DOCTOR OF SCIENCES

Investigation of the surface lipoprotein export machinery of Bacteroidetes

Lauber, Frédéric

Award date:
2016

Awarding institution:
University of Namur

Université de Namur, Faculté des Sciences, Département de Biologie
Unité de Recherche en Biologie des Microorganismes

10th October 2016

Investigation of the surface lipoprotein export machinery of Bacteroidetes

Submitted by Frédéric Lauber
For the degree of PhD in Science

JURY
Prof. Wilbert Bitter (VU University Amsterdam, The Netherlands)
Prof. Jean-François Collet (UCLouvain, Belgium)
Prof. Johan Wouters (UNamur, Belgium)
Prof. Xavier De Bolle (UNamur, Belgium)
Prof. Guy R. Cornelis (UNamur, Belgium)

Supervisors:
Prof. Guy R. Cornelis
Dr. Francesco Renzi

Foreword

Bacteria of the phylum Bacteroidetes harbor abundant surface exposed multi-protein membrane complexes (Sus-like systems) involved in carbohydrate acquisition. These complexes are key determinants for commensalism and also play a role in pathogenesis. Interestingly, Sus-like systems are mainly composed of lipoproteins anchored to the outer membrane and facing the external milieu. This lipoprotein localization is uncommon in most Gram-negative bacteria while it is widespread in Bacteroidetes. To date however, little is known on how these complexes assemble and in particular on how lipoproteins reach the bacterial surface.

The work presented in this manuscript thus aims at providing a better understanding of how lipoproteins are transported to the cell surface in Bacteroidetes as well as how they are distinguished from intracellular lipoproteins using our model organism *Capnocytophaga canimorsus*.

During my master thesis, I explored the function of *C. canimorsus* Sus-like systems. In particular, I could show that one of these systems is devoted to iron acquisition. While this discovery was done before my PhD, the additional experiments required during the publication process took a substantial amount of my time during my PhD. Hence, although it is not directly linked to the main topic of my thesis but nevertheless in relation to it, we decided to include this manuscript in the thesis.

Table of contents

Acknowledgements

First of all, I would like to thank all the members of my group …

... my supervisor **Guy R. Cornelis** for shipping me to Belgium alongside the lab equipment, for supporting me throughout my thesis, for letting me a lot of freedom while at the same time providing guidance and healthy criticism, for believing that I was indeed on the right track and for being so enthousiatic (often more than me) when I finally found what we were looking for.

... **Francesco "Franz" Renzi**, my second supervisor, without whom all of this would not have been possible. Thank you for all your helpful advices and technical tricks, you quick thinking and great problem solving skills, your many corrections even late in the evening, and of course your jokes. Thank you also for being a friend outside of the lab, for showing me a bit of Italian lifestyle and for pushing me to do sports (boxing and running), although we sometimes both regretted the latter. I will miss all this for sure. I am happy to welcome you in England whenever you want !

... **Estelle Hess** for showing me the basics of fluorescence microscopy and flow cytometry and for letting me use your cell cultures. Thank you also for sometimes tempering the wild character of Franz.

… **Katrin Hack** for helping with my first ultracentrifugation ever, for you English corrections, for helping me keeping up my German skills, for deciphering with me the rules of the doctoral school and for being the calm center of the group.

... **Emeline Lawarée** for your advices in terms of thesis writing, FACS data presentation and your sense of humor. See you soon in Oxford !

... **Mélanie Dol** and **Dounia Koudad** for helping me out in the lab and making my life much easier, for bringing me food when I couldn't go myself and for a good laugh.

I am especially grateful to **all members of my jury** who accepted to critically review this thesis and to all my **PhD committee members** who gave me much needed feedback.

My special thanks go to **Dr. Paul Jenö** and **Suzette Moes** for their incredibly fast and efficient analysis of my MS samples.

I have of course to thank **Pablo Manfredi**, who initiated me to the world of *Capnocytophaga canimorsus* back in Basel during my master thesis as well as to the Voodoo aspects of science.

Thanks to all the members of the **URBM** for creating a nice working atmosphere, for always being helpful, for explaining me all sorts of things and for lending me reagents.

I also want to thank **Michel Jadot** and **Isabelle Hamer** for their explanations regarding the ultracentrifuges and for letting me use them extensively.

I would like to thank **Simon Moussa**, basically for staying true to himself: always being in a good mood, ready to do stupid things and for being the only person with whom I could discuss anime/manga.

I also thank **Jérome Coppine** and **Mathieu Waroquier** for their very special Belgian sense of humor.

A big thank you as well to **Leo-Paul Tisserant** and **Thomas Spitz** for being great friends, for allowing me to test my latest decks in Magic the Gathering whenever we met and for always acting as if we saw each other just yesterday.

A special thanks to **Romain Gasser** who always listened to my complains, shared my sometimes weird sense of humor and agreed to spend some of his holidays with me (Japan, we'll see you again !).

Last but not least, I am especially grateful to my family: my parents **Elke** and **André Lauber**, my sister **Catherine Lauber** and my brother in law **William Meriau** as well as their little monsters **Oliver** and **Alexis**. Thank you all for offering me much needed holidays, for pulling me out of the world of science for a couple of weeks each year and for always supporting me.

Abbreviations

ABC transporter	ATP binding cassette transporter
Bam	Beta barrel assembly machinery
Cc	*Capnocytophaga canimorsus*
CFU	Colony forming unit
GlcNAc	N-acetylglucosamine
HA	Hemagglutinin
HIHS	Heat-inactivated human serum
IM	Inner membrane
Lgt	Lipoprotein diacylglyceryl transferase
Lnt	Lipoprotein N-acyl transferase
Lol	Localization of lipoprotein
LPS	Lipopolysaccharide
Lpt	Lipopolysaccharide transport
LspA	Lipoprotein signal peptidase
MS	Mass spectrometry
MucG	*Capnocytophaga canimorsus* mucinase
OM	Outer membrane
OMP	Outer membrane protein
OstA	Organic solvent tolerance protein A (LptD)
PFA	Paraformaldehyde
PMF	Proton motive force
PNA	Peanut Agglutinin
POTRA	Polypeptide translocation associated
PUL	Polysaccharide utilization loci
Sec	Secretion
SiaC	*Capnocytophaga canimorsus* sialidase
Skp	Seventeen kilodalton protein
SPase	Signal peptidase
SPI	Signal peptide I
SPII	Signal peptide II
Strep	Streptavidin
Sus	Starch utilization system
Tat	Twin-arginine translocase
TMH	Transmembrane helix
TPR	Tetratricopeptide
wt	Wild type

Summary

This work focuses at exploring the mechanisms underlying lipoprotein surface localization in Gram-negative bacteria of the phylum Bacteroidetes using as model organism the dog commensal and human pathogen *Capnocytophaga canimorsus* (*Cc*). While unusual in most studied bacteria, this lipoprotein localization is crucial both for commensalism and pathogenicity of many Bacteroidetes.

By *in silico* analyses of *Cc* surface exposed lipoproteins, we identified an N-terminally conserved motif (QKDDE). We show that this motif is sufficient for cell surface localization when introduced in an intracellular lipoprotein and thus represents the *Cc* lipoprotein export signal (LES). We further demonstrate that the overall negative charge of the LES is essential for protein transport. We also determined the minimal composition for a functional LES as well as its optimal positioning. Finally, an *in silico* analysis performed on the lipoproteins of two other Bacteroidetes species, namely *Bacteroides fragilis* and *Flavobacterium johnsoniae*, revealed that the LES is broadly distributed among the phylum. The derived LES of each species was tested and found to be functional in *Cc*, indicating strong conservation of the signaling and the putative lipoprotein transport mechanisms in Bacteroidetes.

We also focused at identifying the underlying lipoprotein transport machinery. We first searched for LolA interaction partners, the periplasmic chaperone of lipoproteins, which led to the identification of several candidates. We found all of them to be involved to some extend in outer membrane biogenesis and/or to be required for growth in liquid medium. In particular, we could show that an Skp homolog is essential in *Cc* and that its depletion leads to early growth arrest. However, their exact function remains to be clarified.

In parallel, we investigated highly conserved proteins unique to Bacteroidetes, *i.e.* putative candidates for the lipoprotein transport machinery. We found that most Bacteroidetes genomes encode more than one BamA

homolog. This additional copy of BamA (Ccan_17810) turned out to be essential in *Cc* and to require an N-terminal lipid anchor for its functioning. Furthermore, when expressed in *E. coli*, Ccan_17810 led to rapid growth arrest and formation of ghost cells. Due to the lack of efficient regulatable expression systems in *Cc* and despite our efforts to generate new ones, the precise function of this protein could not be determined.

Finally, we also investigated the function of surface exposed lipoproteins in *Cc*. Following *in silico* analyses, we identified and characterized a new type of iron acquisition system essential for growth of *Cc* in human serum. This was of particular interest due to the broad substrate specificity of the system, targeting several iron carrying proteins found in humans and other mammals, as well as its pathogen specific distribution among Bacteroidetes. Interestingly, this system was found to have the classical architecture of Sus-like systems, outer membrane anchored complexes usually devoted to polysaccharide degradation. These systems being mostly composed of surface exposed lipoproteins, this study thus showed for the first time that Sus-like systems can target other substrates than carbohydrates, in this case iron, and that surface exposed lipoproteins can be virulence factors in the phylum Bacteroidetes.

Introduction

The Gram-negative bacterial membrane

Since 1884, bacteria have been classified into two distinct groups, Gram-positive and Gram-negative, based on the eponymous staining procedure reflecting their membrane architecture[1]. While nowadays this method of classification is not as clear-cut anymore, the terms Gram-negative to refer to diderm-LPS bacteria and Gram-positive to refer to both monoderm as well as diderm-mycolate bacteria will be used here for the sake of simplicity.

The cell envelope of Gram-positive bacteria is composed of a plasma membrane delimiting the cytoplasm and a thick peptidoglycan layer in which are inserted teichoic and lipoteichoic acids (Fig. 1)[1]. On the other hand, the Gram-negative bacterial cell is composed of two compartments, the cytoplasm and the periplasm, delimited by an inner membrane (IM) and an outer membrane (OM) respectively. Residing in between these two membranes is an additional thin layer made of peptidoglycan (Fig. 1)[1]. Taken together, this architecture provides Gram-negative bacteria with a cell wall strong enough to resist temperature and pH variations and elastic enough to withstand osmotic changes. This formidable protective layer also shields the cell from noxious compounds, allowing it to survive extracellular stresses and to proliferate in changing, sometimes toxic, environments[1].

The most notable feature of the Gram-negative cell wall is the outer membrane. Unlike the IM, the OM is an asymmetric bilayer, composed of phospholipids on the inner leaflet and lipopolysaccharide (LPS) on its outer leaflet (Fig. 1)[1]. LPS molecules tightly interact with each other in order to form a dense network that is impermeable to most compounds[1]. While this assures protection of the cell, this barrier function also strongly limits its ability to release or uptake various molecules. The OM thus also comprises outer membrane proteins (OMPs) called porins, water filled channels that allow import of nutrients[1]. Additionally, Gram-negative bacteria harbor proteins directly exposed at the cell surface, often involved in motility, adhesion, virulence or nutrient acquisition. Finally, lipoproteins, anchored to either leaflet

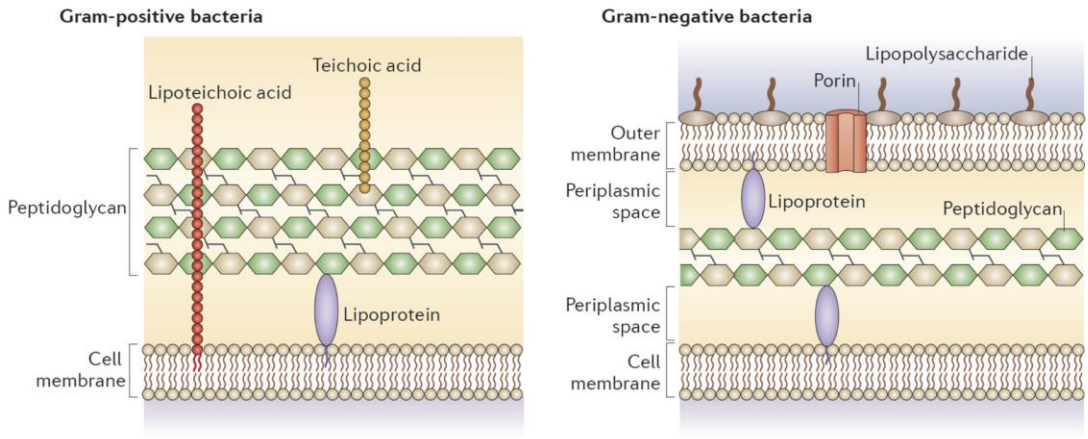


Figure 1. Cell envelope structures of Gram-positive and Gram-negative bacteria
The cell wall of Gram-positive bacteria is composed of a single lipid membrane surrounded by a thick layer of peptidoglycan in which are inserted teichoic and lipoteichoic acids. The cell wall of Gram-negative bacteria consists of two lipid membranes, the cell membrane or inner membrane and the outer membrane, separated by the periplasmic space in which lies a thin layer of peptidoglycan. The outer membrane is an asymmetric bilayer that contains phospholipids in the inner leaflet and lipopolysaccharide on its outer leaflet. Additionally, channel proteins such as porins facilitate exchange with the outside environment. Adapted from [2].

of the membrane by their lipid anchor, participate in a multitude of functions including nutrient acquisition, stress sensing or cell morphology[1].

While the OM is critical for the survival of Gram-negative bacteria, effectively acting as a selective diffusion barrier, its biogenesis poses several obstacles. First, all constituents of the OM are synthesized in the cytoplasm of the cell. This means that they first have to cross the IM before they can be transported towards the OM[1,3]. Second, the periplasm is a hydrophilic aqueous environment; yet, most OM components are hydrophobic in nature, which implies the necessity of a specific way of transport[1,4]. Finally, energy sources such as ATP or the proton motive force (PMF) are unavailable at the level of the OM, thus raising the question of how insertion and/or folding of OM components is accomplished[1,4].

In recent years, scientists have provided detailed insights into how Gram-negative bacteria build their complex cell wall. Although many questions have been answered (Fig. 2)[1,4-7], some remain unsolved while at the same time new organisms and ways of interaction are discovered, raising that many new questions. Indeed, while the early days of microbiology focused on bacterial

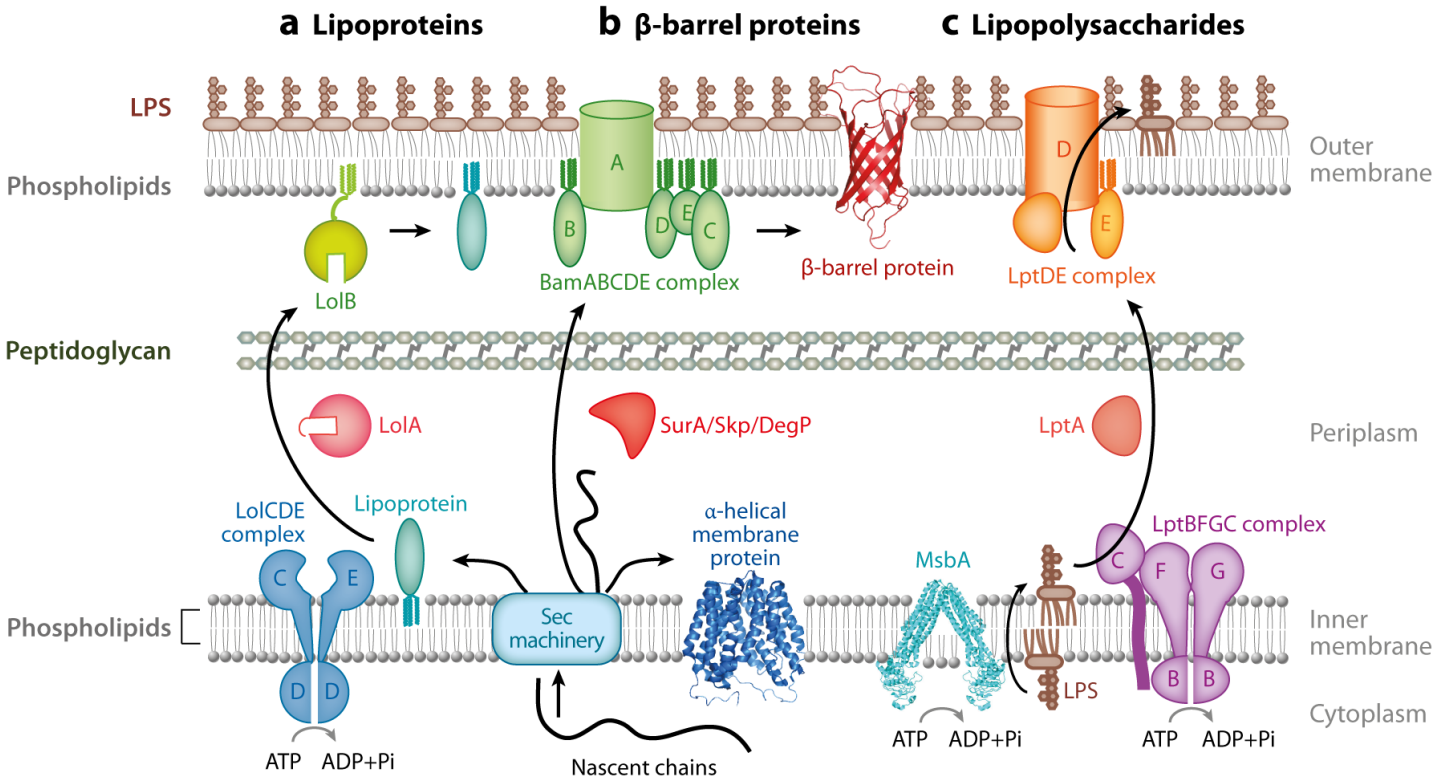


Figure 2. Biogenesis of the outer membrane of *E. coli*

All OM as well as periplasmic proteins are synthesized in the cytoplasm as precursors with a signal peptide at their N-terminus allowing their translocation across the IM by the Sec or Tat machineries. (a) Lipoproteins are transported to the OM by the Localization of lipoproteins (Lol) system. Lipoproteins are first extracted from the IM thanks to the LolCDE ABC transporter. They are then transferred to the periplasmic chaperone LolA that shuttles them across the periplasm before transferring them to the OM lipoprotein LolB. LolB finally inserts lipoproteins into the OM. (b) β-barrel proteins are inserted into the OM by a complex consisting of one β-barrel protein, BamA, and four lipoproteins, BamB, -C, -D and -E. Periplasmic chaperones such as SurA, Skp and DegP transport OM proteins across the periplasm and prevent their aggregation. (c) Similar to proteins, LPS is synthesized in the cytoplasm. LPS is therefore first flipped across the IM by the ABC transporter MsbA and then transported to the OM by the lipopolysaccharide transport (Lpt) system. LPS is extracted from the IM and transferred to LptC by the LptBFG ABC transporter. LptC then transfers LPS to the periplasmic protein LptA, which forms a bridge across the periplasm to deliver LPS to the OM complex formed by LptD and the lipoprotein LptE. Finally, LPS is flipped to the cell surface. The mechanism transporting phospholipids to the OM remains to be clarified. Adapted from [6].

pathogens, commensals have gained increasing interest over the last decade, especially in relation to human health and the benefits they provide their host[8]. Understanding OM biogenesis is thus not only crucial for the development of new antimicrobial compounds to fight bacterial pathogens but also to promote commensals that often rely on complex surface structures for their growth.

Biogenesis of the outer membrane

Taking the first hurdle - protein translocation across the inner membrane

IM proteins and preproteins (defined here as periplasmic, OM and lipoproteins) are all synthesized as precursors in the cytoplasm of the cell. To reach their final destination, these proteins face the major challenge of either inserting into or crossing the IM, an energetically unfavorable process[3,9-11].

IM proteins and preproteins therefore mostly rely on the Sec machinery to achieve this translocation step, an IM protein complex composed of six membrane proteins (SecD, -E, -F, -G, -Y and YidC) and interacting with two cytoplasmic proteins (SecA and -B) in *E. coli* (Fig. 3)[10]. Before translocation across the IM initiates, the Sec translocase first recognizes its substrates by the presence of a signal peptide located at the N-terminus of the preproteins (Fig. 4) [11-13]. This signal peptide (SPI) is composed of three distinct domains: the N domain, containing one to three positively charged amino acids; the H domain, a hydrophobic core region; and the C domain, containing a signal peptidase cleavage site[12]. Integral IM proteins do not contain a signal peptide and are instead recognized by the Sec translocase via their hydrophobic trans-membrane domains[13]. This difference in recognition also results in separate targeting routes towards the Sec machinery: IM proteins depend on SRP (signal recognition particule) to reach the Sec translocase while most preproteins rely on the homotetrameric SecB protein to do so (Fig. 3a and b)[3,9,14,15]. Additionally, IM proteins are inserted into the membrane co-translationally, inducing the formation of a ribosome nascent chain complex, while preproteins are generally transported post-translationally[3,9,16]. Since the Sec translocase only transports unfolded proteins, the SecB-preprotein interaction therefore not only initiates

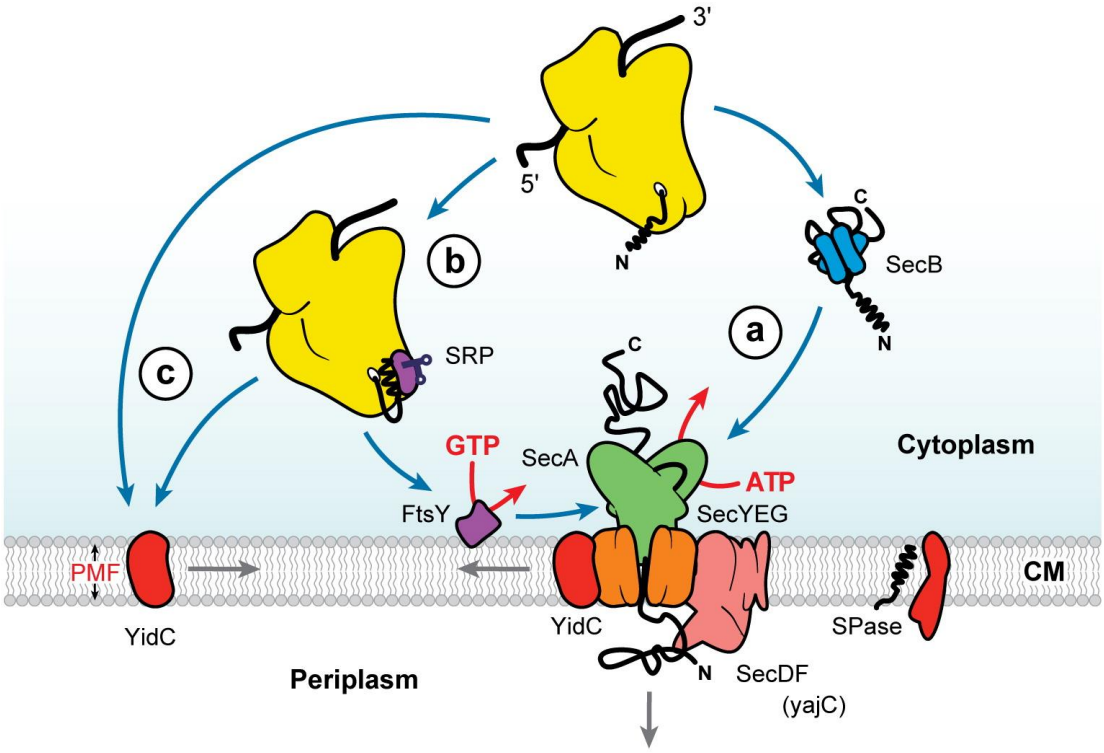


Figure 3. Protein translocation across the IM by the Sec machinery
The Sec machinery is an IM protein complex translocating unfolded preproteins across the IM (CM) and is composed of the motor protein SecA (green), a channel made of SecY, -E and -G (orange), and the accessory proteins SecDF (pink) and YidC (red). (a) Preproteins are mainly targeted to the Sec translocase post-translationally following the recognition of their signal peptide by the SecB chaperone (blue). They are then delivered to SecA that will thread them across the SecYEG channel using ATP as driving force. During or following translocation across the SecYEG channel, the membrane-bound signal peptidase (SPase) cleaves the signal sequence from preproteins at the periplasmic face of the membrane. (b) Membrane proteins are co-translationally targeted to the Sec translocase as ribosome-bound nascent chains by the SRP and the SRP-receptor FtsY (purple). (c) Some membrane proteins insert into the cytoplasmic membrane via YidC. Abbreviation: PMF, proton motive force. Adapted from [3].

the translocation of precursors across the IM but also assures that they remain in an unfolded, transport compatible state[17,18].

The SecB-precursor complex is initially targeted towards the SecAYEG IM translocase. The precursor first interacts with the SecA homodimer[19], inducing release of SecB upon ATP hydrolysis by SecA[20-22], and the unfolded precursor is then threaded through the SecYEG channel (Fig. 3)[9,23,24]. The driving force for this process is provided by SecA ATP hydrolysis and by the PMF[20,25-28]. During or after the translocation step, the signal peptide is cleaved off by signal peptidase I (SPase I), freeing the protein from the membrane and releasing it into the periplasm[12].

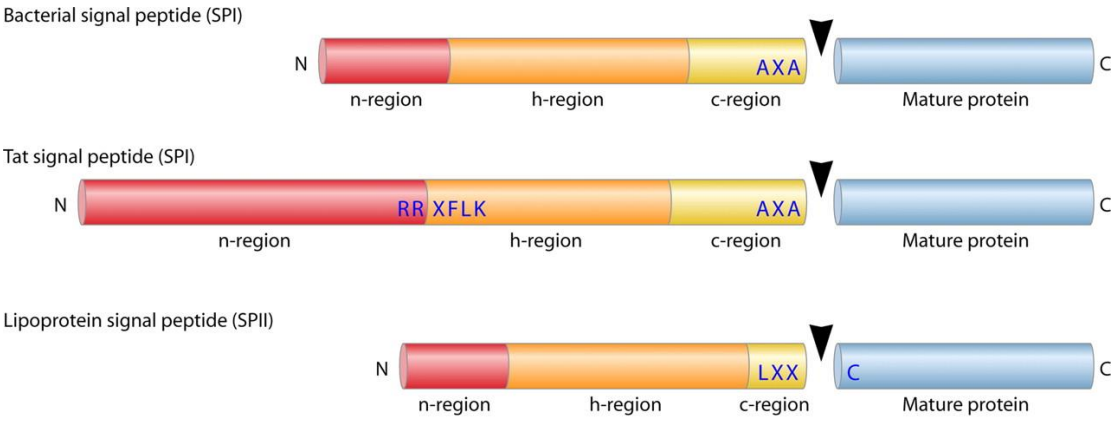


Figure 4. Bacterial preprotein signal peptides
Schematic representation of bacterial signal peptides. Shown are the bacterial (Sec-type) signal peptide, the twin-arginine (Tat) signal peptide and the lipoprotein signal peptide. The N, H, and C regions as well as the peptidase recognition sequences of the respective signal peptides are indicated. The cleavage site is marked with a black arrow. Adapted from [29].

While most *E. coli* preproteins are transported across the IM by the Sec machinery, a secondary route, the twin-arginine translocase (Tat) pathway, exists (Fig. 5)[30]. The Sec and Tat pathways have three main differences. First, the signal peptide of this subset of preproteins contains two adjacent arginines (Fig. 4), hence the name of the system. The second, most remarkable difference is the fact that the Tat system is able to transport (partially) folded proteins rather than unfolded polypeptide chains. Finally, the PMF rather than ATP allows the translocation of Tat substrates[31,32].

Tat transport is accomplished by a multimeric complex made of three IM proteins, TatA, -B and -C. Similar to the Sec pathway, Tat transport is initiated when the signal peptide of the precursor is recognized by a TatBC oligomer (Fig. 5 step 1)[33,34]. This triggers recruitment and oligomerization of TatA, leading to the formation of a membrane pore and the assembly of a complete TatABC translocase (Fig. 5 step 2)[33,35-39]. While this assembly process requires the PMF, the subsequent preprotein translocation seems to depend on TatA only (Fig. 5 step 3). After completion of the transport, the signal peptide is cleaved off by SPase I[40] and the TatABC pore disassembles (Fig. 5 step 4)[35].

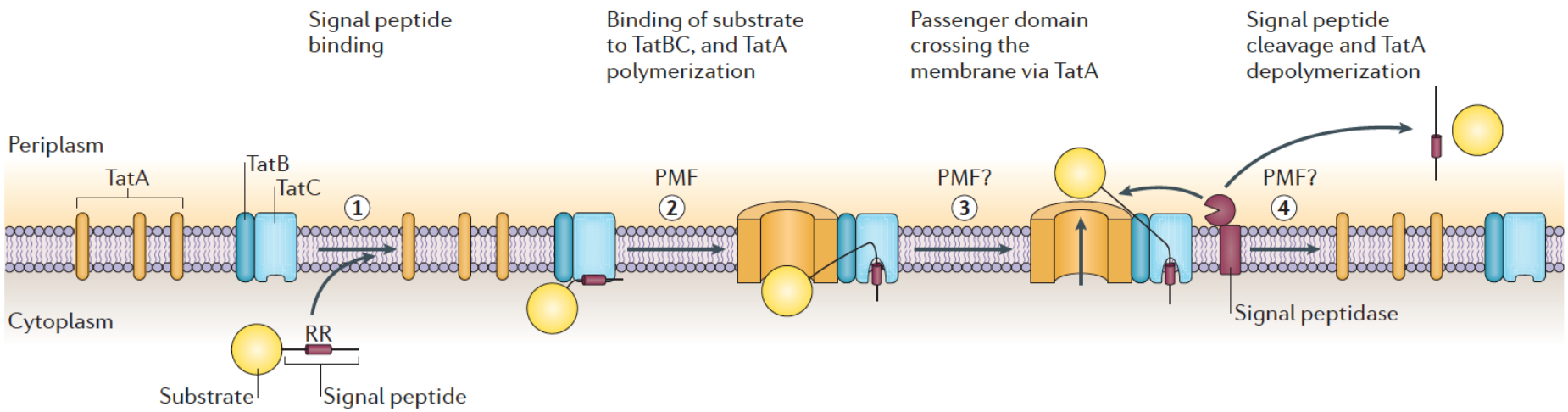


Figure 5. Protein translocation across the IM by the Tat machinery

The Tat machinery is an IM protein complex translocating folded preproteins across the IM and is composed of three membrane spanning proteins, TatA, -B and -C. Transport is initiated upon recognition of the signal peptide of a substrate by the TatBC subcomplex (step 1). This triggers recruitment and oligomerization of TatA in a PMF dependent manner, leading to the formation of a complete Tat translocase (step 2). Transport of substrate (the passenger domain) is then believed to be achieved in an energy-independent way by TatA (step 3). After completion of transport, the TatABC complex disassembles and the signal peptide is cleaved by the SPase I (step 4). Abbreviation: PMF, proton motive force; SPaseI, signal peptidase I. Adapted from [41].

The periplasm – a no-man's land for membrane proteins

Following their cytoplasmic synthesis and translocation through the IM, preproteins do not all share the same fate. Periplasmic proteins remain soluble and fold either spontaneously or with the help of chaperones and/or folding catalysts (*i.e.* for disulphide bond formation)[11,42,43]. In contrast, OMPs and lipoproteins, hydrophobic in nature due to either their amino acid composition or their lipid anchor, cannot traverse the hydrophilic periplasm to reach the OM without assistance[4]. Additionally, while lipoproteins most likely fold similarly to periplasmic proteins, OMPs have to be maintained in an unfolded state till they reach the OM. Gram-negative bacteria therefore evolved specific targeting routes and transport machineries to circumvent these problems[32].

The journey of outer membrane proteins across the periplasm

OMPs carry out a diverse number of vital functions in the OM and allow the cell to interact and to exchange with the outside environment. In *E. coli*, the most abundant OMPs are porins, trimeric transporters with relatively low specificity that allow diffusion of small compounds of up to 600 Da[44]. Other OMPs are involved in OM biogenesis[5,45-47], peptidoglycan binding[48,49], (active) transport[50-52], efflux[53] etc.

While integral IM proteins are characterized by the presence of trans-membrane α-helices, OMPs are almost exclusively composed of β-strands and adopt a so-called β-barrel conformation[54,55]. A β-barrel can be seen as a water-filled cylinder made of an antiparallel β-sheet closed by interactions between its first and last β-strands. The strands are connected to each other by short linkers on the perisplamic side of the barrel while long extracellular loops link the strands on the outside, often folding into and closing the barrel pore. Since β-barrels are integral membrane proteins, they have to insert into a lipid environment. This is achieved thanks to the distribution of hydrophobic residues throughout the β-strands, creating a hydrophobic surface on the outside of the barrel. On the opposite, residues facing the barrel interior are often polar, allowing the entry of water and other substrates[54,55].

After their translocation through the IM, the hydrophobic domains of OMPs have to be protected in order to prevent misfolding or aggregation in the aqueous periplasm[43]. This is achieved by several periplasmic chaperones, the most prominent ones being SurA, Skp and DegP in *E. coli*.

SurA (survival protein A) has been extensively characterized since its discovery and, at least in *E. coli*, seems to be the main OMP chaperone[56-58]. In brief, SurA is able to bind unfolded OMPs[59-61], it accelerates the folding process of OMPs[62-64], and its deletion activates the σ^E stress response (characteristic of OMP misfolding)[63,65], leading to a decreased OMP gene transcription and resulting in lower OM density[56]. Furthermore, SurA interacts with BamA, the central component of the OMP assembly machinery (see following sections), highlighting its function as general OMP chaperone[56]. However, the perhaps most critical function of SurA in *E. coli* is assisting the folding of the LPS transporter LptD (see following sections)[57]. Indeed, the effect of a *surA* and a *lptD* mutation is similar and results in strong OM permeability and growth defects. Furthermore, while *lptD* transcription increases upon *surA* deletion, the total LptD amount in the membrane is still decreased as compared to a wild type strain, indicating that there is no alternative route for LptD transport[57]. SurA is therefore highly important for OM biogenesis as it is required for OMP transport and indirectly acts on LPS insertion into the OM.

Skp (seventeen kilodalton protein) has a broad spectrum of substrates[66-68] and was shown to bind unfolded OMPs and to prevent aggregation of lysozyme *in vitro*[66,69]. In addition, its deletion causes moderate activation of the σ^E stress response and a slight decrease in OMP levels, prompting its role as OMP chaperone[65]. The chaperone function of Skp was further demonstrated by an elegant genetic approach that showed that a *surA skp* double mutant presents a synthetic lethal phenotype[56,70]. This indicates that although Skp is not essential in presence of SurA, they do have similar functions and overlapping sets of substrates. Recent work indeed demonstrates that Skp is involved in LptD biogenesis, although in a different way than SurA since overexpression of SurA did not prevent the *skp* deletion phenotype[71]. The crystal structure of Skp has been solved more than 10 years ago and shows that Skp forms a homotrimer with a shape similar to a jellyfish[69]. The “head” is composed of 12 β-strands

forming the hydrophobic core of the protein while three α-helical "tentacles" expand from the body. These tentacles are quite flexible and allow Skp to interact with substrates much bigger than itself in a one to one complex, ranging from 20 to 90 kDa[67,69]. Additionally, the Skp structure revealed a site that is proposed to be a LPS binding site[69], which confirmed earlier observations that Skp interacts with LPS[72,73]. However, further investigation showed that this interaction is unspecific and is likely related to the role of Skp in the folding of LptD[71,74].

DegP (also known as HtrA) is a periplasmic serine protease that also possesses chaperone activity[75]. DegP is mainly seen as a stress response protein, preventing unfolded or misfolded OMPs to accumulate in the periplasm and to form toxic aggregates[75,76]. Its chaperone activity however, which is independent of the protease activity[76], remains poorly understood. Nevertheless, as for Skp, a double *surA degP* mutant presents a synthetic lethal phenotype[56,70], showing that DegP is essential for proper folding of at least a subset of OMPs. It is therefore assumed that SurA and Skp/DegP work in parallel, partially overlapping pathways.

Inserting OMPs into the membrane – the Bam machinery

Following synthesis and transport across the periplasm, OMPs have to face one final challenge: insertion into an asymmetric lipid bilayer. While OMPs have been shown to fold spontaneously into membranes *in vitro*, the kinetic of the process is far too slow to faithfully reflect the *in vivo* condition[77]. This thus suggested the need of a folding catalyst, which was indeed discovered a few years later: the Bam (Beta Barrel Assembly Machinery) complex (Fig. 6)[4,78].

The Bam complex is composed of five proteins: BamA, a β-barrel with a N-terminal periplasmic domain and BamB, -C, -D and -E, OM anchored lipoproteins[4,58,78]. BamA homologs have been identified in all Gram-negative bacteria[79-82] as well as bacterial-derived compartments such as mitochondria and chloroplasts, highlighting its pivotal role in membrane biogenesis[82]. This is further demonstrated by the fact that BamA is essential in all tested bacteria to date[79,83].

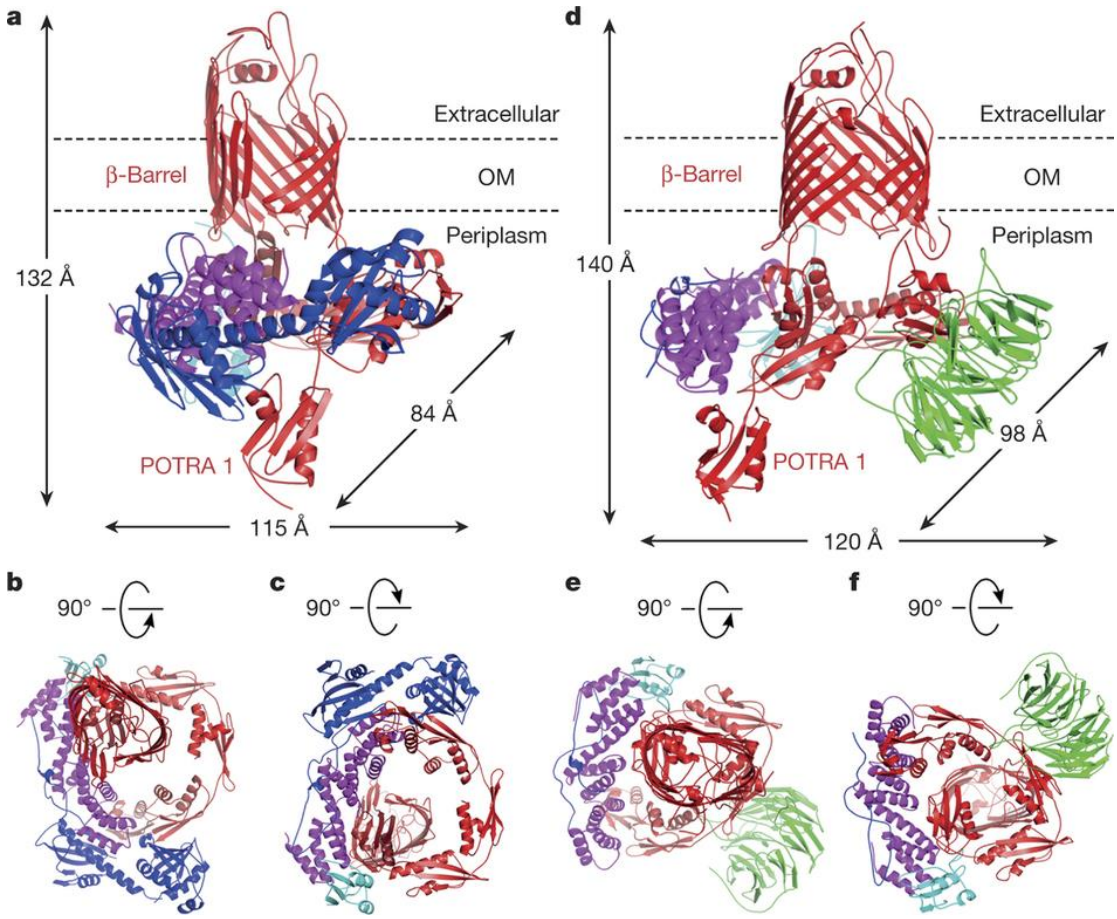


Figure 6. Structure of the *E. coli* β-barrel assembly machinery
Two structures of the *E. coli* Bam complex (BamACDE and BamABCDE) have been solved, giving novel insights into OM β-barrel insertion. The BamA (red) barrel is embedded in the OM while its N-terminal domain extends into the periplasm, forming a circular structure with lipoproteins BamB (green), BamC (blue), BamD (magenta) and BamE (cyan), resembling a top hat. (a–c) Structure of the BamACDE complex, viewed from the membrane plane (a), extracellular side (b) and periplasm (c). The dimensions of the complex are indicated. In this complex, the periplasmic side of the ring is fully closed, while the barrel is open laterally towards the membrane. (d–f) Structure of the BamABCDE complex, viewed from the membrane plane (d), extracellular side (e) and periplasm (f). In the presence of all Bam constituents, the periplasmic ring is open, enabling substrate entry, while the barrel is capped by extracellular loops and remains closed. Due to high flexibility of the protein, BamC, although present in the BamABCDE crystal, could not be mapped clearly. Adapted from [5].

The BamA C-terminus has a classical β-barrel structure while its N-terminus is composed of five structural repeats of a so-called POTRA (polypeptide translocation associated) domain[84-86], numbered P1 to P5 starting from the N-terminus[84]. These POTRA domains are involved both in substrate recruitment and functional assembly of the Bam complex. Indeed, BamA

interacts with all other components of the complex and essentially serves as docking platform to recruit both the Bam lipoproteins and the nascent OMPs. Although structurally similar, the POTRA domains nevertheless differ both in amino acid sequence and in function; not all POTRA domains are essential and they do not all mediate interaction with the same Bam lipoproteins[58,84,87]. Of note, some residues in the P1 POTRA domain favor interaction with SurA, reinforcing the core function of BamA in OMP biogenesis[88].

All Bam lipoproteins interact with BamA, however they do not do it in the same way[78]. Prior to interaction, BamC, -D and -E form a subcomplex in order to bind BamA. On the other hand, BamB associates with BamA independently from the remaining lipoproteins. As a result, absence of the BamCDE subcomplex does not affect BamA-B interaction and vice versa[58,84,87]. Deletion of any of the Bam lipoproteins has at least a slight effect on β-barrel assembly[80,87,89-91], but only BamD is essential in *E. coli*; depletion of BamD basically induces OMP biogenesis arrest[90]. However, mutation of *bamC* in combination with *bamE* induces strong OMP defects and substantial σ^E stress response while mutation of *bamB* in combination with *bamE* is lethal[87]. This points to the fact that while not essential, all Bam lipoproteins are important *in vivo* for correct OMP folding.

Several structural studies have been carried out over the years, giving insight into single or multiple subunits of the Bam machinery[84,92-100], but the structure of the complete BamABCDE complex has only been solved very recently (Fig. 6)[5,45]. This work revealed two main findings. First, rather than expanding into the periplasm, the BamA POTRA domains tightly interact with all Bam lipoproteins, leading to the formation of a periplasmic ring underneath the inner leaflet of the OM. The structure of the complete complex therefore looks like a hat, the BamA β-barrel inserted in the OM being the top and the POTRA domains and lipoproteins forming the brim (Fig. 6).

Second, these studies revealed that the complex exists in two different conformations, an inward-open and a lateral-open state. The inward-open conformation shows that the BamABCDE complex is open on the periplasmic side, supposedly allowing entry of unfolded OMPs, while the extracellular loops of BamA close the barrel to the extracellular milieu. On the other hand, the lateral-open conformation (obtained with a BamACDE complex missing BamB)

shows that the periplasmic access pore is completely closed due to the rotation of the periplasmic ring. In addition, the extracellular loops L1 to L3 are displaced, resulting in strands β1 to β6 of the β-barrel to move away from the barrel core, opening the inside of the barrel to the OM. Site-specific cross-linking confirmed that this lateral gate opening is vital for cell viability and is not an artifact due to the absence of BamB in this complex. The structure observed with the BamACDE complex thus represents a conformation of the Bam machinery that exists *in vivo*. The combination of the opening of the barrel towards the membrane and the rotation of the ring is therefore proposed to provide the driving force for OMP membrane insertion in the absence of an energy source such as ATP. This newly identified ring architecture and the two conformational states also highlight the important role of each subunit of the complex for optimal protein folding and insertion[5,45].

Synthesis and localization of bacterial lipoproteins

Lipoproteins are found in both Gram-negative and Gram-positive bacteria and fulfill many different cellular functions, including membrane biogenesis and homeostasis[90,101-106], cell division[107], substrate transport[108], drug efflux[109], motility[110], adhesion and implication in pathogenicity[111]. Lipoproteins are therefore essential components for the survival of bacteria and their proper localization is critical[112,113]. Lipoprotein synthesis has been mostly studied in *E. coli* and extensive knowledge about the maturation and the transport machinery of lipoproteins has been gained in the last three decades (Fig. 7)[6]. *E. coli* encodes approximately 90 lipoproteins, most of which are facing the periplasm anchored into the IM or OM[6,114].

Lipoproteins are synthesized in the cytoplasm as precursors called pre-prolipoproteins (Fig. 7a). Similar to periplasmic and OM proteins, pre-prolipoproteins have an N-terminal located signal peptide (called SPII) (Fig. 4) [115] of around 20 amino acids that allows their recognition by and transport through the Sec machinery[116,117]. Alternatively, some lipoproteins are transported by the Tat system[118-121]. The SPII signal peptide is, as the SPI, composed of three distinct domains: N, H and C. The C domain additionally

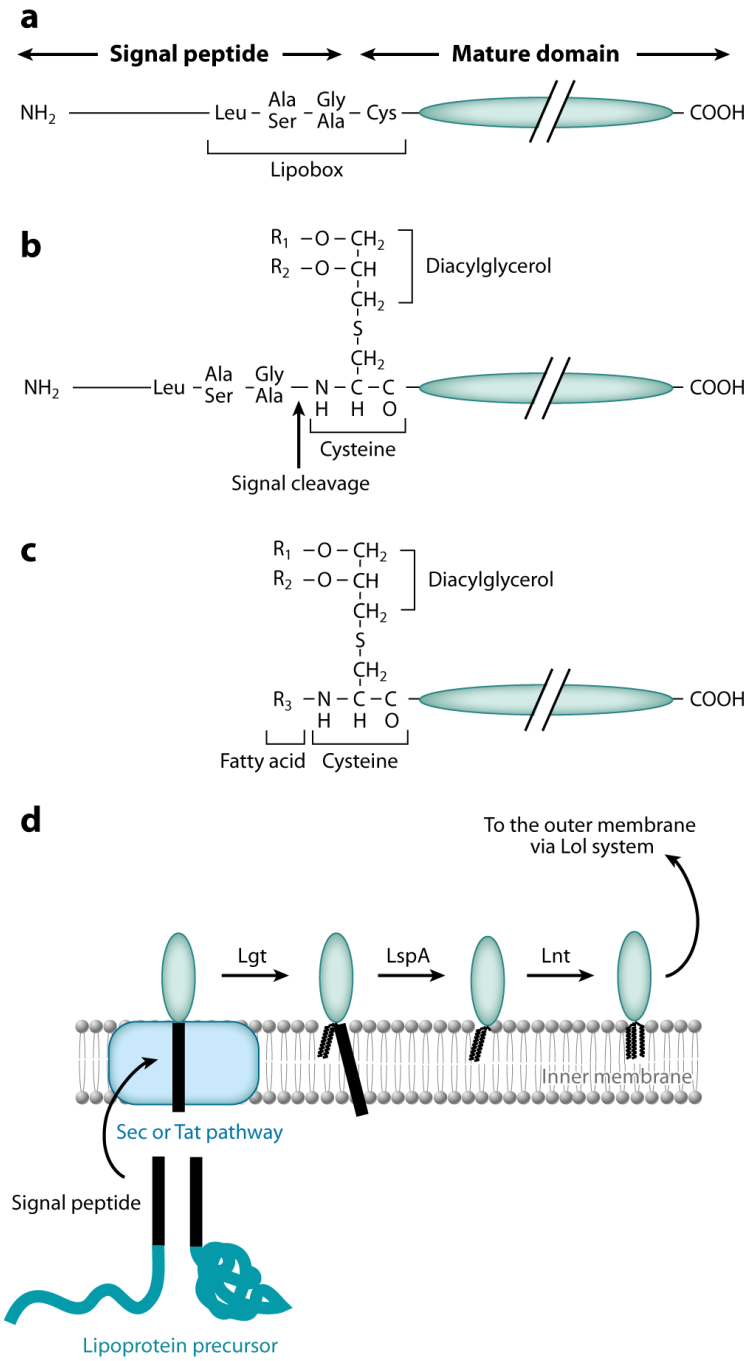


Figure 7. Lipoprotein maturation in bacteria

Following their translocation across the IM, lipoprotein precursors are sequentially processed by three enzymes in order to complete their maturation. (a) A representative lipoprotein precursor. The lipobox containing the conserved cysteine to which the lipid anchor will be attached is indicated. (b) In the first step of maturation, Lgt attaches a diacylglycerol moiety to the side chain of the conserved cysteine of the lipobox. The signal peptide is subsequently removed by LspA, leaving the lipidated cysteine as first amino acid of the mature lipoprotein. In Gram-positive bacteria, this completes the lipoprotein synthesis process. (c) In Gram-negative bacteria, removal of the signal peptide is followed by attachment of a third fatty acid to the conserved cysteine by Lnt, rendering a mature tri-acylated lipoprotein. (d) Summary of the lipoprotein maturation steps in Gram-negative bacteria. Adapted from [6].

contains a conserved consensus sequence, the lipobox [LVI][ASTVI]C, overlapping with the peptidase cleavage site and containing an invariant cysteine residue to which the lipid moieties are attached in the subsequent steps of maturation (Fig. 7)[122-124].

Following translocation across the IM, processing of pre-prolipoproteins into mature forms takes place on the periplasmic side of the IM[6]. First, lipoprotein diacylglyceryl transferase (Lgt) catalyzes the covalent attachment of a diacylglycerol moiety derived from the IM to the side chain of the conserved cysteine of the lipobox[125,126], rendering a prolipoprotein (Fig. 7b). Addition of this lipid anchor serves to maintain the lipoprotein attached to membrane for the following steps of maturation. Lipoprotein signal peptidase (LspA) then cleaves off the signal peptide immediately upstream of the lipidated cysteine[127-129], leaving this amino acid as the new N-terminal residue of mature lipoproteins (Fig. 7b). For most Gram-positive bacteria, this step completes the lipoprotein synthesis process. However, in Gram-negative[130-132] and some Gram-positive bacteria[121,133,134], lipoprotein N-acyl transferase (Lnt) catalyzes the attachment of a third fatty acid to the cysteine, rendering a mature tri-acylated lipoprotein (Fig. 7c).

While in Gram-positive bacteria mature lipoproteins remain attached to the cytoplasmic membrane, in Gram-negative bacteria they can be anchored either to the IM or the OM (Fig. 8)[6]. Following maturation, OM lipoproteins therefore have to be transported through the aqueous environment of the periplasm to reach the inner leaflet of the OM. Since lipoproteins are overall hydrophobic due to their lipid anchor and because the periplasm is a hydrophilic environment, this transport step requires the dedicated localization of lipoprotein (Lol) machinery, composed of five essential proteins, LolA, -B, -C, -D and -E (Fig. 8)[4,6].

The LolC, -D and -E proteins form an IM ABC transporter in a stoichiometry of 1:2:1 responsible for the extraction of OM lipoproteins from the IM (Fig. 8)[135-137]. LolCDE is therefore considered an atypical ABC transporter in the sense that rather than catalyzing substrate transport across a membrane, it induces the release of substrate from the membrane. LolD is a nucleotide-binding subunit with Walker A and B motifs; LolC and -E are IM proteins having

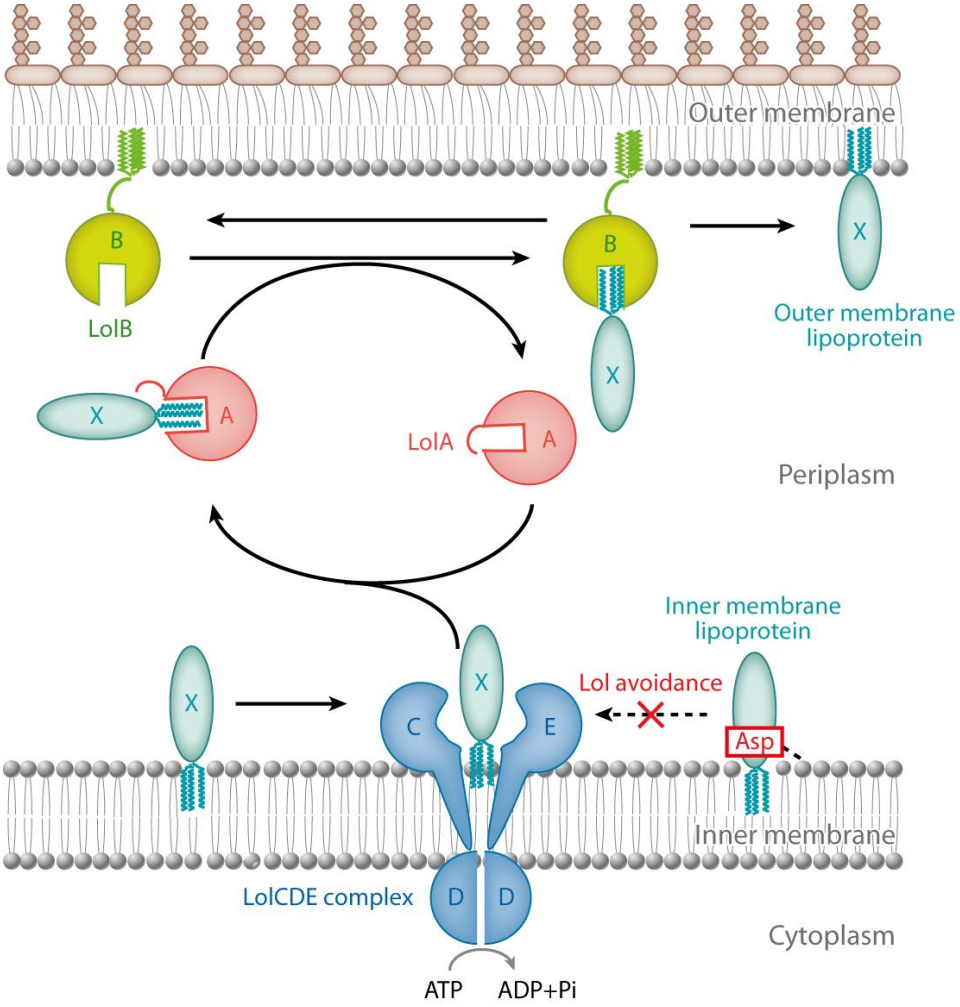


Figure 8. The localization of lipoprotein (Lol) system of Gram-negative bacteria
Unlike IM lipoproteins, lipoproteins destined to the OM have to be transported across the aqueous environment of the periplasm. This is achieved in several steps by the Lol system, composed of five proteins, LolA to -E. First, OM lipoproteins are extracted from the IM by the ABC transporter LolCDE. They are then transferred to the periplasmic chaperone LolA that shuttles them across the periplasm. Finally, LolA delivers its cargo to LolB, which then inserts the lipoprotein into the OM. IM lipoproteins avoid interaction with the LolCDE complex by a so-called lol-avoidance signal (generally Asp at position +2 in *E. coli*) and remain therefore attached to the IM. Adapted from [6].

each four membrane spanning domains as well as one big periplasmic loop[138]. These loops contain amino acids with hydrophobic side chains predicted to form a hydrophobic cavity, allowing LolC and -E to interact with the lipid anchor of lipoproteins[6]. Lipoprotein release from the IM occurs in three distinct steps: i) the lipoprotein interacts with LolE, increasing the affinity of LolD for ATP as well as of LolC for LolA[139]; ii) ATP binds to LolD while LolC interacts with LolA; iii) LolD hydrolyses ATP, thereby weakening the interaction between LolE and the

lipoprotein, which causes the transfer of the lipoprotein to LolA (Fig. 9 step 1 to 3)[140]. Although it is currently not well established how this transfer takes place, it is predicted to be similar to the affinity driven mouth-to-mouth model for the LolA-LolB lipoprotein transfer (outlined below)[139].

LolA is a periplasmic chaperone that shuttles lipoproteins from the periplasmic side of the IM to the periplasmic side of the OM in a one to one complex (Fig. 8)[141,142]. The structure of LolA has been solved several years ago, consisting of an incomplete β-barrel made of eleven antiparallel β-strands closed by a lid formed by three α-helices (Fig. 9b)[143,144]. The incomplete β-barrel and the lid form a closed hydrophobic cavity containing aromatic residues, allowing LolA to accommodate the lipid anchor of lipoproteins. In addition, the C-terminus contains a short helix and a twelfth β-strand that together form a loop critical for lipoprotein localization. Indeed, this loop region prevents interaction of LolA with membrane lipids, therefore assuring that no retrograde transfer of lipoproteins from LolA to the IM is possible[145]. In its unloaded state, the cavity of LolA is closed, while interaction with LolCDE induces opening of the lid, thus allowing the transfer of a lipoprotein from LolCDE to LolA[146]. The soluble LolA:lipoprotein complex then crosses the periplasmic space to reach the OM where LolB inserts the lipoprotein into the membrane.

LolB, an OM anchored lipoprotein (Fig. 8), is structurally very similar to LolA, being composed of an incomplete β-barrel closed by a lid[102]. However, two significant differences distinguish LolA and LolB. First, LolB has no additional C-terminal loop, which allows the protein to interact with phospholipids and therefore to insert lipoproteins into the OM. The precise mechanism by which lipoproteins are inserted into the OM is yet poorly understood, but evidence that a conserved leucine residue at position 68 of LolB initiates membrane targeting has been gained[147]. Second, the amino acids forming the hydrophobic cavity of LolB (Leu and Ile) have more flexible side chains than the aromatic residues forming the LolA cavity. This results in a difference of affinity that allows one-way, energy-independent transfer of lipoproteins from LolA to LolB (Fig. 9 step 4)[148,149]. Moreover, cross-linking experiments showed that LolA and -B partially overlap during lipoprotein transfer in what is described as mouth-to-mouth model, thus connecting the entrances of the hydrophobic cavities (Fig. 9b)[139].

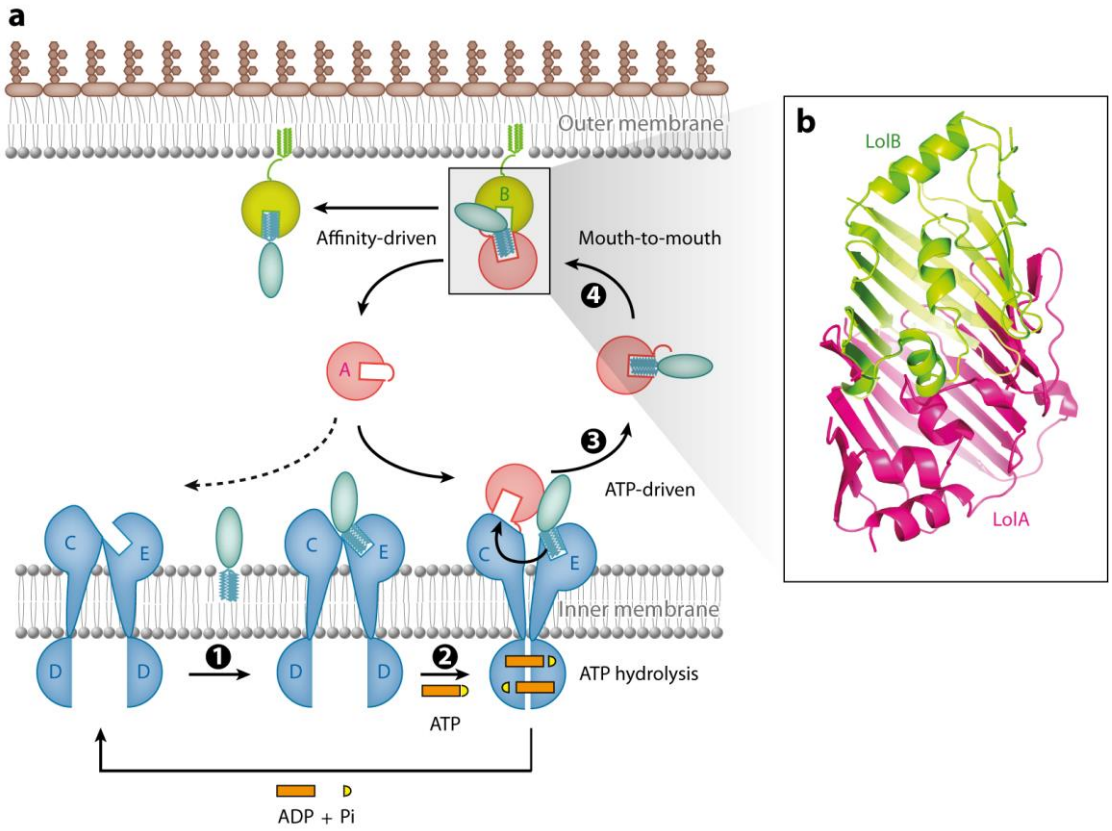


Figure 9. Molecular details underlying the lipoprotein transport by the Lol system
(a) OM lipoproteins are extracted from the IM by their interaction with LolE, which leads to an increased affinity of LolA for the LolCDE complex (step 1). After LolA binding to LolC (step 2), ATP hydrolysis induces opening of the LolA cavity, which can now accept the lipoprotein from LolE (step 3). After crossing the periplasm, LolA and LolB interact with each other in a mouth-to-mouth manner allowing the transfer of the lipoprotein from LolA to LolB (step 4). Finally, LolB insert the lipoprotein into the OM. (b) Structural view of LolA-LolB interaction. Adapted from [6].

Whether a lipoprotein is destined to the IM or the OM is determined by the nature of the amino acids immediately downstream of the lipidated cysteine, referred to as +1 cysteine. These amino acids can prevent interaction between the lipoprotein and the LolCDE complex[150], thereby acting as lol-avoidance signal. In *E. coli* and closely related species, this is referred to as the "+2 rule" because Asp at position +2 is the most common amino acid serving as lol-avoidance signal[151]. Additional work showed that other amino acids at position +2 could serve as lol-avoidance signal, but they are much less frequent in native *E. coli* lipoproteins[152,153]. The impact of the +3 residue on lipoprotein localization has also been investigated, showing that some amino acids at this position can either weaken or strengthen the lol-avoidance properties of Asp[153]. However,

these rules apply only to Enterobacteriaceae[154], as for example in *Pseudomonas aeruginosa* the lol-avoidance signal is determined by the nature of the +3 and +4 residues[155-157]. The precise mechanism by which Asp prevents lipoproteins from interacting with LolCDE is still poorly understood, although the negative charge of its side chain as well as its potential interaction with membrane lipids seems to be crucial[158].

Although the enzymes involved in lipoprotein processing are well conserved in both Gram-positive and Gram-negative bacteria (with the exception of Lnt), the components of the lol machinery are completely absent from Gram-positive bacteria and only partially conserved within Gram-negative bacteria. This is especially true for LolB, which is only found in β- and γ-Proteobacteria[159]. It remains therefore unknown how other bacteria outside of these phyla insert lipoproteins into their OM. Interestingly, the absence of a LolB-like protein might be linked to the fact that some species abundantly expose lipoproteins at their surface[160,161], unlike what is observed in *E. coli*[6] and most Proteobacteria. Surface transport of these lipoproteins could indeed require a protein functionally able to replace LolB but that would also possess additional functions.

Lipopolysaccharide synthesis and transport

While proteins represent the overall main component of the OM, lipopolysaccharide (LPS) is the main constituent of the outer leaflet of the OM and is responsible for the barrier function of the membrane[7]. Indeed, LPS forms a thigh network that is impermeable to most compounds thanks to the interaction between its negatively charged phosphate groups and divalent cations, bridging adjacent LPS molecules[44]. This ensures efficient protection of the bacterial cell from harmful substances, such as antibiotics. Additionally, LPS provides protection against complement killing and macrophages[162]. The mechanisms of LPS synthesis and transport have been elucidated thanks to model organisms such as the γ-Proteobacteria *Escherichia coli* and *Salmonella enterica* or the β-Proteobacterium *Neisseria meningitidis.* LPS is composed of three parts: lipid A, a core oligosaccharide and an O-antigen (Fig. 10)[163].

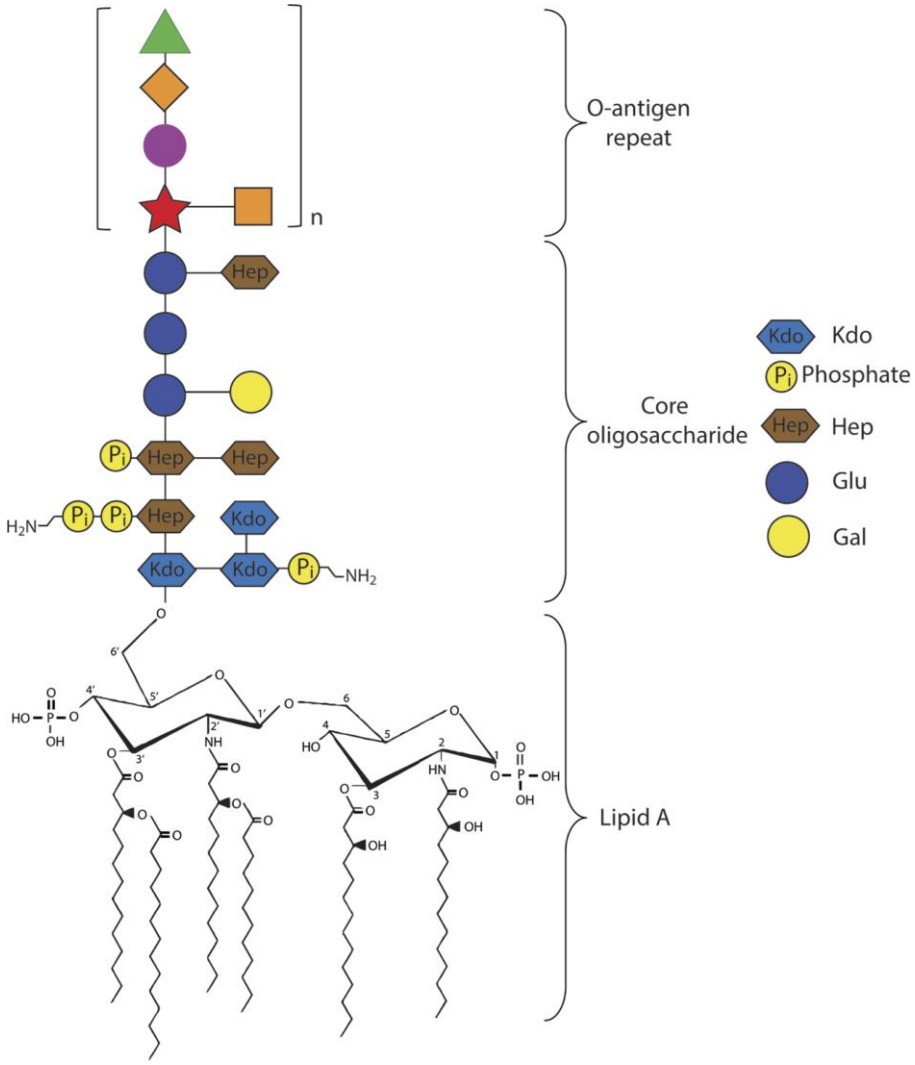


Figure 10. Composition of a typical lipopolysaccharide (LPS) molecule
LPS is composed of three distinct structural elements: lipid A, the core oligosaccharide and the O-antigen. Lipid A is composed of a disaccharide backbone to which fatty acids of various lengths are attached. The core saccharide is divided in inner and outer core. The inner core is well conserved and composed of Kdo and heptose while the outer core can be more variable, containing in this example glucose and galactose. Finally, the O-antigen, if present, is the most variable part of LPS and composed of a variable numbers of sugar repeats. The lipid A and core structures depicted correspond to those of *E. coli* K-12. Kdo, 3-deoxy-D-*manno*-oct-2-ulosonic acid; Hep, heptose; Glu, glucose; Gal, galactose. Adapted from [7].

Lipid A represents the membrane anchor of LPS. It is composed of a disaccharide backbone, usually made of N-acetylglucosamine, to which fatty acids of various lengths are attached (Fig. 10)[163]. Additionally, the 1 and 4' positions of the disaccharide are generally phosphorylated, which, in the presence of divalent cations (Mg^{2+}, Ca^{2+}), allows tight packing and crosslinking of adjacent LPS molecules[44,163]. Although the basic architecture of lipid A is well

conserved, its precise composition can vary from one species to another, including the length, number and composition of fatty acids attached to the disaccharide backbone as well as its phosphorylation[164]. The core oligosaccharide of LPS is often divided into inner core (attached to lipid A) and outer core (attached to the O-antigen). The inner core, composed of Kdo (3-deoxy-D-*manno*-oct-2-ulosonic acid) and heptose (L-*glycero*-D-*manno*-heptose), is relatively well conserved within a species, while the outer core is much more variable (Fig. 10)[163]. As for lipid A, phosphorylation of Kdo and heptose is essential for the barrier function of the OM as it allows strong lateral interactions between LPS molecules. The O-antigen, absent in a number of species, is the least conserved region of LPS and is composed of various repeats of one to six sugars units (Fig. 10)[165].

Lipid A linked to the core oligosaccharide (lipid A-core) and the O-antigen are synthesized independently from each other in the cytoplasm[163]. The Lipid A-inner core is first synthesized by the Raetz pathway, followed by the attachment of additional glycan residues to form the complete Lipid A-core[163]. This moiety is then flipped to the periplasmic leaflet of the IM by the MsbA ABC transporter[166-170]. In parallel, the O-antigen is synthesized by either of two pathways. Using undecaprennyl phosphate as lipid scaffold, the O-antigen repeats are generated on the cytoplasmic side of the IM and are then flipped across the membrane by Wzx. This is followed by their polymerization into complete O-antigen molecules on the perisplasmic side of the IM[163,165]. Alternatively, polymerization takes place in the cytoplasm and the complete O-antigen is then transported across the IM by the ABC-transporter composed of Wzm and Wzt[163,165]. Finally, WaaL attaches the O-antigen to the lipid A-core on the periplasmic side of the IM, rendering the final LPS molecule.

As for lipoproteins, the presence of a lipid anchor in LPS prevents it to cross the periplasm on its own. Thus, the transport of LPS from the IM to the cell surface is performed by the LPS transport (Lpt) machinery, composed of 7 proteins, LptA to G (Fig. 11)[7].

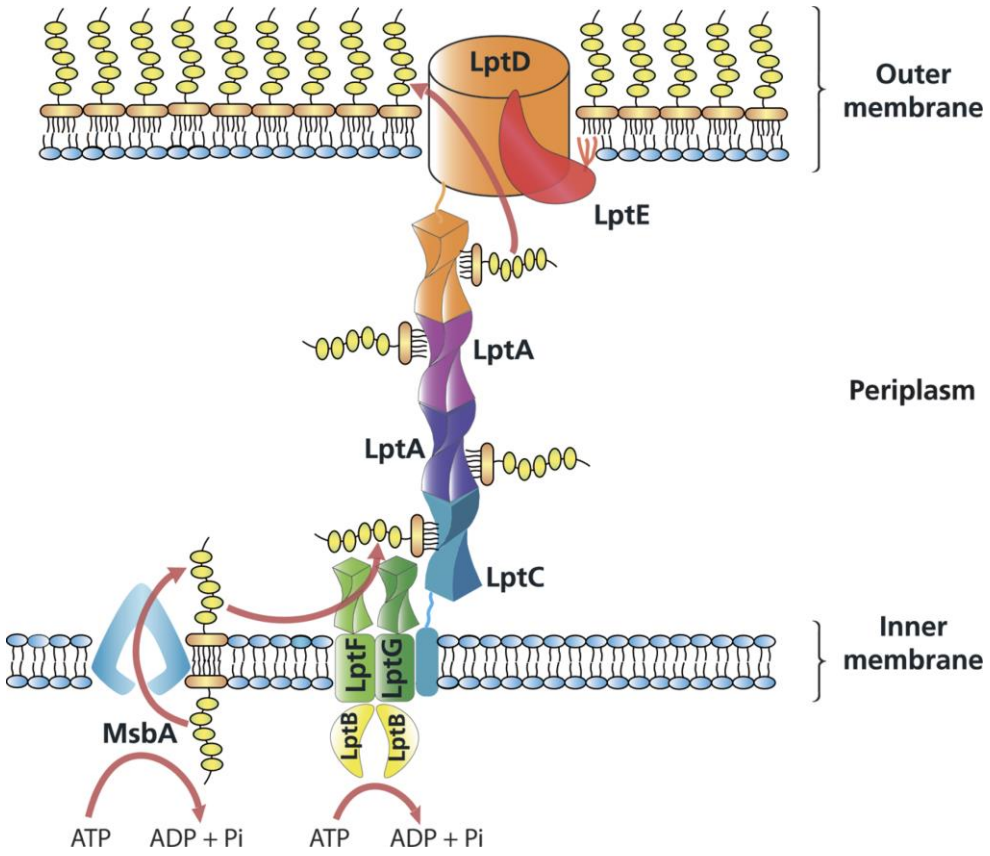


Figure 11. The LPS transport (Lpt) machinery of Gram-negative bacteria
Following flipping across the IM by MsbA, LPS is extracted from the membrane by the LptBFG ABC transporter. It is then transported in a continuous stream via a periplasmic bridge made of LptC, LptA and the periplasmic domain of LptD to the OM. LptD, in combination with LptE, then insert LPS into the outer leaflet of the OM. The entire transport process from the IM to the cell surface is powered by LptB-mediated ATP hydrolysis. Adapted from [7].

LptA, belonging to the OstA family of proteins, is a periplasmic chaperone able to bind LPS[171-173], which consists of 16 consecutive anti-parallel β-strands arranged into a β-jellyroll conformation[174]. Cross-linking experiments showed that it is inside this groove that LPS is binding[175]. *In vitro* data also show that LptA is able to oligomerize[174,176,177]. LptB[172] is a cytoplasmic nucleotide-binding protein with characteristic Walker A and B motifs[178,179]. Due to the absence of trans-membrane domains anchoring it to the IM, LptB is in a 2:1:1 complex with LptF and -G to form a functional ABC transporter[7,171,180,181]. To date, no structural data are available for LptF and -G, but predictions suggest that the periplasmic domains of both proteins have a similar fold than LptA[182]. LptC, which interacts with the LptBFG transporter[180], has a N-terminal transmembrane helix anchoring it to the IM and a periplasmic C-terminal domain[183,184]. Interestingly,

this C-terminal domain is annotated as OstA-like and indeed has a similar structure than LptA, allowing it to bind LPS in a similar way[175,184,185]. Finally, LptD and -E form a one to one complex in the OM that transports LPS to the cell surface[186]. LptD[187,188] is the largest monomeric β-barrel reported to date, composed of 26 anti-parallel β-strands[46,47]. Its periplasmic N-terminal domain has structural similarity to LptA and -C and contains hydrophobic residues, suggestive of its ability to bind the lipid moiety of LPS[46,47]. LptE is an OM lipoprotein that has a dual function for LPS transport. First, in *E. coli*, it was shown to be essential for proper folding and insertion of LptD into the OM[101,189]. Second, it serves as a plug for the LptD barrel, with up to 75% of the protein inside of LptD[46,47,190]. LptE also directly interacts with LPS[191].

Two opposing models have long been proposed for LPS transport across the periplasm, the first working in a similar way as the Lol system (see previous section), the second suggesting a periplasm-spanning multi-protein bridge. In recent years, thanks to the advances in structural biology, the second model of a trans-membrane machine has strongly been favored[7]. Cross-linking experiments showed that the periplasmic domain of LptC interacts with the N-terminal region of LptA[192,193]. Similarly, the C-terminal domain of LptA interacts with the N-terminus of LptD[194]. It was also shown that transfer of LPS from LptBFG to LptC and from LptC to LptA requires the hydrolysis of ATP[175]. Thus, LPS transport is proposed to occur in the following steps (Fig. 11). First, LPS interacts with the LptBFG transporter, inducing ATP hydrolysis by LptB and its transfer from LptFG to LptC. Subsequent ATP hydrolysis then powers the transfer of LPS from LptC to LptA and eventually to LptDE in a continuous stream. The core and O-antigen are proposed to enter the lumen of the LptD β-barrel, probably binding to LptE[190]. The N-terminal domain of LptD is then suggested to form an intramembrane hole through which the lipid moiety of LPS could be inserted into the OM (Fig. 12)[195]. In parallel, the core and O-antigen trigger the opening of a lateral gate between the strands β1 and β26 of LptD, allowing transport to the cell surface (Fig. 12)[46,47,195]. In summary, LPS is transported from the IM to the cell surface by a periplasmic bridge formed by the interacting β-jellyrolls of LptF, -G, -C, -A and -D and powered by the hydrolysis of ATP, providing a continuous stream of LPS (Fig. 11)[7].

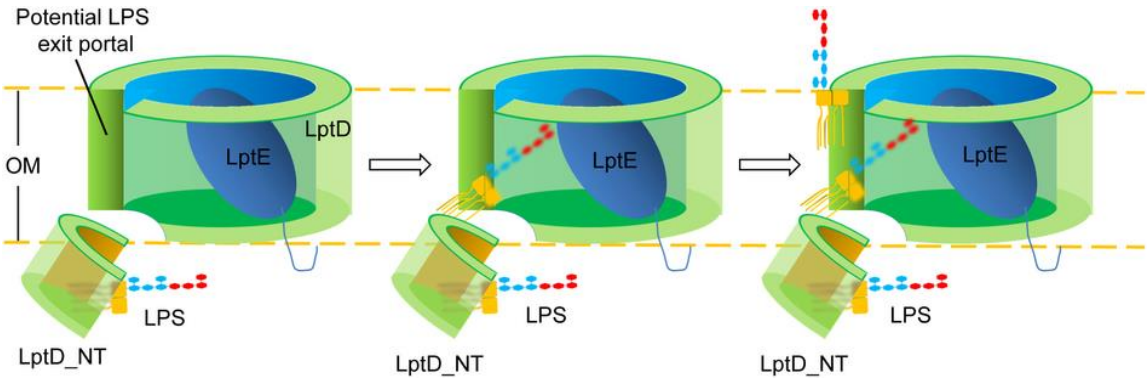


Figure 12. Molecular details underlying the insertion of LPS into the OM
Once LPS has crossed the periplasm, it interacts with both LptD and LptE. Presumably, the LptD N-terminus (LptD_NT) induces the formation of an intramembrane hole that allows insertion of lipid A into the OM. In parallel, LptE would assist the threading of the O-antigen through the LptD pore (LptD). Together, these events trigger the opening of the LptD barrel between the strands β1 and β26, allowing the core oligosaccharide to slide to the surface of the bacteria. Adapted from [47].

Crossing the cell wall – the many faces of bacterial secretion

The OM of Gram-negative bacteria is a formidable protective layer that shields the cell from noxious compounds and allows it to proliferate in even harmful environments. However, this barrier function also strongly limits the cell's ability to release molecules, such as metabolites or proteins, in the environment or to interact with other nearby cells. Gram-negative bacteria therefore have developed a multiplicity of so-called secretion systems, membrane-spanning nanomachines, that allow transport of proteins from the inside to the outside of the cell[196,197]. While the function, composition and overall structure can greatly vary among secretion systems (see below), they can be divided into two categories: single and double membrane spanning or one- and two-step secretion systems (Fig. 13)[198]. For our purposes, the second category will be used. One-step secretion systems (T1SS, T3SS, T4SS and T6SS) are multiprotein machines that span both the inner and outer membrane, therefore directly translocating proteins from the cytoplasm of the cell to the outside. Two-step secretion systems (T5SS, T9SS, chaperone-usher pathway and curli

biogenesis system) on the other hand only span the OM and therefore depend on the Sec- and Tat-machineries to transport their substrates to the periplasm[196,197]. This category also includes the T2SS and most likely the T7SS. Except for the T2SS, T6SS, T7SS and the chaperone-usher pathway, all above-mentioned secretions systems transport unfolded proteins[196,197]. A short overview of the different secretion systems is given below.

One-step secretion systems

Type 1 secretion system (T1SS)

Proteins secreted by the T1SS are involved in nutrient acquisition (proteases, lipases and iron scavengers)[199-201], pathogenesis (haemolysins and leukotoxins)[202] and bacterial competition (bacteriocins)[203].

The T1SS is composed of three distinct parts: an IM component (IMC), a membrane fusion protein (MFP) and the OM β-barrel channel TolC (Fig. 13a) [196,204]. The IMC is an IM protein that belongs to the ABC-transporter family of proteins. The MFP is an IM anchored periplasmic protein that forms a hexameric tunnel-like structure and links the IMC to the OM protein TolC. The IMC and MFP are responsible for substrate recognition, each IMC-MFP pair only interacting with one specific set of substrates, while TolC can associate with different IMC-MFP pairs and forms the OM channel through which the substrates will be secreted[196,204]. Recent data suggests that the IMC:MFP:TolC complex adopts a stoichiometry of 3:6:3[205], which led to the following working model (Fig. 13a): the IMC-MFP recognizes its substrates via their N-terminally located glycine-rich motif; substrate binding triggers ATP hydrolysis by the IMC leading to translocation of the substrate from the cytoplasm to the periplasm; IMC-MFP-substrate interaction induces TolC recruitment; formation of a complete complex leads to opening of the TolC channel and release of the substrate[198,203].

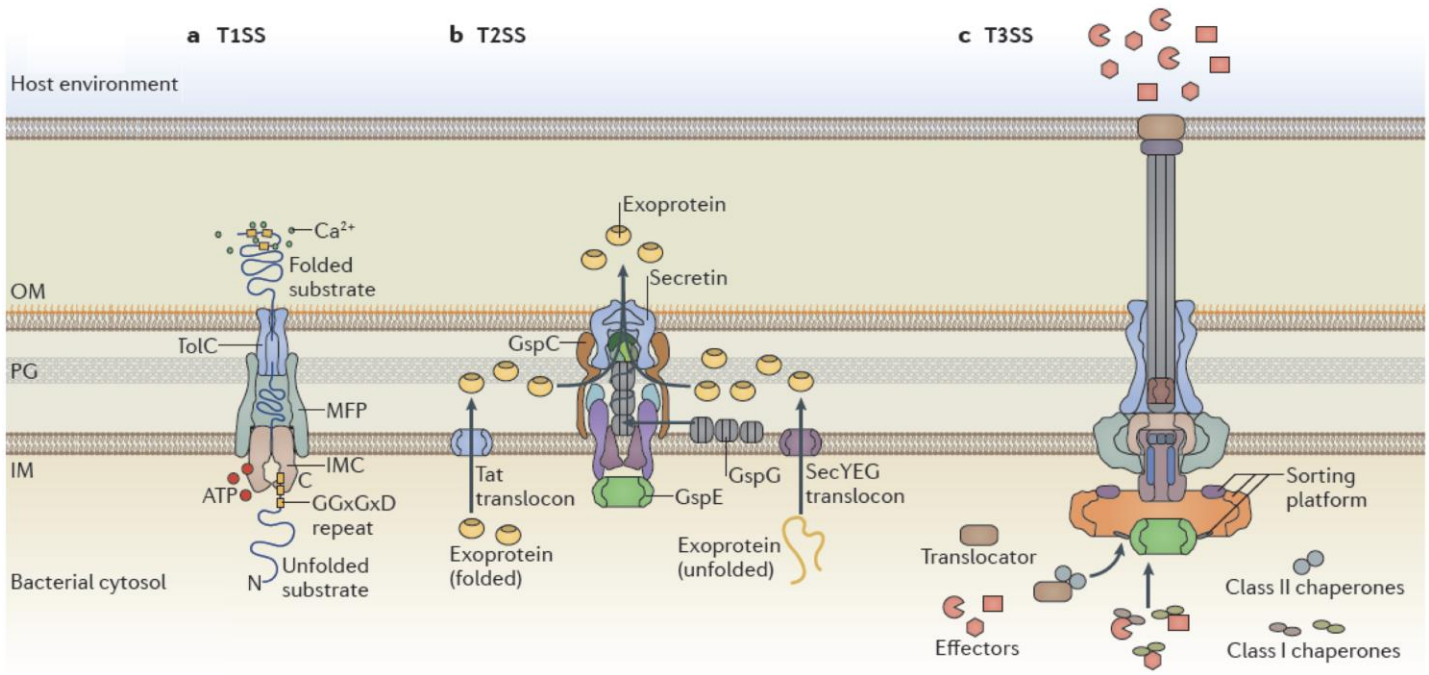


Figure 13. Structural organization of type 1, 2 and 3 secretion systems of Gram-negative bacteria

(a) The T1SS is composed of an outer membrane TolC component, a periplasmic membrane fusion protein (MFP) and an inner membrane component (IMC) that supplies energy for transport. Following recognition of the substrate, ATP hydrolysis by the IMC triggers the transfer of the unfolded substrate to the MFP. TolC then secretes the substrate through the OM. (b) The T2SS consists of an OM complex (the secretin GspD, blue), a periplasmic pseudopilus (composed of the major pseudopilin subunit GspG and additional minor pseudopilin subunits), and an IM platform that is tightly associated with the cytoplasmic ATPase GspE (green). GspC (brown) recruits the substrate from the periplasmic space to the secretin. Thanks to the ATPase activity of GspE, the pseudopilus then pushes the substrate through the secretin channel, releasing the substrate. (c) The T3SS is composed of an OM secretin (blue), an IM complex and a needle (grey). The basal body interacts with a cytoplasmic ATPase (green) and a sorting platform (orange). The secretin extends from the OM to the periplasm, forming a series of protective rings that surround the needle. The sorting platform and the secretin are connected by the IM complex. Following contact with a host cell, the secretion of so called translocators is initiated. After insertion into the host cell membrane, the translocators form a functional pore that assists the subsequent transport of effectors into the host cell cytosol. Effector secretion through the T3SS is subject to temporal regulation mediated by interaction with different classes of cytoplasmic chaperones. Adapted from [196].

Type 3 secretion system (T3SS)

The T3SS is one of the most complex prokaryotic nanomachines described to date and is found in pathogenic bacteria such as *Yersinia*, *Shigella* and *Salmonella*[206]. The T3SS mediates the transfer of so-called effector proteins into eukaryotic host cells, thereby hijacking host cellular functions in order to allow the bacterium to colonize and proliferate[207-209]. The main constituent of the system is syringe-like in shape (hence often referred to as injectisome or needle complex), composed of up to 25 different proteins and being 3.5 MDa in size (Fig. 13c). The system can be divided into 2 subunits: the basal body, composed of an IM complex, an OM complex and a periplasmic bridge; and the needle itself[196,197,210].

The IM complex is made of 2 oligomeric rings inserted into the membrane, one facing the cytoplasm and one facing the periplasm[196,211]. Additionally, a cytoplasmic sorting platform interacts with the IM complex and is involved in recruitment of effector proteins[212,213]. The OM complex, composed of stacked rings forming the so-called secretin, and the periplasmic bridge, termed the neck, are made of one protein and associate with the periplasmic ring of the IM base[201,210,211]. The needle is basically a hollow tube inserted through the IM and OM rings and extending into the extracellular milieu[196,210,211].

Unlike other secretion systems, the secretion of T3SS effectors is subject to temporal regulation[213]. Upon contact of the needle tip with a target cell, the IM complex and sorting platform will first recruit effector proteins called translocators[209,214]. Following transport across the needle, these proteins will form a pore in the membrane of the host cell, paving the way for other effector proteins that will then subdue the cellular machinery of the host. T3SS substrates therefore present a hierarchy in their secretion. This temporal regulation is thought to be dependent on cytoplasmic chaperones that associate wit T3SS effectors and that have different affinities towards the cytoplasmic sorting platform[209,214,215]. Effectors that need to be secreted first therefore interact with chaperones that have higher affinity. Finally, the entire secretion process is powered by several copies of a cytoplasmic ATPase.

Type 4 secretion system (T4SS)

The T4SS has the unique ability among secretion systems to transport both proteins and DNA; it is thus mainly described in respect to conjugation of plasmid DNA into bacterial and eukaryotic cells [216-218]. Since conjugation is extremely widespread among prokaryotes, the T4SS is the most commonly found secretion system known to date and can be encountered in Gram-negative, Gram-positive and even archaea[219].

The T4SS is composed of 12 proteins: VirB1 to VirB11 and VirD4[196,220,221]. While VirB1 is a periplasmic peptidoglycan hydrolase required for the T4SS pilus biogenesis[222], all other proteins are integral parts of the secretion apparatus. Recent structural analysis allowed to determine the structure of a VirB3 to VirB10 comprising complex[220] and, in combination with previous data, showed that it can be divided into 3 distinct parts (Fig. 14a): the IM complex (VirB3, 4, 6, 8 and 10)[220], the stalk (composition unknown) and the core-OM complex (VirB7, 9 and 10)[220,223]. Additionally, the pilus, composed of VirB2 and 5, extends into the extracellular medium and establishes contact between mating cells[222].

The IM complex is divided into two cytoplasmic barrels, each composed of 6 subunits of VirB4, linked by an IM inserted bridge made of VirB6, 8 and 10[220]. The core-OM complex is arranged into 2 stacked rings[220,223]. The stalk is an extended structure that connects the IM and OM complexes and allows translocation of substrates, which is powered by ATP hydrolysis[196,197]. Due to the absence of VirB11 and VirD4 in the analyzed complex, a precise mechanism of substrate transport is so far not proposed. However, one model suggests that the T4SS switches between 2 modes of functioning[222]. First, VirB11 would bind to VirB4, inducing formation of the pilus (Fig. 14a step 1). Once the pilus interacts with a receptor cell, VirB11 switches from VirB4 to VirD4, leading to substrate secretion (Fig. 14a step 2)[222,224].

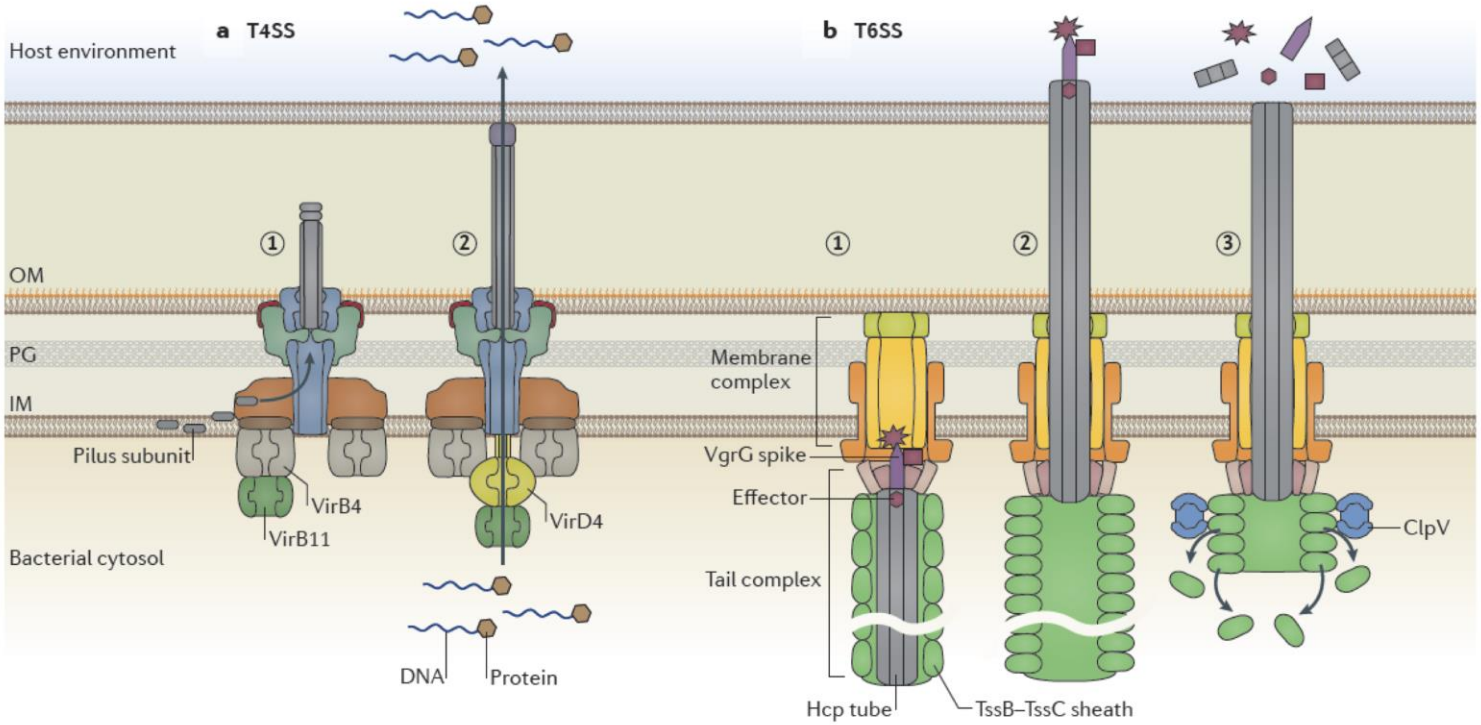


Figure 14. Structural organization of type 4 and 6 secretion systems of Gram-negative bacteria

(a) The T4SS is composed of three ATPases (VirD4, VirB4 and VirB11) that, together with VirB3 (dark brown), VirB6 and VirB8 (light brown), form the IM complex. VirB7 (red), VirB9 (light green) and VirB10 (blue) form the core–OM complex, with VirB10 extending from the IM to the OM. The conjugative pilus is composed of VirB2 (grey) and VirB5 (purple). Supposedly, the association of VirB11 with VirB4 promotes pilus subunit assembly (step 1), whereas the association of VirB11 with VirD4 facilitates substrate translocation (step 2). (b) The T6SS is composed of a membrane complex spanning the periplasm and a cytoplasmic tail complex, comprising a phage-like tube (Hcp), sheath and baseplate. The two complexes are connected in the cytoplasm through the baseplate of the tail complex. Effectors are recruited to the tube through interaction with the VgrG spike at the tip of the tube (step 1). An unknown extracellular signal then triggers sheath contraction, which leads to the ejection of the spike–tube complex across the target membrane, thereby delivering effector proteins into the cell (step 2). The ATPase ClpV (blue) disassembles the contracted sheath, which enables a new T6SS complex to be reassembled from the released subunits (step 3). Adapted from [196].

Type 6 secretion system (T6SS)

The T6SS is one of the more recently described Gram-negative secretion systems[225,226] and is involved in toxin delivery into eukaryotic[227,228] but also prokaryotic[229,230] cells, therefore having a major role in bacterial competition[231,232]. Predominantly found in Proteobacteria[233], such as *Vibrio*, *Pseudomonas* and *Francisella*, a new class of T6SS has also recently been described in the phylogenetically distant Bacteroidetes[234].

The core complex of the T6SS is composed of 13 conserved proteins in combination with several accessory proteins that together define T6SS subclasses[226,233,235,236]. The T6SS machinery can be divided into 2 main components: a double membrane spanning complex[237] and a cytoplasmic tail complex (Fig. 14b)[237,238]. The membrane complex is composed of at least 3 proteins and is bridging the IM and the OM[237,239,240]. The tail complex, which appears to be highly similar in structure and function to the bacteriophage tail, can be subdivided into tail sheath, tube and baseplate[238]. Like the viral protein delivery system, the baseplate is believed to act as building platform for the tube and its surrounding sheath (Fig. 14b). The tube is topped by a so-called spike that is involved in tube polymerization as well as substrate recruitment (Fig. 14b step 1)[238,241].

While the exact mechanism of protein delivery, especially what signal triggers secretion, is not yet described, it is though to be similar to phage tail contraction[241-243]. Following interaction with a target cell, contraction of the sheath leads to translocation of the tube through the membrane complex and across the membrane of the receptor cell. T6SS substrates are delivered by interaction with the spike before sheath contraction (Fig. 14b step 2) [196,231,236,237]. A cytoplasmic ATPase then induces sheath disassembly and the resulting subunits are available for a new round of sheath construction and contraction (Fig. 14b step 3)[196,236].

Two-step secretion systems

Type 2 secretion system (T2SS)

The T2SS, also referred to as general secretion system, is to date the only two-step secretion system known to span both the inner and outer membranes[196,244,245]. Additionally, the T2SS is one of the few machineries able to transport folded substrates that are first translocated to the periplasm via the Sec and Tat pathways[196,244,245]. This system in widely distributed among Gram-negative bacteria and secretes both enzymes and toxins, among which the well studied *Vibrio cholera* toxin[244,245].

The T2SS is composed of up to 15 components that together form 4 distinct parts (Fig. 13b): an OM complex, referred to as secretin, that constitutes the secretion channel[246]; a periplasmic pseudopilus[247]; an IM spanning complex interacting with both OM complex[248] and pseudopilus; and a cytoplasmic ATPase[196,244,245,249]. While the complete structure of the T2SS has yet to be determined, the secretion process is believed to occur in 2 steps (Fig. 13b). First, the periplasmic substrates are recruited to the secretin channel located at the level of the OM. The substrates are then secreted by being pushed across the OM pore by the pseudopilus in an ATP-dependent manner[196,244,245].

Type 5 secretion system (T5SS)

The T5SS is unique among secretion systems because it is the only one in which the translocation machinery and the substrate is one and the same protein, hence the more commonly used name autotransporter system[196,250-252]. As a result, a given autotransporter is composed of only 2 domains: a C-terminal translocator domain, corresponding to an OM-inserted β-barrel[253,254], and a N-terminal passenger domain, which represents the secreted protein (Fig. 15a). The T5SS is responsible for the secretion of various virulence factors and sometimes adhesins[250,252]. Consequently, while most passenger domains are cleaved off after their transport across the OM, adhesins remain attached to the membrane via their translocator domain[250,252].

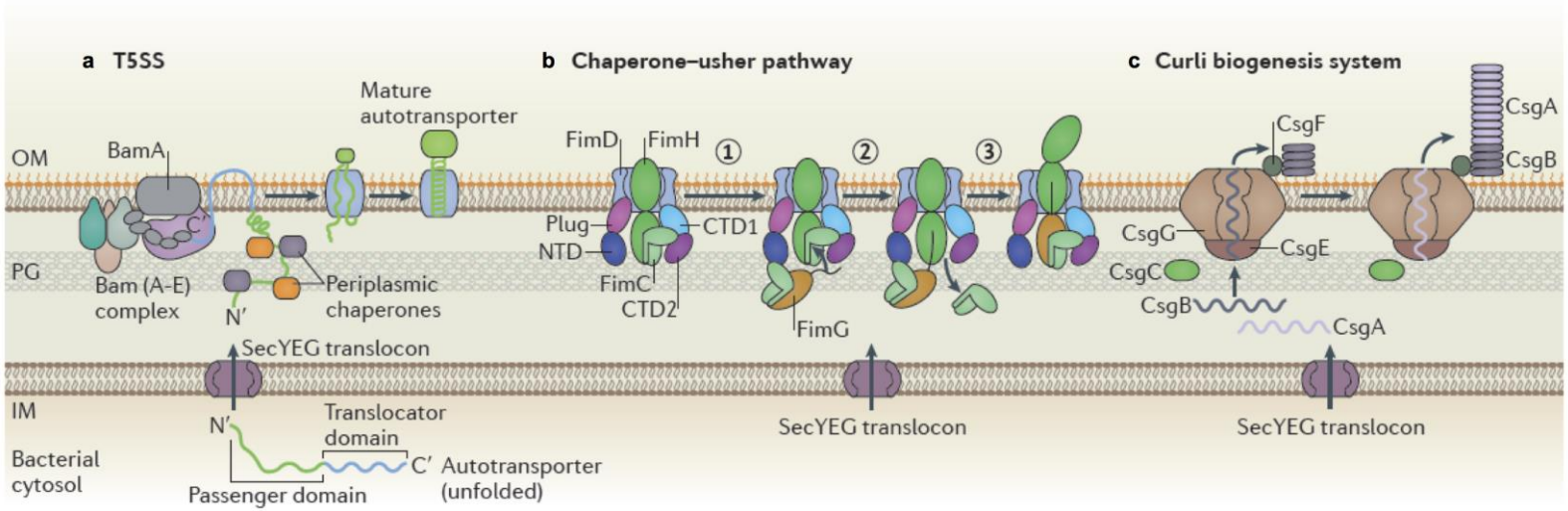


Figure 15. Structural organization of type 5, chaperone-usher pathway and curli biogenesis secretion systems of Gram-negative bacteria

(a) The T5SS or autotransporter system is a single protein secretion system composed of a C-terminal translocator domain inserted into the OM as a β-barrel and a N-terminal passenger domain exposed to the extracellular space after translocation through the β-barrel. The unfolded autotransporter is transported to the periplasm by the Sec machinery, where several chaperones stabilize its unfolded structure to prevent aggregation. It is then inserted into the OM with the assistance of the Bam complex. The unfolded passenger domain then passes through the pore created by the translocator domain and folds on the surface of the cell. (b) The chaperone–usher pathway is responsible for the synthesis of a pilus made of several subunits (FimH, -G, -F and -A) and is composed of the usher protein FimD, containing a pore, a plug, a N-terminal domain (NTD) and two C-terminal domains (CTD1 and CTD2) and the periplasmic chaperone FimC, involved in targeting the pilus subunits towards the usher. Pilus biogenesis is initiated by FimC–FimH interaction with the FimD CTDs. A FimC–FimG complex is then recruited to the FimD NTD (step 1). This leads to interaction between FimH and FimG, triggering dissociation of the FimC–FimH complex (step 2). FimH-FimG are then partially translocated through the FimD pore, thereby freeing the FimD NTD that is now able to bind a new FimF-FimC complex. This cycle is then continued by the addition of approximately 1,000 copies of the FimA subunit to the pilus. (c) The curli biogenesis system is composed of 6 proteins, CsgA and -B being the curli subunits and CsgC, -E, -F and -G assembling them. CsgA and -B are transported across the OM by a pore made of CsgE and -G. CsgB is then anchored to the cell surface thanks to CsgF before acting itself as nucleation factor for CsgA. The role of CsgC remains unclear. Adapted from [196].

Similar to porins and other OM β-barrels, autotransporters are transported to the periplasm via the Sec machinery, interact with SurA to remain in an unfolded state and are ultimately delivered to the Bam complex for membrane insertion (Fig. 15a)[255]. Following completion of the β-barrel, the passenger domain is then threaded through the pore and in most cases released into the extracellular environment (Fig. 15a)[256]. Recently however, several data suggest that autotransporters might not be as autonomous as initially described and point out an active role of the Bam or the Tam (translocation and assembly module) machineries in secretion of the passenger domain[257-259]. The exact secretion mechanism thus remains to be clarified.

The chaperone-usher pathway (CU)

The CU pathway is responsible for the synthesis and membrane anchorage of surface structures called pili, often found in uropathogenic *E. coli* (UPEC)[260-262]. These appendages mediate cell adhesion and biofilm formation and are therefore important pathogenicity factors[260,262].

The CU pilus is made of two components (Fig. 15b): the pilus itself, made of a long cylindrical structure with a flexible tip at its end[263,264], and the usher protein, an OM protein that polymerizes the pilus subunits[260,262,265]. As for other two-step secretion systems, the usher and the pilus subunits are secreted by the Sec machinery before their assembly[266,267]. The usher is a 24 stranded β-barrel occluded by a plug domain[265]. Additionally, one N-terminal and two C-terminal domains involved in pilus subunit binding are extending into the periplasm[265]. In order to remain unfolded, the pilus subunits form complexes with dedicated periplasmic chaperones, hence the name of the system[267-269]. These 1:1 complexes then interact with the usher protein, inducing opening of the channel by displacing the barrel plug followed by polymerization of the pilus subunits[196,270-274]. First, a chaperone:subunit complex associates with the C-terminal domains of the usher (Fig. 15b step 1). This is followed by a second complex interacting with the usher N-terminal domain, leading to dissociation of the first chaperone:subunit complex (Fig. 15b step 2). This pilus subunit is then inserted into the barrel lumen while the subunit of the second complex is

transferred from the N- to the C-terminal domains of the usher (Fig. 15b step 3) [196,270-274]. A new round of polymerization can thus begin. Once polymerization is complete, the pilus remains anchored to the OM via the usher.

The curli biogenesis

The curli biogenesis system, sometimes referred to as T8SS or extracellular nucleation-precipitation (ENP) pathway, is involved in the biogenesis of amyloids called curli, long proteic nanofibers essential for biofilm formation[275,276]. Curli protect bacteria from harmfull environments and promote host invasion in species such as *E. coli* and *Salmonella*[275,276].

Curli biogenesis is accomplished thanks to 6 proteins, CsgA, -B, -C, -E, -F and -G (Fig. 15c)[276,277]. CsgA is the major curli subunit and is secreted as a soluble unfolded protomer by the Sec machinery[278,279]. To reach the bacterial surface, CsgA is transported across the OM by a pore formed by CsgE and -G[280,281]. CsgG is a lipoprotein possessing a transmembrane α-helix which assembles into oligomeres to form the OM channel[281], while CsgE is a periplasmic protein that tighly interacts with CsgG and is required for recruitment of CsgA subunits[282,283]. Once secreted through the pore, CsgA interacts with CsgB and assembles into amyloid fibers (Fig. 15c)[284,285]. Assembly of CsgB itself depends on the surface exposed CsgF[286]. The precise function of CsgC is still not clear, although it is suspected to target CsgA to the OM pore[196,276].

Type 7 secretion system (T7SS)

The T7SS has first been described in Mycobacteria[287], where it is required for virulence of species such as *M. tuberculosis*. Although Mycobacteria are not Gram-negative bacteria *per se*, they do possess a diderm membrane architecture (Fig. 16a)[288,289]. Indeed, they have an additional membrane surrounding their plasma membrane, called mycomembrane, composed of a waxy lipid coat made of mycolic acids[288,289] while the periplasmic space contains a peptidoglycan as well as arabinogalactan layer[288,289]. Up to 5 subclasses of T7SS are reported in

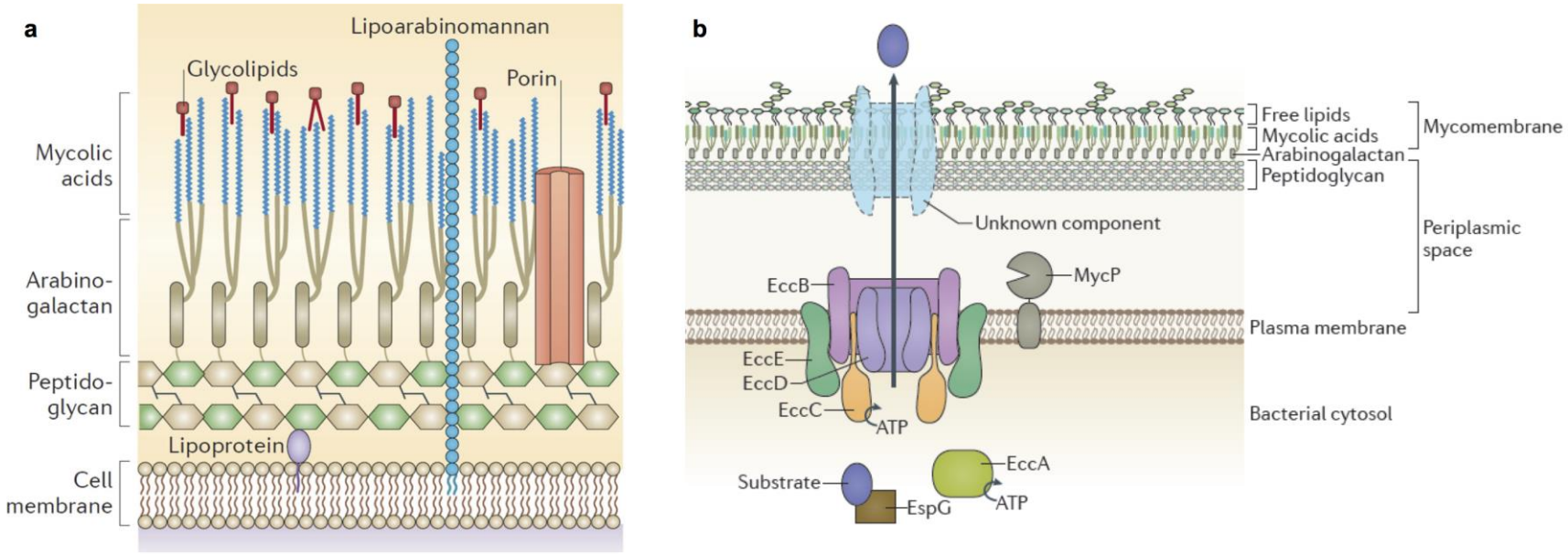


Figure 16. The mycobacterial cell envelope and the T7SS

(a) Mycobacteria, belonging to the high G+C Gram-positive bacteria, have a cell envelope similar to Gram-negative bacteria. It consists of a plasma membrane (equivalent to the IM of Gram-negative bacteria), a periplasmic space containing peptidoglycan and arabinogalactan, and a thick, complex OM that contains a waxy lipid coat of mycolic acids called the mycomembrane. Glycolipids, porins as well as lipoarabinomannan are inserted into this OM. (b) The T7SS is composed of 4 conserved core proteins (EccB, -C, -D and -E) forming an IM complex associated to the periplasmic protease MycP implicated in substrate processing. The cytoplasmic ATPase EccA and the chaperone EspG are presumably involved in substrate guidance and secretion. To date, only the IM component of the T7SS has been identified; the putative OM transporter, if any, remains unknown. Hence, the precise transport mechanism remains to be determined. Interestingly, T7SS substrates have been shown to form heterodimers which seems to be a prerequisite for secretion. Adapted from [2,196].

Mycobacteria[290] and homologous systems have been identified in Gram-positive bacteria such as *Staphylococcus aureus* or *Bacillus anthracis*[291-293].

T7SSs are composed of 4 conserved core proteins (EccB, -C, -D and -E) forming an IM complex associated to the periplasmic protease MycP, which is essential for secretion (Fig. 16b)[294-297]. Additional cytoplasmic proteins such as EccA and EspG are thought to aid secretion, act as chaperones and assure substrate specificity (Fig. 16b)[230,298,299]. Interestingly, T7SS substrates are secreted as heterodimers that assume a four-helix bundle conformation, which seems to be a common feature to all substrate pairs[300-303]. To date, only the IM complex of the T7SS has been identified[295,296] and it is therefore not clear whether secretion occurs in a one- or two-step mechanim. Consequently, the putative transporter (if any) inserted into the mycomembrane has yet to be identified.

Type 9 secretion system (T9SS)

The T9SS is the most recently described Gram-negative bacterial secretion system and is exclusive to the phylum Bacteroidetes[304,305]. This system is involved both in secretion and gliding motility[306,307]. Gliding motility is a way of movement that allows bacteria to rapidly crawl over surface[308]. Different types of gliding motility have been described, but they all work independently of flagella and require energy for translocation of the cell[308]. In the case of Bacteroidetes such as *Flavobacterium johnsoniae*, gliding motility requires the presence of 11 conserved proteins to promote cell movement, while only a subset (7 proteins) is required to form the secretion apparatus[305]. T9SS substrates are involved in pathogenesis of species such as *Porphyromonas gingivalis* that causes severe periodontitis[309,310].

While the complete structure of the T9SS still requires further investigation, a recent study provided first insights into its structural organization[311]. Gorasia *et al*. showed that two components of the *P. gingivalis* T9SS, PorK and PorN, form a 50 nm diameter ring-shaped complex attached to the periplasmic side of the OM, containing 32 to 36 subunits of each protein (Fig. 17). The authors also demonstrated that PorL and PorM form a separate stable

complex anchored to the IM proposed to provide energy for substrate transport (Fig. 17). Additionally, substrate recognition and processing have been mostly elucidated. Following their secretion by the Sec machinery, T9SS substrates are transported across the OM after recognition of a specific signal sequence in their C-terminus[312,313]. Once they reach the bacterial cell surface, this C-terminal signal is cleaved off and the mature proteins may or may not be linked to a short form of LPS, called A-LPS, that allows their anchorage to the OM[314,315].

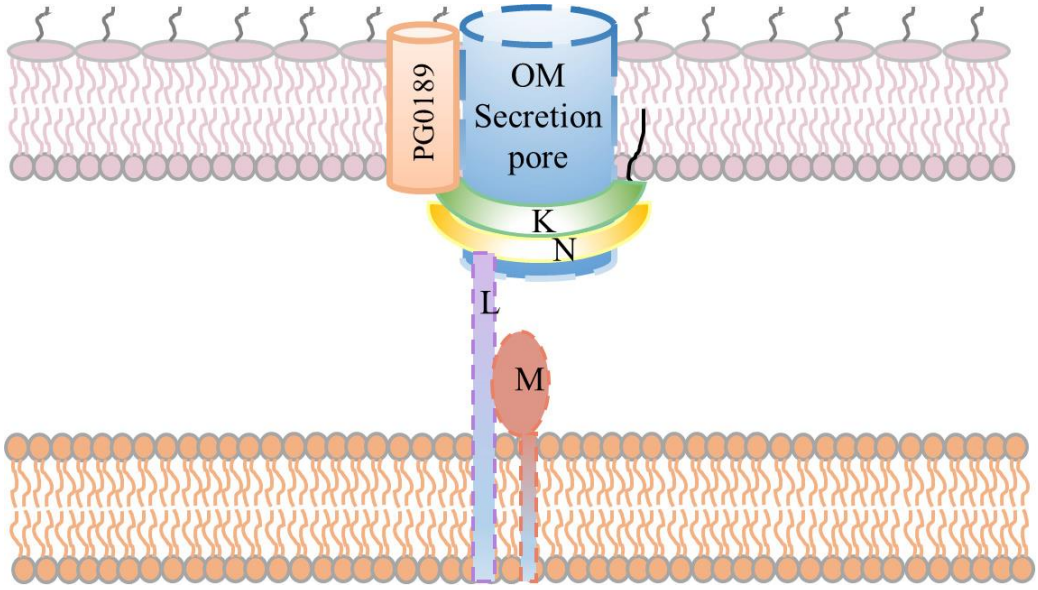


Fig 17. Proposed model for the structural organization of the *P. gingivalis* T9SS
PorK and PorN interact to form a ring-shaped structure that is localized in the periplasm and tethered to the OM via the PorK lipid anchor (black line). This structure may be further stabilized by its association with the PG0189 outer membrane protein. It is proposed that the PorK and PorN rings assemble around the periplasmic extensions of a yet unknown OM secretion pore. PorL and PorM have transmembrane spanning domains and are proposed to power secretion of the T9SS substrates through transient interactions with the PorK/N complex. The topology of the PorL and PorM inner membrane proteins is not known. Additional components of the secretion system without assigned structure or interaction are not displayed. Adapted from [311].

Surface exposed lipoproteins in bacteria

As previously described, most bacterial secretion systems are involved in transport of soluble proteins that are either released into the environment or a nearby cell, polymerized at the cell surface to form complex structures or covalently attached to the OM. However, to date, little attention has been paid to

proteins that are attached to membranes prior to transport initiation, especially how proteins such as lipoproteins reach the bacterial surface[161,316,317]. While this lipoprotein localization is rather rare in Proteobacteria and only a few cases have been elucidated[161,316], it is common in species belonging to the Spirochaetes or Bacteroidetes phylum and is a fundamental aspect of their respective biology[161,316,317].

Rather unusual …

RcsF in *E. coli*

RcsF is part of the Rcs phosphorelay, a signaling system composed of 6 proteins that detects and responds to OM and cell wall damages[106,318-321]. RcsF is the sensor of the system and activates the signaling cascade by interacting with the IM-anchored periplasmic protein IgaA[104]. Recent work has shown that RcsF monitors BamA activity by interacting with BamA as well as OmpA, one of the major *E. coli* porins (Fig. 18)[104]. More precisely, Cho *et al*. showed that BamA passes RcsF over to OmpA and that OmpA exposes, at least partially, RcsF to the cell surface. By doing so, the Rcs phosphorelay remains inactive (Fig. 18)[104]. However, absence of OmpA or decreased activity levels of BamA (resulting from an encountered stress) triggers the Rcs stress response as the protein remains periplasmic and interacts with IgaA[104]. This is so far the first example where BamA activity is linked to lipoprotein surface localization.

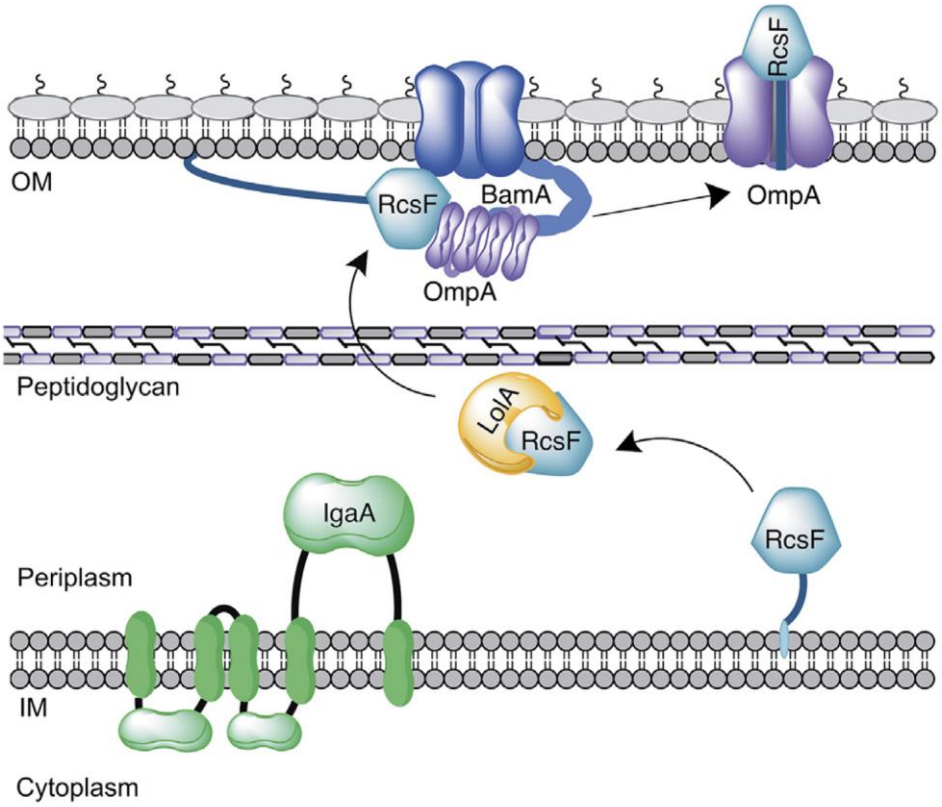


Figure 18. Model of *E. coli* RcsF surface exposure
Following maturation, RcsF is transported to the OM by the Lol system similarly to other OM lipoproteins in *E. coli*. RcsF then interacts with BamA, the central component of the Bam machinery, leading to the formation of a complex between RcsF and OmpA, an abundant β-barrel protein. This results in surface display of RcsF, thus unable to interact with its cognate partner IgaA. The Rcs stress response is effectively shut down. Upon stress, affecting for exemple the activity of BamA, the RcsF-OmpA complex cannot be assembled anymore, leading to periplasmic localization of RcsF. It then interacts with IgaA, inducing the Rcs stress response. Adapted from [104].

TbpB, LbpB and fHbp in *Neisseria meningitidis*

Neisseria meningitidis presents several surface exposed lipoproteins at its surface, namely TbpB, LbpB and fHbp. Although their function has since long been elucidated (TbpB and LbpB are involved in iron acquisition in blood[322-324] while fHbp binds factor H in order to prevent killing by complement[325]), the question of how theses proteins are transported to the cell surface remained unsolved. Recent work has now shed light on this aspect and showed that a novel transporter, called Slam1 (surface lipoprotein assembly modulator), is responsible for their transport[326]. Slam1 is an OM protein with 2 TPR (tetratricopeptide) domains in its N-terminus and a 14-stranded β-barrel in its

C-terminus. The function of the TPR domains remains to be clarified as the β-barrel domain on its own is sufficient to mediate protein transport to the bacterial surface. However, Slam1 is not a general lipoprotein transporter, as HpuA (another surface exposed lipoprotein[327,328]) requires the Slam1 homolog Slam2 for transport, indicating that each transporter has specific substrates[326]. Interestingly, Slam1 homologs were identified in other Proteobacteria and secretion of TbpB could be reconstituted in *E. coli* by expressing Slam1, indicating that this system, and therefore surface exposed lipoproteins, might be more common in this phylum than generally assumed[326].

NalP in *Neisseria meningitidis*

NalP, similar to SphB1 in *Bordetella pertussis*[329,330], is a surface exposed serine protease belonging to the family of autotransporters (Fig. 19 step 3c)[331]. It is thus composed of a C-terminal β-barrel domain responsible for transport and a N-terminal secreted domain harboring the enzymatic activity. NalP is involved in autocatalytic processing, leading to its own release from the cell, as well as for the processing of other cell surface associated proteins[331,332]. Interestingly, NalP is also a lipoprotein, therefore being anchored into the OM. While the lipidation is not necessary for transport, absence of lipidation does however increase NalP autocatalytic processing, thus decreasing the amount of NalP bound to the OM, which in turn decreases the processing of its other substrates[333].

PulA in *Klebsiella oxytoca*

PulA, a starch debranching enzyme, is attached to the cell surface before being slowly released into the extracellular medium[11,334,335]. Interestingly, PulA presents the classical Lol-avoidance signal (asparate at position +2 in the mature protein) and should therefore remain anchored into the IM, suggesting that surface exposure is not achieved by the Lol-system or extension of thereof. Rather, it has been shown that PulA is transported via the T2SS, although how exactly the protein is extracted from the IM remains unclear (Fig. 19 step 2b)

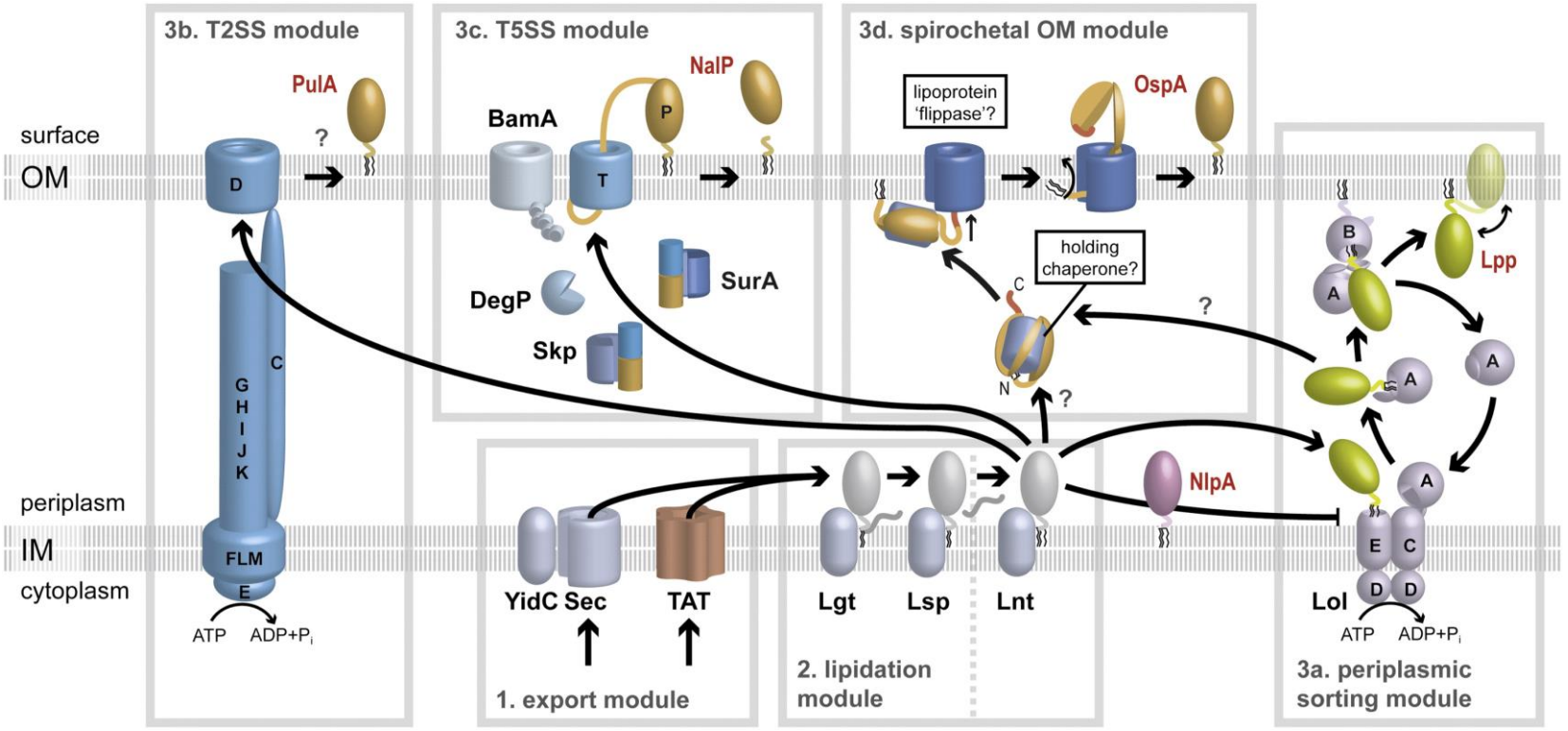


Figure 19. Model of lipoprotein surface exposure in *Klebsiella*, *Neisseria* and *Borrelia*

As explained previously, the classical lipoprotein synthesis pathway in Gram-negative bacteria involves translocation of a precursor across the IM by either Sec or Tat machinery (step 1), maturation of the precursor into a tri-acylated lipoprotein (step 2) and Lol-dependent transport to the OM (step 3a). In this model, *E. coli* NlpA and Lpp represent typical IM and OM lipoproteins. However, in some specific cases, lipoproteins are transported to the cell surface independently of the Lol system. In *Klebsiella oxytoca*, the PulA protein is surface localized by the T2SS. PulA is first recruited to the T2SS pseudopilus before being pushed in a piston-like manner through the OM pore. How the lipid anchor of PulA is accommodated during this transport is unknown (step 3b). *Neisseria* spp NalP is transported by the T5SS. The unfolded polypeptide is first escorted to the OM by periplasmic chaperones where its translocator domain is then inserted into the membrane by the Bam complex. The passenger domain is threaded through the pore of the translocator domain and anchored to the OM. NalP is then slowly released from the cell by autolytic cleavage (step 3c). Spirochaetes such as *Borrelia burgdorferi* abundantly expose lipoproteins at their surface, like for example OspA. While the nature of the OM flippase machinery and the involvement of the Lol system in this process are still unclear, it has been shown that surface lipoproteins have to remain in an unfolded, transport compatible state. This led to the hypothesis of the involvement of a putative "holding" chaperone. Adapted from [161].

[11,334,336]. Noteworthy, lipidation is not required for surface transport but does however improve its efficiency at high expression levels[337]. To date, only few other lipoproteins have been reported to be transported by the T2SS, among which the *E. coli* SslE protein[338] and cytochromes in S*hewanella oneidensis*[339].

… or maybe not ?

Lipoproteins in *Borrelia burgdorferi*

Spirochaetes are diderm bacteria with a double membrane architecture reminiscent of a classical Gram-negative bacterial cell, *i.e.* presence of an IM, a periplasmic space and an OM[340]. However, in *Borrelia burgdorferi*, the causative agent of Lyme disease, striking differences exist, such as the presence of periplasmic flagella[341,342], absence of phosphatidylethanolamine in the OM[343], and perhaps most notably, absence of LPS[340,344]. Additionally, *Borrelia burgdorferi* abundantly expose lipoproteins at their surface that are involved in pathogenesis[340]. These proteins have therefore been investigated in detail due to their importance in vaccine development. In parallel, the underlying surface transport mechanism has been studied.

While the OM transport machinery (if any) remains unknown, progress has been made in regard to the signal involved in lipoprotein surface localization in *B. burgdorferi*. First, unlike described in Proteobacteria, the +2 rule (Lol-avoidance signal) or variations of thereof do not apply in Spirochaetes[345,346]. Second, there seems to be no specific signal responsible for surface targeting, hence the assumption that surface lipoproteins are transported by default across the OM[346,347]. Third, the folding state of lipoproteins is crucial for their surface export. Indeed, lipoproteins are transported as unfolded polypeptides, evidenced by the fact that prematurely folded lipoproteins are anchored to the OM but remain periplasmic[348]. On the other hand, fold destabilizing mutations allow surface export of otherwise periplasmic retained proteins[349]. This led to the hypothesis of a putative “holding” chaperone that would prevent folding of surface lipoproteins before their transport across the OM (Fig. 19 step 3d)[348]. Additionally, the fact that folded lipoproteins are attached to the OM but not

surface localized suggests that OM anchoring and surface transport might be two uncoupled steps[348]. Further investigation is needed in order to understand how *Borrelia* targets its lipoproteins to the cell surface as well as to identify the corresponding transport machinery.

Apart from Spirochaetes, the only other phylum in which lipoproteins are abundantly surface exposed is Bacteroidetes. Being the main subject of this work, a detailed description is provided within the next section.

Of Bacteroidetes, lipoproteins and dogs

Bacteroidetes – your (sometimes) friendly neighborhood bacterium

The phylum Bacteroidetes comprises a high diversity of Gram-negative bacterial species that have colonized all types of habitats. They can be found in soil[350], aquatic environments[351] and as commensals of mammals or insects[352,353]. Among them, *Bacteroides* spp. are common members of the intestinal flora where they play a major role in gut homeostasis[354-358] while *Capnocytophaga* and *Porphyromonas* spp. are part of the oral flora[359,360]. However, this phylum also includes opportunistic pathogens such as *Bacteroides fragilis* and *Capnocytophaga canimorsus* that cause acute systemic infections in humans[361-366] or *Porphyromonas gingivalis* that causes severe periodontal diseases[359]. The wide distribution of these organisms is in part due to their high adaptability to their ecological niche, especially thanks to their vast array of glycosylhydrolases. Indeed, genome analysis of members of this phylum has revealed their incredibly diverse arsenal of enzymes allowing them to degrade nearly all types of carbohydrates they can encounter[160,305,358,367,368]. Interestingly, these enzymes are often surface exposed lipoproteins and are part of multi-protein OM complexes devoted to nutrient acquisition. These complexes, facing the outside environment[369,370], are encoded in genetic regions named Polysaccharide Utilization Loci (PUL)[358,368] that represent a hallmark of this phylum.

Polysaccharide utilization loci in Bacteroidetes

PUL-encoded complexes are involved in the binding, degradation and internalization of a wide variety of carbohydrates[368,371,372], including complex glycans such as starch[372] and hemicellulose[373,374], highly glycosylated proteins like mucins[375,376] and even iron from sero-transferrin[377]. The first PUL described encodes the so-called Sus (Starch utilization system)[372,378] from *Bacteroides thetaiotaomicron* (Fig. 20), a human gut commensal able to degrade a wide range of host dietary glycans [357,379,380]. The Sus membrane complex[371,372,381,382] is composed of five proteins: SusC, a TonB-dependent receptor spanning the outer membrane[369,383]; SusD, -E and -F, cell surface exposed lipoproteins involved in starch binding [370,383,384] and SusG, a cell surface exposed lipoprotein with α–amylase activity[385,386]. Upon binding of starch, SusG hydrolyses the long polymeric glucose chains into smaller oligosaccharides that are then imported through the SusC channel into the periplasm where they are further processed by SusA and B (Fig. 20)[372]. The expression of the *sus* operon is under the control of the SusR protein, an IM spanning sensor/regulator that induces *sus* gene expression upon detection of maltose, a glucose disaccharide, in the periplasm (Fig. 20). Thanks to the increasing availability of genomic data, a significant number of PUL-encoded systems have been identified not only in *B. thetaiotaomicron* (representing as much as 18% of its genome)[358,375], but in all major groups of Bacteroidetes, including commensals[387,388], saprophytes[305,367] as well as pathogens[160]. This indicates that independently of the bacterial lifestyle, PUL are critical for nutrient acquisition in this phylum. Furthermore, since many Bacteroidetes species are gut commensals, their PUL-derived ability to digest complex glycans into short fatty acids directly benefits their host, prompting a predominant role of these species in gut homeostasis and host nutrition[356]. PUL-encoded complexes therefore play a pivotal role in the physiology of both Bacteroidetes and host[372,379].

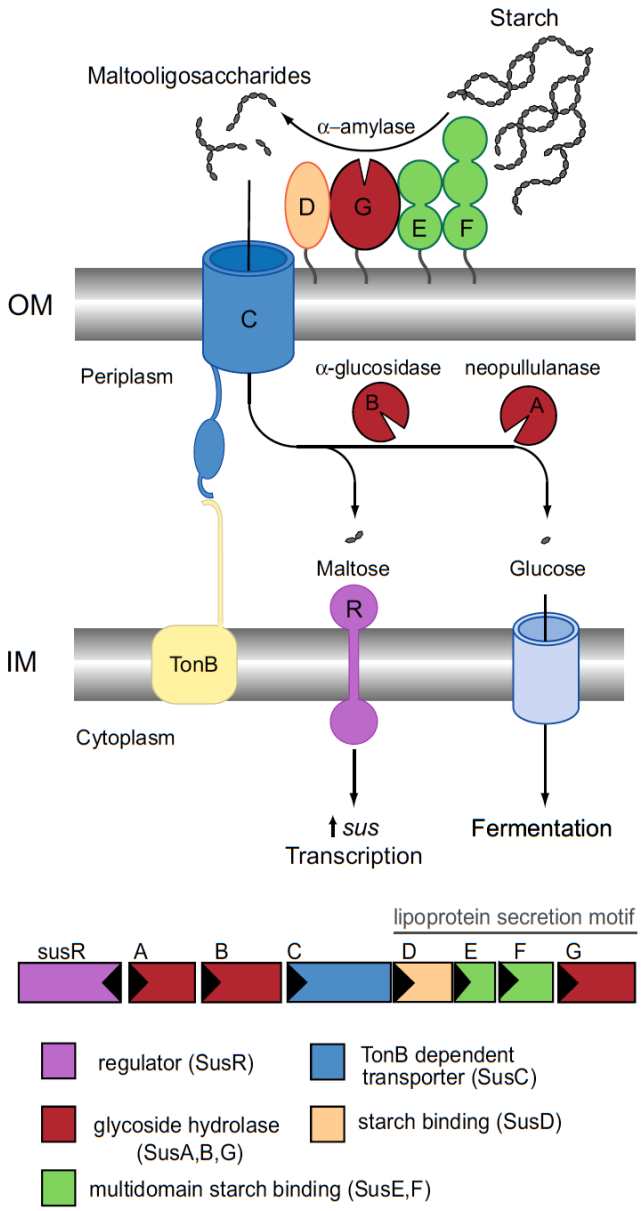


Figure 20. The Sus (starch utilization system) of *Bacteroides thetaiotaomicron*
Functional model of Bacteroidetes glycan foraging based on the *B. thetaiotaomicron* Sus complex. Starch is bound to the cell by the surface exposed lipoproteins SusD, -E and -F. The surface exposed amylase SusG then cleaves starch into smaller oligosaccharides that are subsequently transported across the outer membrane by the TonB-dependent receptor SusC. Oligosaccharides are further degraded into mono- or disaccharides by the periplasmic enzymes SusA and SusB. Liberated saccharides serve as signal for the transcriptional regulator SusR that activates *sus* gene expression. Monosaccharides are finally imported across the cytoplasmic membrane to be metabolized. The genetic organization of the *sus* operon is shown below. Black arrowheads indicate transcription orientation of each gene. Adapted from [371].

Capnocytophaga canimorsus – dog commensal and human pathogen

C. canimorsus is a commensal of the oral flora of dogs and cats and an opportunistic pathogen for humans, causing septicemia as well as meningitidis upon transmission[362,363,389]. Although the incidence of the infection is rather low, with a total of 484 confirmed cases since 1961, the outcome can be fatal in up to 26 % of the infections even with appropriate antibiotic treatment. Our reference strain *C. canimorsus* 5[160], a strain isolated from a human fatal septicemia[390], has 13 PUL and these encode more than half of all proteins exposed at the bacterial surface[160].

The *C. canimorsus* Sus homolog is encoded by PUL12 and was shown to degrade starch as well as glycogen (unpublished data). Interestingly, 3 other of these PUL-encoded complexes play critical roles in the biology of *C. canimorsus* 5. PUL3 encodes a novel iron acquisition system (Ics) allowing the bacterium to fetch iron from human transferrin, thus being indispensable for growth in human serum and therefore representing a potential virulence factor (see chapter 4)[377]. PUL5, encoding the Gpd complex, enables the bacterium to harvest amino sugars from the surface of eukaryotic cells as well as from soluble serum glycoproteins[391-393]. This capacity has recently been linked to the inability of the bacterium to synthesize N-acetylglucosamine, a key component for peptidoglycan assembly, highlighting its reliance on external amino sugar availability[376]. Finally, PUL9 has been shown to encode the Muc complex, devoted to mucin degradation[376]. This complex has also been shown to be able to compensate for the loss of the PUL5 encoded Gpd complex if an excess amount of mucin is provided, indicating a high degree of adaptation of the bacterium to its ecological niche, the dog's mouth, rich in mucin[376].

Aim of the thesis

Recent studies showed that surface-exposed lipoproteins, and in particular PUL-associated lipoproteins, are highly abundant in Bacteroidetes. Indeed, in human pathogens such as *C. canimorsus* and *B. fragilis*, they represent more than half of all surface displayed proteins[160,394]. To date however, little is known on how these lipoproteins reach the bacterial surface. As already mentioned, while some lipoprotein surface localization mechanisms have been explored in Proteobacteria, these often represent unique cases of a given secretion system and do not seem to operate at a larger scale as would be required in Bacteroidetes (see "Surface exposed lipoproteins in bacteria"). In addition, with the exception of the T1SS, T9SS and some occurrences of T6SS, none of the above-described secretion systems seem to be present in Bacteroidetes[395,396]. Similarly, homologs to the recently described Slam1 and 2 proteins, transporting lipoproteins to the surface of *N. meningitidis*[326], seem to be absent in this phylum. Finally, no protein homologous to LolB, which inserts lipoproteins in the OM of *E. coli*, could be identified in Bacteroidetes.

Due to the high physiological importance of PUL-encoded complexes, both for commensalism and pathogenesis, the question of how Bacteroidetes massively transport lipoproteins across their outer membrane and present them at their surface represents an interesting and fascinating topic. In this regard, it is also important to note that not all lipoproteins transported to the OM of Bacteroidetes are necessarily surface exposed; some lipoproteins remain intracellular and thus face the periplasm[392]. This therefore raises the question of how surface exposed and periplasmic lipoproteins are distinguished from each other and thus correctly targeted to their final destinations.

The work in this thesis therefore focuses on two main aspects: i) the characterization of the signal discriminating intracellular and extracellular lipoproteins, ultimately resulting in their final subcellular localization and ii) the identification of the machinery that transports lipoproteins across the OM to the bacterial surface in Bacteroidetes. This work was essentially performed in our model organism *C. canimorsus* 5.

References

1 Silhavy, T. J. *et al.* The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**, (2010).
2 Brown, L. *et al.* Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat Rev Microbiol* **13**, (2015).
3 Driessen, A. J. & Nouwen, N. Protein translocation across the bacterial cytoplasmic membrane. *Annu Rev Biochem* **77**, (2008).
4 Bos, M. P. *et al.* Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**, (2007).
5 Gu, Y. *et al.* Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531**, (2016).
6 Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, (2011).
7 Putker, F. *et al.* Transport of lipopolysaccharide to the Gram-negative bacterial cell surface. *FEMS Microbiol Rev* **39**, (2015).
8 Gensollen, T. *et al.* How colonization by microbiota in early life shapes the immune system. *Science* **352**, (2016).
9 Collinson, I. *et al.* Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Philos Trans R Soc Lond B Biol Sci* **370**, (2015).
10 du Plessis, D. J. *et al.* The Sec translocase. *Biochim Biophys Acta* **1808**, (2011).
11 Pugsley, A. P. The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**, (1993).
12 von Heijne, G. The signal peptide. *J Membr Biol* **115**, (1990).
13 Lee, H. C. & Bernstein, H. D. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc Natl Acad Sci U S A* **98**, (2001).
14 Koch, H. G. *et al.* In vitro studies with purified components reveal signal recognition particle (SRP) and SecA/SecB as constituents of two independent protein-targeting pathways of Escherichia coli. *Mol Biol Cell* **10**, (1999).
15 Xu, Z. *et al.* Crystal structure of the bacterial protein export chaperone secB. *Nat Struct Biol* **7**, (2000).
16 Randall, L. L. Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation. *Cell* **33**, (1983).
17 Arkowitz, R. A. *et al.* Translocation can drive the unfolding of a preprotein domain. *EMBO J* **12**, (1993).
18 Bechtluft, P. *et al.* Direct observation of chaperone-induced changes in a protein folding pathway. *Science* **318**, (2007).
19 Fekkes, P. *et al.* Preprotein transfer to the Escherichia coli translocase requires the co-operative binding of SecB and the signal sequence to SecA. *Mol Microbiol* **29**, (1998).
20 Lill, R. *et al.* SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. *EMBO J* **8**, (1989).
21 Lill, R. *et al.* The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell* **60**, (1990).
22 Fekkes, P. *et al.* The molecular chaperone SecB is released from the carboxy-terminus of SecA during initiation of precursor protein translocation. *EMBO J* **16**, (1997).
23 Brundage, L. *et al.* The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62**, (1990).
24 Hizlan, D. *et al.* Structure of the SecY complex unlocked by a preprotein mimic. *Cell Rep* **1**, (2012).
25 Duong, F. & Wickner, W. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *EMBO J* **16**, (1997).
26 Driessen, A. J. Precursor protein translocation by the Escherichia coli translocase is directed by the protonmotive force. *EMBO J* **11**, (1992).
27 Schiebel, E. *et al.* Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64**, (1991).
28 van der Wolk, J. P. *et al.* The catalytic cycle of the escherichia coli SecA ATPase comprises two distinct preprotein translocation events. *EMBO J* **16**, (1997).
29 Dalbey, R. E. *et al.* Membrane proteases in the bacterial protein secretion and quality control pathway. *Microbiol Mol Biol Rev* **76**, (2012).
30 Berks, B. C. The twin-arginine protein translocation pathway. *Annu Rev Biochem* **84**, (2015).
31 Mould, R. M. & Robinson, C. A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *J Biol Chem* **266**, (1991).
32 Yahr, T. L. & Wickner, W. T. Functional reconstitution of bacterial Tat translocation in vitro. *EMBO J* **20**, (2001).
33 Alami, M. *et al.* Differential interactions between a twin-arginine signal peptide and its translocase in Escherichia coli. *Mol Cell* **12**, (2003).

34 Cline, K. & Mori, H. Thylakoid DeltapH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. *J Cell Biol* **154**, (2001).
35 Mori, H. & Cline, K. A twin arginine signal peptide and the pH gradient trigger reversible assembly of the thylakoid [Delta]pH/Tat translocase. *J Cell Biol* **157**, (2002).
36 Dabney-Smith, C. & Cline, K. Clustering of C-terminal stromal domains of Tha4 homo-oligomers during translocation by the Tat protein transport system. *Mol Biol Cell* **20**, (2009).
37 Dabney-Smith, C. *et al.* Oligomers of Tha4 organize at the thylakoid Tat translocase during protein transport. *J Biol Chem* **281**, (2006).
38 Alcock, F. *et al.* Live cell imaging shows reversible assembly of the TatA component of the twin-arginine protein transport system. *Proc Natl Acad Sci U S A* **110**, (2013).
39 Rose, P. *et al.* Substrate-dependent assembly of the Tat translocase as observed in live Escherichia coli cells. *PLoS One* **8**, (2013).
40 Luke, I. *et al.* Proteolytic processing of Escherichia coli twin-arginine signal peptides by LepB. *Arch Microbiol* **191**, (2009).
41 Palmer, T. & Berks, B. C. The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* **10**, (2012).
42 Goemans, C. *et al.* Folding mechanisms of periplasmic proteins. *Biochim Biophys Acta* **1843**, (2014).
43 Merdanovic, M. *et al.* Protein quality control in the bacterial periplasm. *Annu Rev Microbiol* **65**, (2011).
44 Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**, (2003).
45 Han, L. *et al.* Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat Struct Mol Biol* **23**, (2016).
46 Dong, H. *et al.* Structural basis for outer membrane lipopolysaccharide insertion. *Nature* **511**, (2014).
47 Qiao, S. *et al.* Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, (2014).
48 Pautsch, A. & Schulz, G. E. Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**, (1998).
49 Arora, A. *et al.* Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* **8**, (2001).
50 Ferguson, A. D. *et al.* Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide. *Science* **282**, (1998).
51 Locher, K. P. *et al.* Transmembrane signaling across the ligand-gated FhuA receptor: crystal structures of free and ferrichrome-bound states reveal allosteric changes. *Cell* **95**, (1998).
52 Schirmer, T. *et al.* Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267**, (1995).
53 Koronakis, V. *et al.* Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **405**, (2000).
54 Wimley, W. C. The versatile beta-barrel membrane protein. *Curr Opin Struct Biol* **13**, (2003).
55 Tamm, L. K. *et al.* Folding and assembly of beta-barrel membrane proteins. *Biochim Biophys Acta* **1666**, (2004).
56 Sklar, J. G. *et al.* Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev* **21**, (2007).
57 Vertommen, D. *et al.* Characterization of the role of the Escherichia coli periplasmic chaperone SurA using differential proteomics. *Proteomics* **9**, (2009).
58 Hagan, C. L. *et al.* Reconstitution of outer membrane protein assembly from purified components. *Science* **328**, (2010).
59 Behrens, S. *et al.* The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J* **20**, (2001).
60 Bitto, E. & McKay, D. B. Binding of phage-display-selected peptides to the periplasmic chaperone protein SurA mimics binding of unfolded outer membrane proteins. *FEBS Lett* **568**, (2004).
61 Hennecke, G. *et al.* The periplasmic chaperone SurA exploits two features characteristic of integral outer membrane proteins for selective substrate recognition. *J Biol Chem* **280**, (2005).
62 Lazar, S. W. & Kolter, R. SurA assists the folding of Escherichia coli outer membrane proteins. *J Bacteriol* **178**, (1996).
63 Rouviere, P. E. & Gross, C. A. SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev* **10**, (1996).
64 Ureta, A. R. *et al.* Kinetic analysis of the assembly of the outer membrane protein LamB in Escherichia coli mutants each lacking a secretion or targeting factor in a different cellular compartment. *J Bacteriol* **189**, (2007).
65 Missiakas, D. *et al.* New components of protein folding in extracytoplasmic compartments of Escherichia coli SurA, FkpA and Skp/OmpH. *Mol Microbiol* **21**, (1996).
66 Chen, R. & Henning, U. A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. *Mol Microbiol* **19**, (1996).

67 Qu, J. *et al.* The trimeric periplasmic chaperone Skp of Escherichia coli forms 1:1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions. *J Mol Biol* **374**, (2007).
68 Jarchow, S. *et al.* Identification of potential substrate proteins for the periplasmic Escherichia coli chaperone Skp. *Proteomics* **8**, (2008).
69 Walton, T. A. & Sousa, M. C. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol Cell* **15**, (2004).
70 Rizzitello, A. E. *et al.* Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. *J Bacteriol* **183**, (2001).
71 Schwalm, J. *et al.* Role for Skp in LptD assembly in Escherichia coli. *J Bacteriol* **195**, (2013).
72 de Cock, H. *et al.* Non-lamellar structure and negative charges of lipopolysaccharides required for efficient folding of outer membrane protein PhoE of Escherichia coli. *J Biol Chem* **274**, (1999).
73 Qu, J. *et al.* Binding regions of outer membrane protein A in complexes with the periplasmic chaperone Skp. A site-directed fluorescence study. *Biochemistry* **48**, (2009).
74 Burmann, B. M. *et al.* Revisiting the interaction between the chaperone Skp and lipopolysaccharide. *Biophys J* **108**, (2015).
75 Meltzer, M. *et al.* Structure, function and regulation of the conserved serine proteases DegP and DegS of Escherichia coli. *Res Microbiol* **160**, (2009).
76 Subrini, O. & Betton, J. M. Assemblies of DegP underlie its dual chaperone and protease function. *FEMS Microbiol Lett* **296**, (2009).
77 Kleinschmidt, J. H. Membrane protein folding on the example of outer membrane protein A of Escherichia coli. *Cell Mol Life Sci* **60**, (2003).
78 Ricci, D. P. & Silhavy, T. J. The Bam machine: a molecular cooper. *Biochim Biophys Acta* **1818**, (2012).
79 Voulhoux, R. *et al.* Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**, (2003).
80 Wu, T. *et al.* Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell* **121**, (2005).
81 Arnold, T. *et al.* Omp85 from the thermophilic cyanobacterium Thermosynechococcus elongatus differs from proteobacterial Omp85 in structure and domain composition. *J Biol Chem* **285**, (2010).
82 Schleiff, E. *et al.* Omp85 in eukaryotic systems: one protein family with distinct functions. *Biol Chem* **392**, (2011).
83 Gentle, I. *et al.* The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J Cell Biol* **164**, (2004).
84 Kim, S. *et al.* Structure and function of an essential component of the outer membrane protein assembly machine. *Science* **317**, (2007).
85 Gentle, I. E. *et al.* Molecular architecture and function of the Omp85 family of proteins. *Mol Microbiol* **58**, (2005).
86 Sanchez-Pulido, L. *et al.* POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem Sci* **28**, (2003).
87 Sklar, J. G. *et al.* Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli. *Proc Natl Acad Sci U S A* **104**, (2007).
88 Bennion, D. *et al.* Dissection of beta-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of Escherichia coli. *Mol Microbiol* **77**, (2010).
89 Charlson, E. S. *et al.* Differential effects of yfgL mutation on Escherichia coli outer membrane proteins and lipopolysaccharide. *J Bacteriol* **188**, (2006).
90 Malinverni, J. C. *et al.* YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol Microbiol* **61**, (2006).
91 Vuong, P. *et al.* Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry. *J Bacteriol* **190**, (2008).
92 Noinaj, N. *et al.* Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* **501**, (2013).
93 Ni, D. *et al.* Structural and functional analysis of the beta-barrel domain of BamA from Escherichia coli. *FASEB J* **28**, (2014).
94 Albrecht, R. *et al.* Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallogr D Biol Crystallogr* **70**, (2014).
95 Noinaj, N. *et al.* The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex. *J Mol Biol* **407**, (2011).
96 Dong, C. *et al.* Structure of Escherichia coli BamB and its interaction with POTRA domains of BamA. *Acta Crystallogr D Biol Crystallogr* **68**, (2012).
97 Chen, Z. *et al.* Structural basis for the interaction of BamB with the POTRA3-4 domains of BamA. *Acta Crystallogr D Struct Biol* **72**, (2016).
98 Sandoval, C. M. *et al.* Crystal structure of BamD: an essential component of the beta-Barrel assembly machinery of gram-negative bacteria. *J Mol Biol* **409**, (2011).
99 Dong, C. *et al.* Structure of Escherichia coli BamD and its functional implications in outer membrane protein assembly. *Acta Crystallogr D Biol Crystallogr* **68**, (2012).

100 Kim, K. H. *et al.* Structural characterization of Escherichia coli BamE, a lipoprotein component of the beta-barrel assembly machinery complex. *Biochemistry* **50**, (2011).
101 Chimalakonda, G. *et al.* Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **108**, (2011).
102 Matsuyama, S. *et al.* A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of Escherichia coli. *EMBO J* **16**, (1997).
103 Clavel, T. *et al.* TolB protein of Escherichia coli K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Mol Microbiol* **29**, (1998).
104 Cho, S. H. *et al.* Detecting envelope stress by monitoring beta-barrel assembly. *Cell* **159**, (2014).
105 Typas, A. *et al.* Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell* **143**, (2010).
106 Laubacher, M. E. & Ades, S. E. The Rcs phosphorelay is a cell envelope stress response activated by peptidoglycan stress and contributes to intrinsic antibiotic resistance. *J Bacteriol* **190**, (2008).
107 Uehara, T. *et al.* LytM-domain factors are required for daughter cell separation and rapid ampicillin-induced lysis in Escherichia coli. *J Bacteriol* **191**, (2009).
108 Maeda, S. & Omata, T. Substrate-binding lipoprotein of the cyanobacterium Synechococcus sp. strain PCC 7942 involved in the transport of nitrate and nitrite. *J Biol Chem* **272**, (1997).
109 Zgurskaya, H. I. & Nikaido, H. Cross-linked complex between oligomeric periplasmic lipoprotein AcrA and the inner-membrane-associated multidrug efflux pump AcrB from Escherichia coli. *J Bacteriol* **182**, (2000).
110 McBride, M. J. Cytophaga-flavobacterium gliding motility. *J Mol Microbiol Biotechnol* **7**, (2004).
111 Kovacs-Simon, A. *et al.* Lipoproteins of bacterial pathogens. *Infect Immun* **79**, (2011).
112 Paradis-Bleau, C. *et al.* Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. *Cell* **143**, (2010).
113 Yakushi, T. *et al.* Lethality of the covalent linkage between mislocalized major outer membrane lipoprotein and the peptidoglycan of Escherichia coli. *J Bacteriol* **179**, (1997).
114 Miyadai, H. *et al.* Effects of lipoprotein overproduction on the induction of DegP (HtrA) involved in quality control in the Escherichia coli periplasm. *J Biol Chem* **279**, (2004).
115 Inouye, S. *et al.* Amino acid sequence for the peptide extension on the prolipoprotein of the Escherichia coli outer membrane. *Proc Natl Acad Sci U S A* **74**, (1977).
116 Sugai, M. & Wu, H. C. Export of the outer membrane lipoprotein is defective in secD, secE, and secF mutants of Escherichia coli. *J Bacteriol* **174**, (1992).
117 Watanabe, T. *et al.* Synthesis and export of the outer membrane lipoprotein in Escherichia coli mutants defective in generalized protein export. *J Bacteriol* **170**, (1988).
118 Hutchings, M. I. *et al.* Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol* **17**, (2009).
119 McDonough, J. A. *et al.* The twin-arginine translocation pathway of Mycobacterium smegmatis is functional and required for the export of mycobacterial beta-lactamases. *J Bacteriol* **187**, (2005).
120 Thompson, B. J. *et al.* Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium Streptomyces coelicolor. *Mol Microbiol* **77**, (2010).
121 Widdick, D. A. *et al.* The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor. *Proc Natl Acad Sci U S A* **103**, (2006).
122 Babu, M. M. *et al.* A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol* **188**, (2006).
123 Braun, V., and H. C. Wu. in *Bacterial cell wall* Vol. 27 319 (Elsevier, 1993).
124 Hayashi, S. & Wu, H. C. Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, (1990).
125 Hantke, K. & Braun, V. Covalent binding of lipid to protein. Diglyceride and amide-linked fatty acid at the N-terminal end of the murein-lipoprotein of the Escherichia coli outer membrane. *Eur J Biochem* **34**, (1973).
126 Sankaran, K. & Wu, H. C. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem* **269**, (1994).
127 Dev, I. K. & Ray, P. H. Rapid assay and purification of a unique signal peptidase that processes the prolipoprotein from Escherichia coli B. *J Biol Chem* **259**, (1984).
128 Hussain, M. *et al.* Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the Escherichia coli outer membrane. *J Biol Chem* **257**, (1982).
129 Tokunaga, M. *et al.* Post-translational modification and processing of Escherichia coli prolipoprotein in vitro. *Proc Natl Acad Sci U S A* **79**, (1982).
130 Fukuda, A. *et al.* Aminoacylation of the N-terminal cysteine is essential for Lol-dependent release of lipoproteins from membranes but does not depend on lipoprotein sorting signals. *J Biol Chem* **277**, (2002).
131 Gupta, S. D. *et al.* Characterization of a temperature-sensitive mutant of Salmonella typhimurium defective in apolipoprotein N-acyltransferase. *J Biol Chem* **268**, (1993).
132 Gupta, S. D. & Wu, H. C. Identification and subcellular localization of apolipoprotein N-acyltransferase in Escherichia coli. *FEMS Microbiol Lett* **62**, (1991).

133 Tschumi, A. *et al.* Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria. *J Biol Chem* **284**, (2009).
134 Widdick, D. A. *et al.* Dissecting the complete lipoprotein biogenesis pathway in Streptomyces scabies. *Mol Microbiol* **80**, (2011).
135 Yakushi, T. *et al.* A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nat Cell Biol* **2**, (2000).
136 Yakushi, T. *et al.* LolA-dependent release of a lipid-modified protein from the inner membrane of Escherichia coli requires nucleoside triphosphate. *J Biol Chem* **273**, (1998).
137 Narita, S. *et al.* Disruption of lolCDE, encoding an ATP-binding cassette transporter, is lethal for Escherichia coli and prevents release of lipoproteins from the inner membrane. *J Bacteriol* **184**, (2002).
138 Yasuda, M. *et al.* Membrane topology and functional importance of the periplasmic region of ABC transporter LolCDE. *Biosci Biotechnol Biochem* **73**, (2009).
139 Okuda, S. & Tokuda, H. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proc Natl Acad Sci U S A* **106**, (2009).
140 Taniguchi, N. & Tokuda, H. Molecular events involved in a single cycle of ligand transfer from an ATP binding cassette transporter, LolCDE, to a molecular chaperone, LolA. *J Biol Chem* **283**, (2008).
141 Matsuyama, S. *et al.* A novel periplasmic carrier protein involved in the sorting and transport of Escherichia coli lipoproteins destined for the outer membrane. *EMBO J* **14**, (1995).
142 Tajima, T. *et al.* Genetic analyses of the in vivo function of LolA, a periplasmic chaperone involved in the outer membrane localization of Escherichia coli lipoproteins. *FEBS Lett* **439**, (1998).
143 Nakada, S. *et al.* Structural investigation of the interaction between LolA and LolB using NMR. *J Biol Chem* **284**, (2009).
144 Takeda, K. *et al.* Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *EMBO J* **22**, (2003).
145 Okuda, S. *et al.* A short helix in the C-terminal region of LolA is important for the specific membrane localization of lipoproteins. *FEBS Lett* **582**, (2008).
146 Oguchi, Y. *et al.* Opening and closing of the hydrophobic cavity of LolA coupled to lipoprotein binding and release. *J Biol Chem* **283**, (2008).
147 Hayashi, Y. *et al.* Roles of the protruding loop of factor B essential for the localization of lipoproteins (LolB) in the anchoring of bacterial triacylated proteins to the outer membrane. *J Biol Chem* **289**, (2014).
148 Taniguchi, N. *et al.* Mechanisms underlying energy-independent transfer of lipoproteins from LolA to LolB, which have similar unclosed {beta}-barrel structures. *J Biol Chem* **280**, (2005).
149 Miyamoto, A. *et al.* Mutant of LolA, a lipoprotein-specific molecular chaperone of Escherichia coli, defective in the transfer of lipoproteins to LolB. *Biochem Biophys Res Commun* **287**, (2001).
150 Masuda, K. *et al.* Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proc Natl Acad Sci U S A* **99**, (2002).
151 Yamaguchi, K. *et al.* A single amino acid determinant of the membrane localization of lipoproteins in E. coli. *Cell* **53**, (1988).
152 Gennity, J. M. & Inouye, M. The protein sequence responsible for lipoprotein membrane localization in Escherichia coli exhibits remarkable specificity. *J Biol Chem* **266**, (1991).
153 Terada, M. *et al.* Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *J Biol Chem* **276**, (2001).
154 Lewenza, S. *et al.* Direct visualization of red fluorescent lipoproteins indicates conservation of the membrane sorting rules in the family Enterobacteriaceae. *J Bacteriol* **188**, (2006).
155 Lewenza, S. *et al.* Novel inner membrane retention signals in Pseudomonas aeruginosa lipoproteins. *J Bacteriol* **190**, (2008).
156 Tanaka, S. Y. *et al.* Characterization of the Pseudomonas aeruginosa Lol system as a lipoprotein sorting mechanism. *J Biol Chem* **282**, (2007).
157 Narita, S. & Tokuda, H. Amino acids at positions 3 and 4 determine the membrane specificity of Pseudomonas aeruginosa lipoproteins. *J Biol Chem* **282**, (2007).
158 Hara, T. *et al.* Mechanism underlying the inner membrane retention of Escherichia coli lipoproteins caused by Lol avoidance signals. *J Biol Chem* **278**, (2003).
159 Sutcliffe, I. C. *et al.* A phylum level analysis reveals lipoprotein biosynthesis to be a fundamental property of bacteria. *Protein Cell* **3**, (2012).
160 Manfredi, P. *et al.* The genome and surface proteome of Capnocytophaga canimorsus reveal a key role of glycan foraging systems in host glycoproteins deglycosylation. *Mol Microbiol* **81**, (2011).
161 Zuckert, W. R. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta* **1843**, (2014).
162 Murray, G. L. *et al.* Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of Salmonella enterica serovar Typhimurium with macrophages and complement. *J Bacteriol* **188**, (2006).
163 Raetz, C. R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, (2002).
164 Raetz, C. R. *et al.* Lipid A modification systems in gram-negative bacteria. *Annu Rev Biochem* **76**, (2007).

165 Kalynych, S. *et al.* Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiol Rev* **38**, (2014).

166 Doerrler, W. T. *et al.* MsbA-dependent translocation of lipids across the inner membrane of Escherichia coli. *J Biol Chem* **279**, (2004).

167 Polissi, A. & Georgopoulos, C. Mutational analysis and properties of the msbA gene of Escherichia coli, coding for an essential ABC family transporter. *Mol Microbiol* **20**, (1996).

168 Zhou, Z. *et al.* Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J Biol Chem* **273**, (1998).

169 Ward, A. *et al.* Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc Natl Acad Sci U S A* **104**, (2007).

170 Doerrler, W. T. & Raetz, C. R. ATPase activity of the MsbA lipid flippase of Escherichia coli. *J Biol Chem* **277**, (2002).

171 Sperandeo, P. *et al.* Characterization of lptA and lptB, two essential genes implicated in lipopolysaccharide transport to the outer membrane of Escherichia coli. *J Bacteriol* **189**, (2007).

172 Sperandeo, P. *et al.* Non-essential KDO biosynthesis and new essential cell envelope biogenesis genes in the Escherichia coli yrbG-yhbG locus. *Res Microbiol* **157**, (2006).

173 Tran, A. X. *et al.* The LptA protein of Escherichia coli is a periplasmic lipid A-binding protein involved in the lipopolysaccharide export pathway. *J Biol Chem* **283**, (2008).

174 Suits, M. D. *et al.* Novel structure of the conserved gram-negative lipopolysaccharide transport protein A and mutagenesis analysis. *J Mol Biol* **380**, (2008).

175 Okuda, S. *et al.* Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. *Science* **338**, (2012).

176 Merten, J. A. *et al.* Concentration-dependent oligomerization and oligomeric arrangement of LptA. *Protein Sci* **21**, (2012).

177 Santambrogio, C. *et al.* LptA assembles into rod-like oligomers involving disorder-to-order transitions. *J Am Soc Mass Spectrom* **24**, (2013).

178 Sherman, D. J. *et al.* Decoupling catalytic activity from biological function of the ATPase that powers lipopolysaccharide transport. *Proc Natl Acad Sci U S A* **111**, (2014).

179 Wang, Z. *et al.* Structural and functional studies of conserved nucleotide-binding protein LptB in lipopolysaccharide transport. *Biochem Biophys Res Commun* **452**, (2014).

180 Narita, S. & Tokuda, H. Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS Lett* **583**, (2009).

181 Ruiz, N. *et al.* Identification of two inner-membrane proteins required for the transport of lipopolysaccharide to the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **105**, (2008).

182 Villa, R. *et al.* The Escherichia coli Lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *J Bacteriol* **195**, (2013).

183 Sperandeo, P. *et al.* Functional analysis of the protein machinery required for transport of lipopolysaccharide to the outer membrane of Escherichia coli. *J Bacteriol* **190**, (2008).

184 Tran, A. X. *et al.* Structure and functional analysis of LptC, a conserved membrane protein involved in the lipopolysaccharide export pathway in Escherichia coli. *J Biol Chem* **285**, (2010).

185 Sestito, S. E. *et al.* Functional characterization of E. coli LptC: interaction with LPS and a synthetic ligand. *Chembiochem* **15**, (2014).

186 Wu, T. *et al.* Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **103**, (2006).

187 Bos, M. P. *et al.* Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface. *Proc Natl Acad Sci U S A* **101**, (2004).

188 Braun, M. & Silhavy, T. J. Imp/OstA is required for cell envelope biogenesis in Escherichia coli. *Mol Microbiol* **45**, (2002).

189 Ruiz, N. *et al.* Nonconsecutive disulfide bond formation in an essential integral outer membrane protein. *Proc Natl Acad Sci U S A* **107**, (2010).

190 Malojcic, G. *et al.* LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface. *Proc Natl Acad Sci U S A* **111**, (2014).

191 Chng, S. S. *et al.* Characterization of the two-protein complex in Escherichia coli responsible for lipopolysaccharide assembly at the outer membrane. *Proc Natl Acad Sci U S A* **107**, (2010).

192 Bowyer, A. *et al.* Characterization of interactions between LPS transport proteins of the Lpt system. *Biochem Biophys Res Commun* **404**, (2011).

193 Sperandeo, P. *et al.* New insights into the Lpt machinery for lipopolysaccharide transport to the cell surface: LptA-LptC interaction and LptA stability as sensors of a properly assembled transenvelope complex. *J Bacteriol* **193**, (2011).

194 Freinkman, E. *et al.* Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. *Biochemistry* **51**, (2012).

195 Gu, Y. *et al.* Lipopolysaccharide is inserted into the outer membrane through an intramembrane hole, a lumen gate, and the lateral opening of LptD. *Structure* **23**, (2015).

196 Costa, T. R. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**, (2015).

197 Gerlach, R. G. & Hensel, M. Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297**, (2007).
198 Rego, A. T. *et al.* Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochem J* **425**, (2010).
199 Akatsuka, H. *et al.* The three genes lipB, lipC, and lipD involved in the extracellular secretion of the Serratia marcescens lipase which lacks an N-terminal signal peptide. *J Bacteriol* **177**, (1995).
200 Letoffe, S. *et al.* Secretion of the Serratia marcescens HasA protein by an ABC transporter. *J Bacteriol* **176**, (1994).
201 Guzzo, J. *et al.* The secretion genes of Pseudomonas aeruginosa alkaline protease are functionally related to those of Erwinia chrysanthemi proteases and Escherichia coli alpha-haemolysin. *Mol Microbiol* **5**, (1991).
202 Bleves, S. *et al.* Protein secretion systems in Pseudomonas aeruginosa: A wealth of pathogenic weapons. *Int J Med Microbiol* **300**, (2010).
203 Fath, M. J. *et al.* Purification and characterization of colicin V from Escherichia coli culture supernatants. *Biochemistry* **33**, (1994).
204 Kanonenberg, K. *et al.* Type I secretion systems - a story of appendices. *Res Microbiol* **164**, (2013).
205 Du, D. *et al.* Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**, (2014).
206 Galan, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**, (2006).
207 Cornelis, G. R. The type III secretion injectisome. *Nat Rev Microbiol* **4**, (2006).
208 Cornelis, G. R. The type III secretion injectisome, a complex nanomachine for intracellular 'toxin' delivery. *Biol Chem* **391**, (2010).
209 Buttner, D. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant- and animal-pathogenic bacteria. *Microbiol Mol Biol Rev* **76**, (2012).
210 Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, (2004).
211 Schraidt, O. & Marlovits, T. C. Three-dimensional model of Salmonella's needle complex at subnanometer resolution. *Science* **331**, (2011).
212 Akeda, Y. & Galan, J. E. Chaperone release and unfolding of substrates in type III secretion. *Nature* **437**, (2005).
213 Lara-Tejero, M. *et al.* A sorting platform determines the order of protein secretion in bacterial type III systems. *Science* **331**, (2011).
214 Izore, T. *et al.* Biogenesis, regulation, and targeting of the type III secretion system. *Structure* **19**, (2011).
215 Parsot, C. *et al.* The various and varying roles of specific chaperones in type III secretion systems. *Curr Opin Microbiol* **6**, (2003).
216 Hao, J. *et al.* Direct visualization of horizontal gene transfer in cotton plants. *J Hered* **105**, (2014).
217 Cascales, E. & Christie, P. J. The versatile bacterial type IV secretion systems. *Nat Rev Microbiol* **1**, (2003).
218 Cascales, E. & Christie, P. J. Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* **304**, (2004).
219 Alvarez-Martinez, C. E. & Christie, P. J. Biological diversity of prokaryotic type IV secretion systems. *Microbiol Mol Biol Rev* **73**, (2009).
220 Low, H. H. *et al.* Structure of a type IV secretion system. *Nature* **508**, (2014).
221 Christie, P. J. *et al.* Mechanism and structure of the bacterial type IV secretion systems. *Biochim Biophys Acta* **1843**, (2014).
222 Trokter, M. *et al.* Recent advances in the structural and molecular biology of type IV secretion systems. *Curr Opin Struct Biol* **27**, (2014).
223 Fronzes, R. *et al.* Structure of a type IV secretion system core complex. *Science* **323**, (2009).
224 Ripoll-Rozada, J. *et al.* Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems. *J Bacteriol* **195**, (2013).
225 Pukatzki, S. *et al.* Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* **103**, (2006).
226 Filloux, A. The rise of the Type VI secretion system. *F1000Prime Rep* **5**, (2013).
227 Ma, A. T. *et al.* Translocation of a Vibrio cholerae type VI secretion effector requires bacterial endocytosis by host cells. *Cell Host Microbe* **5**, (2009).
228 Chow, J. & Mazmanian, S. K. A pathobiont of the microbiota balances host colonization and intestinal inflammation. *Cell Host Microbe* **7**, (2010).
229 Hood, R. D. *et al.* A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. *Cell Host Microbe* **7**, (2010).
230 MacIntyre, D. L. *et al.* The Vibrio cholerae type VI secretion system displays antimicrobial properties. *Proc Natl Acad Sci U S A* **107**, (2010).
231 Ho, B. T. *et al.* A view to a kill: the bacterial type VI secretion system. *Cell Host Microbe* **15**, (2014).
232 Schwarz, S. *et al.* Burkholderia type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. *PLoS Pathog* **6**, (2010).

233 Boyer, F. *et al.* Dissecting the bacterial type VI secretion system by a genome wide in silico analysis: what can be learned from available microbial genomic resources? *BMC Genomics* **10**, (2009).
234 Russell, A. B. *et al.* A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host Microbe* **16**, (2014).
235 Zheng, J. & Leung, K. Y. Dissection of a type VI secretion system in Edwardsiella tarda. *Mol Microbiol* **66**, (2007).
236 Journet, L. & Cascales, E. The Type VI Secretion System in Escherichia coli and Related Species. *EcoSal Plus* **7**, (2016).
237 Zoued, A. *et al.* Architecture and assembly of the Type VI secretion system. *Biochim Biophys Acta* **1843**, (2014).
238 Leiman, P. G. *et al.* Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc Natl Acad Sci U S A* **106**, (2009).
239 Ma, L. S. *et al.* An IcmF family protein, ImpLM, is an integral inner membrane protein interacting with ImpKL, and its walker a motif is required for type VI secretion system-mediated Hcp secretion in Agrobacterium tumefaciens. *J Bacteriol* **191**, (2009).
240 Felisberto-Rodrigues, C. *et al.* Towards a structural comprehension of bacterial type VI secretion systems: characterization of the TssJ-TssM complex of an Escherichia coli pathovar. *PLoS Pathog* **7**, (2011).
241 Basler, M. *et al.* Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, (2012).
242 Taylor, N. M. *et al.* Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, (2016).
243 Kudryashev, M. *et al.* Structure of the type VI secretion system contractile sheath. *Cell* **160**, (2015).
244 Nivaskumar, M. & Francetic, O. Type II secretion system: a magic beanstalk or a protein escalator. *Biochim Biophys Acta* **1843**, (2014).
245 Korotkov, K. V. *et al.* The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat Rev Microbiol* **10**, (2012).
246 Reichow, S. L. *et al.* Structure of the cholera toxin secretion channel in its closed state. *Nat Struct Mol Biol* **17**, (2010).
247 Gray, M. D. *et al.* In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae. *Mol Microbiol* **79**, (2011).
248 Korotkov, K. V. *et al.* Structural and functional studies on the interaction of GspC and GspD in the type II secretion system. *PLoS Pathog* **7**, (2011).
249 Lu, C. *et al.* Hexamers of the type II secretion ATPase GspE from Vibrio cholerae with increased ATPase activity. *Structure* **21**, (2013).
250 Leo, J. C. *et al.* Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philos Trans R Soc Lond B Biol Sci* **367**, (2012).
251 Leyton, D. L. *et al.* From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. *Nat Rev Microbiol* **10**, (2012).
252 Wells, T. J. *et al.* Autotransporter proteins: novel targets at the bacterial cell surface. *FEMS Microbiol Lett* **274**, (2007).
253 Oomen, C. J. *et al.* Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23**, (2004).
254 van den Berg, B. Crystal structure of a full-length autotransporter. *J Mol Biol* **396**, (2010).
255 Roman-Hernandez, G. *et al.* Reconstitution of bacterial autotransporter assembly using purified components. *Elife* **3**, (2014).
256 Junker, M. *et al.* Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol Microbiol* **71**, (2009).
257 Pavlova, O. *et al.* Mechanistic link between beta barrel assembly and the initiation of autotransporter secretion. *Proc Natl Acad Sci U S A* **110**, (2013).
258 Gruss, F. *et al.* The structural basis of autotransporter translocation by TamA. *Nat Struct Mol Biol* **20**, (2013).
259 Selkrig, J. *et al.* Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol* **19**, (2012).
260 Wright, K. J. *et al.* Development of intracellular bacterial communities of uropathogenic Escherichia coli depends on type 1 pili. *Cell Microbiol* **9**, (2007).
261 Lillington, J. *et al.* Biogenesis and adhesion of type 1 and P pili. *Biochim Biophys Acta* **1840**, (2014).
262 Waksman, G. & Fronzes, R. Molecular architecture of bacterial type IV secretion systems. *Trends Biochem Sci* **35**, (2010).
263 Mu, X. Q. & Bullitt, E. Structure and assembly of P-pili: a protruding hinge region used for assembly of a bacterial adhesion filament. *Proc Natl Acad Sci U S A* **103**, (2006).
264 Hahn, E. *et al.* Exploring the 3D molecular architecture of Escherichia coli type 1 pili. *J Mol Biol* **323**, (2002).

265 Phan, G. *et al.* Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. *Nature* **474**, (2011).
266 Hultgren, S. J. *et al.* Chaperone-assisted assembly and molecular architecture of adhesive pili. *Annu Rev Microbiol* **45**, (1991).
267 Vetsch, M. *et al.* Pilus chaperones represent a new type of protein-folding catalyst. *Nature* **431**, (2004).
268 Barnhart, M. M. *et al.* PapD-like chaperones provide the missing information for folding of pilin proteins. *Proc Natl Acad Sci U S A* **97**, (2000).
269 Sauer, F. G. *et al.* Structural basis of chaperone function and pilus biogenesis. *Science* **285**, (1999).
270 Busch, A. & Waksman, G. Chaperone-usher pathways: diversity and pilus assembly mechanism. *Philos Trans R Soc Lond B Biol Sci* **367**, (2012).
271 Geibel, S. & Waksman, G. The molecular dissection of the chaperone-usher pathway. *Biochim Biophys Acta* **1843**, (2014).
272 Geibel, S. *et al.* Structural and energetic basis of folded-protein transport by the FimD usher. *Nature* **496**, (2013).
273 Remaut, H. *et al.* Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol Cell* **22**, (2006).
274 Verger, D. *et al.* Molecular mechanism of P pilus termination in uropathogenic Escherichia coli. *EMBO Rep* **7**, (2006).
275 Blanco, L. P. *et al.* Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol* **20**, (2012).
276 Evans, M. L. & Chapman, M. R. Curli biogenesis: order out of disorder. *Biochim Biophys Acta* **1843**, (2014).
277 Barnhart, M. M. & Chapman, M. R. Curli biogenesis and function. *Annu Rev Microbiol* **60**, (2006).
278 Hammar, M. *et al.* Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12. *Mol Microbiol* **18**, (1995).
279 Gibson, D. L. *et al.* AgfC and AgfE facilitate extracellular thin aggregative fimbriae synthesis in Salmonella enteritidis. *Microbiology* **153**, (2007).
280 Robinson, L. S. *et al.* Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein. *Mol Microbiol* **59**, (2006).
281 Goyal, P. *et al.* Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* **516**, (2014).
282 Taylor, J. D. *et al.* Atomic resolution insights into curli fiber biogenesis. *Structure* **19**, (2011).
283 Nenninger, A. A. *et al.* CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation. *Mol Microbiol* **81**, (2011).
284 Hammar, M. *et al.* Nucleator-dependent intercellular assembly of adhesive curli organelles in Escherichia coli. *Proc Natl Acad Sci U S A* **93**, (1996).
285 Hammer, N. D. *et al.* The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization. *Proc Natl Acad Sci U S A* **104**, (2007).
286 Nenninger, A. A. *et al.* Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF. *Proc Natl Acad Sci U S A* **106**, (2009).
287 Stanley, S. A. *et al.* Acute infection and macrophage subversion by Mycobacterium tuberculosis require a specialized secretion system. *Proc Natl Acad Sci U S A* **100**, (2003).
288 Hoffmann, C. *et al.* Disclosure of the mycobacterial outer membrane: cryo-electron tomography and vitreous sections reveal the lipid bilayer structure. *Proc Natl Acad Sci U S A* **105**, (2008).
289 Zuber, B. *et al.* Direct visualization of the outer membrane of mycobacteria and corynebacteria in their native state. *J Bacteriol* **190**, (2008).
290 Gey Van Pittius, N. C. *et al.* The ESAT-6 gene cluster of Mycobacterium tuberculosis and other high G+C Gram-positive bacteria. *Genome Biol* **2**, (2001).
291 Anderson, M. *et al.* EsaD, a secretion factor for the Ess pathway in Staphylococcus aureus. *J Bacteriol* **193**, (2011).
292 Chen, Y. H. *et al.* Characterization of EssB, a protein required for secretion of ESAT-6 like proteins in Staphylococcus aureus. *BMC Microbiol* **12**, (2012).
293 Garufi, G. *et al.* ESAT-6-like protein secretion in Bacillus anthracis. *J Bacteriol* **190**, (2008).
294 Ates, L. S. *et al.* Type VII Secretion: A Highly Versatile Secretion System. *Microbiol Spectr* **4**, (2016).
295 Houben, E. N. *et al.* Composition of the type VII secretion system membrane complex. *Mol Microbiol* **86**, (2012).
296 Houben, E. N. *et al.* Take five - Type VII secretion systems of Mycobacteria. *Biochim Biophys Acta* **1843**, (2014).
297 Solomonson, M. *et al.* Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion system. *J Biol Chem* **288**, (2013).
298 Korotkova, N. *et al.* Structure of the Mycobacterium tuberculosis type VII secretion system chaperone EspG5 in complex with PE25-PPE41 dimer. *Mol Microbiol* **94**, (2014).
299 Ekiert, D. C. & Cox, J. S. Structure of a PE-PPE-EspG complex from Mycobacterium tuberculosis reveals molecular specificity of ESX protein secretion. *Proc Natl Acad Sci U S A* **111**, (2014).

300 Renshaw, P. S. *et al.* Structure and function of the complex formed by the tuberculosis virulence factors CFP-10 and ESAT-6. *EMBO J* **24**, (2005).
301 Abdallah, A. M. *et al.* PPE and PE_PGRS proteins of Mycobacterium marinum are transported via the type VII secretion system ESX-5. *Mol Microbiol* **73**, (2009).
302 Ilghari, D. *et al.* Solution structure of the Mycobacterium tuberculosis EsxG.EsxH complex: functional implications and comparisons with other M. tuberculosis Esx family complexes. *J Biol Chem* **286**, (2011).
303 Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proc Natl Acad Sci U S A* **103**, (2006).
304 Nakayama, K. Porphyromonas gingivalis and related bacteria: from colonial pigmentation to the type IX secretion system and gliding motility. *J Periodontal Res* **50**, (2015).
305 McBride, M. J. *et al.* Novel features of the polysaccharide-digesting gliding bacterium Flavobacterium johnsoniae as revealed by genome sequence analysis. *Appl Environ Microbiol* **75**, (2009).
306 Sato, K. *et al.* A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A* **107**, (2010).
307 McBride, M. J. & Nakane, D. Flavobacterium gliding motility and the type IX secretion system. *Curr Opin Microbiol* **28**, (2015).
308 McBride, M. J. Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. *Annu Rev Microbiol* **55**, (2001).
309 Saiki, K. & Konishi, K. Identification of a Porphyromonas gingivalis novel protein sov required for the secretion of gingipains. *Microbiol Immunol* **51**, (2007).
310 Sato, K. *et al.* Identification of a new membrane-associated protein that influences transport/maturation of gingipains and adhesins of Porphyromonas gingivalis. *J Biol Chem* **280**, (2005).
311 Gorasia, D. G. *et al.* Structural Insights into the PorK and PorN Components of the Porphyromonas gingivalis Type IX Secretion System. *PLoS Pathog* **12**, (2016).
312 Veith, P. D. *et al.* Protein substrates of a novel secretion system are numerous in the Bacteroidetes phylum and have in common a cleavable C-terminal secretion signal, extensive post-translational modification, and cell-surface attachment. *J Proteome Res* **12**, (2013).
313 de Diego, I. *et al.* The outer-membrane export signal of Porphyromonas gingivalis type IX secretion system (T9SS) is a conserved C-terminal beta-sandwich domain. *Sci Rep* **6**, (2016).
314 Shoji, M. *et al.* Por secretion system-dependent secretion and glycosylation of Porphyromonas gingivalis hemin-binding protein 35. *PLoS One* **6**, (2011).
315 Chen, Y. Y. *et al.* The outer membrane protein LptO is essential for the O-deacylation of LPS and the co-ordinated secretion and attachment of A-LPS and CTD proteins in Porphyromonas gingivalis. *Mol Microbiol* **79**, (2011).
316 Konovalova, A. & Silhavy, T. J. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos Trans R Soc Lond B Biol Sci* **370**, (2015).
317 Wilson, M. M. & Bernstein, H. D. Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. *Trends Microbiol*, (2015).
318 Evans, K. L. *et al.* Eliminating a set of four penicillin binding proteins triggers the Rcs phosphorelay and Cpx stress responses in Escherichia coli. *J Bacteriol* **195**, (2013).
319 Farris, C. *et al.* Antimicrobial peptides activate the Rcs regulon through the outer membrane lipoprotein RcsF. *J Bacteriol* **192**, (2010).
320 Majdalani, N. & Gottesman, S. The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* **59**, (2005).
321 Castanie-Cornet, M. P. *et al.* RcsF is an outer membrane lipoprotein involved in the RcsCDB phosphorelay signaling pathway in Escherichia coli. *J Bacteriol* **188**, (2006).
322 Morgenthau, A. *et al.* Bacterial receptors for host transferrin and lactoferrin: molecular mechanisms and role in host-microbe interactions. *Future Microbiol* **8**, (2013).
323 Irwin, S. W. *et al.* Preparation and analysis of isogenic mutants in the transferrin receptor protein genes, tbpA and tbpB, from Neisseria meningitidis. *Mol Microbiol* **8**, (1993).
324 Pettersson, A. *et al.* Molecular characterization of LbpB, the second lactoferrin-binding protein of Neisseria meningitidis. *Mol Microbiol* **27**, (1998).
325 Welsch, J. A. & Ram, S. Factor H and neisserial pathogenesis. *Vaccine* **26 Suppl 8**, (2008).
326 Hooda, Y. *et al.* Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in Neisseria. *Nature Microbiology* **1**, (2016).
327 Lewis, L. A. & Dyer, D. W. Identification of an iron-regulated outer membrane protein of Neisseria meningitidis involved in the utilization of hemoglobin complexed to haptoglobin. *J Bacteriol* **177**, (1995).
328 Lewis, L. A. *et al.* Molecular characterization of hpuAB, the haemoglobin-haptoglobin-utilization operon of Neisseria meningitidis. *Mol Microbiol* **23**, (1997).
329 Coutte, L. *et al.* Surface anchoring of bacterial subtilisin important for maturation function. *Mol Microbiol* **49**, (2003).

330 Coutte, L. *et al.* Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *EMBO J* **20**, (2001).
331 van Ulsen, P. *et al.* A Neisserial autotransporter NalP modulating the processing of other autotransporters. *Mol Microbiol* **50**, (2003).
332 Roussel-Jazede, V. *et al.* NalP-mediated proteolytic release of lactoferrin-binding protein B from the meningococcal cell surface. *Infect Immun* **78**, (2010).
333 Roussel-Jazede, V. *et al.* Lipidation of the autotransporter NalP of Neisseria meningitidis is required for its function in the release of cell-surface-exposed proteins. *Microbiology* **159**, (2013).
334 Pugsley, A. P. *et al.* Analysis of the subcellular location of pullulanase produced by Escherichia coli carrying the pulA gene from Klebsiella pneumoniae strain UNF5023. *Mol Microbiol* **4**, (1990).
335 Pugsley, A. P. *et al.* Extracellular pullulanase of Klebsiella pneumoniae is a lipoprotein. *J Bacteriol* **166**, (1986).
336 d'Enfert, C. *et al.* Cloning and expression in Escherichia coli of the Klebsiella pneumoniae genes for production, surface localization and secretion of the lipoprotein pullulanase. *EMBO J* **6**, (1987).
337 Kornacker, M. G. *et al.* Outer membrane translocation of the extracellular enzyme pullulanase in Escherichia coli K12 does not require a fatty acylated N-terminal cysteine. *J Biol Chem* **266**, (1991).
338 Baldi, D. L. *et al.* The type II secretion system and its ubiquitous lipoprotein substrate, SslE, are required for biofilm formation and virulence of enteropathogenic Escherichia coli. *Infect Immun* **80**, (2012).
339 Shi, L. *et al.* Direct involvement of type II secretion system in extracellular translocation of Shewanella oneidensis outer membrane cytochromes MtrC and OmcA. *J Bacteriol* **190**, (2008).
340 Radolf, J. D. & Samuels, D. S. *Borrelia: Molecular Biology, Host Interaction and Pathogenesis*. (Caister Academic Press, 2010).
341 Barbour, A. G. & Hayes, S. F. Biology of Borrelia species. *Microbiol Rev* **50**, (1986).
342 Motaleb, M. A. *et al.* Borrelia burgdorferi periplasmic flagella have both skeletal and motility functions. *Proc Natl Acad Sci U S A* **97**, (2000).
343 Belisle, J. T. *et al.* Fatty acids of Treponema pallidum and Borrelia burgdorferi lipoproteins. *J Bacteriol* **176**, (1994).
344 Takayama, K. *et al.* Absence of lipopolysaccharide in the Lyme disease spirochete, Borrelia burgdorferi. *Infect Immun* **55**, (1987).
345 Chen, S. *et al.* Determination of Borrelia surface lipoprotein anchor topology by surface proteolysis. *J Bacteriol* **193**, (2011).
346 Kumru, O. S. *et al.* Surface localization determinants of Borrelia OspC/Vsp family lipoproteins. *J Bacteriol* **193**, (2011).
347 Schulze, R. J. & Zuckert, W. R. Borrelia burgdorferi lipoproteins are secreted to the outer surface by default. *Mol Microbiol* **59**, (2006).
348 Chen, S. & Zuckert, W. R. Probing the Borrelia burgdorferi surface lipoprotein secretion pathway using a conditionally folding protein domain. *J Bacteriol* **193**, (2011).
349 Schulze, R. J. *et al.* Translocation of Borrelia burgdorferi surface lipoprotein OspA through the outer membrane requires an unfolded conformation and can initiate at the C-terminus. *Mol Microbiol* **76**, (2010).
350 Lauber, C. L. *et al.* Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**, (2009).
351 Kirchman, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**, (2002).
352 Moran, N. A. *et al.* Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl Environ Microbiol* **71**, (2005).
353 Chang, H. H. *et al.* Complete Genome Sequence of "Candidatus Sulcia muelleri" ML, an Obligate Nutritional Symbiont of Maize Leafhopper (Dalbulus maidis). *Genome Announc* **3**, (2015).
354 Bjursell, M. K. *et al.* Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J Biol Chem* **281**, (2006).
355 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, (2005).
356 Koropatkin, N. M. *et al.* How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**, (2012).
357 Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, (2005).
358 Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, (2003).
359 Mysak, J. *et al.* Porphyromonas gingivalis: major periodontopathic pathogen overview. *J Immunol Res* **2014**, (2014).
360 Socransky, S. S. *et al.* Capnocytophaga: new genus of gram-negative gliding bacteria. III. Physiological characterization. *Arch Microbiol* **122**, (1979).
361 Brook, I. The role of anaerobic bacteria in bacteremia. *Anaerobe* **16**, (2010).
362 Butler, T. Capnocytophaga canimorsus: an emerging cause of sepsis, meningitis, and post-splenectomy infection after dog bites. *Eur J Clin Microbiol Infect Dis* **34**, (2015).

363 Gaastra, W. & Lipman, L. J. Capnocytophaga canimorsus. *Vet Microbiol* **140**, (2010).
364 Sears, C. L. Enterotoxigenic Bacteroides fragilis: a rogue among symbiotes. *Clin Microbiol Rev* **22**, (2009).
365 Sears, C. L. *et al.* Bacteroides fragilis subverts mucosal biology: from symbiont to colon carcinogenesis. *J Clin Invest* **124**, (2014).
366 Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, (2007).
367 Bauer, M. *et al.* Whole genome analysis of the marine Bacteroidetes'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**, (2006).
368 Terrapon, N. *et al.* Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, (2015).
369 Reeves, A. R. *et al.* A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol* **178**, (1996).
370 Reeves, A. R. *et al.* Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol* **179**, (1997).
371 Foley, M. H. *et al.* The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell Mol Life Sci*, (2016).
372 Martens, E. C. *et al.* Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, (2009).
373 Valentine, P. J. *et al.* Cloning and partial characterization of two chromosomal loci from Bacteroides ovatus that contain genes essential for growth on guar gum. *Appl Environ Microbiol* **58**, (1992).
374 Salyers, A. A. *et al.* Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Appl Environ Microbiol* **33**, (1977).
375 Martens, E. C. *et al.* Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, (2008).
376 Renzi, F. *et al.* Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. *MBio* **6**, (2015).
377 Manfredi, P. *et al.* New iron acquisition system in Bacteroidetes. *Infect Immun* **83**, (2015).
378 Cho, K. H. & Salyers, A. A. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **183**, (2001).
379 Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**, (2011).
380 Rogers, T. E. *et al.* Dynamic responses of Bacteroides thetaiotaomicron during growth on glycan mixtures. *Mol Microbiol* **88**, (2013).
381 Anderson, K. L. & Salyers, A. A. Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **171**, (1989).
382 Anderson, K. L. & Salyers, A. A. Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J Bacteriol* **171**, (1989).
383 Cameron, E. A. *et al.* Multidomain Carbohydrate-binding Proteins Involved in Bacteroides thetaiotaomicron Starch Metabolism. *J Biol Chem* **287**, (2012).
384 Koropatkin, N. M. *et al.* Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, (2008).
385 Koropatkin, N. M. & Smith, T. J. SusG: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* **18**, (2010).
386 Shipman, J. A. *et al.* Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol* **181**, (1999).
387 Xu, J. *et al.* Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* **5**, (2007).
388 Rosewarne, C. P. *et al.* Analysis of the bovine rumen microbiome reveals a diversity of Sus-like polysaccharide utilization loci from the bacterial phylum Bacteroidetes. *J Ind Microbiol Biotechnol* **41**, (2014).
389 Bobo, R. A. & Newton, E. J. A previously undescribed gram-negative bacillus causing septicemia and meningitis. *Am J Clin Pathol* **65**, (1976).
390 Shin, H. *et al.* Resistance of Capnocytophaga canimorsus to killing by human complement and polymorphonuclear leukocytes. *Infect Immun* **77**, (2009).
391 Mally, M. *et al.* Capnocytophaga canimorsus: a human pathogen feeding at the surface of epithelial cells and phagocytes. *PLoS Pathog* **4**, (2008).
392 Renzi, F. *et al.* The N-glycan glycoprotein deglycosylation complex (Gpd) from Capnocytophaga canimorsus deglycosylates human IgG. *PLoS Pathog* **7**, (2011).
393 Renzi, F. *et al.* Only a subset of C. canimorsus strains is dangerous for humans. *Emerg Microbes Infect* **4**, (2015).
394 Wilson, M. M. *et al.* Analysis of the outer membrane proteome and secretome of Bacteroides fragilis reveals a multiplicity of secretion mechanisms. *PLoS One* **10**, (2015).
395 Abby, S. S. *et al.* Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**, (2016).
396 McBride, M. J. & Zhu, Y. Gliding motility and Por secretion system genes are widespread among members of the phylum bacteroidetes. *J Bacteriol* **195**, (2013).

1. Defining the lipoprotein export signal of Bacteroidetes

1.1. Manuscript submitted: Identification of a new lipoprotein export signal in Gram-negative bacteria

Identification of a new lipoprotein export signal in Gram-negative bacteria

Frédéric Lauber[a], Guy Richard Cornelis[a], Francesco Renzi[a#]

[a]Département de Biologie, Unité de Recherche en Biologie des Microorganismes (URBM), Université de Namur, 5000 Namur, Belgium

#Address correspondence to Francesco Renzi, francesco.renzi@unamur.be

Abstract

Bacteria of the phylum Bacteroidetes, including commensals and opportunistic pathogens, harbor abundant surface-exposed multi-protein membrane complexes (Sus-like systems) involved in carbohydrate acquisition. These complexes have been mostly linked to commensalism and in some instances they have also been shown to play a role in pathogenesis. Sus-like systems are mainly composed of lipoproteins anchored to the outer membrane and facing the external milieu. This lipoprotein localization is uncommon in most studied Gram-negative bacteria while it is widespread in Bacteroidetes. Little is known on how these complexes assemble and in particular on how lipoproteins reach the bacterial surface. Here, by bioinformatic analyses, we identify a lipoprotein export signal (LES) at the N-terminus of surface-exposed lipoproteins of the human pathogen *Capnocytophaga canimorsus* corresponding to K-(D/E)$_2$ or Q-A-(D/E)$_2$. We show that, when introduced in sialidase SiaC, an intracellular lipoprotein, this signal is sufficient to target the protein to the cell surface. Mutational analysis of the LES in this reporter system showed that the amino acid composition, the position of the signal sequence and the global charge are critical for lipoprotein surface transport. These findings were further confirmed by the analysis of the LES of mucinase MucG, a naturally surface exposed *C. canimorsus* lipoprotein. Furthermore, we identify a LES in *Bacteroides fragilis* and *Flavobacterium johnsoniae* surface lipoproteins that allow *C. canimorsus* surface protein exposure, thus suggesting that Bacteroidetes share a common new bacterial lipoprotein export pathway that flips lipoproteins across the outer membrane.

Importance

Bacteria of the phylum Bacteroidetes are important human commensals and pathogens. Understanding their biology is therefore a key question for human health. A main feature of these bacteria is the presence of abundant lipoproteins at their surface that play a role in nutrient acquisition. To date, the underlying mechanism of lipoprotein transport is unknown. We show for the first time that Bacteroidetes surface lipoproteins share an N-terminal signal that drives surface localization. The localization and the overall negative charge of the lipoprotein export signal (LES) are crucial for its role. Overall, our findings provide the first evidence that Bacteroidetes are endowed with a new bacterial lipoprotein export pathway that flips lipoproteins across the outer membrane.

Introduction

Among Gram-negative bacteria**,** the phylum Bacteroidetes is composed of a large diversity of organisms widely distributed in the environment. Some are saprophytes such as *Flavobacteria,* found in soil[1] and aquatic environments[2], while others are commensals of animals. Among them, *Bacteroides* spp. are common members of the intestinal flora where they play a major role in gut homeostasis[3-7] while *Capnocytophaga* and *Porphyromonas* spp. are part of the oral flora[8,9]. *Bacteroides fragilis*, a commensal of the human intestine and *Capnocytophaga canimorsus*, a common member of the dog oral flora can cause severe systemic human infections[10-15] while *Porphyromonas gingivalis* causes severe periodontal diseases[8]. The wide distribution of these organisms reflects their high adaptability, partially due to their vast array of glycosylhydrolases allowing them to degrade nearly all types of carbohydrates they can encounter [7,16-19]. Interestingly, these enzymes are often surface exposed lipoproteins and are part of multi-protein outer membrane (OM) complexes devoted to nutrient acquisition. These complexes, facing the outside environment[20,21], are encoded in genetic regions named Polysaccharide Utilization Loci (PUL)[19] that represent a hallmark of this phylum.

To date, most studies focused at identifying and characterizing the function of these Bacteroidetes surface complexes[5,7,16-18,22,23] but little is known

on how they assemble[24] and in particular on how lipoproteins reach the bacterial surface. In Gram-negative Proteobacteria lipoprotein synthesis and transport has been well studied in model organisms such as *Escherichia coli*[25]. Lipoproteins are first synthesized as a precursor in the cytoplasm before their translocation to the periplasm via the Sec[26,27] or Tat machinery[28-30]. This recognition is mediated by the N-terminally located signal peptide II [31], which contains a conserved cysteine residue critical for the subsequent steps of maturation[32,33]. After crossing the inner membrane (IM), lipoprotein precursors remain anchored to the periplasmic side of the IM where they are then processed by three enzymes, rendering a final tri-acylated lipoprotein[34-37]. Lipoproteins destined to be inserted into the OM are transported through the aqueous environment of the periplasm via the dedicated Lol (localization of lipoproteins) transport machinery, composed of five proteins, LolA, -B, -C, -D and -E[25,38]. In Proteobacteria most OM lipoproteins are inserted in the inner leaflet of the OM and thus face the periplasm. The surface localization of OM lipoproteins in Bacteroidetes thus implies the existence of a yet unknown dedicated recognition and transport mechanism.

The present study deals with the reference strain *C. canimorsus* 5[39], which encodes 13 PUL. Three of them were recently shown to play critical roles in the biology and pathogenesis of this bacterium[40-42]. We address the question of how lipoproteins are targeted to the bacterial surface. We identify a signal sequence (LES) present at the N-terminus of surface exposed lipoproteins and we show that this signal is sufficient to target an intracellular lipoprotein to the cell surface. We extend our findings to other Bacteroidetes species, namely *Flavobacterium johnsoniae* and *Bacteroides fragilis*, identifying their specific LES thus showing that they share a new bacterial lipoprotein export pathway that flips lipoproteins across the outer membrane.

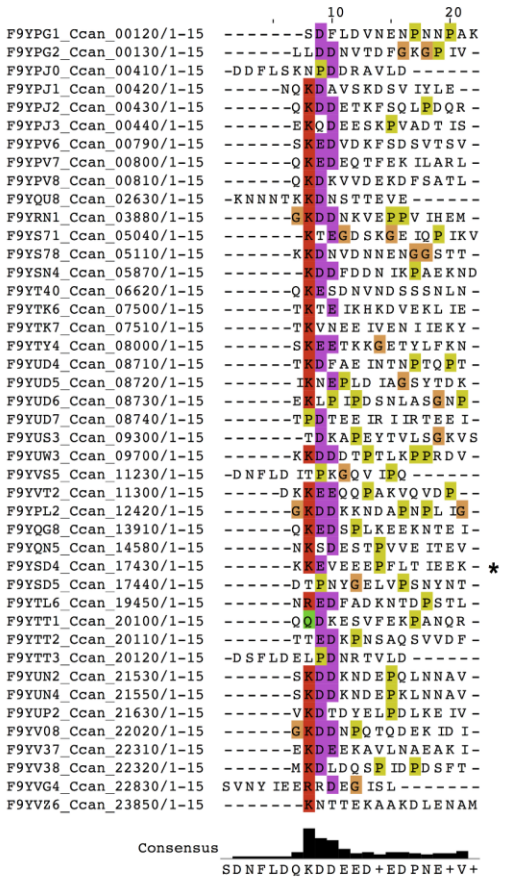
Results

In silico identification of a putative lipoprotein export signal

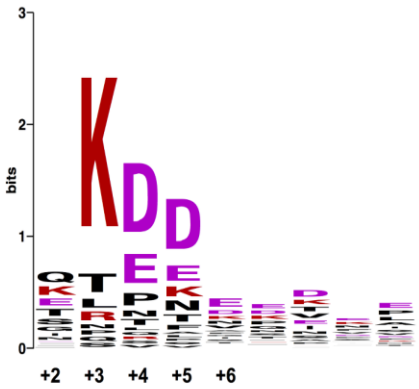
In order to see if a specific amino acid motif would be responsible for the targeting of lipoproteins to the bacterial surface, we examined in detail the

sequences of the lipoproteins detected at the surface of *C. canimorsus* 5[17]. When aligning the mature lipoproteins, a lysine (K), followed by either an aspartate (D) or a glutamate (E) residue, appeared to be conserved in close proximity to the N-terminal cysteine at position +1 (Fig. S1). This was refined by a second alignment considering only the 15 N-terminal residues of the mature lipoproteins and excluding the invariant first cysteine (Fig. 1A). The resulting consensus motif corresponded to Q-K-D-D-E, located between positions +2 and +6 (Fig. 1B) with a conservation of 16, 72, 48, 44 and 23 % respectively (Fig. 1C). In order to see whether this motif is specific to the surface-exposed lipoproteins, the same analysis was performed on OM lipoproteins facing the periplasm[17]. No highly conserved residues were identified in this set of proteins (Fig. S2), suggesting that the QKDDE consensus motif could be indeed a *bona fide* lipoprotein export signal (LES).

A



B



C

Position	+2			+3			+4		
Amino acid	Q	K	E	K	T	L	D	E	P
Percentage	16	13	11	72	9	4	48	20	9

Position	+5			+6		
Amino acid	D	E	K	E	D	K
Percentage	44	13	9	23	11	9

Figure 1. Alignment of *C. canimorsus* surface exposed lipoproteins reveals the presence of an N-terminal conserved motif
(A) MAFFT alignment of the first 15 N-terminal amino acids of mature surface exposed lipoproteins. The first invariant cysteine residue of each sequence was removed before performing the alignment. Highly conserved residues are highlighted according to Clustal color code (R, K in red; D, E in magenta; P in yellow; G in orange; Q, N, S, T in green, C in pink; A, I, L, M, F, W, V in blue; H, Y in cyan)[43]. The derived consensus sequence is shown below. (B) Generated WebLogo of the consensus sequence determined in (A). Positions relative to the +1 cysteine are indicated below. Charged residues are indicated in color. The color code is the same as in (A). (C) Amino acid frequency for each position of the consensus sequence, expressed as percentage. The three most represented amino acids for each position are shown. MucG (Ccan_17430) is indicated by a star.

The LES leads to surface localization of the periplasmic lipoprotein sialidase

To verify this hypothesis, we introduced the QKDDE motif in the sequence of the *C. canimorsus* sialidase (SiaC), an OM lipoprotein that faces the periplasm[42,44]. SiaC harboring the LES ($SiaC_{+2QKDDE+6}$) (Fig. 2A and B) was detected at the bacterial surface by immunolabeling followed by flow cytometry and microscopy (Fig. 2D and E and Fig. S3). In contrast, wild type SiaC and the soluble $SiaC_{C17G}$ variant were undetectable. Furthermore, proteinase K accessibility assays on intact cells showed that more than 90% of $SiaC_{+2QKDDE+6}$ was surface exposed (Fig. S4). This indicated that the addition of the consensus to an OM periplasmic lipoprotein is sufficient to drive its efficient transport to the bacterial surface and hence that this consensus represents a LES.

Determination of the minimal consensus allowing surface localization of sialidase

We next determined the minimal sequence required to constitute a functional LES. We first substituted the least conserved amino acids of the LES, namely the +2 Q and +6 E, by alanine residues generating constructs $SiaC_{+2AKDDE+6}$ and $SiaC_{+2AKDDA+6}$ (Fig. 2A). After monitoring protein expression (Fig. 2B), immunolabeling showed that both constructs localized to the bacterial surface (Fig. 2D and E), although to a lower extent than $SiaC_{+2QKDDE+6}$, thus indicating that the KDD motif is sufficient to target lipoproteins to the cell surface. We then tested if glutamate was able to functionally replace aspartate ($SiaC_{+2AKEEA+6}$) (Fig. 2A) since both residues were enriched in the alignment (Fig.

1C). Substitution of the two aspartate with two glutamate residues did not prevent surface localization but led to a clear reduction of fluorescence (Fig. 2D and E) in line with the lower conservation of glutamate at position +4 and +5 (Fig. 1C), explaining that in *C. canimorsus* surface lipoproteins aspartate is preferred over glutamate.

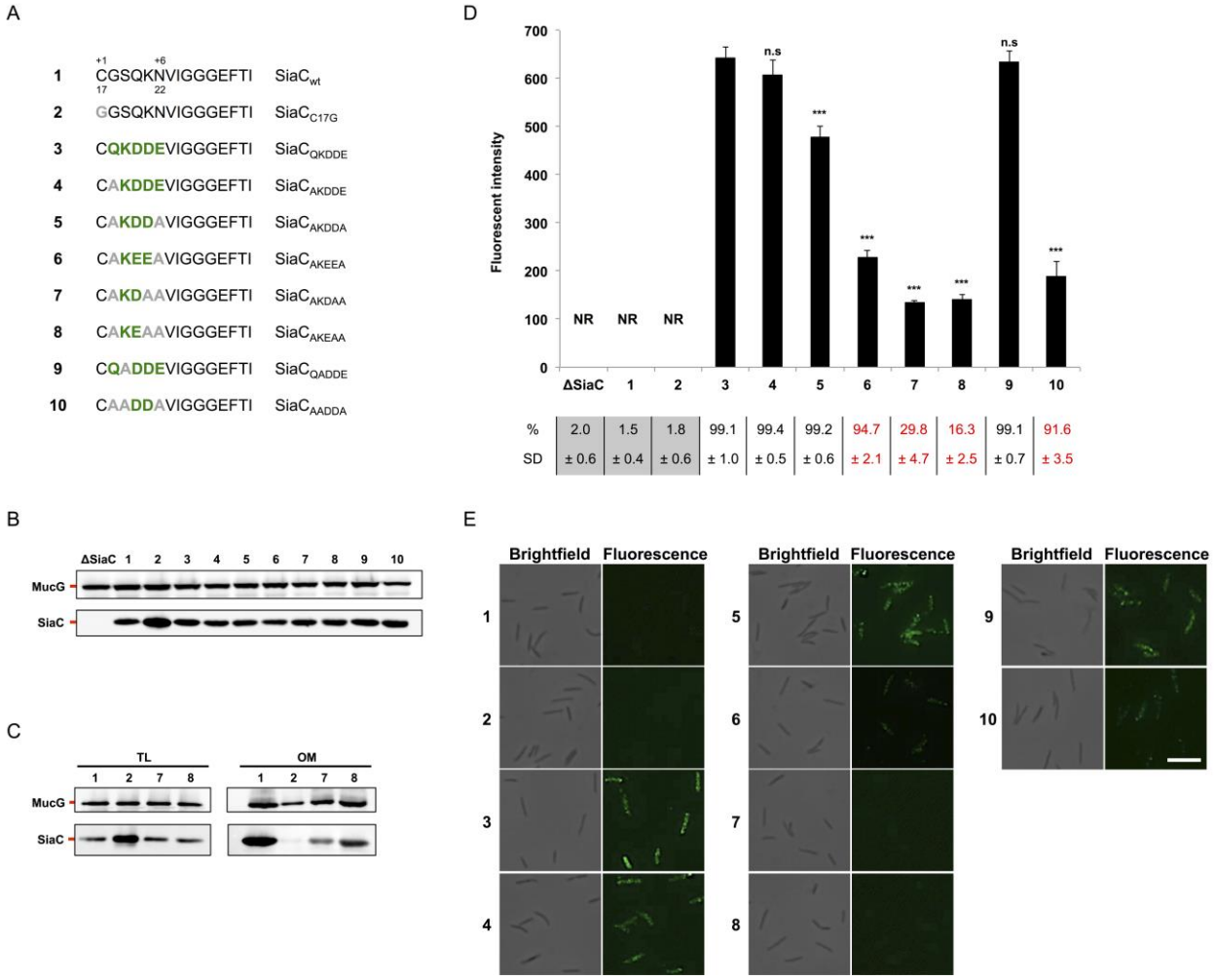


Figure 2. The LES allows SiaC surface exposure
(A) SiaC wt and consensus sequence mutant constructs. Amino acids derived from the consensus are indicated in bold green, point mutations are indicated in bold grey. (B) Detection of SiaC by western blot analysis of total cell extracts of strains expressing the SiaC constructs described in (A). Expression of MucG was monitored as loading control. (C) Detection of SiaC by western blot analysis of total lysates (TL) and outer membrane (OM) fractions of bacteria expressing different SiaC constructs. Expression of MucG was monitored as loading control. (D) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 3; n.s: not significant. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq$

0.001 as compared to reference construct 3). (E) Immunofluorescence microscopy images of bacteria labeled with anti-SiaC serum. Scale bar: 5 µm.

We then generated two SiaC constructs harboring only either KD or KE ($SiaC_{+2AKDAA+6}$ and $SiaC_{+2AKEAA+6}$) (Fig. 2A) but these two couples of residues alone turned out to be very weak LES since only 29.8 ± 4.7 ($SiaC_{+2AKDAA+6}$) and 16.3 ± 2.5 % ($SiaC_{+2AKEAA+6}$) of the cells displayed the proteins at their surface (Fig. 2D). In addition, the fluorescence intensity was weak: 28.2 and 29.4 % respectively of the intensity observed for the $SiaC_{+2AKDDA+6}$ reference (Fig. 2D). In order to verify that these constructs were not impaired in their transport to the OM, we monitored their presence in isolated outer membrane fractions. Both mutant proteins were found to be anchored to the OM although at lower levels than the wt protein, in particular for the construct $SiaC_{+2AKDAA+6}$, suggesting that these mutations could also impact to a minor extend OM localization of SiaC (Fig. 2C). Overall these data supported our hypothesis that K-$(D/E)_2$ represents the minimal LES. These findings also suggested that a functional LES might require an overall negative charge, supported by the fact that KDD is allowing efficient transport of SiaC to the surface while KD only is not (Fig. 2D).

We finally investigated the importance of the highly conserved lysine residue at position +3 of the LES (Fig. 2A). Unexpectedly, substitution of K alone ($SiaC_{+2QADDE+6}$) had no impact on the display of SiaC at the bacterial surface (Fig. 2D and E). However, removal of both K and Q ($SiaC_{+2AADDA+6}$) led to more than 60 % decrease of fluorescence intensity as compared to $SiaC_{+2AKDDA+6}$. Since the glutamine residue itself was not found to be critical ($SiaC_{+2AKDDA+6,}$ Fig. 2D), we conclude that either the +2 Q or the +3 K is required to form a functional LES. Taken together, these data indicate that the minimal export motif allowing surface localization of SiaC is composed of only two negatively charged amino acids preceeded by a positively charged or polar residue. Based on the consensus, we thus defined the minimal *C. canimorsus* LES as being K-$(D/E)_2$ or Q-A-$(D/E)_2$.

Positional effect of the minimal LES on sialidase surface localization

We next addressed the question of the importance of the position of the LES. The initial alignment showed that K is mainly conserved at position +3 (72 %), to a lower extent at position +2 (13 %) and is completely absent from position +4 (Fig. 1C). In contrast, D and E were conserved at positions +4, +5 and +6 (48, 44 and 11 % for D and 20, 13 and 23 % for E respectively) and completely absent from position + 3 (Fig. 1C). This suggested that not only the composition of the export signal could be crucial but also its position relative to the +1 cysteine. We therefore generated constructs in which the KDD motif was separated from the +1 cysteine by zero, two, three or four alanine residues (Fig. 3A) and compared their surface localization to the construct in which the KDD motif is separated from the +1 cysteine by only one alanine residue ($SiaC_{+2AKDDA+6}$). Although the four proteins were expressed (Fig. 3B), none of them were exported as efficiently as the one where only one alanine separated the KDD motif from the +1 cysteine ($SiaC_{+2AKDDA+6}$) (Fig. 3C and D). All proteins were anchored to the OM, thus again indicating that only the last step of transport to the surface was affected by these mutations (Fig. 3E). The position of the K-$(D/E)_2$ signal relative to the +1 cysteine is thus critical for the *C. canimorsus* LES and the optimal situation is C-X-K-$(D/E)_2$-X.

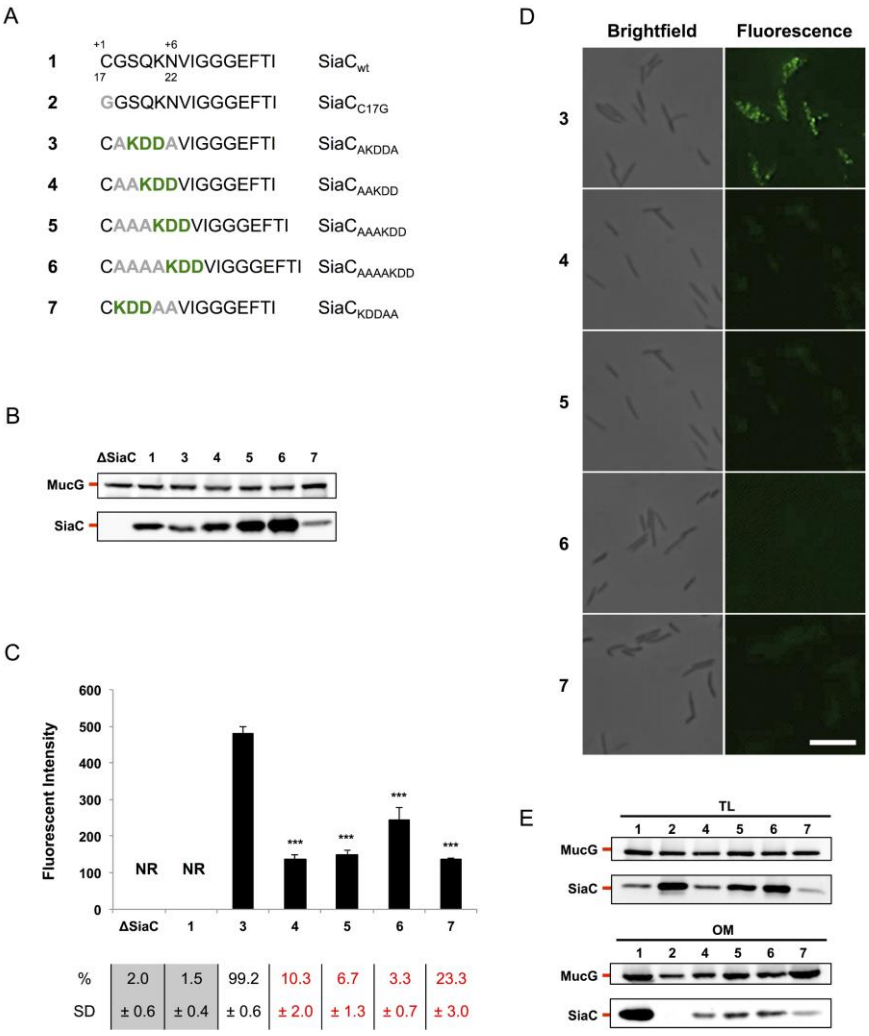


Figure 3. The position of the minimal LES is crucial for its function
(A) SiaC wt and consensus sequence mutant constructs. Amino acids derived from the consensus are indicated in bold green, point mutations are indicated in bold grey. (B) Detection of SiaC by western blot analysis of total cell extracts of strains expressing the SiaC constructs described in (A). MucG expression was monitored as loading control. (C) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 3. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 3). (D) Immunofluorescence microscopy images of bacteria stained with anti-SiaC serum. Scale bar: 5 µm. (E) Detection of SiaC by western blot analysis of total lysates (TL) and outer membrane (OM) fractions of bacteria expressing different SiaC constructs. MucG expression was monitored as loading control.

Characterization of the LES of the surface exposed lipoprotein MucG

Looking at the LES of different *C. canimorsus* surface lipoproteins (Fig 1), it appeared that some were quite divergent from the consensus. Among these is

the LES of mucinase MucG (Ccan_17430)[41], KKEVEEE (Fig 1A and Fig S5A). We first confirmed that MucG is indeed a surface exposed lipoprotein (Fig. S5) and then we tested whether this poorly conserved LES would drive the export of sialidase to the surface of *C. canimorsus*. We introduced the MucG LES, KKEVEEE or part of this sequence, in SiaC, giving $SiaC_{+2KKEVE+6}$, $SiaC_{+2KKEVEE+7}$ and $SiaC_{+2KKEVEEE+8}$ (Fig. S6A and B), and monitored their surface localization (Fig. S6C and D). $SiaC_{+2KKEVE+6}$, was only poorly transported to the cell surface while $SiaC_{+2KKEVEE+7}$ and $SiaC_{+2KKEVEEE+8}$ showed clear surface localization. Although the overall protein amount of $SiaC_{+2KKEVE+6}$ was reduced, the protein appeared to be anchored to the OM (Fig. S6E). The only difference between these constructs being the number of negatively charged amino acids in the LES, this strongly supported our initial findings that the LES requires an overall negative charge to drive transport of lipoproteins to the bacterial surface (Fig 2).

We next wanted to study the MucG LES in its native background. To this aim we systematically substituted residues 22 to 28 of the MucG LES by alanines (Fig. 4A). After verifying that all mutant proteins were expressed (Fig. 4B), we monitored the surface exposure of the MucG variants by flow cytometry (Fig. 4C). Alanine substitution of K22, V25 and E27 did not significantly alter surface exposition of MucG, while mutation of K23, E24, E26 or E28 resulted in a 25 to 50% decrease of surface exposure. None of these single mutations completely abolished surface localization, suggesting that the MucG motif is redundant, presumably due to the presence of two lysines and four glutamates. The mutation of one of those residues could therefore be compensated by the presence of another one in close proximity and indeed all protein variants we generated harbor an overall negatively charged functional LES.

Because of this, we generated two additional constructs by mutating simultaneously either all negatively or all positively charged residues in the MucG LES (Fig. 4A). After having confirmed their correct expression (Fig. 4B), we analyzed their surface localization by flow cytometry (Fig. 4C). As expected, substitution of the two lysine residues ($MucG_{AAEVEEE}$) led to MucG surface exposure in only 23.1 ± 4.5 % of the cells (Fig. 4C). Furthermore, the fluorescence intensity in this subset of cells was markedly decreased as compared to the wt strain (23.8 %), indicating that the efficiency of the transport

was also strongly affected in this subpopulation. This is in good agreement with our previous findings showing the importance of the K/Q residues for surface export (Fig. 2)

Similarly, $MucG_{KKAAAAA}$ was surface localized in only 41.9 ± 6.9 % of the cells (Fig. 4C) and the fluorescence intensity in this subpopulation was decreased as compared to the wt strain (24.5 %). This is in agreement with our findings in SiaC that an overall negatively charged LES is critical for efficient surface localization.

By combining the data obtained from single and multiple alanine substitutions, the minimal LES for optimal MucG surface exposure appears to be X-K-$(D/E)_3$ downstream from the +1 cysteine, hence resembling the one deduced from previous experiments (X-K-$(D/E)_2$-X) (Fig. 2, Fig. 3 and Fig. S6).

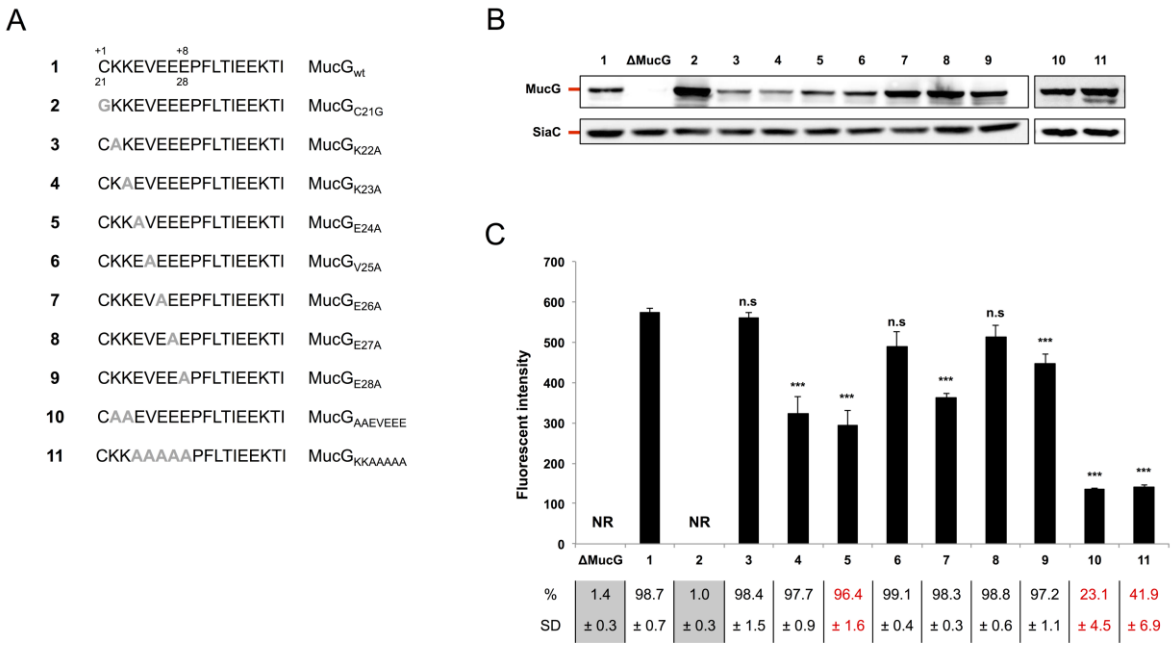


Figure 4. MucG LES mutational analysis
(A) MucG wt and mutant constructs. Point mutations are indicated in bold grey. (B) Detection of MucG by western blot analysis of total cell extracts of strains expressing the MucG constructs described in (A). Expression of SiaC was monitored as loading control. (C) Quantification of MucG surface exposure by flow cytometry of live cells labeled with anti-MucG serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 1; n.s: not significant. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 1).

The LES is conserved in the Bacteroidetes phylum

To see if the LES identified in *C. canimorsus* would be conserved in other Bacteroidetes, we took advantage of the recently published *B. fragilis* NCTC 9343 surfome study[45] and performed an *in silico* analysis on the N-terminus of the identified surface lipoproteins (Fig. S7A). We found an enrichment in negatively charged amino acids in close proximity to the +1 cysteine (SDDDD) (Fig. S7A). However, unlike in the *C. canimorsus* LES, the aspartate residues were majorly located at position +3 and +4 instead of +4 and +5. Additionally, this region was not enriched in positively charged amino acids at position +3 but harbored a polar serine residue at position +2. This sequence is different from the LES identified in *C. canimorsus* but it has nevertheless a clear similarity with the *C. canimorsus* LES. Indeed, it starts with a polar residue followed by several negatively charged residues and in *C. canimorsus*, the lysine residue could be substituted by an alanine provided that a glutamine was present at position +2 (Fig. 2D and E). We thus hypothesize that SDDDD represents the consensus LES of *B. fragilis*. We then searched for the LES of *Flavobacterium johnsoniae* UW101 that belongs to the same family as *C. canimorsus*, the Flavobacteriaceae. Since no surfome analysis has been performed on this bacterium, we recovered the sequences of all predicted SusD homologs[19], supposedly surface exposed lipoproteins. We next aligned their N-termini and derived the consensus sequence SDDFE (Fig. S7B). Interestingly, this motif seems closer to the LES of *B. fragilis* than to the *C. canimorsus* one in the sense that is enriched in a polar residue (S) rather than in a positively charged one. However, negatively charged amino acids are still predominant in this LES.

The LES from *B. fragilis* and *F. johnsoniae* is functional in *C. canimorsus*

To validate our findings, we tested if the consensus sequences predicted for *B. fragilis* (SDDDD) and *F. johnsoniae* (SDDFE) would represent a functional LES in *C. canimorsus.* Both sequences were inserted in SiaC (Fig. 5A) and the recombinant proteins were tested in *C. canimorsus* (Fig. 5B). Both constructs were found to be surface localized (Fig. 5C and D), although at lower levels than SiaC harboring the *C. canimorsus* LES, indicating that the LES from *Bacteroides* and *Flavobacteria* allow surface transport of lipoproteins in *Capnocytophaga*.

Overall these data confirm the evidence of a shared novel pathway for lipoprotein export in this phylum of Gram-negative bacteria.

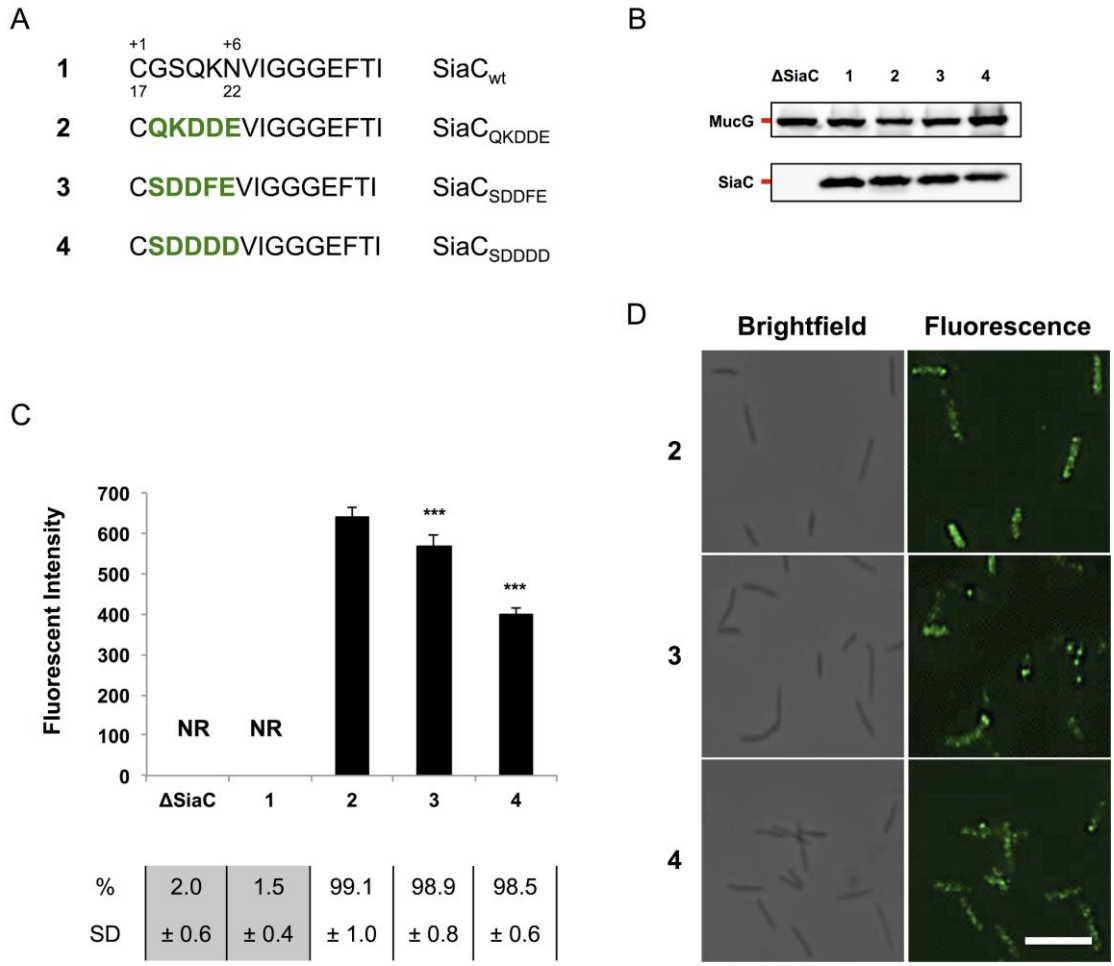


Fig. 5. *B. fragilis* and *F. johnsoniae* LES allow SiaC surface localization
(A) SiaC wt and consensus sequence mutant constructs. Amino acids derived from the *B. fragilis* or *F. johnsoniae* consensus are indicated in bold green. (B) Detection of SiaC by western blot analysis of total cell extracts of strains expressing the SiaC constructs described in (A). MucG expression was monitored as loading control. (C) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 2. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey. (D) Immunofluorescence microscopy images of bacteria labeled with anti-SiaC serum. Scale bar: 5 μm.

Discussion

In conclusion, we show for the first time that surface exposed lipoproteins of Bacteroidetes harbor a specific signal at their N-terminus that drives their transport to the bacterial surface. In addition, we derived the canonical LES sequence that represents the most common choice of amino acid at each position for *C. canimorsus*, *B. fragilis* and *F. johnsoniae*. For *C. canimorsus* it is C-X-K-(D/E)$_2$-X, where X can be any amino acid as long as the overall negative charge of the LES is maintained. Interestingly, this is different from what has been described in the Spirochaetes *Borrelia burgdorferi*. This bacterium, also harboring a high proportion of surface lipoproteins, seems to transport them at its surface by default without the requirement of a specific signal[46-48]. This suggests that Bacteroidetes and Spirochaetes evolved different lipoprotein transport machineries and corresponding signaling pathways. The LES of Bacteroidetes is in direct proximity to the +1 cysteine, a region that acts as lol-avoidance signal in Proteobacteria[49-51] thus indicating that also the sorting rules distinguishing inner and outer membrane lipoproteins are different in Bacteroidetes.

The discovery of the LES implies the existence of a novel export pathway in bacteria and represents the starting point for the identification of the machinery that allows surface lipoproteins localization. In this regard, it is interesting to note that Bacteroidetes do not encode any homolog of LolB, the OM lipoprotein responsible of the insertion of lipoproteins into the inner leaflet of the OM in *E. coli* and most studied bacteria[52,53]. The function of LolB in Bacteroidetes might therefore be fulfilled by another protein or protein complex that would also be able to flip surface exposed lipoproteins across the OM.

Recently a novel lipoprotein export system has been discovered in the human pathogen *Neisseria meningitidis*[54]. This bacterium displays several lipoproteins at the bacterial surface, among which the TbpA and HupA proteins that are involved in iron uptake from transferrin and haemoglobin respectively [55,56]. Hooda *et al.* have shown that TbpA is transported to the bacterial surface by an integral outer membrane protein named Slam1 (lipoprotein assembly modulator 1) while HupA is transported by the paralog Slam2. While homologs of Slam could be found in several Proteobacteria[54], we could not identify any

homolog in Bacteroidetes thus suggesting that in this phylum lipoproteins are transported to the bacterial surface via a different mechanism. Furthermore, in *Neisseria* no conserved signal sequence has so far been identified in surface exposed lipoproteins and the evidence that TbpA and HupA require each a specific Slam transporter suggests that in this bacterium the recognition between the lipoprotein and the transporter is different from Bacteoidetes where a common specific sequence would address the lipoproteins to the bacterial surface. We believe the discovery of the LES represents a step forward in understanding the complex biology of Bacteroidetes, being at the same time commensal and opportunistic pathogens, and is the starting point for the identification of the machinery that allows surface lipoproteins localization.

Materials and Methods

Bacterial strains and growth conditions

Bacterial strains used in this study are listed in Table S1. *Escherichia coli* strains were routinely grown in lysogeny broth (LB) at 37°C. *C. canimorsus* strains were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) plates (SB plates) for 2 days at 37°C in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 100 µg/ml ampicillin (Amp), 50 µg/ml kanamycin (Km) for *E. coli* and 10 µg/ml erythromycin (Em), 10 µg/ml cefoxitin (Cfx), 20 µg/ml gentamicin (Gm) for *C. canimorsus*.

Construction of *siaC* and *mucG* expression plasmids

Plasmids and primers used in this study are listed in Table S2 and S3 respectively*. siaC* (*Ccan_04790*) was amplified from 100 ng *C. canimorsus* 5 genomic DNA with primers 4159 and 7696 using the Q5 High-Fidelity DNA Polymerase (M0491S; New England Biolabs). The initial denaturation was at 98°C for 2 min, followed by 30 cycles of amplification (98°C for 30 s, 52°C for 30 s, and 72°C for 2 min) and finally 10 min at 72°C. After purification, the fragment was digested using NcoI and XhoI restriction enzymes and cloned into plasmid pMM47.A, leading to plasmid pFL117. *mucG* (*Ccan_17430*) was cloned in the same way except that primers 7182 and 7625 were used for amplification and that the fragment was cloned into plasmid pPM5, leading to plasmid pFL43.

Site-specific point mutations were introduced by amplifying separately the N- and C-terminal part of each gene using forward and reverse primers harboring the desired mutations in their sequence in combination with primers 4159 and 7696 for *siaC* and 7182 and 7625 for *mucG*. Both PCR fragments were purified and then mixed in equal amounts for PCR using the PrimeStar HS DNA Polymerase (R010A; Takara). The initial denaturation step was performed at 98°C for 2 min, followed by 30 cycles of amplification (98°C for 10 s, 60°C for 5 s, and 72°C for 3 min 30 s) and finally 10 min at 72°C. Final PCR products were then cleaned, digested using NcoI and XhoI restriction enzymes and cloned into plasmids pMM47.A or pPM5 for *siaC* and *mucG* respectively. The incorporation of the desired point mutations in all inserts was confirmed by sequencing. Plasmids

expressing *siaC* and *mucG* variants were transferred to *C. canimorsus* 5 *siaC* and *mucG* deletion strains respectively by electroporation[39].

Immunofluorescence labeling for flow cytometry and microscopy analysis

Bacteria grown for 2 days on SB plates were collected, washed once with PBS, and resuspended in one ml PBS to an OD_{600} of 0.1. 5 µl of bacterial suspension (approximately 3 x 10^5 bacteria) were used to inoculate 2.5 ml of DMEM (41965-039; Gibco) containing 10% heat-inactivated human serum (HIHS) in 12-well plates (665 180; Greiner Bio-one). Bacteria were harvested after 23h of growth at 37°C in the presence of 5% CO_2, washed twice with PBS, and resuspended in 1 ml PBS. The optical density at 600 nm of bacterial suspensions was measured and approximately 3 x 10^7 bacteria were collected for each strain. Bacteria were resuspended in 200 µl PBS containing 1% BSA (w/v) and incubated for 30 min at room temperature. Bacteria were then centrifuged, resuspended in 200 µl of a primary antibody dilution (rabbit anti-SiaC or rabbit anti-MucG antiserum) and incubated for 30 min at room temperature. Following centrifugation, bacterial cells were washed 3 times before being resuspended in 200 µl of a secondary antibody dilution (donkey anti-rabbit coupled to Alexa Fluor 488; A-21206; Invitrogen) and incubated for 30 min at room temperature in the dark. Following centrifugation, bacteria were washed 3 times, resuspended in 200 µl of 4% PFA (w/v) and incubated for 15 min at room temperature in the dark. Finally, bacteria were centrifuged, washed once and resuspended in 700 µl of PBS. For flow cytometry analysis, samples were directly analyzed with a BD FACSVerse™ (BD Biosciences) and data were processed with BD FACSuite™ (BD Biosciences). Analysis was performed on all events without previous gating. For microscopy analysis, labeled bacteria were added on top of poly-L-lysine-coated coverslips and were allowed to adhere for 30 min at room temperature. After removal of bacterial suspension, coverslips were washed 3 times, mounted upside down on glass slides and allowed to dry overnight at room temperature in the dark. All microscopy images were captured with an Axioscop (Zeiss) microscope with an Orca-Flash 4.0 camera (Hamamatsu) and Zen 2012 software (Zeiss). As control, samples were prepared in parallel as described above except that rabbit pre-immunization serum was used for labeling.

Proteinase K accessibility assay

Bacteria were grown overnight as described above for immunofluorescence labeling. Bacteria were harvested, washed twice with PBS and resuspended in 1 ml PBS. The optical density at 600 nm of bacterial suspensions was measured and approximately 6 x 10^7 bacteria were collected for each strain. Bacteria were then resuspended in 500 µl PBS containing 200 µg/ml Proteinase K (P2308, Sigma) and incubated for 30 min at 37°C. Bacteria were then centrifuged, whased once with PBS containing 5mM PMSF (P7626, Sigma), washed once with PBS and finally resuspended in 20 µl SDS PAGE buffer. Untreated control samples were realized in parallel. Eight microliter samples were loaded on 12% SDS PAGE gels. After gel electrophoresis, proteins were transferred onto nitrocellulose membrane and analyzed by Western blot.

***In vivo* radiolabeling with [^{3}H] palmitate, immunoprecipitation and fluorography**

Bacteria were grown overnight as described above for immunofluorescence labeling, except that bacteria were grown in 5 ml medium in 6-well plates (657 160; Greiner Bio-one). After 18 h of incubation, [9,10-^{3}H] palmitic acid (32 Ci/mmol; NET043; Perkin-Elmer Life Sciences) was added to a final concentration of 50 µCi/ml and incubation was continued for 6 h. Bacteria were then collected by centrifugation, washed 2 times with 1 ml PBS and pellets were stored at -20°C until further use. Pellets were resupended in 300 µl PBS containing 1% Triton™ X-100 (28817.295; VWR) and vortexed 10 sec to lyse bacteria. Lysates were centrifuged 2 min at 14,000 g and the supernatant was transferred into a new tube. MucG proteins were immuno-precipitated by addition of 15 µl MucG antiserum for 90 min at room temperature with constant agitation. In parallel, 20 µl of Protein A agarose slurry (P3476; Sigma-Aldrich) were washed 2 times with 500 µl wash buffer (0.1% Triton™ X-100 in PBS), saturated with 500 µl 0.2% BSA (w/v) for 30 min and washed again 2 times with wash buffer. The Protein A agarose slurry was then added to the cell lysate and incubation was continued for 30 min at room temperature with constant agitation. Samples were then centrifuged at 14,000 g for 2 min and the supernatant was discarded. Pellets were washed 5 times with 500 µl wash

buffer. Bound proteins were eluted by addition of 50 µl SDS PAGE buffer and heating for 10 min at 95°C. Samples were centrifuged again and supernatants were carefully separated from the agarose beads and loaded on 10% SDS PAGE gels. After gel electrophoresis, gels were fixed in a 25/65/10 isopropanol/water/acetic acid solution overnight and subsequently soaked for 30 min in Amplify (NAMP100; Amersham) solution. Gels were vacuum dried and exposed to SuperRX autoradiography film (Fuji) for 13-21 days until desired signal strength was reached.

Human salivary mucin degradation

Fresh human saliva was collected from healthy volunteers and filter-sterilized using 0.22 µm filters (Millipore). Bacteria grown for 2 days on SB plates were collected, washed once with PBS, and set to an OD_{600} of 1. One hundred µl of bacterial suspension (approximately 5 x 10^7 bacteria) were then mixed with 100 µl of human saliva and incubated for 240 min at 37°C. As negative control, 100 µl of saliva was incubated with 100 µl PBS. Samples were then centrifuged for 5 min at 13,000 g, the supernatant carefully collected and loaded on 10% SDS PAGE gels. Mucin degradation was monitored by lectin staining with PNA agglutinin (DIG glycan differentiation kit, 11210238001; Roche) according to manufacturer's instructions. Mucin degradation was estimated by loss or reduction of PNA staining as compared to the negative control.

Outer membrane protein purification

Outer membrane proteins were isolated as described in references [45] and [57] with several modifications. All steps were carried out on ice unless otherwise stated. All sucrose concentrations are expressed as percentages of w/v in 10 mM HEPES (pH 7.4). Bacteria collected from 2 plates were washed 2 times with 30 ml 10 mM HEPES (pH 7.4) before being resuspended in 4.5 ml of 10% sucrose. Bacterial cells were then disrupted by 2 passages through a French press at 35,000 psi. The lysate was collected and centrifuged for 10 min at 16,500 g to remove insoluble material. The crude cell extract was then layered on top of a sucrose step gradient composed of 1.33 ml of 70% sucrose and 6 ml of 37% sucrose and centrifuged at 100,000 g (28,000 rpm) for 70 min at 4°C in a SW41

Ti rotor. The yellow material above the 37% sucrose solution and at the 10%/37% interface, corresponding to soluble and enriched inner membrane proteins, was collected and diluted to 7 ml with 10 mM HEPES (pH 7.4). The high density band at the 37%/70% interface, corresponding to enriched outer membrane proteins, was collected and diluted to 7 ml with 10 mM HEPES (pH 7.4). Membranes from both fractions were then centrifuged at 320,000 g (68,000 rpm) for 90 min at 4°C in a 70.1 Ti rotor. The supernatant of the yellow material fraction, corresponding to soluble proteins, was transferred to a fresh tube and stored at -20°C. The pellet of the same tube, corresponding to a mixture of inner and outer membrane fractions, was resuspended in 1 ml of 40% sucrose and stored at -20°C. The supernatant of the outer membrane proteins band was discarded, the pellet resuspended in 7 ml of 10 mM HEPES (pH 7.4) containing 1% Sarkozyl (L5777; Sigma-Aldrich) and incubated at room temperature for 30 min with constant agitation. The outer membrane fraction was then centrifuged at 320,000 g for 60 min at 4°C in a 70.1 Ti rotor, resuspended in 7 ml of 100 mM Na_2CO_3 (pH 11) and incubated at 4°C for 20 min with constant agitation. The outer membrane fraction was then centrifuged, washed with 7 ml unbuffered 40 mM Tris and centrifuged again. Finally, the purified outer membrane was resuspended in 200 to 400 µl unbuffered 40 mM Tris and stored at -20°C. Protein concentration of all fractions was assessed using the Bio-Rad Protein Assay (500-0006; Bio-Rad) according to the manufacturer's instructions. One to 2 µg of total protein of total cell lysates and outer membrane fractions were loaded on 12% SDS PAGE gels. After gel electrophoresis, proteins were transferred onto nitrocellulose membrane and analyzed by Western blot. Purity of the outer membrane fraction was assessed by measuring the SDH activity as described previously [58]. Total lysate and enriched inner membrane fractions served as controls (data not shown).

Lipoprotein multiple sequence alignment

The sequences of 40 lipoproteins previously identified as being part of the surface proteome of *C. canimorsus* 5[17] were retrieved from the Uniprot database[59] (Release 2015_12). Additionally, 2 *C. canimorsus* 5 proteins (F9YSD4 and F9YTT3) detected at the bacterial surface but predicted to harbour an SPI

signal were reanalysed with the PATRIC database[60] and found to possess an SPII signal and thus considered lipoproteins, rendering a final list of 43 surface exposed predicted lipoproteins (Table S4). The SPII cleavage site of each protein was then predicted using the LipoP software[61] (1.0 Server, default settings), showing that all proteins possess one clear SPII cleavage site. Accordingly, protein sequences were trimmed to their predicted mature form. Lists corresponding to either full-length protein sequences or 15 amino acids downstream of the +1 cysteine were generated. Datasets were then submitted to multiple sequence alignment using the MAFFT online tool[62] (version 7.268, default settings) and the output was analysed using the Jalview software[63] (version 2.9.0b2). The final consensus sequence logo was drawn using WebLogo[64] (version 2.8.2, default settings). The sequences of the 17 *C. canimorsus* outer membrane lipoproteins presumably facing the periplasm[17] were processed in the same way (Table S5). The sequences of the 22 previously identified proteinase K sensitive *Bacteroides fragilis* NCTC 9343 surface exposed lipoproteins[45] were processed in the same way (Table S5). Forty-two *Flavobacterium johnsoniae* UW101 predicted SusD-like lipoproteins were identified in the PULDB of the CAZY database[19], the corresponding sequences extracted from the Uniprot database and processed as described above (Table S5).

Statistical analysis

All data is presented as mean ± standard deviation (SD). Statistical analyses were done by one-way ANOVA followed by Bonferroni test using the GraphPad Prism version 5.00 for Windows, GraphPad Software, La Jolla California USA, www.graphpad.com. A P value ≤ 0.05 was considered statistically significant.

Funding information

This work was financed by advanced grant 293605-CAPCAN from the European Research Council to Guy R. Cornelis; Francesco Renzi is a post-doctoral fellow “chargé de recherche” of the Belgian National Fund for Research (FNRS). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

The authors declare no competing financial interests. We thank P. Manfredi, K Hack, E. Hess and E. Lawarée for stimulating discussions; M. Dol, J. Coppine, M. Jadot and I. Hamer for technical assistance.

References

1 Lauber, C. L. *et al.* Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol* **75**, (2009).
2 Kirchman, D. L. The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**, (2002).
3 Bjursell, M. K. *et al.* Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J Biol Chem* **281**, (2006).
4 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, (2005).
5 Koropatkin, N. M. *et al.* How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* **10**, (2012).
6 Sonnenburg, J. L. *et al.* Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, (2005).
7 Xu, J. *et al.* A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**, (2003).
8 Mysak, J. *et al.* Porphyromonas gingivalis: major periodontopathic pathogen overview. *J Immunol Res* **2014**, (2014).
9 Socransky, S. S. *et al.* Capnocytophaga: new genus of gram-negative gliding bacteria. III. Physiological characterization. *Arch Microbiol* **122**, (1979).
10 Brook, I. The role of anaerobic bacteria in bacteremia. *Anaerobe* **16**, (2010).
11 Butler, T. Capnocytophaga canimorsus: an emerging cause of sepsis, meningitis, and post-splenectomy infection after dog bites. *Eur J Clin Microbiol Infect Dis* **34**, (2015).
12 Gaastra, W. & Lipman, L. J. Capnocytophaga canimorsus. *Vet Microbiol* **140**, (2010).
13 Sears, C. L. Enterotoxigenic Bacteroides fragilis: a rogue among symbiotes. *Clin Microbiol Rev* **22**, (2009).
14 Sears, C. L. *et al.* Bacteroides fragilis subverts mucosal biology: from symbiont to colon carcinogenesis. *J Clin Invest* **124**, (2014).
15 Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, (2007).
16 Bauer, M. *et al.* Whole genome analysis of the marine Bacteroidetes'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**, (2006).
17 Manfredi, P. *et al.* The genome and surface proteome of Capnocytophaga canimorsus reveal a key role of glycan foraging systems in host glycoproteins deglycosylation. *Mol Microbiol* **81**, (2011).
18 McBride, M. J. *et al.* Novel features of the polysaccharide-digesting gliding bacterium Flavobacterium johnsoniae as revealed by genome sequence analysis. *Appl Environ Microbiol* **75**, (2009).
19 Terrapon, N. *et al.* Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* **31**, (2015).
20 Reeves, A. R. *et al.* A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol* **178**, (1996).
21 Reeves, A. R. *et al.* Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol* **179**, (1997).
22 Koropatkin, N. M. *et al.* Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**, (2008).
23 Martens, E. C. *et al.* Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**, (2009).
24 Foley, M. H. *et al.* The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell Mol Life Sci*, (2016).
25 Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, (2011).
26 Sugai, M. & Wu, H. C. Export of the outer membrane lipoprotein is defective in secD, secE, and secF mutants of Escherichia coli. *J Bacteriol* **174**, (1992).
27 Watanabe, T. *et al.* Synthesis and export of the outer membrane lipoprotein in Escherichia coli mutants defective in generalized protein export. *J Bacteriol* **170**, (1988).
28 Hutchings, M. I. *et al.* Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol* **17**, (2009).
29 Thompson, B. J. *et al.* Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium Streptomyces coelicolor. *Mol Microbiol* **77**, (2010).
30 Widdick, D. A. *et al.* The twin-arginine translocation pathway is a major route of protein export in Streptomyces coelicolor. *Proc Natl Acad Sci U S A* **103**, (2006).
31 Inouye, S. *et al.* Amino acid sequence for the peptide extension on the prolipoprotein of the Escherichia coli outer membrane. *Proc Natl Acad Sci U S A* **74**, (1977).
32 Hayashi, S. & Wu, H. C. Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, (1990).
33 Braun, V., and H. C. Wu. in *Bacterial cell wall* Vol. 27 319 (Elsevier, 1993).

34 Dev, I. K. & Ray, P. H. Rapid assay and purification of a unique signal peptidase that processes the prolipoprotein from Escherichia coli B. *J Biol Chem* **259**, (1984).

35 Hantke, K. & Braun, V. Covalent binding of lipid to protein. Diglyceride and amide-linked fatty acid at the N-terminal end of the murein-lipoprotein of the Escherichia coli outer membrane. *Eur J Biochem* **34**, (1973).

36 Hussain, M. *et al.* Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the Escherichia coli outer membrane. *J Biol Chem* **257**, (1982).

37 Sankaran, K. & Wu, H. C. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem* **269**, (1994).

38 Bos, M. P. *et al.* Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* **61**, (2007).

39 Mally, M. & Cornelis, G. R. Genetic tools for studying Capnocytophaga canimorsus. *Appl Environ Microbiol* **74**, (2008).

40 Manfredi, P. *et al.* New iron acquisition system in Bacteroidetes. *Infect Immun* **83**, (2015).

41 Renzi, F. *et al.* Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. *MBio* **6**, (2015).

42 Renzi, F. *et al.* The N-glycan glycoprotein deglycosylation complex (Gpd) from Capnocytophaga canimorsus deglycosylates human IgG. *PLoS Pathog* **7**, (2011).

43 Thompson, J. D. *et al.* Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**, (2002).

44 Mally, M. *et al.* Capnocytophaga canimorsus: a human pathogen feeding at the surface of epithelial cells and phagocytes. *PLoS Pathog* **4**, (2008).

45 Wilson, M. M. *et al.* Analysis of the outer membrane proteome and secretome of Bacteroides fragilis reveals a multiplicity of secretion mechanisms. *PLoS One* **10**, (2015).

46 Chen, S. & Zuckert, W. R. Probing the Borrelia burgdorferi surface lipoprotein secretion pathway using a conditionally folding protein domain. *J Bacteriol* **193**, (2011).

47 Kumru, O. S. *et al.* Surface localization determinants of Borrelia OspC/Vsp family lipoproteins. *J Bacteriol* **193**, (2011).

48 Schulze, R. J. & Zuckert, W. R. Borrelia burgdorferi lipoproteins are secreted to the outer surface by default. *Mol Microbiol* **59**, (2006).

49 Seydel, A. *et al.* Testing the '+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. *Mol Microbiol* **34**, (1999).

50 Terada, M. *et al.* Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *J Biol Chem* **276**, (2001).

51 Yamaguchi, K. *et al.* A single amino acid determinant of the membrane localization of lipoproteins in E. coli. *Cell* **53**, (1988).

52 Matsuyama, S. *et al.* A novel outer membrane lipoprotein, LolB (HemM), involved in the LolA (p20)-dependent localization of lipoproteins to the outer membrane of Escherichia coli. *EMBO J* **16**, (1997).

53 Tanaka, K. *et al.* Deletion of lolB, encoding an outer membrane lipoprotein, is lethal for Escherichia coli and causes accumulation of lipoprotein localization intermediates in the periplasm. *J Bacteriol* **183**, (2001).

54 Hooda, Y. *et al.* Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in Neisseria. *Nature Microbiology* **1**, (2016).

55 Morgenthau, A. *et al.* Bacterial receptors for host transferrin and lactoferrin: molecular mechanisms and role in host-microbe interactions. *Future Microbiol* **8**, (2013).

56 Lewis, L. A. & Dyer, D. W. Identification of an iron-regulated outer membrane protein of Neisseria meningitidis involved in the utilization of hemoglobin complexed to haptoglobin. *J Bacteriol* **177**, (1995).

57 Kotarski, S. F. & Salyers, A. A. Isolation and characterization of outer membranes of Bacteroides thetaiotaomicron grown on different carbohydrates. *J Bacteriol* **158**, (1984).

58 Kasahara, M. & Anraku, Y. Succinate dehydrogenase of Escherichia coli membrane vesicles. Activation and properties of the enzyme. *J Biochem* **76**, (1974).

59 Consortium, U. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, (2015).

60 Wattam, A. R. *et al.* PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**, (2014).

61 Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**, (2003).

62 Katoh, K. *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, (2002).

63 Waterhouse, A. M. *et al.* Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, (2009).

64 Crooks, G. E. *et al.* WebLogo: a sequence logo generator. *Genome Res* **14**, (2004).

Supplemental materials

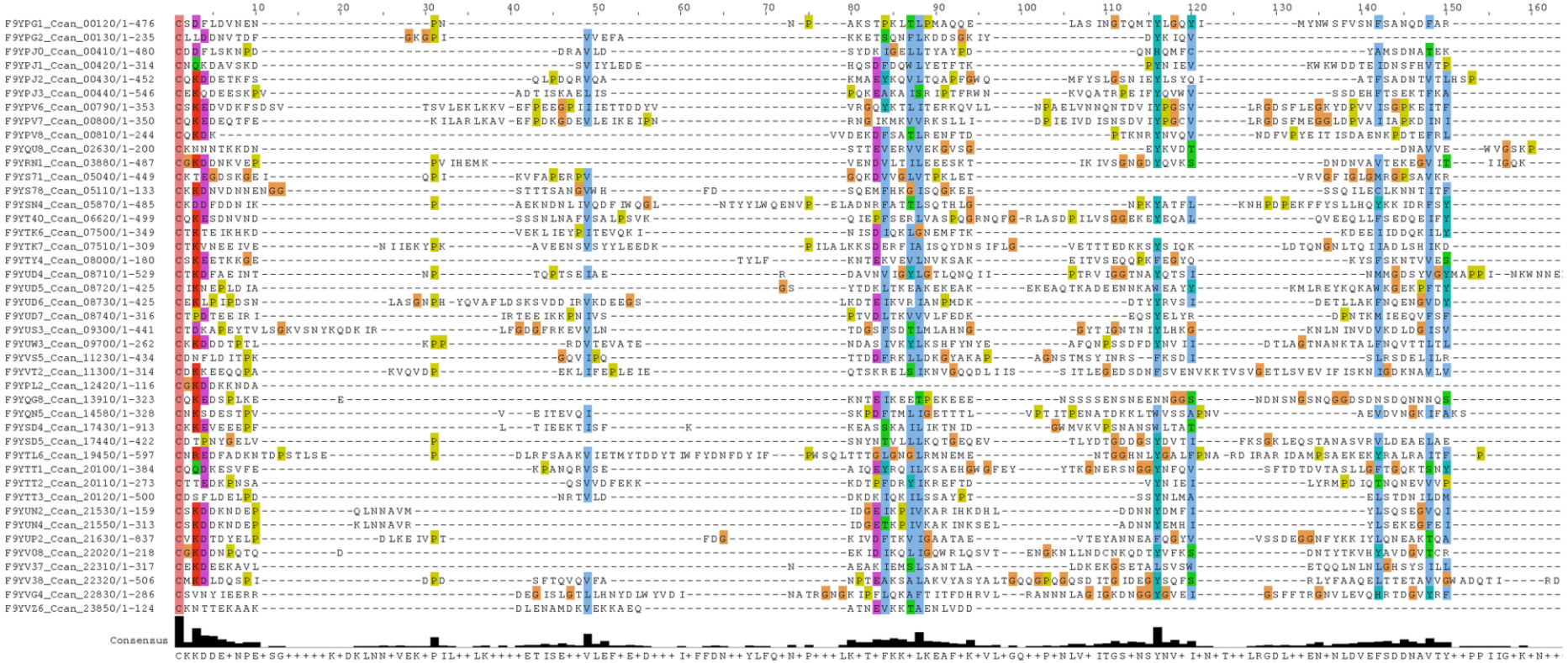


Figure S1. Multiple sequence alignment of full length *C. canimorsus* surface lipoproteins

MAFFT alignment of mature surface exposed lipoproteins. Only the N-terminal region, showing the conserved K-(D/E) motif, is displayed. Highly conserved residues are highlighted according to Clustal color code (R, K in red; D, E in magenta; P in yellow; G in orange; Q, N, S, T in green, C in pink; A, I, L, M, F, W, V in blue; H, Y in cyan). The derived consensus sequence is shown below.

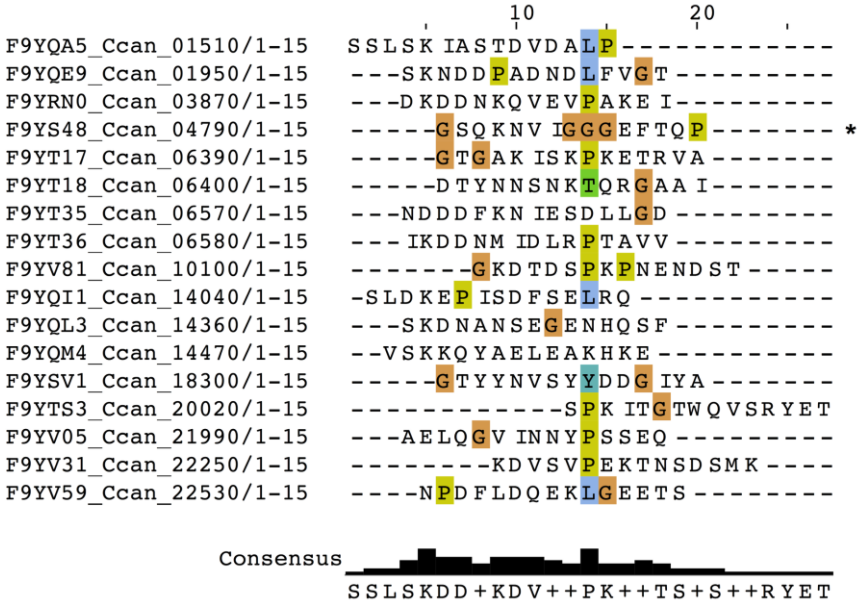


Figure S2. Multiple sequence alignment of *C. canimorsus* periplasmic outer membrane lipoproteins

MAFFT alignment of the first 15 N-terminal amino acids of intracellular OM lipoproteins. The first invariant cysteine residue of each sequence was removed before performing the alignment. Highly conserved residues are highlighted according to Clustal color code (for details, see Fig. S1). The derived consensus sequence is shown below. SiaC (Ccan_04790) is indicated by a star.

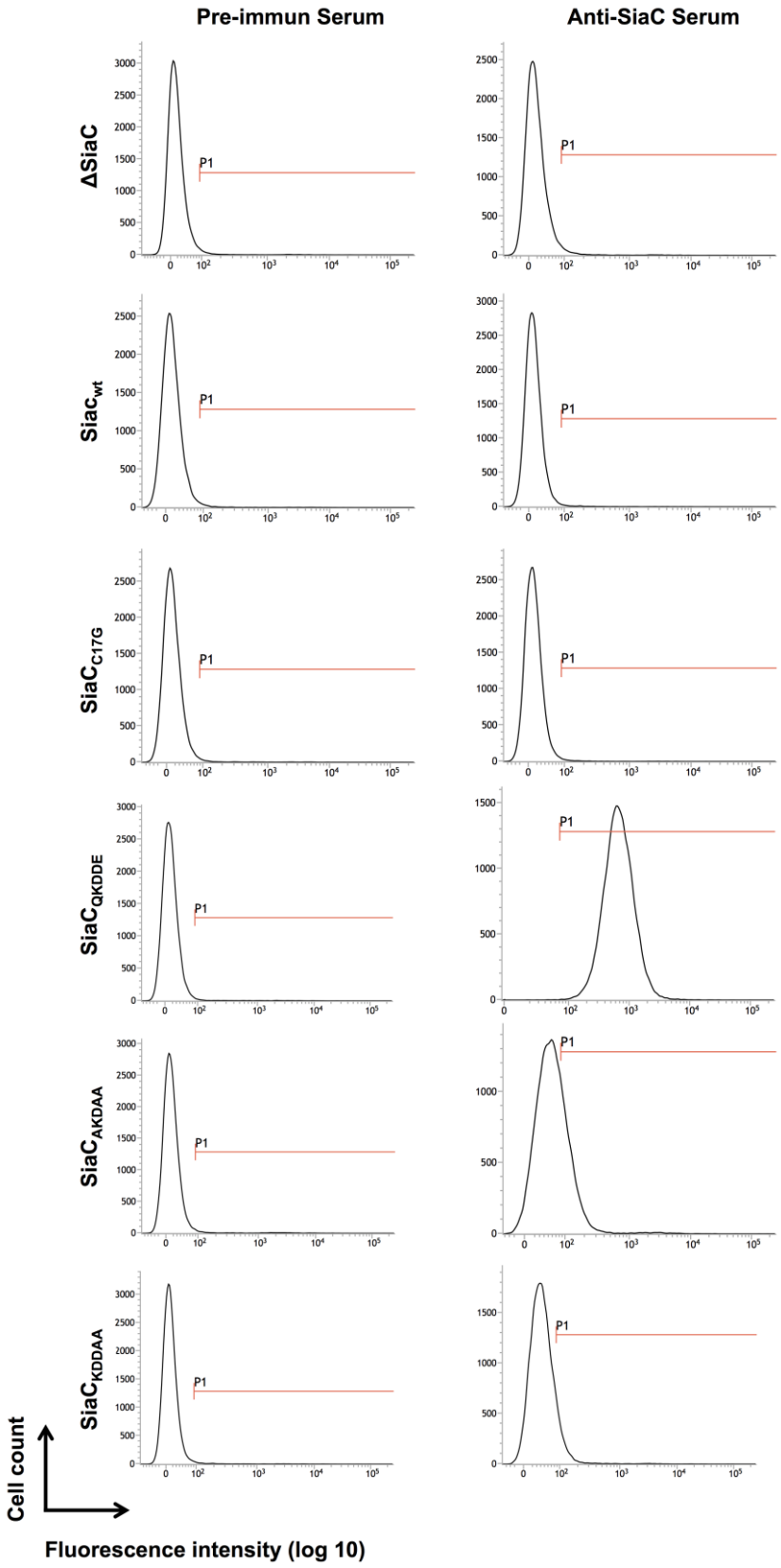


Figure S3. Flow cytometry analysis of anti-SiaC stained cells

Shown are representative experiments for selected strains stained with either pre-immunization serum (negative control) or anti-SiaC serum. A shift of the fluorescence intensity in the P1 channel indicates that cells are stained, *i.e.* SiaC is surface exposed

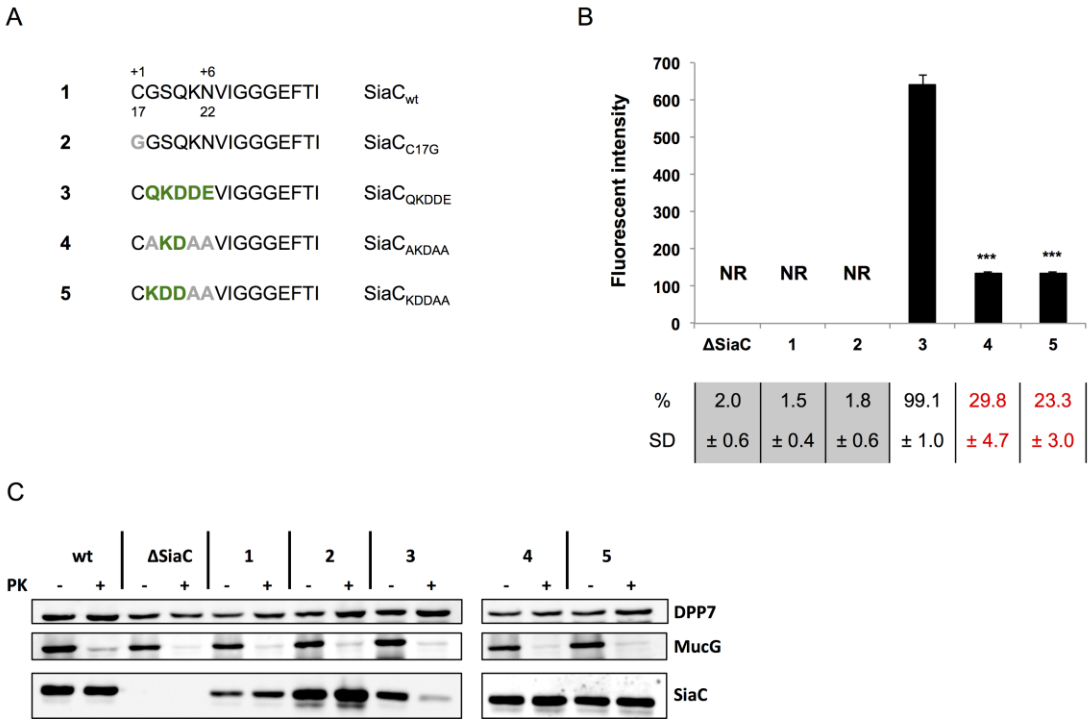


Figure S4. The *C. canimorsus* LES leads to efficient transport of SiaC to the cell surface

(A) SiaC wt and consensus sequence mutant constructs. Amino acids derived from the consensus sequence are indicated in bold green, point mutations are indicated in bold grey. (B) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 3. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 3). (C) Detection of SiaC by western blot analysis of intact cell expressing the SiaC constructs shown in (A) treated with proteinase K (+) or with reaction buffer (-). DPP7 expression was monitored as loading control, MucG was used as positive control for proteinase K accessibility. The wt strain was used as external control.

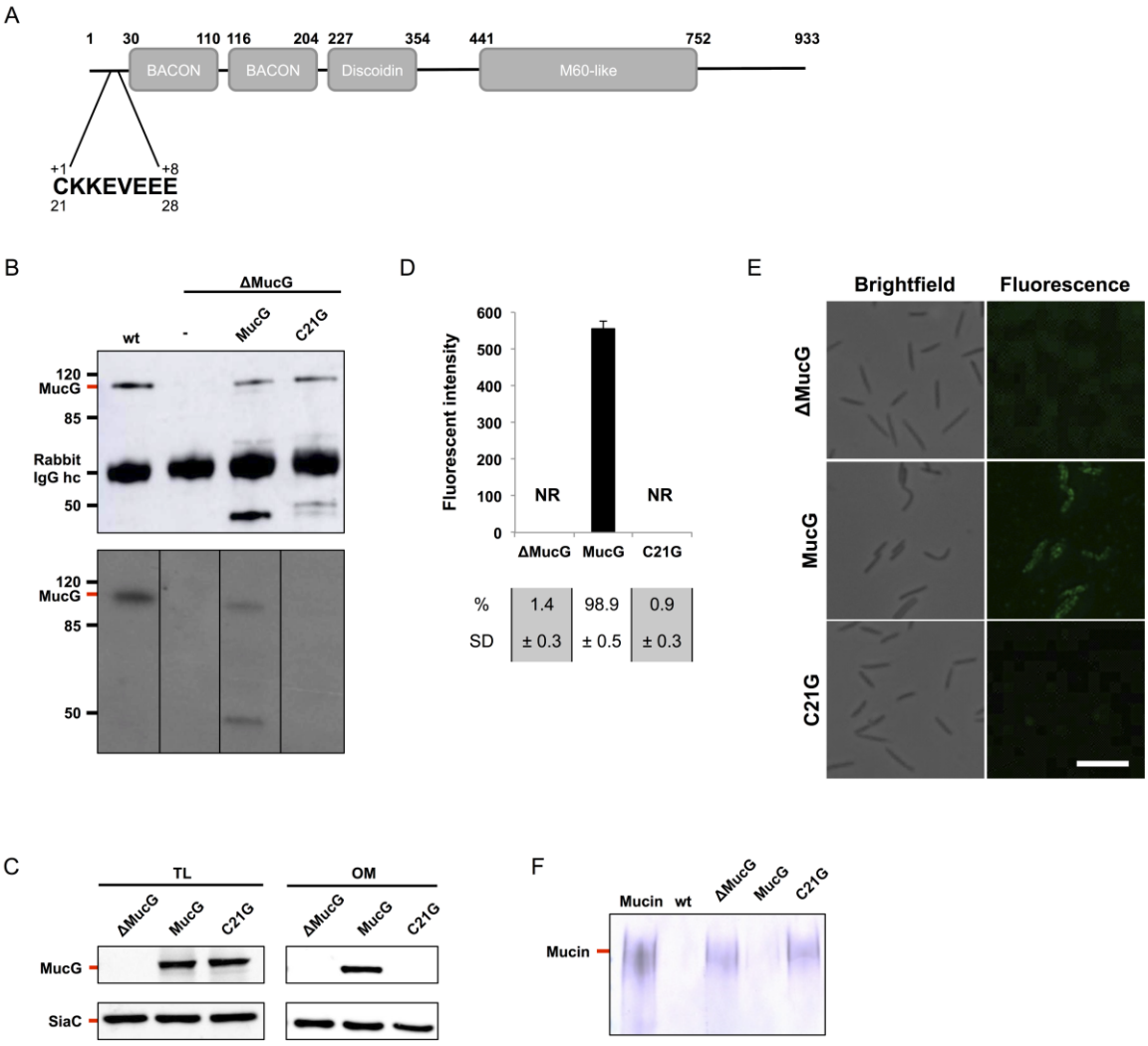


Figure S5. MucG is a surface exposed lipoprotein

(A) MucG domain annotation. Predicted structural domains are indicated by grey boxes, amino acid positions are indicated on top. The predicted LES is shown below. (B) Western blot analysis (top) and fluorography (bottom) of the elution fraction of MucG immunoprecipitation of ^{3}H palmitate labeled bacteria. MucG is lipidated in the wt and Δ*mucG* + MucG strains but not in the Δ*mucG* + $MucG_{C21G}$ strain in which the predicted site of lipidation is mutated, showing that MucG is a lipoprotein. Rabbit IgG hc correspond to the heavy chain of the rabbit MucG antiserum present in the analyzed elution fraction. The low molecular weight band in the MucG strain likely represents a truncated MucG form due to overexpression. This band being radiolabeled, this indicates that the truncation takes place at the C-terminus of MucG. The two low molecular weight bands in the $MucG_{C21G}$ mutant likely represent two different MucG truncated forms that are generated when the protein overexpressed is not lipidated and periplasmic. (C) MucG detection by western blot analysis of total cell lysates (TL) and outer membrane (OM) fractions of bacteria expressing different MucG constructs. MucG but not the soluble $MucG_{C21G}$ is detected in the OM fraction, showing that MucG is a *bona fide* OM lipoprotein. SiaC expression was monitored as loading control. (D) Quantification of MucG surface exposure by flow cytometry of live cells labeled with anti-MucG serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean.. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤2.5 %) are highlighted in grey. (E)

Immunofluorescence microscopy images of bacteria labeled with anti-MucG serum. Scale bar: 5 μm. (F) Detection of mucin by PNA lectin staining of human saliva following incubation with bacteria expressing different MucG constructs performed as in reference 41. Untreated saliva serves as negative control. Reduction of PNA staining indicates mucin degradation by surface localized MucG.

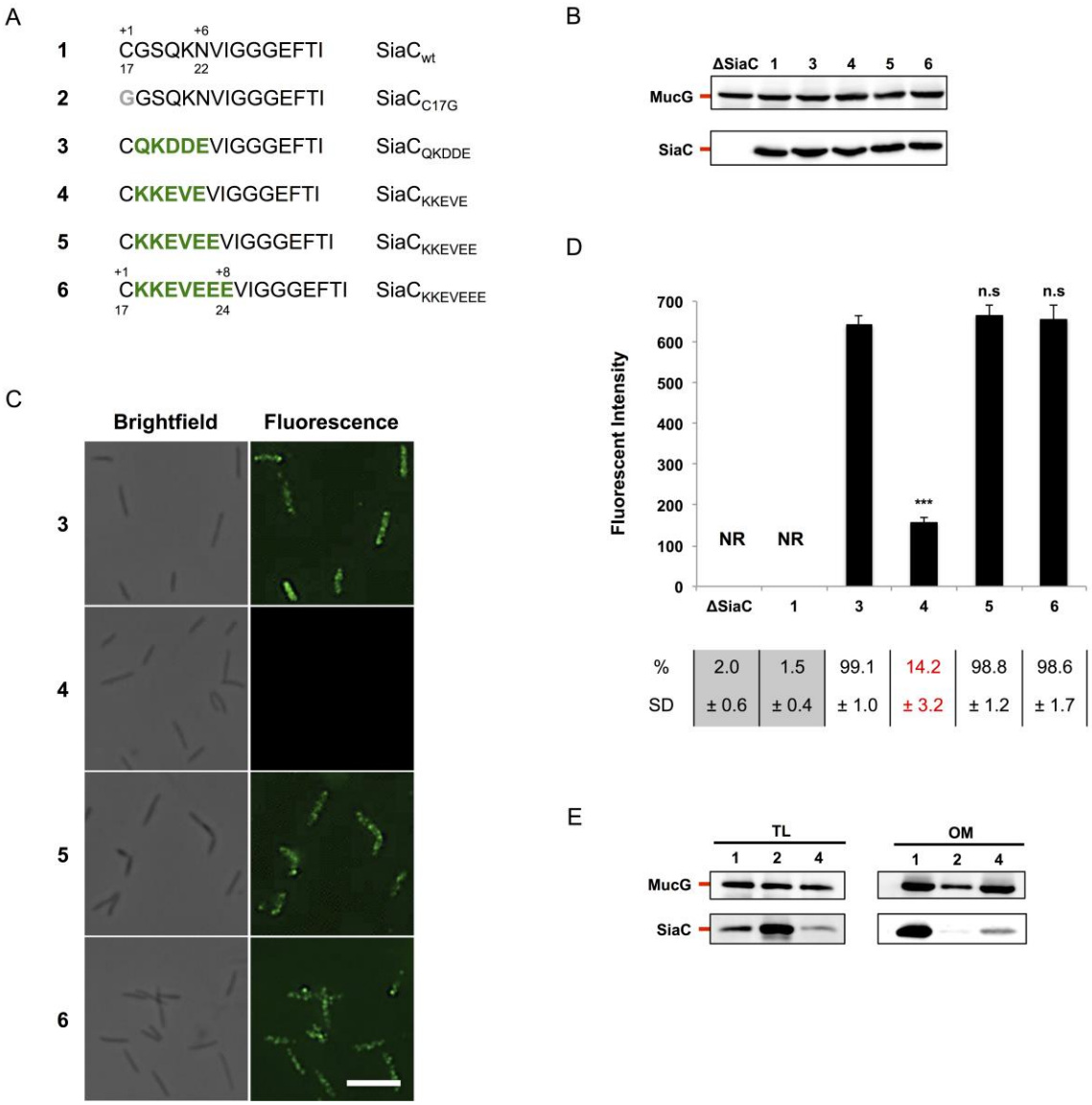
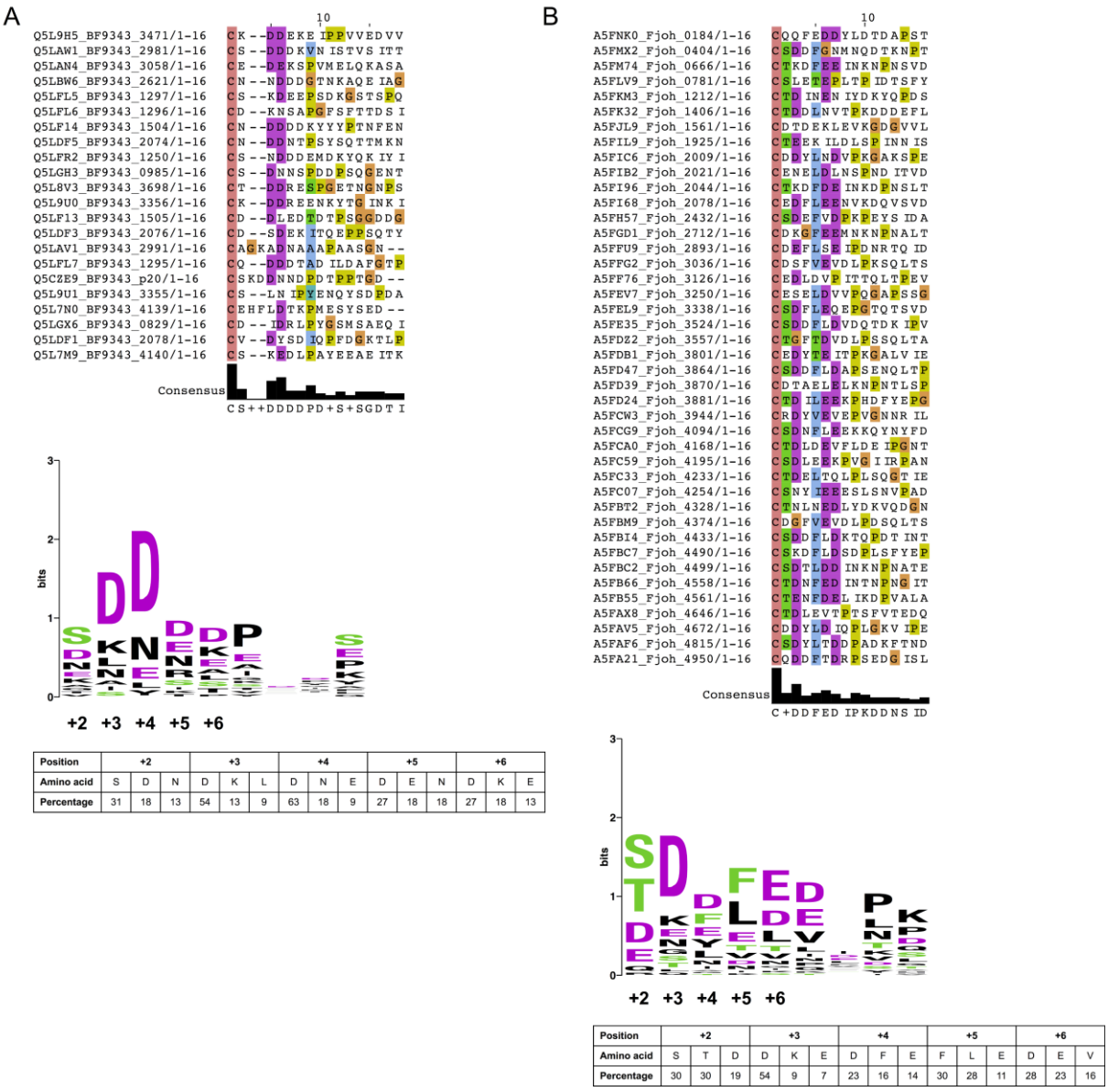


Figure S6. The MucG LES allows SiaC surface localization

(A) SiaC wt and MucG LES sequence mutant constructs. Amino acids derived from the consensus or MucG LES are indicated in bold green, point mutations are indicated in bold grey. (B) Detection of SiaC by western blot analysis of total cell extracts of strains expressing the SiaC constructs shown in (A). MucG expression was monitored as loading control. (C) Immunofluorescence microscopy pictures of bacteria labeled with anti-SiaC serum. Scale bar: 5 µm. (D) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$ as compared to reference construct 3; n.s: not significant. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 3). (E) Western blot analysis of total lysates (TL) and outer membrane (OM) fractions of bacteria expressing different SiaC constructs. MucG expression was monitored as loading control.



Position	+2			+3			+4			+5			+6		
Amino acid	S	D	N	D	K	L	D	N	E	D	E	N	D	K	E
Percentage	31	18	13	54	13	9	63	18	9	27	18	18	27	18	13

Position	+2			+3			+4			+5			+6		
Amino acid	S	T	D	D	K	E	D	F	E	F	L	E	D	E	V
Percentage	30	30	19	54	9	7	23	16	14	30	28	11	28	23	16

Figure S7. Multiple sequence alignment of *B. fragilis* and *F. johnsoniae* surface lipoproteins
(A) MAFFT alignment of the first 16 N-terminal amino acids of proteinase K sensitive *B. fragilis* lipoproteins. (B) MAFFT alignment of the first 16 N-terminal amino acids of SusD-like *F. johnsoniae* lipoproteins. Highly conserved residues are highlighted according to Clustal color code (for details, see Fig. S1). Corresponding Weblogo and amino acid frequencies are indicated below.

Table S1. Bacterial strains used in this study

Strain	Genotype and/or description	Reference
E. coli		
Top10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lac*ZΔM15 Δ*lacX*74 *recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL endA1 nupG*; Sm^r	Invitrogen
C. canimorsus		
Cc5	Wild type (BCCM-LMG 28512)	(39)
Δ*siaC*	Replacement of *Ccan_04790* by *ermF*; Em^r	(39)
Δ*mucG*	Replacement of *Ccan_17430* by *ermF*; Em^r	(41)

Table S2. Plasmids used in this study

Plasmid	Description	Reference
Vectors[a]		
pMM47.A	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ermF* promoter	(39)
pPM5	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ompA* promoter	(40)
Expression plasmids		
pFL43	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL44	Full length *mucG* C21G with a C-terminal HA tag amplified with primers 7259/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL71	Full length *mucG* K22A with a C-terminal HA tag amplified with primers 7182/7487 and 7486/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL72	Full length *mucG* K23A with a C-terminal HA tag amplified with primers 7182/7489 and 7488/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL73	Full length *mucG* E24A with a C-terminal HA tag amplified with primers 7182/7491 and 7490/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL74	Full length *mucG* V25A with a C-terminal HA tag amplified with primers 7182/7493 and 7492/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL75	Full length *mucG* E26A with a C-terminal HA tag amplified with primers 7182/7495 and 7494/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL76	Full length *mucG* E27A with a C-terminal HA tag amplified with primers 7182/8048 and 8047/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL77	Full length *mucG* E28A with a C-terminal HA tag amplified with primers 7182/8050 and 8049/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	This study
pFL79	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7510 and 7509/7625 and cloned into pPM5 using NcoI/XhoI restriction sites. Replacement of aa 22-28 by AAEVEEE	This study
pFL84	Full length *mucG* with a C-terminal HA tag amplified with primers	This study

	7182/7899 and 7898/7625 and cloned into pPM5 using NcoI/XhoI restriction sites. Replacement of aa 22-28 by KKAAAAA	
pFL117	Full length *siaC* amplified with primers 4159 and 7696 and cloned into pMM47.A using NcoI/XhoI restriction sites	This study
pFL118	Full length *siaC* C17G amplified with primers 5545 and 7696 and cloned into pMM47.A using NcoI/XhoI restriction sites	This study
pFL132	Full length *siaC* amplified with primers 4159/8017 and 8016/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KKEVE	This study
pFL133	Full length *siaC* amplified with primers 4159/8054 and 8052/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KKEVEE	This study
pFL134	Full length *siaC* amplified with primers 4159/7972 and 7971/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KKEVEEE	This study
pFL143	Full length *siaC* amplified with primers 4159/8058 and 8057/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by QKDDE	This study
pFL144	Full length *siaC* amplified with primers 4159/8086 and 8085/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKDDE	This study
pFL145	Full length *siaC* amplified with primers 4159/8084 and 8083/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKDDA	This study
pFL146	Full length *siaC* amplified with primers 4159/8153 and 8152/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKEEA	This study
pFL147	Full length *siaC* amplified with primers 4159/8149 and 8148/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKDAA	This study
pFL148	Full length *siaC* amplified with primers 4159/8151 and 8150/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKEAA	This study
pFL149	Full length *siaC* amplified with primers 4159/8157 and 8156/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AAKDD	This study
pFL150	Full length *siaC* amplified with primers 4159/8159 and 8158/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AAAKDD	This study
pFL151	Full length *siaC* amplified with primers 4159/8161 and 8160/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AAAAKDD	This study
pFL152	Full length *siaC* amplified with primers 4159/8169 and 8168/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KDDAA	This study
pFL153	Full length *siaC* amplified with primers 4159/8165 and 8164/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by QADDE	This study
pFL154	Full length *siaC* amplified with primers 4159/8167 and 8166/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AADDA	This study
pFL155	Full length *siaC* amplified with primers 4159/8164 and 8163/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by SDDFE	This study
pFL156	Full length *siaC* amplified with primers 4159/8173 and 8172/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by SDDDD	This study

[a]: Selection markers for *C. canimorsus* are in between brackets

Table S3. Primers used in this study

Ref.	Sequence 5'-3'	Restriction[a]
4159	cataccatgggaaatcgaatttttatctt	NcoI
5545	catgccatgggaaatcgaatttttatcttttattcgcttttgttcttttgtcggctggtggaagcc aaaaaaacg	NcoI
7182	ggccatggggaaaaaaatagtatccattagc	NcoI
7259	ggccatggggaaaaaaatagtatccattagcttatttttccttatctcagcaactatttggttag ccggtaaaaaggaag	NcoI
7625	ggctcgagctaagcgtaatctggaacatcgtatgggtaaaacgtaacttgagttctc	XhoI
7696	ggctcgagttagttcttgataaattcctcaactgg	XhoI
7486	tggttagcctgtgcaaaggaagttgaagaagaacc	
7487	ggttcttcttcaacttcctttgcacaggctaacca	
7488	ttagcctgtaaagcggaagttgaagaagaacctttttc	
7489	gaaaaggttcttcttcaacttccgctttacaggctaa	
7490	gcctgtaaaaaggcagttgaagaagaacctttttctaac	
7491	gttagaaaaggttcttcttcaactgccttttttacaggc	
7492	tgtaaaaaggaagctgaagaagaacctttttctaac	
7493	gttagaaaaggttcttcttcagcttccttttttaca	
7494	aaaaaggaagttgcagaagaacctttttctaacaatag	
7495	ctattgttagaaaaggttcttctgcaacttccttttt	
7509	tggttagcctgtgcagcggaagttgaagaagaacc	
7510	ggttcttcttcaacttccgctgcacaggctaacca	
7898	gcagctgcagcggctccttttctaacaatagaagaaaaaacc	
7899	agccgctgcagctgccttttttacaggctaaccaaatagttgc	
7971	aaaaaggaagttgaagaagaagtaatcggcggaggcgaatttacacaacccg	
7972	ttcttcttcaacttccttttttacaagccgacaaaagaacaaaagcg	
8016	aaaaaggaagttgaagtaatcggcggaggcgaatttacacaacccg	
8017	ttcaacttccttttttacaagccgacaaaagaacaaaagcg	
8047	aggaagttgaagcagaacctttttctaacaatagaagaaaaaacc	
8048	gaaaaggttctgcttcaacttccttttttacaggctaacc	
8049	ggaagttgaagaagcacctttttctaacaatagaagaaaaaacc	
8050	gaaaaggtgcttcttcaacttccttttttacaggctaaccaaatagttg	
8052	aaaaaggaagttgaagaagtaatcggcggaggcgaatttacacaacccg	
8054	ttcttcaacttccttttttacaagccgacaaaagaacaaaagcg	
8057	caaaaggacgatgaagtaatcggcggaggcgaatttacacaacccg	
8058	ttcatcgtccttttgacaagccgacaaaagaacaaaagcg	
8083	gcaaaggacgatgcagtaatcggcggaggcgaatttacacaacccg	
8084	tgcatcgtcctttgcacaagccgacaaaagaacaaaagcg	
8085	gcaaaggacgatgaagtaatcggcggaggcgaatttacacaacccg	
8086	ttcatcgtcctttgcacaagccgacaaaagaacaaaagcg	
8148	gcaaaggacgctgcagtaatcggcggaggcgaatttacacaacccg	

8149	tgcagcgtcctttgcacaagccgacaaaagaacaaaagcg	
8150	gcaaaggaagctgcagtaatcggcggaggcgaatttacacaacccg	
8151	tgcagcttcctttgcacaagccgacaaaagaacaaaagcg	
8152	gcaaaggaagaggcagtaatcggcggaggcgaatttacacaacccg	
8153	tgcctcttccctttgcacaagccgacaaaagaacaaaagcg	
8156	gctgcaaaggacgatgtaatcggcggaggcgaatttacacaacccg	
8157	atcgtcctttgcagcacaagccgacaaaagaacaaaagcg	
8158	gcagctgcaaaggacgatgtaatcggcggaggcgaatttacacaacccg	
8159	atcgtcctttgcagctgcacaagccgacaaaagaacaaaagcg	
8160	gccgcagctgcaaaggacgatgtaatcggcggaggcgaatttacacaacccg	
8161	atcgtcctttgcagctgcggcacaagccgacaaaagaacaaaagcg	
8162	tctgatgacttcgaagtaatcggcggaggcgaatttacacaacccg	
8163	ttcgaagtcatcagaacaagccgacaaaagaacaaaagcg	
8164	caagcggacgatgaagtaatcggcggaggcgaatttacacaacccg	
8165	ttcatcgtccgcttgacaagccgacaaaagaacaaaagcg	
8166	gcagctgacgatgcagtaatcggcggaggcgaatttacacaacccg	
8167	tgcatcgtcagctgcacaagccgacaaaagaacaaaagcg	
8168	aaggacgatgcagctgtaatcggcggaggcgaatttacacaacccg	
8169	agctgcatcgtccttacaagccgacaaaagaacaaaagcg	
8172	agtgatgacgacgatgtaatcggcggaggcgaatttacacaacccg	
8173	atcgtcgtcatcactacaagccgacaaaagaacaaaagcg	

[a]: Restriction sites are underlined

Table S4. *C. canimorsus* 5 surface exposed lipoproteins

Uniprot Accession	ORF name	Annotation	SPII cleavage site[c]	% of surfome[d]
F9YPG1	Ccan_00120	Uncharacterized protein	22-23	8.35
F9YPG2	Ccan_00130	Uncharacterized protein	19-20	4.25
F9YPJ0	Ccan_00410	Uncharacterized protein	18-19	0.23
F9YPJ1	Ccan_00420	Uncharacterized protein	17-18	0.32
F9YPJ2	Ccan_00430	Uncharacterized protein	20-21	0.27
F9YPJ3	Ccan_00440	Uncharacterized protein	19-20	0.14
F9YPV6	Ccan_00790	Uncharacterized protein	19-20	12.80
F9YPV7	Ccan_00800	Tetanolysin O	19-20	0.58
F9YPV8	Ccan_00810	Uncharacterized protein	12-13	0.46
F9YQU8	Ccan_02630	UPF0312 protein	19-20	3.63
F9YRN1	Ccan_03880	TvBspA-like-625	20-21	1.24
F9YS71	Ccan_05040	Glycosyl hydrolase family 109 protein 5 (EC 3.2.1.49)	26-27	/
F9YS78	Ccan_05110	Uncharacterized protein	18-19	0.69
F9YSN4	Ccan_05870	Carboxyl-terminal-processing protease (EC 3.4.21.102)	16-17	1.02
F9YT40	Ccan_06620	Thiol-activated cytolysin	21-22	1.37

F9YTK6	Ccan_07500	Uncharacterized protein	16-17	0.45
F9YTK7	Ccan_07510	Uncharacterized protein	15-16	0.20
F9YTY4	Ccan_08000	Uncharacterized protein	19-20	/
F9YUD4	Ccan_08710	GpdD	16-17	3.99
F9YUD5	Ccan_08720	GpdG	20-21	3.43
F9YUD6	Ccan_08730	GpdE	16-17	1.28
F9YUD7	Ccan_08740	GpdF	17-18	3.25
F9YUS3	Ccan_09300	Thioredoxin family protein (EC 1.8.1.8)	16-17	/
F9YUW3	Ccan_09700	Peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8)	19-20	0.71
F9YVS5	Ccan_11230	Uncharacterized protein	17-18	0.17
F9YVT2	Ccan_11300	Uncharacterized protein	17-18	1.11
F9YPL2	Ccan_12420	Uncharacterized protein	18-19	2.57
F9YQG8	Ccan_13910	Uncharacterized protein	21-22	0.27
F9YQN5	Ccan_14580	Internalin-J (EC 3.2.1.83)	23-24	0.23
F9YSD4	Ccan_17430[a]	MucG mucinase	20-21	1.29
F9YSD5	Ccan_17440	MucE	18-19	8.99
F9YTL6	Ccan_19450	Uncharacterized protein	18-19	5.15
F9YTT1	Ccan_20100	Uncharacterized protein	19-20	/
F9YTT2	Ccan_20110	Uncharacterized protein	20-21	1.64
F9YTT3	Ccan_20120[b]	Uncharacterized protein	20-21	2.08
F9YUN2	Ccan_21530	Uncharacterized protein	23-24	/
F9YUN4	Ccan_21550	Uncharacterized protein	23-24	0.09
F9YUP2	Ccan_21630	Uncharacterized protein	24-25	11.30
F9YV08	Ccan_22020	Uncharacterized protein	17-18	0.03
F9YV37	Ccan_22310	Uncharacterized protein	21-22	0.17
F9YV38	Ccan_22320	Uncharacterized protein	20-21	0.19
F9YVG4	Ccan_22830	Uncharacterized protein	16-17	0.12
F9YVZ6	Ccan_23850	Uncharacterized protein	17-18	/
			Total	**84,06**

[a]: Using the annotated translational start site Ccan_17430 is predicted to be a cytoplasmic protein, but if translation begins at an AUG 13 codons downstream then it is predicted to be a lipoprotein

[b]: Using the annotated translational start site Ccan_20120 is predicted to be a cytoplasmic protein, but if translation begins at an AUG 18 codons downstream then it is predicted to be a lipoprotein.

[c]: SPII cleavage site predicted by the LipoP software; numbers indicate the position of the last amino acid of the signal peptide and the position of the +1 cysteine.

[d]: Quantitative contribution to surfome composition, expressed in percentage, as described in (17).

'/' stands for not quantified.

Table S5. Bacteroidetes outer membrane lipoproteins

Uniprot Accession	ORF name	Annotation	SPII cleavage site[c]
***C. canimorsus* 5 periplasmic outer membrane lipoproteins**			
F9YQA5	Ccan_01510	Putative Subtilisin (EC 3.4.21.62)	18-19
F9YQE9	Ccan_01950	Uncharacterized protein	19-20
F9YRN0	Ccan_03870	Surface antigen BspA	20-21
F9YS48	Ccan_04790	Neuraminidase	16-17
F9YT17	Ccan_06390	Membrane or secreted protein	15-16
F9YT18	Ccan_06400	Inner membrane lipoprotein yiaD	16-17
F9YT35	Ccan_06570	Uncharacterized protein	19-20
F9YT36	Ccan_06580	Uncharacterized protein	22-23
F9YV81	Ccan_10100	Uncharacterized protein	19-20
F9YQI1	Ccan_14040	Uncharacterized protein	16-17
F9YQL3	Ccan_14360	Uncharacterized protein	32-33
F9YQM4	Ccan_14470	OmpA/MotB C-terminal like outer membrane protein	17-18
F9YSV1	Ccan_18300	Uncharacterized protein	25-26
F9YTS3	Ccan_20020	Uncharacterized protein	20-21
F9YV05	Ccan_21990	Uncharacterized protein	16-17
F9YV31	Ccan_22250	TvaII (EC 3.2.1.1)	36-37
F9YV59	Ccan_22530	Uncharacterized protein	20-21
***B. fragilis* NCTC 9343 proteinase K sensitive surface exposed lipoproteins**			
Q5L9H5	BF9343_3471	Uncharacterized protein	21-22
Q5LAW1	BF9343_2981	Putative lipoprotein	22-23
Q5LAN4	BF9343_3058	Putative lipoprotein	18-19
Q5LBW6	BF9343_2621	Putative lipoprotein	22-23
Q5LFL5	BF9343_1297	Uncharacterized protein	18-19
Q5LFL6	BF9343_1296	Uncharacterized protein	20-21
Q5LF14	BF9343_1504	Uncharacterized protein	25-26
Q5LDF5	BF9343_2074	Putative exported protein	21-22
Q5LFR2	BF9343_1250	Uncharacterized protein	22-23
Q5LGH3	BF9343_0985	Conserved hypothetical lipoprotein	24-25
Q5L8V3	BF9343_3698	Putative exported protein	20-21
Q5L9U0	BF9343_3356	Putative lipoprotein	23-24
Q5LF13	BF9343_1505	Uncharacterized protein	37-38
Q5LDF3	BF9343_2076	Putative lipoprotein	25-26
Q5LAV1	BF9343_2991	Putative exported protein	19-20
Q5LFL7	BF9343_1295[b]	Uncharacterized protein	24-25
Q5CZE9	BF9343_p20[c]	Uncharacterized protein	18-19
Q5L9U1	BF9343_3355	Uncharacterized protein	29-30
Q5L7N0	BF9343_4139	Putative outer membrane protein	28-29
Q5LGX6	BF9343_0829	Possible outer membrane protein	16-17
Q5LDF1	BF9343_2078	Conserved hypothetical lipoprotein	21-22

Q5L7M9	BF9343_4140	Uncharacterized protein	25-26
***F. johnsoniae* UW101 SusD-like lipoproteins**			
A5FNK0	Fjoh_0184	RagB/SusD domain protein	22-23
A5FMX2	Fjoh_0404	RagB/SusD domain protein	17-18
A5FM74	Fjoh_0666	RagB/SusD domain protein	19-20
A5FLV9	Fjoh_0781	RagB/SusD domain protein	26-27
A5FKM3	Fjoh_1212	RagB/SusD domain protein	19-20
A5FK32	Fjoh_1406	RagB/SusD domain protein	24-25
A5FJL9	Fjoh_1561	RagB/SusD domain protein	21-22
A5FIL9	Fjoh_1925	RagB/SusD domain protein	19-20
A5FIC6	Fjoh_2009	RagB/SusD domain protein	20-21
A5FIB2	Fjoh_2021	RagB/SusD domain protein	18-19
A5FI96	Fjoh_2044	RagB/SusD domain protein	22-23
A5FI68	Fjoh_2078	RagB/SusD domain protein	20-21
A5FH57	Fjoh_2432	RagB/SusD domain protein	21-22
A5FGD1	Fjoh_2712	RagB/SusD domain protein	21-22
A5FFU9	Fjoh_2893	RagB/SusD domain protein	18-19
A5FFG2	Fjoh_3036	RagB/SusD domain protein	34-35
A5FF76	Fjoh_3126	RagB/SusD domain protein	20-21
A5FEV7	Fjoh_3250	RagB/SusD domain protein	19-20
A5FEL9	Fjoh_3338	RagB/SusD domain protein	24-25
A5FE35	Fjoh_3524	RagB/SusD domain protein	17-18
A5FDZ2	Fjoh_3557	RagB/SusD domain protein	27-28
A5FDB1	Fjoh_3801	RagB/SusD domain protein	18-19
A5FD47	Fjoh_3864	RagB/SusD domain protein	21-22
A5FD39	Fjoh_3870	RagB/SusD domain protein	20-21
A5FD24	Fjoh_3881	RagB/SusD domain protein	21-22
A5FCW3	Fjoh_3944	RagB/SusD domain protein	19-20
A5FCG9	Fjoh_4094	RagB/SusD domain protein	20-21
A5FCA0	Fjoh_4168	RagB/SusD domain protein	23-24
A5FC59	Fjoh_4195	RagB/SusD domain protein	17-18
A5FC33	Fjoh_4233	RagB/SusD domain protein	21-22
A5FC07	Fjoh_4254	RagB/SusD domain protein	17-18
A5FBT2	Fjoh_4328	RagB/SusD domain protein	18-19
A5FBM9	Fjoh_4374	RagB/SusD domain protein	34-35
A5FBI4	Fjoh_4433	RagB/SusD domain protein	20-21
A5FBC7	Fjoh_4490	RagB/SusD domain protein	24-25
A5FBC2	Fjoh_4499	RagB/SusD domain protein	17-18
A5FB66	Fjoh_4558	RagB/SusD domain protein	20-21
A5FB55	Fjoh_4561	RagB/SusD domain protein	18-19
A5FAX8	Fjoh_4646	RagB/SusD domain protein	17-19
A5FAV5	Fjoh_4672	RagB/SusD domain protein	19-20
A5FAF6	Fjoh_4815	RagB/SusD domain protein	25-26
A5FA21	Fjoh_4950	RagB/SusD domain protein	24-25

[a]: SPII cleavage site predicted by the LipoP software; numbers indicate the position of the last amino acid of the signal peptide and the position of the +1 cysteine.
[b]: As described in reference (45), the translational start site of BF9343_1295 was moved 15 codons downstream, resulting in a predicted lipoprotein.
[c]: As described in reference (45), the translational start site of BF9343_p20 was moved 38 codons downstream, resulting in a predicted lipoprotein.

1.2. Extended results : Identification of a new lipoprotein export signal in Gram-negative bacteria

1.2.1. Characterization of the MucG LES in SiaC

Our *in silico* analysis identified as MucG LES the sequence 22-KKEVEEE-28 (Fig. S5A). This was confirmed by mutational analysis in MucG (Fig. 4) as well as by introducing this sequence into SiaC resulting in protein surface exposure (Fig. S6). However, insertion of the sequence 22-KKEVE-26 into SiaC led to very poor surface localization of the protein (Fig. S6C and D) thus indicating the requirement of a negatively charged LES. Indeed, the 22-KKEVE-26 peptide is neutral in charge due to the presence of two positively and two negatively charged residues while 22-KKEVEE-27 and 22-KKEVEEE-28, both leading to clear surface localization of SiaC (Fig. S6C and D), have an overall negative charge thanks to the one or two additional glutamate residues.

In order to further confirm this hypothesis, we generated two SiaC contructs by replacing the lysine residues at position 22 or 23 by alanines ($SiaC_{+2AKEVE+6}$ and $SiaC_{+2KAEVE+6}$ respectively) thus rendering the signal's overall charge negative (Fig. E1A). Following western blot analysis to confirm protein expression (Fig. E1B), we monitored the presence of these SiaC variants at the cell surface by flow cytometry (Fig. E1C). Interestingly, $SiaC_{+2AKEVE+6}$ was surface localized in 79.3 ± 3.4 % of the cells (Fig. E1C), although the total amount of SiaC displayed by each cell was lower than in the $SiaC_{+2KKEVEE+7}$ and $SiaC_{+2KKEVEEE+8}$ constructs (approximately 25 %). This represents however a dramatic increase as compared to $SiaC_{+2KKEVE+6}$ and confirmed that removal of one positively charged amino acid (K22) does indeed favor surface targeting. The fact that only a small amount of $SiaC_{+2AKEVE+6}$ was transported to the surface could reflect our previous finding that glutamate is less efficient at promoting SiaC surface export than aspartate (Fig. 2D and E). In contrast, $SiaC_{+2KAEVE+6}$ behaved as $SiaC_{+2KKEVE+6}$, with very little protein transported to the surface (Fig. E1C). This result highlighted the fact that, although the introduced peptide motif is overall negatively charged, the position of the positively charged amino acid, K at

position +2 rather than +3 as in the consensus sequence, appears critical for proper surface localization of MucG.

To validate this point, we constructed an additional hybrid protein by replacing amino acids 18 to 22 of SiaC by amino acids 23 to 27 of MucG ($SiaC_{+2KEVEE+6}$), shifting the added MucG peptide by one amino acid as compared to $SiaC_{+2KKEVE+6}$. This generates a signal peptide with only one positively charged residue (K) but located at position +2 rather than +3 (Fig. E1A). Similar to the $SiaC_{+2KAEVE+6}$ contruct and in good agreement with our previous results, this construct only localized at the cell surface of 47.9 ± 1.9 % of the cells (Fig. E1C). Additionally, the fluorescence intensity was low, confirming a positional effect of the lysine residue on surface transport.

Taken together, our data with the MucG LES in SiaC confirm the results obtained with the consensus LES in SiaC, namely the compositional as well as positional requirements of the *C. canimorsus* LES.

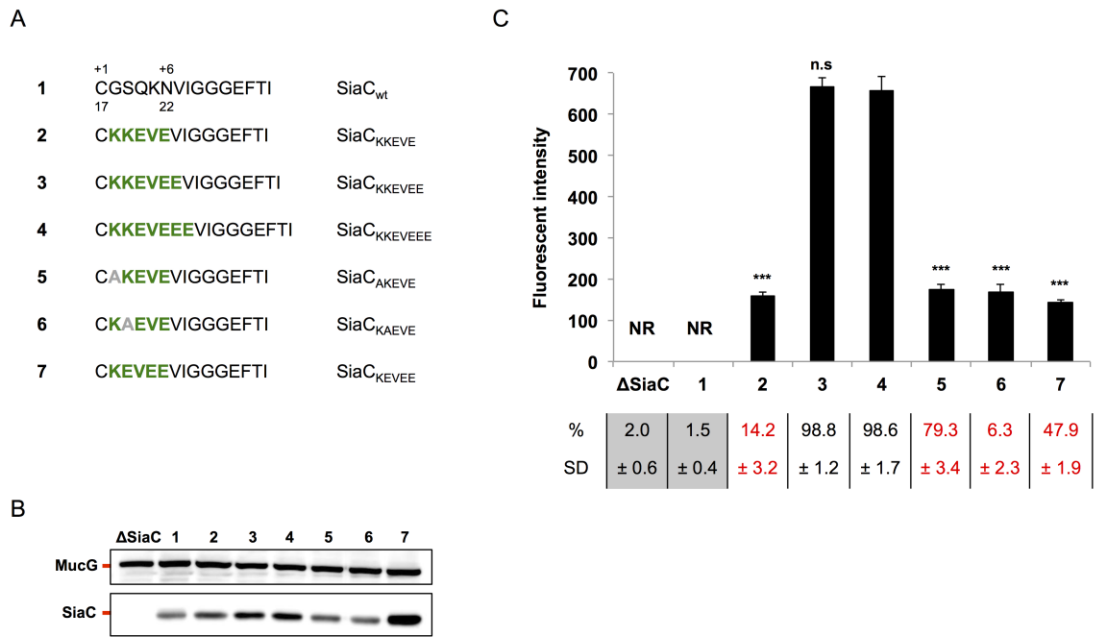


Figure E1. Characterization of the MucG LES in SiaC

(A) SiaC wt and MucG LES sequence mutant constructs. Amino acids derived from MucG are indicated in bold green, point mutations are indicated in bold grey. (B) Detection of SiaC by western blot analysis of total cell extracts of strains expressing the SiaC constructs described in (A). Expression of MucG was monitored as loading control. (C) Quantification of SiaC surface exposure by flow cytometry of live cells labeled with anti-SiaC serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; ***, $p \leq 0.001$

as compared to reference construct 4; n.s: not significant. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 4).

1.2.2. Arginine can functionally replace lysine in the MucG LES

In our initial *in silico* analysis, the lysine located at position +3 was found to be the most conserved residue in *C. canimorsus* surface exposed lipoproteins (Fig. 1B and C). Surprisingly, point mutation of this residue did not affect surface exposure of SiaC unless the +2 residue was also mutated (Fig. 2D and E). In order to clarify whether the high conservation of lysine was linked to the nature of the amino acid itself or solely to its charge, we replaced singularly or simultaneously the lysine residues in the MucG LES by arginine residues (Fig. E2A). The expression of the resulting constructs, $MucG_{+2RREVEEE+8}$, $MucG_{+2RAEVEEE+8}$ and $MucG_{+2AREVEEE+8}$, was then confirmed by western blot (Fig. E2B). Interestingly, substitution of both lysines by arginines led to a clear surface localization of $MucG_{+2RREVEEE+8}$, although slightly lower than in the wt construct (Fig. E2C). This could be explained by the fact that arginine at postion +3 is only rarely found in *C. canimorsus* surface lipoproteins (Fig 1A and B). This also indicated that it is indeed the charge of the amino acid rather than the amino acid itself that is important for surface targeting. $MucG_{+2RAEVEEE+8}$ and $MucG_{+2AREVEEE+8}$ were also both surface exposed, 22-RAEVEEE-28 being even more efficient in MucG export than the wt LES sequence (Fig. E2C). On the other hand, $MucG_{+2AREVEEE+8}$ was less efficiently transported (Fig. E2C).

Taken together, these data show that the charge rather the nature of the amino acid in position +2 or +3 of the LES is involved in MucG surface exposure.

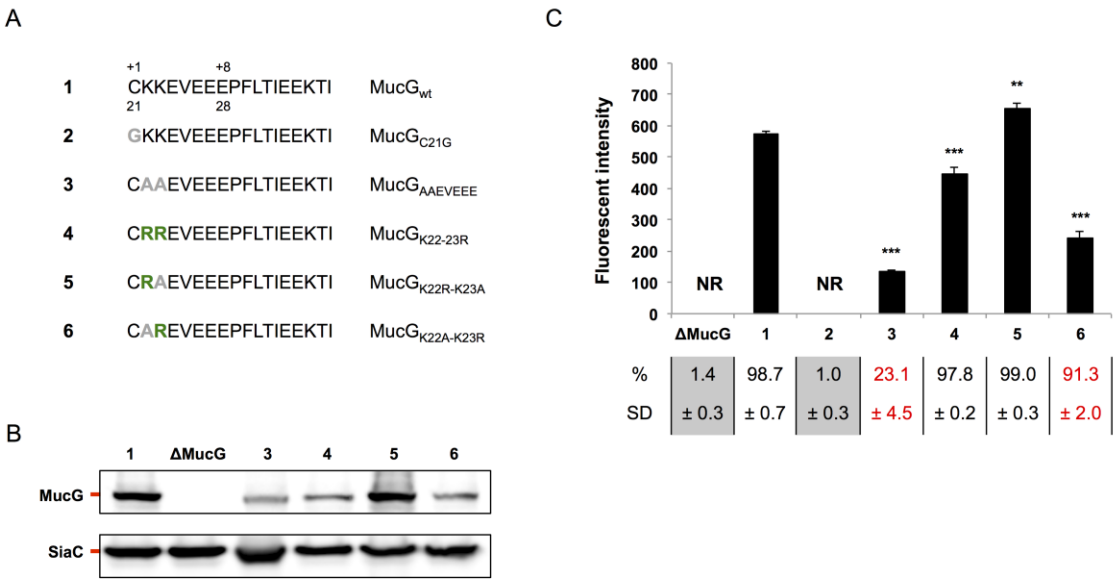


Figure E2. Arginine can functionnaly replace lysine in the MucG LES
(A) MucG wt and mutant constructs. Arginine substitutions are indicated in bold green, alanine substitutions are indicated in bold grey. (B) Detection of MucG by western blot analysis of total cell extracts of strains expressing the MucG constructs described in (A). Expression of SiaC was monitored as loading control. (C) Quantification of MucG surface exposure by flow cytometry of live cells labeled with anti-MucG serum. Shown is the fluorescence intensity of stained cells only; NR: not relevant. The averages from at least three independent experiments are shown. Error bars represent 1 standard deviation from the mean; **, $p \leq 0.01$, ***, $p \leq 0.001$ as compared to reference construct 1. The percentage of stained cells is indicated below; SD: standard deviation. Strains below detection limit (≤ 2.5 %) are highlighted in grey, strains with a statistically significant lower stained population are in red ($p \leq 0.001$ as compared to reference construct 3).

1.2.3. Discussion

Here we pursued the analysis of the MucG LES, a model surface exposed lipoprotein. By introducing a truncated, non-functional derivative of the MucG LES (22-KKEVE-26) into SiaC, we could show by site directed mutagenesis that both the amino acid composition as well as the position of each amino acid relative to the +1 cysteine greatly impact the functionality of the MucG LES. This confirms our previous data obtained with the *C. canimorsus* consensus LES inserted into SiaC (Fig. 2 and 3). Additionally, we showed that arginine is able to functionally replace lysine as positively charged amino acid in this signal sequence. Based on the high conservation of lysine at position +3 in the *C. canimorsus* LES consensus and the almost complete absence of arginine at the same position, it is clear that in *C. canimorsus* lysine is preferred over arginine

although the reasons behind this compositional biais remain unknown. In the same line, it would also be interesting to test if a histidine residue could replace the lysine.

In this regard, identification of the putative lipoprotein export machinery (see chapter 2) would likely help to clarify this matter. Indeed, understanding how the LES interacts with this putative transporter could shed light on why a positive charge next to the lipidated cysteine, although not absolutely required as seen in some constructs of SiaC, favors lipoprotein surface translocation. The same is true for the requirement of an overall negative charge of the LES. One could for exemple envision a specific binding pocket in the transporter composed of positively and negatively charged amino acids that would be complementary to the residues in the LES. Similarly, identification of the transporter could also clarify how lipoproteins such as SiaC remain intracellular while others are transported to the surface. A mechanism similar to the Lol-avoidance signal might take place in which surface lipoproteins interact with their dedicated transporter while periplasmic lipoproteins avoid interaction with it.

In conclusion, the identification of a conserved signal sequence in surface exposed lipoproteins (lipoprotein export signal - LES) in Bacteroidetes strongly indicates the existence and the conservation of a dedicated transport mechanism. Accordingly, the finding of the Bacteroidetes LES will certainly benefit the identification of this novel lipoprotein transport machinery. Considering that this transport system could a least partially protrude into the outside environment, it could represent an interesting target for the developement of new antibiotics. Indeed, generation of antimicrobials specifically targeting Bacteroidetes could be of major interest for the treatment of anaerobic infections (*Bacteroides fragilis*) and peridontal diseases (*Porphyromonas gingivalis*) in humans as well as economical relevant poultry and fish pathogens such as *Riemerella anatipestifer* and *Flavobacterium columnare*.

1.2.4. Supplemental materials

Table E1. Plasmids used in extended results

Plasmid	Description	Reference
Expression plasmids		
pFL97	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7897 and 7896/7625 and cloned into pPM5 using NcoI/XhoI restriction sites. Replacement of aa 22-28 by RREVEEE	This study
pFL98	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7893 and 7892/7625 and cloned into pPM5 using NcoI/XhoI restriction sites. Replacement of aa 22-28 by RAEVEEE	This study
pFL99	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7895 and 7894/7625 and cloned into pPM5 using NcoI/XhoI restriction sites. Replacement of aa 22-28 by AREVEEE	This study
pFL140	Full length *siaC* amplified with primers 4159/8029 and 8028/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by AKEVE	This study
pFL141	Full length *siaC* amplified with primers 4159/8031 and 8030/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KAEVE	This study
pFL142	Full length *siaC* amplified with primers 4159/8082 and 8081/7696 and cloned into pMM47.A using NcoI/XhoI restriction sites. Replacement of aa 18-22 by KEVEE	This study

Table E2. Oligonucleotides used in extended results

Ref.	Sequence 5'-3'	Restriction[a]
4159	cataccatgggaaatcgaatttttatctt	NcoI
7182	ggccatggggaaaaaaatagtatccattagc	NcoI
7625	ggctcgagctaagcgtaatctggaacatcgtatgggtaaaacgtaacttgagttctc	XhoI
7696	ggctcgagttagttcttgataaattcctcaactgg	XhoI
7892	tggttagcctgtagagcggaagttgaagaagaacctttc	
7893	gaaaaggttcttcttcaacttccgctctacaggctaacca	
7894	ttagcctgtgcaagagaagttgaagaagaacctttc	
7895	gaaaaggttcttcttcaacttctcttgcacaggctaa	
7896	tggttagcctgtagaagagaagttgaagaagaacctttc	
7897	gaaaaggttcttcttcaacttctcttctacaggctaacca	
8028	gcaaaggaagttgaagtaatcggcggaggcgaatttacacaacccg	
8029	ttcaacttcctttgcacaagccgacaaaagaacaaaagcg	
8030	aaagcggaagttgaagtaatcggcggaggcgaatttacacaacccg	
8031	ttcaacttccgctttacaagccgacaaaagaacaaaagcg	
8081	aaggaagttgaagaagtaatcggcggaggcgaatttacacaacccg	
8082	ttcttcaacttccttacaagccgacaaaagaacaaaagcg	

[a]: Restriction sites are underlined

2. Identification of the export machinery of surface exposed lipoproteins in Bacteroidetes

2.1.Abstract

Bacteroidetes display many lipoproteins at their cell surface and to date, little is known on how these proteins reach the bacterial surface. In addition, since Bacteroidetes do not encode a LolB homolog, also how lipoproteins are inserted into the OM remains unknown.

Here, we address the question of how surface exposed lipoproteins reach the bacterial surface in the human pathogen *C. canimorsus.* Using LolA as bait protein to perform pull-down experiments, we identify three candidate proteins putatively involved in lipoprotein export (Ccan_02550, Ccan_09090 and Ccan_13690). Characterization of these candidates showed that all are involved to some extend in growth of *C. canimorsus*, *Ccan_02550* being critical for proliferation in liquid medium and *Ccan_09090* being an essential gene. Furthermore, we could show that absence of Ccan_02550 and Ccan_09090 affects OM protein abundance as well as LPS synthesis and/or transport, suggesting a potential chaperone activity of these proteins. However, their precise function and involvement in lipoprotein export, if any, remains to be clarified.

In parallel, we also made an educated guess approach based on the predicted characteristics of a putative surface lipoprotein transporter. We selected five highly conserved Bacteroidetes proteins, which, except one, all turned out to be essential in *C. canimorsus*. We could thus show for the first time that a BamA homolog possessing a lipid anchor (Ccan_17810) is essential in *C. canimorsus* and that its lipidation is crucial for its function. However, due to the lack of efficient genetic tools in *C. canimorsus* and the toxicity of Ccan_17810 in *E. coli*, we could not further investigate this candidate.

In conclusion, we could identify several essential genes of *C. canimorsus* linked to OM biogenesis and that could have a potential role in lipoprotein export. However, further analysis is required to precisely determine their function.

2.2. Identification and characterization of LolA interaction partners

2.2.1. Introduction

C. canimorsus encodes homologs to all previously described components of the Lol machinery, except for LolB. Therefore, while lipoprotein synthesis, maturation and IM release are assumed to be mostly similar to what has been observed in *E. coli*[1], nothing is known about how *C. canimorsus* lipoproteins are inserted into the OM after their transport across the periplasm nor how surface exposed lipoproteins are translocated across the OM. A likely hypothesis would be that a LolB-like protein inserts lipoproteins into the OM and that the same or another protein (or protein complex) then flips some of them to the cell surface. Such a predicted transporter would have several characteristics that would allow it to interact with lipoproteins. First, it is unknown whether or not surface lipoproteins are translocated in a folded or unfolded state; therefore, the transporter might have to interact with its substrates in a similar way than does BamA, thus involving protein-protein interaction domains such as POTRAs[2], TPRs (tetratricopeptide repeat)[3] or binding sites as found in chaperones[4-7]. Second, the lipid anchor of lipoproteins must be accommodated in some way during insertion into the OM and surface translocation; this could be achieved in a LolB-like fashion[1] or involve several proteins similar to LptD and -E[8], suggesting the presence of a hydrophobic cavity in the transporter. Third, in order to facilitate surface transport of lipoproteins across the OM, this transporter is likely a membrane protein itself, either completely or partially embedded into the OM, as BamA[9] and LptD[10]. Finally, since surface-exposed lipoproteins are a hallmark of Bacteroidetes, this implies that the lipoprotein transporter is conserved in the phylum.

Additionally, we assume that independently of their final localization, all OM and surface lipoproteins should cross the periplasm via LolA. In *E. coli*, LolA and LolB partially overlap during lipoprotein transfer[11]; this could also be the case for the putative OM lipoprotein transporter in Bacteroidetes. We thus

performed pull-down experiments using LolA as bait with the aim to find OM component(s) of the unknown lipoprotein export pathway.

2.2.2. Identification of LolA interaction partners

In order to perform a pull-down experiment using LolA as bait, we first needed to generate a *C. canimorsus* strain expressing a double tagged LolA-Strep-His protein and delete the chromosomal wt *lolA*. Since LolA is essential for cell viability, this was achieved by deleting the LolA encoding gene (*Ccan_16490*) in a strain expressing a plasmid-born copy of wt LolA, followed by plasmid exchange with the vector encoding the double-tagged LolA protein (for details, see methods section). We next proceeded to purify LolA-Strep-His and its interaction partners from *C. canimorsus* by Histidine-Streptavidine tandem affinity purification. Taking into account that the interaction between LolA and its OM partner(s) might be transient, we performed in parallel purifications in the presence or absence of two different crosslinking reagents, formaldehyde (Fig. 1A and D) and dithiobis succinimidyl propionate (DSP) (Fig. 1B and E). In addition, we also tested different lysis protocols, *i.e.* Triton X-100 (Fig. 1A and D) and French press (Fig. 1C and F), to evaluate the effect of detergents on the co-purified proteins. The quality of the elution fractions was assessed by Western blot and silver staining. The co-purified proteins were then identified by mass spectrometry and subsequently analyzed *in silico* (Table 1).

2.2.3. *In silico* characterization of LolA interaction partners

Independent of crosslinking and lysis method, the majority of identified peptides belonged to cytoplasmic proteins, ranging from 73.30 to 93.73 % in abundance (excluding LolA) (Table 1). Most of these proteins were predicted to be involved in protein synthesis or maturation, a fact that can be explained by the high expression level of the plasmid-borne LolA-Strep-His protein. Since cytoplasmic proteins are unlikely to play a role in lipoprotein surface export, this group of proteins was not further investigated.

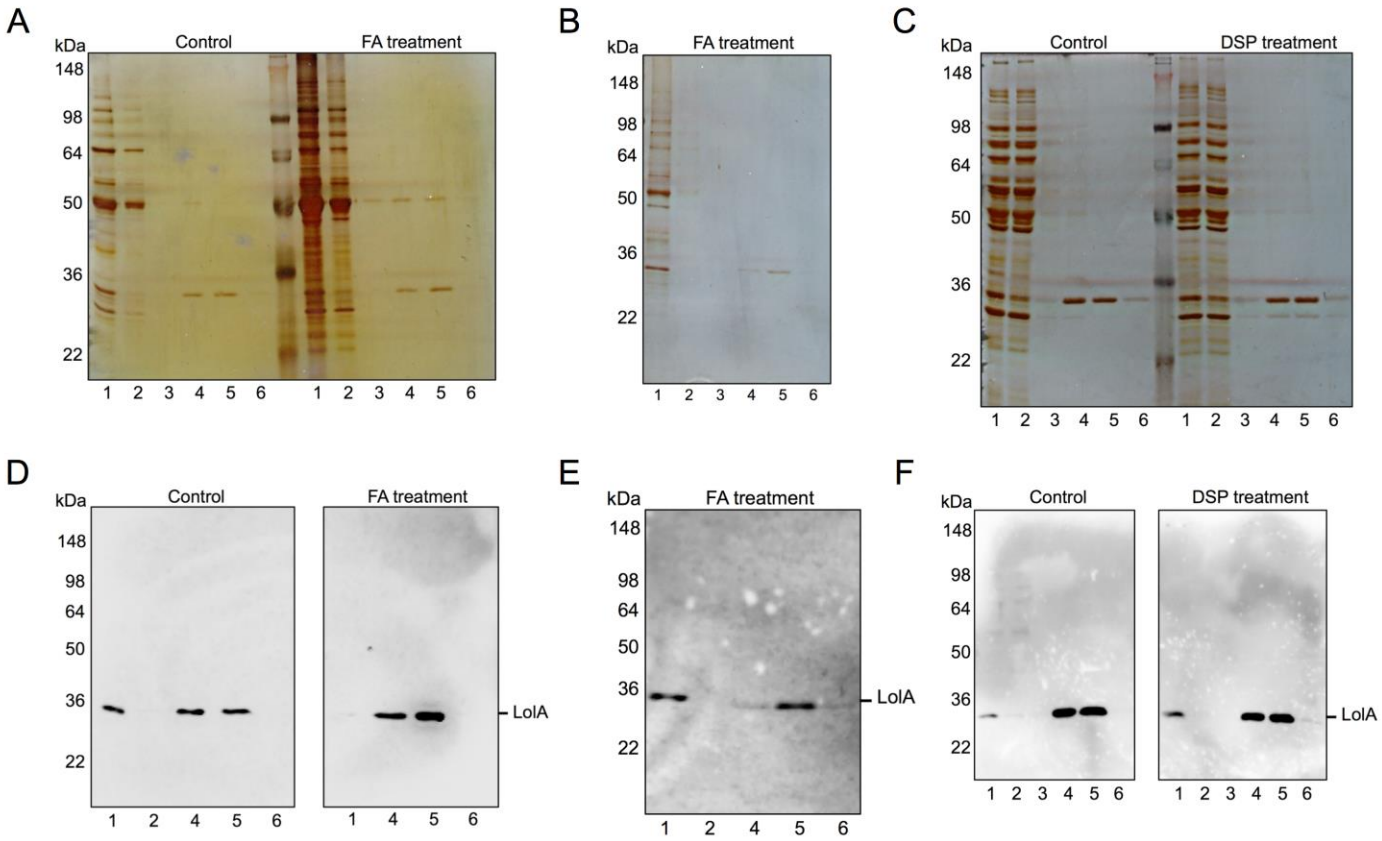


Fig. 1. LolA-Strep-His pull-down experiments

Analysis of LolA-Strep-His pull-down elution fractions by Western blot and silver staining. (A - C) Silver staining of elution fractions from formaldehyde (FA) (A and B) or DSP (C) crosslinked samples lyzed by addition of Triton X-100 (A and C) or by French press (B). (D - F) Detection of LolA by Western blot analysis in elution fractions from formaldehyde (D and E) or DSP (F) crosslinked samples lyzed by addition of Triton X-100 (D and F) or by French press (E). Proteins were detected using an anti-Strep antibody. Numbers throughout the figure refer to: 1. His column elution fraction; 2. Strep column flow through fraction; 3. Strep column wash fraction; 4. Strep column elution fraction 1; 5. Strep column elution fraction 2; 6. Strep column elution fraction 3.

The remaining peptides corresponded to proteins with either a signal peptide I (periplasmic and OM proteins), a signal peptide II (lipoproteins) or proteins with predicted TMHs (transmembrane helix) (IM proteins). Periplasmic and OM proteins were the most abundant species (3.18 to 20.56 %) followed by lipoproteins (1.09 to 4.20 %) and TMH-containing proteins (0.22 to 2.49 %) (Table 1). Overall, the type of crosslink reagent and the lysis method used did not significantly affect the amount or the type of proteins co-purified with LolA. Interestingly, and in good agreement with our starting hypothesis, both intracellular and surface exposed lipoproteins were crosslinked to LolA, suggesting that LolA indeed transports both subtypes of lipoproteins across the periplasm (Table 1). Surprisingly however, the amount of lipoproteins crosslinked to LolA was relatively low, a fact that could reflect the efficiency of transport of LolA and its rapid cargo delivery to its unknown OM partner(s). Alternatively, this could also reflect the growth condition in which the crosslinking was performed (*i.e*. bacteria that are not actively growing or do not require lipoproteins for their growth in this specific condition).

In order to find potential OM lipoproteins transporters, a first selection process was carried out based on the relative protein abundance and their annotation. This analysis showed that 7 proteins were particularly enriched in samples treated with a crosslinking reagent, representing between 2.49 and 31.44% of all non-cytoplasmic proteins (Table 2). These proteins were then further analyzed for domain conservation, structural prediction as well as taxonomic conservation. This resulted in a final list of 4 candidates, comprising Ccan_02550, Ccan_02920, Ccan_09090 and Ccan_13690 (Table 2).

Ccan_02920, annotated as an OmpA-like protein, has been previously analyzed in our lab and deletion of the gene encoding this protein did not result in any growth of morphological defects (F. Renzi, personal communication). As surface exposed lipoproteins are essential for the growth of *C. canimorsus* in the tested conditions, one would expect a strong phenotype upon removal of a protein involved in their localization. Hence, Ccan_02920 was not further investigated.

Ccan_02550 and Ccan_13690 are both annotated as TPR-containing proteins and are well conserved throughout the Bacteroidetes phylum.

Table 1. General statistics of LolA crosslink experiments

				Composition[b]				
Sample	**Crosslinker**	**Lysis**	**Total[a]**	**Cytoplasmic**	**TMH**	**SpI**	**SpII In[c]**	**SpII Out[c]**
Control 1	-	Triton X-100	64	73.44 (89.98)	3.13 (2.49)	20.31 (12.32)	1.56 (1.03)	1.56 (0.17)
Crosslink 1A	Formaldehyde	Triton X-100	160	61.25 (73.37)	3.75 (1.88)	23.13 (20.56)	5.63 (2.49)	6.25 (1.71)
Crosslink 1B	Formaldehyde	French press	134	75.37 (83.95)	0.75 (0.22)	16.42 (13.22)	3.73 (1.50)	3.73 (1.11)
Control 2	-	Triton X-100	118	86.44 (93.73)	4.24 (1.53)	6.78 (3.18)	1.69 (0.30)	0.85 (1.26)
Crosslink 2	DSP	Triton X-100	118	72.88 (89.32)	5.08 (1.04)	17.80 (8.54)	1.69 (0.19)	2.54 (0.90)

[a]: Total number of identified proteins

[b]: Relative number and relative abundance (combined peak area of corresponding peptides) of identified proteins expressed in percentage

[c]: Localization of lipoproteins based on *C. canimorsus* surface composition[12]. SpII In: proteins facing the periplasm; SpII Out: proteins facing the outside

Table 2. Annotation of potential candidates

Candidate	**Length (aa)**	**Signal peptide**	**Relative abundance[a]**	**Conservation[b]**	**Annotation[c]**
Ccan_02550	461	SpI	4.91 / 7.10 / 8.48	19/29	TPR containing protein, structural similarity to Tom70/Tom71
Ccan_02920	453	SpI	3.15 / 2.49 / ND	13/29	OmpA-like protein
Ccan_09090	177	SpI	25.24 / 25.97 / 31.44	29/29	Chaperone protein Skp, OmpH-like family
Ccan_13690	418	SpI	8.99 / 9.19 / 8.05	26/29	TPR-containing protein, structural similarity to Tom70/Tom71, DnaJ-like protein
Ccan_02630	219	SpII	0.81 / 1.30 / 1.22	18/29	YceI-like protein family
Ccan_15210	238	SpI	ND / 0.91 / ND	25/29	MotA/TolQ/ExbB proton channel family
Ccan_15800	537	SpI	6.75 / 7.11 / 7.48	9/29	DUF4139, DUF4140

Grey text corresponds to additional non-investigated potential candidates

[a]: Relative abundance (combined peak area of corresponding peptides) in crosslink samples 1A / 1B / 2 among non-cytoplasmic proteins. ND: not detected

[b]: Conservation of candidates among 29 Bacteroidetes reference genomes

[c]: Predicted annotation derived from Pfam, CDD, HHPred and Phyre2 analyses

TPR domains are known to mediate protein-protein interaction as well as multiprotein complex formation[3]. Furthermore, structural prediction indicates that both proteins could have a similar fold to Tom70/Tom71[13-15], two proteins involved in protein import across the outer membrane of mitochondria. Additionally, Ccan_13690 was identified in the OM fraction of *C. canimorsus* in a previous study[12], suggesting that it could be an integral membrane protein. These proteins represented interesting candidates and were further investigated (see below).

Ccan_09090, the most abundant candidate detected in all samples, is a homolog to the periplasmic *E. coli* Skp chaperone and is part of the OmpH family of proteins[4,12]. Structurally predicted to be very close to its *E. coli* homolog, Ccan_09090 (called Skp_{Cc} hereafter) has however been reported to be inserted into the OM of *C. canimorsus*[12], a fact that could match its involvement in lipoprotein export. Furthermore, Skp_{Cc} is part of a conserved operon encoding a second OmpH-like protein (Ccan_09080) as well the *C. canimorsus* BamA homolog Ccan_09070, hinting a potential role in OM biogenesis[16]. The Skp_{Cc} protein thus represents an interesting candidate and was further characterized (see below).

2.2.4. Generation of putative lipoprotein transporter deletion strains

In order to characterize the three selected LolA interaction partners Ccan_02550, Ccan_13690 and Ccan_09090 (Skp_{Cc}), we generated the corresponding deletion strains. While this was readily achieved for *Ccan_02550* and *Ccan_13690*, giving the Δ*Ccan_02550* and Δ*Ccan_13690* strains respectively, no deletion mutant could be obtained for skp_{Cc} suggesting that its function is essential for cell viability. We therefore introduced a plasmid encoding skp_{Cc} under the control of a *C. canimorsus* IPTG-inducible promoter (see chapter 3) into the wild type strain and then deleted the chromosomal copy of skp_{Cc}. This resulted in the generation of an Skp conditional mutant (cΔskp_{Cc}), dependent on the presence of IPTG for growth, and confirmed that Skp_{Cc} is essential in *C. canimorsus*. This result was somehow unexpected since in *E. coli*, absence of Skp

causes only moderate effects on growth and OMP levels, therefore suggesting that the chaperone networks in the two organisms are different.

2.2.5. Growth of putative lipoprotein transporter mutants

We next investigated the effect of these deletions on bacterial growth in heat inactivated human serum (HIHS) and on human embryonic kidney (HEK) cells, two conditions where PUL-encoded complexes are required for optimal growth[17-19]; mislocalization of PUL-encoded surface exposed lipoproteins would thus lead to a growth defect. As shown in Figure 2A and B, deletion of *Ccan_13690* had no effect when bacteria were grown on HEK cells but led to a 10-fold decrease of biomass in HIHS. When the growth of Δ*Ccan_02550* was tested, only few slow growing colonies were recovered from the plated inoculum and no colonies were recovered after incubation of the mutant in both the HIHS and HEK conditions (Fig. 2A and B). This suggested that although Ccan_02550 is not crucial for the growth on plates, its absence might compromise membrane integrity or permeability, resulting in a lethal phenotype in liquid medium. Alternatively, one could envision a decreased resistance to osmotic changes in the Δ*Ccan_02550* strain, which would inhibit growth in rich medium such as HIHS. Regarding Ccan_09090 (Skp_{Cc}), growth of cΔskp_{Cc} in non-permissive condition (without IPTG) had a dramatic effect in both HIHS and in the presence of HEK cells thus confirming its importance for cell viability. Addition of IPTG fully restored the growth on HEK cells and partially in HIHS (Fig. 2A and B), showing that the observed phenotype is indeed due to the depletion of Skp_{Cc}. We also monitored growth of cΔskp_{Cc} in permissive and non-permissive condition in 10% HIHS over time, and found that depletion of Skp_{Cc} results in early growth arrest after approximately 12h while wt bacteria grew till 24h (Fig. 2C). Yet, microscopic analysis revealed only minor morphological differences between cΔskp_{Cc} and wt bacteria after 24h (Fig. 2D), suggesting that depletion of Skp_{Cc} blocks growth altogether rather than leading to growth or division defects.

We then tested whether addition of GlcNAc, a known substrate of PUL-encoded complexes that has been shown to be critical for *C. canimorsus* growth[17,18], would rescue the growth impairment of the different mutants (Fig.

2A and B). However, addition of this compound had no effect, indicating that the observed phenotypes are not solely due to GlcNAc starvation.

Taken together, these results indicate that all three LolA interaction partners are involved in *C. canimorsus* growth in liquid medium. Strikingly, deletion of *Ccan_02550* completely abolished growth in liquid medium while depletion of *skp*$_{Cc}$, found to be an essential gene, led to early growth arrest in non-permissive conditions.

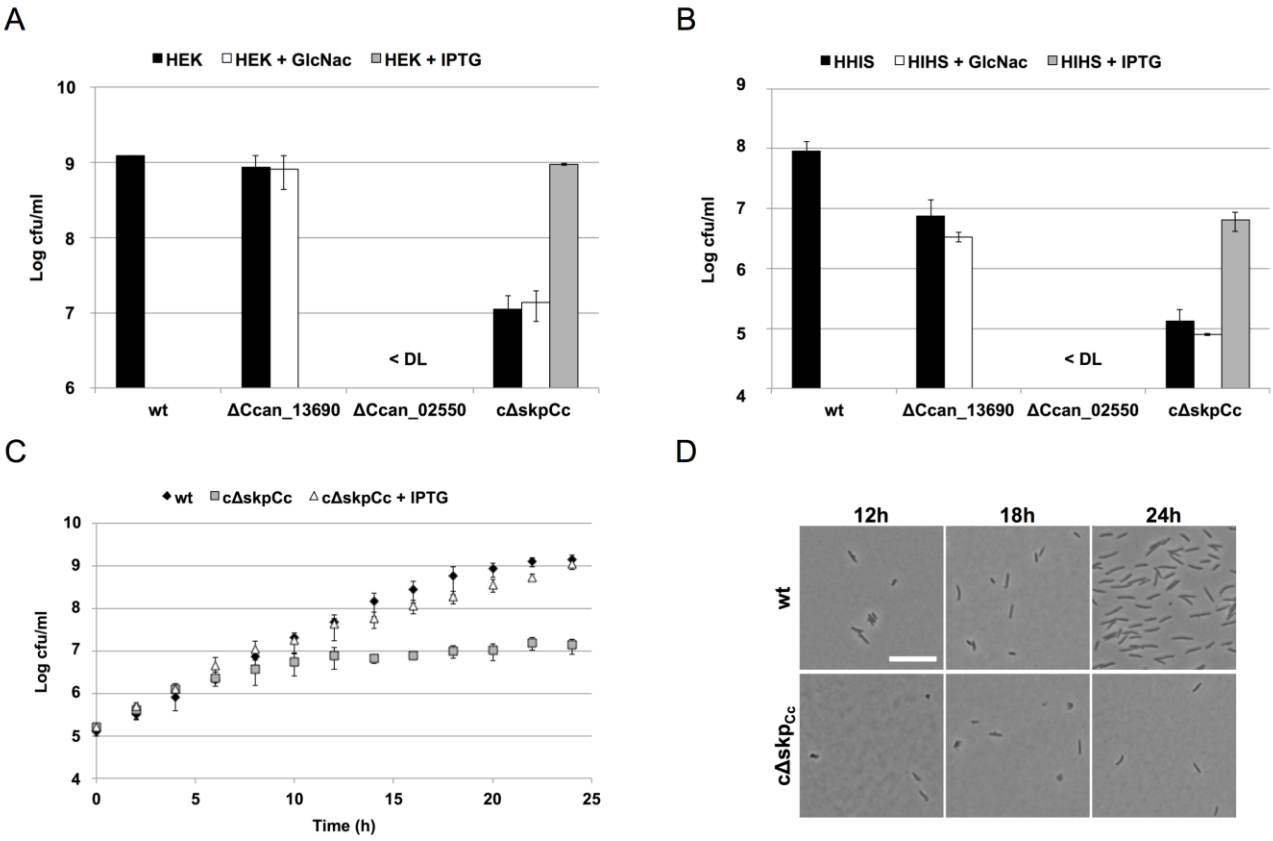


Fig. 2. Growth phenotypes of putative lipoprotein transporter mutants
(A) Counts of wt, Δ*Ccan_13690,* Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ bacteria after 23 hours of growth on HEK293 cells (MOI, 0.05) with or without GlcNAc (0.01%) or IPTG (0.5 mM). The averages from three independent experiments are shown. Error bars represent 1 standard deviation from the mean. < DL: below detection limit. (B) Counts of wt, Δ*Ccan_13690,* Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ bacteria after 23 hours of growth in HIHS with or without GlcNAc (0.01%) or IPTG (0.5 mM). The averages from three independent experiments are shown. Error bars represent 1 standard deviation from the mean. < DL: below detection limit. (C) Growth curve of wt and cΔ*skp*$_{Cc}$ bacteria grown in DMEM containing 10% HIHS with or without IPTG (0.5 mM). The averages from three independent experiments are shown. Error bars represent 1 standard deviation from the mean. (D) Bright-field microscopy pictures of wt and cΔ*skp*$_{Cc}$ bacteria grown as in (C) for 12, 18 and 24 hours. Scale bar: 5 μm.

2.2.6. OMP composition and LPS profile of putative lipoprotein transporter mutants

Since lipoprotein maturation and/or localization defects can have an impact on OMP assembly[20] (Bam complex) and LPS transport[21] (Lpt complex), we analyzed the OMP composition and the LPS profile of all deletion strains. Following isolation of the OM fraction of each strain (Fig. 3A), we compared the OM protein composition and the relative protein abundance of the deletion strains to the wt by Western blot and silver staining. While no major difference could be observed in the OM composition of the Δ*Ccan_13690* strain compared to the wt, the deletion of *Ccan_02550* and depletion of Skp_{Cc} resulted in an overall decrease of OMP levels (Fig. 3B and C). However, very few proteins seemed to be directly affected by the deletion of these proteins as no band clearly disappeared or was shifted in size due to proteolytic cleavage. We thus hypothesized that both Ccan_02550 and Skp_{Cc} might work as general chaperones involved in lipoprotein or OMP biogenesis, similarly to Skp and SurA in *E. coli*, and that deletion of these factors would have a global effect on the bulk mass of OMPs rather than only on surface lipoproteins [22].

We next analyzed the LPS profile of the mutant strains. In *C. canimorsus* 5, two independent LPS have been identified: a major LPS (migrating at approximately 20 kDa, Fig. 3D band C) and a second, smaller structure (migrating at approximately 15 kDa, Fig. 3D band D)[23]. Similarly to the OMP composition, deletion of *Ccan_13690* had no impact on either LPS form as compared to the wt (Fig. 3D). The same was observed for the Δ*Ccan_02550* and cΔskp_{Cc} mutants. Interestingly however, these mutants showed the presence of an additional band in respect to the wt (Fig. 3D band X) that could represent a different LPS form. Noteworthy, the *Ccan_02550* gene is located immediately upstream of the *C. canimorsus* LptC homolog Ccan_02560, which could indicate its implication in LPS synthesis or transport. Alternatively, the deletion of *Ccan_02550* could also have a polar effect on *Ccan_02560*, resulting in accumulation of LPS at the IM and leading to partial LPS degradation. While this later hypothesis remains to be clarified, the inability of the Δ*Ccan_02550* strain to grow in liquid medium might therefore be explained by LPS assembly or transport defects resulting in compromised OM permeability.

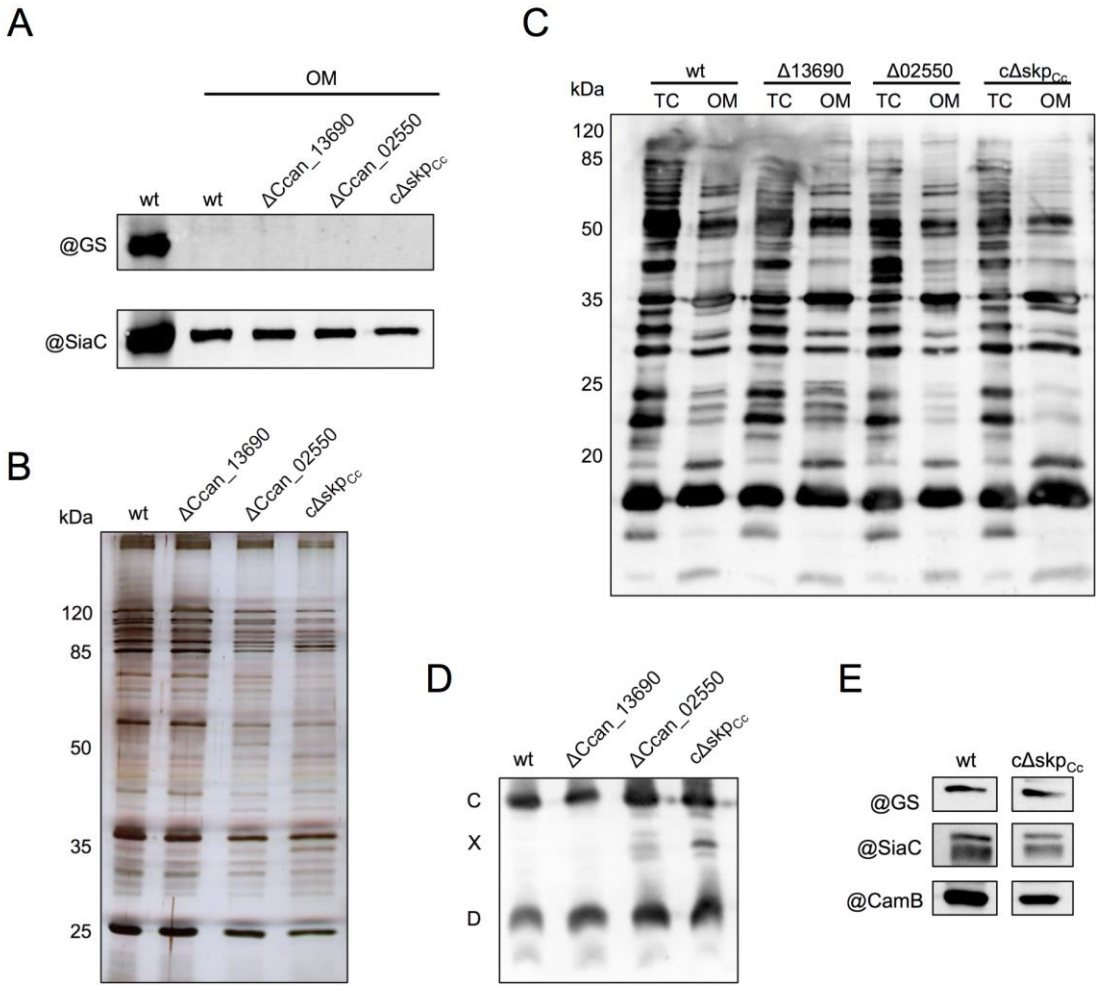


Fig. 3. Outer membrane composition and LPS pattern of putative lipoprotein transporter mutants

(A) Western blot analysis of OM fractions of wt, Δ*Ccan_13690,* Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ bacteria. Glutamine synthase (GS) serves as cytoplasmic control, SiaC serves as OM control. A total cell extract of wt bacteria was loaded in lane 1 as reference. (B) Silver staining of OM fractions analyzed in (A). (C) Western blot analysis of total cell extracts (TC) and OM fractions of wt, Δ*Ccan_13690,* Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ bacteria. Anti-*C. canimorsus* 5 antiserum was used for detection. (D) Western blot analysis of LPS preparations of wt, Δ*Ccan_13690,* Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ bacteria. Bands corresponding to the previously described LPS structures are indicated (band C and D). The additional band (band X), presumably LPS, detected in Δ*Ccan_02550* and cΔ*skp*$_{Cc}$ strains is also shown. Anti-*C. canimorsus* 5 antiserum was used for detection. (E) Quantification of lipoprotein abundance by Western blot analysis of total cell extracts of wt and cΔ*skp*$_{Cc}$ bacteria grown for 23h in HIHS. Glutamine synthase (GS) serves as loading control, SiaC is used to estimate OM lipoprotein abundance, CamB is used to estimate surface lipoprotein abundance.

Since deletion of *Ccan_13690* had no visible effect on OMP composition or LPS pattern and *Ccan_02550* seemed to be related to LPS rather than to lipoprotein transport, we decided to focus on *skp*$_{Cc}$ and monitored the abundance of several OM lipoproteins after growth in non-permissive condition of this strain. Western blot analysis showed that, as observed for the OMP profile, the levels of the periplasmic OM lipoprotein SiaC and the surface exposed lipoprotein CamB (belonging to PUL11) were decreased by 2-fold in the cΔ*skp*$_{Cc}$ strain but were nevertheless present even after 23h of depletion (Fig. 3E). This suggested that Skp$_{Cc}$ might be involved in a more general way in OM biogenesis rather than be specifically dedicated to lipoprotein localization.

Taken together, these results, although indicating that *Ccan_02550* and *skp*$_{Cc}$ could be involved in OM biogenesis, did not allow a precise definition of their exact function. It is therefore not clear if the observed phenotypes result directly from impaired lipoprotein localization or if it is the indirect result of altered OMP and/or LPS biogenesis.

2.2.7. Localization of Skp$_{Cc}$ and identification of its interaction partners

In order to clarify this, we decided to perform pull-down experiments using Skp$_{Cc}$ as bait as previously done for LolA. We generated a strain expressing a plasmid born copy of Skp$_{Cc}$-Strep-His, followed by the deletion of the chromosomal wt copy (Fig. 4A). We first addressed the question of Skp$_{Cc}$ localization by performing cell fractionation, using CamB and SiaC as outer membrane markers and SiaC$_{C17G}$ as soluble periplasmic marker. This analysis showed that, similarly to *E. coli*, Skp$_{Cc}$ was majorly located in the soluble fraction of the cell lysate, indicating that the protein is periplasmic rather than membrane anchored (Fig. 4B). This is in contradiction with previously obtained data that showed that Skp$_{Cc}$-derived peptides were released from intact cells when treated with trypsin[12]. However, this could be explained by leakage of periplasmic content into the medium during cell surface shaving. While the evidence that Skp$_{Cc}$ is soluble indicated that it is likely not the OM lipoprotein

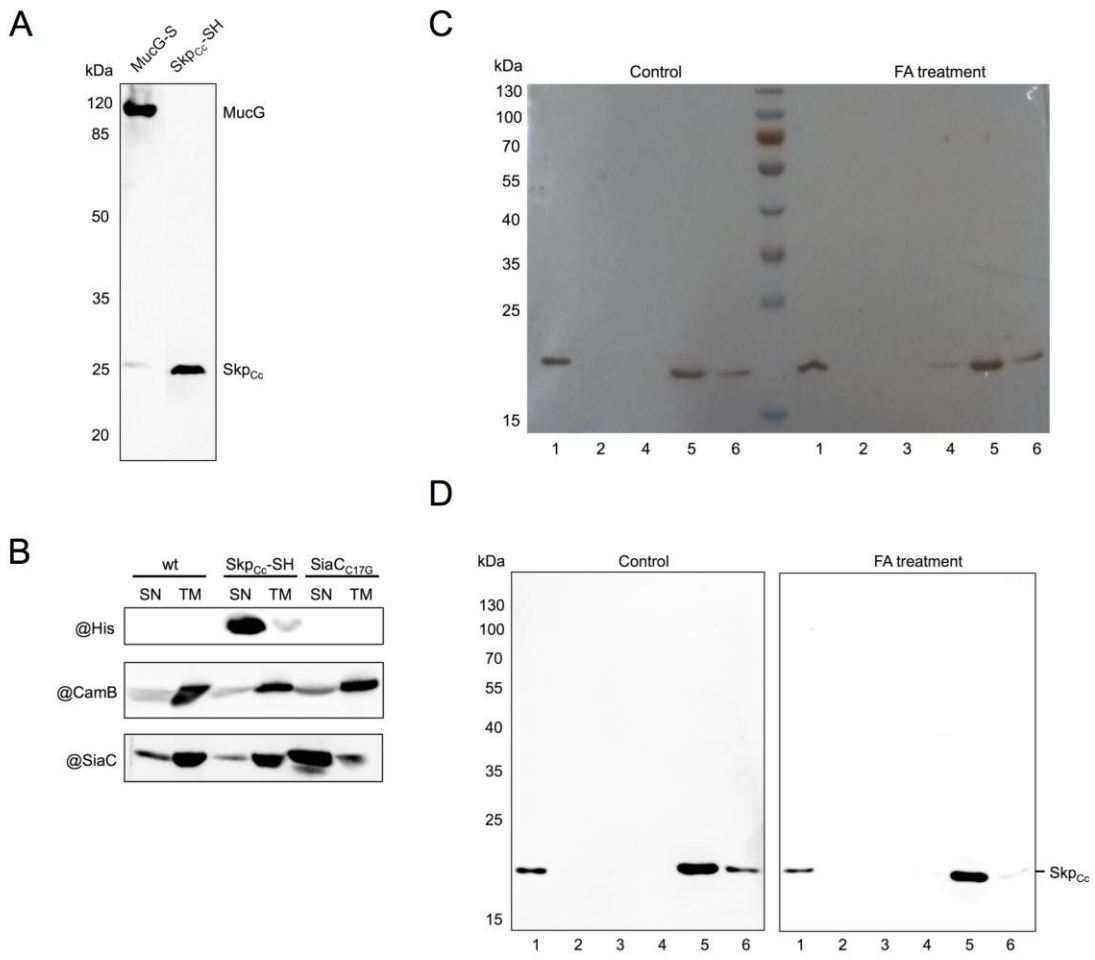


Fig. 4. Skp$_{Cc}$-Strep-His localization and pull-down experiments
(A) Detection of Skp$_{Cc}$-Strep-His by Western blot analysis of total cell extracts. The protein was detected using an anti-Strep antibody. MucG-Strep serves as positive control. (B) Western blot analysis of supernatant (SN) and total membrane (TM) fractions of wt, Skp$_{Cc}$-Strep-His and SiaC$_{C17G}$ bacteria. SiaC$_{C17G}$ is a soluble periplasmic variant of SiaC, an OM anchored lipoprotein. An anti-Strep antibody was used to detect of Skp$_{Cc}$-Strep-His. SiaC serves as OM lipoprotein control, CamB serves as surface lipoprotein control. (C) Silver staining of elution fractions from formaldehyde crosslinked samples (FA treatment) lyzed by addition of Triton X-100. (D) Detection of Skp$_{Cc}$-Strep-His by Western blot analysis of samples shown in (D). Proteins were detected using an anti-Strep antibody. Numbers in panel (C) and (D) refer to: 1. His column elution fraction; 2. Strep column flow through fraction; 3. Strep column wash fraction; 4. Strep column elution fraction 1; 5. Strep column elution fraction 2; 6. Strep column elution fraction 3.

transporter itself, we did not exclude the possibility that Skp$_{Cc}$ might still be involved in some way in lipoprotein surface exposure, for example by keeping lipoproteins in a transport compatible state or assisting the folding of the transporter. We therefore proceeded to purify Skp$_{Cc}$-Strep-His and its interaction partners by tandem affinity purification followed by identification by mass spectrometry (Fig. 4C and D). The *in silico* analysis was performed as described

before and led to the selection of 4 potential candidates (Table 3 and 4). Among them, 3 corresponded to previously identified LolA interaction partners, namely Ccan_02550, Ccan_02920 and Ccan_13690 (Table 4). The last candidate, Ccan_18290 is a predicted β-barrel protein homologous to *E. coli* FadL[24] and *Pseudomonas putida* TodX[25], a long-chain fatty acid and a toluene transporter respectively (Table 4). This type of transporter allows the entry of hydrophobic molecules and their insertion into the OM through lateral gate opening of the β-barrel[24]. While these proteins are normally involved in import of substrates, the overall transport mechanism could well fit a putative lipoprotein transporter. However, deletion of this protein did not lead to any growth defect in HIHS (data not shown), thus excluding this protein as potential lipoprotein transporter.

Interestingly, among the identified Skp_{Cc} putative interaction partners, we detected two *C. canimorsus* homologs of the Bam machinery, namely Ccan_09070 (BamA) and Ccan_09420 (BamD) (Table 4). Strikingly, Ccan_09420 was the most abundant protein among all non-cytoplasmic proteins, representing in total 10.63 %. This suggested either that Skp_{Cc} specifically interacts with the *C. canimorsus* Bam machinery, presumably delivering polypeptide chains to the complex, and that BamD is its preferred interaction partner, or that Skp_{Cc} is the main chaperone of BamD and assures its proper folding.

Finally, to our surprise, LolA was not crosslinked to Skp_{Cc}. The most likely hypothesis is that since Skp_{Cc} seems to be a periplasmic chaperone, overexpression of LolA-Strep-His during the first pull-down experiment induced periplasmic stress and/or crowding, leading to recruitment of Skp to prevent LolA-Strep-His misfolding or aggregation. We thus assume that the observed LolA-Skp_{Cc} complex would result from a non-specific interaction.

Table 3. General statistics of Skp_{Cc} crosslink experiments

				Composition[b]				
Sample	**Crosslinker**	**Lysis**	**Total[a]**	**Cytoplasmic**	**TMH**	**SpI**	**SpII In[c]**	**SpII Out[c]**
Control	-	Triton X-100	9	44.44 (19.09)	0.00 (0.00)	44.44 (53.40)	11.11 (27.49)	0.00 (0.00)
Crosslink	Formaldehyde	Triton X-100	156	35.90 (29.77)	2.56 (1.10)	46.15 (38.95)	11.54 (23.66)	3.85 (6.49)

[a]: Total number of identified proteins

[b]: Relative number and relative abundance (combined peak area of corresponding peptides) of identified expressed in percentage

[c]: Localization of lipoproteins based on *C. canimorsus* surface composition [12]. SpII In: proteins facing the periplasm; SpII Out: proteins facing the outside

Table 4. Annotation of potential candidates

Candidate	**Length (aa)**	**Signal peptide**	**Relative abundance[a]**	**Conservation[b]**	**Annotation[c]**
Ccan_02550	461	SpI	2.25	19/29	TPR containing protein, structural similarity to Tom70/Tom71
Ccan_02920	453	SpI	5.97	13/29	OmpA-like protein
Ccan_13690	418	SpI	3.88	26/29	TPR-containing protein, structural similarity to Tom70/Tom71, DnaJ-like protein
Ccan_18290	503	SpI	7.04	29/29	FadL-like protein, TodX-like protein
Ccan_01510	538	SpII	2.07	23/29	Peptidase_S8 family, Serine protease, Subtilisin-like
Ccan_09070	845	SpI	0.34	29/29	Omp85/BamA family
Ccan_09420	286	SpII	10.63	29/29	BamD

Grey text corresponds to additional non-investigated potential candidates

[a]: Relative abundance (combined peak area of corresponding peptides) in crosslink sample among, non-cytoplasmic proteins

[b]: Conservation of candidates among 29 Bacteroidetes reference genomes

[c]: Predicted annotation derived from Pfam, CDD, HHPred and Phyre2 analyses

2.2.8. Discussion

In conclusion, we could show by pull-down experiments that LolA interacts with both periplasmic OM lipoproteins as well as surface exposed lipoproteins, indicating that both subsets of proteins are processed the same way until they reach the OM.

Our data also suggest that LolA interacts to some extend with Ccan_13690, Ccan_02550 and Skp_{Cc}, although the observed interaction might be due to LolA overexpression. Further experiments indicated that Ccan_02550 and Skp_{Cc} could be involved in OM biogenesis, however their precise role remains to be clarified. Our results suggest that Ccan_02550 might take part in LPS synthesis and/or transport or in the folding of another component related to these pathways. Skp_{Cc} was found to be an essential gene and to interact with the Bam machinery, especially with BamD, prompting its role in OMP assembly. While both proteins still require further investigation, this therefore indicates that their involvement in lipoprotein surface exposure is either minor or indirectly linked through OMP assembly and LPS transport. In this regard, it is interesting to note that a recent study showed that an Skp homolog in *Porphyromonas gingivalis* is involved in the maturation and processing of T9SS substrates[26]. Indeed, deletion of this protein resulted in decreased T9SS substrate activity as well as decreased virulence of the mutant strain. While a role of Skp_{Cc} in T9SS cannot be ruled out, its homolog in *Porphyromonas gingivalis* was however not found to be essential. This could indicate that Skp has different functions in these two organisms.

In order to identify the putative lipoprotein transporter, several options can be investigated. For instance, the conditions in which the crosslink experiments were conducted are not optimal to detect transient interactions. Indeed, while *C. canimorsus* grows readily on plates, growth in liquid medium is much more fastidious. It is therefore difficult to grow large numbers of cells to exponential phase in this condition. As a result, the delay between the recovery of the bacteria from plates and the crosslinking reaction is probably too long to accurately fix transient interactions. Furthermore, a significant proportion of bacteria grown on plate could be in stationary phase, thus not actively growing, a

condition where lipoprotein export might be limited. An alternative would therefore consist in applying our pull-down approach using as model organism another Bacteroidetes species such as *Flavobacterium johnsoniae* that grows readily in defined media. Finally, use of alternative crosslinking reagents that do not cross the IM, such as DTSSP, would reduce the amount of cytoplasmic contaminants and hence increase the probability to specifically crosslink OM LolA interaction partners.

Another option would be based on our identification of the lipoprotein export signal (LES). Indeed, one could generate a bait protein or peptide harboring the LES, overexpress it to saturate the transporter and then perform pull-down experiments followed by mass spectrometry. Using a peptide-based system would have the additional benefit of limiting non-specific interactions with other cellular proteins. A third option would consist in fusing the LES to an easily detectable reporter that would thus be surface localized. A transposon library could then be generated and screened for absence or decreased amount of the reporter at the cell surface, followed by mapping of the genetic regions involved. This approach could also be used on known surface exposed lipoproteins such as MucG, although the screening process would be more fastidious. Considering the fact that the lipoprotein transporter might be essential, one could also perform a Tn-seq analysis focusing on the genetic regions without transposon insertions, likely essential genes, to identify potential candidates.

2.3. An educated guess approach to the surface lipoprotein transporter

2.3.1. Preface

In order to identify the putative lipoprotein export machinery and in parallel to the work focusing on LolA, we also made an educated guess approach. Based on the previously defined characteristics of such a transporter (see section 2.2.1), *i.e.* a highly conserved Bacteroidetes OM protein with protein-protein interaction domains and/or presence of a hydrophobic cavity, we selected five *C. canimorsus* candidate proteins.

2.3.2. Selected lipoprotein export machinery candidates

According to these parameters we selected the proteins encoded by genes: *Ccan_09070*, *Ccan_17810*, *Ccan_20230*, *Ccan_16770* and *Ccan_06900* (Table 1).

Ccan_09070, Ccan_17810 and Ccan_20230 belong to the omp85/TpsB protein superfamily[27-29], which can be divided into the omp85 (e.g. BamA) and TpsB (e.g. FhaC, two-partner secretion system) protein families (Fig. 1)[28]. While structurally similar (presence of N-terminal POTRA domains and of a C-terminal β-barrel), there is a clear separation between the two protein families at the sequence level and in respect to the number of POTRA domains they harbor (Fig. 1)[28]. The omp85 family can be further divided into nine subfamilies based on domain architecture. Two domain architectures referred to as TamA and "Lipo" (termed LipoBamA hereafter) stand out as being the only ones showing a taxonomic distribution indicating vertical inheritance rather than horizontal gene transfer (Fig. 1), suggesting a phylum specific function[28]. TamA (translocation and assembly module) has been described recently and was shown to be indispensable for membrane insertion of some autotransporter proteins[30-33]. Interestingly, TamA is mostly restricted to Proteobacteria[28] and is linked to virulence in these organisms[31,34-36]. On the other hand, LipoBamA is of yet unknown function but shows a clear taxonomic restriction to the Chlorobi

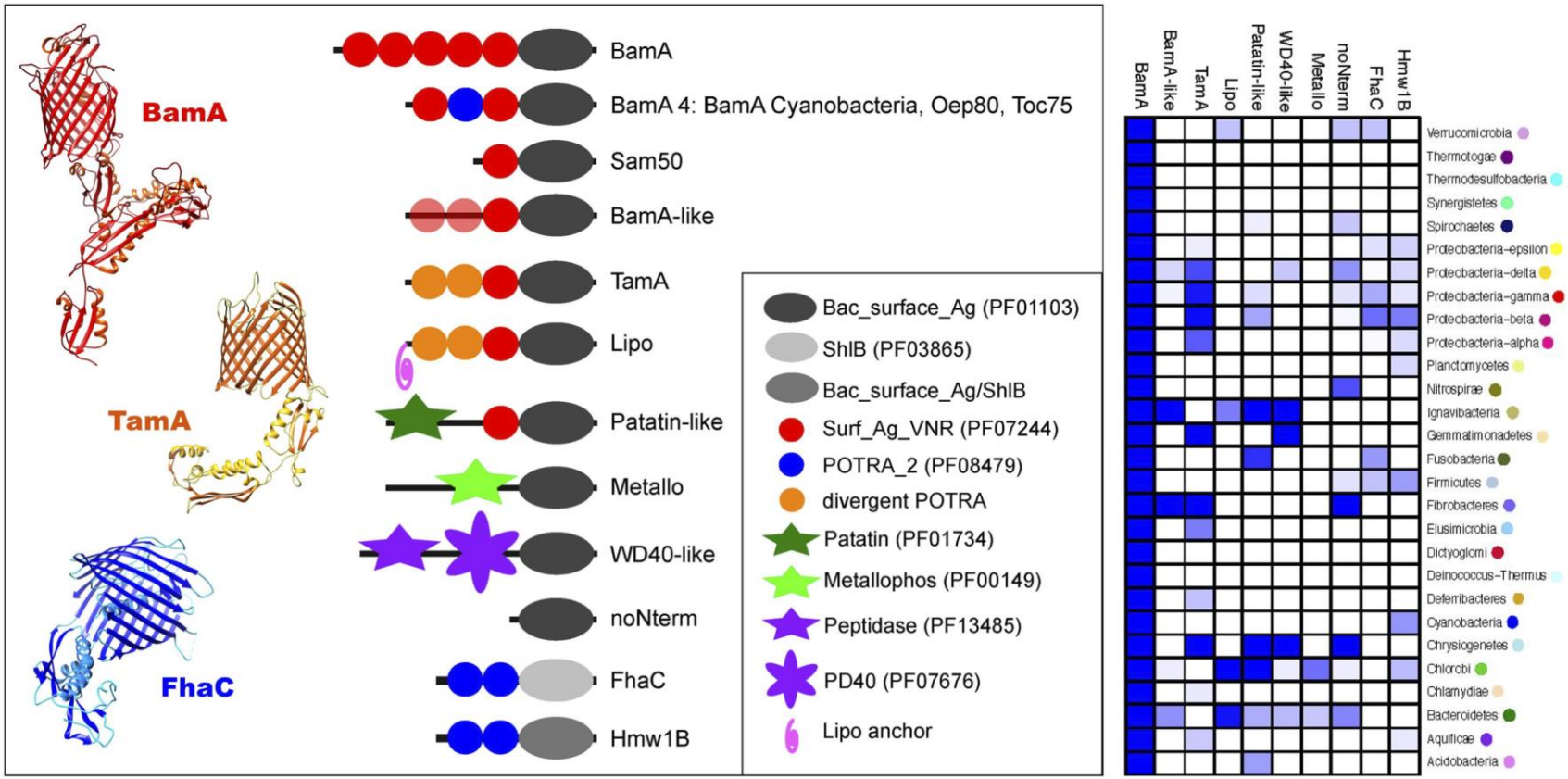


Fig. 1. Structural diversity of the omp85/TpsB protein superfamily

Schematic representation of the eleven subfamilies of the omp85/TpsB superfamily in Bacteria. The mitochondrial Sam50 is indicated for comparison. Crystal structures of BamA (PDB 4K3B), TamA (PDB 4C00) and FhaC (PDB 2QDZ) are indicated on the left, taxonomic distribution of each protein family is indicating on the right. Adapted from [28].

and Bacteroidetes phyla (Fig. 1)[28]. From a structural point of view, TamA and LipoBamA are very similar to each other and differ from the classical BamA protein family essentially by the number of N-terminal POTRA domains they carry (three rather than five) as well as the presence of a lipid anchor for LipoBamA (Fig. 1).

Based on our *in silico* analysis, Ccan_09070 is the closest *C. canimorsus* homolog of BamA and likely performs the same function as the *E. coli* protein (Table 1 and Fig. 2).

Ccan_17810 shares structural similarities with both *E. coli* BamA and TamA (Table 1). However, structural prediction also indicates that Ccan_17810 harbors only three POTRA domains (Fig. 2). Additionally, Ccan_17810 is a predicted lipoprotein, suggesting that it is part of the LipoBamA subfamily. Given the taxonomic distribution of proteins with this type of domain architecture (Fig. 1), this is of particular interest in respect to the putative lipoprotein export machinery.

Ccan_20230 showed structural homology to *Haemophilus ducreyi* BamA and *E. coli* TamA (Table 1). DeltaBlast analysis also suggests some remote homology to *E. coli* FhaC. TamA and FhaC can be easily discriminated based on their number of POTRA domains (three for TamA, two for FhaC) as well as their β-barrel Pfam profile[28]. However no precise prediction could be obtained for either of these criteria likely due to Ccan_20230 sequence degeneration. Hence, it is not clear whether this protein belongs to the TamA or to the FhaC family.

Ccan_16770 is a homolog of *E. coli* LptD (Table 1), having an OstA-like domain in its N-terminus and a β-barrel in its C-terminus (Fig. 2). However, unlike *E. coli* LptD, Ccan_16770 is a predicted lipoprotein. Similarly to Ccan_17810, this means that its N-terminus is anchored either to the IM or the OM, therefore limiting its flexibility.

Finally, Ccan_06900 is a so far uncharacterized protein with only one annotated OstA-like domain in its N-terminus (Table 1) (Fig. 2). This is of great interest because in *C. canimorsus* we could not identify any homolog of LptA, the periplasmic protein bridging the IM and OM complexes allowing LPS transport to the cell surface in *E. coli*, by a classical Blast analysis. As already mentioned,

Table 1. Homology and structural prediction of potential lipoprotein transporters

	DeltaBlast against *E. coli* K12			HHpred			Phyre2			
ORF name	**Homolog**	**Coverage[a]**	**E value**	**Homolog**	**Species[b]**	**E value**	**Homolog**	**Species[b]**	**Confidence[c]**	**Conservation[d]**
Ccan_09070	BamA	100	0	BamA	Ec	9.80E-72	BamA	Ng	100	29/29
	TamA	99	2.00E-74	BamA	Ng	1.00E-77	BamA	Ec	100	
				TamA	Ec	1.30E-52	TamA	Ec	100	
Ccan_17810	BamA	87	3.00E-143	BamA	Ec	1.00E-55	BamA	Ng	100	28/29
	TamA	79	4.00E-34	BamA	Ng	1.90E-55	BamA	Ec	100	
				TamA	Ec	2.40E-49	TamA	Ec	100	
Ccan_20230	BamA	63	2.00E-29	BamA	Hd	4.80E-27	BamA	Ng	100	10/29
	FhaC	52	1.69E-03	TamA	Ec	5.90E-26	BamA	Ec	100	
	TamA	50	8.00E-02	BamA	Ec	1.70E-25	TamA	Ec	100	
Ccan_06900	LptA	22	3.00E-33	LptD N-ter	Sf	7.70E-15	LptD N-ter	Sf	99.6	28/29
	LptD	20	1.80E-02	LptA	Pa	1.30E-13	LptH	Pa	99.5	
				LptA	Ec	7.10E-11	LptA	Ec	99.5	
Ccan_16770	LptD	52	4.00E-112	LptD	Sf	3.40E-70	LptD	Sf	100	29/29
				LptD	Pa	4.60E-56	LptD	Pa	100	
				LptD	Yp	6.20E-54	LptD	Yp	100	

[a]: query coverage in percentage

[b]: *Ec: E. coli; Ng: Neisseria gonorrhoeae; Hd: Haemophilus ducreyi; Sf: Shigella flexneri; Pa: Pseudomonas aeruginosa; Yp: Yersinia pestis*

[c]: model confidence in percentage

[d]: Conservation of candidates among 29 Bacteroidetes reference genomes

LptA harbors an N-terminal OstA domain that is responsible for LPS binding. It could thus be that Ccan_06900 plays a role in LPS transport in *C. canimorsus*, maybe overtaking the function of LptA. Noteworthy, this protein is predicted to be much bigger than LptA, being 552 amino acids in length rather than 185. This could indicate that while Ccan_06900 might fulfill the LptA function, it could also have additional ones. Alternatively, this difference in size could also be the result of an adaptive process whereby Ccan_06900 evolved a long C-terminal domain in order to interact with the membrane anchored N-terminus of Ccan_16770 (the LptD homolog) in a different manner than what is observed between LptA and LptD in *E. coli*, while still allowing transfer of LPS.

Except for Ccan_20230, all of the above-described candidates are highly conserved throughout the Bacteroidetes phylum, making them targets of choice for the putative OM lipoproteins transporter (Table 1).

2.3.3. Generation of candidate mutants and their characterization

We first attempted to generate the corresponding deletion strain of each candidate gene. With the exception of *Ccan_20230*, all genes turned out to be essential. We therefore proceeded as for *Ccan_09090* (*skp*$_{Cc}$) by transforming the wt strain with a plasmid encoding each gene of interest downstream of an IPTG inducible promoter prior to deletion of the chromosomal copy. We thus generated the conditional mutants cΔ*Ccan_09070,* cΔ*Ccan_17810,* cΔ*Ccan_06900 and* cΔ*Ccan_16770*. Next, we tested the growth of the *Ccan_20230* deletion mutant and the four conditional mutants in HIHS in permissive and non-permissive conditions. Surprisingly, none of the conditional mutants had a growth defect in absence of IPTG (Fig. 3). Since all genes with the exception of *Ccan_20230* appeared to be essential, we concluded that the IPTG-regulatable expression system used was leaky and that basal expression was sufficient to sustain normal growth. This is also in agreement with the relatively low expression level of these genes as compared to *skp*$_{Cc}$ (4 to 5 times lower) as determined by mRNA sequencing (K. Hack, unpublished data).

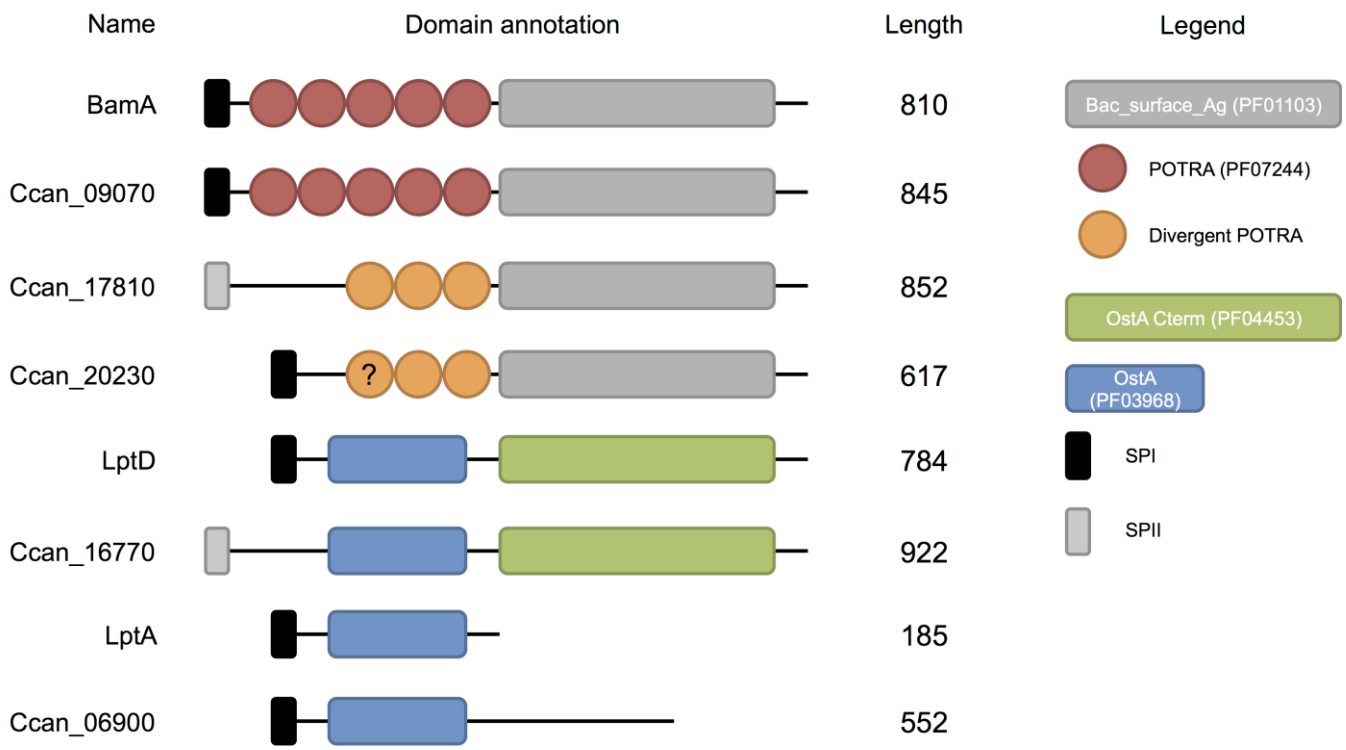


Fig. 2 Domain annotation of potential lipoprotein transporters

Schematic representation of the domain architecture of potential *C. canimorsus* lipoprotein surface transporters. *E. coli* proteins BamA, LptD and LptA serve as references. The length of each protein is indicated in number of amino acids. Identified domains and their corresponding Pfam references are indicated in the legend. SPI: signal peptide I; SPII: signal peptide II.

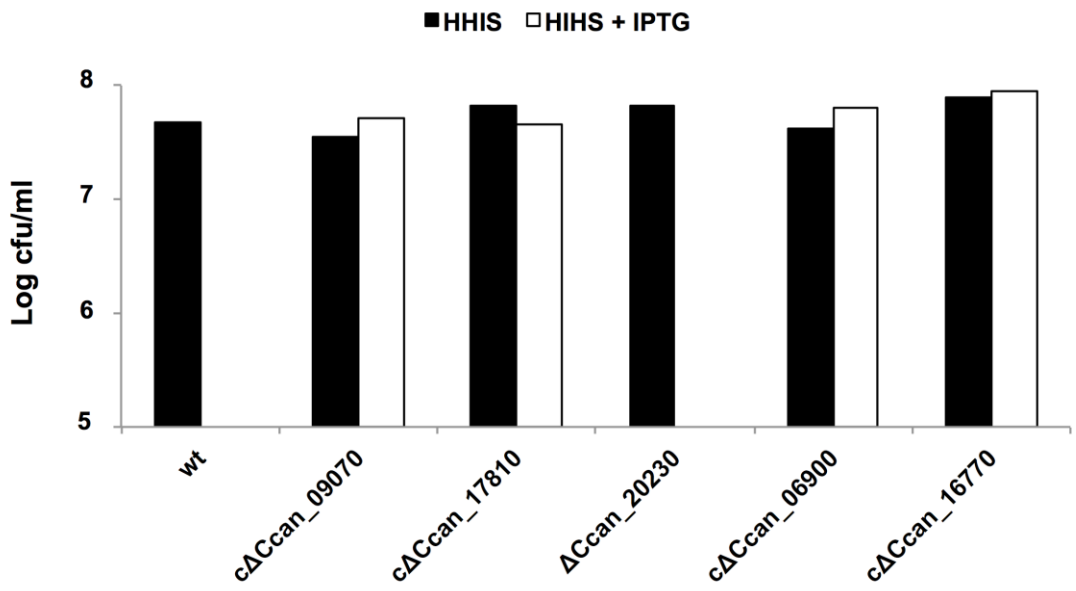


Fig. 3. Growth in heat inactivated human serum of lipoprotein surface transporter candidate mutants
Counts of wt, cΔ*Ccan_09070,* cΔ*Ccan_17810,* Δ*Ccan_20230,* cΔ*Ccan_06900* and cΔ*Ccan_16770* bacteria after 23 hours growth in HIHS with or without IPTG (0.5 mM).

As to date no other inducible promoter is available for *C. canimorsus* and attempts to build a tetracycline based regulated system failed (see chapter 3), we then decided to try to identify the interaction partners of these candidates. We selected protein Ccan_17810 (belonging to the LipoBamA subfamily) as it seemed to be the best candidate for the surface lipoprotein transporter. Indeed, Ccan_06900 and Ccan_16770 are likely involved in LPS transport rather than lipoprotein localization, Ccan_09070 is most likely the homolog of BamA and Ccan_20230, while being an interesting candidate, is not essential in a growth condition where surface lipoproteins are required, indicating that it is not involved in their transport. However, it is not immediately clear why *C. canimorsus* would encode an additional essential copy of a BamA homolog. We therefore hypothesized that Ccan_17810 might have a novel, unrelated function to OMP assembly, namely lipoprotein surface transport.

We started the analysis of the *C. canimorsus* LipoBamA Ccan_17810 by first investigating the role of the lipid anchor. We generated a variant of

Ccan_17810 in which the predicted site of lipidation was mutated (Ccan_17810$_{C20G}$) and expressed it in the wt strain. We then tried to delete the chromosomal copy of Ccan_17810 in this strain but no mutant was obtained, indicating that the anchorage of the N-terminus of Ccan_17810 to the inner or outer membrane has an important biological function (Fig. 4). Presumably, if, unlike BamA, this protein works independently of other partners[9], anchoring its N-terminus to the OM could allow more efficient substrate interaction and/or membrane targeting.

We next generated strains expressing Ccan_17810 in fusion with a Strep and His tag either at its N- or C-terminus in order to perform pull-down experiments and identify possible interaction partners. However, when cell lysates of these strains were tested, no tagged Ccan_17810 proteins nor their degradation products could be observed (data not shown). This indicated that either the tagged proteins are not properly folded and thus undergo complete degradation or that the presence of the tags prevents the biological function of Ccan_17810, leading to their proteolysis.

In conclusion, we could show genetically that Ccan_09070, Ccan_17810, Ccan_16770 and Ccan_06900 are essential in *C. canimorsus*. However, since so far no functional conditional mutants could be obtained, mainly because of the lack of efficient genetic tools in *C. canimorsus*, the role of these proteins could not be further investigated. We found that the lipid anchor of Ccan_17810 is probably critical for its biological function, which represents a novelty in respect to other BamA-like proteins. Unfortunately, no strain expressing a tagged Ccan_17810 derivative could be generated to date, prompting us to produce an antibody against this protein in order to perform co-immunoprecipitation assays to identify Ccan_17810 interaction partners.

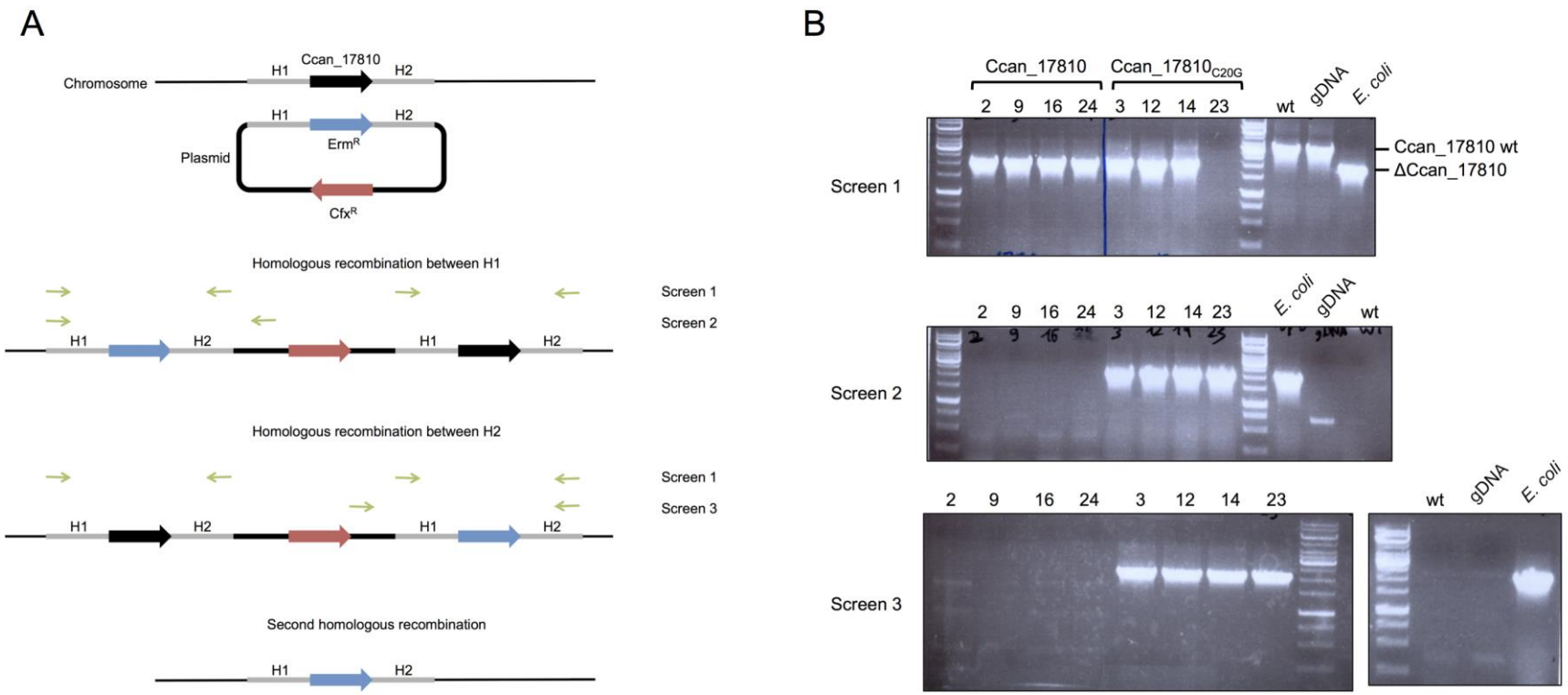


Fig. 4. Chromosomal deletion of *Ccan_17810*

(A) Schematic representation of the possible recombination events leading to chromosomal deletion of *Ccan_17810*. The gene replacement is achieved in a two-step process by homologous recombination. First, the plasmid is integrated into the chromosome by recombination between the homologous regions H1 located on the plasmid and on the chromosome. Alternatively, the recombination takes place between the H2 regions. The second recombination then leads to removal of the backbone of the plasmid, the *Ccan_17810* gene and its replacement by the Erm^R resistance cassette. The gene of interest (*Ccan_17810*) is represented by a black arrow, erythromycin and cefoxitin resistance cassettes by blue (Erm^R) and red (Cfx^R) arrows respectively. Homologous regions on the chromosome and the suicide plasmid are indicated by grey boxes. For sake of simplicity, the plasmid borne copy of Ccan_17810 is not displayed. The PCR strategy used to screen for second recombination events (removal of chromosomal *Ccan_17810*) is indicated by green arrows. Depending on which recombination event takes place, screens 1 and 2 (H1 recombination) or screens 1 and 3 (H2 recombination) are used to determine successful deletion of *Ccan_17810*. (B) PCR screening of 4 individual clones recovered following conjugation of *C. canimorsus* expressing either Ccan_17810 or $Ccan\_17810_{C20G}$ and *E. coli* harboring the suicide plasmid for *Ccan_17810* chromosomal deletion. Presence of the Erm^R cassette is indicated by a band of 1800 bp while the wt gene corresponds to 2700 bp. Similarly, presence of the plasmid backbone in the chromosome is indicated by a band of 2,000 bp while no amplification is seen if the second recombination event occurred. *C. canimorsus* wt cells and genomic DNA (gDNA) serve as positive controls for screen 1, *E. coli* harboring the suicide plasmid serve as positive control for screens 2 and 3.

2.3.4. Expression analysis of Ccan_17810 in *E. coli*

Since the study of Ccan_17810 in *C. canimorsus* was technically limited, we decided to express the protein in *E. coli* in order to better understand its function. Previous work in our group had shown that expression of *C. canimorsus* surface exposed lipoproteins in *E. coli* does not lead to their surface localization (F. Renzi, personal communication). We therefore reasoned that if Ccan_17810 is indeed the surface lipoprotein transporter, expressing it together with a known substrate, *i.e.* a *C. canimorsus* surface exposed lipoprotein, should lead to surface localization of this protein in *E. coli*.

We first cloned *Ccan_17810* downstream of an arabinose inducible promoter in the low copy plasmid pBAD33 and transferred it into *E. coli* MG1655 cells, giving the strain *E. coli* pFL116. We then monitored the growth of this strain over time in the presence or absence of arabinose. As shown in Fig. 5A, induction of Ccan_17810 synthesis led to rapid growth arrest of *E. coli* cells until exhaustion of arabinose, at which point growth resumed normally. Alternatively, when Ccan_17810 expression was started at the time of inoculation, it led to a long lag phase (Fig. 5B). Microscopy analysis revealed that addition of arabinose induced formation of "ghost" cells in part of the population that could explain the apparent growth arrest (Fig. 5C).

We reasoned that expression of Ccan_17810 in *E. coli* could be detrimental because it could cause misfolding or mislocalization of native *E. coli* proteins. We thus hypothesized that in the presence of a putative substrate the expression of Ccan_17810 would be tolerated by *E. coli* cells. We therefore cloned *mucG* (the mucinase of PUL9) downstream of an arabinose inducible promoter in the high copy vector pBAD24 and transferred this construct in *E. coli* pFL116 (expressing Ccan_17810). The growth of the resulting *E. coli* pFL116-pFL114 strain was then assessed with and without induction. However, even in the presence of the putative substrate MucG, expression of Ccan_17810 led to rapid growth arrest and appearance of ghost cells (Fig. 5D). Monitoring MucG expression after induction showed that growth arrest was likely not linked to MucG protein synthesis blockage or induction of proteolysis (Fig. 5E and F).

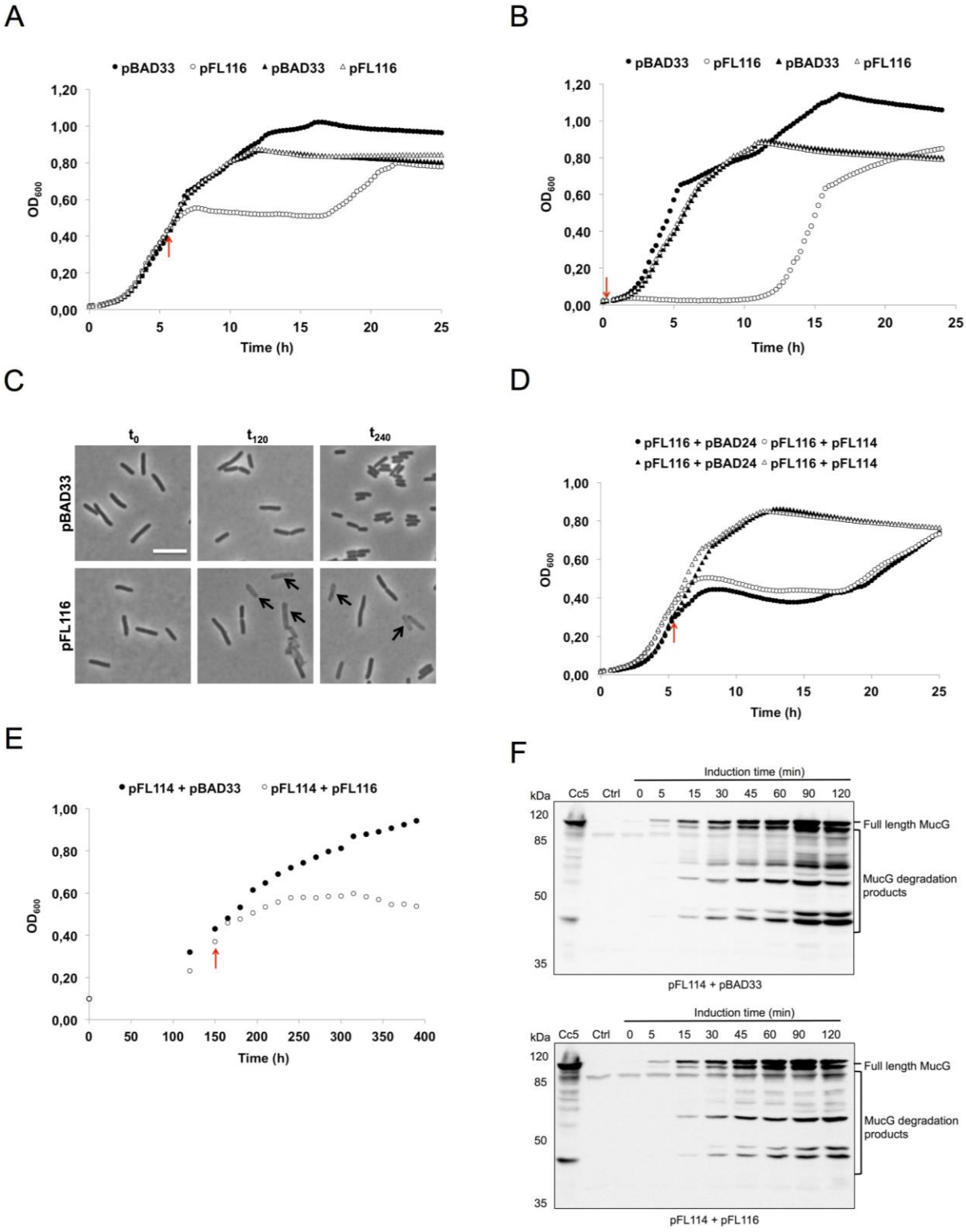


Fig. 5. Expression of Ccan_17810 in *E. coli* leads to growth arrest and ghost cells formation
(A) Growth curve of *E. coli* MG1655 in cM63 medium. Once exponential phase was reached (around OD=0.4), 0.2% arabinose (red arrow) was added (round shapes) or not (triangular shapes) to the culture. pFL116 refers to *E. coli* expressing Ccan_17810, pBAD33 refers to the control strain harboring the empty vector. (B) Growth curve of *E. coli* MG1655 in cM63 medium. Bacteria were inoculated (red arrow) in cM63 medium (triangular shapes) or in cM63 medium

containing 0.2% arabinose (round shapes). Strains are the same as in (A). (C) Bright-field microscopy pictures of *E. coli* pFL116 and *E. coli* pBAD33 grown in cM63 medium following induction with 0.2% arabinose Ghost cells are indicated by a black arrow. Scale bar: 5 µm. (D) Growth curve of *E. coli* MG1655 in cM63 medium. Once exponential phase was reached, 0.2% arabinose (red arrow) was added (round shapes) or not (triangular shapes) to the medium. pFL116 + pFL114 refers to *E. coli* expressing Ccan_17810 and MucG, pFL116 + pBAD24 refers to *E. coli* expressing only Ccan_17810. (E) Growth curve of *E. coli* MG1655 in cM63 medium. Once exponential phase was reached, 0.2% arabinose (red arrow) was added to the medium. pFL114 + pFL116 refers to *E. coli* expressing MucG and Ccan_17810 (white spheres), pFL114 + pBAD33 refers to *E. coli* expressing only MucG (black spheres). (F) Western blot analysis of samples from panel (E). MucG expression is followed over time upon arabinose induction in the presence or absence of Ccan_17810. A representative experiment is shown for each panel.

Taken together, our results indicate that expression of Ccan_17810 is toxic for *E. coli* and induces growth arrest. The exact reasons for this phenotype have not yet been addressed and require further investigations. For example, one could monitor the activation of different periplasmic stress pathways upon Ccan_17810 expression, such as σ^E, Cpx or Rcs. Depending on which response is triggered, this could help to better understand the role of Ccan_17810 in *C. canimorsus*.

2.3.5. Discussion

Here, we tried to identify the lipoprotein export machinery of *C. canimorsus* by investigating several candidates based on an educated guess. Through a genetic approach, we showed that Ccan_09070 and Ccan_16770, orthologs of BamA and LptD respectively, are essential in *C. canimorsus*, which was somewhat expected. The same was true for Ccan_06900, a putative distant homolog of LptA and presumably involved in LPS transport. Interestingly, we found that a second BamA homolog, Ccan_17810, is essential in *C. canimorsus*, and we thus tried to identify its function. However, despite several attempts (including genetics and protein engineering) and different approaches (expression of Ccan_17810 in *E. coli*), we could not elucidate the precise role of this protein. The next step should therefore consist in generating antibodies against different fragments of Ccan_17810. This would allow us to monitor its expression and to perform pull-down assays in the wt genetic background, thereby avoiding all possible effects of protein overexpression. Additionally,

since the C-terminus of Ccan_17810 is predicted to form a β-barrel, extracellular loops might be exposed at the cell surface. Incubation of *C. canimorsus* cells with anti-Ccan_17810 antibody could thus inhibit the function of the protein by binding to these loops, mimicking a depletion strain. One could then assess growth and OM composition over time, thereby helping to define the role of Ccan_17810. Finally, a recent study showed that RNA silencing takes place in Bacteroidetes and that it can be used for gene regulation[35]. This could be used as alternative method to generate a Ccan_17810 conditional mutant.

2.3.6. Materials and Methods

Bacterial strains and growth conditions

(i) Conventional bacterial growth conditions and selective agents

Bacterial strains used in this study are listed in Table S1. *Escherichia coli* strains were routinely grown in lysogeny broth (LB) at 37°C. For growth curves and induction tests, *E. coli* strains were grown in complete M63 medium (2 g/l $(NH_4)_2SO_4$, 13.6 g/l KH_2PO_4, 0.5 mg/l $FeSO_4$ $7H_2O$) containing 1 mM $MgSO_4$, 0.2% (w/v) glycerol and 0.1 % (w/v) casamino acids (cM63). Glucose and arabinose were added separately at 0.2 % (w/v) final concentration where indicated. *C. canimorsus* strains were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) plates (SB plates) for 2 days at 37°C in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 100 µg/ml ampicillin (Amp), 50 µg/ml kanamycin (Km), 10 µg/ml chloramphenicol (Cm) for *E. coli* and 10 µg/ml erythromycin (Em), 10 µg/ml cefoxitin (Cfx), 20 µg/ml gentamicin (Gm), 10 µg/ml tetracycline (Tc) for *C. canimorsus*.

(ii) *E. coli* MG1655 derivatives growth curve in cM63 medium

E. coli MG1655 containing pFL114 and pFL116 were grown overnight in 5 ml cM63 containing 0.2 % glucose at 37°C. Cultures were then diluted to an OD_{600} of 0.05 with fresh medium containing 0.2 % glucose or arabinose, dispensed into 200 µl aliquots in 96-well plates and incubated at 37°C. Growth was assessed by

measuring the absorbance at 600 nm using an xMark microplate spectrophotometer (Bio-Rad) and Microplate Manager 6 software (version 6.0; Bio-Rad) over time. Alternatively, cultures were diluted with fresh medium containing 0.2 % glycerol and dispensed into 96-well plates. Once exponential phase was reached (OD_{600} of approximately 0.4), 0.2 % arabinose was added and incubation was continued. *E. coli* MG1655 containing empty vectors pBAD24 and pBAD33 served as control strains. All conditions were tested in triplicate.

(iii) *E. coli* MG1655 derivatives induction in cM63 medium

E. coli MG1655 containing pFL114 and pFL116 were grown overnight in 5 ml cM63 containing 0.2 % glucose. Cultures were then diluted to an OD_{600} of 0.1 with fresh medium containing 0.2 % glucose and incubated at 37°C. Once exponential phase was reached (OD_{600} of approximately 0.4), a sample corresponding to 1 ml of an OD_{600} of 0.5 was taken (negative control). The remaining culture was then washed once with cM63 medium before being resuspended to its initial volume in cM63 medium containing 0.2 % arabinose. Growth was then monitored over time and samples corresponding to 1 ml of an OD_{600} of 0.5 were taken. All samples were centrifuged and resuspended in 50 µl SDS PAGE buffer (1% SDS, 10% glycerol, 50 mM dithiothreitol, 0.02% bromophenol blue, 45 mM Tris, pH 6.8) for Western blot analysis. In parallel, 500 µl samples were taken for microscopy analysis, washed once with PBS and fixed with 0.8% PFA. All microscopy images were captured with an Axioscop (Zeiss) microscope with an Orca-Flash 4.0 camera (Hamamatsu) and Zen 2012 software (Zeiss).

(iv) End point growth of *C. canimorsus* in heat-inactivated human serum (HIHS)

Growth assays were performed in 96-well plates. Inocula were prepared from cultures grown on SB plates, set to an OD_{600} of 0.2, and serially diluted 1:10 four times. Twenty µl of bacterial suspension (around 2 x 10^2 bacteria) were then used to inoculate 180 µl of HIHS (S1-Liter, Millipore). HIHS was supplemented with IPTG (0.5 mM final concentration) or GlcNAc (A8625; Sigma; 0.01% final concentration) where indicated. Cultures were then incubated statically for 23 h

at 37°C in the presence of 5% CO_2. Serial dilutions were plated on SB plates (containing 0.5 mM IPTG if required), and CFU were determined.

(v) Growth curve of *C. canimorsus* in DMEM 10% HIHS

Growth curves were performed in 24-well plates. Inocula were prepared from cultures grown on SB plates, set to an OD_{600} of 1, and serially diluted 1:10 two times. Two hundred µl of bacterial suspension (around 1 x 10^6 bacteria) were then used to inoculate 10 ml of Dulbecco's modified Eagle's medium (DMEM; Invitrogen) containing 10% (v/v) HIHS and dispensed into 1 ml aliquots. Cultures were then incubated statically at 37°C in the presence of 5% CO_2. At different time points, one aliquot of medium was collected and serially diluted. Dilutions were plated on SB plates (containing 0.5 mM IPTG if required), and CFU were determined.

(vi) Growth of *C. canimorsus* bacteria with HEK293 cells

HEK293 cells were cultured in DMEM containing 10% (v/v) fetal calf serum (Invitrogen) and 1 mM sodium pyruvate. Cells were grown in medium without antibiotics in a humidified atmosphere enriched with 5% CO_2 at 37°C. A total of 1 x 10^4 bacteria were incubated with 2 x 10^5 HEK293 cells (multiplicity of infection (MOI) of 0.05) in a final volume of 1 ml. Cultures were supplemented with IPTG (0.5 mM final concentration) or GlcNAc (0.01% final concentration) where indicated. Cultures were incubated statically for 23 h at 37°C in the presence of 5% CO_2. Serial dilutions were plated on SB plates (containing 0.5 mM IPTG if required) and CFU were determined.

(vii) End point growth of *C. canimorsus* for lipoprotein quantification

Cultures for Western blot analysis of the wt and cΔ*skp*$_{Cc}$ conditional mutant were performed in 24-well plates. Inocula were prepared from cultures grown on SB plates and set to an OD_{600} of 0.02. Two µl of wt suspension were then used to inoculate 2 ml of DMEM containing 10% (v/v) HIHS and dispensed into 1 ml aliquots. One hundred µl of cΔ*skp*$_{Cc}$ suspension were used to inoculate 5 ml of DMEM containing 10% HIHS and dispensed into 1 ml aliquots. Cultures were incubated statically for 23h at 37°C in the presence of 5% CO_2. Serial dilutions

were plated on SB plates (containing 0.5 mM IPTG if required) and CFU were determined. In parallel, bacteria were collected, washed once with PBS and pellets stored at -20°C. Following CFU count, pellets were resuspended in SDS-PAGE buffer. Equivalent amounts corresponding to approximately 5 x 10^7 bacteria were loaded on 12% SDS PAGE gels and analyzed by Western blot.

Genetic manipulation of *C. canimorsus* and *E. coli* MG1655

Plasmids and primers used in this study are listed in Table S2 and S3.

(i) Gene deletion in *C. canimorsus*

Mutagenesis of *C. canimorsus* was performed as described previously with slight modifications[19]. Briefly, replacement cassettes with flanking regions spanning approximately 500 bp homologous to regions directly framing the genes to be deleted were constructed with a three-fragment overlapping PCR strategy. First, two Q5 High-Fidelity DNA Polymerase (M0491S; New England Biolabs) PCRs were performed on 100 ng of *C. canimorsus* genomic DNA with primers for the upstream (oligonucleotides 1.1 and 1.2) and downstream (oligonucleotides 2.1 and 2.2) regions flanking the sequence targeted for deletion. Primers 1.2 and 2.1 included a 20-bp extension at their 5' extremities corresponding to both ends of the *ermF* gene (including the promoter). The *ermF* resistance gene and its promoter were amplified from pMM106 with primers 3.1 and 3.2, which included approximately 20-bp extensions for further annealing to amplify homologous regions. All three products were cleaned and then mixed in equal amounts for PCR using Q5 High-Fidelity DNA Polymerase. The initial denaturation was at 98°C for 2 min, followed by 10 cycles without primers to allow annealing and extension of the overlapping fragments (98°C for 30 s, 52°C for 30 s, and 72°C for 60 s). After the addition of external primers (1.1 and 2.2), the program was continued for 25 cycles (98°C for 30 s, 52°C for 30 s, and 72°C for 60 s) and finally for 2 min at 72°C. Final PCR products consisted of locus::*ermF* insertion cassettes and were digested with PstI and SpeI restriction enzymes for cloning into the corresponding sites of the *C. canimorsus* suicide vector pMM25[37]. The resulting plasmids were transferred by RP4-mediated conjugation from *E. coli* S17-1 to *C. canimorsus* to allow the integration of the

insertion cassette. Transconjugants were selected for the resistance to erythromycin and subsequently checked for sensitivity to cefoxitin.

(ii) Construction of LolA expression vector

The *Flavobacterium johnsoniae ompA* promoter was amplified from pPM5 DNA using 7201 and 7202 and cloned into pMM104.A using BamHI and XbaI restriction sites, leading to plasmid pFL62. *lolA* (*Ccan_16490*) was amplified from *C. canimorsus* genomic DNA using 7203 and 7204 and cloned into pFL62 using NcoI and XbaI restriction sites, leading to plasmid pFL63.

(iii) Construction of LolA-Strep-His and Skp_{Cc}-Strep-His expressing strains

To engineer LolA with a C-terminal Strep and His tag, *lolA* (*Ccan_16490*) was amplified from *C. canimorsus* genomic DNA using 7203 and 7205 and cloned into pPM5 using NcoI and XbaI restriction sites in frame with the 6 x His sequence in the vector, leading to plasmid pFL64. Skp_{Cc} was engineered the same way, except that 7284 and 7302 were used for amplification, leading to plasmid pFL67.

To generate the mutant expressing LolA-Strep-His, the wt strain was first transformed with pFL63 and tetracycline resistant colonies were selected. The resulting complemented strain was then used for conjugation with *E. coli* S17 harboring the *lolA* mutator plasmid. Erythromycin and tetracycline resistant colonies were selected and checked for sensitivity to cefoxitin. Removal of the chromosomal *lolA* copy was further verified by PCR. This strain was then transformed with pFL64 and cefoxitin resistant colonies were selected. Colonies were then sub-cultured twice in DMEM containing 10 % (v/v) fetal calf serum and cefoxitin for 23h. Serial dilutions were plated on SB plates containing cefoxitin and colonies were checked for sensitivity to tetracycline. Expression of LolA-Strep-His was confirmed by Western blot using a mouse-HRP anti-Strep antibody (MCA2489P, AbD serotec).

To generate the mutant expressing Skp_{Cc}-Strep-His, the wt strain was first transformed with pFL67 and cefoxitin resistant colonies were selected. The resulting complemented strain was then used for conjugation with *E. coli* S17 harboring the *skp_{Cc}* mutator plasmid. Erythromycin resistant colonies were selected and removal of the chromosomal *skp_{Cc}* copy was verified by PCR.

Expression of Skp_{Cc}-Strep-His was confirmed by Western blot using a mouse-HRP anti-Strep antibody (MCA2489P, AbD serotec).

(iv) Generation of *Ccan_06900*, *Ccan_09070*, *Ccan_09090*, *Ccan_16770* and *Ccan_17810* conditional mutants

The genes of interest were amplified from *C. canimorsus* genomic DNA and cloned into pFL32 using NcoI and XbaI restriction sites for *Ccan_06900*, *Ccan_09070*, *Ccan_09090* and *Ccan_16770* or NcoI and XhoI restriction sites for *Ccan_17810*. The wt strain was first transformed with each expression plasmid and cefoxitin resistant colonies were selected. The resulting complemented strains were then used for conjugation with *E. coli* S17 harboring the corresponding mutator plasmid on IPTG containing medium. Erythromycin resistant colonies were selected and checked by PCR for removal of the chromosomal copy of each gene.

(v) Construction of *E. coli* MG1655 derivatives expressing *C. canimorsus* MucG and Ccan_17810

mucG (*Ccan_17430*) was amplified from *C. canimorsus* genomic DNA using 7182 and 6925 and cloned into pBAD24 using NcoI and XbaI restriction sites, leading to plasmid pFL114. *Ccan_17810* was amplified from *C. canimorsus* genomic DNA using 7900 and 7901 and cloned into pBAD33 using KpnI and SphI restriction sites, leading to plasmid pFL116. pBAD33 having no Shine Dalgarno (SD) sequence, the SD box from pBAD24 was included into primer 7900. Plasmids were then transferred to *E. coli* MG1655 by electroporation.

SDS PAGE, Western blotting and silver staining

Bacteria grown for 2 days on SB plates were collected, washed once with PBS, and resuspended in 1 ml PBS at an OD_{600} of 1, corresponding to approximately 5 x 10^8 bacteria. The cΔ*skp_{Cc}* conditional mutant was grown for 2 days on SB plates containing IPTG, bacteria were then seeded on SB plates without IPTG and grown for 1 day before being collected. Bacteria were centrifuged for 3 min at 5,000 g and resuspended in 100 µl SDS PAGE buffer. Samples were heated for 5 min at 96°C and 5 µl were loaded on 12% SDS PAGE gels. After gel

electrophoresis, proteins were transferred onto nitrocellulose membrane (1060008; GE Healthcare) and analyzed by Western blot using rabbit antisera as primary antibodies and swine-HRP anti-rabbit (P0217; Dako) as secondary antibody. Proteins were detected using LumiGLO (54-61-00; KPL) according to manufacturer's instructions. Alternatively, gels were analyzed by silver staining following electrophoresis. Briefly, gels were fixed for 2 h in 50% methanol, 12% acetic acid and 0.05% formaldehyde. Gels were then washed three times for 20 min in 35% ethanol, sensitized for 2 min with 0.02% sodium thiosulfate and washed three times for 5 min in ddH_2O. Gels were then stained with 0.2% silver nitrate, washed twice for 1 min in ddH_2O and finally developed in 6% sodium carbonate. The developing reaction was stopped by washing extensively with ddH_2O.

Analysis of LPS profiles

Bacteria grown for 2 days on SB plates were collected, washed once with PBS, and resuspended in 1 ml PBS at an OD_{600} of 0.75. The cΔ*skp*$_{Cc}$ conditional mutant was grown for 2 days on SB plates containing IPTG and passaged for 1 day on SB plates without IPTG before being collected. Bacteria were centrifuged for 3 min at 5,000 g, resuspended in 125 µl SDS PAGE buffer and heated for 10 min at 99°C. Proteinase K was then added at a final concentration of 50 µg/ml and samples were incubated overnight at 37°C. Samples were again heated, proteinase K added as before and incubation was continued for 3 h at 56°C. Finally, samples were heated before being loaded on 15% SDS PAGE gels. The LPS profiles were then analyzed as previously described for Western blotting using a rabbit anti-Cc5 antiserum.

Outer membrane purification

Bacteria grown for 2 days on SB plates were collected, washed once with ice cold 10 mM HEPES, pH 7.4, and resuspended in 3 ml ice cold HEPES at an OD_{600} of 1. The cΔ*skp*$_{Cc}$ conditional mutant was grown for 2 days on SB plates containing IPTG and passaged for 1 day on SB plates without IPTG before being collected. Bacterial suspensions were then sonicated on ice until clearance and centrifuged at 5,000 g for 5 minutes at 4°C to pellet insoluble material. Supernatants were

collected and centrifuged again for 30 minutes at 20,000 g at 4°C. Pellets were resuspended in 2 ml 10 mM HEPES containing 1% (w/v) sarcosyl (N-Lauroylsarcosine sodium salt, Sigma) and incubated at room temperature for 30 minutes. Finally, samples were centrifuged at 20,000 g for 30 min at 4°C and pellets resuspended in 100 µl SDS PAGE buffer. Samples were heated for 5 min at 96°C and 10 µl were loaded on 12% SDS PAGE gels. Samples were then analyzed by silver staining.

LolA and Skp_{Cc} crosslink and tandem affinity purification

Bacteria expressing either LolA-Strep-His or Skp_{Cc}-Strep-His were grown on SB plates for 2 days at 37°C in the presence of 5% CO_2. Bacteria from 6 plates were harvested and resuspended in 25 ml 10 mM HEPES, pH 7.4. The suspension was then split in two and formaldehyde at a final concentration of 1% was added to one of the tubes. Both tubes were then incubated for 25 min at room temperature under constant agitation. Glycine at a final concentration of 0.5 M was added to both tubes to stop the crosslink reaction. Alternatively, the crosslinking reaction was performed using DSP (22585; ThermoFisher Scientific) at a final concentration of 80 µg/ml and the reaction was stopped by addition of Tris-HCl at a final concentration of 20 mM. Bacteria were collected by centrifugation and resuspended in 35 ml lysis buffer (25 mM Tris-HCl, 150 mM NaCl, 0.4% Triton, 1% sodium deoxycholate, pH 7.6). Alternatively, bacteria were disrupted by 2 passages through a French press at 35,000 psi. The lysates were clarified by centrifugation (12 min at 16,000 g at 4°C) and the supernatant was diluted 1:2 in PBS containing 10 mM imidazole and proteinase inhibitor (cOmplete Mini, EDTA-free protease Inhibitor cocktail tablets; Roche). Aliquots of 3.5 ml of a 50% slurry of chelating Sepharose fast flow beads (GE Healthcare) were coupled to Ni^{2+} according to the manufacturer's instructions, added to the cleared lysates and incubated overnight at 4°C under constant agitation. The solution was then loaded into a column, and the resin was washed with 25 column volumes (CV) of high-salt buffer (50 mM Tris, 500 mM NaCl; pH 8.0) followed by 5 CV of low-salt buffer (50 mM Tris, 100 mM NaCl; pH 8.0). Proteins were eluted from the resin with 2 CV of elution buffer (50 mM Tris, 100 mM NaCl, 350 mM imidazole; pH 8.0). The elution fraction was then diluted 1:2 in

PBS and added on top of 1 ml of a 50% slurry (0.5 ml CV) of Strep-Tactin superflow resin (2-1206-002; IBA). The flowthrough was reloaded twice. The resin was then washed 4 times with 10 CV of buffer W (100 mM Tris, 150 mM NaCl, 1 mM EDTA; pH 8.0) and proteins were eluted in 3 steps with 0.5 ml of elution buffer (100 mM Tris, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin; pH 8.0). The elution fractions were then analyzed by Western blot and silver staining and the proteins were identified by LC/MS as described previously[18].

In silico analyses

Candidate proteins identified in elution fractions of pull-down experiments or from genome mining were selected following *in silico* analyses using DeltaBlast for homology detection[38], CD-search[39] and Pfam[40] for domain annotation, HHpred[41] and Phyre2[42] for structural prediction and LipoP[43] for signal peptide identification. All applications were run with default settings. Protein sequences were recovered from Uniprot[44]. Conservation of proteins was assessed using the MaGe Platform (http://www.genoscope.cns.fr/agc/mage) with the *Capnocytophaga canimorsus* Cc5 genome set as query and the search carried out against the following genomes: *Bacteroides caccae* ATCC 43185, *Bacteroides fragilis* NCTC 9343, *Bacteroides ovatus* ATCC 8483, *Bacteroides thetaiotaomicron* VPI-5482, *Bacteroides vulgatus* ATCC 8482, *Capnocytophaga gingivalis* ATCC 33624, *Capnocytophaga ochracea* DSM 7271, *Cellulophaga algicola* DSM 14237, *Cellulophaga lytica* DSM 7489, *Chitinophaga pinensis* DSM 2588, *Cytophaga hutchinsonii* ATCC 33406, *Dyadobacter fermentans* DSM 18053, *Dysgonomonas mossii* DSM 22836, *Flavobacterium johnsoniae* UW101, *Flavobacterium psychrophilum* JIP02/86, *Gramella forsetii* KT0803, *Leadbetterella byssophila* DSM 17132, *Odoribacter splanchnicus* DSM 20712, *Parabacteroides distasonis* ATCC 8503, *Pedobacter heparinus* DSM 2366, *Pedobacter saltans* DSM 12145, *Porphyromonas gingivalis* ATCC 33277, *Prevotella denticola* F0289, *Prevotella melaninogenica* ATCC 25845, *Prevotella ruminicola* 23, *Rhodothermus marinus* DSM 4252, *Riemerella anatipestifer* DSM 15868 and *Zobellia galactanivorans*. Homology constraints were as follows: Query coverage ≥ 50%, query identity ≥ 20%.

2.3.7. References

1 Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, (2011).
2 Sanchez-Pulido, L. *et al.* POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem Sci* **28**, (2003).
3 Blatch, G. L. & Lassle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21**, (1999).
4 Korndorfer, I. P. *et al.* Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. *Nat Struct Mol Biol* **11**, (2004).
5 Walton, T. A. & Sousa, M. C. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol Cell* **15**, (2004).
6 Meltzer, M. *et al.* Structure, function and regulation of the conserved serine proteases DegP and DegS of Escherichia coli. *Res Microbiol* **160**, (2009).
7 Fischer, G. & Schmid, F. X. The mechanism of protein folding. Implications of in vitro refolding models for de novo protein folding and translocation in the cell. *Biochemistry* **29**, (1990).
8 Putker, F. *et al.* Transport of lipopolysaccharide to the Gram-negative bacterial cell surface. *FEMS Microbiol Rev* **39**, (2015).
9 Gu, Y. *et al.* Structural basis of outer membrane protein insertion by the BAM complex. *Nature* **531**, (2016).
10 Botos, I. *et al.* Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. *Structure* **24**, (2016).
11 Okuda, S. & Tokuda, H. Model of mouth-to-mouth transfer of bacterial lipoproteins through inner membrane LolC, periplasmic LolA, and outer membrane LolB. *Proc Natl Acad Sci U S A* **106**, (2009).
12 Manfredi, P. *et al.* The genome and surface proteome of Capnocytophaga canimorsus reveal a key role of glycan foraging systems in host glycoproteins deglycosylation. *Mol Microbiol* **81**, (2011).
13 Brix, J. *et al.* The mitochondrial import receptor Tom70: identification of a 25 kDa core domain with a specific binding site for preproteins. *J Mol Biol* **303**, (2000).
14 Schlossmann, J. *et al.* Tom71, a novel homologue of the mitochondrial preprotein receptor Tom70. *J Biol Chem* **271**, (1996).
15 Li, J. *et al.* The structural plasticity of Tom71 for mitochondrial precursor translocations. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **66**, (2010).
16 Manfredi, P. *et al.* Complete genome sequence of the dog commensal and human pathogen Capnocytophaga canimorsus strain 5. *J Bacteriol* **193**, (2011).
17 Renzi, F. *et al.* Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. *MBio* **6**, (2015).
18 Renzi, F. *et al.* The N-glycan glycoprotein deglycosylation complex (Gpd) from Capnocytophaga canimorsus deglycosylates human IgG. *PLoS Pathog* **7**, (2011).
19 Manfredi, P. *et al.* New iron acquisition system in Bacteroidetes. *Infect Immun* **83**, (2015).
20 Ricci, D. P. & Silhavy, T. J. The Bam machine: a molecular cooper. *Biochim Biophys Acta* **1818**, (2012).
21 Chimalakonda, G. *et al.* Lipoprotein LptE is required for the assembly of LptD by the beta-barrel assembly machine in the outer membrane of Escherichia coli. *Proc Natl Acad Sci U S A* **108**, (2011).
22 Sklar, J. G. *et al.* Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev* **21**, (2007).
23 Shin, H. *et al.* Resistance of Capnocytophaga canimorsus to killing by human complement and polymorphonuclear leukocytes. *Infect Immun* **77**, (2009).
24 van den Berg, B. *et al.* Crystal structure of the long-chain fatty acid transporter FadL. *Science* **304**, (2004).
25 Wang, Y. *et al.* Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in Pseudomonas putida F1. *Mol Gen Genet* **246**, (1995).
26 Taguchi, Y. *et al.* Involvement of an Skp-Like Protein, PGN_0300, in the Type IX Secretion System of Porphyromonas gingivalis. *Infect Immun* **84**, (2016).
27 Gentle, I. E. *et al.* Molecular architecture and function of the Omp85 family of proteins. *Mol Microbiol* **58**, (2005).
28 Heinz, E. & Lithgow, T. A comprehensive analysis of the Omp85/TpsB protein superfamily structural diversity, taxonomic occurrence, and evolution. *Front Microbiol* **5**, (2014).
29 Webb, C. T. *et al.* Evolution of the beta-barrel assembly machinery. *Trends Microbiol* **20**, (2012).
30 Selkrig, J. *et al.* Assembly of beta-barrel proteins into bacterial outer membranes. *Biochim Biophys Acta* **1843**, (2014).
31 Selkrig, J. *et al.* Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol* **19**, (2012).
32 Gruss, F. *et al.* The structural basis of autotransporter translocation by TamA. *Nat Struct Mol Biol* **20**, (2013).

33 Shen, H. H. *et al.* Reconstitution of a nanomachine driving the assembly of proteins into bacterial outer membranes. *Nat Commun* **5**, (2014).
34 Struve, C. *et al.* Application of a novel multi-screening signature-tagged mutagenesis assay for identification of Klebsiella pneumoniae genes essential in colonization and infection. *Microbiology* **149**, (2003).
35 Burall, L. S. *et al.* Proteus mirabilis genes that contribute to pathogenesis of urinary tract infection: identification of 25 signature-tagged mutants attenuated at least 100-fold. *Infect Immun* **72**, (2004).
36 Kelly, M. *et al.* Essential role of the type III secretion system effector NleB in colonization of mice by Citrobacter rodentium. *Infect Immun* **74**, (2006).
37 Mally, M. & Cornelis, G. R. Genetic tools for studying Capnocytophaga canimorsus. *Appl Environ Microbiol* **74**, (2008).
38 Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, (2008).
39 Marchler-Bauer, A. & Bryant, S. H. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, (2004).
40 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, (2016).
41 Hildebrand, A. *et al.* Fast and accurate automatic structure prediction with HHpred. *Proteins* **77 Suppl 9**, (2009).
42 Kelley, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, (2015).
43 Juncker, A. S. *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* **12**, (2003).
44 Consortium, U. UniProt: a hub for protein information. *Nucleic Acids Res* **43**, (2015).
45 Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science* **277**, (1997).
46 Guzman, L. M. *et al.* Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, (1995).

2.3.8. Supplemental materials

Table S1. Bacterial strains used in this study

Strain	Genotype and/or description	Reference
E. coli		
Top10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lac*ZΔM15 Δ*lacX*74 *recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL endA1 nupG*; Sm^r	Invitrogen
MG1655	K-12 F^- λ^- *ilvG*$^-$ *rfb-50 rph-1*	(45)
C. canimorsus		
Cc5	Wild type (BCCM-LMG 28512)	(37)
ΔCcan_02550	Replacement of *Ccan_02550* by *ermF*; Em^r	This study
cΔCcan_06900	Replacement of *Ccan_06900* by *ermF* in *C. canimorsus* harboring pFL167; Em^r Cfx^r	This study
cΔCcan_09070	Replacement of *Ccan_09070* by *ermF* in *C. canimorsus* harboring pFL169; Em^r Cfx^r	This study
cΔCcan_09090	Replacement of *Ccan_09090* by *ermF* in *C. canimorsus* harboring pFL68; Em^r Cfx^r	This study
ΔCcan_13690	Replacement of *Ccan_13690* by *ermF*; Em^r	This study
cΔCcan_16770	Replacement of *Ccan_16770* by *ermF* in *C. canimorsus* harboring pFL168; Em^r Cfx^r	This study
cΔCcan_17810	Replacement of *Ccan_17810* by *ermF* in *C. canimorsus* harboring pFL170; Em^r Cfx^r	This study
ΔCcan_20230	Replacement of *Ccan_20230* by *ermF*; Em^r	This study

Table S2. Plasmids used in this study

Plasmid	Description	Reference
Vectors[a]		
pMM25	ColE1 *ori*; Km^r (Cf^r); suicide vector for *C. canimorsus*	(37)
pMM47.A	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ermF* promoter	(37)
pPM5	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ompA* promoter	(19)
pFL32	ColE1 ori; (pCC7 ori); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with IPTG inducible *cfxA* promoter	Chapter 3
pFL62	ColE1 ori; (pCC7 ori); Ap^r; (Tc^r). *E. coli-C. canimorsus* expression shuttle plasmid. *ompA* promoter inserted into BamHI/XbaI sites of pMM104.A	This study
pFL172	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid. Addition of Strep and His tag sequences for in frame cloning. Sequences were amplified using primers 7369/7370 and cloned into pPM5 using XhoI/SpeI restriction sites.	This study
pBAD24	pBR *ori*; Ap^r. High copy *E. coli* expression plasmid with arabinose inducible promoter	(46)
pBAD33	pACYC *ori*; Cm^r. Low copy *E. coli* expression plasmid with arabinose inducible promoter	(46)
Mutator plasmids		

pMM106	Δ*siaC*::*ermF*. Cassette for replacement of *siaC* by *ermF.*	(37)
pFL37	Δ*Ccan_16490*::*ermF*. Cassette for replacement of *Ccan_16490* by *ermF* using 7207/7208 and 7209/7210 on gDNA, 7211/7212 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL47	Δ*Ccan_02550*:: *ermF*. Cassette for replacement of *Ccan_02550* by *ermF* using 7286/7287 and 7288/7289 on gDNA, 7290/7291 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL48	Δ*Ccan_06900*:: *ermF*. Cassette for replacement of *Ccan_06900* by *ermF* using 7236/7237 and 7238/7239 on gDNA, 7240/7241 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL49	Δ*Ccan_16770*:: *ermF*. Cassette for replacement of *Ccan_16770* by *ermF* using 7230/7231 and 7232/7233 on gDNA, 7234/7235 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL50	Δ*Ccan_09070*:: *ermF*. Cassette for replacement of Ccan_09070 by *ermF* using 7309/7310 and 7311/7312 on gDNA, 7313/7314 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL51	Δ*Ccan_17810*:: *ermF*. Cassette for replacement of *Ccan_17810* by *ermF* using 7315/7316 and 7317/7318 on gDNA, 7319/7320 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL52	Δ*Ccan_20230*:: *ermF*. Cassette for replacement of *Ccan_20230* by *ermF* using 7321/7322 and 7323/7324 on gDNA, 7325/7326 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL53	Δ*Ccan_09090*:: *ermF*. Cassette for replacement of *Ccan_09090* by *ermF* using 7278/7279 and 7280/7281 on gDNA, 7282/7283 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL166	Δ*Ccan_13690*:: *ermF*. Cassette for replacement of *Ccan_13690* by *ermF* using 7292/7293 and 7294/7295 on gDNA, 7296/7297 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
pFL175	Δ*Ccan_18290*:: *ermF*. Cassette for replacement of *Ccan_18290* by *ermF* using 8174/8175 and 8176/8177 on gDNA, 8178/8179 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
Expression plasmids		
pFL63	Full length Ccan_16490 amplified with primers 7203/7204 and cloned into pFL62 using NcoI/XbaI restriction sites.	This study
pFL64	Full length *Ccan_16490* with a C-terminal Strep and His tag amplified with primers 7203/7205 and cloned into pPM5 using NcoI/XbaI restriction sites	This study
pFL67	Full length *Ccan_09090*-StrepHis amplified from gDNA with 7284/7302 and cloned into pPM5 using NcoI/XbaI. Addition of C-terminal Strep tag in frame with His tag.	This study
pFL68	Full length *Ccan_09090* amplified with primers 7284/7285 and cloned into pFL32 using NcoI/XbaI restriction sites	This study
pFL167	Full length *Ccan_06900* amplified with primers 7222/7229 and cloned into pFL32 using NcoI/XbaI restriction sites	This study
pFL168	Full length *Ccan_16770* amplified with primers 7219/7228 and cloned into pFL32 using NcoI/XbaI restriction sites	This study
pFL169	Full length *Ccan_09070* amplified with primers 7328/7329 and cloned into pFL32 using NcoI/XbaI restriction sites	This study
pFL170	Full length *Ccan_17810* amplified with primers 7330/7331 and cloned into pFL32 using NcoI/XhoI restriction sites	This study
pFL114	Full length *mucG* amplified with primers 7182/6925 and cloned into pBAD24 using NcoI/XbaI restriction sites	This study
pFL116	Full length *Ccan_17810* amplified with primers 7900/7901 and cloned into pBAD33 using KpnI/SphI restriction sites	This study
pFL171	Full length *Ccan_17810* with N-terminal His and Strep tag amplified with primers 7330/7582 and 7581/7331 and cloned into pPM5 using NcoI/XhoI restriction sites. Addition of His-	This study

	Strep tag between amino acids 25 and 26	
pFL173	Full length *Ccan_17810* with C-terminal Strep and His tag amplified with primers 7330/7688 and cloned into pFL172 using NcoI/XhoI restriction sites	This study
pFL174	Full length Ccan_17810 C20G with N-terminal His and Strep tag amplified with primers 7330/7582 and 7583/7331 and cloned into pPM5 using NcoI/XhoI restriction sites. Addition of His-Strep tag between amino acids 25 and 26	This study

[a]: Selection markers for *C. canimorsus* are in between brakets

Table S3. Oligonucleotides used in this study

Ref.	**Sequence 5'-3'**	**Restriction[a]**
7201	ccggatccttttttttaacatttgattttgtatttaaaaaatttgg	BamHI
7202	ggtctagaatatccatggttaatttttttaattacaatttagttaattacaagc	XbaI
7207	ggctgcagtagatgcgcaaggcgccgatcaactcatcgg	PstI
7208	gagtagataaaagcactgttcgtttttataaaatatttttatttttctgcc	
7209	aaaaatttcatccttcgtagaagataagctttgagaattttagaccg	
7210	ccactagtgctactcttatacttaattcatcacc	SpeI
7211	ggcagaaaataaaaatattttataaaaacgaacagtgcttttatctactc	
7212	cggtctaaaattctcaaagcttatcttctacgaaggatgaaatttttt	
7286	ggctgcagcgtgataatattaacattgatttgaatgc	PstI
7287	ctatgatgttgcaaataccgatgagcccaacagccaaagctacgattaaatttttttc	
7288	gaaaaatttcatccttcgtaggcaaatggcagacaaaaaatcagcatatatttttaaaag	
7289	ccactagtgcccgaacggtgctaaatttcttgtcg	SpeI
7290	gaaaaaatttaatcgtagctttggctgttgggctcatcggtatttgcaacatcatag	
7291	taaaatatatgctgatttttttgtctgccatttgcctacgaaggatgaaatttttcag	
7236	ggctgcagtgctttggtttgtaattatatgc	PstI
7237	gagtagataaaagcactgttggtatctaatttttctttgacgtgc	
7238	aaaaatttcatccttcgtagtgtacatattttattatgaataaaaccatagc	
7239	ggactagtcttgataatttgctgcaatgccttcgc	SpeI
7240	gcacgtcaaagaaaattagataccaacagtgcttttatctactc	
7241	gctatggttttattcataataaaatatgtacactacgaaggatgaaatttttcaggg	
7230	ggctgcaggataaaaacatcagctttggctctgttggc	PstI
7231	gagtagataaaagcactgttgctatattcgccttttgaacgctg	
7232	aaaaatttcatccttcgtagattataaatatcacataaccgtatg	
7233	ggactagtggggtatattcaaatcgtatcttaagaaagc	SpeI
7234	cagcgttcaaaaggcgaatatagcaacagtgcttttatctactc	
7235	catacggttatgtgatatttataatctacgaaggatgaaatttttcaggg	
7309	ggctgcagagcttttggtcaagtttcttaaaaagaaatacg	PstI
7310	ctatgatgttgcaaataccgatgagcaattttatttgagttgttcacttgtttttcc	
7311	cctgaaaaatttcatccttcgtagttttggcacgatattttcttaatcac	
7312	ccactagtgaatacaacattgaaacatcactcttttc	SpeI
7313	ggaaaaacaagtgaacaactcaaataaaattgctcatcggtatttgcaacatcatag	

7314	caaaagtgattaagaaaatatcgtgccaaaactacgaaggatgaaatttttcagg	
7315	ggctgcaggaagataatgcacttgaacgcg	PstI
7316	ctatgatgttgcaaataccgatgagcaaataagacaaattaggcttca	
7317	cctgaaaaatttcatccttcgtagataattgctctttatcaatcaattttcatac	
7318	ccactagtctgtcagattatcaatcactgcgg	SpeI
7319	tgaagcctaatttgtcttatttgctcatcggtatttgcaacatcatag	
7320	gtatgaaaattgattgataaagagcaattatctacgaaggatgaaatttttcagg	
7321	ggctgcaggctttattcttttacttttgc	PstI
7322	ctatgatgttgcaaataccgatgagcggttactttctttttttattaaagaattgg	
7323	cctgaaaaatttcatccttcgtagttttctatgaatgatttattgaaaaatagg	
7324	ccactagtgcgttttccgtggcgactcacagc	SpeI
7325	ccaattctttaataaaaaaagaaagtaaccgctcatcggtatttgcaacatcatag	
7326	cctatttttcaataaatcattcatagaaaactacgaaggatgaaatttttcagg	
7278	ggctgcaggcaagaaaaatttggagcagaaaatggcgagtacg	PstI
7279	ctatgatgttgcaaataccgatgagcctgtgtaaattaaagtgttaatataatc	
7280	cctgaaaaatttcatccttcgtagtaaaaaattaaataaaatcgttattaaaagtgttcg	
7281	ccactagtaaaattccgttttttcgcaccgaatcgttactgg	SpeI
7282	gattatattaacactttaatttacacaggctcatcggtatttgcaacatcatag	
7283	cgaacacttttaataacgattttatttaatttttttactacgaaggatgaaatttttcagg	
7292	ggctgcaggggcgaaactcacaaggggtacgtgg	PstI
7293	ctatgatgttgcaaataccgatgagcatctcttcgaatttatttttaaaatttttattc	
7294	cctgaaaaatttcatccttcgtagttttgtttcataataaataattaaaccccctc	
7295	ccactagtcaagttttaaatcacgattttaaagg	SpeI
7296	gaataaaaatttttaaaaataaattcgaagagatgctcatcggtatttgcaacatcatag	
7297	ggaggggtttaattatttattatgaaacaaaactacgaaggatgaaatttttcag	
7203	ccccatggggaaaaagatactattgttaatatc	NcoI
7204	ggtctagattatagttctgaaatatagtatcc	XbaI
7205	ggtctagagctttttcgaactgcgggtggctccatagttctgaaatatagtatccttc	XbaI
7284	ggccatggggatgagaaacattaaaactattatgattgcg	NcoI
7302	ggtctagagctttttcgaactgcgggtggctccagaatcctaattccttttaacg	XbaI
7285	ggtctagattagaatcctaattccttttaacgc	XbaI
7222	ggccatggggttgatactcaacatagcatacatacg	NcoI
7229	ggtctagattatttttgttcagatgtttttttcttg	XbaI
7219	ggccatggggttgcataaaaaaatttaattcaaacataaaaaaagttactattttgc	NcoI
7228	ggtctagattaattcaaacgtcggtcgggttctcgg	XbaI
7328	ggccatggggtcattgaagaaatttttgtctg	NcoI
7329	cctctagattaaaattgttgtccaaaaataaagtgtgtttgcc	XbaI
7330	ggccatggggaaatctcgttttaaaatatattgc	NcoI
7331	ggctcgagttagaaaggataatttatccc	XhoI
7182	ggccatggggaaaaaaatagtatccattagc	NcoI
6925	gctctagactaaaacgtaacttgagttctctctccg	XhoI
7900	atcgggtaccaggaggaattcaccatgaaatctcgttttaaaatatattgc	KpnI
7901	atcggcatgcttagaaaggataatttatccc	SphI
7581	catcacgctgcctggagccacccgcagttcgaaaaagtaccagaaaacacgcatttgc	
7582	gctccaggcagcgtgatgatggtgatgatgtcgtttggttgcattacaagccactaacg	

7688	ggctcgaggaaaggataatttatcccaaaattatatac	XhoI
7369	ccgctcgagtggagccacccgcagttcgaaaaaggcgcac	XhoI
7370	ggactagttcaatgatgatgatgatgtgcgccttttttcg	SpeI
8147	ggctgcagtggaagatactatgatagttcaggacg	PstI
8145	ctatgatgttgcaaataccgatgagccttttcttctttagtttaattg	
8176	cctgaaaaatttcatccttcgtagatatttaattttgttagttttaaaaaagtcagc	
8177	cc$_{actagt}$cctcctatggagatattgatattgaac	SpeI
8178	caattaaactaaagaagaaaaggctcatcggtatttgcaacatcatag	
8179	gctgacttttttaaaactaacaaaattaaatatctacgaaggatgaaatttttcagg	

[a]: Restriction sites are underlined

3. Development of regulatable expression systems for *C. canimorsus*

3.1.Introduction

C. canimorsus was first isolated in 1976[1] and named dysgonic fermentor 2 (DF2). Its actual name was given in 1989 by Brenner[2] that described the main features of this species. Since 2001, our lab started the in depth characterization of this organism and to this aim a set of genetic tools was developed. This included the engineering of expression vectors as well as suicide plasmids in order to perform gene replacement by conjugation and homologous recombination[3]. Thanks to these tools, we succeeded in the identification and characterization of key genes involved in both bacterial metabolism and pathogenesis. Unfortunately the study of essential genes could not be afforded without genetic tools allowing to control the expression of these genes, *i.e.* regulatable promoters. Indeed, while many genetic systems have been developed for Proteobacteria, these often cannot be used in Bacteroidetes due to promoter incompatibility. We therefore sought to engineer vectors with inducible and/or repressible promoters for *C. canimorsus* that would allow switching on/off genes of interest. This would be of particular interest in respect to the identification of the lipoprotein export machinery that is likely essential in this bacterium.

3.2.Adaptation of an IPTG-inducible expression system

A recent study described the construction of an IPTG-inducible expression vector for *Bacteroides fragilis*[4]. This was achieved by engineering the promoter of the *cfxA* resistance gene (P*cfxA*), adding the LacI binding sequences *lacO₃* and *lacO₁* upstream of the -33 binding box and downstream of the transcription initiation site (TIS) of P*cfxA*, giving the *lacO₃* P*cfxA lacO₁* construct. This fragment was then inserted into an expression vector encoding *lacI* under the control of the constitutive *tetQ* promoter (P*tetQ*). This rendered the final plasmid pFD1146 in which the expression of genes cloned downstream of *lacO₃* P*cfxA lacO₁* is induced by IPTG. Briefly, in the absence of IPTG (non permissive condition), LacI binds to the *lacO* binding sites in the modified P*cfxA* promoter[4,5], thereby blocking transcription. In the presence of IPTG (permissive condition),

LacI binds to IPTG which in turn blocks its binding to *lacO*, thus allowing transcription of the gene of interest[4,5].

Taking advantage of this, we recreated the above system by amplifying the *lacO₃* P*cfxA lacO₁* and P*tetQ-lacI* fragments from pFD1146 and inserting them into the *C. canimorsus* expression vector pMM47.A[3], giving the final vector pFL32 (Fig. 1A and B). We then cloned the *C. canimorsus mucG* gene downstream of the *lacO₃* P*cfxA lacO₁* promoter (plasmid pFL157) and tested MucG IPTG dependent expression. As shown in Fig. 1C, addition of IPTG significantly increased the expression of MucG in the strain harboring pFL157. The plasmid pFL32 allowed us to characterize for the first time an essential *C. canimorsus* gene (*Ccan_09090*, encoding Skp_{Cc}) by constructing a conditional mutant that relied on IPTG for growth (see chapter 2). However, as already observed by the authors that constructed pFD1146, the expression of this system is leaky, having a basal expression even in the absence of IPTG[4]. This was for us a major issue when we wanted to generate conditional mutants for *Ccan_09070*, *Ccan_17810*, *Ccan_06900* and *Ccan_16770*, candidate genes for the lipoprotein export machinery. Indeed, while all these genes are essential in *C. canimorsus*, no growth defect could be observed in non-permissive conditions (see chapter 2). We reasoned that the basal expression of these genes from the leaky *lacO₃* P*cfxA lacO₁* promoter might be sufficient to sustain growth of *C. canimorsus*. This hypothesis is also supported by the relatively low expression level of these genes as compared to *Ccan_09090* (4 to 5 times lower) as determined by mRNA sequencing (K. Hack, unpublished data). Because the use of this system could not allow us to study most of the essential lipoprotein transporter candidates, we decided to construct new regulatable expression vectors for *C. canimorsus.*

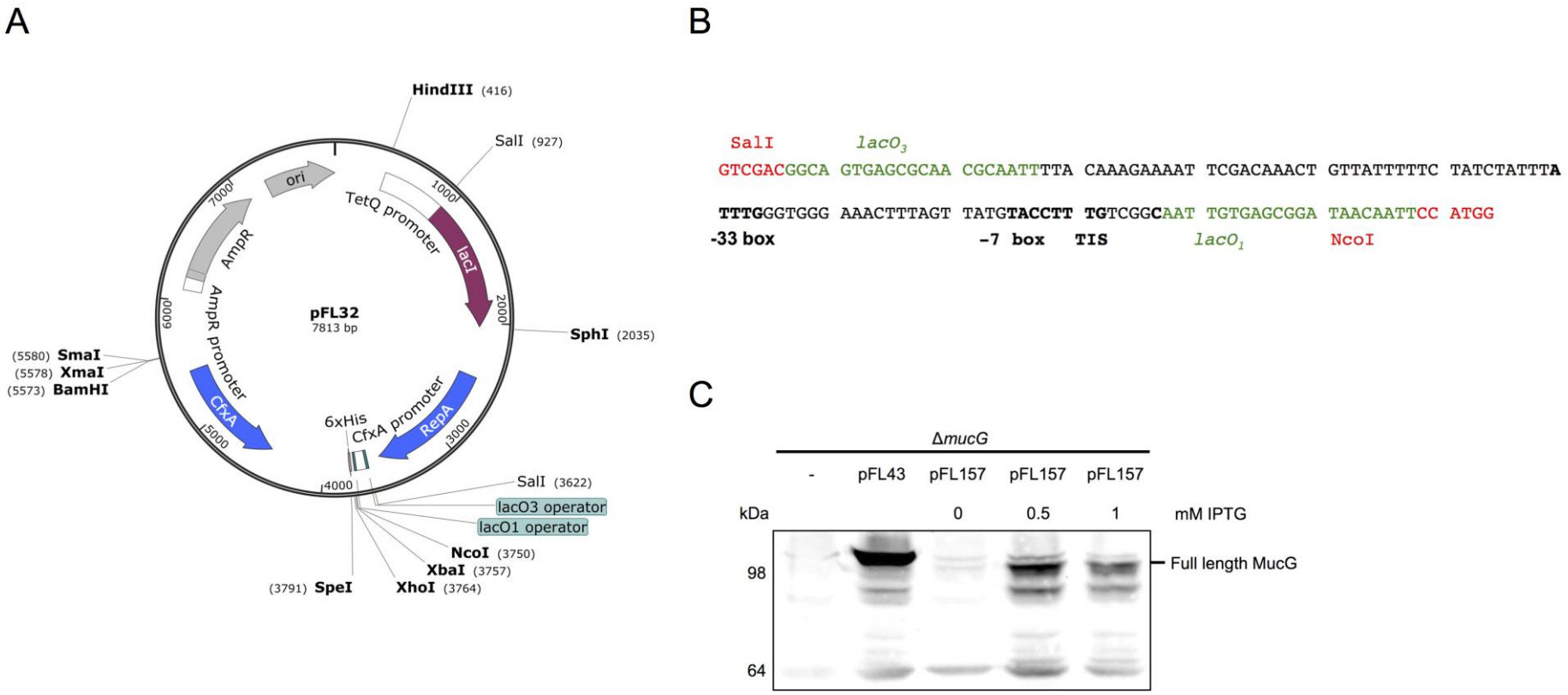


Fig. 1 Construction of an IPTG-regulatable expression system for *C. canimorsus*

(A) Map of the expression vector pFL32. *E. coli* specific genetic elements are indicated in grey, *C. canimorsus* specific genetic elements in blue. The *lacI* and *lacO* binding sites are indicated in purple and green, respectively. (B) Nucleotide sequence of the $lacO_3$ P*cfxA* $lacO_1$ promoter amplified from pFD1146. $lacO_1$ and $lacO_3$ binding sites are indicated in green, SalI and NcoI restriction sites in red, the -33 and -7 boxes and the transcription initiation site (TIS) in bold. (C) Western blot analysis of IPTG induction test of pFL157 (pFL32 encoding *mucG*). *C. canimorsus* cells harboring pFL157 were grown for one day on IPTG-containing plates before being analyzed. Proteins were detected using an anti-HA antibody. Δ*mucG* serves as negative control; Δ*mucG* + pFL43, constitutively expressing MucG, serves as positive control.

3.3.Construction of TetR-based expression systems

As already mentioned, IPTG-inducible expression systems are known to be leaky, even in *E. coli*. More tightly regulated systems have therefore been developed to circumvent this problem. One of the most widespread systems, both for prokaryotes and eukaryotes, is based on the use of tetracycline (Tc)[6].

In bacteria, after entering the cytoplasm of the cell, Tc binds to the small subunit of the ribosome resulting in protein synthesis inhibition and ultimately growth arrest[7]. The most common tetracycline resistance mechanism in bacteria, for example in *E. coli*, relies on two proteins, TetR and TetA[7]. TetR is a homodimeric transcription regulator that binds Tc once it reaches the cytoplasm. TetA is an IM spanning efflux pump that exports the TetR-Tc complex out of the cytoplasm[7]. In the absence of Tc, TetR binds to its cognate nucleotide binding site *tetO*, thereby blocking the transcription of TetA. Upon addition of Tc, the TetR-Tc complex is formed, resulting in dissociation of TetR from *tetO*. Transcription of TetA can thus resume, leading to active export of Tc from the cytoplasm[7]. The classical vector system therefore usually consists of one promoter harboring the *tetO* binding sites downstream of which the gene of interest is cloned and of the *tetR* gene under the control of a constitutive promoter. In brief, absence of Tc inhibits gene expression while addition of Tc induces it.

Several features have contributed to make this system popular. First, TetR is not subject to catabolic repression as it can be the case for LacI-based systems[8]. Second, an artificial analog of Tc, anhydrotetracycline (ATc), has been developed[9]. This compound has the dual advantage of having a higher affinity than Tc towards TetR and at the same time poorly binds to the small ribosomal subunit. Compared to Tc, ATc is therefore a more potent regulator of gene expression whit the advantage of having almost no bacteriostatic effect[9]. Finally, a reverse variant of TetR (revTetR) has been isolated and found to bind to *tetO* only in the presence of Tc, thus allowing switching off gene expression instead of switching it on[10]. For these reasons, we decided to adapt Tc-based vector systems to *C. canimorsus*.

To this aim, we introduced an *E. coli tetO* binding site upstream of the -33 box of the P*cfxA* promoter (*tetO* P*cfxA*) and replaced the constitutive promoter of

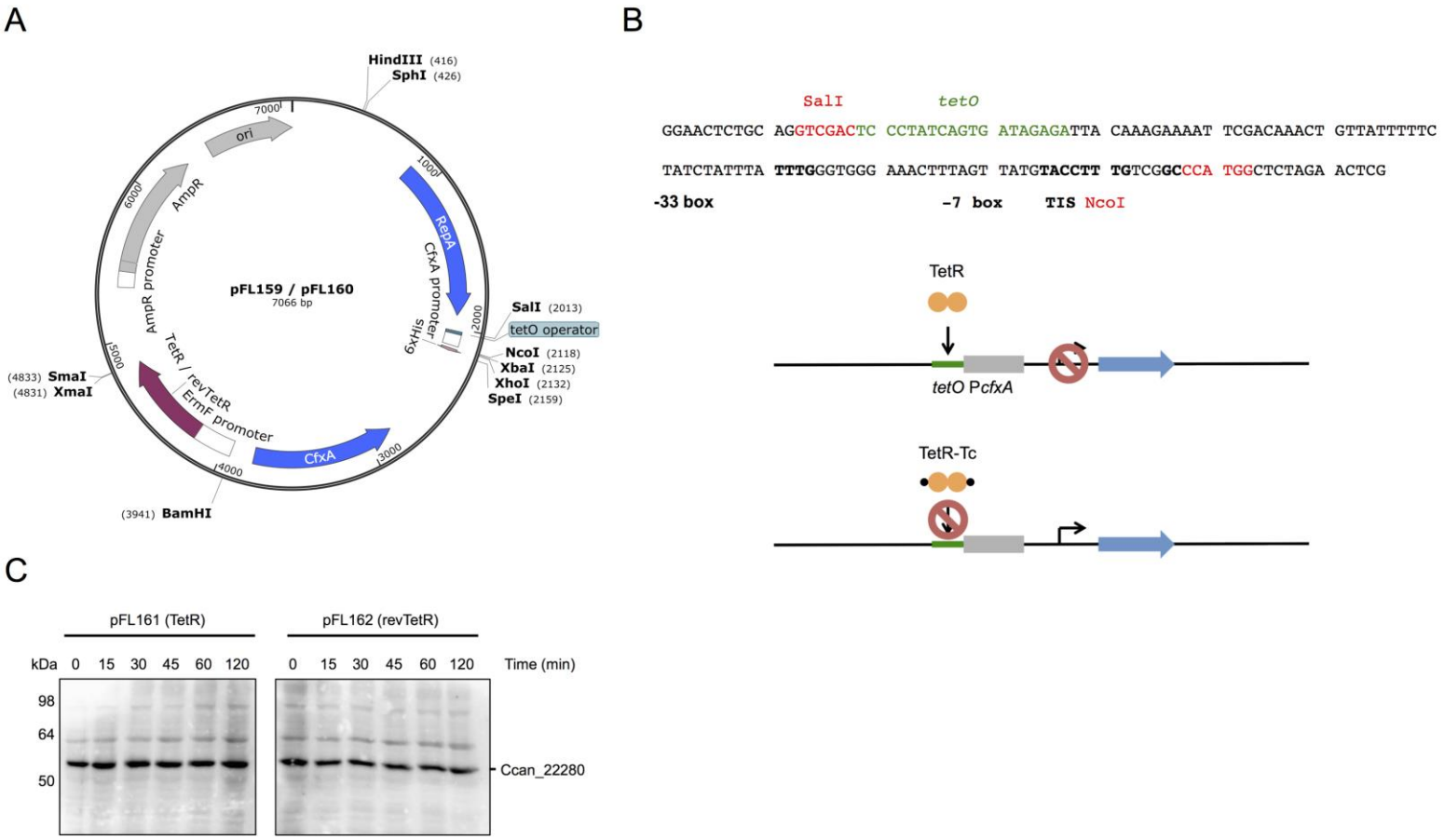


Fig. 2 Construction of a TetR-based expression system for *C. canimorsus* using one *tetO* binding site

(A) Map of the expression vectors pFL159 and pFL160. *E. coli* specific genetic elements are indicated in grey, *C. canimorsus* specific genetic elements in blue. TetR/revTetR and the *tetO* binding site are indicated in purple and green, respectively. (B) Nucleotide sequence of the *tetO* P*cfxA* promoter amplified from pMM47.A. The *tetO* binding site is indicated in green, SalI and NcoI restriction sites in red, the -33 and -7 boxes and the transcription initiation site (TIS) in bold. A schematic reprensentation of the regulation mechanism is shown below. Note that only the TetR regulation is depicted and that revTetR functions in the opposite way. (C) Western blot analysis of pFL161 and pFL162 (pFL159 and pFL160 encoding Ccan_22280 respectively) induction test using 100 ng/ml of ATc. Proteins were detected using an anti-HA antibody.

pMM47.A by this construct. The resulting plasmid was then restricted in order to insert either the *tetR* or *revtetR* encoding gene under the control of the constitutive *ermF* promoter (P*ermF*), leading to pFL159 (TetR inducible system) and pFL160 (revTetR repressible system) vectors (Fig. 2A and B). We then monitored expression of the *C. canimorsus* amylase Ccan_22280 over time upon addition of ATc to assess the performance of these vectors. Surprisingly, the amount of protein detected at t_0 and at later time points was constant in both constructs (Fig. 2C) thus showing absence of regulation. The presence of Ccan_22280 throughout the course of the experiment in the revTetR-based system might be explained by an insufficient amount of ATc, resulting in incomplete repression of gene expression. However, the presence of a band at t_0 in the TetR-based system, where no expression should occur unless ATc is provided, rather suggests that the TetR protein is either not expressed or is not recognizing the *tetO* binding sequence. Alternatively, the presence of only one *tetO* binding site might not be enough to assure control over gene expression.

To rule out this possibility, we decided to mimic the previously constructed LacI-based system by inserting *tetO* binding sites upstream and downstream of the P*cfxA* promoter, giving *tetO* P*cfxA tetO*. This construct was then inserted in pFL161 instead of the *tetO* P*cfxA* promoter leading to the vector pFL163 (Fig. 3A and B). We then monitored expression of Ccan_22280 over time following addition of ATc to the medium (Fig. 3C). As for plasmid pFL161, the protein was already detectable at t_0 and no increase of protein amount was observed at later time points, even in presence of a high concentration of inducer thus indicating complete absence of regulation. Since the *tetO* binding sites are inserted at the same positions as the *lacO* binding sites, we hypothesized that TetR could not be expressed or could undergo proteolysis in *C. canimorsus*. However, when a HA-tagged version of TetR was cloned downstream of a constitutive promoter and tested in *C. canimorsus*, the protein was clearly detectable (data not shown). The most likely conclusion therefore is that although expressed, *E. coli* TetR might not be folded properly in *C. canimorsus* and hence does not bind to *tetO*.

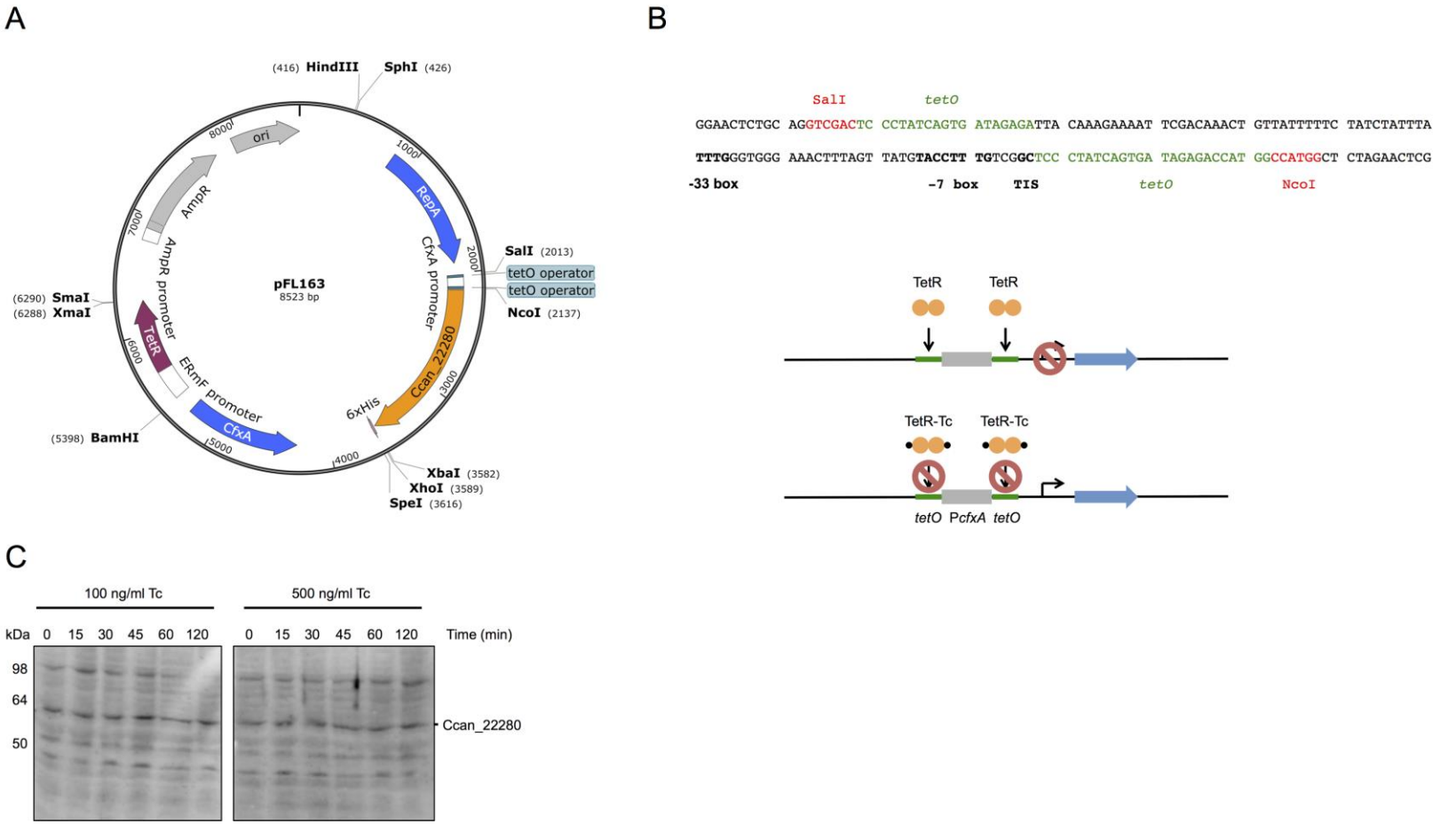


Fig. 3 Construction of a TetR-based expression system for *C. canimorsus* using two *tetO* binding sites

(A) Map of the expression vector pFL163. *E. coli* specific genetic elements are indicated in grey, *C. canimorsus* specific genetic elements in blue. TetR and the *tetO* binding sites are indicated in purple and green, respectively. Ccan_22280 is indicated in orange. (B) Nucleotide sequence of the *tetO* P*cfxA tetO* promoter. The *tetO* binding sites are indicated in green, SalI and NcoI restriction sites in red, the -33 and -7 boxes and the transcription initiation site (TIS) in bold. A schematic reprensentation of the regulation mechanism is shown below. (C) Western blot analysis of amylase Ccan_22280 expression from pFL163 induction test using 100 and 500 ng/ml of ATc. Proteins were detected using an anti-HA antibody.

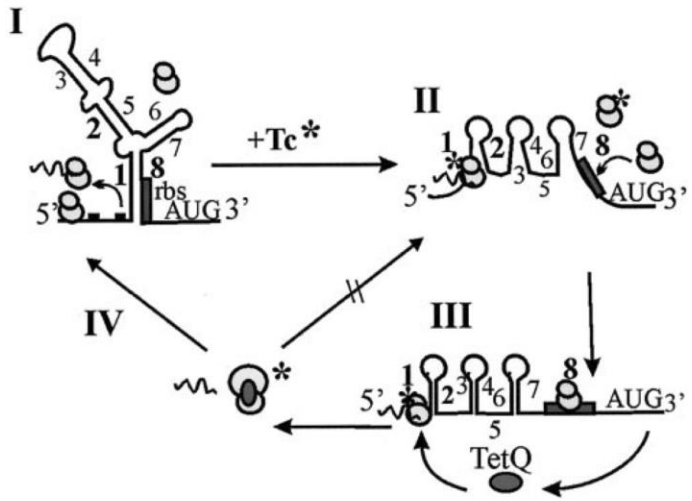
3.4.Construction of a TetQ-based expression system

We thus thought of an alternative approach based on the tetracycline resistance mechanism described in Bacteroidetes. Unlike *E. coli*, Bacteroidetes species do not actively export Tc out of the cytoplasm but modify their ribosomes to prevent Tc binding thanks to the TetQ resistance protein[7,11]. It is important to note that in this system, the amount of TetQ is controlled at the level of translation rather than of transcription, meaning that a certain amount of *tetQ* mRNA is always present in the cell. This translational regulation is achieved thanks to the organization of the *tetQ* mRNA that contains two regions, the leader sequence and the *tetQ* sequence[11] (Fig. 4A).

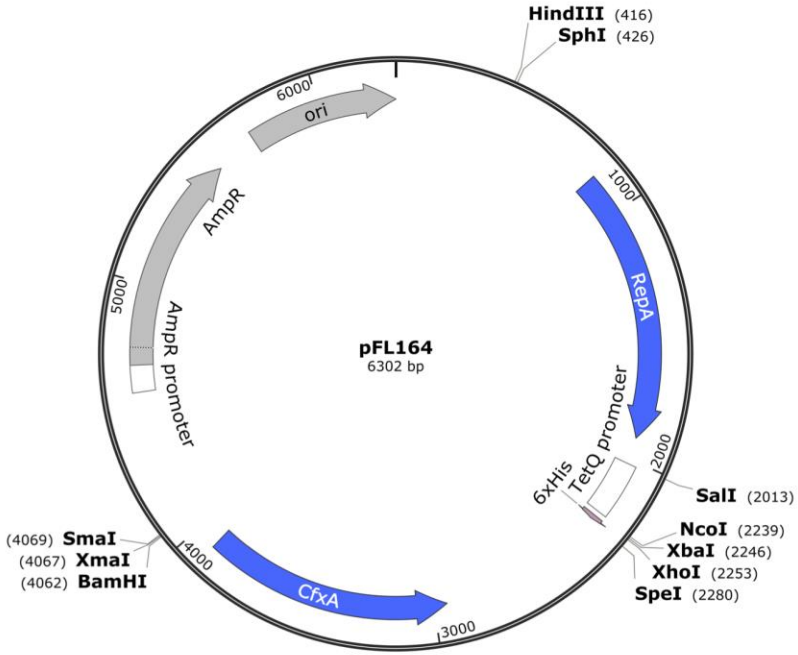
In the absence of Tc, ribosomes bind to the leader sequence of the *tetQ* mRNA and translate a so-called leader peptide[11]. This results in the formation of a hairpin structure that makes the ribosome binding site (RBS) of the *tetQ* sequence inaccessible (Fig. 4A, step I). Once Tc binds to ribosomes, translation of the leader sequence stalls. This allows formation of alternative hairpin structures, freeing the *tetQ* RBS and thus leading to translation of the transcript[11] (Fig. 4A, step II). TetQ then interacts with ribosomes in order to prevent their Tc binding[7,11] (Fig. 4A, step III), thus enabling resuming of protein synthesis. Interestingly, this leads to a negative feedback loop, as the now Tc-resistant ribosomes will again be able to translate the leader sequence, which in turn blocks translation of *tetQ* (Fig. 4A, step IV). The resistance conferred by TetQ therefore leads to a balance between Tc-resistant and Tc-sensitive ribosomes.

In order to see whether this regulation could be exploited in *C. canimorsus*, we amplified the *tetQ* promoter and leader sequence of pMM104 and inserted it into pMM47.A, giving plasmid pFL164 (Fig. 4B and C). We then cloned *mucG* downstream of the *tetQ* promoter, originating plasmid pFL165. We did not add the *tetQ* resistance gene itself in this vector in order to avoid the above mentioned negative feedback loop. Since the regulation mechanism depends on the ability of Tc to bind to ribosomes, ATc, because of its low affinity for ribosomes, cannot be used in this system. We thus tested several sub-inhibitory concentrations of Tc and monitored MucG expression (Fig. 4D). Surprisingly, MucG was not expressed in any condition tested. This result could be due to the

A



B



C

```
            SalI
GGAACTCTGC AGGTCGACAC CTACGTTTCC CTAATAAAAT GTCTATGGTA AAAAGTTAAA AAATCCTCCT
   -33 box                    -7 box   TIS              Hp1
ACTTTTGTTA GATATATTTT TTTGTGTAAT TTTGTAATCG TTATGCGGCA GTAATAATAT ACATATTAAT

ACGAGTTAGG AATCCTGTAG TTCTCATATG CTACGAGGAG GTATTAAAAG GTGCGTTTCG ACAATGCATC

TATTGTAGTA TATTATTGCT TAATCCAACC ATGGCTCTAG AACTCG
          Hp8           NcoI
```

D

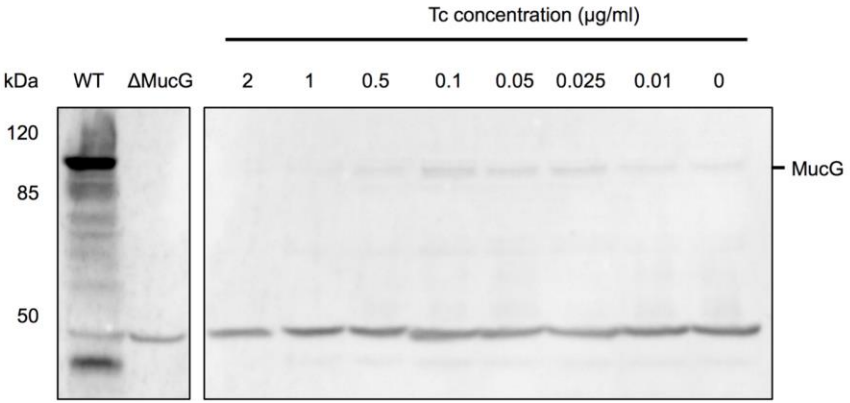


Fig. 4 Construction of a Tc-regulatable expression system for *C. canimorsus* using the *tetQ* promoter
(A) Proposed model of TetQ provided tetracycline resistance. In the absence of Tc, ribosomes bind to the *tetQ* mRNA and translate a leader peptide (dashed line) (step I). This results in the formation of a hairpin structure between regions Hp1 and Hp8 of the mRNA, rendering the RBS of *tetQ* inaccessible to ribosomes (shaded box). Once Tc (asterisk) binds to ribosomes, translation of the leader sequence stalls (step II). This induces the formation of alternative hairpin structures in the mRNA, rendering the RBS of *tetQ* accessible to ribosomes that have not bound Tc. This leads to TetQ synthesis, which will interact with ribosomes to make them Tc-resistant (step III). Tc-resistant ribosomes are then able to resume translation of the leader peptide, which results in the formation of the hairpin structure blocking *tetQ* translation. TetQ therefore regulates its own expression by a negative feedback loop. (B) Map of the expression vector pFL164. *E. coli* specific genetic elements are indicated in grey, *C. canimorsus* specific genetic elements in blue. The *tetQ* promoter is indicated by a white box. (C) Nucleotide sequence of the *tetQ* promoter. The Hp1 and Hp8 sequences involved in haiprin formation are indicated in green, SalI and NcoI restriction sites in red, the -33 and -7 boxes and the transcription initiation site (TIS) in bold. (D) Western blot analysis of pFL165 (pFL164 encoding *mucG*) induction test using various concentration of Tc. Proteins were detected using an anti-HA antibody.

fact that sub-inhibitory concentrations of Tc might still affect a large proportion of ribosomes thus attenuating protein synthesis. Alternatively, the Tc concentrations used might be too low to efficiently induce ribosome stalling on the leader sequence, thereby preventing MucG expression. We are therefore currently testing if the addition of the TetQ encoding gene would improve the expression of MucG in this system. While this would add the risk of decreasing the translation of MucG due to the negative feedback loop, it would allow us to use higher Tc concentrations. This could generate equilibrium between Tc-sensitive and Tc-resistant ribosomes, resulting in an overall constant expression.

3.1.Conclusion and perspectives

Development of new genetic tools and in particular of regulatable protein expression systems for *C. canimorsus* is critical and so far represents the bottleneck for the study of essential genes in this organism. While construction of new regulatory systems has so far proven difficult, in the future efforts will be

pursued in order to be able to characterize physiological processes such as lipoprotein surface transport in detail. So far, our attempts have mainly focused on promoter activity regulation; however, a very recent study showed that RNA silencing takes place in Bacteroidetes and that it can be used for gene regulation[12]. This therefore opens the door for new regulatory strategies that we will try to adapt for *C. canimorsus*.

3.2.Materials and Methods

Chemicals and reagents

Tetracycline (Tc, T7660, Sigma) and anhydrotetracycline (ATc, 37919, Sigma) were stored at -20°C as 10 mg/ml and 2.5 mg/ml stock solutions respectively in 50% ethanol. IPTG (I6758, Sigma) was sterile filtered and stored at -20°C as 0.5 M stock solution in ddH_2O. Western blot analysis of MucG-HA and Ccan_22280-HA was performed using rat anti-HA (11867423001, Roche) and goat-HRP anti-rat (629520, Invitrogen) antibodies.

Bacterial strains and growth conditions

(i) Conventional bacterial growth conditions and selective agents

Bacterial strains used in this study are listed in Table S1. *E. coli* strains were routinely grown in lysogeny broth (LB) at 37°C. *C. canimorsus* strains were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) plates (SB plates) for 2 days at 37°C in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 100 µg/ml ampicillin (Amp), 50 µg/ml kanamycin (Km), for *E. coli* and 10 µg/ml erythromycin (Em), 10 µg/ml cefoxitin (Cfx), 20 µg/ml gentamicin (Gm) for *C. canimorsus*.

(ii) IPTG induction assay

Bacteria harboring pFL157 were grown for 2 days on SB plates before being passage for 1 day on SB containing 0, 0.5 or 1 mM IPTG. Bacteria were then collected, washed once with PBS, and resuspended in 1 ml PBS at an OD_{600} of 1.

Bacteria were centrifuged for 3 min at 5,000 g and resuspended in 100 µl SDS PAGE buffer, followed by Western blot analysis.

(iii) Tetracycline/anhydrotetracycline induction assays

Strains harboring pFL161 or pFL162 were grown for 2 days on SB plates, collected, washed once with PBS, and resuspended in 1 ml PBS at an OD_{600} of 10. The bacterial suspensions were then used to inoculate 10 ml PBS containing 20% (v/v) fetal calf serum. ATc was added at a final concentration of 100 ng/ml and cultures were incubated at 37°C with constant agitation. One-ml samples were collected over time, washed once with PBS and resuspended in 100 µl SDS PAGE buffer followed by Western blot analysis. Bacteria harboring pFL163 were treated in the same way, except that 100 and 500 ng/ml ATc concentrations were tested in parallel.

Tetracycline-dependent expression of bacteria harboring pFL165 was assessed in 96-well plates. Inocula were prepared from cultures grown on SB plates, set to an OD_{600} of 1, and serially diluted 1:10 three times. Two hundred µl of bacterial suspension were then used to inoculate 10 ml DMEM containing 10% (v/v) HIHS and dispensed into 1 ml aliquots. After 18 h of incubation, 0.01, 0.025, 0.05, 0.1, 0.5, 1 or 2 µg/ml of Tc were added and incubation was continued for 6 h. Bacteria were then collected by centrifugation, washed twice with PBS and resuspend in 75 µl SDS PAGE buffer followed by Western blot analysis.

Genetic manipulation of *C. canimorsus* and *E. coli* MG1655

Plasmids and primers used in this study are listed in Table S2 and S3.

(i) Construction of pFL32 expression vector

In order to adapt the IPTG-inducible promoter system described in [4], the *lacO*$_3$ P*cfxA lacO*$_1$ promoter from pFD1146 was amplified using 7144 and 7145 and cloned into pMM47.A using SalI and NcoI restriction sites, leading to plasmid pFL31. P*tetQ-lacI* was amplified from pFD1146 DNA using 7195 and 7149 and cloned into pFL31 using HindIII and SphI restriction sites, leading to plasmid pFL32. To test IPTG induction, *mucG* (*Ccan_17430*) was amplified from *C.*

canimorsus genomic DNA using 7182 and 7183 and cloned into pFL32 using NcoI and XbaI restriction sites, leading to plasmid pFL157.

(ii) Construction of *tetO*-P*cfxA* expression vector

P*cfxA* was amplified from pMM47.A using 7242 and 7243 and cloned into pPM5 using SalI and NcoI restriction sites, leading to plasmid pFL158. In order to insert the *tetO* binding sequence upstream of the -33 box of the P*cfxA* promoter, the *tetO* sequence 5'-TCCCTATCAGTGATAGAGA-3' was included into primer 7242. The *tetR* and *revtetR* genes were fused to the P*ermF* promoter by overlapping PCR. First, P*ermF* was amplified from pMM47.A using 7244 and 7245 for fusion with *tetR* or using 7244 and 7246 for fusion with *revtetR*. In parallel, *tetR* was amplified from pBBR-IBA3plus using 7247 and 7249 while *revtetR* was amplified from pDH624 using 7248 and 7250. For each overlapping PCR, both products were cleaned and then mixed in equal amounts for PCR as described previously. The resulting fragments were cloned into pFL158 using BamHI and XmaI restriction sites, leading to pFL159 (TetR) and pFL160 (RevTetR). Finally, *C. canimorsus* amylase (*Ccan_22280*) with a C-terminal HA tag was amplified from genomic DNA using 7184 and 7185 and cloned into pFL159 or pFL160 using NcoI and XbaI restriction sites, leading to pFL161 and pFL162 respectively.

(iii) Construction of *tetO*-P*cfxA*-*tetO* expression vector

The *tetO*-P*cfxA* promoter was amplified from pFL158 using 7242 and 7258 and cloned into pFL161 using SalI and NcoI restriction sites, leading to plasmid pFL163. In order to insert the *tetO* binding sequence downstream of the transcription initiation site of the P*cfxA* promoter, the *tetO* sequence 5'-TCCCTATCAGTGATAGAGA-3' was included into reverse primer 7258, resulting in a *tetO*-P*cfxA*-*tetO* construct.

(iv) Construction of the P*tetQ* expression vector

The *tetQ* promoter and leader sequence was amplified from pMM104 using 7888 and 7889 and cloned into pMM47.A using SalI and NcoI restrictions sites, leading to plasmid pFL164. *C. canimorsus mucG* (*Ccan_17430*) was amplified from

genomic DNA using 7182 and 7183 and cloned into pFL164 using NcoI and XbaI restriction sites, leading to pFL165.

3.3.References

1 Bobo, R. A. & Newton, E. J. A previously undescribed gram-negative bacillus causing septicemia and meningitis. *Am J Clin Pathol* **65**, (1976).

2 Brenner, D. J. *et al.* Capnocytophaga canimorsus sp. nov. (formerly CDC group DF-2), a cause of septicemia following dog bite, and C. cynodegmi sp. nov., a cause of localized wound infection following dog bite. *J Clin Microbiol* **27**, (1989).

3 Mally, M. & Cornelis, G. R. Genetic tools for studying Capnocytophaga canimorsus. *Appl Environ Microbiol* **74**, (2008).

4 Parker, A. C. & Jeffrey Smith, C. Development of an IPTG inducible expression vector adapted for Bacteroides fragilis. *Plasmid* **68**, (2012).

5 Kercher, M. A. *et al.* Lac repressor-operator complex. *Curr Opin Struct Biol* **7**, (1997).

6 Saenger, W. *et al.* The Tetracycline Repressor-A Paradigm for a Biological Switch. *Angew Chem Int Ed Engl* **39**, (2000).

7 Chopra, I. & Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Biol Rev* **65**, (2001).

8 Gorke, B. & Stulke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* **6**, (2008).

9 Degenkolb, J. *et al.* Structural requirements of tetracycline-Tet repressor interaction: determination of equilibrium binding constants for tetracycline analogs with the Tet repressor. *Antimicrob Agents Chemother* **35**, (1991).

10 Scholz, O. *et al.* Activity reversal of Tet repressor caused by single amino acid exchanges. *Mol Microbiol* **53**, (2004).

11 Wang, Y. *et al.* Translational control of tetracycline resistance and conjugation in the Bacteroides conjugative transposon CTnDOT. *J Bacteriol* **187**, (2005).

12 Cao, Y. *et al.* Cis-encoded sRNAs, a conserved mechanism for repression of polysaccharide utilization in the Bacteroides. *J Bacteriol*, (2016).

13 Renzi, F. *et al.* Glycan-foraging systems reveal the adaptation of Capnocytophaga canimorsus to the dog mouth. *MBio* **6**, (2015).

14 Thanbichler, M. & Shapiro, L. MipZ, a spatial regulator coordinating chromosome segregation with cell division in Caulobacter. *Cell* **126**, (2006).

15 Zilio, N. *et al.* A new versatile system for rapid control of gene expression in the fission yeast Schizosaccharomyces pombe. *Yeast* **29**, (2012).

3.4.Supplemental materials

Table S1. Bacterial strains used in this study

Strain	Genotype and/or description	Reference
E. coli		
Top10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lac*ZΔM15 Δ*lacX*74 *recA1 araD139* Δ(*araleu*)7697 *galU galK rpsL endA1 nupG*; Sm^r	Invitrogen
C. canimorsus		
Cc5	Wild type (BCCM-LMG 28512)	(3)
Δ*mucG*	Replacement of *Ccan_02550* by *ermF*; Em^r	(13)
Δ*Ccan_22280*	Replacement of *Ccan_22280* by *ermF*; Em^r	This study

Table S2. Plasmids used in this study

Plasmid	Description	Reference
Vectors[a]		
pMM25	ColE1 *ori*; Km^r (Cf^r); suicide vector for *C. canimorsus*	(3)
pMM47.A	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ermF* promoter	(3)
pMM104	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Tc^r). *E. coli-C. canimorsus* shuttle plasmid	(3)
pMM106	Δ*siaC*::*ermF*. Cassette for replacement of *siaC* by *ermF.*	(3)
pFD1146	ColE1 *ori*; Sp^r; (Em^r). *E. coli-B. fragilis* expression shuttle plasmid with *lacO₃-PcfxA-lacO₁* promoter	(4)
pBBR-IBA3plus	pBBR1 *ori*, Gm^R; Broad host-range vector allowing expression of genes under control of the *tet* promoter	(14)
pDM291	*Schizosaccharomyces pombe* vector allowing regulatable expression of genes via revtetR-*tup11Δ70*	(15)
Suicide plasmids		
pFL42	Δ*Ccan_22280*::*ermF*. Cassette for replacement of *Ccan_22280* by *ermF* using 7170/7171 and 7172/7173 on gDNA, 7174/7175 on pMM106 and cloned into pMM25 using PstI/SpeI.	This study
Expression plasmids		
pFL31	P*tetQ-lacI* amplified from pFD1146 with primers 7195/7149 and cloned in pMM47.A using SphI/HindIII restriction sites	This study
pFL32	*lacO₃-PcfxA-lacO₁* amplified from pFD1146 with primers 7144/7145 and cloned into pFL31 using SalI/NcoI restriction sites	This study
pFL43	Full length *mucG* with a C-terminal HA tag amplified with primers 7182/7625 and cloned into pPM5 using NcoI/XhoI restriction sites	Chapter 1
pFL157	Full lenght *mucG* with a C-terminal HA tag amplified with primers 7182/7183 and cloned into pFL32 using NcoI/XbaI restriction sites	This study
pFL158	*tetO*-P*cxfA* amplified from pMM47.A using 7242/7243 and cloned into pPM5 using SalI/NcoI restriction sites. Addition of *tetO* sequence (5'-TCCCTATCAGTGATAGAGA-3') upstream of -33	This study

	box	
pFL159	P*ermF-tetR* inserted into pFL158 using BamHI/XmaI restriction sites. P*ermF* was amplified from pMM47.A with primers 7244/7245 and *tetR* was amplified from pBBR-IBA3plus with primers 7247/7249	This study
pFL160	P*ermF-revretR* fusion inserted into pFL158 using BamHI/XmaI restriction sites. P*ermF* was amplified from pMM47.A with 7244/7246 and revTetR was amplified from pDH624 with primers 7248/7250.	This study
pFL161	Full length Ccan_22280 with a C-terminal HA tag amplified with primers 7184/7185 and cloned into pFL159 using NcoI/XbaI restriction sites	This study
pFL162	Full length Ccan_22280 with a C-terminal HA tag amplified from with primers 7184/7185 and cloned into pFL160 using NcoI/XbaI restriction sites	This study
pFL163	*tetO*-P*cxfA-tetO* amplified from pFL158 using primers 7242/7258 and cloned into pFL161 using SalI/NcoI restriction sites. Addition of *tetO* sequence (5'-TCCCTATCAGTGATAGAGA-3') upstream of -33 box and downstream of TIS	This study
pFL164	P*tetQ* promoter and leader sequence amplified from pMM104 using primers 7888/7889 and cloned into pMM47.A using SaI/NcoI restriction sites	This study
pFL165	Full lenght *mucG* 24-28A with a C-terminal HA tag amplified with primer 7182/7183 and cloned into p164 using NcoI/XbaI restriction sites	This study

[a]: Selection markers for *C. canimorsus* are in between brakets

Table S3. Oligonucleotides used in this study

Ref.	**Sequence 5'-3'**	**Restriction[a]**
7144	gggtcgacggcagtgagcgcaacgc	SalI
7145	ggccatggaattgttatccgctcacaattgc	NcoI
7149	gggcatgctcactgcccgctttccagtcgg	SphI
7170	cgctgcagcttctaataatcgtagccaatatatcg	PstI
7171	aaaaatttcatccttcgtagttacgaattctaaattacaaatttcaaattacg	
7172	gagtagataaaagcactgttaagtatgaaatcaattacgaattacg	
7173	ggactagtggtactaatcaacgaaaaatatgg	SpeI
7174	cgtaatttgaaatttgtaatttagaattcgtaactacgaaggatgaaatttttcagg	
7175	cgtaattcgtaattgatttcatacttaacagtgcttttatctactccg	
7182	ggccatggggaaaaaaatagtatccattagc	NcoI
7183	ggtctagactaagcgtaatctggaacatcgtatgggtaaaacgtaacttgagttctc	XbaI
7184	ggccatggggaaaaaaaatattttaacaatgg	NcoI
7185	ggtctagactaagcgtaatctggaacatcgtatgggtattttgaaaccgaccaaacc	XbaI
7195	ccaagcttgaattcccaaaaggtctaaaagtaaattttatcc	HindIII
7242	gggtcgactccctatcagtgatagagattacaaagaaaattcgacaaactg	SalI
7243	ccccatgggccgacaaaggtacataactaaagtttcccaccc	NcoI
7244	ggggatccgctcatcggtatttgcaacatcatagaaattgc	BamHI

7245	cactttacttttatctaaacgagacatcatgtaacttcttacaggtgaatacttcttg	
7246	cactttacttttatctaatctggacatgtaacttcttacaggtgaatacttcttg	
7247	caagaagtattcacctgtaagaagttacatgatgtctcgtttagataaaagtaaagtg	
7248	caagaagtattcacctgtaagaagttacatgtccagattagataaaagtaaagtg	
7249	ggcccgggttaagacccactttcacatttaagttg	SmaI
7250	ggcccgggttatccactttcacatttaagttg	SmaI
7258	ccccatggtctctatcactgatagggagccgacaaaggtacataactaaagtttcccacc	NcoI
7888	atgtcgacacctacgtttccctaataaaatgtctatgg	SalI
7889	atccatggttggattaagcaatatatactacaatagatgc	NcoI

[a]: Restriction sites are underlined

4. A new iron acquisition system in Bacteroidetes

4.1. Manuscript published: A new iron acquisition system in Bacteroidetes

Author contributions: PM, FL, FR, KH, EH and GC conceived and designed the experiments. PM, FL, FR, KH, EH performed the experiments. PM, FL, FR, KH, EH and GC analyzed the data and wrote the paper.

Statement of my work: My contribution was the data of Figures 2, 3D and S3

IAI Journals.ASM.org

New Iron Acquisition System in *Bacteroidetes*

Pablo Manfredi,[a] Frédéric Lauber,[a,b] Francesco Renzi,[a,b] Katrin Hack,[b] Estelle Hess,[b] Guy R. Cornelis[a,b]

Biozentrum der Universität Basel, Basel, Switzerland[a]; Université de Namur, Namur, Belgium[b]

Capnocytophaga canimorsus*, a dog mouth commensal and a member of the *Bacteroidetes* phylum, causes rare but often fatal septicemia in humans that have been in contact with a dog. Here, we show that *C. canimorsus* strains isolated from human infections grow readily in heat-inactivated human serum and that this property depends on a typical polysaccharide utilization locus (PUL), namely, *PUL3* in strain Cc5. PUL are a hallmark of *Bacteroidetes*, and they encode various products, including surface protein complexes that capture and process polysaccharides or glycoproteins. The archetype system is the *Bacteroides thetaiotaomicron* Sus system, devoted to starch utilization. Unexpectedly, *PUL3* conferred the capacity to acquire iron from serotransferrin (STF), and this capacity required each of the seven encoded proteins, indicating that a whole Sus-like machinery is acting as an iron capture system (ICS), a new and unexpected function for Sus-like machinery. No siderophore could be detected in the culture supernatant of *C. canimorsus*, suggesting that the Sus-like machinery captures iron directly from transferrin, but this could not be formally demonstrated. The seven genes of the ICS were found in the genomes of several opportunistic pathogens from the *Capnocytophaga* and *Prevotella* genera, in different isolates of the severe poultry pathogen *Riemerella anatipestifer*, and in strains of *Bacteroides fragilis* and *Odoribacter splanchnicus* isolated from human infections. Thus, this study describes a new type of ICS that evolved in *Bacteroidetes* from a polysaccharide utilization system and most likely represents an important virulence factor in this group.

Capnocytophaga canimorsus is a commensal bacterium from the oral cavity of dogs that is regularly isolated, since its description in 1989, from extremely severe human infections worldwide (1, 2). Following contact with a dog, these infections generally start with vague influenza symptoms, and patients enter the hospital with fulminant septicemia often associated with peripheral gangrene. Mortality is as high as 40% in spite of adequate antibiotherapy and frequent amputations (1, 3–7). Infections do not necessarily occur after severe injuries, which generally are followed by a preventive antibiotic treatment, but rather after small bites, scratches, or even licks (8). Many cases involve splenectomized, alcoholic, or immunocompromised patients, but more than 40% of the cases concern healthy people with no obvious risk factors (5, 8–12), indicating that *C. canimorsus* infections are not restricted to immunocompromised individuals. It is worth noting that there is no report of a dog having been infected by *C. canimorsus*, although 74% of the dogs carry it (13–15). Thus, evolution shaped these bacteria essentially as commensals of the mouth and not as pathogens. Besides *C. canimorsus*, the oral cavity of dogs also harbors *Capnocytophaga cynodegmi* (14), the species most closely related to *C. canimorsus*, with a difference in the 16S RNA sequence only in the range of 1.5% (13). Interestingly, *C. cynodegmi* is not reported to cause human infections (13, 16). Other bacteria from the genus *Capnocytophaga* colonize the oral cavity of diverse mammals, including humans (17, 18). *Capnocytophaga* are fastidious capnophilic (i.e., CO_2 loving) Gram-negative bacteria that belong to the family of *Flavobacteriaceae* in the phylum *Bacteroidetes*. *Flavobacteriaceae* include a variety of environmental and marine bacteria, such as *Flavobacterium johnsoniae* (19), and a few severe animal pathogens, like *Flavobacterium psychrophilum*, the causative agent of cold water disease in salmonid fish (20), and *Riemerella anatipestifer*, which causes duckling disease in waterfowl and turkeys (21, 22). Besides the *Flavobacteriaceae*, the phylum *Bacteroidetes* includes the *Bacteroidaceae*, which contain many anaerobic commensals of the mammalian intestinal flora, such as *Bacteroides thetaiotaomicron* and *Bacteroides fragilis* (23). The phylum *Bacteroidetes* is taxonomically remote from the *Proteobacteria* group, including most studied human pathogens, and the biology of these bacteria reveals a number of original features. One of these features is the presence of many systems resembling the archetypal starch utilization system (Sus) discovered in *B. thetaiotaomicron* (24). The Sus system is a cell envelope-associated multiprotein complex characterized by the coordinated action of several proteins and lipoproteins involved in substrate binding, degradation, and internalization into the periplasm (14, 24–31). Subsequent microbial genome sequencing projects revealed the presence of many polysaccharide utilization loci (PUL) encoding Sus-like systems in the genome of *B. thetaiotaomicron* and other saccharolytic *Bacteroidetes* (26, 31, 32), targeting all major classes of host and dietary glycans (33). The genome of saprophytic *Bacteroidetes* like *F. johnsoniae* also contains a large number of PUL (34), indicating that they are a hallmark of the *Bacteroidetes* phylum rather than of the commensal *Bacteroides* only. The genome of the clinical isolate type strain *C. canimorsus* 5 (also called strain

Received 13 May 2014 **Returned for modification** 16 June 2014
Accepted 26 October 2014

Accepted manuscript posted online 3 November 2014

Citation Manfredi P, Lauber F, Renzi F, Hack K, Hess E, Cornelis GR. 2015. New iron acquisition system in *Bacteroidetes*. Infect Immun 83:300–310. doi:10.1128/IAI.02042-14.

Editor: S. M. Payne

Address correspondence to Guy R. Cornelis, guy.cornelis@unamur.be.

P.M. and F.L. are joint first authors and contributed equally to this work.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/IAI.02042-14.

doi:10.1128/IAI.02042-14

Cc5) (35, 36) contains 13 such PUL, which may encode surface feeding machineries (37). At least 10 of them are expressed, accounting for more than half of the surface-exposed proteins, when Cc5 bacteria are grown on HEK293 cells. All of these findings indicate that surface-exposed complexes specialized in foraging complex glycans or other macromolecules play a central role in the biology of *C. canimorsus* (37). Indeed, *C. canimorsus* has the unusual property of harvesting *N*-linked glycan chains of soluble proteins like immunoglobulins and even of surface glycoproteins from animal cells, including phagocytes. This capacity depends on a Sus-like complex encoded by *PUL5* (38). However, the function of the other PUL is not known yet, and their impact on pathogenicity is unclear. In the present study, we aimed at identifying *C. canimorsus* virulence factors implicated in septicemia, and we demonstrate that *PUL3* encodes a Sus-like system devoted to the acquisition of iron from transferrins, including human serotransferrin (STF).

MATERIALS AND METHODS

Ethics statement. Blood samples from healthy volunteers who had signed a written informed consent were provided by the Blutspendezentrum SRK Beider Basel. The experiments were approved by the Ethikkommission Beider Basel EKBB (no. EK398/11).

Bacterial strains. This study was carried out with *C. canimorsus* strains isolated from human infections (35, 36) and *C. canimorsus* and *C. cynodegmi* strains isolated from dogs in two areas of Switzerland. One strain of *C. cynodegmi* was purchased from the ATCC. *Escherichia coli* S17-1, *Pseudomonas aeruginosa* PAO1, and the *C. canimorsus* mutant strains are described in Table S1 in the supplemental material.

Conventional bacterial growth conditions and selective agents. *C. canimorsus* bacteria were routinely grown on heart infusion agar (Difco) supplemented with 5% sheep blood (Oxoid) (SB plates) for 2 days at 37°C in the presence of 5% CO_2. *Escherichia coli* strains were grown routinely in lysogeny broth (LB) at 37°C. *Pseudomonas aeruginosa* PAO1 (39) was grown on SB plates at 37°C in the presence of 5% CO_2. To select for plasmids, antibiotics were added at the following concentrations: 10 $\mu g \cdot ml^{-1}$ erythromycin, 10 $\mu g \cdot ml^{-1}$ cefoxitin, 20 $\mu g \cdot ml^{-1}$ gentamicin.

Mutagenesis by allelic exchange and *trans*-complementation. Mutagenesis of the Cc5 wild type (wt) was performed as described in reference (40), with slight modifications. Briefly, replacement cassettes with flanking regions spanning approximately 500 bp homologous to regions directly framing targeted genes were constructed with a three-fragment overlapping PCR strategy. First, two PCRs using Phusion polymerase (M0530S; New England BioLabs) were performed on 100 ng of Cc5 genomic DNA with primers for the upstream (oligonucleotides 1.1 and 1.2) and downstream (oligonucleotides 2.1 and 2.2) regions flanking the sequence targeted for deletion. Primers 1.2 and 2.1 included a 20-bp extension at their 5′ extremities corresponding to both ends of the *ermF* gene (including the promoter). The *ermF* resistance gene was amplified from pMM13 with primers 3.1 and 3.2, which included approximately 20-bp extensions for further annealing to amplify homologous regions. All three PCR products were cleaned and then mixed in equal amounts for PCR using Phusion polymerase. The initial denaturation was at 98°C for 2 min, followed by 10 cycles without primers to allow annealing and extension of the overlapping fragments (98°C for 30 s, 50°C for 40 s, and 72°C for 2 min). After the addition of external primers (1.1 and 2.2), the program was continued for 20 cycles (98°C for 30 s, 50°C for 40 s, and 72°C for 2 min 30 s) and finally for 10 min at 72°C. Final PCR products consisted of locus::*ermF* insertion cassettes and were digested with PstI and SpeI for cloning into the appropriate sites of the *C. canimorsus* suicide vector, pMM25 (40). The resulting plasmids were transferred by RP4-mediated conjugative DNA transfer from *E. coli* S17-1 to Cc5 to allow the integration of the insertion cassette. Transconjugants then were selected for the presence of the *ermF* resistance cassette and checked for sensitivity to cefoxitin, indicating the loss of the pMM25 backbone, and the mutated regions were sequenced with primers 1.1 and 2.2. *Trans*-complementation of the different knockouts was done by introducing the relevant genes cloned in the *C. canimorsus* expression vector pPM5. Mutant strains are listed in Table S1 in the supplemental material, primers are in Table S2, and plasmids are in Table S3.

PCR screen for *PUL3*. For PCR screen of *PUL3* genes, strains were grown for 2 days on SB plates, collected, and resuspended in 400 μl phosphate-buffered saline (PBS) at an OD_{600} of 2. Bacterial suspensions then were centrifuged at 6,000 relative centrifugal forces (RCF) for 5 min and resuspended in 400 μl H_2O. Five-μl aliquots of bacterial suspensions then were used in 35-cycle PCRs as described in reference 13. Primers used for the amplification of genes *Ccan_03640* (*icsA*), *Ccan_03650* (*icsC*), *Ccan_03680* (*icsD*), *Ccan_03690* (*icsE*), *Ccan_03700* (*icsF*), *Ccan_03710* (*icsG*), and *Ccan_03720* (*icsH*) are listed in Table S2 in the supplemental material. 16S rRNA genes were amplified as a control (see Table S2).

Sera and protein-depleted serum derivatives. Batches of fresh human blood pooled from 20 individuals were collected at the University Hospital of Basel (Blutspendezentrum). The pooled blood was clotted and centrifuged for 10 min at 6,000 RCF, and the supernatant (serum) was collected for further analyses. Alternatively, human serum collected off the clot from healthy normal humans was purchased from EMD Millipore (S1-liter; Billerica, MA, USA). Serum then was heat inactivated (HIHS) at 55°C for 1 h when required. Protein-depleted human serum (PDHS) was obtained by collecting the flowthrough of 15 ml human serum passed through a single Amicon filter unit with a nominal molecular mass limit of 50 kDa (UFC905024; Millipore) by spinning at 4,000 RCF for 40 min at 20°C. Protein depletion then was monitored by SDS-PAGE (41) and silver staining (42). Transferrin depletion was checked by anti-STF immunoblotting (goat anti-human transferrin; T2027; Sigma-Aldrich). Filtered serum then was heat inactivated as described above.

Growth in heat-inactivated and protein-depleted human sera. Growth assays were performed in 96-well plates. Inocula were prepared from cultures grown on SB plates, set to an OD_{600} of 0.2, and serially diluted 1:10 four times. Twenty-two- and 10-μl bacterial suspensions then were used to inoculate 200 μl of HIHS and 50 μl of PDHS, respectively. HIHS was supplemented with iron (III) chloride ($FeCl_3$; 0.25 mM), iron (III) citrate ($FeC_6H_5O_7$; 0.25 mM), or iron (II) sulfate ($FeSO_4$; 0.25 mM) if required. PDHS was supplemented with iron (III) chloride (0.25 mM), human STF (3 $g \cdot liter^{-1}$; 16-16-032001; Athens Research), human ApoSTF (3 $g \cdot liter^{-1}$; 16-16-A32001; Athens Research), human lactoferrin (1.5 $g \cdot liter^{-1}$; 30-1147; Fitzgerald), bovine STF (3 $g \cdot liter^{-1}$; PRO-510; Prospecbio), hemin (0.25 mM; H9039; Sigma-Aldrich), or hemoglobin (0.1 mM; H7379; Sigma-Aldrich) if required. Equivalent volumes of inocula also were plated in order to precisely determine bacterial concentrations by CFU counting at the inoculation time point. Infections then were incubated statically for 23 h at 37°C in the presence of 5% CO_2. Serial dilutions were plated on SB plates, and CFU were determined. The number of generations was calculated according to the following formula: CFU in the well = inoculum $\times 2^{\text{Number of generations}}$. Cocultures were performed essentially in same the way, except that 200 μl of HIHS was inoculated with both 22 μl of wild-type *C. canimorsus* 5 and 22 μl of the deletion strain set at an OD_{600} of 0.2 and serially diluted 1:10 four times. In addition, serial dilutions following incubation were plated in parallel on SB plates and on SB plates containing erythromycin for the selection of deletion strains. The total growth of the deletion strains corresponded to the CFU on erythromycin-containing plates, while the total growth of the wt was determined by subtracting the CFU counts on erythromycin-containing plates from the CFU counts on SB plates.

Protein concentrations were checked using a Bio-Rad protein assay kit (500-0002; Bio-Rad), and iron concentrations, except for hemoglobin, were checked using the ferrozine assay (43). The iron concentration of hemoglobin was specifically determined using a modified ferrozine assay (44). Protein and iron concentrations are given in Table S4 in the supplemental material.

Monitoring of transcription by real-time RT-PCR. The Cc5 wt was inoculated at a density of 5×10^5 bacteria · ml^{-1} in 15 ml of HIHS at 37°C in the presence of 5% CO_2 with or without 0.25 mM iron (III) citrate. Bacteria were harvested after 6 h (corresponding to the mid-log growth phase) by centrifugation at 7,000 RCF at 4°C for 5 min. The pellet was resuspended in RNAprotect bacterial reagent (76506; Qiagen) and centrifuged again at 5,000 RCF for 10 min. The Δ*furA* deletion strain was grown under the same conditions without the addition of iron. Bacteria were lysed in 200 μl Tris-EDTA (TE) buffer containing proteinase K (60 mAU · ml^{-1}; 19131; Qiagen) and lysozyme (1 mg · ml^{-1}; 10837059001; Roche) for 10 min at 25°C on a shaker. RNA was extracted with the miRNeasy minikit (217004; Qiagen). One ml of QIAzol reagent was heated up to 65°C and added to each sample. Samples were vortexed for 3 min and incubated for 5 min at room temperature. Two hundred μl chloroform was added. The following steps were performed according to the manufacturer's instructions. To remove genomic DNA, an on-column DNase digestion and an additional DNase digestion postextraction were performed using an RNase-free DNase set (79254; Qiagen). RNA was purified with the RNeasy MinElute cleanup kit (74204; Qiagen). RNA integrity was verified by nondenaturing agarose gel electrophoresis (1% agarose [EP-0010-05; Eurogentec] in Tris-acetate-EDTA [TAE]). The absence of genomic DNA was tested by PCR for 16S rRNA. One hundred to 500 ng of RNA was reverse transcribed using Superscript II reverse transcriptase (200 U) (18064-014; Invitrogen) and random primers (100 ng · ml^{-1}) (48190011; Invitrogen) according to the manufacturer's instructions. A no-enzyme control was included for all RNA samples to confirm the absence of genomic DNA. Quantitative PCR (qPCR) was performed using FastStart Universal SYBR Master (Rox) (04913850001; Roche) and primers at 0.3 μM. Primers were designed with NCBI primer-BLAST. Three technical replicates were run for each target and condition. Before performing the actual qPCR, serial cDNA dilutions were amplified, and PCR and primer efficiencies were evaluated by means of a standard curve. All qPCRs were performed on a StepOne machine (Applied Biosystems) using the following thermal cycling conditions: 2 min at 50°C, 10 min at 95°C, 40 cycles 15 s at 95°C, and 1 min at 60°C. Fold change was calculated as described in reference 45, with the $\Delta\Delta C_T$ method (where C_T is threshold cycle) considering the efficiency of the PCR for each target. 16S rRNA served as a reference gene.

Transferrin deglycosylation analyses and lectin stainings. For the assessment of the deglycosylation of STF by Cc5, bacteria were collected from SB plates and resuspended in PBS at an OD_{600} of 1. One hundred microliters of bacterial suspensions then was incubated with 100 μl of a transferrin (16-16-032001; Athens Research) solution (0.2 g · $liter^{-1}$) for 180 min at 37°C. As a negative control, 200 μl of a 1:2-diluted transferrin solution alone was incubated for 180 min at 37°C. Samples then were centrifuged for 5 min at 13,000 RCF, supernatant was collected, and a 12-μl aliquot was loaded in a 12% SDS-PAGE gel. Samples were analyzed by Coomassie brilliant blue R250 (B0149; Sigma) and lectin stainings with *Sambucus nigra* lectin (SNA) according to the manufacturer's recommendations (digoxigenin glycan differentiation kit; 11210238001; Roche). For the deglycosylation of human STF by PNGase F, 9 μl of human STF (2 g · $liter^{-1}$; 16-16-032001; Athens Research) was incubated with 2 μl of either fresh or heat-inactivated (10 min at 75°C) enzyme (P0704L; New England BioLabs) in the presence of 1.2 μl of 10× G7 buffer (B3704; New England BioLabs) for 2 h at 37°C. Deglycosylation then was monitored by immunoblotting and lectin stainings with SNA as described above. For subsequent growth assays, PDHS was supplemented with 4 μl of deglycosylated STF for a minimal final concentration required for growth of 0.1 g · $liter^{-1}$.

Siderophore detection assay. Siderophore production was assayed using a modified chrome azurol S (CAS) procedure (46, 47). CAS reagent was prepared as described in reference 46. In order to reach the same final count, *C. canimorsus* 5 and *Pseudomonas aeruginosa* PAO1 were inoculated at approximately 10^4 and 10^7 bacterial cells, respectively, in 1 ml HIHS in 12-well plates and incubated for 23 h at 37°C in the presence of 5% CO_2. Serial dilutions were plated on SB plates to determine the final growth by CFU counting. Bacterial cells were removed by two successive centrifugation steps at 12,000 RCF for 5 min at 20°C. Supernatants then were dialyzed overnight at 4°C (3,500 molecular weight cutoff [MWCO]; 133110; Spectra/Por Biotech) against 4 ml double-distilled water (ddH_2O) containing 0.02% sodium azide. An uninfected control sample of HIHS was treated in parallel. Finally, dialysates were concentrated for approximately 8 h at 37°C to 200 to 250 μl using a Concentrator plus centrifuge (Eppendorf). Fifty μl of dialysate was mixed with an equal volume of CAS solution in a 96-well plate and incubated for 4 h at 37°C. Absorbance at 630 nm was measured using an xMark microplate spectrophotometer (Bio-Rad) and Microplate Manager 6 software (version 6.0; Bio-Rad), ddH_2O containing 0.02% sodium azide serving as a blank, and uninfected HIHS serving as the reference. All measurements were realized in duplicates. Siderophore production was estimated by comparing the ratio [(A_{630} of sample)/(A_{630} of reference)] of Cc5 and PAO1 dialysates to a desferrioxamine mesylate (Desferal) (252750; Calbiochem) standard curve in ddH_2O containing 0.02% sodium azide.

Uptake of iron from transferrin by *C. canimorsus*. ^{55}Fe-transferrin was prepared according to references 48 and 49. ApoSTF at 1 mg · ml^{-1} in 40 mM Tris-HCl buffer (pH 7.4) containing 2 mM sodium carbonate was mixed with 0.075 μmol of sodium citrate and 0.0075 μmol of $^{55}Fe\text{-}Cl_3$ (Perkin-Elmer) and incubated at room temperature for 30 min. The solution then was transferred into dialysis tubing (6,000 to 8,000 MWCO; 132665; Spectra/Por Biotech) and dialyzed four times against 250 ml 40 mM Tris-HCl buffer (pH 7.4) containing 2 mM sodium carbonate for 16 h. The final protein and iron concentrations were evaluated as described above, and transferrin was found to be 20% iron saturated. Five hundred μl ^{55}Fe-STF (3.25 μM) was mixed with 500 μl of bacterial suspension in RPMI (R8758; Sigma-Aldrich) with 2.5% HIHS set to an OD_{600} of 1. The mixture was incubated statically at 37°C for 24 h, and a control sample was incubated in parallel on ice. Cells then were harvested by centrifugation (6,000 RCF, 3 min), washed four times with 1 ml PBS, and resuspended in a final volume of 1 ml of PBS. The OD_{600} was measured for each sample, and equivalent amounts of bacteria were transferred into scintillation vials. Four ml of scintillation liquid (Ultima Gold; 6013329; Perkin-Elmer) was added, and vials were incubated overnight in the dark. Radioactivity associated with bacteria was quantified with a Beckman LS6500 liquid scintillation counter (Beckman Coulter, Fullerton, CA).

Identification of *PUL3* gene members in the genome of other organisms. *icsA* (*Ccan_03640*; gi|340621142; YP_004739593.1), *icsC* (*Ccan_03650*; gi|340621143; YP_004739594), *Ccan_03660* (gi|340621144; YP_004739595.1), *Ccan_03670* (gi|340621145; YP_004739596.1), *icsD* (*Ccan_03680*; gi|340621146; YP_004739597.1), *icsE* (*Ccan_03690*; gi|340621147; YP_004739598.1), *icsF* (*Ccan_03700*; gi|340621148; YP_004739599.1), *icsG* (*Ccan_03710*; gi|340621149; YP_004739600.1), and *icsH* (*Ccan_03720*; gi|340621150; YP_004739601.1) from *C. canimorsus* 5 were blasted against the nr database and clustered at 70% identity (50). Hits above the threshold (high-scoring segment pair E value of $<10^{-5}$) were aligned with ClustalW (default settings) (51). Alignments were used to build hidden Markov models with HMMER.3 (http://hmmer.org/). Models were calibrated and searched against a local copy of the microbial complete genome database, including approximately 2,100 genomes (NCBI) with HMMER.3 and an E value cutoff of 0.0001. A series of Perl scripts was used to sort the outputs and to count occurrences of complete or partial systems. Occurrences of homologous systems then were reported on an illustrative phylogenetic tree based on 16S rRNA sequences from the ribosomal database project (RDP; http://rdp.cme.msu.edu/index.jsp). All sequences were 1,200 nucleotides long and tagged as good quality according to RDP. Type strains and isolated samples were preferred. At least two sequences per genus were downloaded as alignment files from RDP. Consensus was inferred using EMBOSS (http://www.sanger.ac.uk/Software/EMBOSS). Genus consensuses were aligned with ClustalW (default settings), and phylogenetic analyses were conducted in MEGA4 (52). Evolutionary history was in-

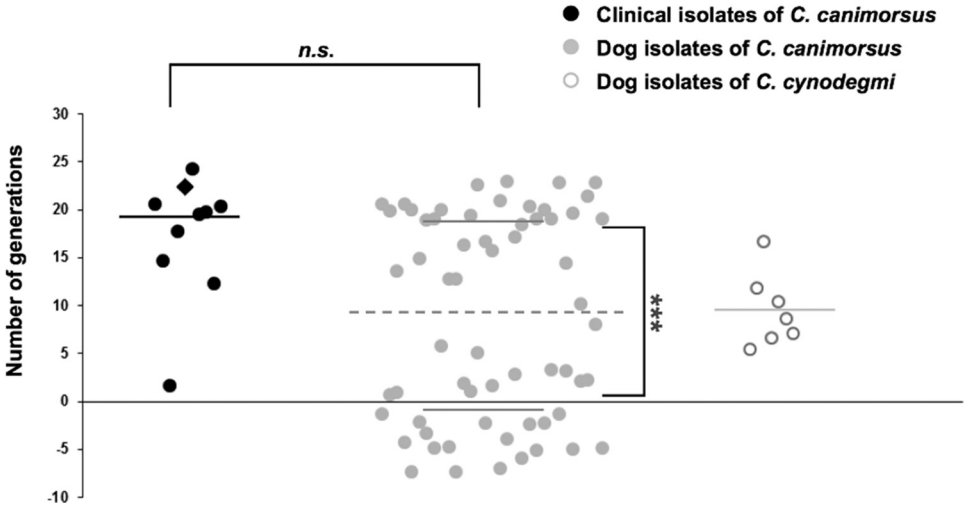


FIG 1 Growth of *C. canimorsus* and *C. cynodegmi* strains in heat-inactivated human serum. The number of generations achieved after 23 h in HIHS for individual *Capnocytophaga* species strains are graphed. Black, clinical isolates of *C. canimorsus*; the diamond shape indicates strain Cc5; gray, dog isolates of *C. canimorsus*; white, dog isolates of *C. cynodegmi*. The best significant expectation-maximization clustering of the *C. canimorsus* dog isolates is reached when clustering the isolates into the two groups (growing and nongrowing) separated by the dotted line. Solid lines indicate the average number of generations for each group (averages from 3 experiments). ***, *t* test error probability below 0.001. The clinical isolates and the growing dog isolates of *C. canimorsus* cannot be discriminated by a *t* test on the sole basis of their growth scores (*n.s.*). The group of *C. cynodegmi* strains displays intermediate growth.

ferred using unweighted-pair group method using average linkages (UPGMA), and evolutionary distances were computed using the maximum composite likelihood method. All positions containing gaps and missing data were eliminated (complete deletion option), leaving a total of 1,143 positions in the final data set. Further searches for *PUL3* genes involved in iron acquisition in organisms absent from the complete genome database were based on PSIBLAST searches with default parameters at the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with two reiterations in total. Only the first 500 hits below an E value of 0.05 were considered. The computations were performed on the CPU cluster of the $[BC]^2$ Basel Computational Biology Center (http://www.bc2.ch/center/index.htm).

Accession numbers for relevant genes and proteins mentioned in the text. The sequences of *icsA* (*Ccan_03640*; gi|340621142; YP_004739593.1), *icsC* (*Ccan_03650*; gi|340621143; YP_004739594), *icsD* (*Ccan_03680*; gi|340621146; YP_004739597.1), *icsE* (*Ccan_03690*; gi|340621147; YP_004739598.1), *icsF* (*Ccan_03700*; gi|340621148; YP_004739599.1), *icsG* (*Ccan_03710*; gi|340621149; YP_004739600.1), and *icsH* (*Ccan_03720*; gi|340621150; YP_004739601.1) were deposited in GenBank previously.

RESULTS

***C. canimorsus* strains isolated from human infections grow readily in heat-inactivated human serum.** While Cc5 bacteria survived in 10% fresh human serum (53), they were killed in 100% fresh human serum (FHS) (data not shown). In contrast, they grew readily in 100% heat-inactivated human serum (HIHS) (Fig. 1), reaching after 23 h a density of about 15×10^9 CFU · ml^{-1} irrespective of the inoculum. In order to assess the relevance of this observation for pathogenesis of human infection, we monitored the growth of 78 different *Capnocytophaga* strains in HIHS. Nine strains of *C. canimorsus* isolated from human infections (referred to as clinical isolates), 62 strains of *C. canimorsus* isolated from dog mouth (dog isolates), and 7 strains of oral canine *C. cynodegmi* were inoculated in HIHS, and colonies were counted after 23 h of incubation. All clinical isolates grew readily, achieving 19 ± 3.7 generations (Fig. 1). In contrast, dog isolates fell into two groups, a first group of strains (31 strains, 50%) performed 18.4 ± 3.2 generations, similar to the clinical isolates, while a second group of 31 strains either did not grow or produced fewer than 8 generations (Fig. 1). The very different proportions of *C. canimorsus* strains able to grow in HIHS among clinical isolates and dog strains strongly suggests that this capacity correlates with pathogenicity and that clinical isolates originate from a subpopulation of dog strains. The strains of *C. cynodegmi* that were tested performed 9.3 ± 4 generations in HIHS (Fig. 1), which is significantly different from both groups of *C. canimorsus* isolated from dogs (*P* values below 10^{-3}). This is somewhat surprising, given that *C. cynodegmi* is not reported to cause systemic human infections. However, the differences between *C. cynodegmi* and the two groups of *C. canimorsus* suggest that several factors can influence the growth of bacteria from this taxon in HIHS.

***PUL3* is crucial for growth in HIHS.** In order to identify the genes underlying the capacity to grow in HIHS, we compared the genomes of Cc5 (35), three additional clinical isolates of *C. canimorsus* (Cc2, Cc11, and Cc12) (36), three *C. canimorsus* dog strains that failed to grow in HIHS (CcD38, CcD93, and CcD95), and three strains of *C. cynodegmi* that displayed moderate growth levels (Ccyn2B, Ccyn49044, and Ccyn74). The genome sequences and their annotations will be described in detail elsewhere. This comparative analysis identified 97 orthologous groups of genes whose presence correlates with the capacity to grow in HIHS. Only 54 of these orthologous clusters included genes with a predicted function. Thirty-eight were involved in a variety of processes, but 16 encoded Sus-like feeding complexes (data not shown). The latter 16 genes belong to only 3 polysaccharide utilization loci, namely *PUL3* (9 genes), *PUL7* (6 genes), and *PUL11* (one gene) (37). Because *PUL* genes represent 16.5% of those differentiating strains that can or cannot grow in HIHS while all of the *PUL* genes represent only about 4% of the Cc5 complete genome, we first tested the *PUL3*, *PUL7*, and *PUL11* knockout mutants for growth in HIHS. In good agreement with the prediction based on genomics, bacteria deprived of the *PUL3* locus were dramatically im-

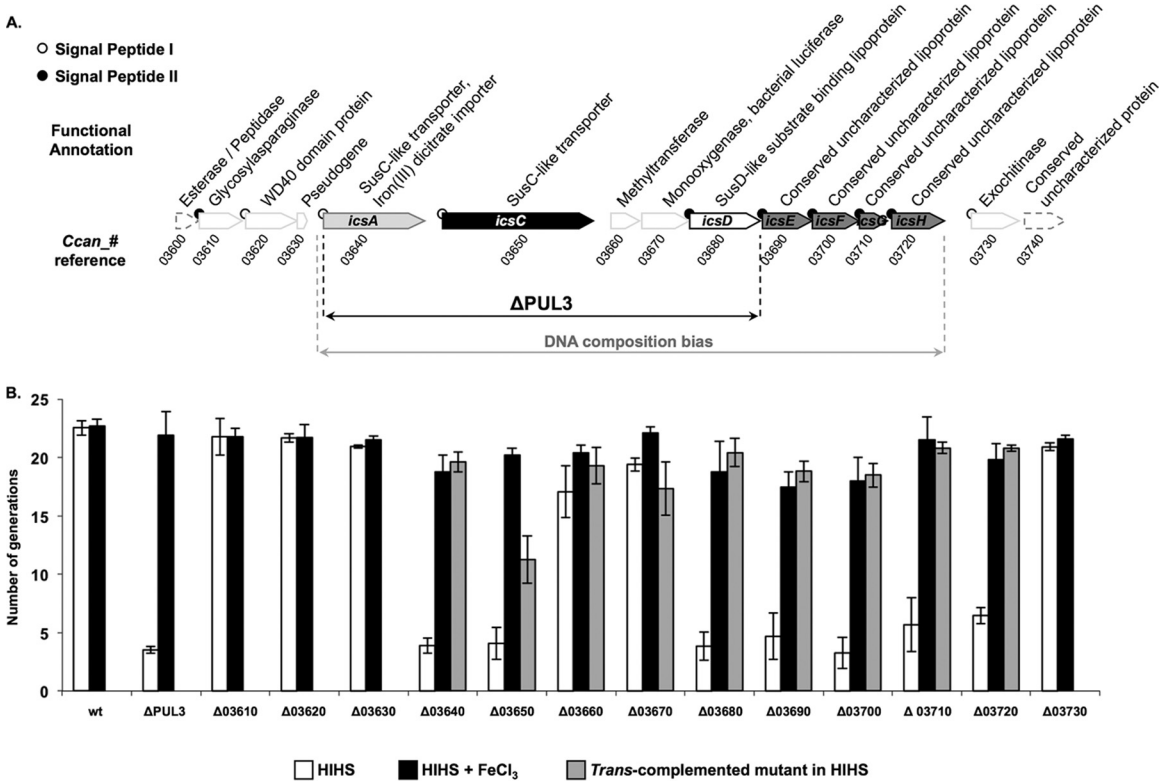


FIG 2 Functional characterization of *PUL3*. (A) Genetic organization and functional annotation of *PUL3*. Genes likely involved in the capture of iron by *C. canimorsus* in human serum are labeled *icsA-H*. Gray-delineated white arrows indicate genes whose deletion had no effect on iron acquisition. The two genes marked by dashed white arrows were not knocked out in this study. White and black circles at the N terminus of the coding sequences indicate that the protein has a type I or type II (lipoprotein) signal peptide, respectively. The numbers under the arrows correspond to the *Ccan_* gene references of strain Cc5. The black double arrow indicates the span of the deletion in the Δ*PUL3* mutant used throughout this study. The gray double arrow shows the range of the region of *PUL3* exhibiting a DNA composition bias with respect to the rest of the chromosome, as computed by Alien_Hunter with a local score of 34.589 (default significance cutoff, 18). (B) Number of generations achieved by the wt and single-gene mutants in HIHS (white bars) and in HIHS supplemented with 0.25 mM iron (III) chloride ($FeCl_3$) (black bars). Gray bars indicate the growth of mutants *trans*-complemented with a plasmid expressing the corresponding deleted gene. Error bars indicate standard deviations (averages from 3 experiments). All differences above 7 generations have *t* test-based error probabilities below 0.008.

paired in their capacity to grow in HIHS, while bacteria deprived of *PUL7* or *PUL11* did not show any significant growth reduction compared to the wt (see Fig. S1 in the supplemental material). We also tested the 10 Cc5 knockout mutants deprived of the other *PUL* genes (37). Not surprisingly, *PUL5* mutants had a moderate growth defect in HIHS compared to the wt (see Fig. S1). This is consistent with the fact that *PUL5* encodes the Gpd glycoprotein deglycosylation system that is essential for aminosugar scavenging (37, 54). The deletion of *PUL1* also led to a moderate growth defect, but this was not investigated further.

***PUL3* has a unique genetic organization compared to the other *PUL* genes of *C. canimorsus* 5.** *PUL3* was annotated as a large locus of 15 genes sharing the same transcriptional orientation (*Ccan_03600* to *Ccan_03740*) (37) (Fig. 2). *PUL3* has two major features that make it different from the other PUL of Cc5. First, it is the only PUL where the *susC*-like gene *Ccan_03650* is separated from the *susD*-like lipoprotein gene by other genes. However, these intervening genes (*Ccan_03660* and *Ccan_03670*) have a functional annotation that is unusual for *PUL* genes (Fig. 2A), suggesting that they have inserted within an ancestral canonical PUL. The second unusual feature of *PUL3* is the presence of two *susC*-like genes instead of a single one (*Ccan_03640* and *Ccan_03650*). *Ccan_03640* is 378 amino acids smaller than *Ccan_03650* and shares some remote similarities with the iron (III) dicitrate transporter FecA of *E. coli* (Uniprot accession number P13036). Significant intergenic regions of around 400 bp frame each *susC*-like gene (Fig. 2A), while in most PUL there is only one large noncoding sequence with promoter activity located upstream from the single *susC* homologue (37). As for most other PUL, the last genes from the putative main operon (*Ccan_03690* to *Ccan_03720*) encode conserved hypothetical lipoproteins for which no function could be assigned (Fig. 2A). Genes at both ends of the locus (*Ccan_03610*, *Ccan_03620*, and *Ccan_03730*) seem to lie outside the putative main operon; nevertheless, their predicted localization and function is compatible with a role in glycan or glycoprotein degradation at the bacterial surface (Fig. 2A). A bias in the DNA K-mer composition, as detected by the Alien_hunter software (55), can be observed from *Ccan_03640* to *Ccan_03720* with respect to the rest of the chromosome (Fig. 2A), suggesting that the central region of *PUL3* has been acquired more recently than the other genes at the periphery.

The *sus*-like genes of *PUL3* are required for iron scavenging in human serum. Since the annotation of *Ccan_03640* pointed to an iron transporter, we tested whether the addition of various iron sources to the HIHS could rescue the growth of the Δ*PUL3* mutant bacteria. When HIHS was supplemented with different iron

salts at a concentration of 250 μM, the growth of Δ*PUL3* mutant bacteria was fully restored to the wt level (data not shown).

In order to investigate whether the whole Sus-like apparatus or only one of the SusC-like proteins was involved in iron uptake, we performed a systematic replacement of each of the 13 genes, ranging from *Ccan_03610* to *Ccan_03730*, by an erythromycin resistance cassette. Interestingly, the substitution of each of the seven typical *PUL* genes had a drastic effect on the growth capacity in HIHS (Fig. 2B). Indeed, the deletion of each of the two *susC* homologs (*Ccan_03640* and *Ccan_03650*), the *susD* homolog (*Ccan_03680*), and each of the four uncharacterized lipoprotein genes (*Ccan_03690*, *Ccan_03700*, *Ccan_03710*, and *Ccan_03720*) reduced the number of generations per 23 h from 22.5 ± 0.8 to an average of 4.5 ± 1.1 (Fig. 2B). As expected, the addition of iron (III) chloride to the HIHS restored the growth capacity of all the mutant strains (19.4 ± 1.7 generations) (Fig. 2B). *Trans*-complementation of the seven individual mutants restored the growth capacity, indicating that each of these genes is involved in the growth process in HIHS (Fig. 2B). These results lead to the conclusion that iron uptake requires not only a putative TonB-dependent outer membrane transporter but also a multiprotein Sus-like complex.

The strains deleted of the two genes with an unusual functional annotation for *PUL* genes (*Ccan_03660* and *Ccan_03670*) and the deletion mutants for upstream (*Ccan_03610*, *Ccan_03620*, and *Ccan_03630*) and downstream (*Ccan_03730*) genes in the locus were able to grow normally in HIHS (Fig. 2B). Thus, the locus encoding the iron capture system (ICS) (gray double arrow in Fig. 2A) is smaller than the whole of *PUL3*, as initially described by Manfredi et al. (37), and corresponds to the genes sharing a similar K-mer bias in their DNA content (55), as mentioned above. We named the seven genes required for iron acquisition *ics*. We called *Ccan_03640* and *Ccan_03650*, the two putative TonB-dependent porins (SusC-like), *icsA* and *icsC*, respectively, and the gene encoding a homolog of *susD* (*Ccan_03680*) was named *icsD*. The four additional lipoproteins were named according to their order in the putative operon of *icsE*, *icsF*, *icsG*, and *icsH* (*Ccan_03690*, *Ccan_03700*, *Ccan_03710*, and *Ccan_03720*, respectively) (Fig. 2A). We suggest limiting *PUL3* to the genes forming an iron capture system that has been acquired at once by horizontal transfer.

***PUL3* expression is regulated by iron and FurA.** If *PUL3* was indeed devoted to iron capture, its expression probably would be modulated by iron. To assess this, we monitored the expression by real-time PCR of three *PUL3* genes (*Ccan_03640*, *Ccan_03650*, and *Ccan_03680*) as representatives of the *PUL3* locus, comparing the expression of these genes in Cc5 bacteria grown in HIHS to those of bacteria grown in HIHS supplemented with iron (III) citrate as a source of free iron. The addition of iron (III) citrate led to a ca. 2-fold decrease in the expression of all three *PUL3* genes, indicating that *PUL3* expression is modulated by the presence of free iron in the serum (see Fig. S2 in the supplemental material).

In many bacteria, the expression of genes involved in iron storage and iron uptake, such as iron channels, as well as transferrin and hemoglobin binding proteins and siderophores, is regulated by the transcriptional regulator FurA. Upon increasing the concentration of free iron, Fe^{2+} cations may bind to FurA, which then activates or represses gene transcription (56). Since the genome of Cc5 encodes a FurA-like protein (*Ccan_15860*), we generated a *furA* deletion mutant. We then quantified *Ccan_03640* (*icsA*), *Ccan_03650* (*icsC*), and *Ccan_03680* (*icsD*) mRNA levels by real-time PCR, in the wt and the *furA* mutant, during growth in HIHS. The expression of *PUL3* genes increased by about 2-fold in the *furA* mutant strain compared to wt levels (see Fig. S2 in the supplemental material). These results suggest that *furA* regulates *PUL3* and reinforces the previous results showing that *PUL3* is modulated by iron.

***PUL3* encodes a system capturing iron from human transferrin.** In order to identify the source of iron exploited by *C. canimorsus* in human serum, we first depleted the HIHS of most of its protein content until the growth of *C. canimorsus* became dependent on the supply of iron (III) chloride. Protein depletion was monitored by silver-stained SDS-PAGEs and mass spectrometry analysis (Fig. 3A). With the exception of small amounts of human serum albumin (Uniprot accession number P02768), only trace amounts of other proteins could be detected in the PDHS. Depletion of STF, the major iron-binding protein in human serum, was confirmed specifically by Western blotting (Fig. 3B). We then tested whether the addition of human serotransferrin could restore growth in this PDHS. As shown in Fig. 3C, human iron-bound STF could restore the growth of wt Cc5 bacteria but not of Δ*PUL3* mutant bacteria. In contrast, when human ApoSTF was used instead of its iron-loaded counterpart, neither wt bacteria nor the Δ*PUL3* bacteria grew. Additionally, we monitored the uptake of iron from ^{55}Fe-loaded transferrin by wt and Δ*PUL3* mutant Cc5 bacteria. As shown in Fig. 3D, over a period of 24 h, wt Cc5 bacteria assimilated around 200-fold more ^{55}Fe at 37°C than on ice, indicating that the capture mechanism is an active mechanism. In good agreement with the previous data, at 37°C, Δ*PUL3* mutant bacteria captured around 80-fold less iron than did wt bacteria. Together, these data demonstrate that *PUL3* encodes an ICS that allows iron scavenging from transferrin.

Given the oral ecology of *C. canimorsus*, we tested whether lactoferrin (LTF), which is abundant in saliva and body fluids, also could serve as an iron source. Like human STF, human LTF could restore the growth of the wt but not of the Δ*PUL3* mutant bacteria in PDHS (Fig. 3E). Since humans are not a natural host for *C. canimorsus*, we suspected that the ICS would not be human specific. Indeed, bovine STF could serve as an iron source in a *PUL3*-dependent manner (Fig. 3E). Despite the broad recognition spectrum among members of the transferrin family, other iron-binding molecules found in the human body, such as hemoglobin or hemin, could not restore the growth defect of wt bacteria in PDHS, indicating that *C. canimorsus* is not able to directly take up heme or to secrete hemophores (Fig. 3E). This suggests that the *PUL3*-encoded system is specific for proteins of the transferrin family.

Iron capture from transferrin does not involve soluble factors. Several attempts to demonstrate the direct binding of transferrin to the ICS turned out to be unsuccessful. Hence, we had to exclude that the ICS could be involved in the synthesis, the release, or the capture of an intermediary siderophore. To do this, we performed a series of cross-feeding experiments between wt and individual *PUL3* gene mutants. We first confirmed that the growth defect of the Δ*PUL3* mutant bacteria in HIHS still could be rescued by the addition of iron in the presence of wt bacteria, indicating that there is no competition between the strains (see Fig. S3A in the supplemental material). We then tested whether the mutants lacking a single *ics* gene could grow in HIHS in the presence of wt bacteria. As shown in Fig. S3B in the supplemental material, the presence of wt bacteria did not allow the growth of

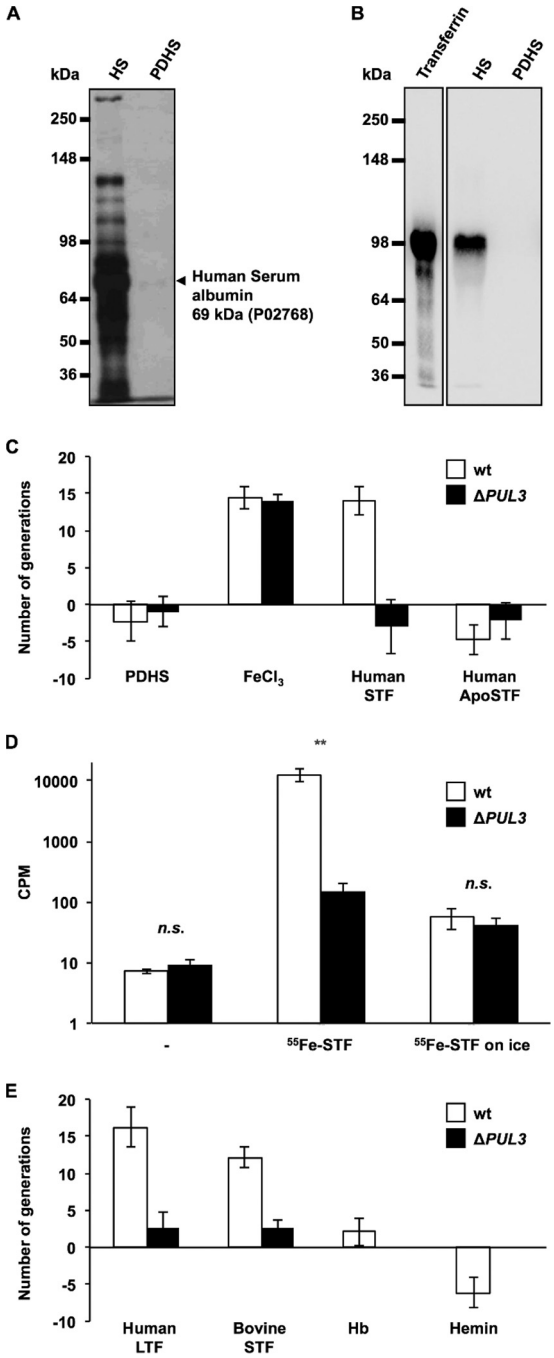


FIG 3 *PUL3* encodes an iron capture system targeting transferrins. (A) Silver-stained SDS-PAGE of normal human serum (HS) (0.1 μl) and protein-depleted human serum (PDHS) (10 μl). The arrow indicates traces of serum albumin as identified by mass spectrometry. Numbers on the left indicate the protein masses of the references in kDa. (B) Anti-transferrin Western blot. The first lane corresponds to purified human serotransferrin (0.3 μg). Lanes two and three were loaded as described for lanes one and two of panel A. Numbers on the left indicate the protein masses of the references in kDa. (C) Number of generations achieved by wt (white bars) and Δ*PUL3* mutant (black bars) bacteria after 23 h in PDHS supplemented with 0.25 mM iron (III) chloride ($FeCl_3$) and human serotransferrin at 3 g · liter^{-1} (human STF) or human apo-serotransferrin at 3 g · liter^{-1} (human ApoSTF). (D) Uptake of iron from transferrin by *C. canimorsus* cells. Number of cpm (counts per minute) measured for wt (white bars) and Δ*PUL3* mutant (black bars) Cc5 bacteria incubated without (−) or with ^{55}Fe-labeled serotransferrin (^{55}Fe-STF) for 24 h at 37°C. Bacteria incubated on ice in the presence of ^{55}Fe-labeled STF serve as the control. Error bars indicate standard deviations (averages from 3 experiments). **, *t* test error probability of <0.01. (E) Number of generations achieved by wt (white bars) and Δ*PUL3* mutant (black bars) bacteria after 23 h in PDHS supplemented with human lactoferrin at 1.5 g · liter^{-1} (human LTF), bovine serotransferrin at 3 g · liter^{-1} (bovine STF), hemoglobin at 0.1 mM (Hb), and hemin at 0.25 mM. Error bars represent standard deviations (averages from 3 experiments). All differences above 9 generations have *t* test-based *P* values below 0.0034.

any *ics* mutant. Thus, we can exclude that *PUL3* gene products serve to export or synthesize a soluble siderophore.

We then examined the genome of Cc5 to detect genes involved in siderophore synthesis. The search included genes encoding the synthesis of enterochelin, vibriobactin, pyochelin, yersiniabactin, mycobactin, corynebactin, bacillibactin, myxochelin A or B, and, more generally, carboxylate, catecholate, and hydroxamate siderophores. No homologs were detected, suggesting that *C. canimorsus* does not produce already-known siderophores, but one cannot exclude that *C. canimorsus* synthesizes a totally new and unknown class of iron-fetching molecules. Therefore, we attempted to detect a siderophore in the concentrated HIHS culture supernatant of Cc5, taking *Pseudomonas aeruginosa* PAO1 (57, 58) as a control. While the chrome azurol technique (46, 47) detected a siderophore in the culture supernatant of PAO1, even after a 10-fold dilution, it gave a negative result for the undiluted Cc5 culture supernatant at a comparable biomass (see Fig. S3C and D in the supplemental material).

Although these observations do not formally rule out that *C. canimorsus* secretes a siderophore that would be captured by the ICS, they make it unlikely.

Iron capture occurs independently of the *N*-glycosylation of transferrin. Since *C. canimorsus* has been shown to deglycosylate *N*-linked glycoproteins through the *PUL5*-encoded GpdG complex (54), we investigated whether the glycosylation state of transferrin plays a role in iron capture. We first monitored the glycosylation state of the protein prior to and after incubation with *C. canimorsus*. Not surprisingly, we observed a strong deglycosylation of the *N*-linked glycan chains of human STF by wild-type *C. canimorsus*, and this deglycosylation turned out to be dependent on *PUL5* (see Fig. S4A and B in the supplemental material). However, deletion of *PUL5* had only a slight effect on growth in HIHS (see Fig. S1), suggesting that the iron capture system is not acting downstream of the *PUL5*-encoded Gpd complex (54). In addition, nondenaturing removal of *N*-linked glycan chains from human STF with a PNGase F treatment prior to PDHS supplementation (Fig. 4A) did not alter iron chelation by STF, as indicated by the low growth level of the Δ*PUL3* mutant, or prevent the ICS activity in the case of wt bacteria (Fig. 4A and B). These observations indicate that *N*-linked glycans of human transferrin do not play any determinant role in the process of iron extraction from STF.

In *C. canimorsus* and *C. cynodegmi*, the capacity to grow in HIHS correlates with the presence of *ics* genes. We mentioned before that among the strains for which the full genome was sequenced, there was a perfect correlation between growth in HIHS and the presence of *PUL3*. We then sought to further validate the hypothesis that growth in HIHS depends on the capacity to acquire iron by testing the effect of iron supplementation on the growth of 15 strains otherwise unable to grow on HIHS. These 15 strains were known to be devoid of *PUL3* because their full ge-

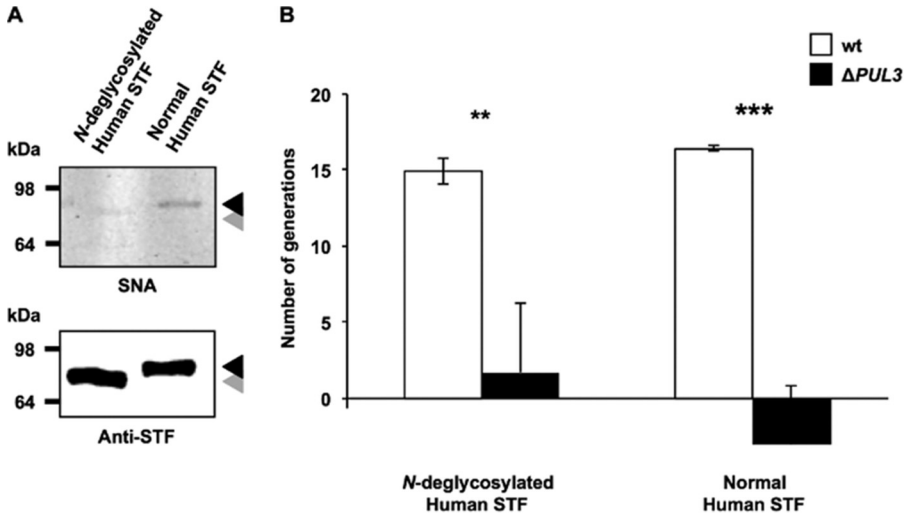


FIG 4 Process of iron capture from STF is independent from *N*-linked glycan chains. (A) *Sambucus nigra* lectin (SNA) staining (top) and anti-serotransferrin immunoblot (bottom) of human serotransferrin (STF) after treatment with fresh (lane 1) and heat-inactivated (lane 2) PNGase F. The black arrow corresponds to the position of the intact protein, while the gray arrow indicates the faint shifted band of the *N*-deglycosylated STF. Numbers on the left indicate the protein mass of the references in kDa. (B) Number of generations achieved by wt (white bars) and *PUL3*-deleted (black bars) bacteria after 23 h in PDHS supplemented with human STF (120 mg · liter^{-1}) treated with either fresh or heat-inactivated PNGase F. Error bars represent standard deviations (averages from 3 experiments). For comparisons to wt values, *t* test-based error probabilities were <0.01 (**) and <0.001 (***), respectively.

nome was sequenced (CcD38, CcD93, and CcD95) or because the individual *ics* genes could not be amplified by PCR (12 strains) (data not shown). As expected, the addition of an excess of free iron strongly enhanced the growth of all of these strains in HIHS (Fig. 5).

In conclusion, growth in HIHS globally correlates with the presence of *PUL3*, and the absence of growth in HIHS correlates with the absence of *PUL3* genes. All of this suggests that the ICS, encoded within the accessory genome of *Capnocytophaga*, is a major factor responsible for iron capture and, by extension, for growth in HIHS.

The complete ICS is found in *Bacteroidetes* species most frequently isolated from human infections. Each of the 9 genes of *PUL3* (*Ccan_03640-Ccan_03720*) was considered to assess the occurrence of the ICS within the bacterial kingdom. Search models for each gene of *PUL3* were built and screened against the complete genome database (http://www.ncbi.nlm.nih.gov/genome/browse/). Out of the 2,100 genomes screened, the two genes which were not involved in the ICS (*Ccan_03660* and *Ccan_03670*) were found in a large taxonomic range and frequently were independent of the occurrence of the other genes of *PUL3* (data not shown). In contrast, the seven *ics* genes were identified only in synteny in the complete genomes of three other *Bacteroidetes* isolated from infected humans: *Bacteroides fragilis* YCH46 (NC_006347), isolated from a human septicemia, *Bacteroides fragilis* NCTC9343 uid57639 (NC_003228), isolated from an abdominal infection, and *Odorib-*

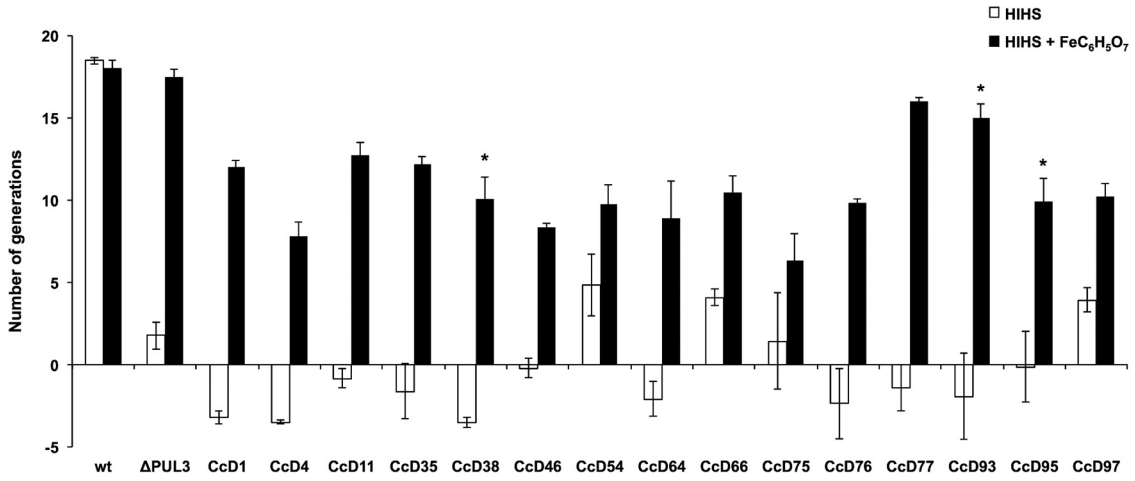


FIG 5 Dog strains unable to grow in HIHS are rescued by the addition of iron. Shown are the number of generations achieved by the wt and Δ*PUL3* isolates, three sequenced dog isolates (CcD95, CcD93, and CcD38), and 12 other randomly picked dog isolates after 23 h in HIHS alone (white bars) or supplemented with 0.25 mM iron (III) citrate ($FeC_6H_5O_7$) (black bars). Error bars indicate standard deviations (averages from 3 experiments). An asterisk indicates sequenced dog strains.

acter splanchnicus DSM20712 uid63397 (NC_015160), isolated from an abdominal abscess (see Fig. S5A in the supplemental material). In addition, *Riemerella anatipestifer* DSM15868 uid60727 (NC_014738), isolated from a duck infectious serositis, also possesses the seven *ics* genes, although the synteny is not entirely conserved (see Fig. S5A).

Additional PSI-BLAST searches for the *ics* genes were carried out against the nonredundant database. The complete set of genes required for the ICS again was exclusively identified in organisms implicated in human or animal infections. These include several *Capnocytophaga* and *Prevotella* species, diverse *Riemerella anatipestifer* isolates, and several additional *Bacteroides fragilis* isolates, *Ornithobacterium rhinotracheale* DSM 15997, *Odoribacter splanchnicus* DSM 20712, and *Porphyromonas* sp. strain F0450, oral taxon 279 (see Fig. S5B in the supplemental material). Thus, the ICS described here is present in a number of *Bacteroidetes* species with pathogenic potential. Interestingly, *PUL3* occurs in bacteria that are able to infect not only mammals but also birds.

DISCUSSION

Here, we showed that nine *C. canimorsus* strains out of nine isolates from human infections grow and survive in HIHS, while only half of the strains isolated from the oral cavity of dogs do so. By genome comparison of representative isolates from groups with distinct growth capacities in HIHS, we could delimit a subset of 97 genes from the *Capnocytophaga* accessory genome potentially involved in growth and survival in human serum. Interestingly, this pool of genes was enriched in genes of the so-called polysaccharide utilization loci of *Bacteroidetes* (16 genes) (30). Out of the 13 *PUL* knockout mutants (37), two of them showed a moderate growth defect, while the deletion of *PUL3* led to a dramatic impairment in the capacity to replicate in HIHS. As suggested by the functional annotation of IcsA, a FecA homologue (59), the *PUL3*-encoded machinery was found to be responsible for the acquisition of iron in human serum. Importantly, iron acquisition did not require IcsA only but also six other *ics*-encoded proteins (IcsC to IcsH). Consistent with its role in iron scavenging in human serum, the *PUL3*-encoded system proved to be essential for fetching iron ions from serotransferrin. Acquisition of iron via heme utilization has been described previously for *Porphyromonas* and *Bacteroides* (60–62); however, in the case of *C. canimorsus*, neither hemin nor hemoglobin was able to rescue iron deprivation in the PDHS, indicating that *C. canimorsus* is not able to directly take up heme or secrete hemophores. Additionally, the hypothesis that *PUL3* is involved in the release of a siderophore was investigated through different approaches and no evidence could be gained, suggesting that iron capture from transferrin does not involve soluble factors. Thus, by analogy with the systems encoded by other PUL, we hypothesize that the iron capture system (ICS) directly interacts with STF, but this could not be formally demonstrated because of the existence of another receptor, still unidentified, that binds many glycoproteins, including STF. Clearly, further work is needed to decipher the mechanism by which the *PUL3*-encoded Sus-like machinery captures iron from transferrin.

Evolutionarily distant from the Tbp or Lbp system of pathogenic *Neisseriaceae* and *Pasteurellaceae* (63) or from the staphylococcal transferrin receptor (64), the *C. canimorsus* ICS initially had been annotated as a polysaccharide utilization system. Indeed, like the canonical starch utilization system (Sus), it consists of SusC and SusD homologs and additional lipoproteins, coencoded within a single putative operon. Despite these classical features of typical polysaccharide-degrading complexes of *Bacteroidetes*, the ICS was shown to function independently from the presence or absence of *N*-linked transferrin glycan moieties. Whether this capture involves some glycan chains of transferrin still needs to be clarified. Another point that requires further clarification is the requirement of the two different SusC-like (putative TonB-dependent porins) proteins IcsA and IcsC.

The presence of a partially conserved *PUL3* devoid of the *fecA*-like transporter gene (*icsA*) in several environmental and plant-associated *Bacteroidetes* spp. suggests the existence of an ancestral version of *PUL3* possibly devoted to a classical carbohydrate substrate. On the other hand, with 341 genome hits, *icsA* is the *ics* gene with the broadest taxonomic occurrence. It can be identified in genomes from diverse taxonomic groups, including *Proteobacteria*, *Spirochetes*, *Bacteroidetes*, or green sulfur bacteria. This contrasts with the occurrence of the other SusC-like gene, *icsC*, which was identified in only 54 genomes, including 48 from the *Bacteroidetes* phylum (data not shown). This taxonomic restriction to the *Bacteroidetes* phylum is typical of *PUL* genes and suggests that *icsA* has been integrated into a classical PUL, which then evolved as a complex iron acquisition system. The other genes of *PUL3* essential for iron capture (*icsD*, *icsE*, *icsF*, *icsG*, and *icsH*) were exclusively identified among *Bacteroidetes*, with *icsG* being found exclusively in genomes including the six other *ics* genes and representing a good marker for the presence of the ICS in other organisms.

Strikingly, the correlation between the occurrence of *PUL3* genes in *C. canimorsus* and the capacity to grow in HIHS strongly suggests a crucial role of the ICS in the process of converting harmless commensal *C. canimorsus* into potential pathogens. The deep compositional DNA bias (55) shared by the genes of *PUL3* (from *Ccan_03640* [*icsA*] to *Ccan_03720* [*icsH*]) with respect to the chromosomal backbone indicates that they were acquired from another organism at the same time. Thus, it is not surprising to repeatedly find a conserved version of *PUL3* in the genome of *Bacteroidetes* species most frequently isolated from human infections (e.g., for many clinical isolates of *B. fragilis*). To our knowledge, this is the first report of a PUL-encoded system serving a purpose other than glycan chain degradation and, by extension, iron acquisition. Besides, the ICS is unique among Gram-negative bacteria in that it can handle a wide range of transferrin isomers, including paralogic (e.g., human STF and human LTF) and orthologic (e.g., human STF and bovine STF) variants, potentially allowing growth in different host environments. This feature consequently explains the taxonomic spread of the ICS among pathogens, which can be considered a key virulence factor of *Bacteroidetes*.

ACKNOWLEDGMENTS

We thank Paul Jenö and Suzanne Moes for their help with mass spectrometry and Simon Ittig, Ulrich Wiesand, and Klaus Handloser for their technical support. We also thank Urs Jenal and Peter Broz for sharing laboratory resources.

This work was supported by grant 3100A0-128659 from the Swiss National Science Foundation and ERC 2011-ADG 20110310 from the European Research Council to G.R.C.

REFERENCES

1. **Bobo RA, Newton EJ.** 1976. A previously undescribed gram-negative *bacillus* causing septicemia and meningitis. Am J Clin Pathol **65:**564–569.

2. **Brenner DJ, Hollis DG, Fanning GR, Weaver RE.** 1989. *Capnocytophaga canimorsus* sp. nov. (formerly *CDC* group *DF-2*), a cause of septicemia following dog bite, and *C. cynodegmi* sp. nov., a cause of localized wound infection following dog bite. J Clin Microbiol **27:**231–235.
3. **Pers C, Gahrn-Hansen B, Frederiksen W.** 1996. *Capnocytophaga canimorsus* septicemia in Denmark, 1982-1995: review of 39 cases. Clin Infect Dis **23:**71–75. http://dx.doi.org/10.1093/clinids/23.1.71.
4. **Le Moal G, Landron C, Grollier G, Robert R, Burucoa C.** 2003. Meningitis due to *Capnocytophaga canimorsus* after receipt of a dog bite: case report and review of the literature. Clin Infect Dis **36:**e42–46. http://dx.doi.org/10.1086/345477.
5. **Janda JM, Graves MH, Lindquist D, Probert WS.** 2006. Diagnosing *Capnocytophaga canimorsus* infections. Emerg Infect Dis **12:**340–342. http://dx.doi.org/10.3201/eid1202.050783.
6. **Westwell AJ, Kerr K, Spencer MB, Hutchinson DN.** 1989. *DF-2* infection. BMJ **298:**116–117.
7. **Bailie WE, Stowe EC, Schmitt AM.** 1978. Aerobic bacterial flora of oral and nasal fluids of canines with reference to bacteria associated with bites. J Clin Microbiol **7:**223–231.
8. **Tierney DM, Strauss LP, Sanchez JL.** 2006. *Capnocytophaga canimorsus* mycotic abdominal aortic aneurysm: why the mailman is afraid of dogs. J Clin Microbiol **44:**649–651. http://dx.doi.org/10.1128/JCM.44.2.649-651.2006.
9. **Lion C, Escande F, Burdin JC.** 1996. *Capnocytophaga canimorsus* infections in human: review of the literature and cases report. Eur J Epidemiol **12:**521–533. http://dx.doi.org/10.1007/BF00144007.
10. **Hantson P, Gautier PE, Vekemans MC, Fievez P, Evrard P, Wauters G, Mahieu P.** 1991. Fatal *Capnocytophaga canimorsus* septicemia in a previously healthy woman. Ann Emerg Med **20:**93–94. http://dx.doi.org/10.1016/S0196-0644(05)81130-8.
11. **Saab M, Corcoran JP, Southworth SA, Randall PE.** 1998. Fatal septicaemia in a previously healthy man following a dog bite. Int J Clin Pract **52:**205.
12. **Deshmukh PM, Camp CJ, Rose FB, Narayanan S.** 2004. *Capnocytophaga canimorsus* sepsis with purpura fulminans and symmetrical gangrene following a dog bite in a shelter employee. Am J Med Sci **327:**369–372. http://dx.doi.org/10.1097/00000441-200406000-00015.
13. **Mally M, Paroz C, Shin H, Meyer S, Soussoula LV, Schmiediger U, Saillen-Paroz C, Cornelis GR.** 2009. Prevalence of *Capnocytophaga canimorsus* in dogs and occurrence of potential virulence factors. Microbes Infect **11:**509–514. http://dx.doi.org/10.1016/j.micinf.2009.02.005.
14. **Suzuki M, Kimura M, Imaoka K, Yamada A.** 2010. Prevalence of *Capnocytophaga canimorsus* and *Capnocytophaga cynodegmi* in dogs and cats determined by using a newly established species-specific PCR. Vet Microbiol **144:**172–176. http://dx.doi.org/10.1016/j.vetmic.2010.01.001.
15. **Umeda K, Hatakeyama R, Abe T, Takakura KI, Wada T, Ogasawara J, Sanada SI, Hase A.** 2014. Distribution of *Capnocytophaga canimorsus* in dogs and cats with genetic characterization of isolates. Vet Microbiol **171:**153–159. 10.1016/j.vetmic.2014.03.023
16. **Blanche P, Bloch E, Sicard D.** 1998. *Capnocytophaga canimorsus* in the oral flora of dogs and cats. J Infect **36:**134.
17. **Kolenbrander PE, Palmer RJ, Jr, Periasamy S, Jakubovics NS.** 2010. Oral multispecies biofilm development and the key role of cell-cell distance. Nat Rev Microbiol **8:**471–480. http://dx.doi.org/10.1038/nrmicro2381.
18. **Frandsen EV, Poulsen K, Kononen E, Kilian M.** 2008. Diversity of *Capnocytophaga* species in children and description of *Capnocytophaga leadbetteri* sp. nov. and *Capnocytophaga* genospecies AHN8471. Int J Syst Evol Microbiol **58:**324–336. http://dx.doi.org/10.1099/ijs.0.65373-0.
19. **McBride MJ.** 2004. *Cytophaga-flavobacterium* gliding motility. J Mol Microbiol Biotechnol **7:**63–71. http://dx.doi.org/10.1159/000077870.
20. **Duchaud E, Boussaha M, Loux V, Bernardet JF, Michel C, Kerouault B, Mondot S, Nicolas P, Bossy R, Caron C, Bessieres P, Gibrat JF, Claverol S, Dumetz F, Le Henaff M, Benmansour A.** 2007. Complete genome sequence of the fish pathogen *Flavobacterium psychrophilum*. Nat Biotechnol **25:**763–769. http://dx.doi.org/10.1038/nbt1313.
21. **Segers P, Mannheim W, Vancanneyt M, De Brandt K, Hinz KH, Kersters K, Vandamme P.** 1993. *Riemerella anatipestifer* gen. nov., comb nov, the causative agent of septicemia anserum exsudativa, and its phylogenetic affiliation within the *Flavobacterium-Cytophaga* rRNA homology group. Int J Syst Bacteriol **43:**768–776. http://dx.doi.org/10.1099/00207713-43-4-768.
22. **Subramaniam S, Huang B, Loh H, Kwang J, Tan HM, Chua KL, Frey J.** 2000. Characterization of a predominant immunogenic outer membrane protein of *Riemerella anatipestifer*. Clin Diagn Lab Immunol **7:**168–174.
23. **Coyne MJ, Chatzidaki-Livanis M, Paoletti LC, Comstock LE.** 2008. Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont *Bacteroides* fragilis. Proc Natl Acad Sci U S A **105:**13099–13104. http://dx.doi.org/10.1073/pnas.0804220105.
24. **Reeves AR, Wang GR, Salyers AA.** 1997. Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. J Bacteriol **179:**643–649.
25. **Cho KH, Salyers AA.** 2001. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol **183:**7224–7230. http://dx.doi.org/10.1128/JB.183.24.7224-7230.2001.
26. **Martens EC, Koropatkin NM, Smith TJ, Gordon JI.** 2009. Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. J Biol Chem **284:**24673–24677. http://dx.doi.org/10.1074/jbc.R109.022848.
27. **Reeves AR, D'Elia JN, Frias J, Salyers AA.** 1996. A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. J Bacteriol **178:**823–830.
28. **Koropatkin NM, Martens EC, Gordon JI, Smith TJ.** 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. Structure **16:**1105–1115. http://dx.doi.org/10.1016/j.str.2008.03.017.
29. **Shipman JA, Berleman JE, Salyers AA.** 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. J Bacteriol **182:**5365–5372. http://dx.doi.org/10.1128/JB.182.19.5365-5372.2000.
30. **Shipman JA, Cho KH, Siegel HA, Salyers AA.** 1999. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. J Bacteriol **181:**7206–7211.
31. **Martens EC, Chiang HC, Gordon JI.** 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe **4:**447–457. http://dx.doi.org/10.1016/j.chom.2008.09.007.
32. **Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI.** 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science **299:**2074–2076. http://dx.doi.org/10.1126/science.1080029.
33. **Bjursell MK, Martens EC, Gordon JI.** 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. J Biol Chem **281:**36269–36279. http://dx.doi.org/10.1074/jbc.M606509200.
34. **McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG, Goltsman E, Wang W, Xu J, Hunnicutt DW, Staroscik AM, Hoover TR, Cheng YQ, Stein JL.** 2009. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. Appl Environ Microbiol **75:**6864–6875. http://dx.doi.org/10.1128/AEM.01495-09.
35. **Manfredi P, Pagni M, Cornelis GR.** 2011. Complete genome sequence of the dog commensal and human pathogen *Capnocytophaga canimorsus* strain 5. J Bacteriol **193:**5558–5559. http://dx.doi.org/10.1128/JB.05853-11.
36. **Shin H, Mally M, Kuhn M, Paroz C, Cornelis GR.** 2007. Escape from immune surveillance by *Capnocytophaga canimorsus.* J Infect Dis **195:**375–386. http://dx.doi.org/10.1086/510243.
37. **Manfredi P, Renzi F, Mally M, Sauteur L, Schmaler M, Moes S, Jeno P, Cornelis GR.** 2011. The genome and surface proteome of *Capnocytophaga canimorsus* reveal a key role of glycan foraging systems in host glycoproteins deglycosylation. Mol Microbiol **81:**1050–1060. http://dx.doi.org/10.1111/j.1365-2958.2011.07750.x.
38. **Mally M, Shin H, Paroz C, Landmann R, Cornelis GR.** 2008. *Capnocytophaga canimorsus*: a human pathogen feeding at the surface of epithelial cells and phagocytes. PLoS Pathog **4:**e1000164. http://dx.doi.org/10.1371/journal.ppat.1000164.
39. **Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV.** 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature **406:**959–964. http://dx.doi.org/10.1038/35023079.
40. **Mally M, Cornelis GR.** 2008. Genetic tools for studying *Capnocytophaga*

canimorsus. Appl Environ Microbiol **74:**6369–6377. http://dx.doi.org/10.1128/AEM.01218-08.
41. **Schagger H, von Jagow G.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem **166:**368–379. http://dx.doi.org/10.1016/0003-2697(87)90587-2.
42. **Rabilloud T, Carpentier G, Tarroux P.** 1988. Improvement and simplification of low-background silver staining of proteins by using sodium dithionite. Electrophoresis **9:**288–291. http://dx.doi.org/10.1002/elps.1150090608.
43. **Valcour AA, Krzymowski G, Onoroski M, Bowers GN, Jr, McComb RB.** 1990. Proposed reference method for iron in serum used to evaluate two automated iron methods. Clin Chem **36:**1789–1792.
44. **Riemer J, Hoepken HH, Czerwinska H, Robinson SR, Dringen R.** 2004. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. Anal Biochem **331:**370–375. http://dx.doi.org/10.1016/j.ab.2004.03.049.
45. **Pfaffl MW.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res **29:**e45. http://dx.doi.org/10.1093/nar/29.9.e45.
46. **Alexander DB, Zuberer DA.** 1991. Use of chrome azurol S reagents to evaluate siderophore production by rhizosphere bacteria. Biol Fertil Soils **12:**39–45. http://dx.doi.org/10.1007/BF00369386.
47. **Schwyn B, Neilands JB.** 1987. Universal chemical assay for the detection and determination of siderophores. Anal Biochem **160:**47–56. http://dx.doi.org/10.1016/0003-2697(87)90612-9.
48. **Grenier D, Tanabe S.** 2011. Transferrin as a source of iron for *Campylobacter rectus.* J Oral Microbiol **2011:**3. http://dx.doi.org/10.3402/jom.v3i0.5660.
49. **Goulet V, Britigan B, Nakayama K, Grenier D.** 2004. Cleavage of human transferrin by *Porphyromonas gingivalis* gingipains promotes growth and formation of hydroxyl radicals. Infect Immun **72:**4351–4356. http://dx.doi.org/10.1128/IAI.72.8.4351-4356.2004.
50. **Li W, Jaroszewski L, Godzik A.** 2002. Tolerating some redundancy significantly speeds up clustering of large protein databases. Bioinformatics **18:**77–82. http://dx.doi.org/10.1093/bioinformatics/18.1.77.
51. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673–4680. http://dx.doi.org/10.1093/nar/22.22.4673.
52. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:**1596–1599. http://dx.doi.org/10.1093/molbev/msm092.
53. **Shin H, Mally M, Meyer S, Fiechter C, Paroz C, Zaehringer U, Cornelis GR.** 2009. Resistance of *Capnocytophaga canimorsus* to killing by human complement and polymorphonuclear leukocytes. Infect Immun **77:**2262–2271. http://dx.doi.org/10.1128/IAI.01324-08.
54. **Renzi F, Manfredi P, Mally M, Moes S, Jeno P, Cornelis GR.** 2011. The N-glycan glycoprotein deglycosylation complex (Gpd) from *Capnocytophaga canimorsus* deglycosylates human IgG. PLoS Pathog **7:**e1002118. http://dx.doi.org/10.1371/journal.ppat.1002118.
55. **Vernikos GS, Parkhill J.** 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. Bioinformatics **22:**2196–2203. http://dx.doi.org/10.1093/bioinformatics/btl369.
56. **Carpenter BM, Whitmire JM, Merrell DS.** 2009. This is not your mother's repressor: the complex role of fur in pathogenesis. Infect Immun **77:**2590–2601. http://dx.doi.org/10.1128/IAI.00116-09.
57. **Cornelis P.** 2010. Iron uptake and metabolism in *pseudomonads.* Appl Microbiol Biotechnol **86:**1637–1645. http://dx.doi.org/10.1007/s00253-010-2550-2.
58. **Cornelis P, Dingemans J.** 2013. Pseudomonas aeruginosa adapts its iron uptake strategies in function of the type of infections. Front Cell Infect Microbiol **3:**75. http://dx.doi.org/10.3389/fcimb.2013.00075.
59. **Hussein S, Hantke K, Braun V.** 1981. Citrate-dependent iron transport system in *Escherichia coli* K-12. Eur J Biochem **117:**431–437. http://dx.doi.org/10.1111/j.1432-1033.1981.tb06357.x.
60. **Olczak T, Sroka A, Potempa J, Olczak M.** 2008. *Porphyromonas gingivalis* HmuY and HmuR: further characterization of a novel mechanism of heme utilization. Arch Microbiol **189:**197–210. http://dx.doi.org/10.1007/s00203-007-0309-7.
61. **Otto BR, Kusters JG, Luirink J, de Graaf FK, Oudega B.** 1996. Molecular characterization of a heme-binding protein of *Bacteroides fragilis* BE1. Infect Immun **64:**4345–4350.
62. **Otto BR, Sparrius M, Wors DJ, de Graaf FK, MacLaren DM.** 1994. Utilization of haem from the haptoglobin-haemoglobin complex by *Bacteroides fragilis*. Microb Pathog **17:**137–147. http://dx.doi.org/10.1006/mpat.1994.1060.
63. **Fuller CA, Yu R, Irwin SW, Schryvers AB.** 1998. Biochemical evidence for a conserved interaction between bacterial transferrin binding protein A and transferrin binding protein B. Microb Pathog **24:**75–87. http://dx.doi.org/10.1006/mpat.1997.0174.
64. **Modun B, Morrissey J, Williams P.** 2000. The staphylococcal transferrin receptor: a glycolytic enzyme with novel functions. Trends Microbiol **8:**231–237. http://dx.doi.org/10.1016/S0966-842X(00)01728-5.

4.2.Supplemental materials

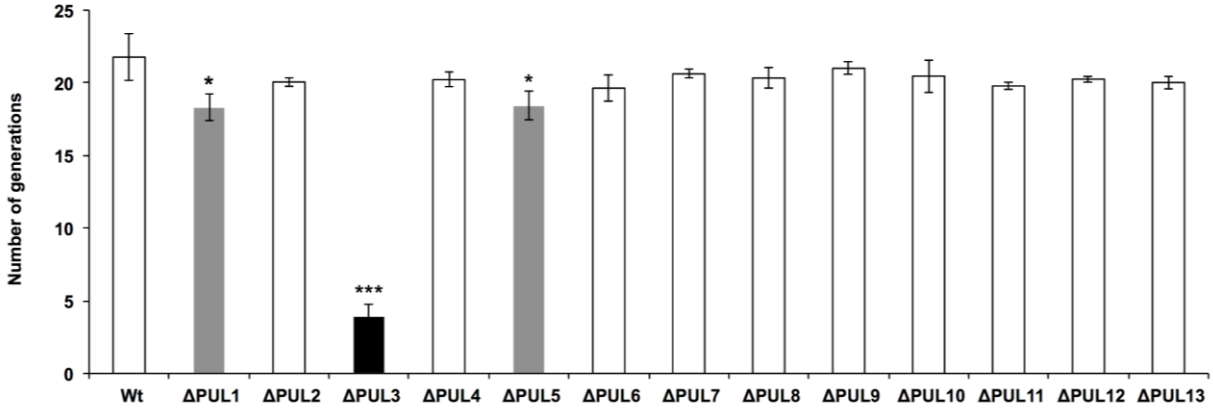


Figure S1. Growth in HIHS of *C. canimorsus 5* mutants deleted from individual PULs
Number of generations achieved by each of the individual *PUL* deletion mutant after 23 hours in HIHS. The black bar and to a lesser extent the dark grey bars indicate significantly reduced growth scores with respect to wt. Error bars indicate standard deviations (average of 3 experiments). (*) and (***) apply to comparisons to wt values and stand for *t*-test based error probabilities of <0.05 and <0.001 respectively.

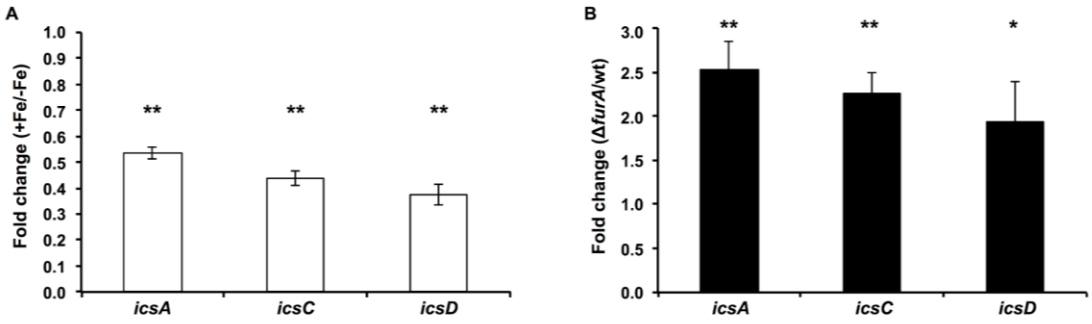


Figure S2. Regulation of transcription of *PUL3* by free iron and the FurA transcriptional regulator
Fold change of mRNA levels of the two *susC* homologues *Ccan_03640* and *Ccan_03650* and the *susD* homologue *Ccan_03680*. (A) Relative mRNA levels from wt bacteria grown in HIHS plus iron (III) citrate (+Fe) *vs.* HIHS with no iron supplementation (-Fe). (B**)** Relative mRNA levels of *ΔfurA vs*. wt *Cc*5 grown in HIHS. Error bars represent the standard deviation (average of 3 experiments). (*) and (**) apply to comparisons to wt values and stand for *t*-test based error probabilities of <0.05 and <0.01 respectively.

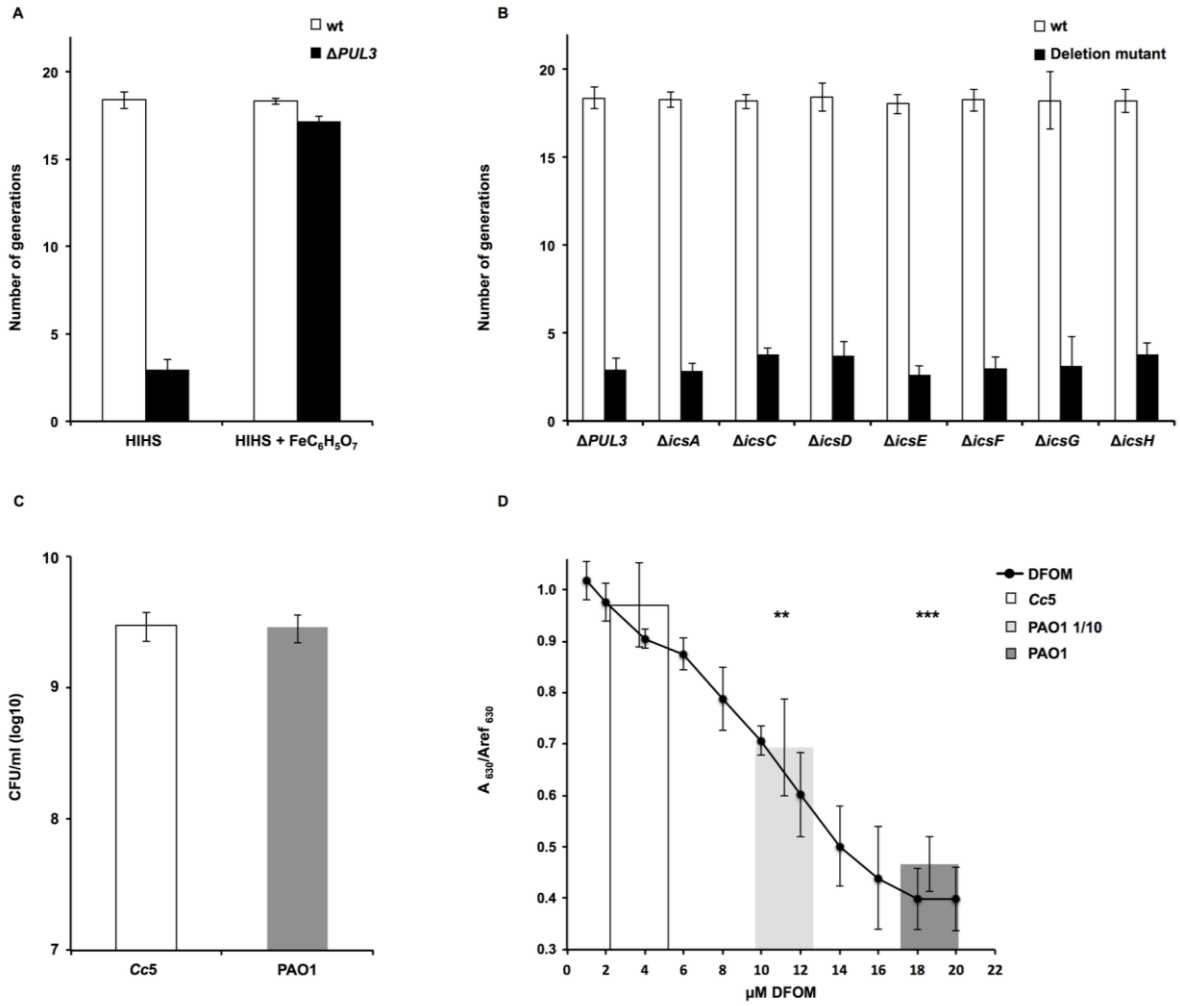


Figure S3. Iron capture from transferrin does not involve soluble factors

(A) Number of generations achieved by wt (white bars) and *ΔPUL3* (black bars) *C. canimorsus* bacteria after 23 hours of co-culture in HIHS alone or supplemented with 0.25 mM iron (III) citrate ($FeC_6H_5O_7$). (B) Number of generations achieved by the wt (white bars) and each of the individual *ics* deletion mutants (black bars) after 23 hours of co-culture in HIHS. (C) Total growth in HIHS of *P. aeruginosa* PAO1 wt (dark grey bar) and *Cc*5 wt (white bar). (D) Siderophore detection in *Cc5* wt (white bar) and *P. aeruginosa* PAO1 wt (light and dark grey bars) HIHS culture supernatants using the chrome azurol S assay. Decrease of the ratio A/A_{ref} at 630 nm indicates presence of siderophore. Dots: dilution series of the iron chelator desferrioxamine mesylate (DFOM) used as standard curve. PAO1 1/10: PAO1 supernatant diluted 1 to 10 in ddH_2O. Error bars in all panels represent standard deviation (average of 3 experiments). (**) and (***) apply to comparisons to *Cc5* wt values and stand for *t*-test based error probabilities of <0.01 and <0.001 respectively.

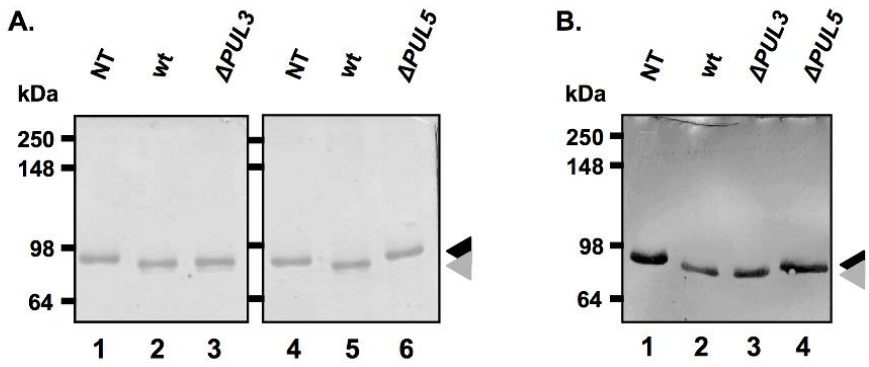


Figure S4. Cleavage of human serotransferrin *N*-glycans by the Gpd complex
(A) Coomassie staining of SDS-PAGE loaded with STF incubated in absence (NT, lane 1 & 4) or presence of wt (lane 2 & 5), *PUL3* deleted (lane 3) or *PUL5* deleted (lane 6) *C. canimorsus*. (B) *Sambucus Nigra* Lectin (SNA) staining of human STF incubated in absence (NT, lane 1) or presence of wt (lane 2), *PUL3* deleted (lane 3) or *PUL5* deleted (lane 4) *C. canimorsus*. Numbers on the left indicate protein mass of the references in kDa. Grey and black arrows indicate a shift in electrophoretic mobility of STF.

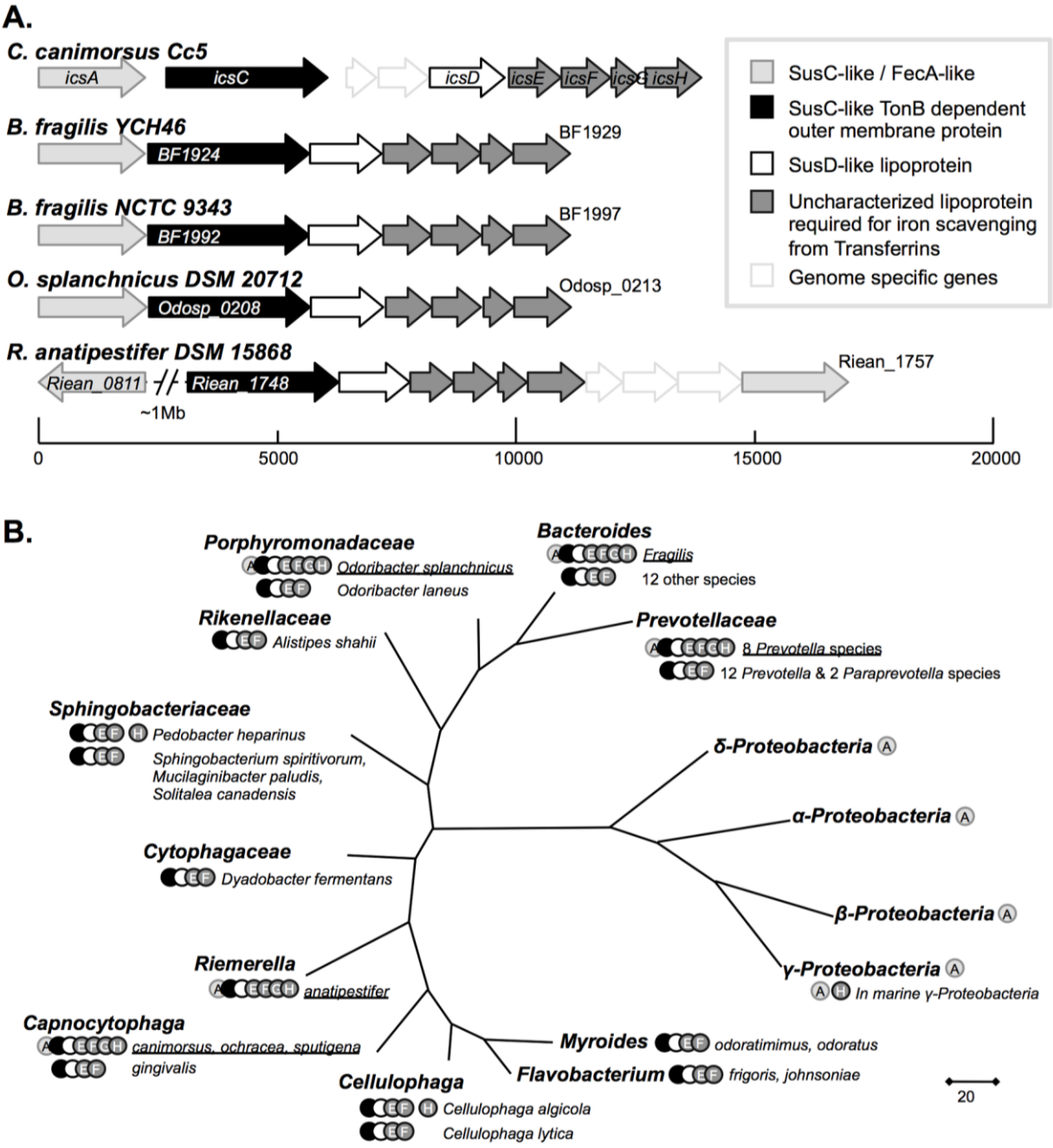


Figure S5. The ICS is mostly found among pathogenic members from the Bacteroidetes phylum and has broad species specificity

(A) Orthologous *PULs* of *Cc5 PUL3* identified in the Complete Genomes database. The *PULs* display has been limited to the genes implicated in iron scavenging with a putative TonB-dependent outer membrane protein ortholog to either *icsA* or *icsC*. For the sake of readability, only the largest SusC like homologs and the last represented genes are tagged here. In the case of *R. anatipestifer* DSM 15868, an additional FecA-like protein that shares higher similarity with other orthologs than with its paralog is found at approximately 1 Mb from the SusC like protein. (B) Non-exhaustive occurrences of *PUL3* genes among other bacteria spotted on a representative 16S rRNA phylogenetic tree. The evolutionary history has been inferred using the Maximum Parsimony method and the consensual 16S rRNA sequences from the different taxa where genes from *PUL3* were found. Spots indicate major combinations of *PUL3* encoded genes found within each taxon. Letters in the round spots refer to the proteins encoded by *PUL3*, namely IcsA (A),

IcsC (C), IcsD (D), IcsE (E), IcsF (F), IcsG (G) and IcsH (H). Underlined taxon names indicate the occurrence of strains exhibiting all seven *ics* genes in their genome.

Table S1 Bacterial strains used in the study

Strain	Description	Reference
E. coli		
S17-1	*hsdR*17 *recA*1 RP4-2-*tet*::Mu-1*kan*::Tn7 Sm^R	(65)
P. aeruginosa		
PAO1	Wild-type *P. aeruginosa*	(39)
C. canimorsus		
*Cc*5	Wild type (BCCM-LMG 28512)	(36)
Δ*PUL1*	Substitution of *PUL1* by *ermF*; Em^R	(37)
Δ*PUL2*	Substitution of *PUL2* by *ermF*; Em^R	(37)
Δ*PUL3*	Substitution of *PUL3* by *ermF*; Em^R	(37)
Δ*PUL4*	Substitution of *PUL4* by *ermF*; Em^R	(37)
Δ*PUL5*	Substitution of *PUL5* by *ermF*; Em^R	(37)
Δ*PUL6*	Substitution of *PUL6* by *ermF*; Em^R	(37)
Δ*PUL7*	Substitution of *PUL7* by *ermF*; Em^R	(37)
Δ*PUL8*	Substitution of *PUL8* by *ermF*; Em^R	(37)
Δ*PUL9*	Substitution of *PUL9* by *ermF*; Em^R	(37)
Δ*PUL10*	Substitution of *PUL10* by *ermF*; Em^R	(37)
Δ*PUL11*	Substitution of *PUL11* by *ermF*; Em^R	(37)
Δ*PUL12*	Substitution of *PUL12* by *ermF*; Em^R	(37)
Δ*PUL13*	Substitution of *PUL13* by *ermF*; Em^R	(37)
Δ*Ccan_03610*	Substitution of *Ccan_03610* by *ermF* using pFL1; Em^R	This study
Δ*Ccan_03620*	Substitution of *Ccan_03620* by *ermF* using pFL2; Em^R	This study
Δ*Ccan_03630*	Substitution of *Ccan_03630* by *ermF* using pFL3; Em^R	This study
Δ*Ccan_03640*	Substitution of *Ccan_03640* by *ermF* using pFL4; Em^R	This study
Δ*Ccan_03650*	Substitution of *Ccan_03650* by *ermF* using pFL5; Em^R	This study
Δ*Ccan_03660*	Substitution of *Ccan_03660* by *ermF* using pFL6; Em^R	This study
Δ*Ccan_03670*	Substitution of *Ccan_03670* by *ermF* using pFL7; Em^R	This study
Δ*Ccan_03680*	Substitution of *Ccan_03680* by *ermF* using pFL8; Em^R	This study
Δ*Ccan_03690*	Substitution of *Ccan_03690* by *ermF* using pFL9; Em^R	This study
Δ*Ccan_03700*	Substitution of *Ccan_03700* by *ermF* using pFL10; Em^R	This study
Δ*Ccan_03710*	Substitution of *Ccan_03710* by *ermF* using pFL11; Em^R	This study
Δ*Ccan_03720*	Substitution of *Ccan_03720* by *ermF* using pFL12; Em^R	This study
Δ*Ccan_03730*	Substitution of *Ccan_03730* by *ermF* using pFL13; Em^R	This study
Δ*Ccan_15860*	Substitution of *Ccan_15860* (*furA*) by *ermF* using pFL61; Em^R	This study

Table S2 Primers used in the study

Ref.	Sequence 5'-3'	Restriction[a]
6953	cgctgcaggctacctatatgatggagcc	PstI
6954	aaaaatttcatccttcgtagaaaaacttcttacgatttttatttag	
7027	gagtagataaaagcactgtttagggacaggacgtg	
7028	ggactagtatccgtctgtgccaataccc	SpeI
6957	ctaaataaaaatcgtaagaagtttttctacgaaggatgaaatttttcagggacaac	
6958	ggacaggacacgtcctgtccctaaacagtgcttttatctactccgatagcttc	
6959	cgctgcagtttacgagcaggacatcc	PstI
6960	aaaaatttcatccttcgtagaaatgataatctttg	
7050	gagtagataaaagcactgttcacttggttacaacgttcc	
7051	ggactagtatccgagtgttttctacc	SpeI
6963	caaagattatcatttctacgaaggatgaaatttttcagggacaac	
7052	ggaacgttgtaaccaagtgaacagtgcttttatctactccgatagcttc	
6965	cgctgcagccaaaacagtttacattgacgg	PstI
6966	aaaaatttcatccttcgtagtctctactatttcctatttttttac	
6967	gagtagataaaagcactgttaataacaatatataaaaatagaatag	
6968	ggactagtacccaaatagcggaaagg	SpeI
6969	gtaaaaaataggaaatagtagagactacgaaggatgaaatttttcagggacaac	
6970	ctattctatttttatatattgttattaacagtgcttttatctactccgatagcttc	
6971	cgctgcagaaatcagtgggaagtaaccgc	PstI
6972	aaaaatttcatccttcgtagttttatgttctttcttgtag	
6973	gagtagataaaagcactgtttttttttagtatttgcccaacg	
6974	ggactagtttttccgttccgtaaggttctgccc	SpeI
6975	ctacaagaaagaacataaaactacgaaggatgaaatttttcagggacaac	
6976	cgttgggcaaatactaaaaaaaacagtgcttttatctactccgatagcttc	
6977	cgctgcagattgggggagagcctcgtgc	PstI
6978	aaaaatttcatccttcgtagatcatctgatatttttattatttgatttgatgc	
6979	gagtagataaaagcactgtttttgtaaggaagggacgtgtcc	
6980	ggactagtccttctcatcgaaattattgacatcg	SpeI
6981	gcatcaaatcaaataataaaaatatcagatgatctacgaaggatgaaatttttcagggacaac	
6982	ggacacgtcccttccttacaaaaacagtgcttttatctactccgatagcttc	
7527	ggctgcaggatttgtacgtaaccaatgtgcttttcacc	PstI
7528	gttgcaaataccgatgagcgattatttttttattttaagcggaaaggacacg	
7529	cctgaaaaatttcatccttcgtagaatatgaaaaaatatcttattctgttggc	
7530	ccactagtgtattcacgagcgggttcaatagaattagtgg	SpeI
7531	cgtgtcctttccgcttaaaataaaaaataatcgctcatcggtatttgcaac	
7532	gccaacagaataagatatttttttcatattctacgaaggatgaaatttttcagg	
7533	ggctgcagcgcaagacttctgattgtacaagagaccg	PstI
7534	gttgcaaataccgatgagcattttattgatttacgtatgatttaagtcgc	
7535	cctgaaaaatttcatccttcgtagaaacaagctaaaaaataatatgac	
7536	ccactagtccatcttttgaaacggctgagatacttgc	SpeI

7537	gcgacttaaatcatacgtaaatcaataaaatgctcatcggtatttgcaac	
7538	gtcatattatttttttagcttgtttctacgaaggatgaaatttttcagg	
6995	cgctgcagcagaaaataatgttcagaaagc	PstI
6996	aaaaatttcatccttcgtagattatttttttagcttgtttctatttgtc	
6997	gagtagataaaagcactgttacgtgttggaatgacagcgg	
6998	ggactagtttcctgcaatcgcacttgatac	SpeI
6999	gacaaatagaaacaagctaaaaaataatctacgaaggatgaaatttttcagggacaac	
7000	ccgctgtcattccaacacgtaacagtgcttttatctactccgatagcttc	
7001	cgctgcagtccattgataatcagcgagag	PstI
7002	aaaaatttcatccttcgtagatttttttctgtttgtaagaacaagaatcgcc	
7003	gagtagataaaagcactgttagtaaaaaggattttcttttc	
7004	ggactagtctcctttgaagagaggaagcc	SpeI
7005	ggcgattcttgttcttacaaacagaaaaaatctacgaaggatgaaatttttcagggacaac	
7006	gaaaagaaaatccttttttactaacagtgcttttatctactccgatagcttc	
7059	cgctgcagaatactctatttacacgg	PstI
7060	aaaaatttcatccttcgtagttttataaattttggtg	
7061	gagtagataaaagcactgtttttttttgaaactgtcatttgg	
7062	ggactagtaagttgcccaatttctgc	SpeI
7063	caccaaaatttataaaactacgaaggatgaaatttttcagggacaac	
7064	ccaaatgacagtttcaaaaaaaacagtgcttttatctactccgatagcttc	
7065	cgctgcagattagtatgttggcattgg	PstI
7066	aaaaatttcatccttcgtagaatattttctttaaagtatgatc	
7067	gagtagataaaagcactgttaatttgtttttttatcttacaatc	
7068	ggactagtttgagacagagtaaaagc	SpeI
7069	gatcatactttaaagaaaatattctacgaaggatgaaatttttcagggacaac	
7070	gattgtaagataaaaaacaaattaacagtgcttttatctactccgatagcttc	
7085	cgctgcagatagggtttatccctgctggggaagg	PstI
7086	aaaaatttcatccttcgtagctctttttctatttatatctg	
7087	gagtagataaaagcactgttaatctgtataaaaatgc	
7088	ggactagtcatcgcgaggatgaagcaaaatataatcc	SpeI
7089	cagatataaatagaaaaagagctacgaaggatgaaatttttcagggacaac	
7090	gcatttttatacagattaacagtgcttttatctactccgatagcttc	
7096	cgctgcagacgctgataccagattgattgattttcaaacagg	PstI
7097	aaaaatttcatccttcgtagatttcaatacttatcatttgtttttaatgc	
7098	gagtagataaaagcactgttgcatcaatcagctacaaccaaaaatcc	
7099	ggactagttacttccgagtatttggttggc	SpeI
7100	gcattaaaaacaaatgataagtattgaaatctacgaaggatgaaatttttcagggacaac	
7101	ggatttttggttgtagctgattgatgcaacagtgcttttatctactccgatagcttc	
7102	cgctgcagggaaatttggataaatacaataatg	PstI
7103	gagtagataaaagcactgttctgcttggtgttttctttttag	
7104	gaaaaatttcatccttcgtagccaagatggcagtagatttattac	
7105	ggactagtattggcaaggttacgataacg	SpeI
7106	ctaaaaaagaaaacaccaagcagaacagtgcttttatctactc	
7107	gtaataaatctactgccatcttggctacgaaggatgaaatttttc	
7036	cgtaccatggcgtgttaccaaaagatagg	NcoI

7037	tgactagttaaaacttcacattcactcc	SpeI
7038	cgtaccatggcgaatcaatcaatgataaagaaactactatatagcg	NcoI
7039	tgactagttaaaacccaacatttacc	SpeI
7040	cgtaccatggcgcccaacgaaagagcatcaaatc	NcoI
7041	tgactagttatcttggattgggtgctaaacc	SpeI
7042	cgtaccatggcgagaagaatatacataatattaacattgg	NcoI
7043	tgactagttattgatttacgtatgatttaagtcgc	SpeI
7044	cgtaccatggcgaaaaaatatcttattctgttggc	NcoI
7045	tgactagttatttgtcaactatttcagc	SpeI
7046	cgtaccatggcgacaatgaatagaaaatatttatttttgataatattactgggg	NcoI
7047	tgactagttatggcaaaataatatactcgc	SpeI
7048	cgtaccatggcgtggaaatatacagttttaatagtgtccc	NcoI
7049	tgactagttatctttttatatcattgattgaaattccg	SpeI
7077	cgtaccatggatagtcatatttggatagaaaagtgg	NcoI
7078	tgactagttattgcttccgtgctacaaatcgg	SpeI
7079	cgtaccatggcgaagtcaaaaaaaatag	NcoI
7080	tgactagttatctttcttcaaaataagc	SpeI
3451	agagtttgatcctggctcag	
3454	gggttgcgctcgttg	
3818	gttttcccagtcacgac	
4730	ggcacgttccagttctttcag	
7125	atgctcaaattgtttgtttgtctcc	
7126	gagcaaacatataaccgaggaacaaagtgc	
7335	caatgcggctcgaatgactg	
7336	gggaatgggcgtagaaacca	
7337	tcggtgaggtggttgttacg	
7338	tacgcgtccttcgagcatac	
7339	caggcaaacaaagcgttgga	
7340	gttccgtaaggttctgccca	
7341	ggacagtgtttacttgttatcaagtc	
7342	gctataatgtgacgagctaaatcac	
7343	tgagtggctaagcgaaagtga	
7344	cttggtaaggttcctcgcgt	
5470	cgatgtcgactttttttttaacatttgattttgtatttaaaaaatttggtgttacttttgc	SalI
5471	cgatccatggttaatttttttaattacaatttagttaattacaagcaaaagtaacacc	NcoI

[a]: Restriction sites are underlined

Table S3 Plasmids used in the study

Plasmid	Description[a]	Reference
Vectors		
pMM13	ColE1 *ori*, Ap^r (Em^r)	(40)
pMM25	ColE1 *ori*; Km^r (Cf^r); suicide vector for *C. canimorsus*	(40)
pMM47.A	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r). *E. coli-C. canimorsus* expression shuttle plasmid with *ermF* promoter	(40)
pPM5	ColE1 *ori*; (pCC7 *ori*); Ap^r; (Cfx^r); promoter of *ompA* (*Fjoh_0697*) from *Flavobacterium johnsoniae* was amplified by PCR using primers 5470 and 5471, digested with *Sal*I and *Nco*I, and inserted into the corresponding sites of pMM47.A, replacing the original *ermF* promoter	This study
Mutator Plasmids		
pFL1	*ermF* framed by the 5' and 3' regions of *Ccan_03610* cloned into pMM25	This study
pFL2	*ermF* framed by the 5' and 3' regions of *Ccan_03620* cloned into pMM25	This study
pFL3	*ermF* framed by the 5' and 3' regions of *Ccan_03630* cloned into pMM25	This study
pFL4	*ermF* framed by the 5' and 3' regions of *Ccan_03640* cloned into pMM25	This study
pFL5	*ermF* framed by the 5' and 3' regions of *Ccan_03650* cloned into pMM25	This study
pFL6	*ermF* framed by the 5' and 3' regions of *Ccan_03660* cloned into pMM25	This study
pFL7	*ermF* framed by the 5' and 3' regions of *Ccan_03670* cloned into pMM25	This study
pFL8	*ermF* framed by the 5' and 3' regions of *Ccan_03680* cloned into pMM25	This study
pFL9	*erm*F framed by the 5' and 3' regions of *Ccan_03690* cloned into pMM25	This study
pFL10	*ermF* framed by the 5' and 3' regions of *Ccan_03700* cloned into pMM25	This study
pFL11	*ermF* framed by the 5' and 3' regions of *Ccan_03710* cloned into pMM25	This study
pFL12	*ermF* framed by the 5' and 3' regions of *Ccan_03720* cloned into pMM25	This study
pFL13	*ermF* framed by the 5' and 3' regions of *Ccan_03730* cloned into pMM25	This study
pFL61	*ermF* framed by the 5' and 3' regions of *Ccan_15860* cloned into pMM25	This study
Expression plasmids		
pFL14	*Ccan_03640* amplified with 7036 & 7037 and cloned into pPM5	This study
pFL15	*Ccan_03650* amplified with 7038 & 7039 and cloned into pPM5	This study
pFL16	*Ccan_03660* amplified with 7077 & 7078 and cloned into pPM5	This study
pFL17	*Ccan_03670* amplified with 7079 & 7080 and cloned into pPM5	This study
pFL18	*Ccan_03680* amplified with 7040 & 7041 and cloned into pPM5	This study
pFL19	*Ccan_03690* amplified with 7042 & 7043 and cloned into pPM5	This study
pFL20	*Ccan_03700* amplified with 7044 & 7045 and cloned into pPM5	This study
pFL21	*Ccan_03710* amplified with 7046 & 7047 and cloned into pPM5	This study
pFL22	*Ccan_03720* amplified with 7048 & 7049 and cloned into pPM5	This study

[a]: Selection markers for *C. canimorsus* are in between brakets

Table S4 Protein and iron concentration of products used in the study

	[Iron] (µM)	**[Protein] (g/l)**
Serotransferrin	41.2 ± 3.0	5.4 ± 0.2
Apotransferrin	0.5 ± 0.2	4.4 ± 0.1
Lactoferrin	37.4 ± 7.2	3.5 ± 0.1
Bovine transferrin	38.1 ± 3.5	4.3 ± 0.2
Hemoglobin	257.4 ± 16.3	5.9 ± 0.6
Human serum, heat inactivated (Millipore)	18.1 ± 1.0	49.0 ± 1.1
Human serum, heat inactivated (University Hospital of Basel)	19.5 ± 1.7	57.2 ± 5.2
Protein depleted human serum, heat inactivated	0.9 ± 0.6	N/A

4.3.Addendum

The genome sequences and annotations of the *C. canimorsus* clinical isolates (Cc2, Cc11, and Cc12)[1], the *C. canimorsus* dog strains (CcD38, CcD93, and CcD95)[2] and the *C. cynodegmi* dog strains (Ccyn2B, Ccyn49044, and Ccyn74)[3] have been deposited in the ENA/NCBI database.

Furthermore, the *C. canimorsus* dog strains (CcD38, CcD93, and CcD95) have now been classified as belonging to a new, separate species, namely *Capnocytophaga canis* [4,5].

1 Manfredi, P. *et al.* Draft Genome Sequences of Three Capnocytophaga canimorsus Strains Isolated from Septic Patients. *Genome Announc* **3**, (2015).
2 Manfredi, P. *et al.* Draft Genome Sequences of Three Capnocytophaga canimorsus Strains Isolated from Healthy Canine Oral Cavities. *Genome Announc* **3**, (2015).
3 Manfredi, P. *et al.* Draft Genome Sequences of Three Capnocytophaga cynodegmi Strains Isolated from the Oral Cavity of Healthy Dogs. *Genome Announc* **3**, (2015).
4 Oren, A. & Garrity, G. M. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* **66**, (2016).
5 Renzi, F. *et al.* Only a subset of C. canimorsus strains is dangerous for humans. *Emerg Microbes Infect* **4**, (2015).

General discussion

The present work aimed at providing a better understanding of the biology of Gram-negative bacteria from the phylum Bacteroidetes using our model organism *Capnocytophaga canimorsus*, a dog commensal and human pathogen. In particular, we focused at exploring the mechanisms underlying lipoproteins surface localization, a characteristic feature of Bacteroidetes. We also studied in detail one PUL-encoded *C. canimorsus* OM protein complex mainly composed of lipoproteins.

We found that this protein complex represents a new type of iron acquisition system (Ics) essential for growth of *C. canimorsus* in human serum. Unlike iron scavenging strategies from other bacteria, this system displayed broad substrate specificity by targeting several different iron carrying proteins found in mammals. Interestingly, the Ics has the classical architecture of Sus-like systems, *i.e.* an outer membrane anchored complex mostly composed of surface exposed lipoproteins. This study thus showed for the first time that Sus-like systems are not limited to carbohydrate acquisition. This also supported our hypothesis that surface exposed lipoproteins, and their underlying transport machinery, are essential in *C. canimorsus*, at least in some growth condition.

We thus investigated how lipoproteins can reach the bacterial surface. Using *in silico* analyses and site directed mutagenesis, we could show that surface exposed lipoproteins in *C. canimorsus* harbor a conserved N-terminal signal sequence that we named LES (Lipoprotein Export Signal). We determined the minimal composition for a functional LES as well as its optimal positioning. We also showed that the derived LES of two other Bacteroidetes species, namely *Bacteroides fragilis* and *Flavobacterium johnsoniae*, are functional in *C. canimorsus.* This indicated strong conservation of the signaling and the putative lipoprotein transport mechanisms in Bacteroidetes, which is in agreement with the fact that surface exposed lipoproteins are widespread in this phylum.

The discovery of the LES prompted us to focus at identifying the underlying lipoprotein transport machinery. By performing pull-down experiments using the lipoprotein chaperon LolA as bait, we identified and investigated several candidate proteins, among which an Skp homolog (Ccan_09090) and a BamA homolog (Ccan_17810). We could show that Ccan_09090 is essential in *C. canimorsus* and that its depletion leads to early growth arrest, a fact that is in line with its potential involvement in outer membrane biogenesis. Ccan_17810, a second copy of BamA, also turned out to be essential and to require a lipid anchor for its functioning. However, direct involvement of these proteins in lipoprotein export remains to be determined.

In order to identify the Bacteroidetes lipoprotein transport machinery, we will now focus on different approaches.

Based on our discovery of the Ics and of several other surface exposed Sus-like systems, we know that *C. canimorsus* relies on several surface exposed lipoprotein complexes to acquire iron and glycans to sustain growth. We thus plan on generating a transposon library that we will test for growth in human serum, followed by selection of the mutants that are strongly impaired in growth. However, this supposes that surface lipoproteins are required in liquid but not solid medium, a question that is so far unresolved. In the event that the lipoprotein transport machinery would be essential even on solid medium, we therefore also plan to perform a Tn-seq experiment, which will allow us to generate a list of essential genes in *C. canimorsus*, including the potential lipoprotein transport machinery.

An alternative approach will be based on our discovery of the LES. By introducing this signal sequence into an easily detectable reporter protein, we can generate a rapid readout system for surface exposure of lipoproteins. We could then use this strain to create a transposon library that we would screen for the exposure of the reporter protein at the cell surface.

In parallel to these genetics based approaches, we also plan to use a more biochemical approach. We will generate different bait proteins or peptides

harboring the LES, followed by pull-down experiments and mass spectrometry. The identified interaction partners will then be further investigated.

We also plan on pursuing our analysis of Ccan_17810, the lipidated BamA homolog, the so far most promising candidate for the export machinery. We will generate antibodies against the entire or parts of this protein that will be used to perform co-immunoprecipitation experiments. Antibodies raised against the extracellular domains of Ccan_17810 could also be used to inhibit its biological function, hence mimicking a depletion strain.

While these approaches would mainly focus on *C. canimorsus*, we will also apply them to less fastidious organisms, such as *Flavobacterium johnsoniae*. Indeed, one of the major difficulties so far is to grow large amounts of *C. canimorsus* in exponential phase, a problem that can be easily circumvented using *F. johnsoniae* since it grows readily in most common growth media. Furthermore, this bacterium can also be grown in defined medium with specific carbon sources. This means that lipoproteins surface exposure in *F. johnsoniae* can be easily monitored in conditions where growth requires the presence of Sus-like systems, *i.e*. surface exposed lipoproteins..

From a more general point of view, the identification of the LES could be used as tool to display a number of different proteins at the bacterial cell surface, which could be of interest for industrial purposes, especially in protein expression and purification. In the same line, Bacteroidetes expressing specific antigens at their surface could be used in probiotic foods or as a new approach for vaccine development.

The fact that surface exposed lipoproteins, and hence their dedicated transport machinery, are widespread in Bacteroidetes could also lead to the generation of new antimicrobials specifically targeting bacteria of this phylum. Indeed, if this machinery is partially surface exposed, as it is the case for the Bam complex and the LPS transporter LpdD, it could represent an interesting candidate for drug development. This would be of major interest for the treatment of anaerobic infections (*Bacteroides fragilis*) and peridontal diseases (*Porphyromonas gingivalis*) in humans as well as economical relevant poultry

and fish pathogens such as *Riemerella anatipestifer* and *Flavobacterium columnare.*