Outer membrane lipoprotein biogenesis: Lol is not the end

Anna Konovalova and Thomas J. Silhavy

Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Washington Road, Princeton, NJ 08544, USA

Bacterial lipoproteins are lipid-anchored proteins that contain acyl groups covalently attached to the N-terminal cysteine residue of the mature protein. Lipoproteins are synthesized in precursor form with an N-terminal signal sequence (SS) that targets translocation across the cytoplasmic or inner membrane (IM). Lipid modification and SS processing take place at the periplasmic face of the IM. Outer membrane (OM) lipoproteins take the localization of lipoproteins (Lol) export pathway, which ends with the insertion of the N-terminal lipid moiety into the inner leaflet of the OM. For many lipoproteins, the biogenesis pathway ends here. We provide examples of lipoproteins that adopt complex topologies in the OM that include transmembrane and surface-exposed domains. Biogenesis of such lipoproteins requires additional steps beyond the Lol pathway. In at least one case, lipoprotein sequences reach the cell surface by being threaded through the lumen of a beta-barrel protein in an assembly reaction that requires the heteropentomeric Bam complex. The inability to predict surface exposure reinforces the importance of experimental verification of lipoprotein topology and we will discuss some of the methods used to study OM protein topology.

1. Introduction

Bacterial lipoproteins are important components of the Gram-negative cell envelope. Although the name ‘lipoprotein’ is often used to describe a non-covalent assembly of proteins and lipids, bacterial lipoproteins belong to the class of so-called lipid-anchored proteins. These proteins contain covalently attached acyl groups as a result of a post-translational modification. Eukaryotic lipid-anchored proteins can be modified at different sites and the modifications are sometimes reversible and may be involved in regulation of the protein’s activity [1]. By contrast, bacterial lipoproteins are modified at a specific site corresponding to the N-terminal Cys residue of the mature lipoprotein [2–4]. This modification is irreversible and happens on the periplasmic site of inner membrane (IM). The lipid moiety causes membrane association of a protein that usually does not contain a hydrophobic membrane targeting domain. Lipoproteins are then sorted to the IM or the outer membrane (OM) based on the sequences which follow the conserved Cys residue. Lipoproteins which are destined to OM are transported there by the localization of lipoproteins (Lol) pathway and inserted in the inner leaflet of the OM [5]. For several decades, lipoproteins were assumed to be simply soluble periplasmic proteins tethered to the OM by their lipid, partly because methods to study OM topology had not been developed. In this review, we will demonstrate, using Escherichia coli as a model organism, that for many lipoproteins biogenesis does not end with the Lol pathway. Lipoproteins can be further translocated onto the cell surface and they can be assembled into OM protein (OMP) complexes with transmembrane or surface-exposed topologies. We will also discuss some of the commonly used techniques to study the topology of proteins in the OM with its unique barrier properties.
2. Lipoprotein maturation and processing

Lipoproteins are synthesized in the cytoplasm as protein precursors with an N-terminal signal sequence (SS) for transport from the cytoplasm. The lipoprotein SS contains a characteristic consensus sequence [LVI][ASTVI][GAS]C known as a lipobox [6,7]. In *E. coli*, lipoproteins are translocated across the IM by the Sec translocon, but in other organisms there are lipoproteins dependent on Tat translocon [8,9]. Maturation and processing of lipoproteins takes place on the periplasmic side of IM (figure 1). The first step is the addition of diacylglycerol to the sulfhydryl group of the conserved Cys residue by phosphatidylglycerol/propolipoprotein diacylglyceryl transferase (Lgt) [10]. This modification is a prerequisite for the SS cleavage by a dedicated lipoprotein signal peptidase (LspA or signal peptidase II) [2,11,12] and ensures that the apolipoprotein remains anchored in the IM after SS cleavage. Signal peptidase generates diacylated apolipoprotein with a new N-terminus provided by the conserved Cys of the lipobox that is referred to as +1 [3]. Finally, the amino group of the Cys residue is N-acylated by phospholipid/apolipoprotein transacylase (Lnt) generating mature triacylated lipoprotein [13,14] (figure 1). Genes encoding for all three enzymes are essential in *E. coli* and highly conserved across Gram-negative bacteria [5]. However, *Int* was recently found to be dispensable for viability of *Francisella tularensis* and *Neisseria gonorrhoeae* [15], suggesting that in some Gram-negative bacteria, mature lipoproteins can exist in a diacylated form similar to what is observed in low GC Gram-positive organisms which do not encode Lnt homologues [16].

3. Export of lipoproteins to the outer membrane

Most of the mature lipoproteins in *E. coli* are targeted for translocation to the OM by the Lol pathway unless they contain a so-called Lol avoidance signal. The Lol avoidance signal in *E. coli* is known as the +2 rule because it is determined by the identity of the amino acid after the conserved Cys [17]. According to this rule, an Asp residue at position +2 causes IM retention of the lipoprotein; it serves as a sorting signal that differentiates IM and OM lipoproteins [17]. Although all native IM lipoproteins in *E. coli* have Asp at +2 and either Asp, Glu or Gln at position +3, additional combinations of +2 and +3 residues could act as IM retention signals [18–20]. Although the +2 rule is generally conserved in enterobacteria [21], the rule does not always apply for species outside this family. For example, in *Pseudomonas aeruginosa* amino acids at positions +3 and +4 also play a critical role in lipoprotein sorting [22,23]. OM lipoproteins in *Borrelia* do not follow the +2 rule either [24].

Lipoproteins destined for the OM are translocated by the Lol proteins (figure 1). The Lol pathway was discovered and characterized by a series of elegant biochemical experiments in the Tokuda laboratory. In *E. coli*, five proteins (LolA–E) are involved in lipoprotein transport. In the first step, the OM lipoprotein is extracted from the IM by the LolICDE complex and released to the periplasm in the form of a soluble complex with the chaperone LolA [25,26]. This step is energy-dependent and driven by ATP hydrolysis by the ATPase LolD [26,27]. LolD is a cytoplasmic ABC-type ATPase tethered to the IM by the interaction with the homologous integral IM proteins LolC and LolE [26]. Although homologous LolC and LolE play different functions in lipoprotein extraction from the IM, LolE recognizes and binds lipoprotein substrates [28], whereas LolC recruits the chaperone LolA [29]. ATP hydrolysis powers the transfer of the lipoprotein directly from LolE to LolA [28]. A large hydrophobic cavity within LolA binds the acyl chains of protein substrates shielding this highly hydrophobic region from the aqueous periplasm [30]. LolA delivers lipoprotein substrates to the OM acceptor protein LolB [31]. LolB is an OM lipoprotein itself and is a structural homologue of LolA [30]. It also contains a hydrophobic cavity that has a higher affinity for the acyl chains than LolA [32]. This difference in affinity allows unidirectional ‘mouth-to-mouth’ transfer of the substrate from LolA to LolB [32]. The hydrophobic cavity of LolA undergoes conformational changes between the ‘open’ and substrate-bound and the ‘closed’ or substrate-free form [33]. After lipoprotein release, LolA in its closed form returns to the LolCDE complex and is recycled (figure 1). Lipoprotein release from LolB in the OM is not well understood, but the protruding loop of LolB is somehow important for this final step in lipoprotein transport [34,35].

The lol genes are essential in *E. coli* and homologues can be found in all Gram-negative bacteria, suggesting that the pathway is conserved. However, conservation of individual genes varies. LolC and LolE are homologues but cannot functionally substitute each other in *E. coli* [28]. However, some bacterial genomes contain only one copy of a lolC/E gene [15]; interestingly, in such cases, the protein product contains sequence motifs of both LolC and LolE and likely represents a functional hybrid of both proteins. The LolF name was proposed to distinguish such proteins from obvious LolC and LolE homologues [15], lolB is the other gene that is only conserved in β- and γ-proteobacteria [5]. It is not clear whether other Gram-negative bacteria contain functional analogues which are not related in sequence to LolC or encode a functional hybrid of LolA and LolB since these proteins also have similar structures.

4. Lipoprotein destiny after outer membrane insertion

Until recently, lipoprotein insertion in the OM by LolB was the last known step for lipoprotein export in *E. coli*. Therefore, the paradigm for OM lipoprotein topology was established that all lipoproteins (at least in *E. coli*) are simply anchored in the inner leaflet of OM by their lipid moiety with their protein domain facing the periplasm (figure 2). Many surface-exposed lipoproteins have been identified in a number of different organisms [36]. In those organisms, surface-exposed lipoproteins have variety of functions: they participate in iron uptake [37–40]; are enzymes such as phospholipases [41], PPlases [42] or glucanases [43]; they participate in adhesion and binding of host factors [44,45] and others. In some organisms, like *Borrelia*, lipoproteins are anchored in the outer leaflet and exposed on the cell surface by default [24].

In most cases, how lipoproteins get to the cell surface is unclear. Surface-associated lipoprotein pullulanase (PulA) of *Klebsiella oxytoca* was the first lipoprotein for which the export pathway was identified [43,46]. PulA contains a Lol avoidance signal and is instead exported directly from the IM to the cell surface by a specific transporter Pul, also
known as a type 2 secretion system [43,47]. Other known lipo-
protein substrates of type 2 secretion systems are surface-
exposed lipoprotein cytochromes in _Shewanella oneidensis_ [48]
and lipoprotein SslE in enteropathogenic _E. coli_ [49].
_Neisseria meningitidis_ NalP and _Bordetella pertussis_ SphB1 are autotransporters which contain a lipoprotein SS
[50–53]. These proteins have a domain architecture typical
of autotransporters: a C-terminal translocation domain that
is required for the secretion of an N-terminal passenger
domain [54]. However, due to the presence of N-terminal
lipid on the passenger domain, it must remain associated
with the OM, at least for a while, where it functions as a
surface protease [50,51].

Many reports of surface-exposed lipoproteins from differ-
et organisms are in contrast to the _E. coli_ periplasmic
paradigm. However, of the 90 lipoproteins in _E. coli_ fewer
than a dozen have been studied experimentally, mainly in
connection to peptidoglycan (PG) and envelope biogenesis.
Of those lipoproteins that have been examined, many adopt
quite complex topologies in the OM: they can exist in com-
plexes with beta-barrel proteins, form transmembrane
channels or be surface exposed.

**5. Complex topologies of outer membrane lipoproteins**

(a) Plug in the barrel

LptE is the OM lipoprotein component of the LPS transport
(Lpt) machine; together with its beta-barrel partner LptD, it
forms an OM subcomplex that is required for LPS insertion
into the outer leaflet of OM [55,56]. N-terminal lipids
anchor LptE in the inner leaflet of OM; however, instead of
facing the periplasmic space, LptE is almost completely
buried inside the large lumen of LptD [57–59] (figure 2).
The interaction between LptE and LptD is highly specific
and the complex is very stable [59]. This is achieved by a large
number of specific contacts sites between LptE and LptD cov-
ering an interaction interface with the surface area of more
than 3000 Å² [57,58]. LptE serves as a scaffold for the folding
of the LptD barrel at the Bam complex, and no LptD protein
can be assembled without its partner; unassembled LptD is
rapidly degraded [60,61]. In addition to its role in LptD fold-
ing, LptE functions as a plug that prevents small molecule
entry through the LptD barrel and it plays a direct role in
the surface assembly of the LPS glycolipid [62,63].

(b) Integral outer membrane lipoproteins

There are two lipoproteins with transmembrane topology
in _E. coli_, CsgG and Wza. Both lipoproteins are OM compo-
nents of secretion channels. CsgG functions in the
transport of curli subunits [64] and Wza transports group 1
(serotype K30) capsular polysaccharides [65]. The lipids of
both CsgG and Wza are anchored in the inner leaflet of
OM and both proteins form oligomeric complexes with
hydrophobic transmembrane domains and large periplasmic
domains [66–68]. Each oligomeric complex features a central
channel used for the secretion of the corresponding substrates
[66–68]. Despite the similarity in topology, CsgG and Wza
Figure 2. Examples of topologies of OM lipoproteins in *E. coli*. OM lipoproteins (green) were assumed to adopt a typical topology with the protein domain in the periplasm and lipid anchored in the inner leaflet of the OM. By contrast, the topology of experimentally studied lipoproteins is often more complex. Lipoproteins can be assembled in interlocked complexes with beta-barrel proteins (blue) like RcsF/OMP and LptE/LptD, or they can form transmembrane multimeric channels like CsgG and Wza. Assembly of the lipoproteins at the OM often results in surface exposure of parts of the protein (RcsF, CsgG and Wza). In the case of RcsF, the lipid moiety anchors the protein in the outer leaflet of OM. (Online version in colour.)

have completely different structures and interact with the OM in different manners.

The CsgG channel is an oligomer of nine CsgG subunits [66,67] (figure 2). Each subunit has alpha–beta structure, and the transmembrane domain is made by amphipathic beta strands. Each subunit donates four beta strands building up a composite 36-stranded beta-barrel which forms a transmembrane channel with an inner width of 40 Å. Although CsgG does not have long extracellular loops typical of many OMPs, it still can be detected on the cell surface by CsgG antibodies [69]. The large periplasmic domain is made of a mixture of α-helices and β-strands, some of which extend from the transmembrane barrel. The periplasmic domain of the CsgG nonamer also features a large cavity that is open on the periplasmic side but gated in *vivo* by a partner protein CsgE [67]. Based on the cryo-EM structure of the CsgG/E complex as well as conductance measurements of the CsgG and CsgG/E channels, a model was proposed in which CsgE binding to CsgG entraps curli subunit protein CsgA inside the CsgG chamber, creating conditions that favour diffusion out through CsgG pore [67,70]. After diffusion, CsgA subunits are nucleated into curli amyloid fibres by CsgB [71]. Curli play an important role in adhesion, biofilm formation and colonization of the host [72].

Wza has different structure and domain organization [68]. Wza is an octamer with four rings (R1–R4) of individual domains (figure 2). The Wza octamer is described as an amphora without handles, in which the R1–R3 domains are formed similar to what are observed in the full-length Wza structure [76,77]. Hence, the lipid moiety anchors the protein in the outer leaflet of OM.

Assembly of CsgG and Wza oligomers at the OM does not require the Bam complex, but the assembly pathways remain known [74]. In both cases, the lipid moiety is required for formation of functional multimers. When lipid is absent, as a result of SS replacement, CsgG proteins accumulate as soluble monomers in the periplasm and the few multimers that can be observed have the incorrect symmetry. By contrast, lipid-less Wza can be still targeted to the OM to some extent, but no stable, functional multimers are formed [75]. Interestingly, when both the transmembrane helix and the lipid are removed, soluble multimers of the R1–R3 domains are formed similar to what are observed in the full-length Wza structure [76,77].

(c) Surface-exposed lipoproteins

Perhaps another reason that contributed to the periplasmic paradigm for lipoproteins is that the two best-studied lipoproteins in *E. coli*, Lpp and Pal, have clearly defined periplasmic functions where they interact with the cell wall. Lpp, also known as Braun’s lipoprotein, is the major lipoprotein in *E. coli*. It is a small 6 kDa protein that exists in two forms: one known as the ‘bound form’ represents Lpp molecules that are covalently cross-linked to PG [78,79]; another fraction is not attached to the PG and is known as the ‘free form’ [80]. The function of the free form is not understood. In contrast to Lpp, Pal interacts with the PG by non-covalent binding [81]. Both lipoproteins contribute to general envelope stability by attaching the OM to the cell wall and null mutations in either gene causes increased OM vesiculation, sensitivity to detergents and other envelope-related defects [82–84]. Both Lpp and Pal were used as model lipoproteins in studies of the Lol pathway [25,26,31]. Because of their clear function in the attachment of the OM to the cell wall, periplasmic localization of these proteins had never been questioned until we discovered that the ‘free form’ of Lpp adopts a cell surface-exposed orientation [85]. We uncovered this topology of Lpp in our effort to characterize the surface proteome of *E. coli* by selective surface biotinylation: Lpp was one of the major proteins that
are labelled in *E. coli*. Although we do not know the precise topology of surface-exposed Lpp, studies of biotinylation sites demonstrate that at least the C-terminus of Lpp is exposed on the cell surface [85]. Since then, the topology of Pal has also been challenged and new data demonstrate that Pal, like Lpp, may exist in this dual orientation with a fraction of the protein being surface exposed [86]. Additional surface-exposed lipoproteins reported in *E. coli* are TraT [87], BamC [88] and YaiW [89].

The surface exposure of Lpp posed the question as to whether there is a specific transporter to localize lipoproteins to the cell surface. However, because the function of free form Lpp is not known, we decided to look for a more experimentally tractable system to probe the mechanism(s) that catalyse lipoprotein surface exposure.

The lipoprotein RcsF stood out because of its function as a sensory component of Rcs envelope stress response [90,91]. The Rcs stress response monitors envelope integrity and can be activated by a number of cues such as LPS stress [92,93], PG stress [94,95], osmotic stress [96,97] and defects in lipoprotein biogenesis [98,99]. Our hypothesis that RcsF may localize to the outer leaflet to sense LPS defects led to the discovery of yet another surface-exposed lipoprotein in *E. coli* [100]. Later, it was shown that RcsF activation to the PG stress does not require surface localization RcsF [101], suggesting distinct mechanisms for monitoring integrity of different envelope compartments.

Not only is RcsF a surface-exposed lipoprotein, it displays a novel OM topology: the N-terminal domain and the lipids of RcsF are surface exposed, while the C-terminal domain resides in the periplasm (figure 2). How could RcsF span the OM if it does not contain a transmembrane domain? The solution to this topological problem proved to rely on RcsF interaction with several OMPs via a short alpha helical sequence which exists in an extended conformation that is threaded through the OMP lumen where it is shielded from the hydrophobic membrane interior. Therefore, like LptE, RcsF is also found in an interlocked complex with a beta-barrel protein. However, there are number of important differences between RcsF/OMP and LptE/D complexes. Unlike LptE, RcsF interacts with the beta-barrel via a short unstructured region only, without extensive site-specific contacts. This lack of barrel-specificity is further evidenced by the fact that RcsF is capable of interacting with several different OMPs such as OmpA, OmpC and OmpF. It is likely that the presence of an unstructured region, rather than sequence specificity, is the requirement for formation RcsF/OMP complexes. Strikingly, a similar type of interaction was observed for the unstructured linker region of a colicin during its entry through the OmpF lumen [102,103]. Finally, LptE is absolutely required for the folding and assembly of LptD. RcsF is not required for the folding and assembly of any OMP.

Having established the surface exposure and topology of RcsF, we wanted to examine the assembly mechanism responsible. How can the unstructured region of RcsF be threaded into the lumen of an OMP? RcsF cannot enter the lumen from N-terminus because even though it is unstructured [104,105], it is tethered to the membrane by a lipid moiety that cannot be extracted from the membrane without an energy source, and even if it could be, the hydrophilic lumen of a folded OMP would not allow the passage of the hydrophobic lipid. Likewise, the folded C-terminal domain of RcsF cannot pass through the lumen because it is simply too big and is folded at an early step of RcsF biogenesis with the help of periplasmic disulphide oxidase DsbA, and the disulphide isomerase DsbC, which catalyse the formation of the two non-consecutive disulphide bonds necessary to stabilize the structure [104,105]. Therefore, to explain the transmembrane topology, it seemed necessary to suggest that the OMP beta-barrel is folded around the RcsF during the process of OMP assembly.

We have provided two lines of evidence to support the hypothesis that the RcsF/OMP interaction is formed during OMP assembly. First, we showed that RcsF cannot form a complex with folded OmpA *in vitro*; however, the complex can be reconstituted when urea-denatured OmpA is refolded in the presence of RcsF. Second, all OMPs are folded *in vivo* by the heteropentomeric Bam complex [106], and we showed that RcsF also interacts with BamA, an essential component of the Bam complex. Although BamA is itself a beta-barrel, site-specific cross-linking revealed that the RcsF/BamA interaction is different from the RcsF/OMP interaction, suggesting that the RcsF/BamA cross-linking observed could represent a step in the assembly pathway of RcsF/OMP at the Bam complex. Indeed, when the function of the Bam complex is compromised, fewer RcsF/OMP complexes are assembled at the OM. These results argue that Bam complex catalyses RcsF surface exposure by assembling the RcsF/OMP complexes at the OM.

The study of RcsF uncovered novel functions for Bam complex in the biogenesis of surface-exposed lipoproteins. BamA can recognize lipoproteins and translocate the lipidated N-terminus into the outer leaflet and then assemble the OMP beta-barrel around the short, unstructured linker between the N- and C-terminus. We think it is likely that other lipoproteins exist in similar OMP complexes in *E. coli* as well as other bacteria and that the Bam complex is involved in their assembly. Clearly, as is demonstrated by PulA, the Bam complex is not required for the surface localization of every lipoprotein. We think it is possible that other structural groups (e.g. a-helical proteins) have their own dedicated assembly pathway.

6. Common methods to study topology of outer membrane lipoproteins

The examples above demonstrate that lipoproteins vary in their structures and topologies, and the latter has to be experimentally determined. Generally, any method commonly used to study membrane protein topology can be employed to study OM lipoproteins [107,108] but the unique features of the OM have to be taken into account when designing the experiment. All methods can be divided into three groups: those based on protein labelling; based on antibody binding; and accessibility for proteolysis.

Techniques based on selective protein labelling use compounds that chemically modify certain amino acids in the protein and allow the detection of the modified protein. The two most commonly used chemistries are N-hydroxysuccinimide (NHS) esters which react with primary amines (e.g. protein lysines), or maleimides which react with sulphydryls (e.g. non-oxidized cysteines) [109]. These functional groups can be coupled chemically to biotin and this allows detection of the modifications using anti-biotin antibodies or streptavidin conjugates. PEG linkers of different length,
which allow detection of the modification based on a protein size shift during immunoblot analysis, can also be used. There are a huge variety of labels that vary in their size and hydrophobicity, and choosing the right one requires understanding the two important properties of the OM that differentiate it from other biological membranes [110]. First, the OM is not permeable to hydrophobic compounds because of the presence of charged, highly interactive lipopolysaccharide molecules that cover the cell surface. Second, the OM is permeable to the small soluble compounds due to the presence of porins which allow water and nutrients to freely enter the cell [110]. Therefore, selection of compounds for labelling OM proteins is somewhat counterintuitive compared to other membrane proteins: hydrophobic compounds will selectively label the cell surface [85], while small hydrophilic compounds can enter the periplasm and can label protein on both sides of the OM. The size exclusion limit of porins in E. coli was determined to be around 600 Da [111], so using water-soluble labels of much larger molecular weight will result in selective surface labelling as well. In our laboratory, we used a hydrophobic biotin label to study Lpp surface exposure [85] as well as for the initial discovery of surface-exposed RcsF [100]. We also used a hydrophilic maleimide with high molecular weight to study the conformation change in extracellular loop 6 of BamA [112]. It is worth mentioning that perhaps not all Gram-negative bacteria have such an asymmetric OM and the natural resistance to hydrophobic OM compounds that E. coli does, and protein labels should be evaluated in each case.

Surface proteolysis is another commonly used method to test protein surface exposure [46,113]. Extracellularly added proteases cannot enter the cell and hence periplasmic proteins are protected from cleavage while surface-exposed proteins are degraded. Cleavage or complete degradation can be analysed by immunoblotting. However, it should be noted that many surface-exposed proteins in bacteria are inherently protease resistant, so negative results cannot be interpreted with confidence, even if proper controls have been done.

Using extracellularly added antibodies follows the same logic as surface proteolysis. Antibodies can bind only surface-exposed epitopes and this binding can be detected either by using antibodies labelled with fluorescent dyes or enzyme conjugates. Whole-cell antibody-based assays include dot blots, whole cell ELISA, immunofluorescence and FACS analysis [114–116]. However, cell fixation techniques can disrupt OM integrity as well as producing other artefacts and it is best for antibody labelling to be performed with live cells [117].

Labelling techniques, especially biotinylation, are widely used for proteomic discovery of surface-exposed proteins because they do not require the generation of protein-specific antibodies: labelled proteins can be easily affinity purified and analysed by mass spectrometry [118,119]. Quantitative proteomics can be also employed to study surface-exposed proteins by comparing relative abundance of the protein before and after the protease treatment [120]. Employing a combination of different methods provides more reliable identification of surface-exposed proteins [113].

All of the methods described can be used to study native proteins and do not require making protein fusions or mutants for site-specific labelling. However, the limitation is that ability to detect a protein with any of the methods depends not only on protein localization to the cell surface but also on the presence of accessible sites for labelling, proteolysis or antibody binding. Lack of such sites can generate false-negative results. One example of such false-negative results that we have encountered is BamC. BamC is an OM lipoprotein component of the Bam complex [106]. We have highly specific polyclonal anti-BamC antibodies [121] which do not bind to whole cells [100] and we could not detect BamC biotinylation [85]. Therefore, we concluded that BamC is a periplasmic lipoprotein, and we have used BamC as a negative-control for lipoprotein surface exposure [85,100]. However, independently raised polyclonal antibodies against BamC can recognize the protein on the cell surface, indicating that at least part of BamC is surface exposed [88]. The apparent discrepancy between these results can be explained by differences in antibody epitope recognition. Identifying these epitopes would help in understanding the topology of BamC. It is likely that only a portion of the protein is exposed on the cell surface; the structure of BamC/BamD complex shows that the N-terminal domain of BamC is involved in the interaction with periplasmically localized lipoprotein BamD [122].

One of the other disadvantages of using these methods to study native proteins is that they seldom give information about precise OM topology. Therefore, a number of approaches were developed to introduce labelling sites (e.g. Cys accessibility assay), proteolytic sites for specific proteases (e.g. TEV protease, enterokinase) or epitopes for antibody-binding (His, HA, FLAG, Strep epitopes) at specific positions of the protein. This method was successfully used to study OMPs. For example, labelling of native and engineered Cys was used in studies of the ferrichrome-binding extracellular loop of FhuA [123], and epitope insertions were used to study the topology of several OMPs in E. coli [124–127]. Insertion of proteolytic sites was used to determine that surface-exposed lipoproteins in Borrelia are anchored to the outer leaflet of OM [128]. These approaches are recommended only when assays to test protein function and/or folding are available to demonstrate that epitope insertions do not interfere with native protein topology. In our study of RcsF, we used biotinylation, proteolysis and polyclonal RcsF antibodies to show that RcsF is surface exposed [100]. However, FLAG epitope walking across the entire protein sequence allowed us to uncover the fact that RcsF has an N-out C-in topology.

7. Concluding remarks

The lipoprotein biogenesis pathway, including lipoprotein maturation and export to the OM by the Lol machine, has been characterized in great molecular detail. Because lipoprotein insertion into the inner leaflet of OM is the last step in the Lol pathway, the periplasmic paradigm for lipoproteins, in which all lipoproteins are viewed as soluble periplasmic proteins simply anchored to the inner leaflet of the OM by N-terminal lipid moieties, was established. This assumption delayed the discovery of many lipoproteins that interact with the OM differently: some lipoproteins are anchored in the outer leaflet of the OM, some are transmembrane proteins and some exist in protein complexes shielded from the periplasm. Identification of such proteins made us acknowledge that our long-held view on lipoprotein topology was oversimplified and our understanding of lipoprotein biogenesis...
was far from complete. We suggest that a periplasmic topology should no longer be assumed for OM lipoproteins but instead a predicted topology should be experimentally tested. We believe that this would lead to the discovery of even more examples of OM lipoproteins with unusual structures and topologies, presenting us with even more challenges for understanding OM biogenesis.

References


70. Bürst, N., and H. Markov. 1997 Nucleator function of CsgB for the assembly of adhesive surface organelles in


