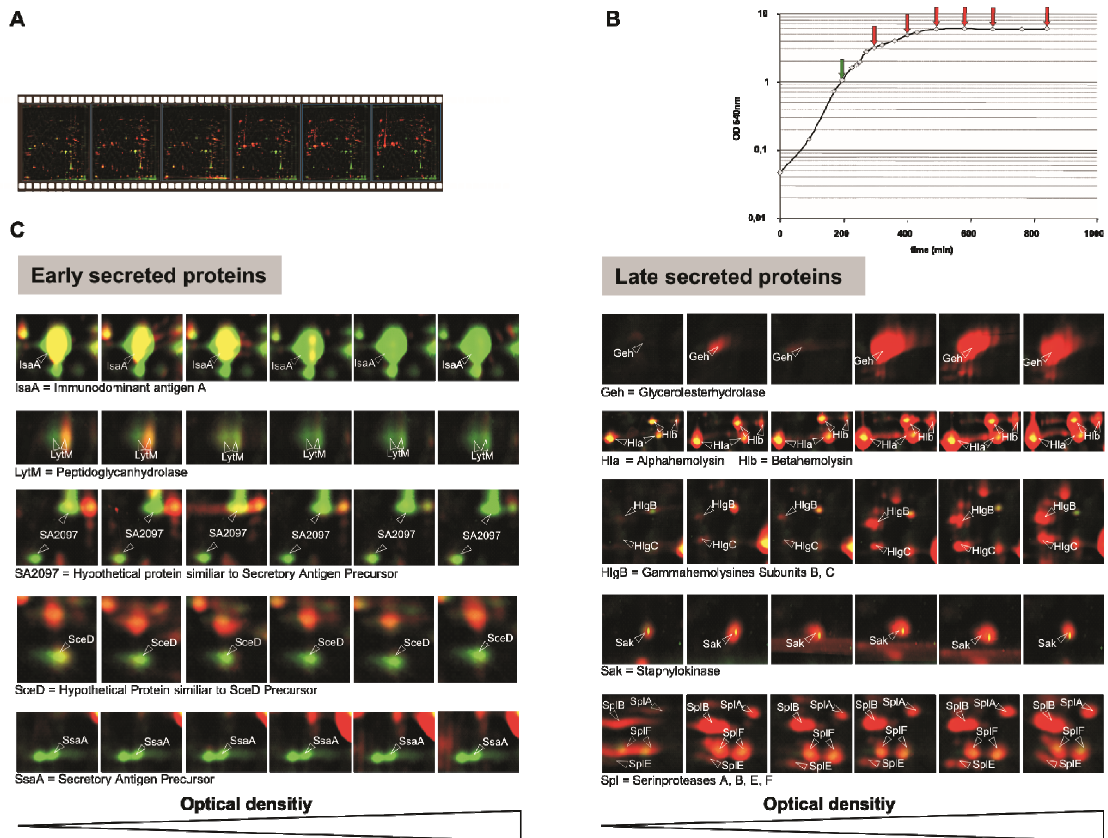

Figure 3. Dynamics of the amount of extracellular proteins during growth of* S. aureus *RN6390 in TSB medium

*(**A**) Individual dual channel 2D patterns of extracellular proteins during the different phases of the growth curve of cells grown in TSB medium were assembled into a movie. The protein pattern at an OD_{540} of 1 (labelled in green) was compared with the protein pattern at the respective higher optical densities (labelled in red). As a consequence of the dual channel labelling, spots of which the intensities do not differ in the compared gels will be yellow; spots of different intensities will be either green or red (Bernhardt et al., 1999). (**B**) Growth curve of* S. aureus *RN6390 grown in TSB medium as determined by OD_{540} readings. The sampling points for proteomics analyses are indicated by arrows. (**C**) Proteomic signatures of selected proteins representing different regulatory groups as revealed by dual channel imaging. The relative amounts of the respective proteins at an OD_{540} of 1 (spots labelled in green) of cells grown in TSB medium were compared with the relative amounts of these proteins at higher optical densities (spots labelled in red). Proteins were stained with Sypro Ruby®.*

Exported staphylococcal virulence factors

S. aureus and *S. epidermidis* are organisms that occur naturally in and on the human body. While *S. epidermidis* is mostly present on the human skin, *S. aureus* can be found on mucosal surfaces. *S. aureus* is carried by 30-40% of the population (Peacock *et al.*, 2001) and can readily be identified in the nose, but the organism can also be detected in other moist regions of the human body, such as axilla, perineum, vagina and rectum, thereby forming a major reservoir for infections. Although most staphylococcal infections are nosocomial (*i.e.* hospital-acquired), an increase in the number of cases of community-acquired antibiotic (methicillin) resistant infections is currently observed world-wide (Centers for Disease Control and Prevention, 2003; Grundmann *et al.*, 2002; Vandenesch *et al.*, 2003). The risk of intravascular and systemic infection by *S. aureus* rises when the epithelial barrier is disrupted by intravascular catheters, implants, mucosal damage or trauma. Interestingly, after infection,

cells of *S. aureus* can persist unnoticed in the human body for long periods of time (years) after which they can suddenly cause another infection. *S. aureus* is primarily an extracellular pathogen whose colonization and invasion of human tissues and organs can lead to severe cytotoxic effects. Nevertheless, *S. aureus* can also be internalized by various cells, including non-phagocytic cells, which seems to induce apoptosis (Hauck and Ohlsen, 2006; da Silva *et al.*, 2004; Mempel *et al.*, 2002). Although *S. aureus* has the potential to form biofilms (Götz, 2002), *S. epidermidis* infections are particularly notorious for the formation of thick multilayered biofilms on indwelling catheters and other implanted devices. The formation of such a biofilm takes place in several steps during which the bacteria first adhere rapidly to the surface of the polymer material that has been coated with a film of proteinaceous and non-proteinaceous organic host molecules (Escher and Characklis, 1990). Bacteria that adhere to this film produce extracellular polymeric substances, mostly polysaccharides and proteins, in turn resulting in a strong attachment to the polymer surface and other bacteria in the growing biofilm. Ultimately, the biofilm is composed of multiple layers of cells, cellular debris, polysaccharides and proteins. *S. epidermidis* proteins that are essential for biofilm formation are, for example, the polysaccharide intercellular adhesin (PIA) (Mack *et al.*, 1996), the accumulation associated protein (AAP) (Rohde *et al.*, 2005) and the biofilm-associated protein (Bap) (Tormo *et al.*, 2005). PIA is most likely identical to the polysaccharide adhesion (PS/A).

Virulence of S. aureus

The pathogenicity of *S. aureus* is caused by the expression of an arsenal of virulence factors (Table 1), which can lead to superficial skin lesions such as styes, furunculosis and paronychia, or to more serious infections such as pneumonia, mastitis, urinary tract infections, osteomyelitis, endocarditis and even sepsis. In very rare cases, *S. aureus* causes meningitis. The virulence factors that *S. aureus* employs to cause these diseases are displayed at the surface of the staphylococcal cell or secreted into the host milieu (Fedtke *et al.*, 2004). Specifically, these virulence factors include (a) surface proteins that promote adhesion to and colonization of host tissues; (b) invasins that are exported to an extracytoplasmic location and promote bacterial spread in tissues (leukocidin, kinases, hyaluronidase); (c) surface factors that inhibit phagocytic engulfment (capsule, Protein A); (d) biochemical properties that enhance staphylococcal survival in phagocytes (carotenoids, catalase production); (e) immunological disguises (Protein A, coagulase, clotting factor); (f) membrane-damaging toxins that disrupt eukaryotic cell membranes (hemolysins, leukotoxin); (g) superantigens that contribute to the symptoms of septic shock (SEA-G, TSST, ET); and (h) determinants for inherent and acquired resistance to antimicrobial agents. Most virulence factors are expressed in a coordinated fashion during the growth cycle of *S. aureus*. The best characterized regulators of virulence factors are the accessory gene regulator (*agr*) (Morfeldt *et al.*, 1988; Peng *et al.*, 1988; Recsei *et al.*, 1986) and the Staphylococcal accessory regulator (SarA) (Cheung *et al.*, 1992; Cheung and Projan, 1994). Ziebandt *et al.* (Ziebandt *et al.*, 2004) showed that extracellular proteins can be divided into two groups, based on the timing of their expression in cells grown in tryptic soy broth (TSB): proteins that are mainly expressed at low cell densities, or proteins exclusively expressed at high cell densities. *Agr* seems to be an important positive regulator of proteins that are expressed at higher optical densities (*e.g.* proteases, hemolysins and lipases) and a negative regulator for proteins that are expressed during the exponential growth phase (*e.g.* immunodominant antigen A, secretory antigen

precursor and several proteins with unknown functions). In addition, Gill *et al.* (Gill *et al.*, 2005) identified 15 other two-component regulatory systems in the genomes of *S. aureus* and *S. epidermidis* that are potentially involved in staphylococcal virulence. In this respect, it is interesting to note that the antibiotic cerulenin, which is known to inhibit protein secretion by *S. aureus* at sub-MIC levels, was recently reported to block transcriptional activation of at least two regulatory determinants, *agr* and *sae*. Thus, it seems that cerulenin inhibits the transcription of genes for secretory proteins rather than the secretion process of these proteins (Adhikari and Novick, 2005). In contrast, it was previously believed that cerulenin would interfere with membrane function through an inhibition of normal fatty acid synthesis.

Table 1. Virulence factors of *S. aureus*

Pathogenic action	Virulence factors	Protein or other compound	Functions
Colonization of host tissues	Surface proteins	ClfA, ClfB, FnbA, FnbB, IsdA SdrC, SdrD, SdrE,	Adhesins, fibronectin and fibrinogen-binding proteins
Lysis eukaryotic cell membranes and bacterial spread	Membrane-damaging toxins, invasins	Geh, Hla, Hld, HlgA-C, HysA, Lip, LukD, LukE, LukF, LukS, Nuc	Hemolysins, hyaluronidase, leukocidin, leukotoxin, lipases, nucleases
Inhibition phagocytic engulfment	Surface factors	CapA-P, Efb, Spa	Capsule, protein A
Survival in phagocytes	Biochemical compounds	KatA, Staphyloxanthin	Carotenoids, catalase production
Immunological disguise and modulation	Surface proteins	ClfA, ClfB, Coa, Spa	Clumping factor, coagulase, protein A
Contribution to symptoms of septic shock	Exotoxins	Eta, Etb, SEA-G, TSST-1	Enterotoxins SEA-G, exfoliative toxin, toxic shock syndrome toxin TSST
Acquired resistance to antimicrobial agents	Resistance proteins	BlaZ, MecA, VanA	MRSA, VRSA

Notably, to date relatively little information is available on the molecular nature of the stimuli that are perceived by the major regulators of the expression of virulence factors. Overall, it should be clear that strain-specific differences in gene regulation by *agr*, *sae* or other regulators may dramatically influence the repetoire of produced virulence factors, thereby having a profound impact on the disease-causing potential of different strains. Since the interplay of different regulators and cell-to-cell communication can impact differently on the expression of virulence factors, even the disease-causing potential of individual *S. aureus* cells within a genetically identical population may vary.

Resistance of S. aureus *to antibiotics*

Resistance of *S. aureus* to antibiotics has been observed very soon after the introduction of penicillin about sixty years ago. In the following years, the amazing ability of staphylococci to develop resistance to antibiotics has resulted in the emergence of methicillin-resistant *S. aureus* (MRSA) and *S. epidermidis* (MRSE) strains. In fact, methicillin resistance was observed already in 1961 in nosocomial isolates of *S. aureus*, one year after the introduction of methicillin (Jevons, 1961). The resistance towards methicillin is a result of the production of an altered penicillin binding protein, PBP2a (or PBP2'), which has less affinity to most β-lactam antibiotics. The PBP2a protein, which is located at the membrane-cell wall interface, is of major importance for cell wall biogenesis by mediating the cross linking of peptidoglycans.

PBP2a is encoded by the *mecA* gene, which is located on a mobile genetic element, also known as the staphylococcal cassette chromosome (SCC) *mec* (Chambers, 1997; Ito *et al.*, 2004). The SCC*mec* element is a basic mobile genetic element that serves as a vehicle for gene exchange among staphylococcal species (Dobrindt *et al.*, 2004). In addition to the *mecA* gene, SCCmec carries the *mecA* regulatory genes *mecI* and *mecR*, an insertion sequence element (IS431*mec*) and a unique cassette of recombinase genes (*ccr*), which are responsible for SCC*mec* chromosomal integration and excision. Eight different types of SCC*mec* elements, type I-V, have been identified so far, based on the classes of *mecA* gene and *ccr* gene complexes (Ito *et al.*, 2009). Notably, type II and III elements contain, besides *mecA*, multiple determinants for resistance against non-β-lactam antibiotics. Accordingly, type II and III SCC*mec* elements are responsible for multidrug resistance in nosocomial MRSA isolates. Some SCC*mec* elements (e.g. type IV SCC*mec*), contain no other resistance gene than *mecA*, and they are significantly smaller compared to for example the type II and III elements. This might serve as an evolutionary advantage, making it easier for these mobile genetic elements to spread across bacterial populations. Phylogenetic analyses of the genes encode by SCC*mec* elements showed distant relationships with homologues in other *S. aureus* genomes and suggest foreign origins for these genes.
Vancomycin resistance has been first reported for *Enterococcus faecium* (Leclercq *et al.*, 1989), and transfer of vancomycin resistance from enterococci, such as *Enterococcus faecalis*, to *S. aureus* has been shown to occur (Noble *et al.*, 1992). Vancomycin has long been a last resort antibiotic for multiple resistant *S. aureus* strains, but already in 1996 a strain was isolated, which showed a reduced sensitivity towards vancomycin (Hiramatsu *et al.*, 1997). Shortly afterwards, additional strains were isolated in different countries that were designated as **v**ancomycin-**i**ntermediately resistant ***S. a**ureus* (VISA). These strains show a significantly thickened cell wall, which allows them to sequester more vancomycin than non-VISA strains, thereby preventing the detrimental effects of this antibiotic (Cui *et al.*, 2003). A search for the genetic basis of the lowered vancomycin sensitivity of the *S. aureus* Mu50 strain revealed that important genes for cell wall biosynthesis and intermediary metabolism have mutations compared to MRSA strains, which might lead to altered expression of genes involved in the cell wall metabolism and a thickened cell wall (Avison *et al.*, 2002). The first highly vancomycin resistant strain was isolated in 2002 (Weigel *et al.*, 2003). This strain was shown to carry a plasmid, which contains, among other resistance genes, the *vanA* gene plus several additional genes required for vancomycin resistance. The proteins encoded by these genes are responsible for replacing the C-terminal D-alanyl-D-alanine (D-ala-D-ala) of the disaccharide pentapeptide cell wall precursor with a depsipeptide, D-alanyl-D-lactate (D-ala-D-lac), thereby lowering the cell wall affinity for vancomycin (Bugg *et al.*, 1991).

Export of virulence factors from the cytoplasm

As most proteinaceous virulence factors are displayed at the surface of the staphylococcal cell or released into the medium, it is important for our understanding of the pathogenic potential of these organisms to map their pathways for protein transport. While specific questions relating to surface display or secretion of particular virulence factors have been addressed for several years, more holistic studies on the genomics and proteomics of these processes have been documented in the scientific literature only very recently. Moreover, no systematic analysis of pathways and cellular machinery for protein transport has thus far been performed for staphylococci. This review is aimed at filling this knowledge gap. To do so, we have taken

full advantage of the availability of six completely sequenced and annotated *S. aureus* genomes and one of the two sequenced *S. epidermidis* strains, as well as recently published data on the analysis of staphylococcal cell wall- and exoproteomes. Additionally, we have combined the published information with bioinformatics-derived data on all potential signals for protein export from the cytoplasm and secretion into the extracellular milieu, or retention in the membrane or cell wall. Since polytopic membrane proteins do not appear to have major direct roles in virulence other than causing drug resistance, such membrane proteins remain beyond the scope of this review. Furthermore, since the secretome of *B. subtilis* has been characterized extensively, both at the level of the protein export machinery and the exoproteome, we have compared the staphylococcal secretomes with that of *B. subtilis*. To our knowledge this has resulted in the first "comparative secretomics" study.

S. aureus strains suitable for comparative secretomics

Fourteen sequenced and fully annotated genomes of *S. aureus* are available in public databases (Table 2; http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi) and thirteen of these genomes were used in the present study. These include one of the first hospital-acquired MRSA isolates, *S. aureus* COL (Gill *et al.*, 2005), which has been widely used in research on staphylococcal methicillin and vancomycin resistance. The sequenced MRSA252 strain (Holden *et al.*, 2004) is a hospital-acquired epidemic strain, which was isolated from a patient who died as a consequence of septicemia. The sequenced MSSA476 strain (Holden *et al.*, 2004) is a community-acquired invasive strain that is penicillin- and fusidic acid-resistant, but susceptible to most commonly used antibiotics. *S. aureus* Mu50 and N315 (Kuroda *et al.*, 2001) are hospital-acquired MRSA strains isolated from Japanese patients. In addition, the Mu50 strain displays vancomycin intermediate sensitivity. The *S. aureus* Mu3 strain was the first isolated strain from a healthy carrier in Brazil that showed vancomycin-resistance (Hiramatsu *et al.*, 1997; Neoh *et al.*, 2008). The community-acquired *S. aureus* strains MW2 (Baba *et al.*, 2002), USA300 and USA300_TCH1516 (Diep *et al.*, 2006) are highly virulent MRSA strains, isolated in the USA. Both *S. aureus* JH1 and JH9 strains are MRSA strains that were isolated from one patient undergoing vanconcomycin treatment on different time points. The JH1 strain was the earliest strain isolated from the patient. The JH9 strain was isolated at a later stage of the treatment and was diagnosed as a vancomycin-resistant strain (Mwangi *et al.*, 2007). Comparison between these two strains would gain insight in the evolution of isogenic strains and the acquirement of vancomycin resistance under antibiotic pressure. The Newman strain (Baba *et al.*, 2008) was isolated from a human infection (Duthie and Lorenz, 1952) and has been widely used as a research strain due to its robust virulence phenotypes. Finally, the NCTC 8325 strain (Gillaspy *et al.*, 2006) is generally regarded as the prototypical strain for all genetic midifications in order to address specific gene regulatory and virulence traits. Furthermore, the sequence of *S. aureus* RF122, a strain that is associated with mastitis in cattle, is now also available in the NCBI database (Herron *et al.*, 2002), but has not been included in the present review which is primarily focused on staphylococcal pathogenicity towards humans. Secretome predictions for this strain are presented in Appendix IIIH. Using Multilocus Sequence Typing with seven housekeeping genes of the different *S. aureus* strains, Holden *et al.* (Holden *et al.*, 2004) showed that the MRSA252 strain is phylogenetically most distantly related to the other sequenced strains, while the Mu50 and N315 strains are indistinguishable by MLST, and the same is true for the MSSA476 and MW2 strains. The COL and NCTC8325 strains are relatively closely related to

each other. However, analysis of the two major pathogenicity islands present in all these strains shows that the distribution of these pathogenicity islands gives contradictory results on phylogenetic relationships of the sequenced *S. aureus* strains (Baba *et al.*, 2008).

Table 2. Sequenced and annotated genomes of *S. aureus* strains

Strain	Origin[a]	Genome size (kbp)		Nr. Of protein encoding genes	
		Chromosome	Plasmid	Chromosome	Plasmid
COL	HA- MRSA	2809	4	2615	3
JH1	HA- MRSA	2907	30	2747	33
JH9	HA- VISA	2907	3	2697	29
MRSA252	HA- MRSA	2903	-	2656	-
MSSA476	CA- MSSA	2800	21	2579	19
Mu3	HA- VISA	2880	-	2698	-
Mu50	HA- VISA	2879	25	2697	34
MW2	CA- MRSA	2820	-	2632	-
N315	HA- MRSA	2815	25	2588	31
NCTC8325	HA- MSSA	2821	-	2892	-
Newman	HA- MRSA	2879	-	2614	-
USA300	CA- MRSA	2873	3,4,37	2560	5,3,36
USA300_TCH1516	CA- MRSA	2873	27,20	2657	26,20
RF122	Bovine mastitis	2743	-	2515	-

[a] *HA-MRSA: Hospital-acquired MRSA; CA-MRSA: Community-acquired MRSA*

Sequenced and annotated genomes of other staphylococcal species, such as *S. epidermidis*, *Staphylococcus haemolyticus* and *Staphylococcus carnosus*, are also publicly available. However, with the exception of the *S. epidermidis* strain ATCC 12228 (Zhang *et al.*, 2003), these are not included in the present review, which is focused primarily on *S. aureus*. A comparative genomic analysis of *S. aureus* COL, Mu50, MW2, N315 and the sequenced *S. epidermidis* strains RP62A and ATCC 12228 has revealed that these species and strains have a set of 1681 genes in common (Gill *et al.*, 2005). In contrast, 454 genes are present in the *S. aureus* strains, but not in *S. epidermidis*, whereas 286 genes are present in *S. epidermidis*, but not in *S. aureus*. Most of the strain-specific and species-specific genes can be related to the presence or absence of particular prophages and genomic islands.

Pathways for staphylococcal protein transport

The bacterial machinery for protein transport is currently best-described for *Escherichia coli* (Gram-negative) and *B. subtilis* (Gram-positive) (for reviews see (de Keyzer *et al.*, 2003; Tjalsma *et al.*, 2000; Tjalsma *et al.*, 2004). Many of the known components that are involved in the different routes for protein export from the cytoplasm and post-translocational modification of exported proteins in these organisms are also conserved in *S. aureus* and *S. epidermidis* (Table 3). In general, proteins that are exported are synthesized with an N-terminal signal peptide, which directs them into a particular transport pathway. Consequently, the presently known signal peptides are classified according to the export pathway into which they direct the corresponding proteins, or the type of signal peptidase that is responsible for their removal (processing) upon membrane translocation.

Table 3. Secretion machinery of *S. aureus*, *S. epidermidis* and *B. subtilis*

Sec-pathway		***S. aureus***	***S. epidermidis***	***B. subtilis***
Chaperone	Ffh	+	+	+
	FtsY	+	+	+
	FlhF	-	-	+
	CsaA	-	-	+
Translocation motor	SecA1	+	+	+
	SecA2	+	+	-
Translocation channel	SecY1	+	+	+
	SecY2	+	+	-
	SecE	+	+	+
	SecG	+	+	+
	SecDF	+	+	+
	YajC (YrbF)	+	+	+
Lipid modification	Lgt	+	+	+
Sec-pathway		***S. aureus***	***S. epidermidis***	***B. subtilis***
Signal peptidase	SpsA (inactive)	+	+	-
	SpsB (SipSTUV)	+	+[a]	+
	SipW (ER-type)	-	-	+
	LspA	+	+[b]	+
Folding catalyst	PrsA	+[c]	+	+
	BdbC	-	-	+
	DsbA (BdbD)	+	+	+
Cell wall anchoring	SrtA	+	+	-
	SrtB	+	-	-
	SrtC	-	+[d]	-
	SrtD	-	-	+
Tat-pathway				
Translocase	TatA	+	-	+
	TatC	+	-	+
Pseudopilin pathway				
	ComGA	+	+	+
	ComGB	+	+	+
	ComC	+	+	+
Bacteriocins				
	Bacteriocin-specific ABC-transporters	?	?	+
Holins				
	CidA (holin)	+	+	+
	LrgA (anitholin)	+	+	+
Ess				
	EsaA	+	-	+
	EsaB	+	-	+
	EsaC	+[e]	-	-
	EssA	+	-	-
	EssB	+	-	+
	EssC	+	-	+

Based on BLAST searches with the corresponding proteins of B. subtilis *in the finished genome database (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi).*
[a] *Two potentially active type I SPases are present in this strain and share homology to* B. subtilis *SipS and SipU*
[b] *Two LspA proteins present in this strain*
[c] *This protein is truncated at the C-terminus in the* S. aureus *JH9 strain*
[d] *The genome of* S. epidermidis *RP62A only contains a srtA gene, whereas the genome of* S. epidermidis *ATCC12228 also contains a srtC gene*
[e] *This protein is missing in the* S. aureus *MRSA252 strain*

The staphylococcal protein export pathways that have been characterized experimentally or that can be deduced from sequenced genomes are schematically shown in Figure 4 and will be discussed in the following sections. Since these pathways are likely to be used for the export

of virulence factors to the cell surface and the milieu of the host, Figure 4 can be regarded as a subcellular road map to staphylococcal pathogenesis.

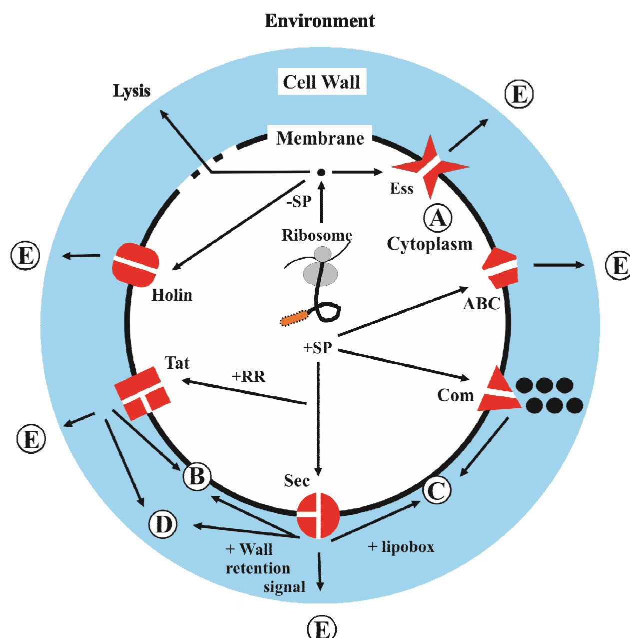

Figure 4. The staphylococcal "pathways to pathogenesis". *Schematic representation of a staphylococcal cell with potential pathways for protein sorting and secretion. (**A**) Proteins without signal peptide reside in the cytoplasm. (**B**) Proteins with one or more transmembrane spanning domains can be inserted into the membrane via the Sec, Tat or Com pathways. (**C**) Lipoproteins are exported via the Sec pathway and after lipid-modification anchored to the membrane. (**D**) Proteins with cell wall retention signals are exported via the Sec, Tat or Com pathways and retained in the cell wall via covalent-, or high-affinity binding to cell wall components. (**E**) Exported proteins with a signal peptide and without a membrane or cell wall retention signal can be secreted into the extracellular milieu via the various indicated pathways.*

Components of the general secretory (Sec) Pathway

The most commonly used pathway for bacterial protein transport is the general secretory (Sec) pathway. Specifically this pathway is responsible for the secretion of the majority of the proteins found in the exoproteome of *B. subtilis* and this is probably also the case for most other Gram-positive bacteria, including *S. aureus* (Tjalsma *et al.*, 2004). Unfortunately, there are only very few published data available concerning the Sec pathway of *S. aureus* and, therefore, we will fill in the current knowledge gaps with data obtained from studies in *B. subtilis* or *E. coli*. Proteins that are exported via the Sec-pathway contain signal peptides with recognition sites for so-called type I or type II signal peptidases (SPases). Notably, type II SPase recognition sites overlap with the recognition sites for the diacylglyceryl transferase Lgt. Precursor proteins with a type II SPase recognition sequence are lipid-modified prior to processing and the resulting mature proteins are retained as lipopoteins in the cytoplasmic membrane via their diacylglyceryl moiety. Furthermore, the Sec-dependent export of proteins can be divided into three stages: a) targeting to the membrane translocation machinery by export-specific or general chaperones, b) translocation across the membrane by the Sec machinery, and c) post-translocational folding and modification. If the translocated proteins of Gram-positive bacteria lack specific retention signals for the membrane or cell wall, they are secreted into the growth medium.

Preprotein targeting to the membrane

In *B. subtilis* the only known secretion-specific chaperone is the signal recognition particle (SRP), which consists of the small cytoplasmic RNA (scRNA), the histon-like protein HBsU and the Ffh protein. Ffh and HBsU bind to different moieties of the scRNA. Studies in *E. coli* have shown that, upon emergence from the ribosome, the signal peptide of a nascent secretory protein can be recognized by several cytoplasmic chaperones and/or targeting factors, such as Ffh or Trigger Factor (TF) (Eisner *et al.*, 2003). In contrast to Ffh, which is required for co-

translational protein export in *E. coli*, the cytoplasmic chaperone SecB has mainly been implicated in post-translational protein targeting. Notably however, SecB is absent from the sequenced Gram-positive bacteria, including *S. aureus* and *B. subtilis*. Most likely, ribosome-nascent chain complexes of *S. aureus* are thus targeted to the membrane by SRP, which, by analogy to *B. subtilis* and *E. coli* will probably involve the SRP receptor FtsY. At the membrane, the nascent preprotein will be directed to the translocation machinery. This process is likely to be stimulated by negatively charged phospholipids (De Leeuw *et al.*, 2000), the Sec translocon (Bibi, 1998; De Leeuw *et al.*, 2000) and/or the SecA protein (Bunai *et al.*, 1999). In this respect SecA may not only function as the translocation motor (see below), but also as a chaperone for preprotein targeting (Herbort *et al.*, 1999). While it has been shown that Ffh is essential for growth and viability in *E. coli* and *B. subtilis*, this does not seem to be the case in all bacteria. For example, Ffh, FtsY and scRNA are not essential in *Streptococcus mutans*. In this organism the SRP is merely required for growth under stressful conditions, such as low pH (<pH 5), high salt (3.5% NaCl) or the presence of H_2O_2 (0.3 mM). This suggests that SRP has an important role in the export of proteins to the membrane or cell wall that protect *S. mutans* against environmental insults (Hasona *et al.*, 2005).
For *B. subtilis*, it has been proposed that the general chaperone CsaA may have a role in preprotein targeting to the membrane, similar to SecB of *E. coli*. This view is supported by the observation that the *B. subtilis* CsaA protein has binding affinity for SecA and preproteins (Müller *et al.*, 2000). However, CsaA is not conserved in *S. aureus*. Therefore, it remains to be investigated whether other chaperones with a preprotein targeting function are present in *S. aureus*.

Translocation across the membrane

As deduced from the known genome sequences, the translocation machinery of *S. aureus* consists of several Sec proteins. The mode of action of these proteins has been studied in great detail in *E. coli* (de Keyzer *et al.*, 2003; Vrontou and Economou, 2004). After binding of a preprotein to a SecA dimer, the SecA molecules will bind ATP, resulting in conformational changes that promote their insertion together with the preprotein into the membrane-embedded translocation channel. Subsequent hydrolysis of ATP causes SecA to release the preprotein, return to its original conformation, and deinsert from the translocation channel. Repeated cycles of ATP-binding and hydrolysis by SecA, together with the proton-motive force, drive further translocation of the preprotein across the membrane. The translocation channel is essentially formed by the SecE and SecY proteins, which are conserved in all bacteria (Veenendaal *et al.*, 2004). An additional non-essential channel component is SecG, which serves to increase the translocation efficiency. While the SecY proteins of different bacteria show a relatively high degree sequence similarity, the SecE and SecG proteins, though present in all bacteria, are less well-conserved. Specifically, the SecE and SecG proteins in *B. subtilis*, *S. aureus* and *S. epidermidis* are considerably shorter than the equivalent proteins of *E. coli*. Although SecA and SecY of *S. aureus* (in this review referred to as SecA1 and SecY1) have not yet been characterized functionally, they are of major importance for growth of *S. aureus*. This was demonstrated with the help of specific antisense RNAs (Ji *et al.*, 2001). Upon *secA* antisense induction a strong growth defect was observed, and *secY* antisense induction turned out to be lethal.
Remarkably, the genome of *S. aureus* contains a second set of *secA* and *secY* genes, referred to as *secA2* and *secY2*, respectively. In contrast to the SecA1 and SecY1 proteins, these

homologues are not essential for growth and viability. It is presently unknown whether SecA2 and SecY2 transport specific proteins across the membrane of *S. aureus*. However, it has been shown for other pathogenic Gram-positive bacteria, which also possess a second set of SecA and SecY, that these proteins are required for the transport of certain proteins related to virulence. In *Streptococcus gordonii*, the export of GspB, a large cell-surface glycoprotein that contributes to platelet binding, seems to be dependent on the presence of SecA2 and SecY2 (Bensing and Sullam, 2002). This protein contains large serine-rich repeats, an LPxTG motif for cell wall anchoring (see below), and a very large signal peptide of 90 amino acids. In *Streptococcus parasanguis* two other proteins, FimA and Fap1, are known to be secreted via SecA2-dependent membrane translocation. FimA is a (predicted) lipoprotein, which is a major virulence factor implicated in streptococcal endocarditis. The FimA homologue in *S. aureus* is a manganese-binding lipoprotein (MntA), associated with an ATP-binding cassette (ABC) transporter. Fap1 of *S. parasanguis* is involved in adhesion to the surface of teeth. Like GspB of *S. gordonii*, Fap1 has a long signal peptide of 50 amino acids, serine-rich repeats and an LPxTG motif for cell wall anchoring. To date, it is not known what determines the difference in the specificity of SecA1/SecY1 and SecA2/SecY2 translocases. However, for *S. gordonii* it has been shown that Gly residues in the signal peptide are important for directing GpsB to the SecA2/SecY2 translocon (Bensing *et al.*, 2007). It is also not known whether the SecA2/SecY2 shares SecE and/or SecG with the SecA1/SecY1 translocase, and whether these translocases function completely independently from each other or whether mixed translocases can occur. Clearly, the *secE* and *secG* genes are not duplicated in *S. aureus*.

In *E. coli*, the heterotrimeric SecYEG complex is associated with another heterotrimeric complex that is composed of the SecD, SecF and YajC proteins (Nouwen *et al.*, 2005). This complex has been shown to be involved in the cycling of SecA (Driessen *et al.*, 1998) and release of the translocated protein from the translocation channel (Matsuyama *et al.*, 1993). SecD and SecF are separate, but structurally related proteins in most bacteria, including *E. coli*. Interestingly, in *B. subtilis* and *S. aureus*, natural gene fusions between the *secD* and *secF* genes are observed. Accordingly, the corresponding SecDF proteins can be regarded as molecular “Siamese twins” (Bolhuis *et al.*, 1998). Unlike SecA, SecY and SecE, the SecDF protein of *B. subtilis* is not essential for growth and viability and its role in protein secretion is presently poorly understood (Bolhuis *et al.*, 1998). *B. subtilis secDF* mutants only showed a mild secretion defect under conditions of high-level synthesis of secretory proteins. The known SecDF proteins have 12 (predicted) transmembrane domains with two large extracytoplasmic loops between the first and second transmembrane segments, and between the seventh and eighth transmembrane segments. For *E. coli* SecD it has been shown that small deletions in the large extracytoplasmic loop result in a malfunctioning of the protein, while the stability of the SecD/F-YajC complex is not affected (Nouwen *et al.*, 2005). It has therefore been proposed that this loop in SecD might provide a protective structure in which translocated proteins can fold more efficiently. The large extracytoplasmic loop in SecF has been proposed to interact with SecY, thereby stabilizing the translocation channel formed by SecYEG. Homologues of the *E. coli* YajC protein are present in many bacteria, including *S. aureus* and *B. subtilis* (YrbF), but their role in protein secretion has not been established yet. It is presently not known whether the *S. aureus* SecDF-YajC complex associates specifically with the SecA1/SecY1 translocase, the SecA2/SecY2 translocase, or both translocases.

Type I Signal peptidases

Signal peptides of preproteins are cleaved during or shortly after translocation by SPase I or SPase II, depending on the nature of the signal peptide (Tjalsma *et al.*, 2001; van Roosmalen *et al.*, 2004). The *B. subtilis* chromosome encodes five type I SPases, named SipS, SipT, SipU, SipV and SipW (van Dijl *et al.*, 1992; Tjalsma *et al.*, 1997; Tjalsma *et al.*, 1998). Two of these, SipS and SipT, are of major importance for the processing of secretory preproteins, growth and viability. In *S. aureus* only two SPase I homologues are present, SpsA and SpsB. The catalytically active SPase I in *S. aureus* is SpsB, which is probably essential for growth and viability (Cregg *et al.*, 1996). This SPase can be used to complement an *E. coli* strain that is temperature-sensitive for preprotein processing. In general, type I SPases recognize residues at the -1 and -3 positions relative to the cleavage site (van Roosmalen *et al.*, 2004). For *B. subtilis* it has been shown that all secretory proteins identified by proteomics have Ala at the -1 position, and 71% of these secretory proteins have Ala at the -3 position (Tjalsma *et al.*, 2004). In contrast, various residues are tolerated at the -2 position, including Ser, Lys, Glu, His, Tyr, Gln, Gly, Phe, Leu, Ala, Asp, Asn, Trp and Pro. Interestingly, Bruton *et al.* (Bruton *et al.*, 2003) studied the cleavage sites in substrates of SpsB of *S. aureus* and showed that this enzyme has a preference for basic residues at the -2 position and tolerance for hydrophobic residues at this position. However, an acidic residue at the -2 position resulted in a significantly reduced rate of processing. The second SPase I homologue of *S. aureus* (SpsA) appears to be inactive, since it lacks the catalytic Ser and Lys residues that are, respectively, replaced with Asp and Ser residues. The presence of an apparently catalytically inactive SpsA homologue is a conserved feature of all staphylococci with sequenced genomes. Notably, in addition to an inactive SpsA homologue, *S. epidermidis* contains two SpsB homologues that respectively show the greatest similarity to SipS and SipU of *B. subtilis*. To date, it is not known whether the inactive SpsA homologues contribute somehow to protein secretion in these organisms.

Lipid-modification of lipoproteins

In *E. coli*, lipid-modification of prolipoproteins involves three sequential steps that are catalyzed by cytoplasmic membrane-bound proteins. The first step involves the transfer of a diacylglyceryl group from phosphatidylglycerol to the sulfhydryl group of the invariant Cys residue that is present at the +1 position of the signal peptide cleavage site in lipoprotein precursors. This step is catalyzed by a phosphatidyl glycerol diacylglyceryl transferase (Lgt) as was shown for *E. coli* by Sankaran *et al.* (Sankaran and Wu, 1994). The recognition sequence for Lgt, which includes the Cys residue that becomes diacylglyceryl-modified, is known as the lipobox. The lipid-modification of the lipobox Cys residue is necessary for the lipoprotein-specific type II signal peptidase (LspA) to recognize and cleave the signal peptide of a prolipoprotein, which represents the second step in lipoprotein modification. The third step involves the transfer of an N-acyl group by an N-acyl transferase (Lnt), resulting in the formation of N-acyl diacylglycerylcysteine at the N-terminus of the mature lipoprotein. Although Lgt and LspA are present in most, if not all bacteria, Lnt is only present in Gram-negative bacteria (Tjalsma *et al.*, 2001). As for other Gram-positive bacteria, no homologue of Lnt could be detected in the genomes of *S. aureus* or *S. epidermidis* (Stoll *et al.*, 2005), which suggests that the lipoproteins of these organisms are not N-acylated.

The *S. aureus* Lgt is a protein of 279 amino acids that contains a highly conserved HGGLIG motif (residues 97 to 102). Although the His residue in this motif was shown to be essential for catalytic activity of the *E. coli* Lgt (Sankaran *et al.*, 1997), it is not strictly conserved in all known Lgt proteins. On the other hand, the strictly conserved Gly at position 103 of *E. coli* Lgt, which is equivalent to Gly98 of *S. aureus* Lgt, is required for activity of this protein. Stoll *et al.* (Stoll *et al.*, 2005) showed that a *S. aureus lgt* mutation has no effect on growth in broth as was also observed for *B. subtilis* (Leskelä *et al.*, 1999). Nevertheless, the absence of Lgt has a considerable effect on the induction of an inflammatory response. Importantly, lipid modification serves to retain exported proteins at the membrane-cell wall interface. This is particularly relevant for Gram-positive bacteria, which lack an outer membrane that represents a retention barrier for exported proteins. In the absence of Lgt, *B. subtilis* cells release a variety of lipoproteins into the extracellular milieu, both in the form of unmodified precursor proteins and alternatively processed mature proteins that lack the N-terminal Cys residue (Antelmann *et al.*, 2001). Similarly, the *S. aureus lgt* mutation resulted in the shedding of certain abundant lipoproteins, such as OppA, PrsA and SitC, into the broth. These lipoproteins are normally retained in the membrane or cell wall of *S. aureus*.

Type II Signal Peptidase

As described above, lipoprotein signal peptides of prolipoproteins are cleaved by type II SPases after the Cys residue in the lipobox is modified by Lgt. Although *B. subtilis* and many other bacteria contain only one copy of the *lspA* gene, some contain a second copy, such as *S. epidermidis*, *Bacillus licheniformis* and *Listeria monocytogenes*. LspA is a membrane protein that spans the membrane four times and both the N- and C-termini are facing the cytoplasmic side of the membrane (Tjalsma *et al.*, 1997; van Roosmalen *et al.*, 2001). Six amino acid residues are important for SPase II activity, of which two Asp residues form the active site (Tjalsma *et al.*, 1997). While processing of lipoproteins by LspA is essential for growth and viability for *E. coli* and other Gram-negative bacteria (Wu, 1996), it is not essential for *B. subtilis* (Tjalsma *et al.*, 1999) and other Gram-positive bacteria, such as *Lactococcus lactis* (Venema *et al.*, 2003). This suggests that processing of prolipoproteins is not essential for their functionality. The latter view is supported by the fact that PrsA, a lipoprotein required for correct folding of translocated proteins, is essential for viability of *B. subtilis* (Kontinen and Sarvas, 1993). In the absence of LspA, some of the lipoproteins of *B. subtilis* are processed in an alternative way by yet unidentified proteases and activity of unprocessed lipoproteins in *lspA* mutants is reduced. Also, in *B. subtilis* the secretion of the non-lipoprotein AmyQ was severely reduced (Tjalsma *et al.*, 1999). This reduction might be the consequence of a malfunction of non-modified PrsA in AmyQ folding. Although most *lspA* mutants have been studied in Gram-negative bacteria and a few non-pathogenic Gram-positive bacteria (Tjalsma *et al.*, 1999; Venema *et al.*, 2003), Sander *et al.* (Sander *et al.*, 2004) showed a severe attenuated phenotype of *lspA* mutants of the pathogen *Mycobacterium tuberculosis*, which implies an important role for lipoprotein-processing by LspA during infection of *M. tuberculosis*. In *S. aureus* both the *lspA* and *lgt* genes are present in single copy in the genomes of all six sequenced strains. Interestingly, one of the two LspA homologues in *S. epidermidis* (125 amino acids) is considerably shorter than other known LspA proteins, including its large paralogue (177 amino acids). This is mainly the result of an additional N-terminal transmembrane domain in the large LspA proteins. As a result the short *S. epidermidis* LspA protein is predicted to have three membrane spanning domains, with the

N-terminus located on the outside of the cell, the C-terminus on the inside of the cell and the (putative) active site Asp residues located on the outer surface of the cytoplasmic membrane.

Signal peptide peptidase

After translocation and processing of the preproteins by signal peptidases, the signal peptides are rapidly degraded by signal peptide peptidases (SPPases). In *B. subtilis* two SPPases, TepA and SppA, are known to be involved in translocation and processing of preproteins (Bolhuis *et al.*, 1999). While TepA is required for translocation and processing of preproteins, SppA is only required for efficient processing of preproteins. Remarkably, no homologues of SppA or TepA were detectable by BLAST searches in the sequenced genomes of *S. aureus* and *S. epidermidis*. As reported by Meima and van Dijl (Meima and van Dijl, 2003), *L. lactis* contains a protein that shows limited similarity to TepA of *B. subtilis* and ClpP of *C. elegans*, suggesting that this protein might be an SPPase-analog of *L. lactis*. In *S. aureus* and *S. epidermidis* this protein homologue also seems to be present and is predicted to be a cytoplasmic membrane protein (our unpublished observations).

Folding catalysts (PrsA and BdbD)

Proteins that are transported across the membrane in a Sec-dependent manner emerge at the extracytoplasmic membrane surface in an unfolded state. These proteins need to be rapidly and correctly folded into their native and protease-resistant conformation, before they are degraded by proteases in the cell wall or extracellular environment (Sarvas *et al.*, 2004). An important folding catalyst in *B. subtilis* is PrsA, which shows homology to peptidyl-prolyl cis/trans-isomerases. PrsA is a lipoprotein (see also section “Lipoproteins”) that is essential for efficient protein secretion and cell viability in *B. subtilis* (Sarvas *et al.*, 2004; Kontinen and Sarvas, 1993). Studies on the effects of PrsA depletion showed that the relative amounts of extracellular proteins from PrsA-depleted cells, were significantly reduced (Vitikainen *et al.*, 2004). The solution structure of PrsA has been solved (Heikkinen *et al.*, 2009), but no data has been published on *S. aureus* mutants lacking PrsA and it will be interesting to investigate whether PrsA is also essential for viability and virulence of this organism. It has already been shown that *S. aureus* lacking Lgt, releases an increased amount of PrsA into the extracellular milieu (Stoll *et al.*, 2005), which might indicate that (most) pre-PrsA is not fully functional, but sufficient for viability. The observation by Stoll *et al.* (Stoll *et al.*, 2005) also shows that, like in *B. subtilis* (Antelmann *et al.*, 2001), the unmodified pre-PrsA is not effectively retained in the cytoplasmic membrane.

Other proteins that are involved in proper folding of extracellular proteins in *B. subtilis* are the membrane proteins BdbC and BdbD, which are involved in the formation of disulfide bonds. Both proteins have been shown to be necessary for the stabilization of the membrane- and cell wall-associated pseudopilin ComGC (Meima *et al.*, 2002). This protein, which is required for DNA binding and uptake during natural competence, contains an intramolecular disulfide bond (Chung *et al.*, 1998). Both BdbC and BdbD are also important for the folding of heterologously produced *E. coli* PhoA, which contains two disulfide bonds, into an active and protease-resistant conformation (Bolhuis *et al.*, 1999; Meima *et al.*, 2002). Though a homologue of BdbD (named DsbA) is present in *S. aureus*, there is no homologue of BdbC in this organism. The same appears to be true for *S. epidermidis*. Nevertheless, measurements of the redox potential of purified DsbA indicate that this protein can act as an oxidase, and this

view is confirmed by complementation studies in a *dsbA* mutant strain of *E. coli* (Dumoulin *et al.*, 2005). The absence of a BdbC homologue from staphylococci is remarkable, since *B. subtilis* BdbC and BdbD are jointly required in the folding of ComGC and *E. coli* PhoA. Notably, all sequenced *S. aureus* genomes encode homologues of ComGC, including the Cys residues that form the disulfide bond in *B. subtilis* ComGC. This raises the question whether ComGC of *S. aureus* does indeed contain a disulfide bond and, if so, which protein(s) are involved in the formation of this disulfide bond. DsbA would be a candidate for this task since it has been shown that this *S. aureus* protein can functionally replace BdbB, BdbC and BdbD in the production of ComGC, *E. coli* PhoA and the S-S bond-containing sublancin 168 in *B. subtilis* (Kouwen *et al.*, 2007). This idea is further supported by the findings of Heras et al. (Heras *et al.*, 2008) that the oxidized and reduced states of DsbA are energetically equivalent, which suggests that this facilitates the reoxidation of DsbA, likely by extracellular oxidants. Notably, *S. aureus* DsbA was shown to be a lipoprotein that does not seem to contribute to the virulence of this organism as tested in mouse and *Caenorhabditis elegans* models (Dumoulin *et al.*, 2005). Furthermore, DsbA was shown to be dispensable for β-hemolysin activity, despite the fact that this protein contains a disulfide bond, which is required for activity (Dziewanowska *et al.*, 1996). Therefore, the biological function of DsbA in staphylococci remains to be elucidated.

Twin-arginine translocation (Tat) pathway

The Tat-pathway exists in many bacteria, archaea, and chloroplasts. This pathway has been named after the consensus double (twin) Arg residues that are present in the signal peptide. The twin Arg residues are part of a motif that directs proteins specifically into the Tat pathway. In contrast to the Sec-machinery where only unfolded proteins are translocated across the membrane, the Tat-machinery is capable of translocating folded proteins. In Gram-negative bacteria, streptomycetes, mycobacteria and chloroplasts, an active Tat-pathway seems to require three core components, named TatA, TatB and TatC (Berks *et al.*, 2005; Dilks *et al.*, 2003; Mori and Cline, 2001; Robinson and Bolhuis, 2001; Yen *et al.*, 2002). In all Gram-positive bacteria except streptomycetes and *Mycobacterium smegmatis*, the Tat pathway involves only TatA and TatC (Dilks *et al.*, 2003; Yen *et al.*, 2002). Recent studies in *E. coli* and chloroplasts have resulted in a model that proposes a key role for TatB-TatC complexes in signal peptide reception and TatA-TatB-TatC complexes in preprotein translocation (Cline and Mori, 2001; Alami *et al.*, 2003). Interestingly, certain mutations in *E. coli* TatA have been shown to allow this protein to compensate for the absence of TatB (Blaudeck *et al.*, 2005). This demonstrated that TatA is intrinsically bifunctional, which is consistent with the fact that most Gram-positive bacteria lack TatB, but have TatA (Jongbloed *et al.*, 2005). In *B. subtilis*, two minimal TatA-TatC translocases with distinct specificities are active (Jongbloed *et al.*, 2004). While the constitutively expressed TatAy-TatCy translocase of *B. subtilis* is required for secretion of the protein with unknown function YwbN, the TatAd-TatCd translocase seems to be expressed only under conditions of phosphate starvation for secretion of the phosphodiesterase PhoD (Tjalsma *et al.*, 2000; van Roosmalen *et al.*, 2001). Most other Gram-positive bacteria that have *tatA* and *tatC* genes, including *S. aureus*, appear to have only one TatA-TatC translocase. The functionality of the *S. aureus* Tat translocase was recently demonstrated (Biswas *et al.*, 2009). In contrast to *S. aureus*, *S. epidermidis* seems to lack a Tat pathway.

Pseudopilin export (Com) pathway

In *B. subtilis* four proteins, ComGC, ComGD, ComGE and ComGG, have been identified with an N-terminal pseudopilin-like signal peptide (Tjalsma *et al.*, 2004; Tjalsma *et al.*, 2000). All four of these proteins are involved in DNA binding and uptake, and are localized in the membrane and cell wall. It is thought that these proteins form a pilus-like structure in the cell wall or modify the cell wall to provide a passage for DNA uptake. Translocation to the extracytoplasmic membrane surface is only possible when these proteins are processed by the pseudopilin-specific SPase ComC in *B. subtilis* (Dubnau, 1999). SPases of this type are bifunctional and do not only catalyze signal peptide cleavage, but also methylation of the N-terminus of the mature protein (Strom *et al.*, 1993). Furthermore, export and functionality of the four ComG proteins depends on the integral membrane protein ComGB and the traffic ATPase ComGA, which is located at the cytoplasmic side of the membrane (Chung and Dubnau, 1998; Hahn *et al.*, 2005). Homologues of ComC, ComGA, ComGB and ComGC, but not ComGD, ComGE and ComGG, are present in the six sequenced *S. aureus* strains. This suggests that the Com system of *S. aureus* is not involved in DNA uptake, but in another solute transport process.

ABC transporters

Bacteriocins are peptides or proteins that inhibit the growth of other bacteria. Most of the characterized bacteriocins can be divided into several classes, depending on specific posttranslational modifications, the presence and processing of particular leader peptides and the machinery for export from the cytoplasm. A well described class of bacteriocins is formed by the lantibiotics. Members of this class are composed of short peptides that contain post-translationally modified amino acids, like lanthionine and β-methyllanthionine (McAuliffe *et al.*, 2001). The production of bacteriocins in *S. aureus* has been described for various strains. *S. aureus* C55 produces the two lantibiotics C55α and C55β (Navaratna *et al.*, 1998). These lantibiotics are both encoded by a 32 kb plasmid, which is readily lost upon growth at elevated temperatures. C55α and C55β showed antimicrobial activity towards other *S. aureus* strains and *Micrococcus luteus*, but not towards *S. epidermidis*. Furthermore, the non-lantibiotics BacR1 (Crupper *et al.*, 1997), aureocin A53 (Netz *et al.*, 2001) and aureocin A70 (Netz *et al.*, 2002a; Netz *et al.*, 2002b) have been identified as bacteriocins with activity against a broad range of bacteria. The genes for both aureocins are located on a plasmid that is present in *S. aureus* strains that were isolated from milk. By analogy with well described bacteriocin export machinery in other organisms (Håvarstein *et al.*, 1995; Peschel *et al.*, 1997), it can be anticipated that all of the afore-mentioned bacteriocins are exported to the external staphylococcal milieu by dedicated ABC transporters. However, no experimental evidence for this assumption has been published for *S. aureus*. Notably, it has been demonstrated that the secretion of the lantibiotics epidermin and gallidermin of *S. epidermidis* Tü3298 and *Staphylococcus gallinarum*, respectively, is facilitated by so-called one-component ABC transporters. Specifically, the ABC-transporter GdmT has been implicated in the transport of these lantibiotics (Peschel *et al.*, 1997).

Holins

Holins are dedicated export systems for peptidoglycan-degrading endolysins that have been implicated in the programmed cell death of bacteria. These exporters, which are composed of homo-oligomeric complexes, can be subdivided into two classes, depending on their number of transmembrane segments. While class I holin subunits have three transmembrane segments, class II holin subunits have two transmembrane segments (Young and Bläsi, 1995). In *S. aureus* the *lrg* and *cid* operons are involved in murein hydrolase activity and antibiotic tolerance (Groicher *et al.*, 2000; Rice *et al.*, 2003). A disrupted *lrg* operon leads to an increase in murein hydrolase activity and a decrease in penicillin tolerance, and a disrupted *cid* operon leads to a decrease in murein hydrolase activity and an increase in penicillin tolerance. It is still unclear how the CidA and LrgA proteins are involved in these mechanisms, but these proteins display significant similarity to the bacteriaphage holin protein family, suggesting that they have a role in protein export. It has therefore been proposed that the CidA and LrgA proteins act on the murein hydrolase activity and antibiotic tolerance analogous to holins and antiholins, respectively (Bayles, 2000; Rice *et al.*, 2003). Sequence similarity searches show that the genes for LrgA and CidA are conserved in the six sequenced *S. aureus* strains, as well as *S. epidermidis* and *B. subtilis*. Notably, none of the three holins of *B. subtilis* was shown to be involved in the secretion of proteins to the extracellular milieu (Westers *et al.*, 2003; Tjalsma *et al.*, 2004).

Ess pathway

The ESX-1 or ESAT-6 secretion system (Ess) pathway has first been described for *M. tuberculosis*. It has been proposed that at least two virulence factors, ESAT-6 (**e**arly **s**ecreted **a**ntigen **t**arget **6** kDa) and CFP-10 (**c**ulture **f**iltrate **p**rotein **10** kDa), are secreted via this pathway in a Sec-independent manner (Berthet *et al.*, 1998; Sørensen *et al.*, 1995). As this pathway was discovered in mycobacteria, it is also known as the Snm pathway (**S**ecretio**n** in **m**ycobacteria; (Converse and Cox, 2005)). The genes for ESAT-6 en CPF-10 are located in conserved gene clusters, which also encode proteins with domains that are conserved in FtsK- and SpoIIIE-like transporters. These conserved **F**tsK/**S**poIIIE **d**omains have therefore been termed FSDs (Burts *et al.*, 2005). In other Gram-positive bacteria including *S. aureus*, *B. subtilis*, *Bacillus anthracis*, *Clostridium acetobutylicum* and *L. monocytogenes*, homologues of ESAT-6 have been identified (Pallen, 2002). The genes for these ESAT-6 homologues are also found in gene clusters that contain at least one membrane protein with a FSD. In *S. aureus*, two proteins named EsxA and EsxB have been identified that seem to be secreted via the Ess pathway (Burts *et al.*, 2005). The *esxA* and *esxB* genes are part of a cluster containing six other genes for proteins that have been implicated in the translocation of EsxA and EsxB. These include the cytoplasmic protein EsaB and the secreted protein EsaC (Burts *et al.*, 2008), as well as the predicted membrane proteins EsaA, EssA, EssB and EssC, of which EssC contains a FSD. Mutations in *essA*, *essB* or *essC* result in a loss of EsxA and EsxB production, which may relate to an inhibition of the synthesis of these proteins, or their folding into a protease-resistant conformation. EsaB is a negative regulator of EsaC and represses the production of EsaC in a post-transcriptional manner. EsaC, although secreted, does not contain the WxG-motif or any other signal peptide and it is still unclear how this protein is recognized by the Ess secretion pathway. All sequenced *S. aureus* strains contain this cluster of *esa*, *ess* and *esx* genes, but it seems to be absent from *S. epidermidis*.

Interestingly, the genes for EsxB and EsaC appear to be absent from the *S. aureus* MRSA252 strain. This implies that the Ess machinery of this strain may be required for the transport of only EsxA and perhaps a few other unidentified proteins. If so, EsaC would be dispensable for an active ESAT-6 pathway and might be specifically involved in the export of EsxB. This view is also suggested from the data published by Burts et al. (Burts *et al.*, 2008), which show that a *S. aureus* Newman strain lacking *esxB* does not produce EsaC. Alternatively, the Ess pathway could be inactive in the *S. aureus* MRSA252 strain due to the absence of EsaC.

Lysis

Various studies have shown that certain proteins with typical cytoplasmic functions and without known signals for protein secretion can nevertheless be detected on the extracellular proteome of different bacteria (Tjalsma *et al.*, 2004). Notably, many of these proteins, such as catalase, elongation factor G, enolase, glyceraldehyde-3-phosphate dehydrogenase, GroEL and superoxide dismutase, are amongst the most highly abundant cytoplasmic proteins. This makes it likely that they are detectable in the extracellular proteome due to cell lysis. Perhaps, such proteins are more resistant to extracytoplasmic degradation than other proteins that are simultaneously released by lysis. However, the possibility that the extracellular localization of typical cytoplasmic proteins is due to the activity of, as yet unidentified, export pathways cannot be excluded. Clearly, until recently this possibility did still apply for the EsxA, EsxB and EsaC proteins, which are now known to be exported via the Ess pathway. A clear indication that the presence of certain "cytoplasmic" proteins in the extracytoplasmic milieu of bacteria may relate to specific export processes was provided by Boël and co-workers (Boël *et al.*, 2004), who showed that 2-phosphoglycerate-dependent automodification of enolase is necessary for its export from the cytoplasm.

Properties of staphylococcal signal peptides and cell retention signals

Signal peptides

All proteins that have to be transported from the cytoplasm across the membrane to the extracytoplasmic compartments of the cell, or the extracellular milieu, need to contain a specific sorting signal for their distinction from resident proteins of the cytoplasm. The known bacterial sorting signals for protein export from the cytoplasm are signal peptides (von Heijne, 1990). These signal peptides can be classified by the transport and modification pathway into which they direct proteins. Presently, four different bacterial signal peptides are recognized that share a common architecture, but differ in details (Figure 5). Two of these direct proteins into the widely used Sec pathway, including the secretory (Sec type) signal peptides and the lipoprotein signal peptides. Proteins with Sec type or lipoprotein signal peptides are processed by different SPases (type I or type II SPases, respectively), and are targeted to different destinations. In *S. aureus* the proteins with Sec type signal peptides are processed by the type I SPase SpsB and targeted to the cell wall or extracellular milieu.

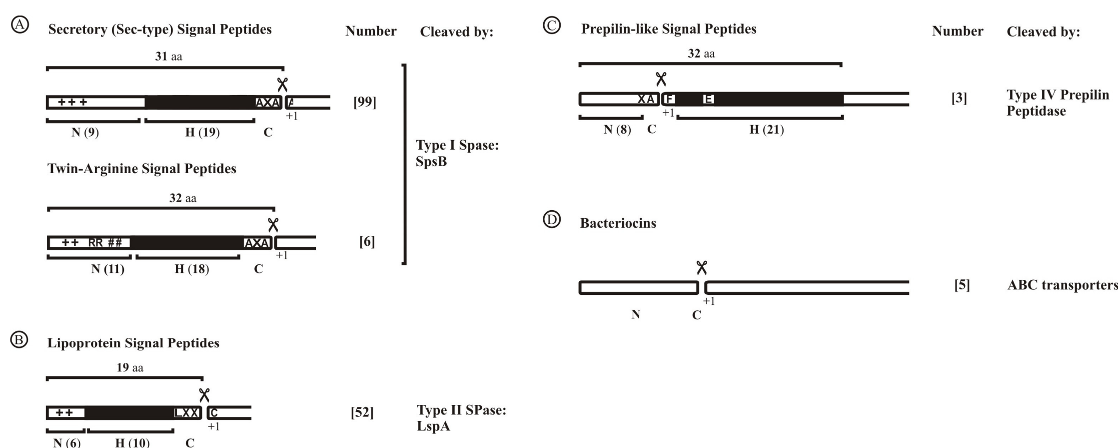

Figure 5. General properties and classification of* S. aureus *signal peptides. *Signal peptide properties are based on SPase cleavage sites and the export pathways via which the preproteins are exported. Predicted signal peptides (144) were divided into five distinct classes: secretory (Sec-type) signal peptides, twin-arginine (RR/KR) signal peptides, lipoprotein signal peptides, pseudopilin-like signal peptides, and bacteriocin leader peptides. Most of these signal peptides have a tripartite structure: a positively charged N-domain (N), containing lysine and/or arginine residues (indicated by +), a hydrophobic H-domain (H, indicated by a black box), and a C-domain (C) that specifies the cleavage site for a specific SPase. Where appropriate, the most frequently occurring amino acid residues at particular positions in the signal peptide or mature protein are indicated. Also, the numbers of signal peptides identified for each class and the respective SPase are indicated.*

The proteins with a lipoprotein signal peptide are lipid-modified by Lgt, prior to processing by the type II SPase LspA. In principle, these lipoproteins are retained at the membrane-cell wall interface, but they can be liberated from this compartment by proteolytic removal of the N-terminal Cys that contains the diacylglyceryl moiety (Antelmann *et al.*, 2001). Proteins with twin-arginine (RR) signal peptides appear to be processed by type I SPases, at least in *B. subtilis*, and targeted to the cell wall or extracellular milieu (Tjalsma *et al.*, 2004). The proteins with a pseudopilin signal peptide are processed by the pseudopilin signal peptidase ComC and most likely localized in the cytoplasmic membrane and cell wall. Finally, bacteriocins contain a completely different sorting and modification signal that is usually called the leader peptide. The known leader peptides show no resemblance to the afore-mentioned signal peptides. The export of bacteriocins via ABC-transporters results in their secretion into the extracellular milieu (Michiels *et al.*, 2001; Schnell *et al.*, 1988).

Sec type, lipoprotein and RR-signal peptides contain three distinguishable domains: the N-, H- and C-domains. The N-terminal domain contains positively charged amino acids, which are thought to interact with the secretion machinery and/or with negatively charged phospholipids in the membrane. The H-domain is formed by a stretch of hydrophobic amino acids which facilitate membrane insertion. Helix-breaking residues in the middle of the H-domain may facilitate H-domain looping during membrane insertion and translocation of the precursor protein. The subsequent unlooping of the H-domain would display the SPase recognition and cleavage site at the extracytoplasmic membrane surface where the catalytic domains of type I and type II SPases are localized (van Roosmalen *et al.*, 2004). Helix-breaking residues just before the SPase recognition and cleavage site would facilitate precursor processing by SPase I or II. In fact, these helix-breaking residues and the SPase cleavage site, respectively, define the beginning and the end of the C-domain. Notably, the C-domain of pseudopilin signal peptides is located between the N- and H-domains (Chung and Dubnau, 1995; Pugsley, 1993; Chung and Dubnau, 1998; Lory, 1998). Accordingly, processing

by pseudopilin-specific SPases, like ComC, takes place at the cytoplasmic side of the membrane and leaves the H-domain attached to the translocated protein.

While many proteins that end up in the extracellular milieu or the cell wall of Gram-positive bacteria have signal peptides, proteins without known export signals can also be found on these locations. The relative numbers of proteins without known signal peptides seem to vary per organism. While these numbers are relatively low for *B. subtilis* and *S. aureus*, they are high for group A *Streptococcus* and *M. tuberculosis* (Tjalsma *et al.*, 2004). As indicated above, some of the proteins without known export signals appear to be liberated from the cell by lysis, while others are actively exported, for example via the Ess pathway. Although the precise export signal in proteins secreted via the Ess pathway has not yet been defined, a WxG motif is shared by many of these proteins and may serve a function in protein targeting (Pallen, 2002). Furthermore, the signal for specific release of lysins via holins is presently not known.

Signal peptide predictions

Several prediction programs that are accessible through the world-wide web are useful tools to predict whether a given protein contains some sort of sorting signal or SPase cleavage site. The programs that we have used in this and other studies were: SignalP-NN and SignalP-HMM version 2.0 (Nielsen *et al.*, 1997), LipoP version 1.0 (Juncker *et al.*, 2003), PrediSi (Hiller *et al.*, 2004) and Phobius (Kall *et al.*, 2004). These programs have been designed to identify Sec type signal peptides, N-terminal membrane anchors (Phobius), or lipoprotein signal peptides in Gram-negative bacteria (LipoP). The TMHMM-program version 2.0 (Cserzö *et al.*, 1997) was used to exclude proteins with (predicted) multiple membrane spanning domains. Predictions for proteins containing a signal peptide were performed with the SignalP program, using the Neural Network (NN) and Hidden Markov Model algorithms (HMM). Version 2.0 of the SignalP program was preferred above Version 3.0 (Bendtsen *et al.*, 2004) for our signal peptide predictions in *S. aureus* and *S. epidermidis*, because the best overall prediction accuracy was obtained with Version 2.0 in a recent proteomics-based verification of predicted export and retention signals in *B. subtilis* (Tjalsma and van Dijl, 2005). Specifically, the Hidden Markov Model in SignalP 2.0 assigns probability scores to each amino acid of a potential signal peptide and indicates whether it is likely to belong to the N-, H-, or C-domains. Proteins with no detectable N-, H-, and C-domain were excluded from the set. Searching for transmembrane domains was performed with the TMHMM program and proteins with more than one (predicted) transmembrane domain were excluded from the set, because they are most likely integral membrane proteins. All proteins with a predicted C-terminal transmembrane segment in addition to a signal peptide were screened for the presence of a conserved motif for covalent cell wall binding. It should be noted that this approach does not automatically result in the exclusion of potential membrane proteins with one N-terminal transmembrane domain. The LipoP-program was used to predict lipoproteins. The combined results of all these programs resulted in a list of proteins, which have: a) signal peptides with distinctive N-, H-, and C-domains, b) no additional transmembrane domains, and c) predicted extracytoplasmic localizations. These proteins were scanned for the presence of proteomics-based consensus motifs for type I, type II or pseudopilin-specific SPase recognition and cleavage sites, twin-arginine motifs, and known leader peptides of bacteriocins by BLAST searches and by use of the PATTINPROT program (http://npsa-pbil.ibcp.fr) as previously described (Tjalsma and van Dijl, 2005). To define the core

exoproteome and variant exoproteome of the *S. aureus* strains, the sets of proteins with predicted signal peptides were used in multiple blasts with the freeware BLASTall from the NCBI. The output was then filtered using Genome2D (Baerends *et al.*, 2004).

Secretory (Sec type) signal peptides

Proteomics-based data sets of membrane, cell wall and extracellular proteins have been extremely valuable for a recent verification of signal peptide predictions in *B. subtilis* (Tjalsma and van Dijl, 2005). Such data sets are now becoming available for *S. aureus*, as exemplified by studies on the membrane *plus* cell wall proteomes of *S. aureus* Phillips (Nandakumar *et al.*, 2005), and the extracellular proteomes of *S. aureus* strains that have been derived from the recently sequenced NCTC8325 and COL strains (Ziebandt *et al.*, 2001; Ziebandt *et al.*, 2004) (Figure 2). Additionally, the extracellular proteomes of several clinical *S. aureus* isolates have been analyzed (Figure 2). The membrane, cell wall and extracellular proteins of *S. aureus* that have been identified by proteomics (Ziebandt *et al.*, 2001; Ziebandt *et al.*, 2004; Nandakumar *et al.*, 2005; Gatlin *et al.*, 2006; Pocsfalvi *et al.*, 2008; Ziebandt *et al.*, submitted), involving 2D-PAGE and subsequent mass spectrometry, are listed in (Supplemental tables IIIa and IIIb). These tables also show the -3 to +1 amino acid sequences of the respective signal peptidase cleavage sites, if present.

Based on the proteomics data for membrane and extracellular proteins of *B. subtilis*, the optimized -3 to +1 pattern [AVSTI] - [SEKYHQFLDGPW] – A - [AQVEKDFHLNS] for signal peptide recognition and cleavage by type I SPases of this organism was identified (Tjalsma and van Dijl, 2005). SPase cleavage occurs C-terminally of the invariant Ala residue at the -1 position. The residues between square brackets in the pattern are listed in the order of their frequency, the most frequently identified residue at each position being placed in first position. By comparing the predicted SPase recognition and cleavage sites in signal peptides of proteomically identified extracellular proteins of *S. aureus* (Supplemental table IIIa) we defined the -3 to +1 pattern [AVST] - [KQNESDHYLFAGR] – A - [AESKDIFLQTY] for productive recognition and cleavage by the type I SPase SpsB. Compared to the equivalent pattern of *B. subtilis* it is interesting to note that the frequencies of certain residues at the -3, -2 and +1 positions differ, as reflected by the most-frequent-first order in which they are listed in the pattern. Moreover, Asn can be present at the -2 position, while Ile is accepted at the +1 position. The latter residues are found in the -2 and +1 positions of certain serine proteases, hemolysins, immunoglobulin G binding protein A and aureolysin (Supplemental table IIIa). It should also be noted that, compared to the optimized SPase recognition pattern of *B. subtilis*, several residues are not found at the -3, -2 and +1 positions of potential SpsB recognition and cleavage sites in identified extracellular proteins of *S. aureus*. As such residues may be present in SPase recognition and cleavage sites of proteins that have escaped identification through proteomics, we have included them in the -3 to +1 search pattern (printed in lowercase) for the identification of potential secretory proteins of staphylococci: [AVSit] - [KHNDQSYEGLRAfpw] – A - [AESDIKLTYfhnqv]. This optimized *S. aureus* search pattern was used as an indicator for the quality of signal peptide predictions that were based on the SignalP-NN, SignalP-HMM, LipoP, PrediSi, Phobius and TMHMM programs. Proteins with potential signal peptides containing this pattern were assigned to have a high probability for an extracytoplasmic localization and a low probability for membrane retention (Supplemental tables IIIc-f). Proteins with potential signal peptides that do not contain this pattern were assigned to have a high probability to be retained in the membrane (data not

shown). In this case, the uncleaved signal peptide could serve as an N-terminal membrane anchor. Following this approach, sets of 186-211 proteins (depending on the *S. aureus* strain) were identified that contain a potential signal peptide or N-terminal transmembrane segment. Scanning for the presence of the SpsB recognition and cleavage motif [AVSit] - [KHNDQSYEGLRAfpw] – A - [AESDIKLTYfhnqv] revealed that, depending on the *S. aureus* strain investigated, 86-106 proteins carry this motif. These proteins are most likely processed by SPase, liberated from the membrane and secreted into the extracellular milieu, unless they contain a cell wall retention signal (see following sections). Most of the other proteins with signal peptides that do not conform to the SpsB recognition and cleavage motif lack the invariant Ala at the –1 position. Also, some of these preproteins contain different residues at the –3, -2 or +1 positions. For example, Asp, Glu, Phe and Lys are highly unlikely residues at the -3 position (van Roosmalen *et al.*, 2004). On the other hand, some preproteins have a Gly at the -3 position (*e.g.* the exotoxins 4 and 5 from *S. aureus* COL) or a Leu (Supplemental tables IIIc and IIId). Since Gly and Leu residues at the -3 position of signal peptides are accepted by the *E. coli* SPase, it seems likely that they are also accepted at this position by SpsB. However, we have not included these residues in the current SpsB recognition and cleavage motif, since we could neither identify these proteins amongst the secreted proteins of *S. aureus* COL (Figure 2), nor find published evidence that these proteins are indeed secreted. Among the proteins with predicted cleavable Sec type signal peptides there are many known extracellular staphylococcal virulence factors, such as exotoxins, enterotoxins (SEM, SEN and SEO), hemolysins, toxic shock syndrome toxin-1 (TSST-1), leukotoxins (LukD, LukE), a secretory antigen SsaA homologue and the immunodominant antigen A (IsaA). Remarkably, the lists of identified extracellular proteins of *S. aureus* COL and RN6390 (Ziebandt *et al.*, 2004) (Supplemental tables IIIa and IIIb) reveal that about 41% of these proteins lack known signal peptides. This percentage is substantially higher than the initial estimate of 10%, which was based on a limited proteomics-derived data set (Tjalsma *et al.*, 2004). It is also interesting to note that the list of identified extracellular proteins without a signal peptide includes enolase, which may be actively exported by an unknown mechanism (Boël *et al.*, 2004), but lacks EsxA and EsxB, which are exported by the Ess pathway (Burts *et al.*, 2005).

Twin-arginine (RR-)signal peptides

The consensus RR-motif that directs proteins into the Tat pathway has previously been defined as [KR]-R-x-#-#, where # is a hydrophobic residue (Cristóbal *et al.*, 1999; Jongbloed *et al.*, 2000). Dilks *et al.* (Dilks *et al.*, 2003) have used a genomic approach to identify possible Tat substrates for 84 diverse prokaryotes using the TATFIND 1.2 program. This study included *S. aureus* Mu50, MW2 and N315. Two potential Tat substrates of unknown function were predicted for *S. aureus* Mu50 and MW2 and one of these was also predicted for *S. aureus* N315 (Dilks *et al.*, 2003). However, both proteins are conserved in all sequenced *S. aureus* strains, including the N315 strain. One of these two predicted Tat substrates has no known function, whereas the other was annotated as a hypothetical protein similar to a ferrichrome ABC transporter (permease). These proteins, however, are not in our list of proteins that have a predicted (RR-)signal peptide. Although they have signal peptides according to the SignalP program, these proteins are localized in the cytoplasm or membrane, respectively, according to the LipoP, PrediSi and Phobius programs. It is therefore unlikely

that these proteins are destined for secretion. Specifically, the hypothetical permease has eight predicted transmembrane helices.

Our own pattern searches for proteins with a possible RR-motif resulted in 24-32 positive hits, depending on the *S. aureus* strain investigated. However, most of these proteins have no detectable N-, H- or C-domains and were, therefore, discarded from our data set. Also, some other proteins with a possible RR-motif are predicted to contain a lipoprotein signal peptide. These predicted lipoproteins were also discarded from the list of potential *S. aureus* Tat substrates, firstly, because none of the identified lipoproteins of *B. subtilis* that have a RR-motif were shown to be secreted via the Tat pathway (Jongbloed *et al.*, 2000; Jongbloed *et al.*, 2002), and secondly, because there is limited published evidence for other bacteria that lipoproteins can be exported Tat-dependently (Widdick *et al.*, 2006). Thus, it appears that only 5-7 proteins, depending on the *S. aureus* strain investigated, are potentially exported by the Tat pathway and cleaved by SpsB. However, it is noteworthy that none of the *B. subtilis* proteins with a KR-motif were so far shown to be secreted Tat-dependently, even though KR-motifs are capable of directing proteins into the Tat pathways of chloroplasts and Gram-negative bacteria, such as *E. coli* and *Salmonella enterica* (Stanley *et al.*, 2000; Hinsley *et al.*, 2001; Molik *et al.*, 2001; Ignatova *et al.*, 2002). If KR-motifs are also rejected by the *S. aureus* Tat pathway, there would not be a single protein in any sequenced *S. aureus* strain that is secreted Tat-dependently. This would be highly remarkable in view of the presence of *tatA* and *tatC* genes in all these strains. Notably, the only known strictly Tat-dependent extracellular proteins of *B. subtilis* are the phosphodiesterase PhoD (Tjalsma *et al.*, 2000) and the protein of unknown function YwbN (Jongbloed *et al.*, 2004). While a homologue of PhoD is not present in any of the six sequenced *S. aureus* strains, homologues of YwbN are present in all these strains. Close inspection of the YwbN homologues of *S. aureus* COL, JH1, JH9, MRSA252, MSSA476, NCTC8325, Newman, USA300 and USA300_TCH1516 revealed the presence of an N-terminal RR-motif, but a potential signal peptide was not identified as such by the SignalP program. In contrast, the YwbN homologues of *S. aureus* Mu3, Mu50, MW2 and N315 appeared to lack this RR-motif. According to comparisons of the deduced amino acid sequences, the latter three YwbN homologues would miss the first 40 residues of *B. subtilis* YwbN. Most likely, this is not the case since the sequences upstream of the annotated *S. aureus* Mu3, Mu50, MW2 and N315 *ywbN* genes encode for a peptide with a RR-motif in the same open reading frame as the *ywbN* structural gene (Supplemental table IIIc). Thus, the RR-motif of *S. aureus* Mu3, Mu30, MW2 and N315 YwbN has escaped identification due to a systematic difference in sequence annotation. Recent studies by Biswas et al. (Biswas *et al.*, 2009) have shown that the YwbN homologue of *S. aureus* is an iron-dependent peroxidase (now named FepB). *Tat* mutant *S. aureus* strains did no longer export active FepB, and the RR-signal peptide of FepB was able to direct Tat-dependent secretion of prolipase or protein A. Interestingly, FepB is needed for iron-acquisition in *S. aureus*, and in a mouse kidney abscess model the bacterial loads of *tat* or *fepB* mutant strains were substantially reduced.

Pseudopilin signal peptides

The signal peptides of pseudopilins differ from the Sec type signal peptides in the location of SPase cleavage sites. In pseudopilin signal peptides, this cleavage site is located between the N- and H-domains (Lory, 1994). The consensus recognition and cleavage motif for pseudopilin SPases, such as ComC, is K-G-F-x-x-x-E. Cleavage by pseudopilin SPases occurs within this motif, between the Gly and Phe residues. Upon cleavage the Phe residue is

methylated. In all sequenced *S. aureus* strains, three proteins were found, which have the canonical pseudopilin SPase recognition and cleavage motif. These proteins are homologues of the cold shock proteins CspB, CspC and CspD of *B. subtilis*. However, even though these proteins do contain the pseudopilin SPase recognition and cleavage pattern, they lack the H-domain. Since the active site of pseudopilin SPases is located in the cytoplasm, cleavage of the CspBCD homologues of *S. aureus* would be possible, but their export via the Com pathway is unlikely. Nevertheless, it should be noted that one of the CspBCD homologues of *S. aureus*, which is known as CspA, was found in the extracellular proteome of a clinical isolate (Figure 2 and Supplemental table IIIb). To verify the absence or presence of pseudopilins in *S. aureus*, BLAST searches with the known ComGC, ComGD, ComGE or ComGG proteins of *B. subtilis* were performed. This revealed the presence of only one potential pseudopilin, which is a homologue of *B. subtilis* ComGC. Although the consensus pseudopilin SPase recognition and cleavage site is absent from *S. aureus* ComGC, a putative cleavage pattern (Q-A-F-T-L-I-E) is present at the position in the ComGC signal peptide where a pseudopilin SPase recognition and cleavage site would be expected. Further analyses revealed that similar observations can be made for ComGC homologues in other Gram-positive bacteria, such as *Bacillus cereus*, *B. anthracis*, *L. monocytogenes*, *S. haemolyticus* and *Oceanobacillus iheyensis*. By comparing the ComGC homologues of these organisms, an expanded search pattern for Gram-positive bacterial pseudopilin SPase recognition and cleavage sites can be defined as [KEQRS]-[GA]-F-x-x-x-E. Interestingly, using this expanded search pattern, two additional potential pseudopilins of *S. aureus* were identified. These potential pseudopilins show similarity to ComGD of *B. cereus* and *B. anthracis*, and ComGF of *Bacillus halodurans* and *L. lactis*. It remains to be shown whether the three identified potential pseudopilins of *S. aureus* are indeed able to assemble into pilin-like structures after processing by the ComC homologue. If so, it will be even more interesting to identify their biological function, for example in adhesion to surfaces, motility, or export of proteins. Such functions could play a role in virulence and have been attributed to type IV pili and pseudopilins of Gram-negative bacteria (Lory, 1998).

Bacteriocin leader peptides

Bacteriocins form a distinct group of proteins with cleavable N-terminal signal peptides, which are often called leader peptides. These leader peptides only have N- and C-domains, and thus completely lack the hydrophobic H-domain. The bacteriocin leader peptides are invoked in posttranslational modification and prevention of premature antimicrobial activity, which would be deleterious to the producing organism. Of the sequenced *S. aureus* bacteriocins, C55α and C55β contain a leader peptide (Navaratna *et al.*, 1999), whereas leader peptides are absent from aureocin A53 (Netz *et al.*, 2001) and aureocin A70 (Netz *et al.*, 2001). Two potential lantibiotics with leader peptides were identified by sequencing the genomes of *S. aureus* MW2 (Baba *et al.*, 2002) (GI numbers 49486642 and 49486641) and MSSA476. In both strains, the corresponding genes are located on the genomic island νSAβ. Both *S. aureus* proteins show similarity to the lantibiotic gallidermin precursor GdmA of *Staphylococcus gallinarum*, and to the lantibiotic epidermin precursor EpiA of *S. epidermidis*. Notably, the *S. aureus* COL strain contains only one of these two potential lantibiotics, which is most similar to the potential MW2 lantibiotic with the accession number 49486641. Two additional putative bacteriocins that were identified by genome sequencing seem to be homologous to *L. lactis* lactococcin 972. The hypothetical protein SAP019 (N315 annotation)

is plasmid-encoded in *S. aureus* N315 and MSSA476, and chromosomally encoded in *S. aureus* MRSA252. The other hypothetical bacteriocin SAS029 is chromosomally encoded in all sequenced *S. aureus* strains. Recently, a program for detecting potential bacteriocins in bacterial genomes (BAGEL) has been released (de Jong *et al.*, 2006). This program is based on the properties of known bacteriocins, including genes that lie close to these bacteriocin genes. Such genes may encode proteins involved in the processing or transport of the bacteriocin. Using the BAGEL program with standard settings to detect potential bacteriocins in the sequenced and annotated *S. aureus* strains, 2-6 proteins were identified as significant. No published data is presently available on the charateristics of these proteins, so it remains to be seen whether they are genuine bacteriocins.

A potential Ess export signal?

As described above, the EsxA, EsxB and EsaC proteins are secreted by *S. aureus* via the Ess route (Burts *et al.*, 2005). All three proteins lack a known signal peptide, but are specifically transported across the membrane nonetheless. This implies that these two proteins must contain an export signal that is recognized by one or more Ess pathway components. The nature of this signal is presently unknown. The most common feature of proteins that are known (or predicted) to be translocated across the membrane via the Ess pathway is a WxG-motif, which is located at ~100 amino acids from the N-terminus of the protein (Pallen, 2002). In particular since the WxG motif appears to be absent from EsaC, the involvement of the WxG-motif in Ess targeting remains to be demonstrated.

Retention signals

Lipoproteins

Lipoproteins appear to be exported via the Sec-pathway. During, or shortly after translocation, the invariant Cys in the lipobox is diacylglyceryl-modified by Lgt. This results in signal peptide cleavage by SPase II and retention of the mature lipoprotein in the membrane. Based on the cleavage sites of lipoproteins that have been identified in various Gram-positive bacteria, Sutcliffe *et al.* (Sutcliffe and Harrington, 2002) have reported the -4 to +2 lipobox pattern [LIVMFESTAG] - [LVIAMGT] - [IVMSTAFG] - [AG] – C - [SGANERQTL]. Furthermore, they reported that neither the charged residues Asp, Glu, Arg, or Lys, nor Gln are present in the region between six and twenty residues N-terminal of the lipobox. Searching the translated proteins encoded by the thirteen *S. aureus* genomes with the pattern shown above using the PATTINPROT program, revealed about 50 proteins with this motif (Supplemental table IIIe and IIIf). A comparison of the PATTINPROT results to the results obtained with the LipoP program shows that 12-18 more potential lipoproteins may be present in *S. aureus*. Most of these extra predicted lipoproteins contain an amino acid at the -1-position (mostly Ser) or the +2-position (mostly Asp) that differs from the lipobox pattern of Sutcliffe *et al.* (Sutcliffe and Harrington, 2002). Recently, Tjalsma and van Dijl (Tjalsma and van Dijl, 2005) have proposed the lipobox search pattern [LITAGMV] - [ASGTIMVF] - [AG] – C - [SGENTAQR] for potential lipoproteins of *B. subtilis* on the basis of published proteomics data. The only difference compared to the pattern by Sutcliffe *et al.* (Sutcliffe and Harrington, 2002) is that Leu absent from the +2 position, which is due to the fact that no

potential *B. subtilis* lipoprotein with Leu at this position was identified by proteomics. Consistently, none of the predicted *S. aureus* lipoproteins has a Leu at the +2 position (Supplemental tables IIIe and IIIf). It is also noteworthy that some lipoproteins contain a [KR]-R-x-#-# motif in their signal peptides, although it has not been shown yet that lipoproteins can be transported via the Tat-pathway. Finally, the hypothetical protein Lpl2 of *S. aureus* N315 and Mu50 was excluded from our lipoprotein predictions, because Asp does not seem to occur at the +2 position of lipoproteins from Gram-positive bacteria (Juncker *et al.*, 2003; Tjalsma and van Dijl, 2005). Nevertheless, the homologues of Lpl2 of the other sequenced *S. aureus* strains are classified as lipoproteins, because they have residues at the +2 position that conform to the lipobox consensus.

Lipoprotein Release Determinant

Although lipoproteins were generally believed to be retained at the membrane-cell wall interface, the presence of lipoproteins in the growth medium of *B. subtilis* was documented by Antelmann *et al.* (Antelmann *et al.*, 2001). This unexpected finding is correlated to the proteolytic removal of the amino-terminal, lipid-modified Cys, which suggests that the observed lipoprotein release into the growth medium is caused by proteolytic "shaving" after processing by LspA. In most of these lipoproteins Tjalsma and van Dijl (Tjalsma and van Dijl, 2005) identified the +1 to +10 consensus sequence C – G - [NSTF] – x - [SGN] – x - [SGKAE] – x – x - [SGA] that might represent the recognition site for a yet unidentified shaving protease. Most probably, a Gly at the +2 position is of major importance for lipoprotein release into the growth medium, while a Ser at this position seems to inhibit this proces. In the sequenced genomes of *S. aureus* only one lipoprotein could be found with the exact motif described above. By searching for patterns with 80% similarity to the consensus sequence (*i.e.* one different residue), five to seven additional lipoproteins can be found, depending on the *S. aureus* strain. In none of these proteins, Thr was found at the +3 position. Instead, a Lys was identified at this position in one or two predicted lipoproteins with a potential release motif, depending on the *S. aureus* strain. In other predicted lipoproteins with a potential release motif, no Gly or Ala residues were found at the +7 position. However, one of these lipoproteins contains a predicted release motif with a Gln at the +7 position. To date, a total number of four potential lipoproteins have been identified in the extracellular milieu of *S. aureus*. The first one was identified by Ziebandt *et al.* (Ziebandt *et al.*, 2004). This protein has been annotated as a thioredoxin reductase (Supplemental table IIIa), but it shows no similarity to known thioredoxin reductases. Instead, it is highly similar to phosphate-binding lipoproteins, such as PstS of *B. subtilis*. It should be noted that this protein was not predicted as a lipoprotein, because the signal peptide contains a Gln residue in the N-domain. According to the search profile of Sutcliffe *et al.* (Sutcliffe and Harrington, 2002) lipoproteins would not contain Gln residues at this position. On the other hand, PstS of *B. subtilis* is a lipoprotein and it would seem quite likely that this is also true for its *S. aureus* homologue. The other three potential lipoproteins were identified in the growth medium of clinical isolates (Tabel 4). Remarkably, none of these four lipoproteins with an extracellular localization contain the complete lipoprotein release motif that was identified in extracellular lipoproteins of *B. subtilis*. However, they do contain a Gly residue at the +2 position, which strengthens the idea that this amino acid residue is probably important for lipoprotein release. It is interesting to note that in lipoproteins of Gram-negative bacteria, an Asp, Gly, Phe, or Trp residue at the +2 position prevents the transport of the mature lipoprotein to the outer

membrane (Tokuda and Matsuyama, 2004; Narita and Tokuda, 2006). In Gram-positive bacteria no outer membrane is present and it is currently not known whether the residues at the +2 position have a role in subcellular protein sorting. However, a Gly at this position does seem to promote lipoprotein release into the extracellular milieu, not only in *B. subtilis*, but also in *S. aureus*.

Cell wall binding domains

Proteins that have to be displayed on the bacterial surface must be retained by non-covalent or covalent binding to the peptidoglycan moiety of the cell wall. In *B. subtilis*, several proteins involved in cell wall turnover contain repeated domains in the C-terminal part of the protein, which have affinity for cell wall components (Ghuysen *et al.*, 1994; Margot and Karamata, 1996; Rashid *et al.*, 1995). Specifically, the *B. subtilis* proteins LytD, WapA, YocH, YvcE, and YwtD have been reported to bind to the cell wall (Ghuysen *et al.*, 1994; Margot and Karamata, 1996; Rashid *et al.*, 1995). While WapA is not conserved in staphylococci, various *S. aureus* proteins with regions that show amino acid sequence similarity to LytD, YocH, YvcE, and YwtD of *B. subtilis* can be found by BLAST searches. Accordingly, these *S. aureus* proteins may be cell wall-bound, but this remains to be shown.

One of the domains that have affinity for cell wall components is the “Lysin Motif”, or LysM domain, which has first been described for bacterial lysins (Ponting *et al.*, 1999). The number of LysM domains can differ for wall-bound proteins from different Gram-positive bacterial species (Steen, 2005). For example, XlyA of *B. subtilis* contains only one LysM domain, whereas three domains can be detected in AcmA of *L. lactis*, or even five or six domains in muramidases from *Enterococcus* species (Joris *et al.*, 1992). Using the LysM domain of AcmA from *L. lactis* in BLAST searches against the six sequenced and annotated *S. aureus* genomes, four proteins with one or more LysM domains were detected. These proteins include a hypothetical protein similar to autolysins (SA0423), the secretory antigen SsaA homologue (SA0620), a conserved hypothetical protein (SA0710), and the LytN protein.

A different domain that can facilitate protein binding to the cell wall is the GW domain. In *L. monocytogenes*, the surface-exposed InlB protein contains three C-terminal GW domains. Each domain consists of ~80 amino acids and starts with a Gly and Trp residue (Braun *et al.*, 1997). This domain specifically binds to lipoteichoic acids in the cell wall (Jonquieres *et al.*, 1999), thereby facilitating the interaction of *L. monocytogenes* with components of human host cells. The only protein found in the sequenced *S. aureus* strains with GW domains is the autolysin protein Atl (Baba and Schneewind, 1998; Baba and Schneewind, 1996). This bifunctional autolysin contains three GW repeats of ~97 amino acids. The protein is exported as a prepro-Atl precursor of 1256 amino acids. Subsequent processing steps result in the removal of the signal peptide and the propeptide, and the separation of the mature region into an amidase and a glucosaminidase (Oshida *et al.*, 1995). A similar separation of the mature region into an amidase and glucosaminidase has been reported for the AtlE protein of *S. epidermidis* (Heilmann *et al.*, 1997). The GW repeats are both necessary and sufficient to direct reporter proteins to the equatorial surface ring of *S. aureus* cells where cell division starts.

Other *S. aureus* wall proteins that contain repeated domains with potential wall binding properties have been described. These include the clumping factors A and B (ClfAB) (Hartford *et al.*, 1997; Ní Eidhin *et al.*, 1998), several serine aspartate repeat proteins (SdrCDE) (Josefsson *et al.*, 1998), the homologue of *S. gordonii* GspB (SasA) (Siboo *et al.*,

2005) (see also the section on covalent cell wall binding below), and the ECM-binding protein homologue (Clarke *et al.*, 2002). Although not documented in the literature, additional proteins with Sec type signal peptides and potential cell wall binding repeats that can be recognized readily. These are the cell wall surface anchor family protein SACOL2505, and the methicillin-resistant surface protein SACOL0050, which both contain C-terminal repeated regions of ~130 amino acids. The latter protein, which shows a high degree of sequence similarity to the SACOL2505 protein, is only found in *S. aureus* COL, but not in the five other sequenced strains. This is due to the fact that the gene for SACOL0050 is localized on the *mec* cassette 1 and therefore not present in the other strains. Notably, the SACOL2505 homologues in *S. aureus* Mu50 and N315 seem to lack the C-terminal part of the protein with the repeats. A close inspection of the sequence of the corresponding genes in these strains revealed that there is a frameshift mutation or sequencing error in these genes, resulting in an apparent or real C-terminal truncation of the corresponding proteins. Thus, the C-terminal cell wall binding repeats are absent or appear to be absent. Interestingly, most of the afore-mentioned proteins with cell wall binding motifs also contain the motif (LPxTG) for covalent attachment to the cell wall by sortase A or sortase B (see below).

It should be noted that a variety of known cell wall binding domains, such as the choline binding domain (Yother and White, 1994), the Cpl-7 cell wall binding domain (Garcia *et al.*, 1990) and the fructosyltransferase cell wall binding domain (Huard *et al.*, 2003; Milward and Jacques, 1990; Rathsam *et al.*, 1993) appear to be absent from staphylococcal proteins.

Covalent attachment to the cell wall

Cell wall sorting proteins, known as sortases, exist in many Gram-positive bacteria and serve to anchor proteins that are destined for cell surface display to the cell wall (Dramsi *et al.*, 2005; Ton-That *et al.*, 2004a). In almost all Gram-positive bacteria there is at least one sortase present and often genes for more than one sortase-like protein can be detected in a single genome. These transpeptidases catalyze the formation of an amidebond between the carboxylgroup of a Thr and the freed amino end of pentaglycine cross-bridges in peptidoglycan precursors. Subsequently, the peptidoglycan precursors with covalently bound proteins are incorporated into the cell wall. More recently, it has been shown that sortases can also be involved in protein polymerization leading to the assembly of pili on the surface of Gram-positive bacteria, such as *Corynebacterium diphtheriae* (Ton-That *et al.*, 2004b; Gaspar and Ton-That, 2006). The 3D-structure of sortase A (SrtA) of *S. aureus* revealed that this protein has a unique β-barrel structure in which a catalytic Cys residue is positioned close to a His residue. This suggests that sortase A forms a thiolate-imidazolium ion pair for catalysis (Ilangovan *et al.*, 2001; Ton-That *et al.*, 2002). Furthermore, it has been shown, that a conserved Arg residue is needed for efficient catalysis (Marraffini *et al.*, 2004). The catalytic cysteine is part of a conserved motif, TLxTC, which can be found in the C-terminal part of the protein (x is usually Val, Thr or Ile). Recently, a classification of sortases has been proposed by Dramsi *et al.* (Dramsi *et al.*, 2005) based on phylogenetic analyses of 61 sortases that are encoded by the genomes of 22 Gram-positive bacteria. These analyses showed that sortases can be grouped into four different classes (A-D). Class A consists of sortases from many low GC% Gram-positive bacteria, including *L. monocytogenes*, *Streptococcus pyogenes* and *S. aureus*. The second class (Class B) is present in only a few low GC% Gram-positive bacteria, including, *L. monocytogenes*, *B. anthracis* and *S. aureus*. Sortases of this class contain three amino acid segments that are not present in the sortases of class A. These

sortases recognize a different motif (NPQTN in *S. aureus*). The genes for substrates of class B sortases are often found at the same locus as the sortase gene. The largest class (Class C) consists of sortases from high GC% and low GC% Gram-positive bacteria. The genes for class C sortases are often present in multiple copies per genome. Characteristic for this class of sortases is a C-terminal hydrophobic domain that might serve as a membrane anchor, and a conserved proline residue behind the catalytic site. Finally, class D sortases are present in high and low GC% Gram-positive bacteria. This class can be divided into three subclusters, depending on whether they are present in bacilli, clostridia or actinomycetales. Since class C and D sortases are absent from *S. aureus*, the (potential) substrates of these enzymes will not be reviewed here.

Sortase A recognition signal

For interaction with host cells during infection, many proteins are anchored to the cell wall of staphylococcal cells, thereby enabling the cells to adhere and invade the host cells, or to evade the immune system. Many of these proteins contain an LPxTG-motif in their C-terminal part, which is recognized by the cell wall sorting protein sortase A. In each of the six sequenced *S. aureus* strains there is only one sortase gene present, which encodes a class A sortase. The LPxTG motif of sortase A substrates is followed by a stretch of hydrophobic amino acids and at least one positively charged amino acid (Lys or Arg) at the C-terminus. After translocation across the membrane, the LPxTG motif is recognized by SrtA and subsequently cleaved between the Thr and Gly residues (Mazmanian *et al.*, 2001; Navarre and Schneewind, 1994). A transpeptidation is then mediated by SrtA through amide-linkage of the C-terminal Thr of the protein to pentaglycine cross-bridges. It has been suggested that SrtA actually uses lipid II as a peptidoglycan substrate and that the proteins that are linked to lipid II are subsequently incorporated into the cell wall. In addition to the canonical LPxTG motif, a LPxAG motif can also be recognized and cleaved by SrtA (Roche *et al.*, 2003). It has been reported that *S. aureus* has 19 proteins that carry the LPxTG motif and 2 proteins that carry a LPxAG motif at their C-termini (Roche *et al.*, 2003) (Table 4). Many of these proteins have been shown to be expressed. These include: protein A (Spa), two clumping factors (ClfA and ClfB; also contain potential wall binding repeats), a collagen-binding protein (Cna), three serine aspartate repeat proteins (SdrC, SdrD and SdrE; also contain potential wall binding repeats), two fibronectin-binding protein (FnbpA and FnbpB) (reviewed by Foster and Hook, (Foster and Hook, 1998)), a plasmin-sensitive protein (Pls) (Savolainen *et al.*, 2001), FmtB (Komatsuzawa *et al.*, 2000), and several *S. aureus* surface (Sas) proteins (Roche *et al.*, 2003). A recent study on the cell wall and membrane proteome by Nandakumar *et al.* (Nandakumar *et al.*, 2005) resulted in the identification of two proteins with an LPxTG cell wall sorting signal. Many of the LPxTG proteins contain in their N-terminal signal peptide a conserved motif, [YF]-SIRK (with some variance) that has also been observed in other proteins that are substrates for SrtA in several Gram-positive bacteria (Bae and Schneewind, 2003). However, this sequence is not found in all of the SrtA substrates and it can also be found in non-cell wall proteins. This suggests that [YF]-SIRK is not a specific SrtA targeting sequence. Interestingly, proteins that contain an LPxTG-motif and a [YF]-SIRK motif in their signal peptides seem to be distributed along the staphyolococcal cell surface in a different manner than those proteins that have an LPxTG-motif, but lack the [YF]-SIRK motif. Those proteins that do contain the [YF]-SIRK motif seems to be positioned in a ring-like structure that forms during cell division for the separation of cells, while the proteins that lack this motif are directed to the cell pole (Dedent *et al.*, 2008). One of the *sas* genes, *sasA* also known as *sraP*, is situated in

the *secA2/secY2* cluster and has an unusually long signal peptide (90 residues). Similar to what has been reported for the cell wall-bound GspB protein in *S. gordonii* (Bensing and Sullam, 2002), it was recently shown that the accessory SecA2/SecY2 system is needed for the transport of SasA/SraP across the membrane (Siboo *et al.*, 2008). Depending on the sequenced *S. aureus* strain, 10-13 proteins with an LPxTG cell wall sorting signal, followed by a hydrophobic stretch of residues and a positively charged C-terminus can be found (Table 4). Among these proteins are the fibrinogen-binding protein A (ClfA), immunoglobulin G binding protein A precursor (Spa), and the Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein (SdrC).

Table 4. Staphylococcal proteins with (potential) Sec type signal peptides and (potential) signals for covalent cell wall binding

PID	Protein	Function	Signal
15925728	AsdA	Adenosine synthase A	LPKTG
15925815	Spa	Immunoglobulin G binding protein A precursor	LPETG
15925838[a]	SasD	hypothetical protein	LPAAG
15926239[b]	SdrC	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	LPETG
15926240[c,d]	SdrD	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	LPETG
15926241[b,d,e]	SdrE	Ser-Asp rich fibrinogen-binding, bone sialoprotein-binding protein	LPETG
15926464	ClfA	fibrinogen-binding protein A, clumping factor	LPDTG
15926713	IsdB	conserved hypothetical protein	LPKTG
15926714	IsdA	cell surface protein	LPKTG
15926715	IsdC	conserved hypothetical protein	NPQTN
15927308[c]	HarA	hypothetical protein	LPKTG
15927333[d,f]	SasC	hypothetical protein, similar to FmtB protein	LPNTG
15927741[b,g]	SasB	FmtB protein	LPDTG
15928076[b,c,h]	Aap	hypothetical protein, similar to accumulation-associated protein	LPKTG
15928081[c,d]	FnbB	fibronectin-binding protein homolog	LPETG
15928082[d,i]	FnbA	fibronectin-binding protein homolog	LPETG
15928174[j]	SasK	hypothetical protein	LPKTG
15928216	ClfB	Clumping factor B	LPETG
15928232[b]	SasF	conserved hypothetical protein	LPKAG
15928240[k]	SasA	hypothetical protein, similar to streptococcal hemagglutinin protein	LPDTG
57652419[l]	Pls	methicillin-resistant surface protein	LPDTG
21284341[m]	Cna	collagen adhesin precursor	LPKTG
27467746[n]	SE0828	Lipoprotein VsaC	LPETG
27468418[n]	SE1500	hypothetical protein	LPKTG
27468419[n]	SE1501	hypothetical protein	LPNTG
27468546[n]	SE1628	hypothetical protein	LPETG
27469070[n]	SE2152	hypothetical protein	LPNTG

[a] *Truncated in* S. aureus *NCTC 8325, thereby missing the C-terminal part containing the LPxTG motif*
[b] *Proteins that are also present in* S. epidermidis
[c] *The genes encoding SdrC, HarA and FnbB are not present in* S. aureus *MRSA252*
[d] *These proteins have a lower SignalP score than our threshold score*
[e] *The gene encoding SdrE is not present in* S. aureus *NCTC8325*
[f] *Truncated in* S. aureus *Mu50, thereby missing the C-terminal part containing the LPxTG motif*
[g] *The gene encoding SasB is not present in* S. aureus *MRSA252, MSSA476 and USA300_TCHC1516; truncated in* S. aureus *MW2*
[h] *Truncated in* S. aureus *Mu50, N315 and Newman, thereby missing the C-terminal part containing the LPxTG motif*
[i] *Truncated in* S. aureus *Newman, thereby missing the C-terminal part containing the LPxTG motif (Grundmeier et al., 2004)*
[j] *The gene encoding SasK is not present in* S. aureus *COL, MRSA252, MSSA476, NCTC 8325, Newman, USA300 and USA300_TCHC1516; truncated in* S. aureus *MW2, thereby missing the N-terminal signal peptide*
[k] *This protein has an unusually long signal peptide (90 amino acids)*
[l] *The gene encoding Pls is only present in* S. aureus *COL*
[m] *The gene encoding Cna is not present in* S. aureus *COL, JH1, JH9, Mu50, N315, NCTC 8325, Newman, USA300 and USA300_TCHC1516*
[n] *Only present in* S. epidermidis

Five additional proteins (SdrD, SdrE, SasC, FnbA and FnbB) with an LPxTG motif can be found among the *S. aureus* strains (Table 4). These five proteins were excluded from our initial list, because the corresponding SignalP scores were lower than our (high) score criteria, or, as was shown for the FnbA and FnbB proteins of the S. aureus Newman strain, the LPxTG motif is missing due to a premature stop codon (Grundmeier *et al.*, 2004). However, since some of the domains present in these proteins (besides the LPxTG motif) are conserved in well-described cell wall proteins, they have been included in Table 4. The remaining proteins with a cell wall sorting signal are either missing in one or more *S. aureus* strains, or have been annotated wrongly. Interestingly, *S. epidermidis* ATCC 12228 contains a gene for a class C sortase (*srtC*), which seems to be absent from other staphylococci. This SrtC is most closely related to sortases of *L. lactis* and *Streptococcus suis*. Two proteins with LPxTG motifs, which are encoded by the same genomic island as SrtC, also seem to be strain-specific (Gill *et al.*, 2005).

Sortase B recognition signal

All sequenced *S. aureus* strains contain sortase B (SrtB) in addition to SrtA. The gene for SrtB is situated at a locus, which is involved in the uptake of haeme iron (Mazmanian *et al.*, 2002). This locus also contains the gene for the cell wall protein IsdC, which contains the SrtB recognition sequence NPQTN. In addition, this locus contains the genes for the SrtA substrates IsdA and IsdB that both contain LPxTG motifs. Notably, IsdC is so far the only known protein known to be anchored to the cell wall by sortase B. IsdC is cleaved by SrtB between the Thr and Asn residues of the NPQTN motif. The only other *S. aureus* protein with a motif that resembles NPQTN is the DNA-binding protein II, but this protein is probably not cell wall bound, because it lacks a signal peptide for export from the cytoplasm.

Comparative secretome analysis

Comparing the predicted secretomes of *S. aureus* and *S. epidermidis* with those of *B. subtilis* and other Gram-positive bacteria revealed that most of the known components of the translocation machinery are present in *S. aureus*. The most notable differences are the second set of *secA* and *secY* genes in *S. aureus*, the absence of known signal peptide peptidases from *S. aureus* and *S. epidermidis*, the absence of a BdbC homologue from *S. aureus* and *S. epidermidis*, the presence of a second *lspA* gene in *S. epidermidis*, the absence of a Tat system from *S. epidermidis*, and the absence of two potential components in the Ess pathway from *S. aureus* MRSA252 (Table 3).. However, it has been shown that the second *secA*/*secY* set is involved in the export of virulence factors in other pathogens (Takamatsu *et al.*, 2005). Though most known determinants for protein export, processing and post-translocational modification in other Gram-positive bacteria are also present in *S. aureus*, in many cases it remains to be investigated to what extent they are necessary for protein export in general and the export of virulence factors in particular.

As shown by multiple BLAST comparisons, the core exoproteome of the sequenced *S. aureus* strains consists of 68 proteins (Supplemental table IIIc). All of these proteins have a signal peptide with a potential SpsB recognition and cleavage site. 46 of these core exoproteins have already been identified in the extracellular milieu and/or membrane/cell wall proteome of different *S. aureus* isolates (Ziebandt *et al.*, 2001; Ziebandt *et al.*, 2004; Nandakumar *et al.*, 2005; Gatlin *et al.*, 2006; Pocsfalvi *et al.*, 2008; Ziebandt *et al.*, submitted). Interestingly, 31 core exoproteins of *S. aureus* are also conserved in *S. epidermidis*, suggesting that they

belong to a core staphylococcal exoproteome, which is presently still poorly defined. Interestingly, the core exoproteome of *S. aureus* seems to be largely composed of enzymes, like proteases, and other factors, like fibrinogen- and IgG-binding proteins, that are required for life in the ecological niches provided by the human host (Supplemental table IIIc). This is particularly true also for the proteins that have the potential to be covalently bound to the cell wall (Table 4). In contrast, the variant exoproteome of *S. aureus* contains most of the known staphylococcal toxins and immunomodulating factors (Supplemental tables IIId and IIIg). This suggests that the components of the variant exoproteome should be regarded as specific gadgets of *S. aureus* that help this organism to conquer certain protected niches of the human host, thereby causing disease. If this idea is correct, proteins of unknown function that belong to the variant exoproteome should be regarded as potentially important virulence factors.

The (predicted) extracellular toxins of *S. aureus* are not present in *S. epidermidis*. This is mainly due to the fact that these toxins are encoded by pathogenicity islands in the genomes of *S. aureus* strains that have, so far, not been observed in *S. epidermidis* genomes. Proteins with predicted signal peptides that are specific for *S. epidermidis* are listed in Supplemental table IIIi. Notably, the majority (*i.e.* 31 out of 37) of predicted *S. epidermidis* exoproteins that have homologues in *S. aureus* share this homology with components of the core exoproteome of *S. aureus* (Supplemental tables IIIc and IIId). This suggests that also in *S. epidermidis* the core exoproteome is involved in housekeeping functions. In contrast to the exoproteome, it is presently difficult to speculate about housekeeping- and disease-causing roles of the constant and variant lipoproteomes of *S. aureus*. This is due to the fact that the function of only few *S. aureus* lipoproteins is known (Supplemental tables IIIe, IIIf and IIIh). In general terms, it is presently not clear why *S. epidermidis* seems to export a lower number of different proteins (94 in total) than *S. aureus* (~165 in total). This difference is all the more remarkable since the total number of proteins encoded by the genomes of *S. aureus* (~2600) and *S. epidermidis* (~2500) are comparable.

Compared to *B. subtilis* and *B. licheniformis* (Voigt *et al.*, 2005), *S. aureus* is also predicted to export a relatively higher number of proteins from the cytoplasm to the membrane-cell wall interface, the cell wall and the extracellular milieu. The genomes of *B. subtilis* and *B. licheniformis* contain ~4100 protein-encoding genes, while the *S. aureus* genomes contain significantly less genes (~2600). Using the most recent prediction protocols (Tjalsma and van Dijl, 2005), *B. subtilis* is predicted to export 190 proteins to an extracytoplasmic location whereas, depending on the strain investigated, *S. aureus* is predicted to export 145-168 proteins (this review). Accordingly, as judged by the relative numbers of protein-encoding genes, *S. aureus* strains appear to export 1-2% more proteins to an extracytoplasmic location than the afore-mentioned bacilli. Most probably, this is related to the fact that *S. aureus* needs an arsenal of virulence factors, such as toxins and surface proteins, for colonization of and survival in its preferred niches in the human host. Such proteins are of less importance for soil bacteria, such as *B. subtilis* and *B. licheniformis*, which thrive predominantly on dead organic matter.

Perspectives

The present review provides a survey of possible protein transport pathways to staphylococcal pathogenesis. In many cases the knowledge gathered from protein secretion studies in other organisms has been projected on *S. aureus*, assuming that similar pathways or pathway components have similar functions in different organisms. Clearly, this leaves room for

surprises when such pathways are investigated thoroughly in *S. aureus*. The same was true for studies on protein secretion in *B. subtilis*. These studies showed, for example, that the absence of SecDF has barely any consequences for protein secretion by *B. subtilis*, whereas SecD and SecF are of key importance for protein translocation in *E. coli*, the organism in which SecD/F was first discovered (Bolhuis *et al.*, 1998). Likewise, LspA was shown to be dispensable in *B. subtilis*, but not in *E. coli* (Tjalsma *et al.*, 1999; Prágai *et al.*, 1997). Thus the relative importance of different secretion machinery components of *S. aureus* needs to be assessed in a systematic manner, preferably in an isogenic background. Such studies would need to address the importance of secretion machinery components for *in vitro* growth on different substrates (*e.g.* broth or blood), and virulence in *in vivo* model systems (*e.g. C. elegans*, *Drosophila melanogaster*, mice and rats) (Bae *et al.*, 2004; García-Lara *et al.*, 2005; Tarkowski *et al.*, 2001). These studies should be complemented with a proteomic verification of our present lipoproteome, wall proteome and exoproteome predictions. Such a verification could involve both gel-based proteomics approaches as outlined in this review, and more sophisticated gel-free proteomics approaches (Völker and Hecker, 2005). This would lead to an improved understanding of the contribution of each protein transport pathway and its substrate proteins to staphylococcal cell physiology and virulence. Since the virulence of different *S. aureus* strains will not only depend on the presence (or absence) of particular genes for virulence factors, but also on their expression, such proteomic studies should also include experiments on the impact of major regulatory systems, such as *agr*, *sae* and *sarA*. On this basis, it should be possible to identify the most promising candidate drug targets in the staphylococcal secretome. Alternatively, secretome components thus identified could represent promising candidates for novel vaccines. For all these efforts, comparative secretomics approaches will provide the essential information on those potential drug targets that are most important for both bacterial housekeeping functions and virulence. Novel drugs and vaccines designed against these targets are likely to have the highest impact (Götz, 2004).

Acknowledgements

We thank Harold Tjalsma and members of the Groningen and European *Bacillus* Secretion Groups and the BACELL Health, Tat machine and StaphDynamics consortia for stimulating discussions. M.J.J.B.S., A.K.Z., S.E., M.H., J.-Y.F.D., and J.M.v.D. were supported by Grants LSHC-CT-2004-503468, LSHG-CT-2004-005257 and LSHM-CT-2006-019064 from the European Union. M.H. was supported from grants of the “Deutsche Forschungsgemeinschaft”, the “Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie”, and the “Fonds der Chemischen Industrie”.

"Music is the medicine of the mind"

-John A. Logan-