Conservation of the PTEN catalytic motif in the bacterial undecaprenyl pyrophosphate phosphatase, BacA/UppP

Justin S. Bickford and Harry S. Nick

Department of Neuroscience, University of Florida, Gainesville, FL 32610, USA

Isoprenoid lipid carriers are essential in protein glycosylation and bacterial cell envelope biosynthesis. The enzymes involved in their metabolism (synthases, kinases and phosphatases) are therefore critical to cell viability. In this review, we focus on two broad groups of isoprenoid pyrophosphate phosphatases. One group, containing phosphatidylic acid phosphatase motifs, includes the eukaryotic dolichyl pyrophosphate phosphatases and proposed recycling bacterial undecaprenol pyrophosphate phosphatases, PgpB, YbjB and YeIu/LpxT. The second group comprises the bacterial undecaprenol pyrophosphate phosphatase, BacA/UppP, responsible for initial formation of undecaprenol phosphate, which we predict contains a tyrosine phosphate phosphatase motif resembling that of the tumour suppressor, phosphatase and tensin homologue (PTEN). Based on protein sequence alignments across species and 2D structure predictions, we propose catalytic and lipid recognition motifs unique to BacA/UppP enzymes. The verification of our proposed active-site residues would provide new strategies for the development of substrate-specific inhibitors which mimic both the lipid and pyrophosphate moieties, leading to the development of novel antimicrobial agents.

Introduction

Polyisoprenyl monophosphates of different lengths and saturation (dolichol or undecaprenol derivatives) are utilized as pyrophosphate lipid carriers (Jones et al., 2009; Surmacz & Swiezewska, 2011; Swiezewska & Danikiewicz, 2005) in N- and O-linked protein glycosylation pathways in all three domains of life, i.e. eukaryotes, bacteria and archaea (Calo et al., 2010; Helenius & Aebi, 2004; Jones et al., 2009; Larkin & Imperiali, 2011; Nothaft & Szymanski, 2010; Schwarz & Aebi, 2011; Weerapana & Imperiali, 2006). Monophosphate lipid carriers are also used in the synthesis of bacterial cell envelope polymers such as peptidoglycan, lipopolysaccharides and teichoic acids (Bouhss et al., 2008; Cantagrel & Lefeber, 2011; de Krijff et al., 2008; Lovering et al., 2012; Reusch & Salton, 1984; Swowoda et al., 2010). As uniquely conserved components of the bacterial cell envelope synthesis and glycosylation pathways, polyisoprenoid lipid carriers are thus essential for life (Hartley & Imperiali, 2012).

Polyisoprenols are categorized into two broad classes based on chain length and saturation of the α-isoprene. The first class includes the undecaprenyl (C55) lipid carriers, which are unsaturated and consist of both trans and cis units in bacteria, while plant undecaprenols vary in length with some in the all-trans configuration (Hartley & Imperiali, 2012; Surmacz & Swiezewska, 2011; Swiezewska & Danikiewicz, 2005). The second class, dolichols, are lipid carriers of varying lengths found in animals, plants and archaea. Dolichols are distinguished by greater length and the presence of a saturated α-isoprene unit (proximal to the phosphoryl or hydroxyl moiety) (Adair et al., 1984; Gough & Hemming, 1970; Skorupinska-Tudek et al., 2008). Each of these lipid carriers must first exist in the monophosphate state to carry out their respective roles. As part of the transfer of the cargo to their lipid carriers, the polyisoprenyl lipid monophosphate is converted to a pyrophosphate containing their respective cargos (Jones et al., 2009); undecaprenyl pyrophosphate carries MurNAc-GlcNAc-peptide complexes for bacterial cell envelope biosynthesis (Figs 1b and S1A, available in Microbiology Online) (Bouhss et al., 2008; Hitchen & Dell, 2006; Lovering et al., 2012), undecaprenyl pyrophosphate binds a heptasaccharide for N-linked glycosylation in bacteria as identified in Campylobacter jejuni (Fig. S1B).
Fig. 1. Lipid carriers in eukaryotic protein glycosylation and bacterial cell envelope biosynthesis. (a) Dol-P is used as the lipid carrier for eukaryotic protein glycosylation. N-acetylglucosamine and mannose are added to Dol-P in the cytoplasm, which is then believed to be flipped to the ER lumen. In the ER lumen, additional mannose and glucose are added to form the final oligosaccharide prior to the transfer to an arginine residue of a protein by an oligosaccharyltransferase. The resulting Dol-PP is converted back into Dol-P by the phosphatase DOLPP1 before being flipped back to the cytosol to begin the cycle anew. (b) Undecaprenyl phosphate is similarly utilized in bacterial cell envelope biosynthesis. In the cytoplasm, the phosphatase BacA/UppP converts de novo synthesized undecaprenyl pyrophosphate (C55-PP) to the monophosphate (C55-P) which is used as the lipid carrier for assembly of peptidoglycan. After addition of the first two sugars, the molecule is flipped to the periplasm for final peptidoglycan assembly. Finally, the peptidoglycan is cleaved from undecaprenyl pyrophosphate, which is recycled to undecaprenyl monophosphate by YbjG, PgpB or YeiU prior to being flipped back to the cytoplasm to repeat the process. PBPs, penicillin-binding proteins. Note: panels are not drawn to scale.
(Hitchen & Dell, 2006; Kumar & Balaji, 2011; Nothaft & Szymanski, 2010) and dolichyl pyrophosphate (Dol-PP), through a GlcNAc linkage, is complexed with a high mannose tree involved in eukaryotic protein glycosylation (Fig. S1C) (Cantagrel & Lefeber, 2011; Helenius & Aebl, 2004; Larkin & Imperiali, 2011; Schwarz & Aebl, 2011; Welti, 2013). This review will focus on the enzymes in eukaryotes and prokaryotes that generate the monophosphorylated lipid carrier either through phosphorylation of the alcohol or dephosphorylation of the pyrophosphate polyisoprenoid moiety, with the latter being coupled to de novo synthesis of the lipid carrier or through a recycling pathway. The high level of conservation in these enzymes will be highlighted with cross-species alignments, derivation of catalytic consensus sequences and 2D structural predictions forming the basis for our structure/function hypotheses.

**Biosynthesis of the eukaryotic dolichyl lipid carriers**

Dolichol, a polyisoprenoid alcohol, was first isolated from pig liver, and is one of the largest lipids in eukaryotes and archaea (Burgos et al., 1963; Pennock et al., 1960). Named for the Greek word for long, dolichos, it can contain 14–24 isoprene units depending on the organism. In human tissues, both dolichol and dolichyl monophosphate (Dol-P) increase significantly with age analogous to another tissues, both dolichol and dolichyl monophosphate (Dol-isoprene units depending on the organism. In human tissues, both dolichol and dolichyl monophosphate (Dol-P) increase significantly with age analogous to another polyisoprenoid alcohol, cholesterol (Carroll et al., 1992). Neutral lipids such as dolichol originate from the mevalonate pathway where condensation of three molecules of acetyl-CoA leads to the production of a key intermediate, mevalonate, by 3-hydroxy-3-methylglutaryl (HMG)-CoA reductase (Ericsson et al., 1993; Goldstein & Brown, 1990). Several subsequent reactions lead to the synthesis of isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). Condensation of IPP units by prenyltransferases first generates geranyl pyrophosphate (C10) followed by farnesyl pyrophosphate (C15), which is the central backbone substrate for ubiquinone, cholesterol and dolichol synthesis through the successive addition of IPP units. In eukaryotes, the isoprenoid lipids, IPP and geranylgeranyl pyrophosphate (C20), also serve as the lipid moieties transferred to proteins through a terminal -SH of a cysteine in the CaaX motif in a process termed isoprenylation. This modification helps tether proteins to membranes due to the hydrophobic nature of the C15 or C20 moieties (Sebit, 2005). Most recently, studies have demonstrated that proteins from bacterial pathogens (Legionella pneumophila) are also prenylated, although this occurs by host cell farnesyltransferase and class I geranylgeranyltransferase (Ivanov et al., 2010; Price et al., 2010).

Dolichol is synthesized by sequential condensation reactions through the cis addition of IPP to trans-IPP. Unlike other polyisoprenoids, the final steps in the synthesis of dolichol involves dephosphorylation and saturation of the x-isoprene unit to yield the alcohol (Adair et al., 1984; Ekström et al., 1984; Gough & Hemming, 1970; Skorupinska-Tudek et al., 2008). The synthesis of a polyisoprenoid alcohol is not unique to eukaryotes in that Gram-positive bacteria also accumulate undecaprenol (Van Horn & Sanders, 2012). Various models have been proposed for these final steps in dolichol biosynthesis (Chojnacki & Dallner, 1988). For example, Heller et al. (1992) proposed that once the appropriate isoprenoid chain length has been attained, the x-unsaturated pyrophosphate derivative is dephosphorylated to the monophosphate, followed by x-saturation to yield Dol-P. However, based on [3H]mevalonate labelling studies in membrane fractions, Ekström et al. (1987) proposed that terminal isoprene condensation and x-saturation occur cooperatively. A third plausible model is based on dehydrodolichol reductase activity in microsomal fractions that achieves x-isoprene saturation preceded by a dephosphorylation of dehydro-Dol-PP (Sagami et al., 1993, 1996).

The final steps in dolichol biosynthesis became better understood with the identification of both human and yeast mutants in the dolichol synthesis pathway (Fig. S2). In 2003, two independent groups identified a human cis-prenyltransferase involved in the synthesis of dolichol through gene complementation in Saccharomyces cerevisiae. Shridas et al. (2003) described the identification of a cDNA for a long-chain cis-isoprenyltransferase located in the endoplasmic reticulum (ER) that they referred to as cis-IPTase, whereas Endo et al. (2003) cloned the same protein, dubbed dehydrodolichyl diphosphate synthase (DHDDS) (Fig. S2). More recently, two publications identified mutations in the DHDDS gene associated with autosomal recessive retinitis pigmentosa in Ashkenazi Jewish families, defining another gene associated with congenital disorders of glycosylation (CDGs) (Zelinger et al., 2011; Züchner et al., 2011). A cross-species phylogenetic comparison in the context of the cloning of the rainbow trout DHDDS gene shows that prokaryotic Z-type prenyltransferases, e.g. undecaprenyl diphosphate synthases (UppsSs), share a remarkable number of conserved positions with avian, mammalian, invertebrate and plant orthologues (Rebi et al., 2009).

The specific enzymes dedicated to the removal of the pyrophosphate from Dol-PP during Dol-P synthesis have not yet been identified. However, Cantagrel and coworkers elegantly demonstrated that the steroid 5x-reductase type 3 (SRD5A3) gene is required for the reduction of the x-isoprene unit in polyisoprenols to generate dolichol (Fig. S2) (Cantagrel et al., 2010; Cantagrel & Lefeber, 2011). These investigators demonstrated SRD5A3 and the yeast orthologue, DFG10, are necessary for conversion of polyisoprenol to dolichol in yeast, mice and humans. Their conclusions were based foremost on mutations in SRD5A3 in a large Emirati family from Baluchi (Southern Iran) and children from seven other families. Subsequently, other studies also identified mutations in patients further supporting...
SRD5A3 as associated with CDGs (Gründahl et al., 2012; Kahrizi et al., 2011; Kasapkara et al., 2012; Morava et al., 2010). Interestingly, these patients exhibited residual dolichol levels, implying that there must be an alternative pathway for dolichol biosynthesis (Cantagrel et al., 2010; Gründahl et al., 2012).

Dolichol must next be phosphorylated to form Dol-P (Fig. S2), the essential lipid carrier for both N- and O-linked glycosylation reactions in the ER (Mookerjea et al., 1983; Volpe et al., 1987). Allen et al. (1978, 1982) first showed CTP-dependent dolichol phosphorylation in mammalian tissues. The CTP-dependent dolichol kinase (DK) in yeast is encoded by sec59 (Heller et al., 1992) which, through complementation studies, led to the cloning of the human DK (Fernandez et al., 2002; Shridas & Waechter, 2006). Recently, mutations in DK have been associated with autosomal recessive forms of dilated cardiomyopathy in paediatric patients, extending the genes in dolichol biosynthesis associated with CDGs (Denecke & Kranz, 2009; Kapusta et al., 2013; Kranz et al., 2007; Lefeber et al., 2011).

Fig. 1(a) summarizes the events involved in protein glycosylation between the cytoplasm and lumen of the ER (Larkin & Imperiali, 2011). This begins with the assembly of UDP-N-acetylglucosamine and GDP-mannose onto the lipid carrier, Dol-P (terminal product in Fig. S2), followed by translocation of the lipid-sugar to the ER lumen for asparagine-linked protein glycosylation. The reaction most relevant to this review is the release of Dol-PP following transfer of the oligosaccharide to the peptidyl arginine by oligosaccharyltransferase. The enzyme dolichyl pyrophosphate phosphatase (DOLPP1) dephosphorylates Dol-PP to Dol-P, which is then flipped to the cytoplasmic leaflet allowing for recycling of Dol-P to reinitiate the glycosylation pathway. Therefore, the two enzymes critical to the generation of the essential lipid carrier, Dol-P, in eukaryotes are DK and DOLPP1 in synthesis and recycling, respectively (Fig. 1a). A protein sequence alignment by ClustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2/) across ~30 diverse mammalian species for DK (Fig. S3A) and DOLPP1 (Fig. S3B) resulted in an astonishing 81 % and 78 % identity, respectively. Although mutations in DK have been associated with autosomal recessive forms of dilated cardiomyopathy (Denecke & Kranz, 2009; Kapusta et al., 2013; Kranz et al., 2007; Lefeber et al., 2011), to date no documented mutations in humans have been found in DOLPP1. Also included on the alignment of DOLPP1 (Fig. S3B) is the consensus phosphatase motif originally defined by Stueck & Carmar (1997) and elaborated on by Neuwald (1997). This tripartite motif is shared by phosphatidic acid phosphatases (PAPs), the mammalian glucose 6-phosphatases and a collection of bacterial non-specific acid phosphatases. The motif has previously been shown to occur in both mouse DOLPP1 (Rush et al., 2002) and the yeast homologue, CWH8 (Fernandez et al., 2001; van Berkel et al., 1999), and shows complete identity in our mammalian species alignment except for the last aspartate (D), which is glutamine (Q) in all DOLPP1 enzymes (Fig. S3B). Interestingly, the glutamine is also found in place of the aspartate in the human, mouse and rat glucose 6-phosphatase proteins (Henrika et al., 1997).

### Biosynthesis of the eubacterial undecaprenyl lipid carriers

In contrast to dolichol biosynthesis, eubacteria synthesize undecaprenyl pyrophosphate (C_{55}-PP) which is then dephosphorylated to undecaprenyl monophosphate (C_{55}-P), the obligate lipid carrier for peptidoglycan, lipopolysaccharides, the enterobacterial common antigen, capsule polysaccharides and teichoic acid biosynthesis (Bouhs et al., 2008; Cantagrel & Lefeber, 2011; de Kruijff et al., 2008; Lovering et al., 2012; Reusch & Salton, 1984; Svoboda et al., 2010). In addition, undecaprenyl monophosphate is the carrier for N-linked protein glycosylation in prokaryotes (Fig. 1b) (Hitchen & Dell, 2006; Kumar & Balaji, 2011; Tabish et al., 2011). Smaller polyisoprenoids such as decaprenyl phosphate (C_{55}-P) are also as carriers for lipid I and II in mycobacteria (Mahapatra et al., 2005).

The mevalonate pathway, which involves the synthesis of IPP and its isomer, DMAPP, was previously believed to be the only source of isoprenoid precursors. However, a non-mevalonate pathway [also termed the 1-deoxy-D-xylulose 5-phosphate (DOXP) or 2C-methyl-D-erythritol 4-phosphate (MEP) pathway] exists in plant chloroplasts, algae, cyanobacteria, eubacteria and apicomplexan parasites (Hunter, 2007; Kuzuyama, 2002). The first step in the non-mevalonate pathway is the condensation of pyruvate and glyceraldehyde 3-phosphate to produce DOXP followed by isomerization to MEP catalysed by 1-deoxy-D-xylulose-5-phosphate reductoisomerase, the target of the antimicrobial, fosmidomycin (Proteau, 2004). Four subsequent steps lead to the synthesis of IPP and DMAPP, the universal precursors to FPP. The synthesis of undecaprenyl pyrophosphate in bacteria originates from all-trans-FPP by the addition of eight isopentenyl units catalysed by the soluble enzyme undecaprenyl pyrophosphate synthase UppS (Guo et al., 2005; Teng & Liang, 2012a, b; Zhu et al., 2013). Undecaprenyl pyrophosphate and dolichol de novo synthesis are distinguished by terminal products (an isoprenyl pyrophosphate versus an alcohol) as well as the saturation state of the α-isoprene units (unsaturated versus saturated). Interestingly, unlike Gram-negative bacteria which possess undetectable levels of free undecaprenol, Gram-positive bacteria maintain a significant level of free undecaprenol, but their biosynthetic pathway remains to be determined (Van Horn & Sanders, 2012). Similar to eukaryotes which encode a DK, Gram-positive bacteria encode an undecaprenol kinase (dgkA) (Hartley et al., 2008; Jerga et al., 2007), indicating that the evolutionarily older Gram-positive bacteria can yield undecaprenyl monophosphate both through phosphorylation and more universally in all eubacteria through phosphatase activity.
Undecaprenyl pyrophosphate phosphatases in bacteria

Fig. 1(b) summarizes the events in peptidoglycan synthesis emphasizing the essential role of undecaprenyl monophosphate as the critical lipid carrier molecule. The initial step is the cytosolic conversion of undecaprenyl pyrophosphate to undecaprenyl monophosphate. We had originally cloned the bacA gene as an enzyme which conferred increased bacitracin resistance to Escherichia coli based on plasmid-encoded overexpression (Cain et al., 1993). We had hypothesized that this gene might encode a protein with undecaprenyl kinase activity. However, El Ghachi et al. (2004) subsequently demonstrated that bacA encodes a protein with undecaprenyl pyrophosphate phosphatase activity, coining the term UppP. They also showed that disruption of the bacA gene alone in E. coli was not lethal despite accounting for 75% of the undecaprenyl pyrophosphate phosphatase activity. These investigators hypothesized that additional enzymes were necessary to account for the remaining ~25% of the undecaprenyl pyrophosphate phosphatase activity, coining the term UppP. They also showed that disruption of the bacA gene alone in E. coli was not lethal despite accounting for 75% of the undecaprenyl pyrophosphate phosphatase activity. These investigators hypothesized that additional enzymes were necessary to account for the remaining ~25% of the undecaprenyl pyrophosphate phosphatase activity and demonstrated that disruption of two additional genes, ybjG and pgpB, along with bacA/uppP caused bacterial cell lysis (El Ghachi et al., 2005). We subsequently used the combined name BacA/UppP to both address activity and distinguish from other enzymes containing undecaprenyl pyrophosphate phosphatase activity. Another study also similarly evaluated alternative enzymes in E. coli with the assumption that phosphatase activity was necessary in the periplasm where undecaprenyl pyrophosphate is released during peptidoglycan transfer and recycled to the cytoplasm after dephosphorylation to undecaprenyl monophosphate (Fig. 1b) (Tatar et al., 2007). In contrast, these investigators proposed that along with BacA/UppP in the cytosol, YbjG and, to a lesser extent, YeiU/LpxT, but not PgpB, were necessary for bacterial viability through the dephosphorylation of undecaprenyl pyrophosphate in the periplasm, allowing the recycling of undecaprenyl monophosphate to the inner leaflet of the membrane. Fig. 1(b) summarizes the tripartite dephosphorylation events by indicating that all three phosphatases, PgpB, YbjB and YeiU/LpxT, may contribute to lipid carrier recycling.

All of these enzymes except for BacA/UppP contain the tripartite motif shared by PAPs (Neuwald, 1997; Stukey & Carman, 1997). This consensus (Kx6Rx2_5x2PSGHx31–34Rx5Hx4D) has been independently highlighted for the bacterial enzymes, PgpB, YbjB and YeiU/LpxT (El Ghachi et al., 2005; Tatar et al., 2007), as well as the human DOLPP1 (Rush et al., 2002) and yeast CWH8 (Fernandez et al., 2001; van Berkel et al., 1999) enzymes. We present alignments of each bacterial protein to the human DOLPP1 protein sequence showing the respective PAP motifs and indicating the positions of identity as compared to the alignment to CWH8, the orthologue from yeast (Fig. S4). We have also included a similar comparison to the YwoA/BcrC protein, a possible subunit of the Bacillus licheniformis bacitracin resistance ABC transporter (Podlesek et al., 1995) that has been argued subsequently to be an undecaprenyl pyrophosphate phosphatase found in various Bacillus strains (Bernard et al., 2005). Human DOLPP1 shares ~29% identity with yeast CWH8, while each of the bacterial enzymes have only 10–19% identity to the human enzyme. Fig. 2 presents the alignment of the PAP motif from each of these enzymes showing the level of variability in this tripartite motif amongst these proteins.

Tatar et al. (2007) demonstrated previously the problems faced by a predictive algorithm in determining topology while mapping some residues in truncated proteins. We have expanded upon their in silico efforts by utilizing an array of programs and algorithms in an attempt to determine a consensus or an algorithm that best predicts the topology of enzymes in these pathways. Table 1 illustrates that amongst eight different software programs [TMHMM (Möller et al., 2001), TMapred (Hofmann & Stoffel, 1993), PRED-TMR2 (Pasquier & Hamodrakas, 1999), SOSUI (Hirokawa et al., 1998), SPLIT4.0 (Juretic et al., 2002), HMMTOP (Tusnady & Simon, 1998), MEMSAT2 (Jones et al., 1994) and OCTOPUS (Viklund & Elofsson, 2008)] that predict transmembrane helices, only human DOLPP1 and the E. coli PgpB protein are consistently predicted to have all three motifs residing on the same side of the membrane, whilst only one program (OCTOPUS) predicts this arrangement for YeiU/LpxT. We provide the 2D structure predictions for human DOLPP1 and E. coli PgpB and YeiU/LpxT using MEMSAT2 to predict transmembrane helices, as illustrated by TMRPres2D (Spyropoulos et al., 2004). These structures illustrate the disparity in predicting the entire tripartite motif on one side of the membrane (Fig. S5). The inconsistency in the prediction for YeiU/LpxT may result from a combination of low sequence identity as compared to DOLPP1 (10%, Fig. S4) and divergence in the tripartite motif (Fig. 2). Our conclusion from this comparative analysis is that, similar to the conclusions of Tatar et al. (2007), no single program predicts consistent topology for all the proteins; however, for some proteins the topological predictions were in agreement, whilst a consensus amongst multiple programs may be necessary to better predict the functional position of the motifs in the absence of structural and mutagenesis studies.

The E. coli pgpB gene product was first proposed as one of three phosphatidylglycerol phosphate (PGP) phosphatases critical in the generation of phosphatidylglycerol (Dillon et al., 1996; Funk et al., 1992; Lu et al., 2011). Interestingly, bacteria with disruption of all three PGP phosphatase genes, pgpA, pgpB and pgpC (Lu et al., 2011), are not viable. However, Touzé et al. (2008a) have argued that PgpB functions as a periplasmic undecaprenyl pyrophosphate phosphatase even though their data demonstrates a 100-fold higher activity towards PGP. The basis for their assumption comes from a fourfold enhancement of PgpB activity on undecaprenyl pyrophosphate in the presence of phospholipids. Similarly, the protein YeiU/LpxT has also been argued to catalyse the transfer of a phosphate group
from the undecaprenyl pyrophosphate donor to the 1-position of lipid A to form lipid A 1-diphosphate on the periplasmic side of the inner membrane (Touzé et al., 2008b; Valvano, 2008). To date, the enzymic specificity towards undecaprenyl pyrophosphate has only been analysed for PgpB where it primarily acts upon PGP and has minimal activity towards undecaprenyl pyrophosphate (Touzé et al., 2008a). These enzymes have therefore not been compared directly with each other to assess their relative activity on undecaprenyl pyrophosphate and their physiologically relevant contributions in vivo. At present, the role of these enzymes is only defined through genomic disruption, which determines the effects on viability and resistance to bacitracin without specifically addressing function, whereas the human DOLPP1 has been shown to complement mutations in the yeast enzyme (Rush et al., 2002). Given the current studies, it is therefore still open to debate as to the importance of the recycling enzymes in the overall cellular production of the essential lipid carrier undecaprenyl monophosphate.

**BacA/UppP: the principal undecaprenyl pyrophosphate phosphatase in eubacteria**

BacA/UppP is a highly hydrophobic protein, whose overexpression on a multicopy plasmid causes a 280-fold increase in UppP activity, whereas a gene disruption mutant retains only 25 % of total UppP activity (Cain et al., 1993; El Ghachi et al., 2004, 2005; Tatar et al., 2007). As the principal *E. coli* undecaprenyl pyrophosphate phosphatase, we therefore predicted that BacA/UppP would be very highly conserved across Gram-negative bacteria. A ClustalW-based alignment of BacA/UppP protein for 36 Gram-negative bacterial species illustrates a 63 % identity (Fig. S6A), while by comparison, similar alignments for PgpB, YbjG and YeU display 28 %, 18 % and 27 % identity, respectively (Fig. S6B–D). A complete nucleotide NCBI BLAST search for YwoA/BrcC (Bernard et al., 2003, 2005; Ohki et al., 2003) identifies apparent orthologues from only a limited number of Gram-positive species. An alignment across these Gram-positive species shows only a 9 % identity, whereas comparison to only the highly represented *Bacillus* genus displays 15 % identity (Fig. S6E). Of particular note, based on complete BLAST searches, obvious homologues/orthologues of YbjG and YeU are not found in Gram-positive bacteria, whilst the PgpB and BacA/UppP proteins are present extensively in the older bacterial phyla. The presence of PgpB across a wide range of eubacteria may be due to its importance as a PGP phosphatase (Lu et al., 2011; Touzé et al., 2008a). The high level of identity for BacA/UppP (63 %) demonstrates clearly the evolutionary pressure to maintain the functionality of this protein and also raises further questions about the biological relevance of the other proteins.

As already indicated, BacA/UppP does not contain the PAP consensus motif found in eukaryotic DOLPP1 and the other putative bacterial undecaprenyl pyrophosphate phosphatase proteins (Fig. S6), yet its inherent activity is dependent on cleavage of a pyrophosphate bond. Given the high degree of identity across the Gram-negative bacteria, it is difficult to predict catalytically relevant residues (Fig. S6A, highlighted residues). However, an alignment using *E. coli* BacA/UppP as the reference sequence compared with a collection of more divergent Gram-positive bacterial sequences, including primarily human pathogens, identifies two regions of the protein that retain a high level of identity (Fig. S7). These regions were noted originally as conserved regions in the identification of *bacA* gene orthologues in *Streptococcus pneumoniae* and *Staphylococcus aureus* based on comparison with six other bacterial species available at this time, including *E. coli* BacA/UppP (Chalker et al., 2000). Supporting the notion of the central importance of BacA/UppP, these investigators demonstrated that a *Strep. pneumoniae bacA* mutant was 160 000-fold more sensitive to bacitracin and that this mutant bacterium displayed a 40 000-fold less bacterial load reflected as c.f.u. in a mouse lung model of infection.

We therefore attempted to better characterize the sequence-dependent motifs underlying the pyrophosphate
phosphatase activity in BacA/UppP. In searching other large phosphatase protein families, we evaluated the catalytic consensus associated with the protein tyrosine phosphatases (PTPs) (Dixon, 1995; Hendriks et al., 2013), a large diverse family with a highly conserved active site with the following consensus motif derived from 151 vertebrate PTP sequences: PxxVHCSAGxGRTG (Tonks, 2013). Included in the PTP family are the dual-specificity phosphatases (DUSPs) defined by the prototypic enzyme, VH1, from vaccinia virus. The DUSPs have been divided recently into three classes based on their sequence homology, with class 2 VH1-like DUSPs including the tumour suppressor, phosphatase and tensin homologue (PTEN), as the most prominent member (Tonks, 2013). PTEN preferentially catalyses the dephosphorylation of phosphatidylinositol 3,4,5-trisphosphate at the 3 position, thus defining a different class of lipid phosphatases relative to the PAP family and differentiating PTEN from the classic PTPs by substrate specificity (Zhang & Yu, 2010). Also included recently in the PTP-lipid phosphatase family is a class 3 VH1-like DUSP mitochondrial enzyme, PTPMT1, whose preferred substrate is PGP (Tonks, 2013; Xiao et al., 2011; Zhang et al., 2011). Interestingly, PGP also serves as the substrate for the bacterial enzymes, PgpA, PgpB and PgpC (Lu et al., 2011), whilst, as noted previously, PgpB also contains partial undecaprenyl pyrophosphate phosphatase activity (Touzé et al., 2008a). The substrates for PTEN, BacA/UppP and PTPMT1 are shown in Fig. 3(a), illustrating the general similarity of a terminal phosphate attached to a lipid tail. We therefore compared the signature P loop motif of PTEN to BacA/UppP, which demonstrated a striking similarity with PTEN and the classic PTP consensus sequence, with the most obvious difference being an increased size in the P loop by three amino acids (Fig. 3b). An upstream acidic residue, E/D, and a conserved proline are present in BacA/UppP, similar to the WPD loop in PTEN that transitions from an ‘open’ to a ‘closed’ state, optimally placing the catalytic acidic residue (E/D) in a favourable position to facilitate substrate dephosphorylation (Alicea-Velazquez & Boggon, 2013; Barr et al., 2009; Zhang & Bishop, 2008). The WPD and P loop regions are completely conserved for BacA/UppP homologues in Gram-negative bacteria (Fig. S6A). An alignment of 67 Gram-positive and -negative bacteria highlighted residues that maintained complete conservation across species, leading to a consensus sequence, E/DxxE/GxxQxxPGxSRSxxT, which is specific to the BacA/UppP family (Fig. 4a).

Crystal structures exist for PTEN (Lee et al., 1999), PTPMT1 (Xiao et al., 2011) and various PTPs (Barr et al., 2009), but no structures have been determined for the bacterial undecaprenyl pyrophosphate phosphatases. We again employed the eight programs to predict transmembrane helices, six of which predicted the consensus catalytic region (Fig. 4a) resides on a single cytoplasmic loop (Fig. 4b). We also scanned the 67 BacA/UppP sequences for highly conserved or identical residues, with these results
overlaid on the 2D structure (Fig. 4b). Notably, the only other region to display a high level of identity resides in the putative first transmembrane helix, which coincides with the conserved region found from the alignment of BacA/UppP among Gram-positive bacteria (Fig. S7). Beyond catalysis, it is likely that this protein contains some specificity for the lipid portion of its substrate. Previous studies have identified a potential isoprenol recognition sequence for dolichol (FI/VxF/YxxIPFxF/Y) which we were unable to identify in BacA/UppP (Albright et al., 1989; Datta & Lehrman, 1993). We therefore hypothesize that the residues exhibiting identity in the first transmembrane domain constitute an undecaprenol recognition motif.

Based on the high level of identity and conservation of the BacA/UppP family, we also generated a phylogenetic tree and compared the results to a tree based on 16S rRNA (Fig. S8). Both alignments provide the expected separation between the Gram-positive and -negative bacteria. However, we have highlighted specific examples of the divergence between these two trees, implying that an alternative evolutionary perspective may be obtained from utilizing a highly conserved protein versus the 16S rRNA sequences.

**Conclusions and future directions**

A comparison of human DOLPP1 and the proposed bacterial recycling enzymes demonstrates the conservation of the classic PAP motif. However, our analysis of the PAPs related to the recycling of bacterial lipid carriers points to the need for further study. Most importantly, it has been reported that one of these enzymes, PgpB, acts primarily as a PGP phosphatase in isolation, whereas the substrate preference for the other enzymes has not yet been determined definitively. Further research may provide additional explanations for the range of specificity of these enzymes as well as the proposed redundancy found in this pathway.

We have identified that BacA/UppP contains potential catalytic residues with a high level of identity to the PTP phosphatase motif found in PTEN whose preferred substrate is phosphatidylinositol 3,4,5-trisphosphate, thus underscoring
the similarity with BacA/UppP by acting upon a lipid-based substrate. Alignments across Gram-positive and -negative bacteria were used to identify both a potential catalytic motif, E/Dx9–14E/Dx9–12GxxQx5PGxSRSxxT, unique to the bacterial undecaprenyl pyrophosphate phosphatase enzyme, BacA/UppP, and a region of high amino acid identity in the putative first transmembrane domain of BacA/UppP, pointing to a likely recognition motif for the undecaprenyl substrate backbone. Given the high level of conservation of BacA/UppP, these insights open the door to new avenues of research targeting bacterial cell envelope biosynthesis and cell viability as a strategy for the development of antibacterial therapeutics.

References

Bacterial monophospholipid carriers


Rush, J. S., Cho, S. K., Jiang, S., Hofmann, S. L. & Waechter, C. J. (2002). Identification and characterization of a cDNA encoding a...


Edited by: S. Spiro


