Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold ‘em, knowing when to fold ‘em

Matthew I. Hutchings1, Tracy Palmer2, Dean J. Harrington3 and Iain C. Sutcliffe4

Gram-positive bacterial lipoproteins are a functionally diverse and important class of peripheral membrane proteins. Recent advances in molecular biology and the availability of whole genome sequence data have overturned many long-held assumptions about the export and processing of these proteins, most notably the recent discovery that not all lipoproteins are exported as unfolded substrates through the general secretion pathway. Here, we review recent discoveries concerning the export and processing of these proteins, their role in virulence in Gram-positive bacteria and their potential as vaccine candidates or targets for new antimicrobials.

Bacterial lipoproteins
Lipoproteins in Gram-positive bacteria are cell envelope proteins anchored into the outer leaflet of the plasma membrane. Lipid modification is achieved through covalent addition of a diacylglyceride to an indispensable cysteine residue in the lipoprotein signal peptide, as originally described for the prototypical Braun’s lipoprotein of Escherichia coli [1]. This provides a common anchoring mechanism for what is now recognized to be an abundant and functionally diverse class of peripheral membrane proteins. In Gram-positive bacteria, lipoproteins function within a subcellular region that is defined at its inner aspect by the plasma membrane and at its outer aspect by the peptidoglycan and other layers of the cell wall. Lipoproteins of Gram-positive bacteria have, thus, been proposed to be functional equivalents of periplasmic proteins in Gram-negative bacteria, a comparison that is most directly sustained by the fact that, in Gram-positive bacteria, substrate binding proteins (SBPs) of ATP-binding cassette (ABC) transporters are typically lipoproteins [2,3]. Moreover, cell fractionation experiments and recent advances in electron microscopy have lent some credibility to the controversial concept of a Gram-positive ‘periplasm’ [4–6]. Cryoelectron microscopy has also provided evidence supporting the presence of an outer membrane permeability barrier in the mycolic acid-containing actinomycete bacteria Mycobacterium smegmatis, Mycobacterium bovis and Corynebacterium glutamicum [7,8]. In addition, since the last review of Gram-positive lipoproteins 13 years ago [3] our understanding of the diverse functions of these proteins has been greatly advanced by the availability of whole genome sequence data. These advances, along with important new insights into the lipoprotein biogenesis pathway, make it timely to revisit this subject.

Crossing the cytoplasmic membrane
Almost all exported proteins are transported across the cytoplasmic membrane of prokaryotes by one of two distinct export pathways. The general secretory (Sec) pathway is the predominant route of protein transport [9]. The Sec machinery recognizes proteins bearing N-terminal signal peptides (Figure 1a) and transports them across the membrane in an unfolded conformation. By contrast, the more recently discovered Tat (twin arginine protein transport) system transports folded and even oligomeric proteins, which often bind redox cofactors [10]. Proteins are also targeted to the Tat system by means of N-terminal signal peptides, which in this case harbour an almost invariant and essential twin-arginine motif (Figure 1b).

Exported proteins that are destined to become lipidated contain a motif in their signal peptides known as a lipobox, which directs them to the lipoprotein biogenesis machinery after transport (Figure 1). It had long been assumed, based primarily on studies in E. coli, that all lipoprotein precursors are synthesized with signal peptides that direct them to the Sec pathway for translocation across the cytoplasmic membrane in an unfolded state [1,11]. More recently, it has also become clear that some putative lipoproteins can be translocated utilizing the SecA2-dependent accessory Sec pathway, which is found in some, but not all, Gram-positive bacteria [12,13]. The first indication that some lipoprotein precursors could be exported in a fully folded state through Tat came during an analysis of the dimethylsulphoxide (Dms) reductase in the Gram-negative bacterium Shewenella oneidensis. DmsA was shown to contain a Tat signal sequence with a lipobox and to be translocated via Tat in a complex with its partner subunit DmsB [14]. The DmsB subunit lacks a signal sequence and the proteins must, therefore, fold and form a complex before export. DmsA is a relatively rare example of an outer...
membrane lipoprotein which faces the extracellular environment and not the periplasm [14]. More recently, it was reported that the HysA subunit of the (NiFeSe) hydrogenase of the Gram-negative bacterium *Desulfovibrio vulgaris* is exported through Tat, despite lacking a signal sequence. In fact, HysA ‘piggybacks’ out through Tat by virtue of binding to its partner protein, the (NiFeSe) hydrogenase subunit, HysB, which harbours a canonical twin arginine signal peptide [15]. After export, HysA is lipidated and retained in the cytoplasmic membrane, facing into the periplasm, in complex with HysB [15]. HysA is remarkable because it does not contain a signal peptide but can still be targeted to the lipoprotein machinery by its first four amino acids (Met-Ser-Gly-Cys), which constitute a lipobox (Figure 1). After lipidation of the cysteine residue, the first three amino acids are cleaved by lipoprotein signal peptidase (Lsp) [15].

In the high Guanine+Cytosine branch of Gram-positive bacteria known as actinomycetes, Tat is also apparently required for the translocation of lipoproteins, including the BlaC β-lactamase putative lipoprotein of *Mycobacterium tuberculosis* when it is expressed in *M. smegmatis* [16] and four putative lipoproteins in the model actinomycete, *Streptomyces coelicolor* [17]. Bioinformatic analysis indicates that Tat translocation of lipoproteins is widespread in the genus *Streptomyces* with up to 20% of putative lipoproteins being exported via Tat in the four currently sequenced *Streptomyces* species (M.I.H and I.C.S, unpublished). In comparison, ~10%–15% of the putative lipoproteins of *M. tuberculosis* [18] are predicted to be Tat substrates, whereas only two of the 41 putative lipoproteins in the actinomycete *Leifsonia xyli* [19] are predicted to be Tat substrates (I.C. S, unpublished). Interestingly, no reports exist of Tat-dependent lipoproteins in the low Guanine+Cytosine (Firmicute) branch of Gram-positive bacteria and, in fact, the model organism *Bacillus subtilis* exports very few of its proteins via Tat. Moreover, some Firmicute genomes (notably those of several streptococci) apparently lack a Tat pathway [20]. Finally, in the archaeon *Haloferax volcanii*, which exports most of its proteins through Tat, it has been demonstrated that some of these Tat exported proteins are lipid modified [21]. Archaeal lipoprotein signal sequences have typical lipoboxes but their genomes do not encode homologues of the bacterial lipoprotein processing enzymes [21]. These recent publications have overturned the assumption that only linear polypeptides can be lipid modified after export from the cell by the Sec pathway. Consequently, it seems likely that Tat export of lipoprotein precursors will be widespread in some bacterial and archaeal lineages and this must lead to a paradigm shift in our understanding of the export and processing of these cell envelope proteins.

**Lipoprotein biogenesis**

The pathway for bacterial lipoprotein biogenesis was established in *E. coli* by the pioneering work of Wu and co-workers [1]. After export through Sec (or Tat), a conserved motif (the ‘lipobox’) in their Type II signal peptides contains the recognition motif for type I (A-X-A, where X is any amino acid) or type II (L₀[A/S/T]₁[G/A]₋₁[Cₓ₁] signal peptides. The type II cleavage site is referred to as the lipoprotein ‘lipobox’. Tat signal peptides have variable length N-regions and a conserved SRRXFLK sequence between the N- and H-regions [88] in which the twin arginine (RR) motif is almost absolutely conserved and gives the transport pathway its name.

![Figure 1](image_url)

**Figure 1.** Type I and type II signal peptides for Sec- and Tat-dependent transport. Both Sec (a) and Tat (b) signal peptides are tripartite in structure with a positively charged N- (N-terminal) region, an H- (hydrophobic) region and a C- (cleavage) region, which contains the recognition motif for type I (A-X-A, where X is any amino acid) or type II (L₀[A/S/T]₁[G/A]₋₁[Cₓ₁]) signal peptides. The type II cleavage site is referred to as the lipoprotein ‘lipobox’. Tat signal peptides have variable length N-regions and a conserved SRRXFLK sequence between the N- and H-regions [88] in which the twin arginine (RR) motif is almost absolutely conserved and gives the transport pathway its name.
N-terminus of the mature lipoprotein. (iii) Lipoprotein N-acyl transferase (Lnt) adds a third fatty acid in an amide linkage to the free amino group of the lipidated cysteine. This step is essential for the release of lipoproteins from the plasma membrane and their transport via the Lol (lipoprotein localisation) pathway to the outer membrane (Figure 2a). Depletion of Lnt results in an accumulation of lipoproteins in the plasma membrane, which is lethal to the cell [22]. Retention of lipoproteins in the plasma membrane of Gram-negative bacteria ('Lol avoidance') is signalled by aspartate at position +2 in E. coli or lysine and serine at positions +3 and +4 in Pseudomonas [23–25]. However, these sorting signals do not interfere with N-acylation because plasma membrane retained lipoproteins are also N-acylated [26]. In Borrelia burgdorferi there is no Lol avoidance signal and lipoproteins are trafficked to the outer membrane by default [27]. The first two steps of this pathway are well conserved in prokaryotes, and lipoprotein biosynthesis in Gram-positive bacteria follows essentially the same path (Figure 2b). However, a recent study reported that Lsp can cleave unlipidated substrates in Listeria monocytogenes, indicating that the pathway does not always occur in strict sequence [28]. Likewise, cleavage of the ScaA lipoprotein precursor at the Lsp cleavage site in a Streptococcus agalactiae lgt mutant [29] also indicates that Lsp could have activity towards non-lipidated precursors, at least in some Gram-positive bacteria.

Lgt is an essential enzyme in Gram-negative bacteria, possibly because of the essential nature of murein lipopeptide and other outer membrane lipopolysaccharides such as YRO [30], but is dispensable for growth in vitro of all Gram-positive bacteria tested to date (see later). Intriguingly, although lgt is present as a single gene in most bacterial genomes, there are two putative lgt paralogues encoded in the genomes of a limited selection of Gram-negative (e.g. Coxiella burnetii) and Gram-positive bacteria (e.g. Bacillus cereus ATCC10987, Clostridium perfringens and S. coelicolor). A full list is accessible via the Pfam database entry for Lgt at http://pfam.sanger.ac.uk/family?acc=PF01790. The roles of these lgt paralogues remains unclear, although paralogous Lgt enzymes might be dedicated to the processing of specific lipoproteins in a manner analogous to the processing of specific wall-anchored proteins by substrate-specific sortase enzymes [31]. For example, the second lgt of B. cereus, ATCC 10987 (BCE_A0191), is notable for being encoded on the pBc10987 plasmid and is part of a locus including several putative lipoproteins (BCE_A0184 Lmb, BCE_A0186 ArsR and BCE_A0188).

After lipidation by Lgt, the signal peptide is cleaved from the prolipoproteins by Lsp leaving the lipobox cysteine at position +1. As with Lgt, Lsp is essential in Gram-negative bacteria and dispensable for growth of Gram-positive bacteria in vitro (see later). Likewise, most organisms seem to only possess a single lsp, although in some cases (e.g. L. monocytogenes, Staphylococcus epidermidis, Nocardia farcinica) a second Lsp paralogue seems to be present. Some separate functionality is suggested for these Lsp paralogues: in the case of N. farcinica, the Lsp paralogue is plasmid encoded, whereas in L. monocytogenes inactivation of LspA (lmo1101) did not seem to be compensated for by the presence of the lmo1110 paralogue [32].

One reason underlying the viability of mutants defective in lipoprotein biosynthetic enzymes in Gram-positive bacteria could be that some lipoprotein precursors retain functionality: the PrsA lipoprotein is an essential protein in B. subtilis and yet both lgt and lsp mutants of this organism are viable [33,34]. Likewise, the lipoproteins PrtM and OppA of Lactococcus lactis are needed for growth of this organism in milk and yet an L. lactis lsp mutant can grow in milk [35].

Studies with lgt and lsp mutants have also revealed alternative pathways for lipoprotein processing. Mutation of lgt or lsp would be predicted to result in a build-up of

---

**Figure 2.** The lipoprotein biogenesis pathway. In Gram-negative bacteria (a) unfolded (red straight line) or folded (red filled circle) lipoproteins are directed to and translocated across the cytoplasmic membrane by the Sec or Tat pathways by their signal sequences (shown in blue) (i). A lipid group (angled black line) is covalently attached to the sulphydryl group of the lipobox cysteine by Lgt (prolipoprotein diacylglycerol transferase) (ii), and the signal peptide is cleaved by Lsp (lipoprotein or type II signal peptidase) (iii). A second lipid group (black line) is then attached to the amino group of the lipoprotein by Lnt (lipoprotein N-acyl transferase) (iv). This occurs in strict order and they are either retained in the cytoplasmic membrane (v) or transported to the outer membrane by the Lol (lipoprotein localisation) pathway (vi). In Gram-positive bacteria (b) the pathway is conserved but does not necessarily occur in strict order. There is some evidence of N-acylation in the low-GC Gram-positive bacteria B. subtilis and S. aureus [39,40] despite an absence of Lnt homologues. Conversely, some high-GC Gram-positive bacteria encode Lnt homologues but there is, as yet, no evidence of lipoprotein N-acylation.
precursor lipoproteins with either uncleaved but non-lipid modified signal peptides (lgt mutants) or uncleaved, lipid modified signal peptides (lsp mutants). Although these phenotypes can be observed, in many cases the typical effects of lgt or lsp mutation seem to be lipoprotein mislocalisation and/or aberrant processing (Table S1 in the supplementary material). Often subsets of lipoproteins have been observed to be processed to yield ‘mature-like’ forms that can remain cell-associated or are released to the culture supernatants (Table S1 in the supplementary material). Indeed, several studies have now used immunoblotting to reveal different effects on the processing of specific lipoproteins in the same mutant background. The identification of the ‘mature-like’ lipoprotein forms indicates that the build-up of lipoprotein precursors in the membranes of lgt and lsp mutant strains could result in alternative processing by other peptidases such as the recently recognized Eep peptidase [36], Lsp (in Lgt mutant backgrounds) [28,29] or type 1 signal peptidases. The release of ‘mature-like’ forms from some lipoprotein precursors in lgt or lsp deletion mutants has been termed ‘shaving’, whereas the release of either lipoprotein precursors or mature, lipidated lipoproteins can be considered as ‘shedding’ [37]. Shaving most likely reflects protein-specific proteolytic cleavage events because N-terminal sequencing of released lipoprotein products, from both mutant and wild-type backgrounds, shows differing cleavage positions with respect to the N-terminal cysteine [2,29,37,38].

Finally, it should be emphasized that the accurate and efficient processing of lipoproteins in wild-type strains is likely to depend upon a close interaction between Lgt, Lsp and the protein translocation machinery. In this respect, it is relevant to note that lipoprotein processing in E. coli requires both the Sec translocase and the YidC membrane insertase [11]. How the lipoprotein biosynthesis pathway operates as a terminal branch of the protein translocation path is an important area for future study (Box 1), particularly because it has now been shown that lipoproteins can translocate through both Sec and Tat. This raises the possibility that Lgt either interacts closely with both Sec and Tat or, alternatively, that lipoprotein signal peptides transiently anchor preprolipoproteins (after their release from either translocase) until they can interact with the membrane-located lipoprotein biosynthetic enzymes.

N-acylation in Gram-positive bacteria

Chemical analyses of Braun’s lipoprotein expressed in B. subtilis and lipoprotein preparations from Staphylococcus aureus are consistent with at least some N-acylation of lipoproteins in these organisms [39,40]. However, BLAST searches for Lnt homologues in Firmicute genomes have failed to yield convincing candidates. Lnt homologues are present in the sequenced genomes of actinomycetes but the functions of these enzymes are unclear. Although the S. coelicolor Int (SCO1336) gene failed to rescue an E. coli lnt deletion strain [41] there are several possible explanations: the expressed protein could be inactive; it could be unable to recognize and modify E. coli substrates or the gene product could have an unrelated function. In fact, there are two Lnt homologues in S. coelicolor, SCO1014 and SCO1336. SCO1014 also shows homology to a domain associated with mycobacterial polyprenol monophosphomannose (PPM) synthases [42]. Notably M. tuberculosis Ppm is a two-domain protein (Rv2051c) in which the N-terminus is similar to Lnt and the C-terminus has PPM synthase activity [42]. PPM is an alkali-stable sugar donor used in the formation of the cell envelope glycolipids lipidmannan (LM) and lipoarabinomannan (LAM) in M. tuberculosis [42]. The function of these mycobacterial Lnt homologues might instead relate to their belonging to the CN hydrolase enzyme superfamily rather than their being directly orthologous to E. coli Lnt.

Role of Lgt and Lsp in virulence of Gram-positive pathogens

The Lgt and Lsp enzymes of the lipoprotein biosynthetic pathway are apparently unique to prokaryotes and are potentially attractive candidates for the development of novel antibacterials because perturbation of this pathway should affect numerous, functionally unrelated lipoproteins. Unsurprisingly, given the variety of functions attributable to lipoproteins, lgt or lsp mutations have pleiotropic effects. For example, lgt mutants of B. subtilis exhibit defects in cytochrome caa₃ activity [43], protein secretion [33], germination and sporulation [44-46] that can be correlated to impaired functions of specific lipoproteins.

The apparent indispensability of Lgt and Lsp in Gram-negative bacteria has precluded the study of virulence of lipoprotein-processing mutants of Gram-negative pathogens. However, several studies have now addressed the virulence of Gram-positive pathogens (Table 1). In many, but not all cases, loss of Lgt or Lsp has led to attenuation of immune activation or virulence either in vitro or in vivo (animal models). Clearly, interpretation of the data on the attenuation or virulence of Lgt or Lsp mutants needs to take into account species and strain variation and also the validity of the animal model(s) employed. However, two surprising observations can be made. First, the only studies that have been carried out on virulence in the natural host (i.e. Streptococcus equi in Welsh mountain ponies [47] and Streptococcus suis in pigs [48]) have failed to show attenuation of lgt or lsp mutants, respectively. The former study [47] is particularly notable as the lgt mutant was attenuated in a mouse infection model. Second, lgt mutants of S. agalactiae [29] and Staphylococcus aureus [49] exhibit hypervirulent phenotypes in mouse models of infection. These phenotypes most likely reflect the failure of non-lipidated lipoprotein precursors to elicit protective effects of Lnt homologues or the lack of PPM synthase activity.

Box 1. Outstanding questions

- Are Gram-positive bacterial lipoproteins cell surface or ‘periplasmic’ proteins (or both)?
- Is the lipoprotein biosynthesis pathway coupled to the secretion apparatus?
- Why are there multiple Lgt enzymes in some Gram-positive bacteria?
- What is the role of Lnt homologues in Gram-positive actinomycetes?
- Which, if any, enzyme catalyses N-acylation in the absence of Lnt?
- Is the lipoprotein biosynthetic pathway a valid target for novel antimicrobials?
immune responses [49–51] possibly because of their inability to activate Toll-like receptor 2-mediated signalling [29]. By contrast, an lgt mutant of Listeria monocytogenes was attenuated in a mouse infection model even though the mutant failed to activate Toll-like receptor 2-mediated immune responses [52]. Thus, in lgt mutants there might be a strain specific balance between effects on immune activation and the functional compromisation because of the loss of lipoprotein lipidation. Individual lipoproteins could represent better prophylactic (vaccine) or therapeutic (drug) targets than the underlying pathway or therapeutic (drug) targets than the underlying pathway.

### Bioinformatic prediction of lipoproteins in Gram-positive bacteria

The sequence features that direct lipoprotein translocation and lipidation (i.e. type II signal peptides) are highly amenable to bioinformatic analyses, through the identification of the well conserved cysteine-containing lipobox [53–55]. Whole genome sequences can be searched for matches to either the Prosite profile PS51257 or the taxon restricted G+LPP sequence pattern (Box 2), which exhibits improved specificity for the identification of lipoproteins from Gram-positive bacteria [54,55]. These sequences can be validated using a range of online tools for lipoprotein identification. Using well-defined datasets of experimentally verified lipoproteins and decoy false-positives, it was recently demonstrated that LipoP [56] (http://www.cbs.dtu.dk/services/LipoP/) is the best performing single tool [55]. However, the most accurate way to analyse putative lipoprotein sequences (either individually or those recovered from whole genome screening) is to use a combination of LipoP with tools that allow recognition of their general signal peptide features (notably SignalP and Phobius). Ambiguous sequences can be further investigated using other online tools for lipoprotein identification [55] and predictors of membrane spanning domains. These predictors of membrane spanning domains can be useful in

### Table 1. Phenotypes exhibited by pathogenic Gram-positive bacteria mutated in the lipoprotein biosynthetic pathway

<table>
<thead>
<tr>
<th>Organism</th>
<th>Mutated gene</th>
<th>Observed phenotypesa</th>
<th>Model</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listeria monocytogenes</td>
<td>lgt</td>
<td>• Slightly lower growth rates in minimal medium</td>
<td>In vitro</td>
<td>[28,53]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Impaired intracellular growth in human epithelial (Caco-2) and mouse fibroblast (3T3) cell lines</td>
<td>In vitro</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Impaired TLR2 mediated immune activation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attenuation in a mouse infection model</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reduced growth and phagosomal escape within macrophages</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Similar intragastric growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Moderately attenuated virulence</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reduced growth in macrophages</td>
<td>In vitro</td>
<td>[32]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attenuated virulence</td>
<td></td>
<td>[85]</td>
</tr>
<tr>
<td></td>
<td>isp</td>
<td>• Growth attenuation in whole human blood and in presence of activated macrophages</td>
<td>In vitro</td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hypervirulent</td>
<td>Mouse intravenous</td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Failure of Lgt mutant to activate innate immune responses</td>
<td>In vitro</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td>lgt</td>
<td>• Growth attenuation in nutrient limited media</td>
<td>In vitro</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Failure of Lgt mutant to activate inflammatory immune responses</td>
<td>In vitro</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td>isp</td>
<td>• Attenuated virulence</td>
<td>Mouse intravenous</td>
<td>[49,66,67]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attenuated virulence</td>
<td>Signature tagged mutagenesis, multiple mouse models</td>
<td>[49,66,67]</td>
</tr>
<tr>
<td>Streptococcus agalactiae</td>
<td>lgt</td>
<td>• Growth attenuation in minimal media</td>
<td>In vitro</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hypervirulent in mice (low dose only)</td>
<td>Mouse subcutaneous</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Failure of Lgt mutant to activate immune responses via released Lpp interaction with TLR2</td>
<td>In vitro</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td>isp</td>
<td>• Growth attenuation in minimal media</td>
<td>In vitro</td>
<td>[29]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Failure of Lsp mutant to activate immune responses via TLR2</td>
<td>In vitro</td>
<td>[29]</td>
</tr>
<tr>
<td>Streptococcus equi</td>
<td>lgt</td>
<td>• Normal colonisation of horse organ cultures</td>
<td>In vitro</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Attenuated virulence in mice</td>
<td>Mouse intranasal</td>
<td>[47]</td>
</tr>
<tr>
<td>Streptococcus pneumoniae</td>
<td>lgt</td>
<td>• Virulence not significantly attenuated in ponies</td>
<td>Mouse intranasal infection (natural host)</td>
<td>[88]</td>
</tr>
<tr>
<td></td>
<td>isp</td>
<td>• 5-log attenuation in survival</td>
<td>Mouse co-infection models of septicemia (intraperitoneal) and pneumonia (intranasal)</td>
<td>[89]</td>
</tr>
<tr>
<td>Streptococcus suis</td>
<td>isp</td>
<td>• Reduced survival in human blood</td>
<td>In vitro</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Increased sensitivity to oxidative stress</td>
<td>In vitro</td>
<td>[48]</td>
</tr>
<tr>
<td>Streptococcus uberis</td>
<td>isp</td>
<td>• Minor colonisation defect in tonsil but virulence not attenuated</td>
<td>Porcine intranasal co-infection (natural host)</td>
<td>[36]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Normal growth in milk</td>
<td>In vitro growth model relevant to mastitis</td>
<td>[36]</td>
</tr>
</tbody>
</table>

*aAbbreviations: Lpp, lipoprotein; TLR2, toll-like receptor 2.*
Box 2. Signal peptide features of Gram-positive bacterial lipoproteins

Signal peptide features can be described using ‘pattern expressions’ written in Prosite syntax as shown in Table I. These patterns can be used for the bioinformatic identification of bacterial lipoproteins. \(<\cdot\rangle\) indicates the pattern is restricted to the N-terminus of the protein and at each position thereafter the amino acids shown are either permitted (square brackets) or prohibited (curly brackets). X is any amino acid. Where stretches of amino acids can vary in length, the range is indicated in parentheses. The original G+LPP pattern was described by analysis of the signal peptide features of 33 experimentally verified lipoproteins [54]. An extended dataset of 90 experimentally verified lipoprotein signal peptides was used to revise this pattern (G+LPVV2; [55]). The essential cysteine is considered the +1 position and, along with amino acids at positions –3 to –1, constitutes the ‘lipobox’. The Prosite profile P51257 (originally pattern PS00013) is based on the analysis of signal peptides from Gram-negative and other bacteria [53] and is notably more relaxed in the –2 and –3 positions.

Table I. Lipoprotein sequence patterns

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Pattern expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>G+LPP</td>
<td>&lt;[MV]:X(10,13):[RK]:(DERKQ)I(6,20)-[LIVMFESTAG]:[LVIAM]-[IVMSTAGF]-[AG]-C</td>
</tr>
<tr>
<td>G+LPVV2</td>
<td>&lt;[MV]:X(0,13):[RK]:(DERK)I(6,20)-[LIVMFESTAGC]:[LVIAMFTG]-[IVMSTAGCP]-[AGS]-C</td>
</tr>
<tr>
<td>P51257b</td>
<td>(DERK)I(6)-[LIVMFWTAG]:[IVMSTAGCQ]-[AGS]-C</td>
</tr>
</tbody>
</table>

*Additional rules apply, i.e. that there must be a K or R in the first seven amino acids and that the cysteine must appear between amino acids 15 and 35.

defining the location of the lipobox cysteine in relation to the putative signal peptide hydrophobic region (Figure 1).

Lipoprotein functions in Gram-positive bacteria

Bioinformatic analyses (as described earlier) have indicated that lipoproteins are a relatively abundant family of proteins, typically representing 2% or more of a Gram-positive bacterial proteome [18,19,43,54,57]. Functional analyses have revealed several recognisable functional groupings of Gram-positive bacterial lipoproteins [3], with these functions reflecting the localisation of lipoproteins within the cell envelope and more specifically at the interface between the membrane and the cell wall (Box 3). It is beyond the scope of this article to review these in detail, but key new observations are highlighted later.

Numerically, the most abundant functional grouping of lipoproteins is the SBPs of ABC importer systems, which typically represent ~40% of the putative lipoproteins in Gram-positive bacteria. Although ABC transport systems are found in all forms of cellular life, importers are only found in bacteria and archaea and these are dependent on an SBP to provide high affinity substrate binding and delivery to the cognate membrane permease components [58,59]. Thus, SBPs are vital to the specificity and directionality of ABC import systems. SBPs have been classified into at least nine subfamilies, reflecting the broad range of substrates transported [60–62] and make a vital contribution to the ability of prokaryotes to acquire diverse substrates from their environments. These include sugars, siderophores, divalent metal ions, anions (such as phosphate and sulphate), amino acids, oligopeptides and nucleosides. These are taken up primarily for nutrient acquisition but the substrates transported (notably in the case of peptide substrates) might also be important signals for environmental sensing, for example in the regulation of processes such as competence, quorum sensing and sporulation [3,63–66]. In an intriguing interplay between sensor and signal, it has been observed that the oligopeptide pheremone signals of enterococci are generated from proteolytic processing of lipoprotein signal peptides and taken up by lipoprotein-dependent oligopeptide ABC permeases [66,67]. These oligopeptides are released from the lipoprotein signal peptide by intramembrane proteases such as Eep, most likely following the cleavage of the signal peptide from the lipoprotein precursor by the action of Lsp [68].

In addition to the sensing of peptides described earlier, lipoproteins have roles in cell envelope sensing processes, including the modulation of two-component signal transduction systems [44,69,70] and the Bacillus germinant receptors GerAC, GerBC, GerD and GerKC [46,71]. Moreover, genome analyses have consistently identified putative lipoproteins predicted to have important roles in cell envelope stability and cell wall cross-linking or remodelling such as penicillin binding proteins and peptidoglycan hydrolases [18,54]. For example, ErfK domain (PFAM PF03734) L,D transpeptidases were recently discovered to have roles in the alternative 3–3 cross-linking of peptidoglycan in *Enterococcus faecium* and *M. tuberculosis* [72,73], perhaps as a way of recycling or remodelling the peptidoglycan. Many bacterial genomes encode several representatives of this family, a subset of which are predicted to be lipoproteins. For example, two of the four ErfK domain proteins in the *M. tuberculosis* H37Rv genome are predicted to be lipoproteins [18], as are all six encoded in the *S. coelicolor* genome. Two ErfK proteins in *E. coli* have also been demonstrated to cross-link Braun’s lipoprotein to the peptidoglycan [74] and so an additional possibility is that some members of this family are responsible for cell wall anchoring of proteins in a manner analogous to sortases [31].

Lipoproteins are also involved in the post-translational steps in the processing of exported proteins. Several Gram-positive members of the membrane insertase YidC family [75] are putative lipoproteins and there are numerous peptidyl-prolyl isomerase (PPIase; also termed foldase or maturase) lipoproteins in Gram-positive bacteria including the essential PrsA in *B. subtilis* [76] and FkbA in *Streptomyces anulatus* (formerly *Streptomyces chrysomallus*) [77] that most likely accelerate protein folding...
outside the cell. It is notable that several of these, including PrsA, belong to the parvulin subfamily of PPIases. However, because of sequence divergence, some representatives of this subfamily might not exhibit appreciable PPIase activity, in contrast to lipoprotein representatives of the cyclophilin family of PPIases [77, 78]. These lipoprotein PPIases are presumably well placed to interact with unfolded substrates emerging from the Sec translocon and could, thus, be important in processing virulence factors of Gram-positive pathogens. Streptococcus pneumoniae requires the SlrA PPIase for efficient murine colonisation [78] whereas a putative maturase lipoprotein mutant of S. equi was attenuated in both a mouse model and in a native equine host [47]. Finally, in staphylococci, the DsbA lipoprotein is a thiol-disulphide oxidoreductase apparently involved in disulphide bond formation in unidentified secreted protein substrates [79].

The role of DsbA illustrates that lipoproteins are also well placed to participate in electron transfer (redox) processes at the membrane surface. Indeed, several Gram-positive cytochrome c oxidase subunit II (CtaC) proteins have been experimentally validated as lipoproteins [43], as has the QoxA menaquinol oxidase [37]. These proteins have additional membrane spanning domains and the role of their lipid modification might be to appropriately orientate the N-terminus. Other small cytochromes in Bacillus (e.g. cytochrome c551) and other Gram-positive bacteria (e.g. Helicobacterium gestii cytochrome c553) are also lipoproteins [54]. In addition to these structural proteins, the B. subtilis Sco1 (YpmQ) accessory protein involved in cytochrome c oxidase assembly is also a lipoprotein [80]. Moreover, genomic analyses have revealed several intriguing instances in which thioredoxin-like lipoproteins are found in conjunction with DsbD/CcdA family proteins. DsbD/CcdA proteins are integral membrane proteins with roles in transferring electrons from the cytoplasm to the outer face of the plasma membrane, typically to allow the reduction of periplasmic or extracytoplasmic disulphide bonds [81]. One example of this is the system II pathway of cytochrome c maturation, which involves a DsbD/CcdA family protein and an associated thioredoxin-like protein (ResA in B. subtilis): these components perform an electron relay that enables the periplasmic or extracytoplasmic reduction of the apocytochrome c such that the prosthetic haem group can be inserted [82]. Several cytochrome c maturation loci have been identified in actinomycete genomes wherein the ResA homologue is a putative thioredoxin-like lipoprotein [19]. A second example of a potentially important electron relay is in the maintenance of secreted methionine sulfoxide reductase activity: several streptococcal genomes contain an operon encoding a secreted methionine sulfoxide reductase, a DsbD-family protein and thioredoxin-like lipoprotein [54]. As an electron donor is needed to maintain the catalytic activity of methionine sulfoxide reductase [83], it can be hypothesized that the DsbD-family proteins transfer electrons across the membrane and the thioredoxin-like lipoprotein relays the electrons to reduce the secreted methionine sulfoxide reductase. This pathway for electron transfer has received experimental support from elegant studies of the periplasmic PilB methionine sulfoxide reductase of Neisseria gonorrhoeae, which has a fused thioredoxin lipoprotein domain at its N-terminus and is likewise maintained by a cognate DsbD family protein [84]. Cumulatively, these examples illustrate that lipoproteins are appropriately localized to participate in cytoplasmic membrane redox processes.

Putative lipoproteins perform a wide variety of other predicted functions, including a diversity of enzymatic activities. Moreover, genomic analyses reveal a considerable proportion (typically ~30%) are conserved hypothetical proteins or hypothetical proteins of unknown function. An important challenge of the post-genomic era will be to assign functions to these proteins. Structural and biochemical analysis of the functions of these proteins will undoubtedly lead to a better understanding of bacterial cell envelope physiology and, most likely, the processing of exported proteins.

Concluding remarks and future perspectives
Our understanding of lipoprotein biogenesis has increased greatly since the pioneering work of Wu and colleagues [1]. Studies with Gram-positive bacteria have revealed that the basic enzymology of Lgt and Lsp are both necessary and sufficient for correct localization of lipoproteins, whereas the extent and significance of N-terminal acylation remains less clear. The advent of whole genome sequencing and our increasing knowledge of bacterial cell biology have changed the way we think about both the Gram-positive bacterial cell envelope and about protein export and localisation. The application of bioinformatic tools to the analysis of microbial genomes has revealed that putative lipoproteins not only represent a notable proportion (~2%) of the typical Gram-positive bacterial proteome but also are notable as cell envelope proteins that interact with membrane associated or exported proteins. Key areas for future study (Box 1) will include the investigation of the lipoprotein biogenesis machinery as a terminal branch of the protein translocation pathway, characterisation of N-acylation and the Lnt homologues in the low and high-GC Gram-positive bacteria, respectively, and the identification of essential lipoproteins as novel drug targets in Gram-positive bacteria. It is clear from recent discoveries, such as the translocation of lipoproteins through Tat, that we still have a lot to learn about lipoprotein biogenesis and the post-translational modification of proteins outside the bacterial cell.

Acknowledgements
We thank members of our laboratories past and present for their contribution to our research on lipoprotein biogenesis and our many colleagues with whom we have discussed lipoprotein assembly and function. Research into lipoprotein biogenesis in the authors’ laboratories is currently supported by the Biotechnology and Biological Sciences Research Council (grant numbers F009224/1, F009429/1, EGH16082), the Medical Research Council (MRC), the Commission of the European Community (grant LSHG-CT-2004–005257) and The Royal Society. T.P. is an MRC Senior Non-Clinical Research Fellow and M.I.H. is a Research Councils UK Academic Fellow.

Appendix A. Supplementary data
References


18 Bengtsson, J. et al. (1999) Bacillus subtilis contains two small c-type cytochromes with homologous heme domains but different types of membrane anchors. J. Biol. Chem. 274, 26179–26184

19 Dartois, V. et al. (1997) KapI is a lipoprotein required for K1B signal transduction and activation of the phosphorelay to sporulation in Bacillus subtilis. Mol. Microbiol. 26, 1097–1108


22 Hamilton, A. et al. (2006) Maturation of the mature lipoprotein attenuates the virulence of Streptococcus equi to a greater extent than does loss of general lipoprotein lipidation. Infect. Immun. 74, 6907–6919


75 Serek, J. et al. (2004) Escherichia coli YidC is a membrane insertase for Sec-independent proteins. EMBO J. 23, 294–301
78 Hermans, P.W. et al. (2005) The streptococcal lipoprotein rotamase A (SltR) is a functional peptidyl-prolyl isomerase involved in pneumococcal colonization. J. Biol. Chem. 281, 968–976
84 Brot, N. et al. (2006) The thioredoxin domain of Neisseria gonorrhoeae PiII can use electrons from DsbH to reduce downstream methionine sulfoxide reductases. J. Biol. Chem. 281, 32668–32675

Reproduction of material from Elsevier articles

Interested in reproducing part or all of an article published by Elsevier, or one of our article figures?
If so, please contact our Global Rights Department with details of how and where the requested material will be used. To submit a permission request online, please visit:

www.elsevier.com/locate/permissions