Review

Correspondence
Pierre Cornelis
pcornel@vub.ac.be

Genome-wide analysis and literature-based survey of lipoproteins in *Pseudomonas aeruginosa*

Kim Remans,¹ Ken Vercammen,² Josselin Bodilis³ and Pierre Cornelis²

¹Department of Molecular and Cellular Interactions, Structural Biology Brussels, VIB, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
²Department of Molecular and Cellular Interactions, Microbial Interactions, VIB, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
³Groupe Microbiologie, Laboratoire M2C, UMR CNRS 6143, UFR des Sciences – Université de Rouen, 76821 Mont Saint Aignan, France

*Pseudomonas aeruginosa* is a Gram-negative opportunistic pathogen able to cause acute or chronic infections. Like all other *Pseudomonas* species, *P. aeruginosa* has a large genome, >6 Mb, encoding more than 5000 proteins. Many proteins are localized in membranes, among them lipoproteins, which can be found tethered to the inner or the outer membrane. Lipoproteins are translocated from the cytoplasm and their N-terminal signal peptide is cleaved by the signal peptidase II, which recognizes a specific sequence called the lipobox just before the first cysteine of the mature lipoprotein. A majority of lipoproteins are transported to the outer membrane via the LolCDEAB system, while those having an avoidance signal remain in the inner membrane. In *Escherichia coli*, the presence of an aspartate residue after the cysteine is sufficient to cause the lipoprotein to remain in the inner membrane, while in *P. aeruginosa* the situation is more complex and involves amino acids at position +3 and +4 after the cysteine. Previous studies indicated that there are 185 lipoproteins in *P. aeruginosa*, with a minority in the inner membrane. A reanalysis led to a reduction of this number to 175, while new retention signals could be predicted, increasing the percentage of inner-membrane lipoproteins to 20%. About one-third (62 out of 175) of the lipoprotein genes are present in the 17 *Pseudomonas* genomes sequenced, meaning that these genes are part of the core genome of the genus. Lipoproteins can be classified into families, including those outer-membrane proteins having a structural role or involved in efflux of antibiotics. Comparison of various microarray data indicates that exposure to epithelial cells or some antibiotics, or conversion to mucoidy, has a major influence on the expression of lipoprotein genes in *P. aeruginosa*.

*Pseudomonas aeruginosa*

The Gram-negative bacterium *Pseudomonas aeruginosa* is an opportunistic pathogen and a notorious nosocomial infective agent. This bacterium infects immunocompromised persons, such as AIDS patients and people with severe burn wounds, and is a serious risk for intensive care unit patients (Lyczak et al., 2000; Meynard et al., 1999; Pirnay et al., 2003). *P. aeruginosa* is also the predominant pathogen found in the airways of cystic fibrosis (CF) patients. In CF lungs, *P. aeruginosa* switches to a mucoid alginate exopolysaccharide producing phenotype, which causes chronic infections that correlate with high mortality rates in CF patients (Lyczak et al., 2002; Ratjen & Doring, 2003). The multiple occurrence of infections caused by this bacterium is largely due to its high degree of antibiotic resistance, which is the result of the synergistic effect of the low permeability of the outer membrane (OM) and the existence of numerous multidrug efflux pumps, which secrete the drugs directly out of the cell (Schweizer, 2003). A class of acylated proteins, called lipoproteins, has been shown to play a role in many fundamental cellular processes and in the pathogenesis of several bacterial infections. In *P. aeruginosa*, microarray analysis has revealed that there is a prominent induction of lipoprotein-encoding genes during mucoid conversion (Firoved et al., 2004). Lipoproteins are also capable of eliciting a strong host immune response. They are recognized by Toll-like receptors present on the surface of several antigen-presenting cells, which leads to the activation of a signal transduction cascade that initiates the expression of pro-inflammatory cytokines (Liang et al.,

Abbreviations: CF, cystic fibrosis; IM, inner membrane; MFP, membrane fusion protein; OM, outer membrane; PG, peptidoglycan; RND, resistance-nodulation-division.
The genome of *P. aeruginosa* comprises about 6200 kb and 3.3% of this genome is predicted to encode 185 different lipoproteins (Lewenza *et al.*, 2005). However, despite the abundance and importance of this class of proteins, very little is known about the numerous *P. aeruginosa* lipoproteins. Almost 60% of the predicted lipoproteins are classified as hypothetical proteins in the *Pseudomonas* Genome Database (Winsor *et al.*, 2009).

### Bacterial lipoproteins

Lipoproteins are synthesized in the cytoplasm as precursors with an N-terminal signal sequence and are translocated across the inner membrane (IM) via the Sec pathway (or in rare instances by the Tat pathway). The signal sequence always ends in a consensus sequence called the lipobox. The lipobox consists of four amino acids, namely Leu-Ala/Ser-Gly/Ala-Cys, with Cys being the first amino acid of the mature lipoprotein (Hayashi & Wu, 1990). In the first step, a diacylglycerol group is attached to the thiol group of the Cys via a thioether linkage. This step is catalysed by the enzyme phosphatidylglycerol/prolipoprotein diacylglycerol transferase (Lgt). Next, the signal sequence is cleaved off just before the Cys by the prolipoprotein signal peptidase (LspA or signal peptidase II), which is specific for lipoproteins (Kiho, 1990). In the last step, the cysteine is also aminoacylated by the phospholipid/apoprotein transacylase (Lnt) to create the mature lipoprotein with three acyl chains attached to its N-terminal Cys residue (Sankaran & Wu, 1994). Although their maturation occurs at the IM, most lipoproteins are located at the periplasmic side of the OM. Work of H. Tokuda and co-workers revealed that the transport of these OM lipoproteins is mediated by the Lol system, which is composed of five proteins, LolABCDE, also present in *P. aeruginosa* (Tanaka *et al.*, 2007). LolCDE releases them out of the IM in an ATP-dependent manner and subsequently a water-soluble complex is formed with the periplasmic chaperone LolA in a 1:1 ratio, and finally lipoproteins are transferred from LolA to LolB, which itself is a lipoprotein (Taniguchi *et al.*, 2005). The homologous proteins composing the lipoprotein biogenesis machinery and Lol system in *P. aeruginosa* are indicated in Table 1. In Escherichia coli, lipoproteins are sorted to the OM by default, unless they have an aspartate residue at position 2 after the cysteine (Asp+2), which acts as the IM retention signal (Masuda *et al.*, 2002; Seydel *et al.*, 1999; Yamaguchi *et al.*, 1988). However, the *E. coli* rule that lipoproteins with Asp+2 stay in the IM, while all lipoproteins with another residue at position 2 are sorted to the OM, is applicable only in the family Enterobacteriaceae (Lewenza *et al.*, 2008; Narita & Tokuda, 2007).

### How many lipoprotein genes are there in *P. aeruginosa* PAO1?

Lewenza *et al.* (2005) predicted 185 lipoprotein-coding genes in *P. aeruginosa* PAO1, which represent 3.3% of the genome. This prediction was made by looking for the presence of a lipoprotein signal peptide containing a lipobox. Analysis of the 185 predicted lipoproteins with LIPOP (Juncker *et al.*, 2003) and LIPO (Berven *et al.*, 2006) together with a manual inspection of all signal peptides predicted by these authors led us to eliminate some of the putative lipoproteins (10 in total). This is the case for PA1228 (putative signal MVTHFLSESGQAAC), PA1373 (MSKRIVVTGMAVSPGLC), PA1864 (MFKIRQRNLQ-LILDAA), PA2137 (MNRRLPGLTILALC) and PA5382 (MALSRQGLIFCAVARC), which have a very short signal lacking a clear h-region composed of hydrophobic amino acids; and PA1465 (MEVVALALAAACLGLAAC), PA3894 (MVGLSVQFLVVFSAISGC) and PA4371 (MSLP-SPSMPACCLTALLGGC) because the positively charged n-region of the signal was missing. Furthermore, for some of these dubious lipoproteins a cytoplasmic localization can be predicted, as for PA1373 (FabF2 3-oxoacyl carrier protein), PA1864 (TetR regulator), PA2137 (a response regulator), PA5382 (LysR regulator). ArgE (N-acetyloethiopiminate), PA5206 (ArgE, ArgF, ArgL) is also unlikely to be a lipoprotein since it is involved in arginine biosynthesis and the corresponding *E. coli* enzyme is cytoplasmic (Meinnel *et al.*, 1992). This gives us a total of 175 lipoproteins in *P. aeruginosa* PAO1, corresponding to 3.2% of the annotated protein-encoding genes (versus 3.3% as previously proposed by Lewenza *et al.*, 2005).

For two lipoproteins, PA1812 (MPPQTRKTPDLADARAVRSVILLIAGALGCM) and PA1969 (MPLVRRS-QKRGGGMPMLQFSSALVLAALLGC), the predicted N-terminal sequence is very long and therefore unlikely to be a signal peptidase II-dependent signal, but analysis of the DNA sequence suggests that translation begins at the second ATG (or GTG) since it is preceded by a good ribosome-binding site, resulting in a shorter signal (14 and 18 residues, respectively, italicized), which fits perfectly with a canonical lipoprotein signal peptide. The possibility also exists that some lipoprotein genes are not correctly annotated. This could be the case for PA1712 (exsB), which is predicted to be a lipoprotein although the signal lacks an n-region with one or two positively charged amino acids. Looking at the nucleotide sequence upstream of the ATG, one finds a GTG followed by an AGG (Arg), TGC (Cys) and TGG (Trp), meaning that indeed a positively charged

### Table 1. The lipoprotein biogenesis machinery and Lol system in *P. aeruginosa*

<table>
<thead>
<tr>
<th>Lipoprotein biogenesis</th>
<th>PA no.</th>
<th>Lol system</th>
<th>PA no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lgt</td>
<td>PA0341</td>
<td>LolA</td>
<td>PA2614</td>
</tr>
<tr>
<td>LspA (signal peptidase II)</td>
<td>PA4559</td>
<td>LolB</td>
<td>PA4668</td>
</tr>
<tr>
<td>Lnt apolipoprotein acyltransferase</td>
<td>PA3984</td>
<td>LolC</td>
<td>PA2988</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LolD</td>
<td>PA2987</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LolE</td>
<td>PA2986</td>
</tr>
</tbody>
</table>
amino acid could be present. However, this protein was not retained in our list. The list of *P. aeruginosa* PAO1 lipoproteins and their localization, when known, is presented in Supplementary Table S1, available with the online version of this paper.

**How many lipoproteins are retained in the inner membrane?**

The majority of lipoproteins are destined for the OM, but some stay in the IM because of a LolCDE avoidance signal, which in *Enterobacteriaceae* is an Asp residue after the Cys (Asp<sup>−2</sup>). However, this appears not to be the case for *P. aeruginosa* since only four lipoproteins have Asp at position +2: the glycosylase MltA (PA1222), one hypothetical lipoprotein encoded by PA1592, the PpiC peptidyl-prolyl isomerase (PA3262) and the NosL NO reductase (PA3396). This anomaly compared to the *E. coli* situation stimulated Lewenza *et al.* (2008) to search for alternative IM retention signals. For example, MexA of *P. aeruginosa* is an IM lipoprotein that forms part of the MexAB-OprM drug efflux pump and possesses Gly and Lys residues at positions 2 and 3, respectively (Yoneyama *et al.*, 2000). To investigate the lipoprotein sorting signals in *P. aeruginosa*, chimeric lipoproteins consisting of various regions of MexA and OprM (an OM lipoprotein) were created and their membrane localization was determined. This analysis revealed that, in contrast to the *E. coli* rule, specific combinations of the residues at position 3 and 4 determine the final membrane destination of the *P. aeruginosa* lipoproteins (Narita & Tokuda, 2007). For instance, Lys<sup>−3</sup>-Ser<sup>+4</sup> was shown to be a potent IM retention signal (Narita & Tokuda, 2007). By fusing the signal peptide and the first four amino acids of the *P. aeruginosa* OM lipoprotein OmlA to the red fluorescent protein mCherry, Lewenza *et al.* (2008) created a very sensitive method to determine the membrane localization of the chimeric lipoprotein by fluorescence microscopy. By mutagenesis of the residues at positions 2, 3 and 4, they confirmed Lys<sup>−3</sup>-Ser<sup>+4</sup> as a Lol avoidance signal and identified several other combinations of residues that act as an IM retention signal, such as Lys<sup>−2</sup>-Val<sup>+3</sup>-Glu<sup>+4</sup>, Gly<sup>−2</sup>-Gly<sup>+3</sup>-Gly<sup>+4</sup>, Gly<sup>−2</sup>-Asp<sup>+3</sup>-Asp<sup>+4</sup> and Gln<sup>−2</sup>-Gly<sup>−3</sup>-Ser<sup>+4</sup> (Lewenza *et al.*, 2008). These IM retention signals are found in approximately 5% of the *P. aeruginosa* lipoproteins, but given the lack of conservation it is unclear how these sequences act as Lol avoidance signals. It was speculated that the secondary structure formed by these residues might play an important role and that the periplasmic loops of LolC and/or LolE could be involved in the recognition of the sorting signal at positions 3 and 4 of *P. aeruginosa* lipoproteins (Tokuda, 2009). In their analysis, Lewenza *et al.* (2008) discovered new retention signals, bringing the number of IM lipoproteins to 13. Assuming that some lipoproteins, such as membrane fusion protein (MFP) components of efflux pumps, and some periplasmic enzymes are IM localized, we can predict more retention signals based on the four amino acids after the Cys residue (Table 2). Seven MFPs are predicted to have a lipoprotein signal peptide with an IM retention signal: PA0156 (TriA), PA0157 (TriB), PA2019, PA2493 (MexE), PA3523, PA3677 and PA4599 (MexC) (Narita & Tokuda, 2007). During the last few years, other IM lipoproteins have been described and their localization confirmed as shown in Table 2. MliC is a periplasmic lysozyme inhibitor, first described by Callewaert *et al.* (2008) in *E. coli*, and its homologue in *P. aeruginosa* (PA0867) has been confirmed to be an IM lipoprotein (Lewenza *et al.*, 2008). The same authors confirmed the IM localization of PscJ (PA1723), a type III secretion protein, the glycosylase MltD (PA1812), the GntK (PA2321) gluconate kinase, the metalloproteinase IcmP (PA4370), the type IV pili biogenesis lipoprotein PilP (PA5041), and two hypothetical lipoproteins, encoded by PA3677 and PA5414, respectively. The following lipoproteins are also likely to be IM proteins: AbpE (PA2993), which has been described to be involved in the biosynthesis of thiamine in *Salmonella* (Beck & Downs, 1999), NlpD (PA3623), an accessory lipoprotein of the OM protein assembly machinery (Uehara *et al.*, 2009), the PA4065 FtsX protein involved in cell division (Reddy, 2007), the PA4367 BifA cyclic-di-GMP phosphodiesterase (Kuchma *et al.*, 2007) and the transglycosylase MltB1 (Reid *et al.*, 2006). In total, three transglycosylases have an IM lipoprotein anchor: MltA (PA1222), MltD (PA1812) and MltB1 (PA4444).

A total of 24 lipoproteins are predicted be inserted in the IM, and we propose to include 16 others in this list (Table 2), which would bring the percentage of IM lipoproteins to 25%. Since this analysis was done taking into account the known or predicted biological function, hypothetical proteins, the function of which could not be predicted, have not been included, so the percentage of IM lipoproteins might still be underestimated. When comparing the sequences of the four residues after the cysteine, it is striking that many IM lipoproteins have a G at the +1 position, often with a negatively charged amino acid (D or E). This is the case for TriA and TriB, which are MFPs involved in the efflux of triclosan, and are associated with TriC antiporter and OpmH OM porin (Mima *et al.*, 2007).

**Proposed inner-membrane lipoproteins and their predicted function**

Two peptidyl-prolyl isomerases are predicted to be IM lipoproteins, PA0699 (PpiC) and PA3262, the latter having a D residue at the position +1 after the cysteine. Some enzymes are involved in carbohydrate metabolism, such as 1-sorbitose dehydrogenase (PA2414), TreA trehalase (PA2416) and GPCD (PA4792), a glycerophosphodiester phosphodiesterase. NirF (PA0516) is involved in the biosynthesis of haem D1, and PA0541 and PA5328 encode c-type cytochromes. QuiP is an acylase encoded by the PA1032 gene, which has been shown to modulate quorum sensing in *P. aeruginosa* by de-acylation of the signal.
One predicted IM lipoprotein is a thioredoxin (PA0953), one is involved in thiamine biosynthesis (PA2993), like the already mentioned AbpE, and one is a putative amidase (PA5485). Two lipoproteins are involved in type VI secretion (PA1048 and PA2364), one is a predicted peptidase (PA4016), and the last one is a sensor protein (PA3044).

**Families of outer-membrane lipoproteins in P. aeruginosa**

Although the function of many lipoproteins is unknown and cannot be predicted at the moment, the characteristics of some of them have been described and these can be grouped into different families.

---

**Table 2. Known and predicted inner-membrane lipoproteins in P. aeruginosa PAO1**

The four amino acid residues after the cysteine are indicated. D residues after the cysteine serving as retention signal in enterobacteria are shown in bold.

<table>
<thead>
<tr>
<th>IM lpp</th>
<th>Protein</th>
<th>Function</th>
<th>Cys 4+</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA0156</td>
<td>TriA</td>
<td>MFP</td>
<td>GAEP</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA0157</td>
<td>TriB</td>
<td>MFP</td>
<td>GDEP</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA0425</td>
<td>MexA</td>
<td>MFP</td>
<td>GKSE</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA0867</td>
<td>MilC</td>
<td>Lysozyme inhibitor</td>
<td>GGSD</td>
<td>Lewenza et al. (2008), Callewaert et al. (2008)</td>
</tr>
<tr>
<td>PA1222</td>
<td>MltA</td>
<td>Glycosylase</td>
<td>DDGK</td>
<td></td>
</tr>
<tr>
<td>PA1398</td>
<td>Hypothetical</td>
<td>Unknown</td>
<td>SKSA</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA1592</td>
<td>Hypothetical</td>
<td>Unknown</td>
<td>DKKE</td>
<td></td>
</tr>
<tr>
<td>PA1723</td>
<td>PscJ</td>
<td>Type III secretion</td>
<td>KVEL</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA1812</td>
<td>MltD</td>
<td>Glycosylase</td>
<td>QGSG</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA2321</td>
<td>GntK</td>
<td>Glucanate kinase</td>
<td>GKSV</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA2493</td>
<td>MexE</td>
<td>MFP</td>
<td>GKAP</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA2993</td>
<td>AbpE</td>
<td>Thiamine biosynthesis</td>
<td>GDTL</td>
<td>Beck &amp; Downs (1999)</td>
</tr>
<tr>
<td>PA3262</td>
<td>PpiC</td>
<td>Peptidyl-prolyl isomerase</td>
<td>DSQT</td>
<td></td>
</tr>
<tr>
<td>PA3396</td>
<td>NosL</td>
<td>NO reductase</td>
<td>DASR</td>
<td></td>
</tr>
<tr>
<td>PA3523</td>
<td>Hypothetical</td>
<td>MFP</td>
<td>GESA</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA3623</td>
<td>NlpD</td>
<td>OM biogenesis</td>
<td>SSSP</td>
<td>Uehara et al. (2009)</td>
</tr>
<tr>
<td>PA3677</td>
<td>Hypothetical</td>
<td>MFP</td>
<td>GNGT</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA4065</td>
<td>FtsX</td>
<td>Cell division</td>
<td>LLLA</td>
<td>Reddy (2007)</td>
</tr>
<tr>
<td>PA4367</td>
<td>BisA</td>
<td>Phosphodiesterase</td>
<td>AQIV</td>
<td>Kuchma et al. (2007)</td>
</tr>
<tr>
<td>PA4370</td>
<td>IcmP</td>
<td>Metalloproteinase</td>
<td>GDDK</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA4444</td>
<td>MltB1</td>
<td>Transglycosylase</td>
<td>SSEP</td>
<td>Reid et al. (2006)</td>
</tr>
<tr>
<td>PA4599</td>
<td>MexC</td>
<td>MFP</td>
<td>GPAE</td>
<td>Narita &amp; Tokuda (2007)</td>
</tr>
<tr>
<td>PA5041</td>
<td>PilP</td>
<td>Type IV pili biogenesis</td>
<td>GGGS</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>PA5414</td>
<td>Hypothetical</td>
<td>Unknown</td>
<td>GGGD</td>
<td>Lewenza et al. (2008)</td>
</tr>
<tr>
<td>Proposed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA0516</td>
<td>NirF</td>
<td>Haem D1 biosynthesis</td>
<td>SQQP</td>
<td></td>
</tr>
<tr>
<td>PA0541</td>
<td>Hypothetical</td>
<td>Cytochrome c</td>
<td>HRMD</td>
<td></td>
</tr>
<tr>
<td>PA0699</td>
<td>PpiC</td>
<td>Peptidyl-prolyl isomerase</td>
<td>GQRE</td>
<td></td>
</tr>
<tr>
<td>PA0953</td>
<td>Hypothetical</td>
<td>Thioredoxin</td>
<td>SADY</td>
<td></td>
</tr>
<tr>
<td>PA1032</td>
<td>QuiP</td>
<td>Acylase</td>
<td>QSWL</td>
<td></td>
</tr>
<tr>
<td>PA1048</td>
<td>Hypothetical</td>
<td>Type VI secretion</td>
<td>SYFQ</td>
<td></td>
</tr>
<tr>
<td>PA2019</td>
<td>Hypothetical</td>
<td>MFP</td>
<td>EEAA</td>
<td></td>
</tr>
<tr>
<td>PA2364</td>
<td>Hypothetical</td>
<td>Type VI secretion</td>
<td>SYFF</td>
<td></td>
</tr>
<tr>
<td>PA2414</td>
<td>Hypothetical</td>
<td>l-Sorbose dehydrogenase</td>
<td>NOGS</td>
<td></td>
</tr>
<tr>
<td>PA2416</td>
<td>TreA</td>
<td>Trehalase</td>
<td>SQGP</td>
<td></td>
</tr>
<tr>
<td>PA2993</td>
<td>ThiA</td>
<td>Thiamine biosynthesis</td>
<td>GDTL</td>
<td></td>
</tr>
<tr>
<td>PA3044</td>
<td>Hypothetical</td>
<td>Sensor</td>
<td>MAGY</td>
<td></td>
</tr>
<tr>
<td>PA4016</td>
<td>Hypothetical</td>
<td>Peptidase</td>
<td>GEGE</td>
<td></td>
</tr>
<tr>
<td>PA4792</td>
<td>GDPD</td>
<td>Glycerophosphodiester phosphodiesterase</td>
<td>VEAR</td>
<td></td>
</tr>
<tr>
<td>PA5328</td>
<td>Hypothetical</td>
<td>Cytochrome c</td>
<td>GEST</td>
<td></td>
</tr>
<tr>
<td>PA5485</td>
<td>Hypothetical</td>
<td>Amidase</td>
<td>SSGP</td>
<td></td>
</tr>
</tbody>
</table>

molecules N-acylhomoserine lactones (Huang et al., 2006). One predicted IM lipoprotein is a thioredoxin (PA0953), one is involved in thiamine biosynthesis (PA2993), like the already mentioned AbpE, and one is a putative amidase (PA5485). Two lipoproteins are involved in type VI secretion (PA1048 and PA2364), one is a predicted peptidase (PA4016), and the last one is a sensor protein (PA3044).
Lipoproteins involved in outer-membrane biogenesis

Lipoproteins have been found to be part of various OM assembly machineries. The Bam complex, which is responsible for the assembly and insertion of proteins into the OM, contains four different lipoproteins in *E. coli*, namely BamB (YfgL), BamC (NlpB), BamD (YfO) and BamE (SmpA) (Knowles et al., 2009; Ruiz et al., 2006). In the *Pseudomonas* genome database, the lipoproteins homologous to BamB (PA3800), BamD (PA4545) and BamE (PA4765) can be identified. In *E. coli* the Lpt system is involved in the lipopolysaccharide trafficking from the IM to the OM, where the essential lipoprotein LptE (RlpB) forms a complex with the integral OM component LptD (Imp) (Narita & Tokuda, 2009; Tokuda, 2009). In the *P. aeruginosa* genome, only the LptE (PA3691) OM lipoprotein was found.

Peptidoglycan-binding lipoproteins

The small lipoprotein OprI (9 kDa) is the homologue of the *E. coli* Braun’s lipoprotein and is one of the most highly produced proteins in *P. aeruginosa* (Braun, 1975; Cornelis et al., 1989; Duchene et al., 1989). In *E. coli*, the Braun’s lipoprotein exists in a free form or is covalently attached to peptidoglycan (PG) via the C-terminal lysine (Braun, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975). Although OprI also has a lysine as the last amino acid, it does not seem to bind covalently to PG (Mizuno, 1975).

OprL is the second most abundant lipoprotein in *P. aeruginosa* and is the equivalent of the *E. coli* peptidoglycan-associated lipoprotein (PAL), which associates non-covalently with peptidoglycan (Bouveret et al., 1995; Lim et al., 1997). OprL is important for the integrity of the cell since an oprl-null mutant formed elongated cells which were sensitive to detergents or EDTA (Rodriguez-Herva & Ramos, 1996; Rodriguez-Herva et al., 1996). OprL is part of the multiprotein Tol-OprL system, which spans the two membranes and is important for the uptake of solutes, as demonstrated in *Pseudomonas putida* (Llamas et al., 2003). It has also recently been shown that OprL contributes to protect cells against oxidative stress (Panmanee et al., 2008). OprL belongs to the OmpA family, which comprises OM lipoproteins with a C-terminal peptidoglycan-binding domain; these are identified via their consensus sequence N-x2-L-x3-RA-x2-V-x3-L (Koebnik, 1995). By associating non-covalently with the PG, these lipoproteins play a role in maintaining the structural integrity of the cell. A first clue to the structural aspects of this interaction came from the NMR structure of the periplasmic domain of *Haemophilus influenzae* Pal (peptidoglycan-associated lipoprotein) bound to a PG precursor (Parsons et al., 2006). Other OM lipoproteins having the PG-binding motif are PA0833 (involved in type VI secretion), PA1041, PA1119 (YfB), PA2900 and PA3692 (also designated LptF, but not related to the LPS biogenesis proteins mentioned above). PA1119 (YfB) is an OM lipoprotein recently described to mediate the cyclic-di-GMP-dependent conversion of *P. aeruginosa* to a small-colony variant phenotype (Malone et al., 2010). The last protein, LptF, has been found to be regulated by the alternative sigma factor AlgU, as will be mentioned later (Damron et al., 2009). It is also worth mentioning that this PG-binding domain is found in non-lipoprotein OM porins such as OprF (Rawling et al., 1998).

Efflux porins

In *P. aeruginosa*, resistance to antibiotics or biocide compounds is often mediated by different efflux systems, including the most represented RND (Resistance-Nodulation-Division) systems (Schweizer, 2003). These efflux systems are generally composed of three proteins, two being anchored in the IM (an efflux antiporter and a membrane fusion protein), and one serving as OM efflux porin (Schweizer, 2003). The first described efflux porin was OprM, which is part of the general efflux system MexAB-OprM (Li et al., 1995; Schweizer, 2003). Other efflux porins were described later and form what is called the Opm family of OM porins, comprising 18 proteins (Hancock & Brinkman, 2002). Out of these, 11 are predicted to be lipoproteins (Table 3) and eight of them are involved in the efflux of molecules, including antibiotics. This is the case for OprM, OprN, OprP, OpmG, OpmB and OpmE (Kohler et al., 1997; Li et al., 1995; Linares et al., 2005; Mima et al., 2005, 2009; Murata et al., 2002; Poole et al., 1996). Other Opm family OM proteins include OpmF, OpmH, OpmI, OpmK, OpmL, AprF and CzcC, but these are not lipoproteins and only one, OpmH, has been described to participate in efflux of an antiseptic, triclosan (Mima et al., 2007). It seems therefore that those Opm family OM porins involved in the efflux of antibiotics are for the most part lipoproteins, while others are not, as shown by the phylogenetic clustering presented in Fig. 1.

Conservation of lipoproteins in the genus *Pseudomonas* and some evolutionary aspects in *P. aeruginosa*

Among the 175 lipoprotein-encoding genes identified in PAO1, 62 have an orthologue in all the sequenced *Pseudomonas* genomes, while 27 are common to and found only in all four sequenced *P. aeruginosa* genomes (PAO1, PA14, PA7 and LES) (Fig. 2). According to the membrane localization (OM, identified IM or proposed IM), the numbers of orthologues do not differ significantly (Table 4), showing the same conservation of the OM and IM lipoproteins in the genus *Pseudomonas*. Fig. 2 shows clearly two major peaks, one corresponding to the genes found in all four *P. aeruginosa*, and the peak corresponding
to the conserved lipoprotein genes, which are present in all analysed representatives, representing the ‘core’ lipoprotein genome.

The already mentioned lipoprotein gene oprI has been previously described to be present in all representatives of the genus Pseudomonas sensu stricto since the oprI coding sequence could be PCR-amplified from different species of Pseudomonas, but not from other species (De Vos et al., 1997). The oprI gene has also been used to study the phylogeny of pseudomonads (De Vos et al., 1998). It should be noted that the oprI gene is not annotated in the P. aeruginosa PA14 and Pseudomonas fluorescens Pf5 genomes although a BLAST search reveals the presence of the gene in these two genomes, which means that oprI is also a lipoprotein gene of the core genome. Another lipoprotein gene of diagnostic value is oprL, which has been used for the specific detection of P. aeruginosa from clinical specimens (De Vos et al., 1997; Pirnay et al., 2003; Rao et al., 2009). The specificity of PCR detection of oprL relies on the fact that the N-terminal portion of the lipoprotein is more divergent, since oprL can be amplified from different Pseudomonas species using primers targeting the more conserved C-terminal part of the protein (unpublished results). An interesting case is the PA1382 gene encoding a type II secretion OM secretin, since this gene is not present in the other P. aeruginosa genomes, but is found in P. fluorescens SBW25 (Pflu_4070), P. syringae 1448a (PSPPH_3054), and P. syringae B728a (Psyr_3141).

The G+C content at the third position of codons can be used to detect a recent lateral transfer from an organism that has a different GC3 content (Muto & Osawa, 1987), especially in the genus Pseudomonas (Bodilis & Barray, 2006). The GC3 content of the lipoprotein-encoding genes is high (87.1 ± 5.6 mol%) and does not differ significantly according to the membrane localization (OM, identified IM or proposed IM), showing a weak impact of recent

![Phylogenetic tree of the OprM family](image)

**Table 3.** The lipoprotein outer-membrane proteins belonging to the OprM family

<table>
<thead>
<tr>
<th>Protein</th>
<th>PA gene no.</th>
<th>Function</th>
<th>Lipoprotein signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>OprM</td>
<td>PA0427</td>
<td>Antibiotic efflux</td>
<td>MKRSFLSLAVAAVVLGSC</td>
</tr>
<tr>
<td>OprN</td>
<td>PA2495</td>
<td>Antibiotic efflux</td>
<td>MIHAQISRSGLASALGFSSLALSLAC</td>
</tr>
<tr>
<td>OprP</td>
<td>PA4597</td>
<td>Antibiotic efflux</td>
<td>MKRPAFGVSALLIALTLGAC</td>
</tr>
<tr>
<td>OprD</td>
<td>PA4208</td>
<td>Quorum sensing</td>
<td>MKRSYPNLSRLALAVGTLGAAC</td>
</tr>
<tr>
<td>OprQ</td>
<td>PA2391</td>
<td>Pyoverdine recycling</td>
<td>MSMKNLSISAICLGLGAC</td>
</tr>
<tr>
<td>OprG</td>
<td>PA5158</td>
<td>Aminoglycoside resistance</td>
<td>MPFPILLHPWPQRLASAILAAGC</td>
</tr>
<tr>
<td>OprA</td>
<td>PA2837</td>
<td>Unknown</td>
<td>MKGTPPLLHASLALGAC</td>
</tr>
<tr>
<td>OprB</td>
<td>PA2525</td>
<td>Antibiotic efflux</td>
<td>MKHTPSLLALVAALGCG</td>
</tr>
<tr>
<td>OprE</td>
<td>PA3521</td>
<td>Antibiotic efflux</td>
<td>MKGPYRSLSALILGCG</td>
</tr>
<tr>
<td>OprJ</td>
<td>PA1238</td>
<td>Unknown</td>
<td>MPLASHLRCVALALGISTALGCG</td>
</tr>
</tbody>
</table>

![Fig. 1. Phylogenetic tree of the OprM family](image)
lateral transfers in the evolutionary history of the lipoprotein-encoding genes in \textit{P. aeruginosa} (Supplementary Table S1).

It has previously been shown that the extracellular and OM protein encoding genes in \textit{P. aeruginosa} evolve faster than cytoplasmic protein encoding genes (Dotsch et al., 2010; Julenius & Pedersen, 2006). As expected, we found average values of the non-synonymous versus synonymous mutations ratio (\(d_N/d_S\)) using the modified Nei–Gojobori method (Tamura et al., 2007), much more important for the lipoprotein genes (0.21) than for the ribosomal genes (0.06) (Bodilis et al., 2009). However, we did not find significant differences either between the \(d_N/d_S\) ratios or between the overall mean variabilities \((P_i)\) of the OM and IM lipoprotein genes (Table 4). Because periplasmic-facing proteins are probably under less selective pressure than surface-exposed proteins, we can assume that most of the lipoproteins in \textit{P. aeruginosa} are exposed to the periplasm whatever their membrane localization, as in \textit{E. coli} (Narita et al., 2004).

We also compared the topologies of the phylogenetic trees (neighbour-joining tree with Kimura two-parameter correction) (Tamura et al., 2007) built from each cluster containing the four \textit{P. aeruginosa} orthologues (Table 4). An unrooted tree built from four \textit{P. aeruginosa} genes has three possible topologies. The topology that groups PAO1 and LESB58 together is the most represented (72.5 \%) among lipoprotein genes, with no significant difference according to the membrane localization (Table 4 and Supplementary Figure S1).

Interestingly, this topology is the right one, i.e. the topology of the organism’s phylogeny, PA7 being an

---

**Table 4.** Orthologues of the PAO1 lipoprotein genes

<table>
<thead>
<tr>
<th>Cellular localization</th>
<th>No. of genes in PAO1*</th>
<th>No. of orthologues in the genus \textit{Pseudomonas}\dagger</th>
<th>Length of sequences (bp)‡</th>
<th>GC3 content (mol%)§</th>
<th>Overall mean variability\dagger</th>
<th>(d_N/d_S)§</th>
<th>Right phylogenetic topology (%)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM</td>
<td>34</td>
<td>12.88 (± 4.60)a</td>
<td>989.65 (± 470.01)a</td>
<td>87.29 (± 4.60)a</td>
<td>0.037 (± 0.034)a</td>
<td>0.154</td>
<td>70.6a</td>
</tr>
<tr>
<td>IM</td>
<td>33</td>
<td>11.09 (± 5.52)ab</td>
<td>1044.91 (± 527.42)b</td>
<td>87.51 (± 5.08)a</td>
<td>0.038 (± 0.024)a</td>
<td>0.178</td>
<td>63.6a</td>
</tr>
<tr>
<td>Not identified</td>
<td>93</td>
<td>9.66 (± 5.68)b</td>
<td>645.78 (± 383.54)b</td>
<td>86.56 (± 6.71)a</td>
<td>0.038 (± 0.024)a</td>
<td>0.248</td>
<td>76.3a</td>
</tr>
</tbody>
</table>

*The 15 genes that did not contain an orthologue in each \textit{P. aeruginosa} genome (PA7, PA14, LESB58 and PAO1) or for which the orthologues in \textit{P. aeruginosa} were identical (no nucleotide sequence variation) were excluded from the study.

†Means ± SD of the number of orthologues in the genus \textit{Pseudomonas}. Values with a different letter (a or b) within the column are significantly different at \(P<0.05\) (Mann–Whitney test).

‡Means ± SD of the values in \textit{P. aeruginosa} (PA7, PA14, LESB58 and PAO1). Values with a different letter (a or b) within the same column are significantly different at \(P<0.05\) (Mann–Whitney test).

§Proportion of neighbour-joining trees (with Kimura two-parameter correction) that were congruent with the organism phylogenetic tree (Topology 1 in Supplementary Fig. S1). Values with a different letter (a or b) within the column are significantly different at \(P<0.05\) (Z test).
outlier (Roy et al., 2010). This observation confirms that the lipoprotein-encoding genes are probably little impacted by lateral transfers. It is worth mentioning that a phylogeny with an atypical topology does not necessarily mean that the gene has undergone lateral transfer since this could also be explained by a variation of the evolution rate, i.e. a high $d_S/d_S$ ratio (Bodilis & Barray, 2006).

**Regulation of lipoprotein genes in P. aeruginosa**

Given the large variety of different lipoproteins, it is not surprising that they have different functions and are differently regulated. Some lipoprotein genes are considered as constitutively expressed, such as oprI and oprL, while others are induced by different factors. Therefore different transcriptome results from the literature were analysed. The factor that has the greatest influence on the expression of lipoprotein genes is exposure to human airway epithelial cells (Chugani & Greenberg, 2007; Frisk et al., 2004). Nineteen lipoprotein genes were found to be upregulated upon contact with epithelial cells, while 10 were downregulated. Among the upregulated lipoproteins, one finds PA1592, PA0737, PA3691 and PA3692. The same genes, plus PA0062 and PA5526, are also induced as a result of the conversion of P. aeruginosa to mucoidy due to the production of alginate extracellular polysaccharide and are known to be dependent on the AlgU ($\sigma^d$) sigma factor (Firoved et al., 2002, 2004). The Lpf F (lipoxin F) gene has recently been described to be also dependent on the AlgU alternative sigma factor, and LpfF contributes to survival in CF lungs (Damron et al., 2009). Another lipoprotein gene upregulated by exposure to epithelial cells and by conversion to mucoidy is osmE (PA4876), which is also induced by a wide range of other environmental changes, such as exposure to $\delta$-cycloserine, and osmotic shock; it is also regulated by the sensor PhoQ, and is quorum sensing induced (Aspedon et al., 2006; Gooderham et al., 2009; Schuster et al., 2003; Wood et al., 2006). $\delta$-Cycloserine exposure results in higher expression of several lipoprotein genes: a type VI secretion protein gene (PA1048), the alginate secretion lipoprotein gene algK (PA3543), PA3962, mexC (PA4599) and PA5107–PA5108 (Keiski et al., 2010; Wood et al., 2006).

Some RND efflux OM lipoprotein genes (oprM, oprN, oprD, oprG) are also induced or repressed by different conditions. The oprM (PA0427) gene is downregulated by exposure to epithelial cells, while oprQ (PA2391) and oprN (PA2495) are induced. The OprM protein is the OM component of the PvdRT-OpmQ efflux pump, which has recently been shown to be involved in the recycling and re-excretion of the siderophore pyoverdine in the periplasm after release of iron (Imperi et al., 2009) and, as expected, these genes are iron-regulated (Ravel & Cornelis, 2003). OprM is the OM protein of the MexGHI-OpmD efflux pump involved in the quorum sensing network of P. aeruginosa, and the oprM gene is induced by quorum sensing, belongs to the stationary-phase sigma factor RpoS regulon, and is SoxR-dependent and upregulated by the antibiotic azithromycin, known to interfere with the quorum sensing system in P. aeruginosa (Aendekerk et al., 2005; Dietrich et al., 2006; Nalca et al., 2006; Narita et al., 2004; Schuster et al., 2004). The oprM gene is induced upon exposure to copper, suggesting that it could be involved in the efflux of this metal (Teitzel et al., 2006). Interestingly, a number of lipoprotein genes are upregulated by the recently described PpyR (PA2663) regulator, including mexE and oprN, which are also under the control of MexT (Attila et al., 2008; Tian et al., 2009). The MexT regulon also includes PA2759, encoding a hypothetical lipoprotein. As already mentioned, some lipoprotein genes are iron regulated since they are repressed under condition of iron sufficiency by the ferric uptake regulator Fur; these include the already mentioned oprM gene, omL (PA4765), icmP (PA4370) and PA4372 (Ochsner et al., 2002; van Oevelen et al., 2008).

**Conclusion**

P. aeruginosa is a highly adaptable bacterium with a rather large genome (>6 Mb) and previous in silico predictions and experimental data suggested that in PA01 185 genes encoded lipoproteins, many of them of unknown function. The present analysis suggests that the number of lipoproteins should be reduced to 175, while more lipoproteins can be predicted to be localized in the IM (23 %). OM porins from RND systems involved in the efflux of antibiotics are lipoproteins, while others belonging to the same family are not, as shown by the present phylogenetic analysis. The reason for a supplementary lipoprotein anchorage for these efflux porins rather than via only transmembrane $\beta$-sheets domains remains to be elucidated. The present analysis of the conservation of lipoprotein genes in the genomes of different pseudomonads showed 62 genes to be conserved, forming the core of lipoprotein genes, while no evidence for large horizontal gene transfer could be found. Finally, lipoprotein genes are mostly regulated by exposure to epithelial cells, conversion to mucoidy, and treatment with membrane-damaging antibiotics such as $\delta$-cycloserine.

**Acknowledgements**

Kim Remans had a fellowship from IWT Vlaanderen.

**References**


