International Journal of Medical Microbiology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

Three-dimensional structures of Lipoproteins from *Streptococcus pneumoniae* and *Staphylococcus aureus*

Sergio G. Bartual[1], Martín Alcorlo, Siseth Martínez-Caballero, Rafael Molina, Juan A. Hermoso*

Department of Crystallography and Structural Biology, Institute of Physical Chemistry "Rocasolano", CSIC, 28006 Madrid, Spain

ARTICLE INFO

Keywords:
Streptococcus pneumoniae
Staphylococcus aureus
Lipoproteins
Protein structure
Virulence

ABSTRACT

Bacterial lipoproteins (Lpp) compose a large family of surface-exposed proteins that are involved in diverse, but critical, cellular functions spanning from fitness to virulence. All of them present a common signature, a sequence motif, known as LipoBox, containing an invariant Cys residue that allows the protein to be covalently bound to the membrane through a thioether linkage. Despite the abundance and relevance of Lpp, there is a scarcity of structural and functional information for this family of proteins. In this review, the updated structural and functional data for Lpp from two Gram-positive pathogenic model organisms, *Staphylococcus aureus* and *Streptococcus pneumoniae* is presented. The available structural information offers a glimpse over the Lpp functional mechanisms. Their relevance in bacterial fitness, and also in virulence and host-pathogen interactions, reveals lipoproteins as very attractive targets for designing of novel antimicrobials, and interesting candidates as novel vaccine antigens.

1. Introduction

A common feature of all bacterial groups is the presence of a single or double membrane that serves, among other functions, as an anchor point for proteins. Membrane proteins can be classified according to their interaction with the membrane as integral membrane proteins, when they are embedded in the phospholipid bilayer by transmembrane segments, or peripheral membrane proteins, when they remain attached to the membrane surface but not inserted. Bacterial lipoproteins (Lpp) are classified as peripheral membrane proteins that have been shown to be directly involved in diverse, but critical, cellular functions like: cell fitness, division, signal transduction, motility, redox and antibiotic resistance, extracytoplasmic folding of proteins, conjugation, sporulation, adhesion and virulence (Abdullah et al., 2014; Alloing et al., 1994; Kohler et al., 2016; Nguyen and Gotz, 2016; Saleh et al., 2013; Schmaler et al., 2009; Shahmirzadi et al., 2016) (Fig. 1A). Indeed, Lpp contribute directly to the pathogen virulence by promoting colonization, invasion, and survival in the host bloodstream (Kovacs-Simon et al., 2011; Pribyl et al., 2014; Saleh et al., 2013).

Gram-negative Lpp could be placed in any of the two membranes but normally are attached to the inner leaflet of the outer membrane (Narita et al., 2004). In Gram-positive bacteria Lpp are translocated from the inner face of the plasma membrane to the extracellular surface (Hutchings et al., 2009). Furthermore, due to the lack of a second membrane, Gram-positive bacteria avoid the diffusion of many vital proteins into the external space by turning them into Lpp.

Bacterial Lpp modification, normally, includes the attachment of an N-Acyl Diacyl Glycerol group that allows Lpp to remain attached to the membranes by direct interaction with the phospholipids polar heads. Some Gram-positive bacteria like *S. aureus* can, in addition to the N-Diacyl Glycerol, attach N-Acyl Triacyl Glycerol depending on environmental conditions (Nakayama et al., 2012). Lpp typically consist of an N-terminus signal peptide, containing positively-charged residues, a hydrophobic sequence and the LipoBox motif [LVI] [ASTVI] [GAS] [C] (Sankaran and Wu, 1994) (Fig. 1B). After synthesis, the N-terminal signal peptide allows preprolipoproteins to be translocated across the cytoplasmic membrane by either a Sec or Tat secretory pathways (Driessen and Nouwen, 2008). Then is anchored to the cell wall by the addition of diacylglyceryl moieties (lipidation) to the thiol group of an invariant cysteine residue in the lipobox motif via the lipoprotein diacylglyceryl transferase (Ltg) (Kohler et al., 2016; Oudega et al., 1993). Finally, a type II signal peptidase (Lps) cleaves the N-terminal signal peptide close to the Cys/diglyceride bond generating the mature lipoprotein.

Interestingly, in many Gram-positive pathogenic bacteria, such as *Streptococcus pneumoniae* or *Staphylococcus aureus*, most Lpp are required for sustaining virulence in pathogen-host interactions (Johnston et al., 2004; Saleh et al., 2013; Shahmirzadi et al., 2016). However,

* Corresponding author.
E-mail address: xjuan@iqfr.csic.es (J.A. Hermoso).
[1] Present address: School of Life Sciences, University of Dundee, DD1 5HL, Dundee, UK.

http://dx.doi.org/10.1016/j.ijmm.2017.10.003
Received 1 August 2017; Received in revised form 21 October 2017

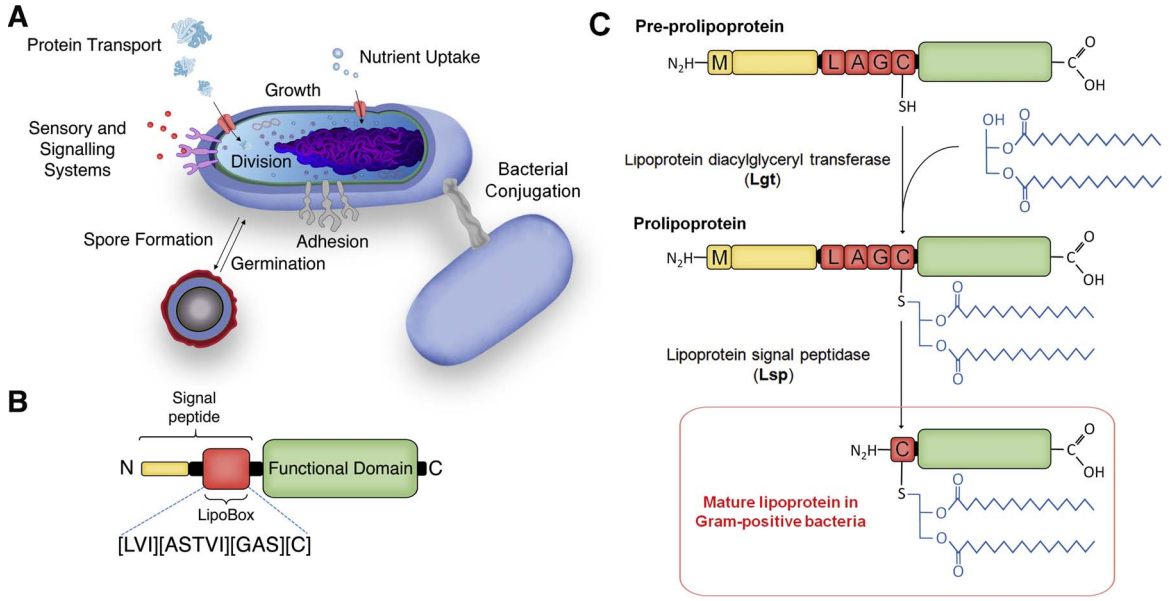

Fig. 1. (A) Schematic representation of the biological roles exerted by bacterial Lipoproteins. Figure adapted from http://www.mrc-lmb.cam.ac.uk/genomes/dolop/(Babu et al., 2006; Madan Babu and Sankaran, 2002). (B) Schematic cartoon of the modular arrangement found in bacterial Lipoproteins. The lipobox contains the invariant Cys that is lipid-modified and the −3 position is mainly Leu. The consensus sequence [LVI][ASTVI][GAS][C] is shown in red. Apart from the lipobox the N-terminal 5–7 residues with mostly two positively charged Lys or Arg residues and the intervening stretch of hydrophobic and uncharged residues of 7–22 amino acid length are important (yellow box). (C) Biosynthesis of lipoproteins in Gram-positive bacteria. The precursor of lipoproteins is the preprolipoprotein, with an N-terminal signal peptide including the consensus sequence of the lipobox at the C-terminal region of the signal peptide (shown in red). During lipoprotein maturation, the thiol group of the invariant cysteine in the lipobox is modified by a diacylglyceryl moiety (in blue) by lipoprotein diacylglyceryl transferase (Lgt), which serves as a membrane anchor. After lipidation, lipoprotein signal peptidase (Lsp) cleaves the signal peptide, leaving the cysteine as the new amino-terminal residue forming the mature lipoprotein in Gram-positive bacteria. M stands for the initial methionine in the precursor forms. Figure adapted from (Kovacs-Simon et al., 2011). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

despite their attractiveness as potential Gram-positive antimicrobial targets and novel vaccine antigens, little is known about their three-dimensional structures. There are different facts that could explain the scarcity of structural information in Lpp family. Among them, the hydrophobic nature of the full-length protein (all the up to now reported crystal structures of Lpp are obtained after deletion of the LipoBox region), the usual presence of a flexible linker connecting the lipidated Cys with the functional domain(s) that hamper crystallization of Lpp, and finally, the insufficient information about *in vivo* functions for many of these Lpp (few of them presenting a catalytic activity), could explain why these proteins have been typically neglected in structural studies. I.e. within the disease model organisms *S. aureus* USA300 and *Pneumococcus* TIGR4, 83% and 49% of their respective Lpp remains structurally uncharacterized. Furthermore, among all the available Lpp structures only half of them have been functional and biochemically characterized. This fact reflects the huge gap in knowledge for this group of critical membrane proteins.

During the last years, there has been intensive bioinformatics analysis of bacterial Lpp. An example of this is the development of DOLOP, a knowledge base for bacterial lipoproteins (http://www.mrc-lmb.cam.ac.uk/genomes/dolop/) that provides a list of functional classification, predictive algorithm for query sequences, primary sequence analysis and lists of predicted lipoproteins from different completed bacterial genomes (Babu et al., 2006; Madan Babu and Sankaran, 2002). DOLOP, accessed on July 2017, contains information of 199 distinct lipoproteins identified among 43 bacterial genomes and classified according to their cellular functions. Although it is a powerful computational tool, DOLOP still doesn't offer functional or structural correlation to published data.

As all the reported three-dimensional structures of Lpp refer to the protein part, and typically just their functional domains, this review focus on the comparison of the Lpp roles and associated three-dimensional structures for two main Gram-positive human pathogens, *S. pneumoniae* TIGR4 and the *S. aureus* MRSA USA300.

2 Overview of the solved *S. pneumoniae* and *S. aureus* lipoproteomes

Lipoproteome constitutes a minimal fraction of the total bacterial genome, 1.76% in the case of *S. pneumoniae* TIGR4 and 2.04% for the bigger *S. aureus* complete genome (Babu et al., 2006). Due to their exposed location and the critical functions under their command, Lpp are being considered targets for the generation of new antimicrobials and interesting candidates for novel vaccine antigens. However, to achieve this goal a deep structural and functional knowledge is required.

Evaluation of the protein data bank (PDB, https://www.rcsb.org/pdb/home/home.do) accessed on July 2017, reveals that, in the case of *S. aureus* USA300, only one third of lipoproteins have been, up to now, structurally characterized. PDB database contains only 18 out of a total of the 67 lipoproteins reported in the strain USA300 that accounts for 0.7% of the total genome (Shahmirzadi et al., 2016). Interestingly, among all the deposit bacterial Lpp structures, there is no structural information of the lipobox or the attached lipid, thus only Lpp catalytic protein cores can be structurally analysed. A large proportion of the staphylococcal Lpp is involved in the uptake of essential ions and nutrients, 24 Lpp are distributed among metal, anion, amino acid and sugar transporters (Fig. 2A). Of them, only 9 have been structurally determined: 4 iron transporters (FhuD2, SirA, SirF and HstA), one manganese and one nickel transporters (MntC and NikA, respectively), one nitrate ABC binding protein, two amino acid transporters (OpuCc and GmpC) and one maltose binding protein (Table 1A). The rest of the *S. aureus* Lpp are divided into other 7 functional categories (Biosynthesis, Respiration, Protein folding, Protein translocation, Phage and plasmid Lpp, Lpl cluster and the unknown Lpp) that accounts for the 62.68% of the total lipoproteome. However, the structures of only 11% of them have been solved (Table 1A). The Biosynthesis category has a single Lpp involved in plasmid uptake whose structure is reported (CamS). Respiration category contains two Lpp but no structure is reported so far. The Protein folding category contains two Lpp of known