

**New insights into the journey of lipoproteins in
the cell envelope of *Escherichia coli*: disorder
matters**

Jessica El Rayes

Members of the jury

Professor Thomas Michiels (President)

Université Catholique de Louvain, Brussels, Belgium

Professor Jean-François Collet (Promoter)

Université Catholique de Louvain, Brussels, Belgium

Assistant Professor Nienke Buddelmeijer

Institut Pasteur, Paris, France

Professor Cedric Govaerts

Université Libre de Bruxelles, Brussels, Belgium

Assistant Professor Francesco Renzi

Université de Namur, Namur, Belgium

Professor Patrice Soumillion

Université Catholique de Louvain, Louvain-la-neuve, Belgium

Professor Emile Van Schaftingen

Université Catholique de Louvain, Brussels, Belgium

"A grandmother is like an angel who takes you under her wings. She watches over you and she would give you anything".

"My dearest Teta Saidé, my grandmother, losing you while being abroad was the hardest thing I have ever faced in this life. However, you helped me and gave the strength and faith I needed the most. I know you are always watching over me and protecting me from above. I offer you this thesis. I am sure you are very proud of me. I love you Teta. Keep watching over me and my family. "

Your granddaughter Jessica

Acknowledgements

As this amazing journey comes to an end, words cannot really express my feelings while writing the acknowledgements section. I cannot believe it has been already five years since I first came to Brussels. I still remember the day I said goodbye to my family and friends in Lebanon. Times flies so quickly when you are surrounded by good people.

First, all my gratitude goes to my PhD supervisor, Prof. Jean-François Collet. JF, you are an amazing person and lab PI. I am not really sure I will be able to express my feelings and how blessed I feel for ending up being a PhD student in the Collet lab. You gave me all the support I needed during these years, in the lab and in my life. You guided me to become the person I am right now. I will never forget how welcome you and your family were when I first landed in Brussels. When I think about my journey in Belgium and how hard it was, I really do not imagine being able to do a PhD abroad with another PI. I am amazingly blessed as you continue to help me through-out my life. One last thing I want to share, I now understand why I did not find a PhD thesis when I first graduated from the university, and why it happened only after two years. I believe God wanted me to be in the right place ☺ Thank you Thank you!!!

To the members of the jury, Professors Michiels, Govaerts, Soumillon and Van Schaftingen, and assistant professors Buddelmeijer and Renzi, thank you for agreeing to review my work, and to participate in the progress of the thesis. I thank you for your time and your effort, and for the amazing scientific discussions we had. I am honored to have you all as members of my PhD thesis committee.

To my scientific dad, Seung-Hyun you are such an amazing scientist. Thank you for your help, your patience and your guidance. I am really grateful for all our discussions and your ideas. You taught me a lot without complaining. The scientific community is not the same without your input. To Joanna, thank you for all your contribution in the linker project and for all your explanation and patience. I am really happy this project ended up in the right way.

To my Spanish sister from another mother, Raquelita thank you for always being there for me. I would have never imagine doing my PhD without having you by my side. I will never forget all our adventures during these years, especially in the T2T conferences, all the selfies you felt obliged to take, all the laughs and the tears. Thank you for all your advices and for being tough when needed. You will always be very special to me. I am very happy I gained a sister in Europe. I wish you all the best, you deserve it. Mil gracias.

Alexandra, my amazing lab partner, thank you for everything! You were really my number one support in the lab. Thank you for always believing in me as a scientist and encouraging me. You are one of the smartest scientists I know. I wish you all the happiness. Nahla, I cannot thank you enough. You helped me a lot during my first year in the lab. You really deserve the best! Juliette thank you for monitoring me and being my chaperone at the beginning of this journey. Mike, you are an amazing person and scientist. Thank you for all your help and ideas. I really enjoy discussing science with you. Petit Kilian, your kindness in the lab is incredible. You are always ready to help others. I will never forget how amazing our New-York adventure was! One of the best trips in my life. I wish you all the best in your Californian journey. To Pauline merci for your continuous help since day one. Elisa and Alexandra merci for the delicious cakes you made to celebrate my birthday! To all the Collet lab members, Emile, Laurie, Naemi, Chi, Alexandre, Elisa, Asma, Reeda, Alix, Bogdan, Antoine, thank you for all the beautiful moments we shared together.

To all my colleagues on the 5th floor, Karim, Vicky, Françoise, Hernan, the Laloux; the Bommer; the Kaneka and the Van Schaftingen labs, many thanks for the scientific and non-scientific discussions we had and the beautiful Christmas lunches we celebrated together.

Chi-Chi, my amazing American-Nigerian friend, I love you so much and thank you for all your encouragement. Your presence is really missed.

To all my Train2Target friends, cheers to all these amazing years we shared together. I will never forget every meeting and discussions we had. Best of luck for each one of you!

My Lebanese family in Belgium,

Aboura and Fredinand you mean a lot to me. Thank you for always being there for me. Being abroad is easier with both of you by my side. Fredinand I will never forget that whenever I needed help, you were always my rock. Grand merci frero!

Al Kai family and my dearest sisters, Leti and Titi, I cannot find the right words to express how much I love you. You are my family here in Belgium. Thank you for everything you did to me. Cheers to all the sleepovers, to always getting weird together and having our "Balthazar" moments!!

To my dearest Allouz, thank you for your kindness. Keep shining beautiful.

Polo, my dear Paul, thank you for always bringing joy and peace in my life.

Sarsouraaaa your presence is missed! Cheers to your lovely smile! Thank you for every time you listened to me. I am very proud of what you are achieving in your life. Keep going.

Elsa, Joanna, Michel, Hiba, Reeda, Shada, Abdallah, Jihane, Elie and Anne, Hussein, Père Nicolas, Sarmad and the "Jeunes de la paroisse" crew, mille merci!! Sarmad thank you for your help and for my scratch profile! Keep going brave man.

To Chantale and Rana, cheers to every dinner we regretted having it! I love you both. Thank you for everything.

My friends in Lebanon,

My Flower Power family, Dais; Philou and Lou, I still remember the amazing farewell you spent months preparing it for me. I am blessed for having you all in my life and sharing everything with you. You make my life more peaceful. I wish you all the best in your life. I hope we will soon be reunited somewhere on this planet.

My dear Racho, my fara, I am very happy for having you as a sister in my life. Despite the distance, our friendship never stops growing. I love you and thank you for everything!

To my number one support, Elie, thank you is not enough. Your presence in my life helped me to overcome all the hard times I have been through. You always know how to make me laugh even when I am crying. Thank you for all the love, the support and the encouragement. I wish you all the best in your life.

My dearest Hanouda, thank you for always taking care of my mental health. Keep going in your life sister. I wish you the best in your future career and life. Mia choukran sadikaty wa jaraty. Thank you for all the support and talks we had. I am very happy for always having you in my life.

A ma famille spirituelle de "Source d'eau vive" et particulièrement à père Jean-Marie, Nabil, Charlotte, Carmen, Charbel, Sandy et Nayla, un grand grand merci! Je suis sûre que j'ai pu arriver au bout de cette aventure grâce à vos prières. Unis toujours à vie!

To Sandra, Ivone and Mike, many thanks for all your prayers and our presence in my life. You taught me what does it really mean to have faith in God in this hard life. You are always in my heart and mind.

My dearest family,

You are my everyday blessing. Your love for me is unconditional. Thank you for supporting me in my whole life. Mom and dad, your sacrifices for me are uncountable. I will never be able to thank you as much as you deserve. I missed you every day during these years of being abroad. I know you are very proud of me and this is my greatest satisfaction. Mom, thank you for your prayers. Dad, thank you for all your hard work and for loving me this much. Joyci ya em Nour bhebik add el dene kella. Thank you for giving me the opportunity of being the aunt of an amazing baby. Chekho I am very proud of the man you became little brother! I love you ktir! Janjoun you make us proud Dr! Thank you Liliane!

Roy sohre el 3aziz, choukran choukran!! Gretta thank you for all the balloons and cakes. Love you beautiful souls.

Therese, Jako, Amoula, Marmoura and Micha thank you for being that amazing. Grand merci for all the support you gave me and for always cheering me on. I love you!

Finally, I cannot finish without thanking the Lord for all his blessings and for surrounding me with all these beautiful people I have in my life. Thank you God for always taking care of me and showing me the right path. Thank you for giving me my guardian angel who never fails of rescuing me.

Table of Contents

<i>List of abbreviations</i>	<i>1</i>
<i>Train2Target research project.....</i>	<i>3</i>
<i>Summary.....</i>	<i>5</i>
<i>Research objectives</i>	<i>9</i>
<i>Section I. Introduction</i>	<i>11</i>
<i>Chapter I. The cell envelope of Gram-negative bacteria.....</i>	<i>13</i>
<i>1.1. Architecture of the cell envelope.....</i>	<i>13</i>
1.1.a. The outer membrane (OM)	14
1.1.b. The periplasm	15
1.1.c. The inner membrane (IM)	18
<i>1.2. Importance of the bacterial envelope</i>	<i>19</i>
<i>Chapter II. Overview on lipoproteins.....</i>	<i>25</i>
<i>2.1. Biogenesis, translocation and maturation.....</i>	<i>25</i>
2.1.a. Biogenesis and translocation	25
2.1.b. Maturation	26
<i>2.2. Sorting and targeting of lipoproteins</i>	<i>36</i>
2.2.a. LolCDE	36
2.2.b. LolA and LolB	44
<i>2.3. Conservation of the lipoprotein pathway.....</i>	<i>47</i>
<i>2.4. Surface exposure of lipoproteins.....</i>	<i>48</i>
<i>2.5. Diverse functions of lipoproteins.....</i>	<i>56</i>
<i>2.6. Inhibition of the journey of lipoproteins.....</i>	<i>64</i>
<i>Chapter III. Intrinsic disorder in proteins.....</i>	<i>69</i>
<i>3.1. Overview on intrinsically disordered proteins.....</i>	<i>69</i>
3.1.a Characteristics of intrinsic disorder	70
3.1.b. Physiological functions of IDPs	72
3.1.c. Informative structural studies of IDPs	75
<i>3.2. IDPs in bacteria</i>	<i>76</i>
<i>Section II. Results</i>	<i>79</i>
<i>Chapter I. Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria</i>	<i>83</i>
<i>Chapter II. Structural insights into the LolA- RcsF complex.....</i>	<i>143</i>

Section III. General Discussion and perspectives	155
Section IV. Conclusion	171
Bibliography.....	175
Annexes	193
<i>I. Published paper: Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria</i>	<i>195</i>
<i>II. Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles</i>	<i>221</i>
<i>III. A tale of two elements: when disorder meets lipoproteins in the cell envelope of Escherichia coli</i>	<i>233</i>

List of abbreviations

ABC: ATP binding cassette

AFM: Atomic force microscopy

AT: Autotransporter

ATP: Adenosine tri phosphate

Bam: β -barrel assembly

CD: Circular dichroism

Cryo-EM: Cryogenic electron microscopy

DDM: Dodecyl maltoside detergent

DNA: Deoxyribonucleic acid

DolP: Division and OM stress-associated lipid-binding protein

DpaA: Peptidoglycan meso-diaminopimelic acid protein amidase

E. coli: *Escherichia coli*

fHbp: Factor-H binding protein

Glc-Nac: N-acetyl-glucosamine

His: Histidine

IDP: Intrinsically disordered protein

IDR: Intrinsically disordered region

IM: Inner membrane

KDO: 3-deoxy-D-manno-octo-2-ulosonic acid

Lgt: Prolipoprotein diacylglycerol transferase

Lnt: Apolipoprotein N-acyltransferase

Lol: Localization of lipoprotein

Lpp: Braun outer membrane lipoprotein

LPS: Lipopolysaccharides

Lpt: Lipopolysaccharide transport

Lsp: Signal peptidase II

MD: Molecular dynamics

MoRF: Molecular recognition feature

Mur-NAc: N-acetyl-muramic acid

NBD: Nucleotides binding domain

NMR: Nuclear magnetic resonance
OM: Outer membrane
OMP: Outer membrane β -barrel protein
Pal: Peptidoglycan associated lipoprotein
PBP: Penicillin binding protein
PG: Peptidoglycan
PH: Periplasmic helix
PMF: Proton motive force
Pop: Polar organizing protein
Psp: Phage shock protein
Rcs: Regulator of capsule synthesis
SAXS: Small angle X-ray scattering
SDS: Sodium dodecyl sulfate
Sec: General secretory pathway
smFRET: Single-molecule fluorescence resonance energy transfer
SRP: Signal recognition particle
SSB: Single-stranded DNA binding proteins
ssDNA: Single-stranded DNA
TAT: Twin arginine translocation pathway
TbpB: Transferrin binding protein B
TM: Transmembrane
TMD: Transmembrane domain
T3SS: Type III secretion system

Train2Target research project

My PhD thesis was supported, in part, by a grant from the European Commission via the International Training Network (ITN) Train2Target (721484).

The ITN research network is part of the Marie- Skłodowska-Curie actions (MSCA) that support training and doctoral programs outside of academia. Thus, it provides a connection between fundamental and applied sciences. The Train2Target was a project funded by the Horizon 2020 Marie-Skłodowska-Curie action. The main objective of this project was to dissect the molecular mechanisms used by Gram-negative bacteria to fight antibiotics, and to find new potential antimicrobials. The consortium of this project grouped different partners from academia and industries. In total, 7 academic research groups and 4 industrial companies participated in the training of 15 PhD students in 9 different European countries.

Antibiotic resistance occurs when the bacteria develops the ability to defeat the drug that was designed to kill this germ. According to the World Health Organization (WHO), at least 700,000 people die each year from drug-resistance diseases. This number is predicted to reach 10 million by 2050. Thus, antimicrobial resistance is considered one of the most dangerous global health threats. The shocking numbers delivered by the WHO call for urgent actions to fight this crisis and prevent a drug-resistance pandemic.

Funding programs like the Train2Target project aim to effectively fight the antibiotic resistance crisis by finding new approaches and solutions, through science and research.

During my PhD journey, I had the chance, thanks to the Train2Target project, to be on secondment at Naicons in Milano in Italy, for one month, under the supervision of Dr. Stefano Donadio and Dr. Margherita Sosio. Naicons was one of the 4 industrial companies of the consortium of the Train2Target project. The mission of this company is to find novel natural compounds with antimicrobial activities. As a PhD student in the Collet lab, I was working on a confidential project that interested both partners: Naicons and the Collet lab.

Summary

The cell envelope is essential for bacterial survival, as it forms a protective barrier that shelters the cell from its environment. This envelope consists of an outer and an inner membrane; both membranes are separated by the periplasm. In the periplasm, a mesh-like layer called peptidoglycan, or murein, is present and is important for cell shape and the protection against osmotic shock. The cell envelope includes three groups of proteins. During my thesis, I focused on studying one of these groups, which is lipoproteins. Lipoproteins, as their name subtly suggest, are globular proteins that are covalently linked to a lipid moiety which anchors them in one of the two membranes. This subset of proteins is involved in a wide range of vital bacterial processes.

Lipoproteins can be retained in the inner membrane (IM) or can be transported to the outer membrane (OM). The sorting and targeting of lipoproteins between the two membranes is controlled by the Lol (Localization of lipoproteins) system. In this thesis, I show that lipoprotein sorting is affected by the presence of an intrinsically disordered region, that we call a linker, that is present at the N-terminal of half OM lipoproteins.

First, I show that deleting the linker re-routes OM lipoproteins to the IM, which has functional consequences. Moreover, I found that the length, but not the sequence, of this linker is important for correct lipoprotein sorting, on the condition that the linker stays unstructured.

Second, I provide mechanistic details on the function of the linker, showing that the disordered region works in a Lol-dependent manner.

Finally, I propose a model where the linker optimizes OM targeting as well as the surface exposure of lipoproteins.

Organization of the thesis

This thesis is organized in four main sections: Introduction, Results, Discussion and Conclusions. Annexes can be found at the end of the manuscript.

The **Introduction** is divided into three chapters. The first chapter provides an overview on the cell envelope of Gram-negative bacteria. The second chapter focuses on lipoproteins and describes the different aspects of their journey, conservation and importance as antibiotic targets. Parts of this chapter were adapted from a review that I co-authored and that was published in *Current Opinion in Microbiology* in March 2021. This review can be found in annex II. The third chapter describes structural disorder in proteins.

The **Results** section is divided into two chapters. *In chapter I*, we investigate the role of the N-terminal intrinsically disordered region, called linker, displayed by half OM lipoproteins. These findings are published in *Nature Chemical Biology* in July 2021. *Chapter II* presents preliminary unpublished data describing the efficient purification of the periplasmic chaperone LolA in complex with a lipoprotein, in this case the stress sensor RcsF devoid of its N-terminal linker.

The **General Discussion** section discusses the main findings of this thesis and the different remaining open questions. This section is followed by the **conclusion**.

The **References** of this thesis, except the ones of the results section, are listed after the conclusion section.

Finally, three **annexes** are included at the end of this thesis. The first one is the published version of the paper presenting most of the results obtained during the thesis (Chapter I of the Results). The second annex is our review on lipoproteins published in Current Opinion in Microbiology. The third one is the Blog Post that we were invited to write, by the Nature Chemical Biology journal, for the Behind the Paper channel of the Nature Research Microbiology community website. This blog post is related to our paper published In Nature Chemical Biology.

Research objectives

The cell envelope of Gram-negative bacteria is the first defense line against antibiotics and environmental stresses. Different groups of proteins are present in the envelope and ensure cellular protection. One of these groups includes globular proteins anchored to the inner or the outer membrane by a lipid moiety and is called lipoproteins. These lipidated proteins are essential for bacterial survival and fulfil a variety of different functions. In consequence, lipoproteins have been drawing a lot of attention lately and are considered a great target for the design of new antibiotics.

Lipoproteins can be retained in the IM or exported to the OM by the Lol system. **The first goal** of my thesis was to study the role of an intrinsically disordered region, present at the N-terminus of OM lipoproteins, in the sorting of lipoproteins between the two membranes of the cell envelope. We found that this disordered region optimizes OM targeting of lipoproteins by Lol. These findings are presented in chapter I of the Results section.

The second goal of my thesis was to optimize the purification of the periplasmic chaperone LolA in complex with a lipoprotein, in an attempt to determine its crystal structure. Preliminary unpublished results are presented in chapter II of the Results section.

Section I. Introduction

Chapter I. The cell envelope of Gram-negative bacteria

The cytoplasm of Gram-negative bacteria is surrounded by the cell envelope. This three-layered structure is therefore exposed to the extracellular environment and thus to antibiotics. Different machineries and proteins are present in this compartment to ensure bacterial protection and resistance to virulence factors and antimicrobials. The cell envelope is consequently very important and essential for the bacteria to survive.

The excessive inappropriate use of antibiotics has led to the appearance of antibiotic resistance as a major global threat. One strategy to fight this resistance is to find new targets and to develop new antimicrobials. Since the cell envelope is the first barrier against toxic compounds, studying the envelope constitutes therefore a good strategy to find novel targets. In this chapter, I will describe in detail the architecture and the importance of the cell envelope of Gram-negative bacteria.

1.1. Architecture of the cell envelope

The cell envelope is the morphological hallmark of Gram-negative bacteria. It is composed of two concentric membranes: the inner membrane (IM), which is in contact with the cytoplasm, and the outer membrane (OM), which constitutes the interface with the environment ¹. The inner and outer membranes are separated by the periplasm, a compartment that represents 10-20% of the total cell volume and contains a thin layer of peptidoglycan ¹. The envelope is required for growth and survival, maintaining cell shape and providing osmotic protection to cells. The main focus of this study will be the *Enterobacteriaceae* family, mainly *Escherichia coli* (*E. coli*), the model bacterium.

In *E. coli*, about one-third of all proteins are targeted to the cell envelope ². Soluble proteins are present in the periplasm where they participate in a variety of processes including peptidoglycan assembly, disulfide bond formation and nutrient import. Integral membrane proteins are present in both the inner and outer membranes; in the OM, they typically adopt a β -barrel conformation ³. Another important group of envelope proteins is the lipoproteins, which are globular proteins anchored to a membrane by a lipid moiety ⁴ (**Fig.1**). Below, we will describe in detail the different compartments of the cell envelope.

1.1.a. The outer membrane (OM)

The OM is the outermost layer of the Gram-negative cell envelope. Gram-positive bacteria lack this layer. The OM is a unique membrane by its asymmetry. While the inner leaflet is formed by phospholipids, the outer leaflet is composed by lipopolysaccharides (LPS) (**Fig.1**). The LPS, also known as endotoxin, is a potent inducer of the host immune system during a bacterial infection ⁵. It is formed of three parts: the hydrophobic lipid A, the oligosaccharide core and the O-antigen (**Fig.1b**). The lipid A is the hydrophobic moiety that anchors the LPS to the OM outer leaflet. In many *Enterobacteriaceae*, lipid A is an acylated β -1-6 linked glucosamine disaccharide. In *E. coli* and *Salmonella*, the glucosamines are phosphorylated at positions 1 and 4' and the lipid A is hexa-acylated ⁶. The oligosaccharide core is divided into two parts: the inner and the outer cores. Within a genus or a family, the inner core is usually conserved: it contains 3-deoxy-D-manno-octo-2-ulonic acid (KDO) and few heptose residues, some of which are phosphorylated. However, the outer core shows more structural diversity and consists mostly of hexoses ⁷. This is perhaps related to the fact that the outer core is exposed to external stresses such as

the host responses, bacteriophages. The outermost component of the LPS is the O-antigen which consists of oligosaccharide repeating units. It is a highly variable chain of carbohydrates. Some strains of *E. coli*, such as the K-12 strain used in this study, lack the O-antigen, which is known to compromise their virulence ⁸. Moreover, the O-antigen is replaced by colanic acid repeats when the capsular colanic acid synthesis is induced ⁹. The LPS molecules are bridged by the action of divalent Mg^{2+} and Ca^{2+} cations. These positive charges counteract repulsion between negatively charged phosphates in lipid A and in the core.

The different subunits of the LPS are exported from the IM to the OM by the lipopolysaccharide transport (Lpt) machinery. Lpt proteins form a complex that spans the cell envelope from the cytoplasm to the OM. The mechanism and the structure of this machinery was recently reviewed in ¹⁰. Two major classes of proteins exist in the OM: outer membrane β -barrel proteins (OMPs) and lipoproteins. Functions of these proteins will be discussed more in detail in section 1.2.

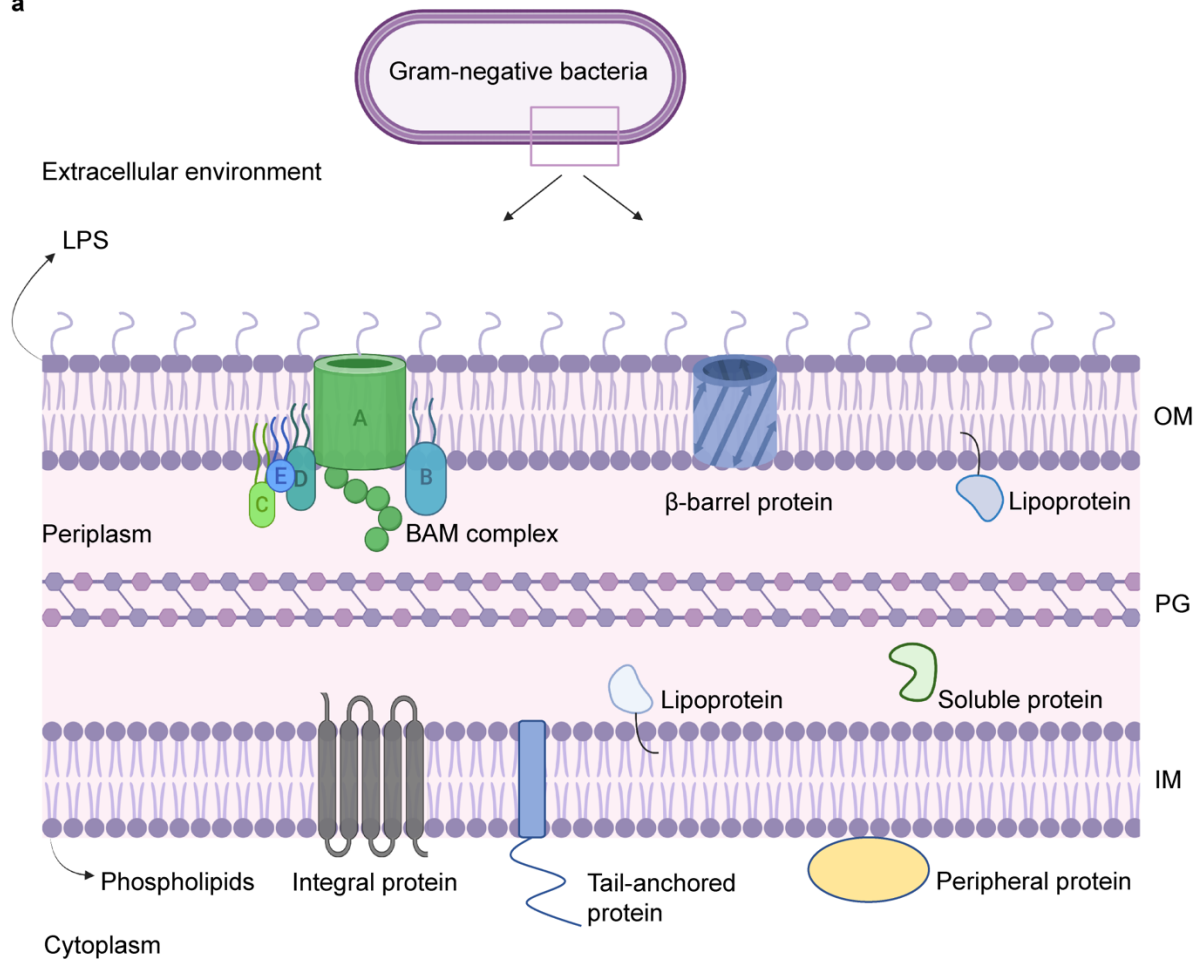
1.1.b. The periplasm

The periplasm is the region that separates the inner and the outer membranes. This hydrophilic space is very important for cell viability because it includes a variety of soluble proteins of different functions: proteases, phosphatases, holdases, chaperones, etc. ¹¹⁻¹³ (**Fig.1**).

A thin layer of crosslinked sugars and amino acids, termed peptidoglycan (PG) or cell wall, lies within the periplasmic space (**Fig.1**). The PG protects the cell from osmotic shocks and helps maintaining cell shape. It is a layered glycopolymer composed of N-acetyl-glucosamine (Glc-Nac) and N-acetyl-muramic acid (Mur-Nac) crosslinked by peptide bridges¹⁴ (**Fig.1**). Many enzymes are involved in PG synthesis and remodeling

during cell division and elongations. Among these enzymes, glycosyltransferases are the glycan chain polymerases and transpeptidases form peptide crosslinks. Penicillin Binding Proteins (PBPs) of class A catalyze both reactions whereas PBPs of class B are monofunctional transpeptidases ¹⁵. Being important for cell division and elongation, PBPs are the target of the β -lactamase antibiotics. The PG mesh is not totally separated from the OM. In fact, many proteins in the OM interact with this glycopolymer. Among these proteins, Lpp (the Braun lipoprotein) is covalently attached by the ϵ -amino group of its C-terminal lysine residue to the peptide stem of the PG ¹⁶. Pal (the peptidoglycan associated lipoprotein) is another OM lipoprotein connected to the PG. However, Pal interacts non covalently with the PG through hydrogen bonds and hydrophobic contacts^{17,18}.

a



b

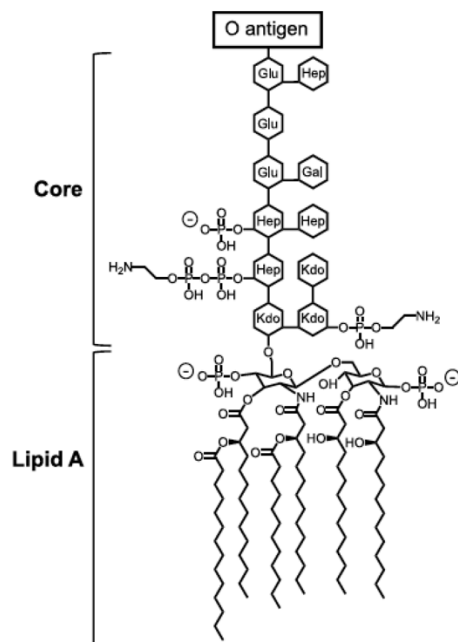


Figure 1. The components of the cell envelope of Gram-negative bacteria. **a.** The cytoplasm of Gram-negative bacteria is surrounded by the cell-envelope, a structure formed by an outer membrane (OM) and an inner membrane (IM) separated by a hydrophilic space termed as periplasm. Integral proteins and few lipoproteins and tail-anchored proteins are present in the IM. Some peripheral proteins are at the edge of the IM. β -barrel proteins and lipoproteins decorate the OM. Soluble proteins, and a thin layer of peptidoglycan are found in the periplasm. **b.** Structure of LPS from *E. coli* K-12 strain. The LPS is composed of lipid A, the oligosaccharide core and the O-antigen. The structure of LPS was adapted from (Wilson & Ruiz, 2021 ¹⁰).

1.1.c. The inner membrane (IM)

The IM, also called cytoplasmic membrane, is the inner most layer of the cell envelope. It forms a hydrophobic barrier against the content of the cytoplasm. It is a phospholipid bilayer. In *E. coli*, phosphatidyl ethanolamine and phosphatidyl glycerol are the main phospholipids, but there is lesser amount of phosphatidyl serine and cardiolipin (diphosphatidylglycerol) ¹. Although few lipoproteins, tail anchored proteins (proteins with a C-terminal domain anchored in the IM) and peripheral associated IM proteins are present in the IM, typical IM proteins are integral membrane proteins with α -transmembrane helices ¹⁹ **(Fig.1)**. IM proteins usually function as enzymes, translocation or extraction machineries. Since the periplasm is deprived of the energy source, the Adenosine Tri Phosphate (ATP), these machineries take advantage of the cytoplasmic ATP and the Proton Motive Force (PMF) to function. Since the IM forms a hydrophobic barrier against the cytoplasm, two IM translocation machineries serve as “bridge” for proteins to pass from the cytoplasm to the cell envelope. These machineries are the general secretion route (Sec) and the Twin arginine translocation pathway (Tat). Sec and Tat are used by unfolded and folded proteins, respectively ²⁰ **(Fig.2)**.

Lipoproteins and secreted proteins are post-translationally translocated through the Sec complex. Briefly, after translation, the protein precursor binds to the cytoplasmic chaperone SecB which keeps it in an unfolded state and prevents aggregation. The complex will then bind to the peripheral ATPase SecA that will thread the polypeptide through the IM translocon SecYEG. Once the protein threaded, the signal peptide of secreted proteins is cleaved by the signal peptidase LepB. As for lipoproteins, the signal peptide may be inserted in the IM through a lateral gate of SecYEG, thereby tethering the lipoprotein to the IM. The signal peptide of lipoproteins is then cleaved by the signal peptidase II (more details are found in Chapter II of the Introduction) **(Fig.2a)** ^{21,22}. Moreover, the translocation of inner membrane integral proteins requires the Sec pathway using the signal recognition particle (SRP) via a co-translational pathway ²³.

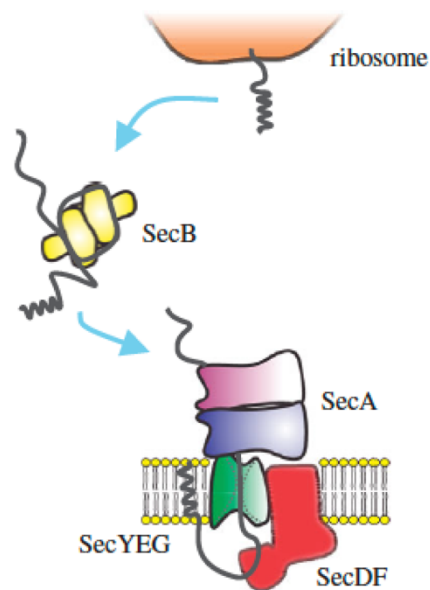
As for the Tat machinery, substrates present a signal peptide with a highly conserved twin-arginine motif (S-R-R-x-F-L-K, where x is a polar amino acid) ²⁴. TAT-routed translocation is an active PMF requiring process. Briefly, the signal peptide binds the receptor TatBC. This binding triggers the PMF-dependent oligomerization of TatA to form an active translocase complex and enable the translocation of the protein ^{25,26} **(Fig.2b)**.

1.2. Importance of the bacterial envelope

The bacterial envelope is very important for cell viability and survival due to the presence of different components, each ensuring a specific role. Some of these functions will be listed below.

- The hydrophobicity of the OM **protects** the cell from toxic large hydrophilic compounds like vancomycin. Vancomycin is a large hydrophilic antibiotic which targets the PG. Moreover, the amphipathic nature of the LPS provides a permeability barrier that protects the cell from different antibiotics ⁷. LPS are also related to the pathogenicity of the bacterial strain.

a



b

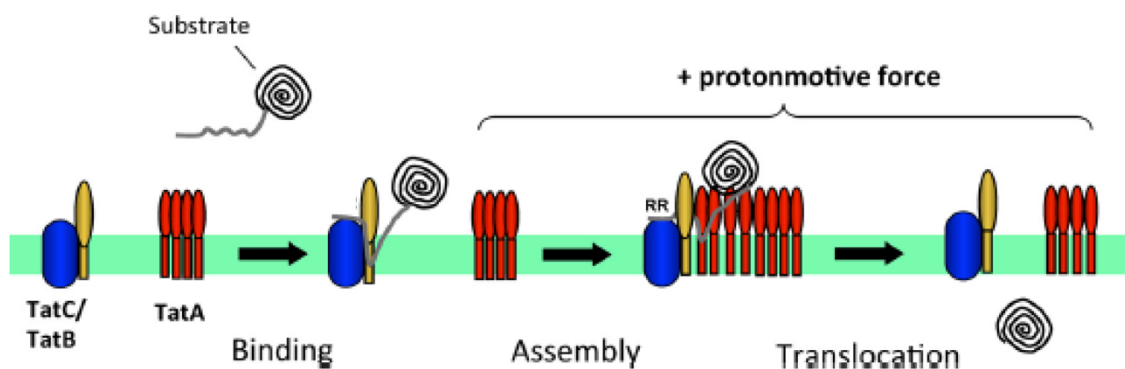


Figure 2. The translocation machineries. Schematic of the Sec and TAT dependent protein translocations. The Sec **(a)** and TAT **(b)** systems are used for the translocation of the unfolded or folded proteins, respectively. The images are adapted from ²⁷ and ²⁶.

- Efflux pumps are present in the cell envelope. Therefore, the cell can **eliminate toxic compounds** that previously entered from the extracellular environment. In *E. coli*, the main efflux pump is AcrAB-TolC ²⁸.
- Many secretions systems span the cell envelope and are used by the pathogen to **deliver DNA or protein** into the extracellular environment or into target cell²⁹.
- Integral OM proteins (OMPs) share a common β -barrel structure and are very important for the **uptake and disposal of small hydrophilic molecules** such as nutrients and waste. OMPs are synthesized in the cytosol, translocated in an unfolded state into the cell envelope through the Sec, transported to the OM by periplasmic chaperones (SurA or DegP/Skp) where they interact with the Bam (β -barrel assembly machinery) complex and fold into an OM integral β -barrel. The Bam complex is formed of BamA, a β -barrel protein which has a periplasmic tail of five flexible POTRA domains, and of four lipoproteins BamB,C,D and E ³⁰ (**Fig.1**). Some OMPs like OmpC and OmpF, function as **diffusion pores** allowing entry of water-soluble molecules through the hydrophobic OM. Therefore, they control cellular uptake of small molecules including nutrients and antibacterial agents ³¹. OmpC and OmpF occur as trimers in the OM ^{32,33}. OmpA, the most abundant OMP in *E. coli*, often adopts a two-domain structure, with an N-terminal OM β -barrel and a C-terminal globular domain in the periplasm ³⁴. Therefore, in this conformation, OmpA is unlikely to function as an efficient large channel. Instead, the C-terminal periplasmic domain of OmpA was shown to interact with the PG leading to the

OM integrity ^{35,36}. However, it has been proposed that OmpA can also fold into a single, larger, OM β -barrel conferring a very low permeability rate to the cell ³⁷⁻³⁹.

- Many lipoproteins are present in the cell envelope ensuring different essential functions. Lipoproteins are proteins with a globular periplasmic domain attached to the membrane by a lipid moiety (**Fig.1**). Having lipoproteins has a major advantage for the bacterial cell. In fact, many lipoproteins decorate the cell surface and are exposed to the extracellular environment while staying anchored to the membrane. Thus, they are used by the bacteria as weapons and virulence factors ⁴⁰⁻⁴³. For instance, the factor-H binding protein at the surface of *Neisseria meningitidis* is a key factor for its virulence and is considered a vaccine antigen ⁴⁴. More details on lipoproteins and their functions are discussed in chapter II of the Introduction.
- Flagellum and pili are found on the surface of the cell envelope. These filamentous proteinaceous structures are key to **virulence**. The flagellum confers **motility**, **adhesion** and, in some cases, acts as a **sensory organ** ⁴⁵. Pili are implicated in cell motility and adhesion, **biofilm formation** and **gene transfer** by conjugation ⁴⁶.
- **Stress response systems** of the cell envelope are used by bacteria as a defense against external stresses (antibiotics, osmotic shock, etc.). These systems detect damages in the cell envelope, and in response modulate genes expression. *E. coli* has five major envelope stress response systems: Bae, the bacterial phage shock protein system (Psp), the σ^E dependent signaling cascade, the Regulator of capsule synthesis (Rcs) system and Cpx. Bae, Rcs and Cpx are two-component systems where a membrane embedded sensor

histidine kinase activates a cytoplasmic response regulator. Bae is activated by the presence of toxic compounds. Psp helps the cell in managing the impacts of agents impairing the IM and the proton motive force. σ^E is activated by the accumulation of misfolded OMPs and LPS. Rcs is activated by stress that triggers the envelope integrity by affecting the OM, the periplasm or the PG. Finally, Cpx monitors the misfolding of proteins and lipoproteins trafficking and maturation in the periplasm^{47,48}. The Rcs and the Cpx stress responses will be detailed in section 2.3 of chapter II.

The functions cited above are among many important functions of the cell envelope, but the list goes on. It is very clear that the cell envelope plays the role of a cargo ship full of different components essential for the survival of bacteria.

In conclusion, the cell envelope plays a crucial role in the survival and the homeostasis of Gram-negative bacteria. Since it encompasses several cellular defense mechanisms, the cell envelope constitutes a weapon of resistance against various antibiotics and toxic compounds. In the next chapter, we will focus on lipoproteins, the main subject of my thesis, and we will discuss their characteristics, biogenesis, surface exposure and importance in the cell.

Chapter II. Overview on lipoproteins

The content of this chapter is based on our review entitled “Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles” that is published in Current Opinion in Microbiology in March 2021 (**Annex II**).

2.1. Biogenesis, translocation and maturation

2.1.a. Biogenesis and translocation

Lipoproteins are synthesized in the cytoplasm as precursors. They display an N-terminal signal peptide that is ~25 amino acids long that generally presents a short, positively charged region followed by a longer stretch of hydrophobic residues ^{49,50}. The last four residues of the signal peptide form a conserved ([LVI]⁻³[ASTVI]⁻²[GAS]⁻¹C⁺¹) motif referred to as the lipobox ⁵⁰. The lipobox functions as a molecular determinant of lipid modification which is unique to bacteria; it can be used to predict genes encoding lipoproteins in bacterial genomes ⁵⁰. Only the cysteine at the last position of the lipobox is strictly conserved: this cysteine will later become the first residue (+1) of the mature protein (**Fig.3a**).

The vast majority of lipoproteins cross the IM in an unfolded state using the general secretory system (Sec) (**Fig.2**) ⁵¹. (See section 1.1.c for more details).

The IM complex SecDF, which usually facilitates the release of the polypeptide from SecYEG, seems also to be important for the translocation of Lpp ⁵². Moreover, a few studies showed that the translocation of Lpp does not require a functional SecB ⁵³, but depends on the SRP and the Sec-associated insertase (YidC) ⁵⁴. This is quite

surprising since SRP and YidC are usually used for the translocation of integral IM proteins ^{21,55}. This result might indicate that some lipoproteins use two translocations pathway. To sum up, it seems that further studies are needed to confirm the final model of lipoprotein's translocation.

Beside the Sec system, few lipoproteins cross the IM in a folded state through the twin-arginine translocation (TAT) machinery ^{56,57}. Predicted TAT-dependent lipoproteins are fast-folding and co-factor containing proteins such as hydrolases, oxido-reductases, transporters and bonding proteins ⁵⁸. For instance, the OM lipoprotein component of the efflux pump of *Pseudomonas sp.*, OprM, is predicted to use the TAT pathway ⁵⁷. Phylogenic analysis showed that pathogenic organisms use TAT less frequently compared to free-living organisms (See section 1.1.c for more details).

2.1.b. Maturation

After translocation, the lipoprotein remains anchored to the IM by its N-terminal signal peptide and undergoes several maturation steps. Three IM enzymes participate in this process ⁵⁹ (**Fig.3**). The first enzyme is the prolipoprotein diacylglyceryl transferase (Lgt), which catalyzes the covalent attachment of a diacylglyceryl moiety to the cysteine of the lipobox by a thioether linkage. After addition of the diacylglyceryl moiety, the signal peptide is cleaved by the signal peptidase II (LspA), which generates an apolipoprotein, with a cysteine as first residue. The third and last step of lipoprotein maturation consists in acylating the N-terminal amino group of the cysteine, yielding a mature lipoprotein with three acyl chains. This reaction is catalyzed by the apolipoprotein N-acyltransferase (Lnt) (**Fig.3**). Lgt, LspA and Lnt are conserved and essential in the majority of Gram-negative organisms, including pathogenic species ⁵⁹⁻

⁶²; they are therefore attractive targets for the design of new antimicrobials. In the last few years, important achievements have been made in the understanding of the catalytic mechanisms of Lgt, LspA and Lnt. Below we will discuss these latest discoveries.

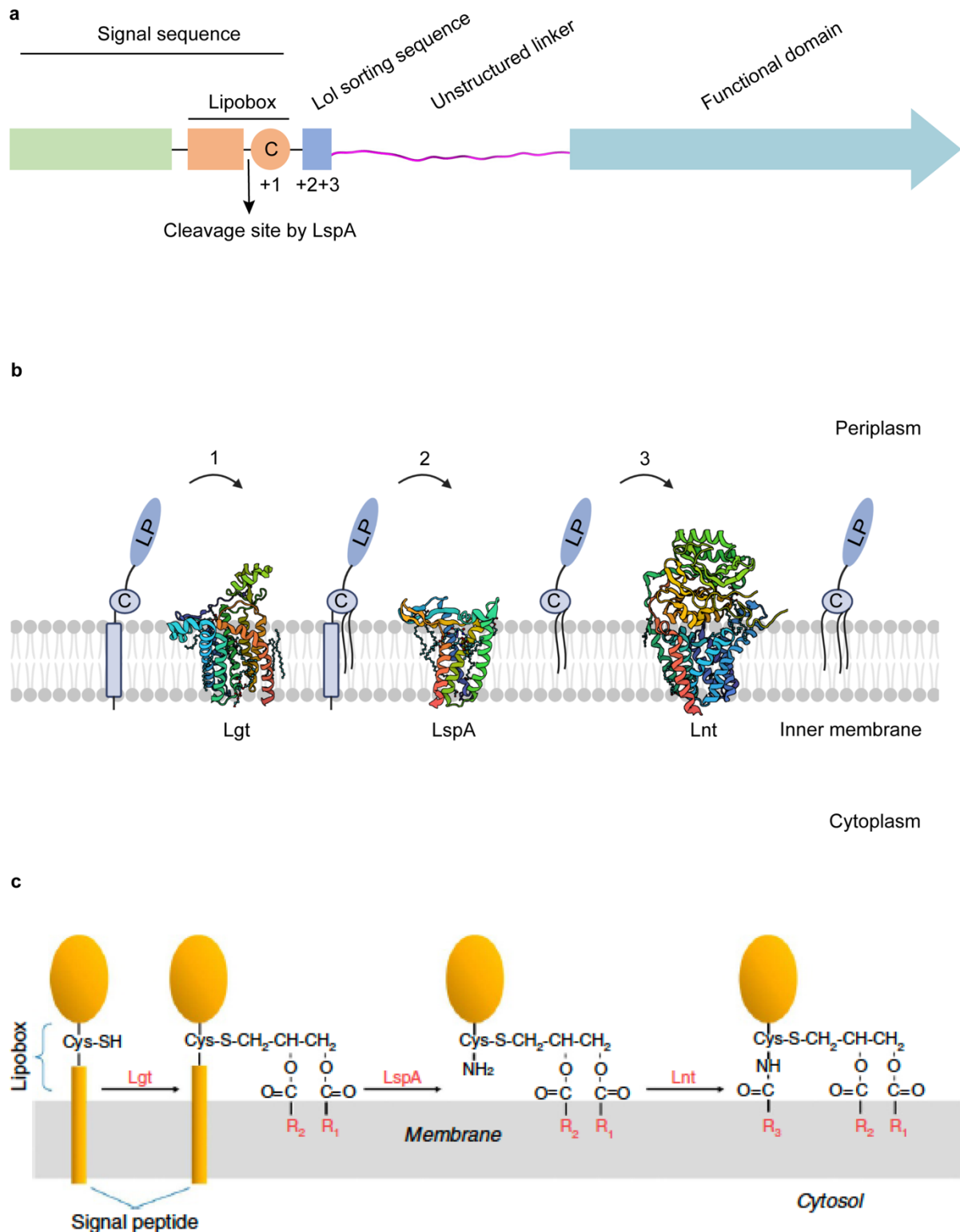


Figure 3. Lipoproteins structural domains and maturation. **a.** Schematic of lipoprotein structural domains. Starting from the N-terminal region, a lipoprotein is formed of a signal sequence with a lipobox, a Lol sorting sequence and a functional domain. Many lipoproteins have an unstructured linker before the functional domain.

b,c. Following translocation of the prelipoprotein across the inner membrane, Lgt catalyzes the addition of diacylglycerol to the thiol group of the cysteine residue in the lipobox, generating 'prolipoprotein' (1). The signal peptide is then cleaved by the peptidase LspA (2), giving rise to 'apolipoprotein'. The last step of the maturation process, the acylation of the amino group of the now N-terminal cysteine, is catalyzed by Lnt (3). The crystal structures of the three enzymes are shown in **b**: Lgt in complex with phosphatidylglycerol (pdb: 5AZC), LspA in complex with globomycin (pdb: 5DIR) and Lnt (pdb: 5VRG). The figure in **c** is adapted from ⁶³.

- **Prolipoprotein diacylglycerol transferase (Lgt)**

Lgt is an integral IM enzyme that catalyzes the first step of the maturation process: the transfer of diacylglycerol moiety from a phosphatidylglycerol of the IM to the sulfhydryl group of the cysteine residue by a thioether linkage (**Fig.3c**) ⁶⁴⁻⁶⁶.

The crystal structure of Lgt from *E. coli* ⁶³ confirms the membrane topology of Lgt ⁶⁷ and shows that Lgt has seven transmembrane (TM) helices which form the body of the protein (or TM core), a periplasmic "head" and two amphipathic arms ⁶³ (**Fig.4**). The TM core forms a major TM and a minor TM domains (**Fig.4**). A central cavity is located between the two TM domains, and it contains two clefts exposed to the membrane serving as potential entrance and exit for the substrates and the product, respectively. Moreover, Lgt has a signature motif, a glycine-rich region with a highly conserved arginine, important for the binding of the negatively charged phosphatidylglycerol ^{63,67,68} (**Fig.4**). Briefly, the proposed mechanism, based on the structure, is that the phosphatidylglycerol, the donor, laterally enters the first binding site from the front-cleft, slides to the second binding site, at the side-cleft, where the prelipoprotein is bound. The lipobox of the prelipoprotein binds a highly conserved HGGL motif on the second binding site of Lgt. The addition of acyl chains

takes place at this site and then the porolipoprotein leaves laterally from this side-cleft
63,68.

Lgt is conserved in all classes of bacteria, it is essential for growth and the majority of the essential residues are located in the membrane ⁶⁷. A recent study suggested that Lgt could be weakly associated to the inner leaflet of the IM ⁶⁹ instead of being a transmembrane protein ^{63,67}. However, these data are not compatible with the structure of Lgt and need to be carefully taken into consideration.

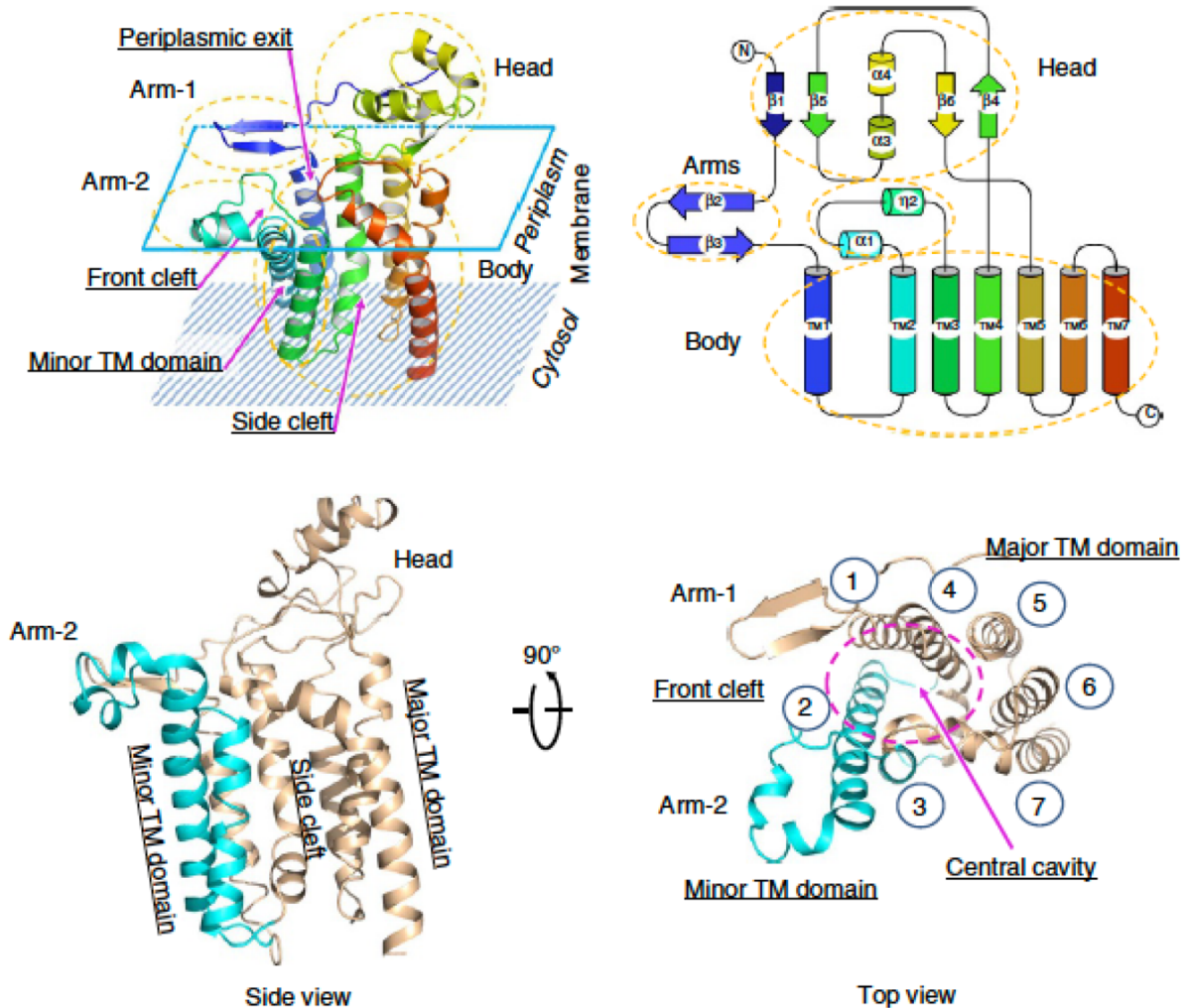


Figure 4. Overall structure of *E. coli* Lgt. The image is adapted from ⁶³ showing the structure of *E. coli* Lgt with its different domains. The TM helices are numbered 1 to 7.

- **The signal peptidase II (LspA)**

LspA is an endopeptidase that catalyzes the second step of lipoprotein maturation. It cleaves the scissile peptide bond between the diacylglyceryl-modified cysteine (+1) and the (-1) (G/A/S) residue forming apolipoprotein with the cysteine as the first residue (**Fig.3**). The lipobox present in the signal sequence of the prolipoprotein is the recognition sequence of LspA.

The reaction has a “check-point”, meaning that LspA will not cleave a non-diacylated lipoprotein. The structure of LspA from two different bacteria, the Gram-negative *Pseudomonas aeruginosa* (*Pa*) and the Gram-positive organism *Staphylococcus aureus* (*Sa*) were recently reported ^{70,71}. In both organisms, LspA appears as a monomer with four transmembrane α -helices and a periplasmic domain with an amphiphilic β -sheet (β -cradle) that rests on the membrane and a periplasmic helix (PH) loop (**Fig.5**). Interestingly, the PH blocks direct access from the periplasm to the active site suggesting that the substrate laterally binds LspA.

The structures were solved in complex with antibiotics (globomycin for PaLspA and SaLspA, and myxovirescin for SaLspA), which were found to block access to a region identified as the catalytic site. The natural cyclic amphiphilic compound, Globomycin, produced by the Gram-positive genus *Streptomyces* ⁷², inhibits LspA by binding its catalytic site and leads to the accumulation of the prolipoprotein in the IM ^{73,74}. Myxovirescin, or antibiotic TA, is another natural inhibitor of LspA produced by *Myxococcus sp.* ⁷⁵ that binds to the catalytic site. The active site contains two aspartate

residues proposed to represent a catalytic dyad; the catalytic aspartates are located in the transmembrane domain and point towards the outer surface of the membrane^{70,71}. Protonation of the NH group of the cysteine, in the catalytic dyad, leads to peptide bonds hydrolysis and therefore the cleavage of the signal peptide⁷⁶.

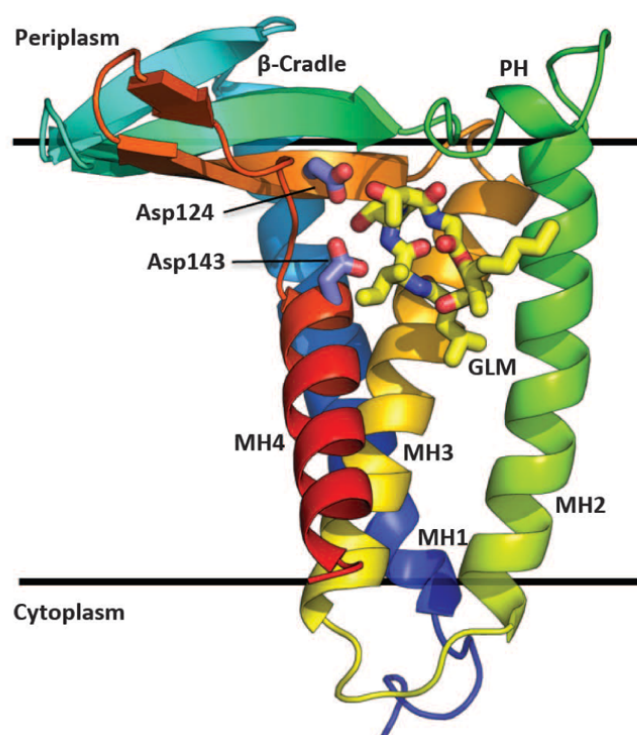


Figure 5. Structure of LspA from *P.aeruginosa* in complex with globomycin. Globomycin (GLM) is shown as stick yellow figure. TM helices are numbered MH1 to MH4. The image is adapted from⁷⁰.

○ **Apolipoprotein N-acyltransferase (Lnt)**

The last maturation step of lipoproteins is catalyzed by Lnt: the addition of a third acyl chain to the diacylated N-terminal cysteine (**Fig.3**). Lnt belongs to the nitrilase superfamily enzymes that exist in animals, plants, fungi and certain prokaryotes. Nitrilases superfamily consists of thiol enzymes involved in natural products synthesis

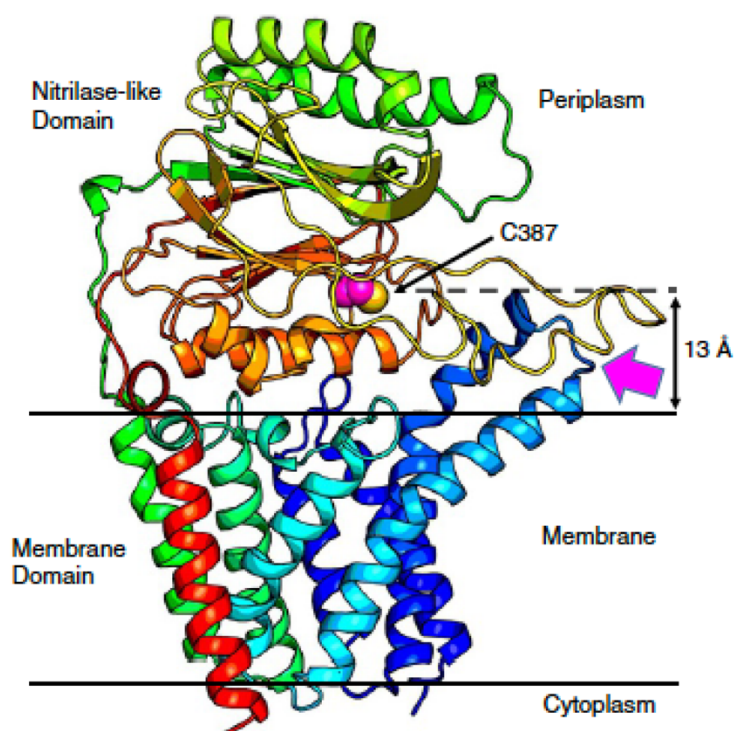
and post-translational modifications. They catalyze nitrile and amide hydrolysis and N-acylation of proteins. Nitrilase-related proteins are multimeric α - β - β - α sandwich proteins with a conserved catalytic site: glutamate-lysine-cysteine ⁷⁷. Global and structure-based sequence analysis classifies members of this superfamily into different branches. Some, like Lnt, have an additional globular domain fused to the nitrilase-related domain ⁷⁷.

Recent publications reported the crystal structures of *E. coli* ⁷⁸⁻⁸¹ and *P. aeruginosa* Lnt ⁷⁹. Both enzymes are remarkably similar and present two domains, although they have only 39% sequence identity. Interestingly, Lnt of *P.aeruginosa* is functional in *E. coli* and can complement *Int* depletion ⁸². Lnt consists of a transmembrane domain with eight α -helices providing support to a periplasmic nitrilase-like domain (**Fig.6a**). Both domains are linked by two brace-like linkers. The region between the two domains is the putative active pocket for substrate to enter and product to leave. Highly conserved hydrophobic residues form a small channel leading from the active site toward the cytoplasmic membrane. This hydrophobic cleft is enclosed by a loop from each side (**Fig.6c**) ^{78,79}. A catalytic triad consisting of a cysteine, a glutamate and a lysine is present in the periplasmic nitrilase domain, residing slightly above the membrane surface (**Fig.6b**). The cysteine of the catalytic triad is about 13 Å away from the center of the transmembrane domain (**Fig.6a**). Interestingly, helix 3 (colored in dark blue in Fig.6 a) has three conserved glycines, close to each other, which provide a certain flexibility. This flexibility might be important to accommodate and orient the vast variety of lipoproteins into the active site ⁷⁹. Moreover, a flexible amphipathic disordered periplasmic loop plays the role of a lid protecting the active site. On the other hand, the movement of this loop enables the catalytic site to be more accessible to the substrates ^{80,81}.

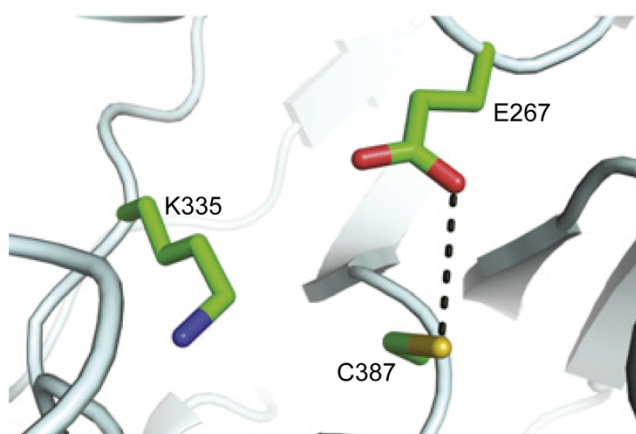
Finally, the available biochemical and structural data support a two-steps ping-pong mechanism for the N-acyltransferase reaction^{78,83-85}. First, an acyl chain is transferred from a phospholipid substrate (preferably phosphatidylethanolamine) to form a thioester linkage with the catalytic cysteine of Lnt. In a second step, the acyl chain is transferred from the cysteine to the α -amino group of the N-terminal cysteine of the lipoprotein (**Fig.3b**).

Altogether, the high-resolution structures of Lgt, LspA and Lnt set the stage to the identification of novel antibiotics targeting lipoprotein maturation^{86,87}; novel fluorescent-based assays that were developed to monitor the activity of Lnt and LspA *in vitro* will facilitate the search for specific inhibitors of these enzymes^{88,89}. The conservation of these enzymes will be detailed in section 2.3 of this chapter.

a



b



c

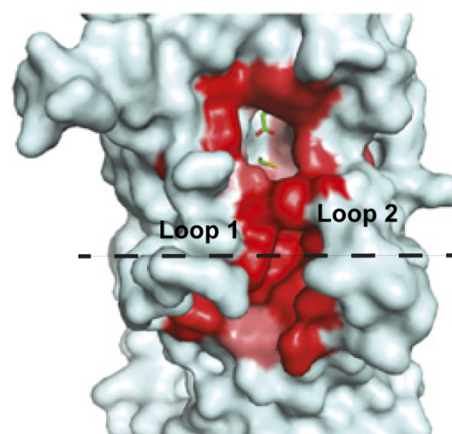


Figure 6. The overall structure of *E. coli* Lnt. **a.** Lnt has a membrane domain and a periplasmic nitrilase-like domain. The catalytic cysteine 387 side chain is shown in sphere representation (magenta: carbon, yellow: sulfur). **b.** Catalytic residues of Lnt are shown in green stick representations. **c.** Surface representation of the hydrophobic groove. Catalytic residues are shown in green stick representation. Residues lining this channel are colored in a white-to-red scale based on the hydrophobicity degree:

red color represents the most hydrophobic residues. Image a is adapted from ⁷⁹ and images b and c are adapted from ⁷⁸.

2.2. Sorting and targeting of lipoproteins

Once matured, lipoproteins either remain in the IM or are exported to the OM. In *E. coli*, most lipoproteins (~90%) are targeted to the OM ^{90,91}. Lipoprotein extraction from the IM and transport to the OM is carried out by the lipoprotein localization (Lol) system ⁹². Below we will discuss in detail sorting and targeting of lipoproteins by the Lol system.

2.2.a. LolCDE

LolCDE is an IM ATP binding cassette (ABC) transporter which extracts lipoproteins from the IM at the expense of cytoplasmic ATP hydrolysis and transfers them to the soluble periplasmic chaperone LolA. LolA then shuttles the lipoprotein across the periplasm for delivery to LolB, an OM lipoprotein in charge of OM insertion (**Fig.7**). It was shown that the interaction between lipoproteins and LolCDE, LolA and LolB is hydrophobic and sensitive to the dodecyl maltoside detergent (DDM) ^{93,94}. This is expected since lipoproteins have the triacylated tail that needs to be protected from the hydrophilic periplasm.

- LolCDE: an ABC transporter

The ABC transporters superfamily is a class of transporters that use ATP hydrolysis as a source of energy to translocate substrates, ranging from ions to macromolecules, across a membrane. They can function either as importers or as exporters. Around 5% of the genome of *E. coli* encodes ABC transporters. Only three of the 79 ABC

transporters of *E. coli* are essential for bacterial growth: LolCDE, as well as MsbA and LptBFG, which are two transporters both involved in the biogenesis and translocation of LPS⁹⁵⁻⁹⁷. The architecture of ABC transporters is conserved: they are formed of two transmembrane domains (TMD) and two nucleotides binding domains (NBD), located in the cytoplasm. TMDs form the ligand binding sites whereas each NBD forms an ATP binding and hydrolysis site. NBDs are characterized by the presence of highly conserved motifs. In contrast, TMDs are quite variable among ABC transporters and their features depend on the translocated substrates⁹⁸. The transporter bounces between an open free state and a closed ATP-bound state: ATP binding causes the dimerization of the two NBDs, which induces a conformational change in the TMDs that is important for the translocation of the substrate. The release of phosphate and NDP, after ATP hydrolysis, weakens the NBD dimer and discharges the substrate⁹⁹.

The LolCDE complex is formed of two TMDs (LolC and LolE), and two copies of LolD which constitute the NBDs (**Fig.7**). LolC and LolE are homologous to each other with a sequence identity of 26%¹⁰⁰. They span the IM four times and open to the periplasm by a large periplasmic domain.

It has been shown that the lipoprotein interacts with LolE while LolC acts as a scaffold for LolA, and that ATP hydrolysis by LolD is important for the transfer of the lipoprotein from LolE to LolA^{101,102}. The LolCDE complex is able to bind one molecule of lipoprotein and this binding increases the affinity of LolD to ATP⁹³.

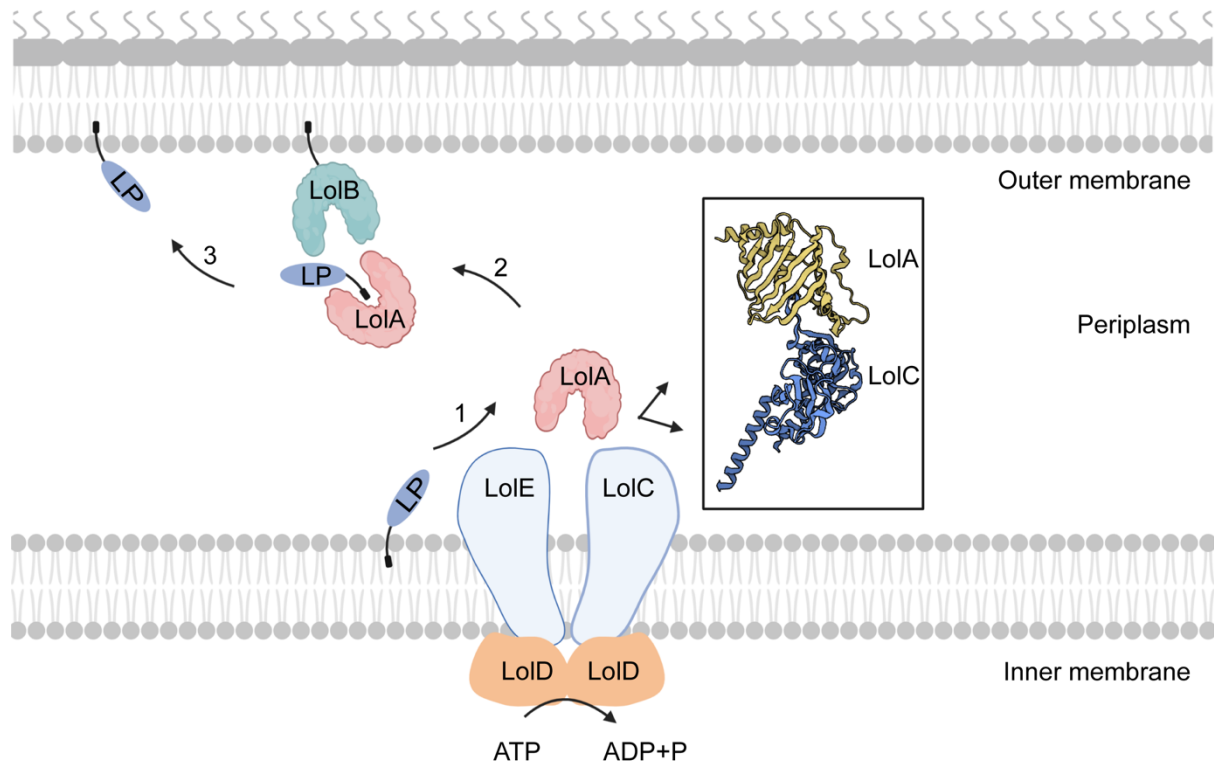


Figure 7. The Lol system transports lipoproteins to the outer membrane. Lipoproteins which are destined to the outer membrane are extracted from the inner membrane by LolCDE, an ABC transporter, and delivered to the periplasmic chaperone LolA (1). The mouth-to-mouth interaction between LolA and LolB delivers lipoproteins to LolB (2), which finally inserts lipoproteins in the outer membrane by a poorly known mechanism (3). The crystal structure of LolA bound to the periplasmic domain of LolC (pdb: 6F3Z) is shown.

- The Lol sorting signal

At the N-terminal part of the lipoprotein, the two residues (+2, +3) that follow the cysteine make what is known as the Lol sorting signal (**Fig.3a**). In *E. coli* and other *Enterobacteriaceae*, the general view is that Lol uses the residues adjacent to the lipidated cysteine to sort inner and outer membrane lipoproteins. If the residue at position +2 is an aspartate, the lipoprotein remains in the IM^{103,104}, most likely because of strong electrostatic interactions between the negative charge of this residue and the positive charge of the amine group of phosphatidylethanolamine that prevent

lipoproteins from interacting with LolCDE ¹⁰⁵. Moreover, this negative charge of the Lol avoidance signal should be within a certain distance, around 2.7 Å, from the Cα of the second residue. Noteworthy, the residue at position +3 also plays a role and interacts with the phosphatidylethanolamine: aspartate, glutamate and glutamine strengthen inner membrane retention, whereas histidine and lysine weaken it ^{103,104}. It is clear that all these interactions form tight bonds between the lipoprotein and the IM, making it difficult for the LolCDE to extract the lipoprotein.

Different sorting “rules” have been described in other bacteria. In *P. aeruginosa* for instance, an aspartate is rarely found at position +2 and inner membrane retention appears to be determined by residues at position +3 and +4 ^{106,107}, while in *Borrelia burgdorferi*, the causative agent of Lyme disease in humans, lipoproteins are secreted to the OM by default ¹⁰⁸. Moreover, the unfolded state of the N-terminal region of lipoproteins in *Borrelia* was shown to be important for the OM targeting ¹⁰⁹. Interestingly, it was recently shown that lipoproteins are correctly sorted in *P. aeruginosa* cells expressing the *E. coli* LolCDE complex, despite differences in the retention signal between these two species ¹¹⁰. To explain this surprising result, the Lory lab proposed a general mechanism in which inner membrane retention results from the ability of the N-terminal domain of IM lipoproteins to adopt a conformation favoring strong interactions with the phospholipid bilayer ¹¹⁰. This speculation might be supported by a previous study where different phospholipids affected the sorting of lipoproteins and the activity of LolCDE in *E. coli* ¹¹¹.

- Possible lipoprotein's extraction mechanism and interaction with LolA

Even though the Lol sorting signals are well characterized in some organisms, as explained above, the molecular details of how a lipoprotein is extracted from the IM by the LolCDE complex are still not well established. However, in the last few years, progress has been made to reveal the molecular details of the interaction between a lipoprotein, LolCDE and LolA.

First, the crystal structure of the periplasmic domain of LolC in complex with LolA was solved and showed that two regions of LolC, known as hook and pad, are important for the interaction with LolA ¹⁰² (**Fig.7**). The hook is formed by a β -hairpin structure and the pad is a trio of surface residues. The structure showed that hydrophobic interactions, hydrogen bonds and salt bridges exist between LolA and LolC ¹⁰². Moreover, it was previously shown that LolA interacts with LolC in a mouth-to-mouth model and uses the same region to interact with LolC and then with LolB ¹¹².

Second, the sequence homology between LolCDE and MacB, a structure determined ABC transporter involved in pumping toxic compounds out of the bacterial cell ¹¹³ helped researchers to study the structure of LolCDE. In fact, modeling of LolCDE in complex with LolA was established ¹⁰², and showed that LolA is located $\sim 60\text{\AA}$ away of the IM, suggesting that mechano-transmission is implicated in lipoprotein extraction and delivery to LolA. Another mechanism was recently proposed based on molecular dynamics simulations that show that LolC facilitates lipoprotein transfer to LolA by destabilizing the IM ¹¹⁴. In this study, the periplasmic domain of LolC, which usually interacts with LolA (**Fig.7**), was shown to be close to the IM in the resting state (ATP free state) which causes a deformation in the IM and a membrane thickness of about

18 Å. Whereas no such deformation was observed in the active state (ATP-bound state) (**Fig.8a**). This observation suggests a model where the lipoprotein is directly transferred from the IM to its binding site in LolA ¹¹⁴ (**Fig.8a**).

Finally, the cryogenic electron microscopy (cryo-EM) structure of the full *E. coli* LolCDE complex was recently solved providing structural insights into the mechanism of lipoprotein extraction ^{115,116} (**Fig.8b,c**). Structures of LolCDE were obtained in different states: lipoprotein-bound, pre-transport state in complex with LolA, ATP-bound state and the apo- or free-state (**Fig.8b**). LolCDE consists of a heterodimer of TMDs; LolC and LolE, and a homodimer of nucleotide binding domains (LolD). LolC and LolE each contains four transmembrane domains with a periplasmic extension around 20 Å away from the IM (**Fig.8b**). The structure of liganded LolCDE with a lipoprotein, likely Lpp, shows outward-open TMDs forming a V-shape central channel with two lateral gates (**Fig.8b**). The lipoprotein is vertically associated with LolCDE, with the triacyl chain pointing down and the ten visible amino acid of the lipoprotein pointing up (**Fig.8b**). This orientation of the lipoprotein in complex with LolCDE is the same as the orientation of the lipoprotein in the IM during its maturation. Thus, this suggests that the lipoprotein is laterally extracted from the IM by the LolCDE. Moreover, the structure shows that the acyl chains of the lipoprotein straddle the channel of LolCDE, the four N-terminal residues of the lipoprotein are buried in the channel whereas the rest of the polypeptide interacts with the periplasmic part of LolE ¹¹⁵ (**Fig.8b,c**). Regarding LolA, it seems that it binds to LolC only after the lipoprotein was extracted from the IM, and that its hydrophobic cavity faces the central channel of LolCDE (**Fig.8b,c**).

The final proposed mechanism, based on the obtained structures, is that the acyl chains of a lipoprotein laterally bind to the channel of LolCDE with the rest of the polypeptide interacting with LolE. LolA binds to LolC and ATP binding induces dimerization of LolD and the closure of the channel (**Fig.8b,c**). Upon ATP hydrolysis, LolD dimers fall apart inducing a twist in LolC which aligns LolA to the hydrophobic interior side of the periplasmic domain of LolC. At this stage, a kind of tunnel is formed between LolA and LolC which enables the acyl chains of the lipoprotein to be transported, from LolCDE to LolA, and the dissociation of the lipoprotein from LolCDE (**Fig.8b,c**). At this point, LolCDE is ready to receive another lipoprotein and to start a new cycle.

Based on these results, the first steps of lipoprotein binding to LolCDE are independent of ATP binding, and the lipoprotein is bound in the upper channel of LolCDE elevated from the IM (**Fig.8b,c**). The remaining question anyone can ask is that how can this earlier step be energy independent? Well, the previous molecular dynamic results ¹¹⁴ suggest that membrane deformation, at the periplasmic domain of LolC, brings phospholipids closer to LolC and thus reduces the energy requirement for lipoprotein extraction ¹¹⁴ (**Fig.8a**). This membrane deformation might be caused by the conformational changes of LolC during ATP hydrolysis (**Fig.8b,c**). Thus, lipoprotein extraction might occur laterally at an elevated membrane level.

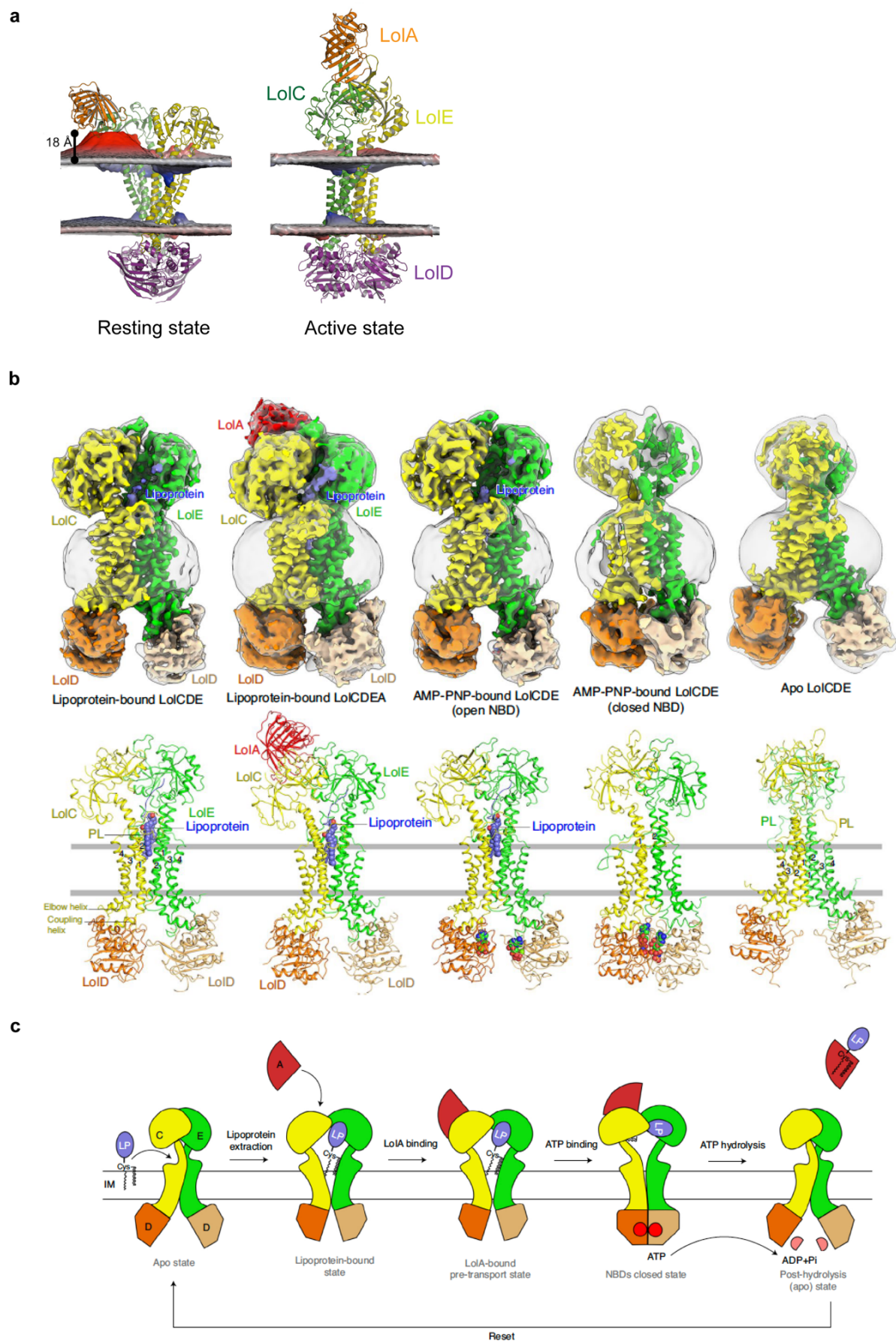


Figure 8. Extraction of the lipoprotein from the IM by the LolCDE complex and its transport to the periplasmic chaperone LolA. **a.** Molecular simulation of the LolCDE complex showing membrane deformation close to the periplasmic domain of LolC, with the position of the phosphate changing by up to 18 Å. LolC, D, E and A are shown in green, purple, yellow and orange, respectively. The image is adapted from ¹¹⁴. **b.** Cryo-EM structures of LolCDE-LolA in complex with a peptide in different states. **c.** Schematic representation of lipoprotein extraction and transfer from LolCDE to LolA. In b and c, LolC, D, E, A and the lipoprotein are shown in yellow, orange, green, red and blue, respectively. (PL: periplasmic loop). The images in b and c are adapted from ¹¹⁵.

2.2.b. LolA and LolB

ATP hydrolysis by LolD causes lipoproteins to be released and delivered to the periplasmic chaperone LolA. After crossing the periplasm, LolA delivers its lipoprotein substrates to LolB by a mouth-to-mouth interaction ^{101,112}.

LolA and LolB share a common structure despite the low sequence identity of 8%. Both proteins have eleven stranded anti-parallel β -sheet which folds into an incomplete β -barrel covered by three α -helices ¹¹⁷. LolA has an extra C-terminal loop and an extra β -sheet important for the efficient release of lipoproteins ¹¹⁸. The inner surface of LolA and LolB forms a hydrophobic cavity that binds the lipid moiety of substrate lipoproteins, thereby protecting the acyl chains from the hydrophilic periplasm. The hydrophobicity of LolB's cavity is higher than LolA's, which can explain the fact that the lipoprotein is transferred from LolA to LolB in a mouth-to-mouth interaction, based on the affinity, without any source of energy ^{94,112,117}. Moreover, the hydrophobic cavity of LolB is always open ready, to receive lipoproteins, whereas LolA's cavity is closed by hydrogen bonds and opens upon binding of a lipoprotein. One additional difference is that LolB has a protruding loop important for membrane targeting ¹¹⁹, whereas LolA has a C-terminal loop that is important for the releasing activity of LolA, and that prevents the LolB independent transfer of lipoproteins to the

membrane ¹¹⁸. These differences might be the reasons why LolA and LolB cannot functionally replace each other ¹²⁰.

Based on the structures, the proposed interaction mechanism is that LolA accommodates the acyl chains on its hydrophobic concave side. Upon interaction with LolB, the concave side of the β -barrel of LolA contacts with the convex side of the β -barrel of LolB. Thus, a kind of hydrophobic tunnel is formed by the hydrophobic insides of LolA and LolB (**Fig.9**). The acyl chains are therefore transferred through this tunnel from LolA to LolB by a mouth-to-mouth interaction ¹²¹. In the last step, LolB inserts the lipoprotein in the inner leaflet of the OM. However, the mechanistic details of this step are yet to be determined. However, it was shown that phosphatidylethanolamine is important for membrane anchoring of lipoproteins by LolB. This might be related to the fact that phosphatidylethanolamine causes curvature stress in the membrane and affect lateral pressure in the membrane ¹²⁰.

Finally, the hydrophobic cavity of LolA was considered large enough to only accommodate one of the acyl chains of the lipoprotein substrates ¹¹⁷, but recent molecular dynamics simulations questioned this idea, suggesting that it could in fact fit up to three ¹²².

At the end, I would like to mention that until this day, there is no structure of a lipoprotein in complex with any of the components of the Lol system that is available. The determination of the structure of such complexes would definitely bring more insights into the real mechanism used by lipoproteins to pass from the IM to the OM, which would be of a great interest to find new inhibitors of the lipoprotein's pathway.

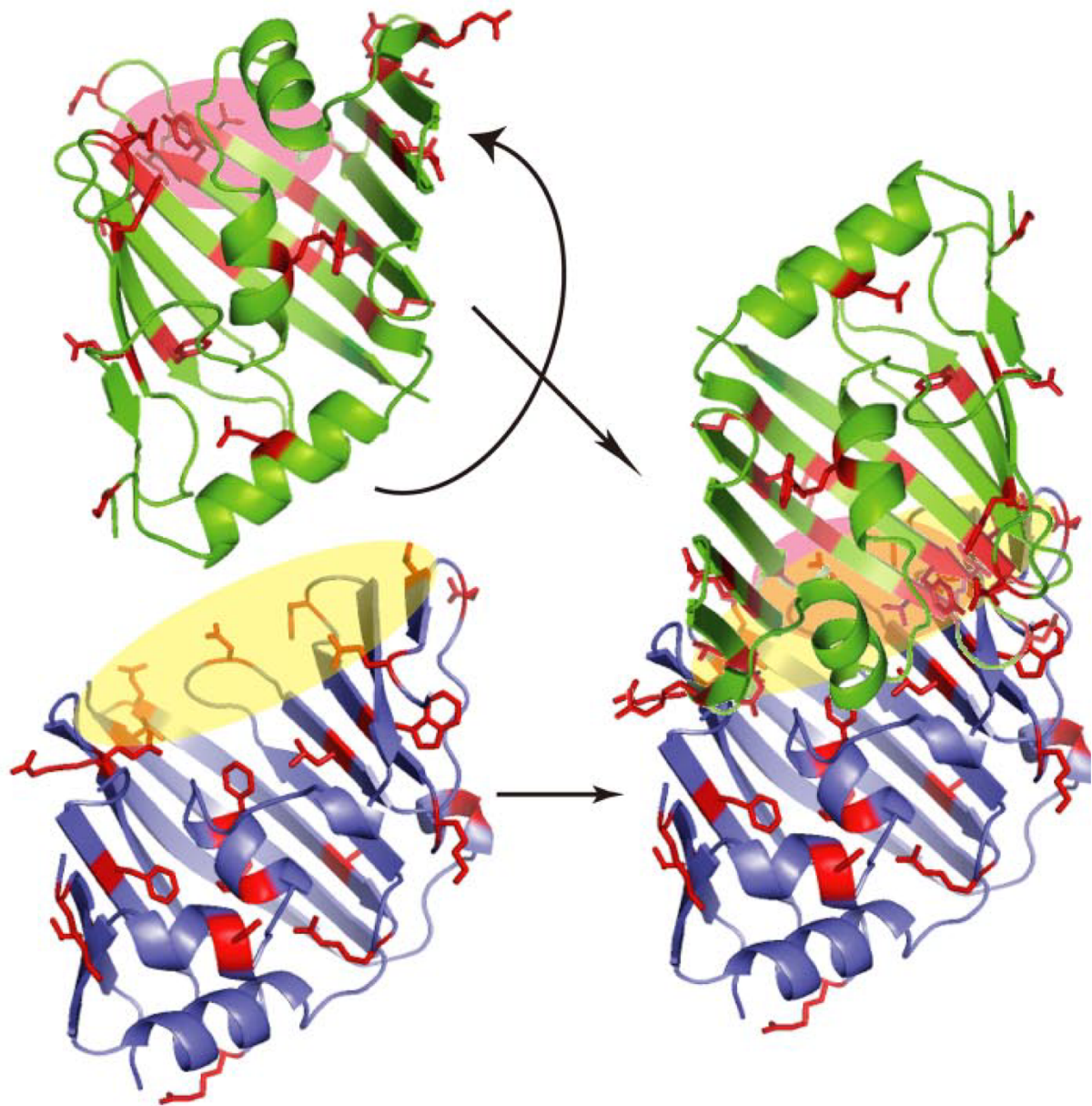


Figure 9. Mouth-to-mouth interaction of LolA and LolB. A hydrophobic tunnel is formed by the inside cavities of LolA and LolB enabling the transfer of the lipoprotein. LolA and LolB are presented in blue and green, respectively. The image is adapted from ¹¹².

2.3. Conservation of the lipoprotein pathway

The evaluation of a protein's conservation is very crucial as it gives us an idea on how important the protein is. Below we will discuss the conservation of all the proteins involved in lipoprotein's pathway.

The two maturation enzymes, Lgt and LspA are conserved in all bacteria. However, Lnt is only present in proteobacteria and actinomycetes ⁵⁹⁻⁶².

LolA is well conserved in Gram-negative bacteria, but LolB is only present in β - and γ -proteobacteria ¹²³, which indicates that other mechanisms mediating lipoprotein insertion into the OM remain to be identified (**Fig.10**). In line with this idea, it was recently reported that LolA and LolB, but not LolCDE, could be deleted under certain specific conditions (including *lpp* deletion) in *E. coli*, which suggests that at least one alternative trafficking route to the OM exists in this bacterium ¹²⁴. This idea will be further discussed in section 2.5.

It is interesting to note that Gram-negative bacteria other than γ -proteobacteria express a homodimer of LolF proteins instead of a heterodimer of LolC and LolE (**Fig.10**). LolF contains sequence features present in LolC and LolE. Bacteria that express LolF, such as *Francisella tularensis*, *Neisseria meningitidis*, *Acinetobacter* species and *Helicobacter pylori* can tolerate Lnt deletion¹²⁵⁻¹²⁸. Therefore, LolF can recognize not only triacylated lipoproteins but also diacylated ones. In this context, LolF seems to be the ancestor of LolCE ¹²⁹.

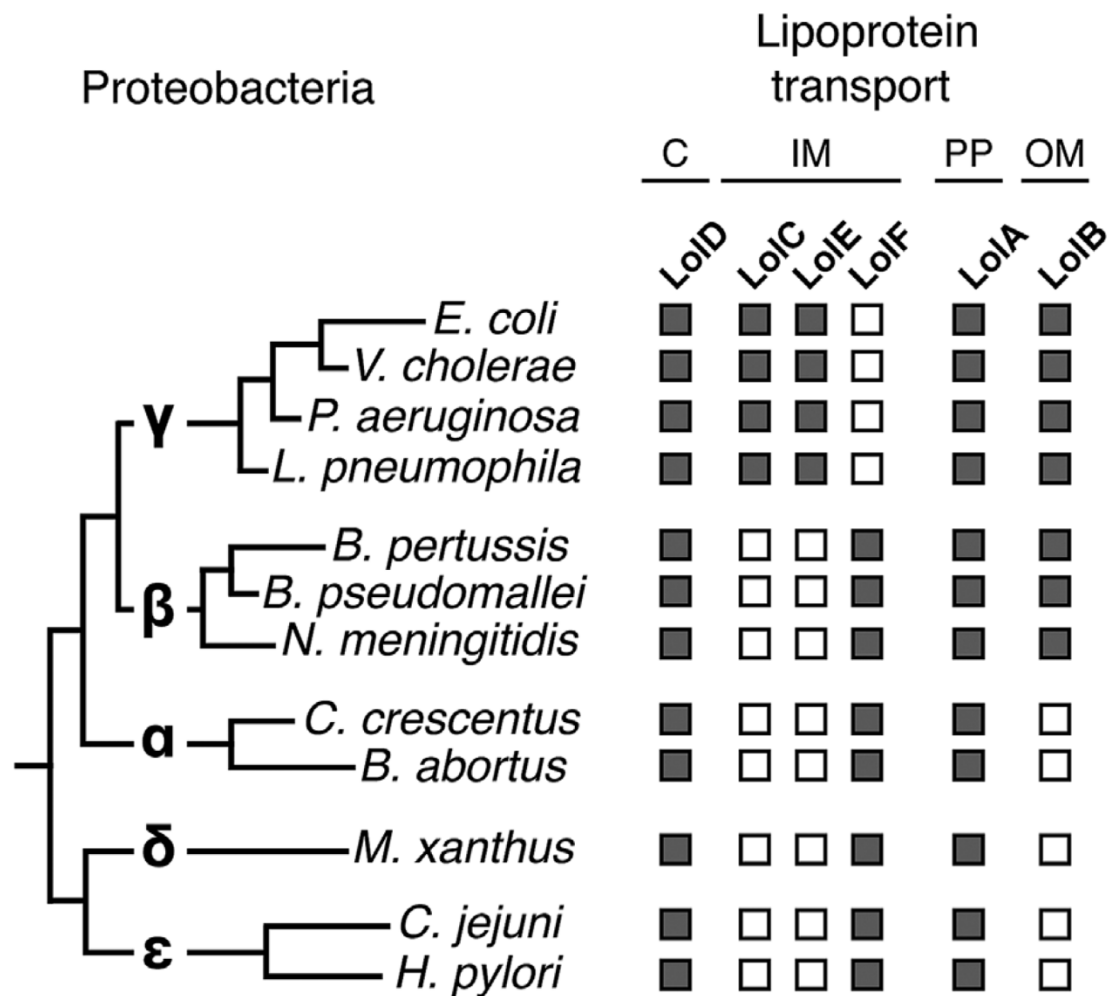


Figure 10. Conservation of the Lol proteins among proteobacteria. LolA and LoID are highly conserved. LoIC and E are only present in γ-proteobacteria and they are replaced by LoIF in other proteobacteria. LoIB is conserved in γ and β-proteobacteria. The image is adapted from ¹²³.

2.4. Surface exposure of lipoproteins

Some lipoproteins, after reaching the OM, cross the asymmetric lipid bilayer and reach the cell surface, using poorly characterized mechanisms ^{61,130}. Surface-exposed lipoproteins, which are attractive candidates for vaccine development ¹³¹ and are often required for virulence ¹³², are common in certain bacterial species such as *B.*

burgdorferi, where as many as two-thirds of all lipoproteins are surface-exposed ¹³³. Other bacteria commonly displaying lipoproteins on their surface include species from the *Bacteroides* and *Neisseria* genus ¹³⁴⁻¹³⁶. In contrast, only a handful lipoproteins have been identified on the *E. coli* cell surface. Caution is required when considering lipoprotein surface exposure: unambiguously determining whether a lipoprotein crosses the OM or not is technically challenging due to the high yields of false positive. Below we will discuss the different putative mechanisms used by lipoproteins to reach the cell surface.

Recently, the Renzi lab discovered the presence of a lipoprotein N-terminal export signal in the sequence of lipoproteins present in the *Bacteroidetes* phylum ¹³⁷. This sequence, which consists of conserved polar residues between positions +2 and +6, is essential for the surface exposure of lipoproteins. This motif was not identified in the sequence of lipoproteins present in the OM facing the periplasm. Interestingly, the introduction of this motif in an OM lipoprotein that faces the periplasm leads to its surface exposure ¹³⁷. Another example of surface exposure-sequence relationship is in the *Borrelia* genus. In fact, the unfolded state of the N-terminal domain of the lipoprotein is essential for the lipoprotein to reach the surface of *Borrelia* ¹⁰⁹. Interestingly, both *Borrelia* and *Bacteroidetes* lack LolB, which implies that other molecular mechanisms exist in these bacteria.

A few lipoproteins, including Pal and Lpp, were detected at the cell surface of *E. coli*. However, little is known about how these lipoproteins reach the surface; LolB was previously described by Tokuda's lab to be responsible of anchoring lipoproteins to the inner leaflet of the OM by a poorly known mechanism ¹¹⁹. The next step, how lipoproteins cross the OM and reach the cell surface, is still ambiguous.

Lpp exists in two forms: covalently bound to the PG and free Lpp, at a ratio of 1:2 ¹³⁸. The C-terminal lysine of Lpp is covalently attached to the PG in the periplasm whereas the free form of Lpp was detected at the cell surface ¹³⁹. Pal, which participates in the Tol-Pal complex that spans the cell envelope and is important for cell integrity and survival, was also shown to interact non-covalently with the PG ^{17,140} and to be surface-exposed. The majority of Pal was shown to be internal and only 10 to 20% was surface-exposed ¹⁴¹.

The most important question is how do these lipoproteins reach the surface? There is no direct evidence on the mechanism used by these lipoproteins to reach the surface. The most evident hypothesis is the existence of a potential lipoprotein-flippase in the OM. It is obviously clear that more studies are required to identify the mechanism used by some lipoproteins to reach the cell surface. This would be a huge step in the microbiology field, as these potential flippase/ transporter would consist of an easy antibacterial target for the design of new antibiotics.

Outer membrane proteins (OMPs) are used for the surface exposure of some lipoproteins

The only *E. coli* lipoprotein that has been shown to reach the cell surface by different groups and complementary techniques is the stress sensor RcsF ¹⁴²⁻¹⁴⁴. Therefore, RcsF has become a paradigm to investigate the mechanisms of lipoprotein surface exposure in *Enterobacteriaceae*. According to the current model, RcsF is exported to the cell surface by the β -barrel assembly machinery (BAM) via the assembly of complexes with β -barrel proteins (**Fig.11**). The crystal structure of a complex between RcsF and BamA, the core BAM component, has recently been solved by our lab¹⁴⁵. In

this complex, which is considered to be an intermediate in the formation of the RcsF- β -barrel complexes, RcsF binds inside the lumen of the BamA β -barrel interacting with regions proposed to undergo a lateral opening during β -barrel insertion. The abundant β -barrels OmpA, OmpC and OmpF were identified as potential vehicles for RcsF surface-exposure ^{142,143}. However, by dissecting the OmpA-RcsF interaction *in vivo* and *in vitro*, a recent publication by our lab showed that RcsF interacts with the C-terminal periplasmic domain of OmpA, not with the N-terminal β -barrel. We therefore concluded that OmpA does not mediate the surface exposure of RcsF and predicted that only OmpC and OmpF (which form large β -barrels) serve as vehicles for surface exposure ¹⁴⁶ **(Fig.11)**.

Beside RcsF, other lipoproteins were also detected on the cell surface. BamC, a lipoprotein of the BAM complex, is surface-exposed ¹⁴⁷. While the unstructured conserved N-terminal domain of BamC is important for the interaction with BamD, another lipoprotein of the BAM complex, the large folded C-terminal domain of BamC is surface-exposed ¹⁴⁸. A hydrophobic conserved motif in the N-terminal domain of BamC is thought to cross the membrane enabling the surface exposure of C-terminal BamC. Interestingly, surface exposure of some lipoproteins in *B. burgdorferi* was shown to be reduced when a putative BamA orthologue was depleted ¹⁴⁹. However, more data are needed to demonstrate if this decrease is a direct result of BamA depletion.

Finally, while BamA appears to be important for the surface exposure of some lipoproteins, others do not need BamA to reach the cell surface. For instance, the surface exposure of the pore-forming lipoproteins, CsgG ¹⁵⁰ and Wza ¹⁵¹, is BamA independent. In fact, monomers of CsgG and Wza are transported to the OM where

they form homo-oligomers in a BamA-independent manner ¹⁵². In the final conformation of CsgG and Wza, parts of the pore are surface-exposed.

Surface Lipoprotein Assembly Modulators (SLAMs) as a surface exposure pathway for lipoproteins

Beside the OMPs, SLAMs is another pathway for lipoproteins to reach the surface ¹⁵³ (**Fig.11**). SLAM consists of an N-terminal domain of tetratricopeptide and a 14-stranded barrel domain (**Fig.11**). The barrel domain is essential for the function. It was first identified in *Neisseria meningitidis*, but it is also presents in β -proteobacteria ^{154,155}. However, SLAM was not identified in *Bacteroidetes* ¹³⁷. Thus, the surface exposure mechanism of lipoproteins in this phylum remains to be discovered.

Neisserial lipoproteins essential for virulence, such as transferrin binding protein B (TbpB) and factor-H binding protein (fHbp), were shown to have a long unstructured N-terminal linker and to use the SLAM to reach the cell surface. Interestingly, expression of TbpB in *E. coli* allowed this lipoprotein to reach the OM but not the surface ¹⁵⁵. However, when fused to Lpp'OmpA chimera of *E. coli* ¹⁵⁶, or when co-expressed with SLAM ¹⁵⁵, neisserial lipoproteins were able decorate the bacterial surface of *E. coli*.

Finally, structural analysis revealed that, despite their sequence diversity, putative lipoproteins that use the SLAM to reach the surface share a common structural domain: a flexible handle domain followed by an 8-stranded barrel domain¹⁵⁴.

Secretion systems are important for some lipoproteins to reach the surface

Beside the abundant OMPs in bacteria and Slam, other systems are used by lipoproteins to get to the surface. Among these systems, we cite the secretion systems.

We count ten different types of secretion systems in Gram-negative bacteria that are important for bacterial survival and virulence ^{29,157,158}. However, a new secretion system, type XI, was recently shown to facilitate the secretion of lipoproteins but also of non-lipidated proteins across the outer membrane ¹⁵⁹. This secretion system is composed of an outer membrane protein of the DUF560 family, a family of unknown function including Slam proteins, that is responsible of the translocation of soluble proteins and lipoproteins across the outer membrane of animal-associated bacteria, like *Neisseria* sp. Thus, this secretion system seems to be important for the virulence of these bacterial species.

These systems are used by bacteria to secrete small molecules, DNA and proteins into the extracellular environment or into target cells. Type V spans only the OM while the other types span both membranes. Here we will focus on types II and V secretion systems since they are used by some lipoproteins to reach the cell surface (**Fig.11**).

Type II secretion systems span the cell envelope and are used by bacteria to secrete hydrolyzing enzymes and toxins into the extracellular environment. As an example of lipoproteins that use this type of secretion system, we cite the lipoprotein PulA, (pullulanase) of *Klebsiella* sp. PulA bypasses the Lol system and uses the type II secretion system to reach the outer face of the OM ^{160,161} (**Fig.11**). Interestingly, the lipidation of PulA was shown not to be required for its secretion ¹⁶⁰. However, the

aspartate at the position +2 of PulA works as an IM retention signal in the absence of a type II secretion system ¹⁶².

Beside the type II, Type V secretion systems are also used by some lipoproteins to reach the surface. Type V, also known as autotransporter (AT), span the OM and are unique in that their substrate, also called the passenger domain, and their secretion pore, called the translocator domain, are fused and form one polypeptide (**Fig.11**). The translocator domain is inserted as a β -barrel in the OM which enables the secretion of the passenger domain through the pore of the β -barrel ¹⁶³. Once again, the Bam complex is important for catalyzing the membrane insertion of the β -barrel domain of the autotransporter. The lipoprotein NalP, of *Neisseria meningitidis*, is surface-exposed by the type V system and the passenger domain of NalP is then secreted by autolytic reaction. NalP is a serine-protease that catalyzes the release of other surface-exposed protein. Interestingly, the lipidation of the (+1) cysteine of this lipoprotein was shown to be important for the function of NalP, not for its biogenesis and secretion ¹⁶⁴.

Finally, a composite secretion system and a new class of lipoproteins have been recently identified in enterotoxigenic *E. coli*. Here, the cysteine lipidation is replaced by the acylation of a highly conserved N-terminal glycine residue ¹⁶⁵. Interestingly, the outer membrane protein CexE is the first member of this new class of lipoproteins. CexE is first cleaved by the signal peptidase I and its glycine residue is acylated by AatD, a Lnt homolog. CexE is then surface-exposed by the Aat composite system that shares similarities with the Lol and type I secretion system. Aap is another example of this new class of lipoproteins that is present in enteroaggregative *E. coli* ¹⁶⁵.

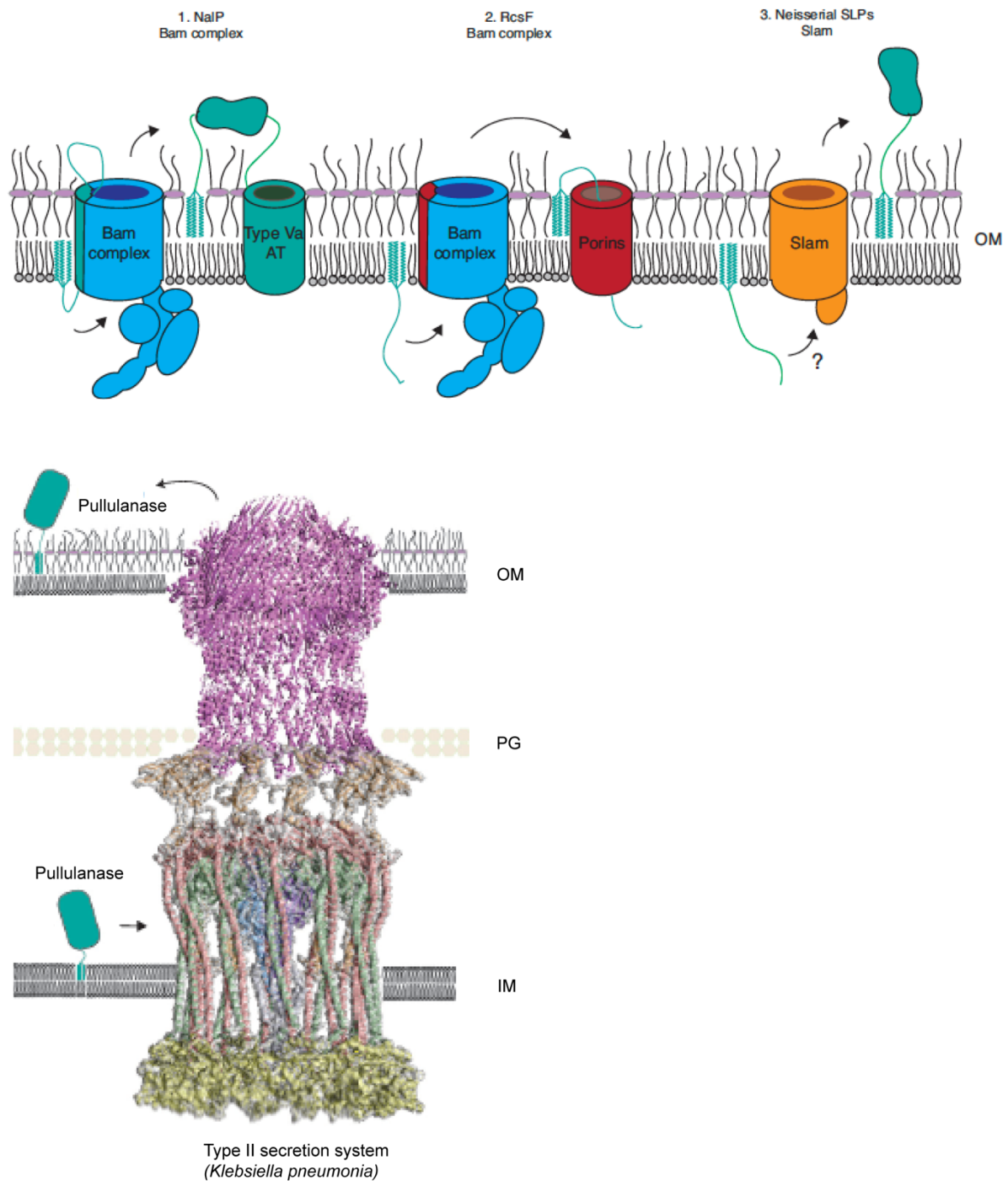


Figure 11. Examples of the different potential surface exposure systems present in Gram-negative bacteria. These systems include the autotransporter system, OMPs, SLAM and the type II secretion system used by NalP, RcsF, Neisserial lipoproteins and the pullulanase of *Klebsiella pneumoniae*, respectively. The image is adapted from ¹⁵³.

2.5. Diverse functions of lipoproteins

Now that we know about the processing of lipoproteins, it is time to go deeper into their functional importance and to discuss why they are attractive targets for the design of new antibiotics. Here I will describe the key features of selected lipoproteins. I focus on the lipoproteins used in my study.

RcsF

RcsF, a small OM lipoprotein, is the stress sensor of the Rcs phosphorelay pathway (**Fig.12a**). The Rcs is a two-component system found in many members of the *Enterobacteriaceae* family. It is activated by a variety of cell envelope perturbations affecting the OM, the PG and the periplasmic space ^{166,167}. These perturbations include exposure to cationic antimicrobial peptides, defects in LPS, treatment with β -lactam antibiotics or lysozymes, osmotic shock and deficiency in membrane-derived oligosaccharides.

The Rcs system is composed of many proteins (**Fig.12a**). RcsF is the first component of the Rcs cascade and is the only OM protein of the system. When RcsF senses envelope stress, it activates the phosphorylation cascade. Thus, RcsC, the IM histidine kinase, autophosphorylates and the phosphoryl group is then transferred to the IM protein, RcsD, before being finally transferred to the response regulator RcsB (**Fig.12a**). RcsB can then bind DNA to regulate gene expression as a homodimer or heterodimer with RcsA, a non-phosphorylatable transcription factor ^{168,169}.

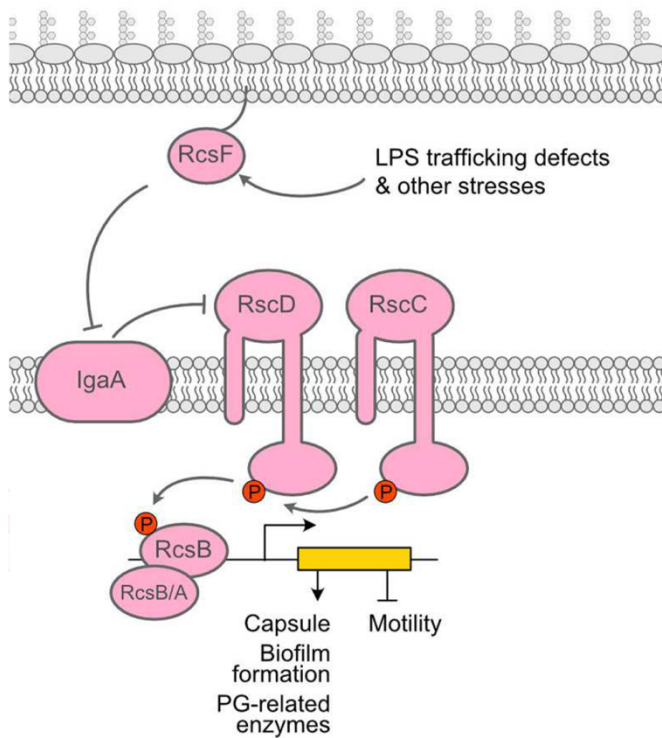
IgaA, an IM protein with a periplasmic domain, is the only essential component of the Rcs system, and it acts as a down regulator of the system by interacting with RcsD ^{142,170,171} (**Fig.12a**). Our lab showed that, in stress conditions, the interaction of RcsF

with the periplasmic domain of IgaA alleviates the inhibitory effect of IgaA on the Rcs pathway, thus inducing the cascade ^{142,170}. Whereas, in non-stressed conditions, RcsF is funneled away from IgaA by interacting with BamA and other OMPs, and therefore being surface-exposed ¹⁴².

In response to Rcs activation, the expression of many genes is modulated such as the genes involved in the biosynthesis of colanic acid, which is a protective exopolysaccharide that is exported to the surface of the bacteria ¹⁷². Production of this capsule enables us to easily detect the colonies where the Rcs is activated, because they appear mucoid. Moreover, many genes were also identified as members of the Rcs regulon such as genes that encode proteins associated with membrane integrity, bacterial motility, biofilm formation and cell division ¹⁷³⁻¹⁷⁶.

Finally, the NMR and the crystal structures of RcsF have been reported few years ago ^{177,178}. The structure shows that RcsF has an N-terminal intrinsically disordered region, or what we call a linker, followed by a C-terminal globular domain which is formed of 4 stranded β -sheets surrounded by 2 α -helices on the sides (**Fig.12b**).

a



b

MRALPICLVALMLSGCSM LSRPVEPVQSTAPQPKAEPKPKAPRAT PVRIYTNAEELVGKPFRLDGEVSGDSCQASN
QDSPPSIPTARKRMQINASKMKANAVLLHSCE VTSGTPGCYRQAVCIGSALNITAK

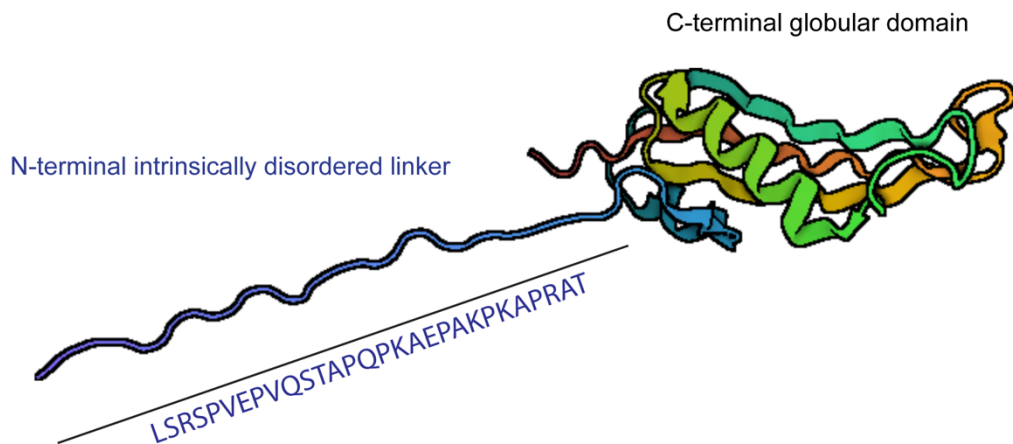


Figure 12. RcsF functional and structural characteristics. a. The Rcs phosphorelay cascade. RcsF senses envelope stresses and activates the

phosphorelay cascade. When the phosphate group reaches the response regulator, RcsB, expression of specific genes is modulated in response of the initial stress. The figure is adapted from ⁴⁷. **b.** NMR structure of RcsF with its amino acids sequence. RcsF is composed of an N-terminal disordered linker followed by a C-terminal globular domain. In the amino acids sequence, the signal sequence, the cysteine with the Lol sorting signal, the N-terminal linker and the globular domain are marked in pink, purple, blue and black, respectively. (pdb: 2L8Y)

NlpE

NlpE is an OM lipoprotein that is related to the Cpx two-components system (**Fig.13**). The Cpx stress response is widespread among *Enterobacteriaceae*. It is activated by perturbations of the envelope, mainly by misfolded proteins in the periplasm and when the PG integrity is challenged, for instance when the cell is exposed to β -lactam antibiotics ¹⁷⁹⁻¹⁸². Cpx is composed of the histidine kinase CpxA that autophosphorylates upon stress, and then transfers the phosphate group to the response regulator CpxR. CpxR then binds to promoters of genes implicated in protein folding and degradation, efflux pumps, periplasmic quality control, cell division, adherence, etc. ^{179,183} (**Fig.13**). Two more proteins, CpxP and NlpE, play important roles in the Cpx pathway. CpxP acts as the inhibitor of the autophosphorylation of CpxA, and NlpE was recently shown to monitor lipoproteins trafficking by activating the Cpx ^{182,184}. In fact, lipoproteins trafficking defects cause the accumulation of NlpE in the IM. Therefore, NlpE activates the Cpx, as a protective response, by interacting with CpxA. Finally, the overexpression of NlpE or its localization in the IM activate the Cpx ¹⁸⁰.

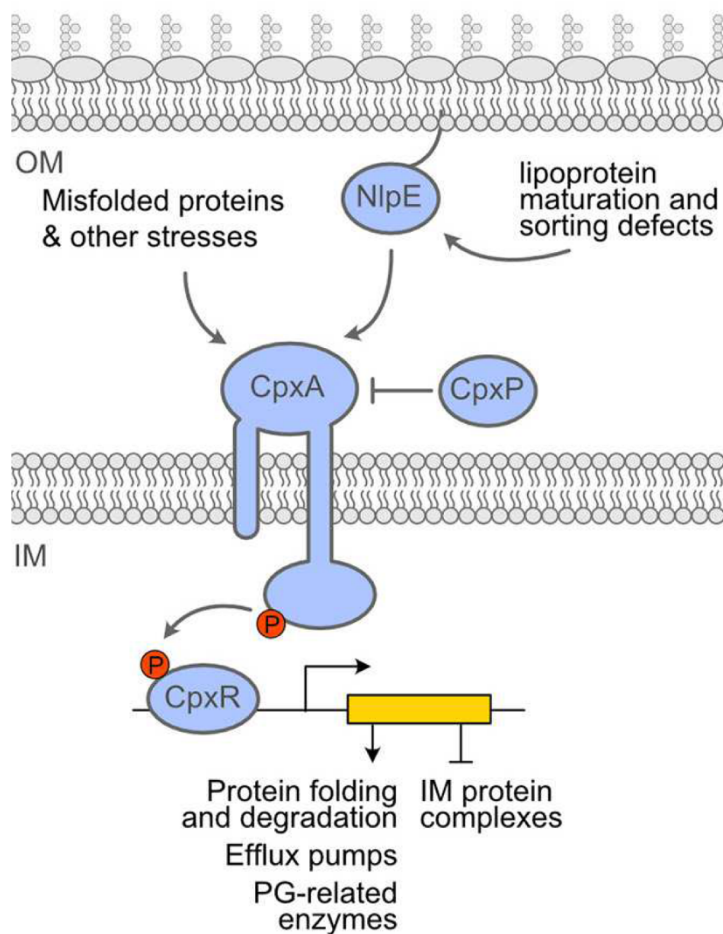


Figure 13. Representation of the Cpx stress response. Upon stress, CpxA autophosphorylates and then phosphorylates the response regulator CpxR. Thus, CpxR modifies the expression levels of many genes. CpxP is the inhibitor of the Cpx pathway, while NlpE is able of activating the Cpx system. The image is adapted from⁴⁷.

Interestingly, the activation of Cpx enables the deletion of *lolB* and *lolA* genes, but not *lolCDE*, in permissible conditions. These conditions include deletion of Lpp crosslinks with the PG (Lpp^{Δk58}) and the deletion of *osmB*, which encodes a small OM lipoprotein of unknown function¹²⁴. This is related to the fact that the accumulation of Lpp and OsmB in the IM is toxic to the cell; Lpp because it crosslinks with the PG, and OsmB is suspected to form ion permeable pores in IM which perturbs the IM integrity. The fact that LolA and LolB are needed in wild-type conditions means that they both ensure

a quick and efficient transfer of these lipoproteins, that become toxic when in the IM. Moreover, finding a condition where LolA and LolB can be deleted indicates that an alternative route for lipoproteins trafficking exists in the cell. However, this route seems to be like a “rescue plan” for the cell to ensure its viability. We could think that this alternative pathway functions only when LolA and LolB are absent or that, it is always active as a helper of the Lol system.

Lpp

Lpp is the numerically most abundant protein in *E. coli* with 1 million copies per cell¹⁸⁵. One third of the Lpp molecules is covalently attached to the PG, therefore Lpp functions as an α -helical tether covalently anchoring the OM to the peptidoglycan. This attachment is mediated by three enzymes of the L-D transpeptidase family: LdtA, LdtB and LdtC. Thus, the distance between the inner and the outer membranes is around 24.6 nm in *E. coli*. In the absence of this attachment, the OM pulls away from the cell body and forms blebs, increasing the distance between the two membranes up to 27.7 nm^{186,187}. Recent study showed that Lpp can also detach from the PG due to the action of the peptidoglycan meso-diaminopimelic acid protein amidase (DpaA)¹⁸⁸. Lpp was also shown to dictate the size of the periplasm: increasing the length of Lpp amplifies the distance between the inner and outer membranes^{187,189}. This IM-OM increased distance renders the cell unable to detect envelope alterations and to trigger the Rcs stress system as a response¹⁸⁷.

Finally, given the crucial role played by Lpp in *E. coli*, it was intriguing that this protein is present in only a small number of γ -proteobacteria. Recent work addressed this problem by discovering that in several Gram-negative species, the OM is covalently attached to the PG by β -barrel proteins^{190,191}. Conserved residues at the N-terminus

of these β -barrels mediate the attachment of these OMPs to the PG. For instance, conserved N-terminal alanyl residues are found in the periplasmic part of the OMPs of *A. tumefaciens*. In *C. burnetii* and in *Brucella abortus*, this tether is formed by N-terminal glycyl residues and an alanyl-aspartyl motif, respectively. Beside OMPs, lipoproteins other than Lpp can also mediate PG attachment. For instance, in *Coxiella burnetii*, attachment is ensured by the OM lipoprotein LimB¹⁹¹. LimB, by its lysine residue, is covalently attached to the mDAP of the PG, in a way similar to Lpp.

Pal

Pal is the OM component of the highly conserved Tol-Pal system¹⁹². The core of Tol-Pal consists of five proteins, three in the IM (TolA, TolQ and TolR), one in the periplasm (TolB) and the OM lipoprotein Pal. Like Lpp, Pal links the OM to the PG by binding to the free diaminopimelic acid residue in the peptidoglycan stem peptides¹⁷. This interaction is however non-covalent but mediated by hydrophobic contacts and hydrogen bonds.

The Tol-Pal system is important for cell division, for maintaining OM integrity and for the uptake of colicins and filamentous phage DNA¹⁹². In cell division, Tol-Pal stabilizes the OM-PG connection during constriction. Interestingly, the Tol-Pal proteins, which are dispersed along the cell envelope in non-dividing cells, are dynamically recruited to the divisome near the time of constriction¹⁹³. The relocation of Pal at the division site by TolB is an active process energized by the PMF¹⁹³⁻¹⁹⁵. An elegant mechanism was recently proposed for the mobilization of Pal by TolB and its deposition at the site of constriction in a process fueled by the PMF^{194,195}.

NlpD

NlpD is an OM lipoprotein that controls the activity of AmiC, one of the amidases that contribute to septum cleavage during cell division ^{196,197}. The OM localization of NlpD is important for regulating its activity ¹⁹⁸. Moreover, NlpD is crucial for coordinating OM invagination and PG remodeling at the division site. Interestingly, the Tol-Pal complex and YraP, an OM lipoprotein important for the OM integrity ^{199,200}, were shown to be required for the proper function of NlpD ¹⁹⁸. YraP, also named DoIP (Division and OM stress-associated lipid-binding protein), was shown to be also related to the divisome, but its exact function remains to be established ^{198,199}.

BamC

BamC is an OM lipoprotein associated with the Bam complex and its absence causes a destabilization of the complex ¹⁴⁸. BamC has a long unstructured N-terminal conserved domain, of 100 residues, that is important for the interaction with BamD, an essential component of the Bam complex ²⁰¹. This interaction seems to play a regulatory role in OMPs biogenesis, as the linker of BamC binds the proposed binding pocket of BamD ²⁰¹. Structural studies of BamC revealed that its globular domain is formed of two parts, with a helix-grip fold, connected by a short flexible linker ^{202,203}.

Finally, after mentioning the characteristics and the roles of some lipoproteins, I would like to say that these are only few examples of the different lipoproteins that decorate the cell envelope, but the list goes on. Each lipoprotein has its own structure and location in the cell envelope that are important for its function.

In conclusion, lipoproteins constitute an essential component of the cell envelope. Their diverse roles have been drawing a lot of attention lately, revealing how crucial these proteins are in a wide range of vital processes. However, this is only the tip of the iceberg: even in the well-studied model organism, *E. coli*, the function of most lipoproteins remains to be established. Finally, as mentioned before, some lipoproteins have an N-terminal intrinsically disordered region. The presence of this region and its role in the lipoprotein's pathway was the focus of my thesis and will be explored in the results section. In chapter III, we will discuss more in detail about the disorder in proteins.

2.6. Inhibition of the journey of lipoproteins

As we previously discussed at the beginning of this chapter, lipoproteins are widespread and play different roles in the cell envelope of Gram-negative bacteria. Since they fulfil a variety of different functions, they are essential for cell viability and survival. Given their important roles, lipoproteins together with their targeting machineries (the maturation enzymes and the Lol system) are considered a main target for the development of new antibiotics. Below, we will summarize the latest antimicrobials that have been discovered against the journey of lipoproteins.

- Molecules targeting the maturation enzymes

The three enzymes Lgt, LspA and Lnt are responsible for the maturation of lipoproteins in the cell envelope of Gram-negative bacteria. Therefore, these enzymes are an attractive target for the design of antibiotics.

The first inhibitor of Lgt, G2824, was recently identified ²⁰⁴. This molecule is a macrocyclic peptide that inhibits the catalytic reaction of Lgt (**Table 1**). Interestingly,

G2824 modestly inhibits growth of wild-type *E. coli* and *A. baumannii*, and this inhibition increases with *lpp* deletion.

Other compounds were found to inhibit the signal peptidase LspA: Globomycin and Myxovirescin (**Table 1**). Globomycin is a natural polar compound produced by different strains of Actinomycetes ⁷². It is specifically active against Gram-negative bacteria and inhibits LspA in a non-competitive manner ⁷⁴. Globomycin binds LspA *via* hydrogen bonds and hydrophobic interactions. Analogs of globomycin were developed and are more potent against LspA ²⁰⁵. These analogs were designed by modifying the backbone of globomycin to become more hydrophobic and by introducing a positive charged side chain, since the active site of LspA contains aspartate. Myxovirescin, also called Myxobacterium-produced antibiotic TA, is a natural compound as well. It is produced by Myxobacterium and inhibits LspA using a similar mechanism as the globomycin ⁷⁵. Interestingly, overexpression of LspA or deletion of *lpp* rescues the cells when exposed to Globomycin or Myxovirescin.

It is quite remarkable how Lpp has a great effect on the mechanisms of all these antimicrobials cited above. It was previously shown that the covalent linkage between Lpp and the PG from the inner membrane is lethal for the cell ²⁰⁶. Therefore, deletion of *lpp* or the expression of an Lpp variant that cannot crosslink to the PG rescue the lethal effect of Globomycin or Myxovirescin ^{75,206}. However, deletion of *lpp* increases the inhibition of Lgt when cells are exposed to G2824 ²⁰⁴. Due to these different effects of Lpp, we can speculate that, during its maturation, Lpp can crosslink to the PG once it gets diacylated by Lgt. However, this interpretation needs to be confirmed by further experimental work.

So far, no inhibitor of Lnt has been discovered. However, the development of novel *in vitro* fluorescent-based assays to monitor the enzymatic activity of Lnt ⁸⁸, as well as LspA ⁸⁹, are promising tools for high-throughput screening libraries of molecules, and the discovery of new inhibitors of these enzymes. Promising results were already reported ⁸⁹. In addition, the use of the high-resolution structures of Lgt, LspA and Lnt can help in the design of novel antibiotics targeting the maturation pathway ^{86,87}.

- Inhibitors of the Lol pathway

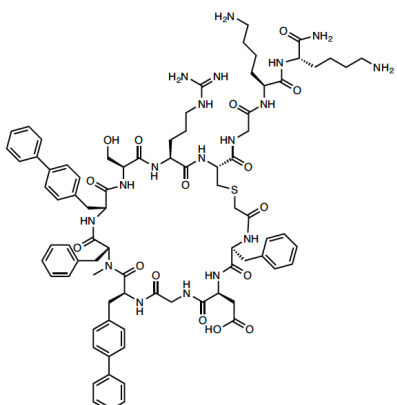
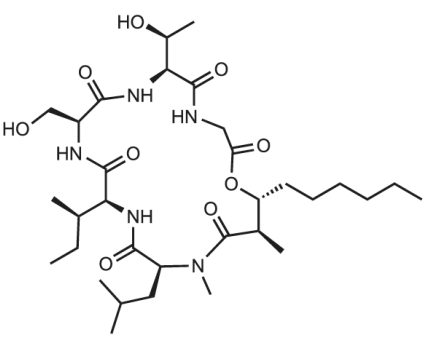
Beside the maturation enzymes, the Lol pathway is also a good target for the design of antimicrobials. Researchers were also focused on finding molecules that inhibit the components of Lol ²⁰⁷.

So far, three inhibitors of the LolCDE complex have been identified. First, pyridineimidazoles inhibit the release of lipoproteins from spheroplasts and inhibit LolCDE ²⁰⁸. Similarly, G0507 is a pyrrolopyrimidinedione that binds to LolCDE and inhibits the release of lipoproteins from the IM ²⁰⁹. Interestingly, the deletion of Lpp confers resistance against G0507 ²⁰⁹. Finally, compound 2 with a pyrazole core was also reported to have the same effect as G0507 ²¹⁰. Remarkably, the three components share high structural similarity (**Table 1**).

A22 is a known inhibitor of the cell wall biogenesis ²¹¹. MAC13243, a novel cyclic compound, and A22 were thought to inhibit the periplasmic chaperone LolA ^{212,213}. MAC13243 was also shown to permeabilize the OM of *E. coli* by inhibiting LolA ²¹⁴. As a result, cells become more sensitive to vancomycin, a large-scaffold antibiotic that usually cannot cross the OM due to its high molecular weight ²¹⁴. However, a recent

study showed that LolA is unlikely to be the target of the two components, A22 and MAC13243 ²¹⁵. These antimicrobials are more likely to inhibit the cell wall elongation machinery by inactivating MreB, the actin-like protein, which leads to the OM permeabilization ²¹⁵.

Finally, it is important to mention that most of these compounds are only active after permeabilization of the OM. It is therefore clear that, despite the significant effort to discover new antimicrobials compounds, developing antibiotics against Gram-negative bacteria remains a big challenge.

Compound	Chemical structure	Mechanism of action	Organism
G2824		Macrocyclic peptide inhibitor of the catalytic reaction of Lgt	Inhibition was shown in <i>E. coli</i> and <i>A. baumannii</i>
Globomycin		Non-competitive inhibitor of LspA	Moderate inhibition was shown in most Gram-negative bacteria and stronger inhibition was observed in <i>E. coli</i> and <i>K. pneumoniae</i>

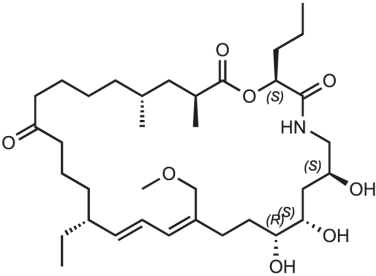
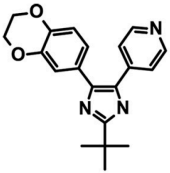
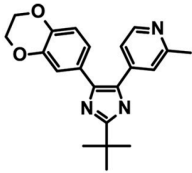
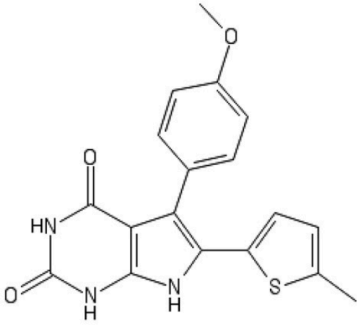
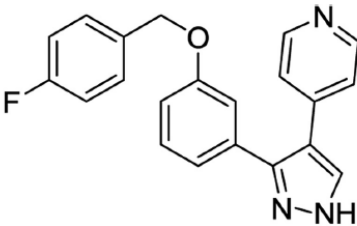
Myxovirescin		Non-competitive inhibitor of LspA	Inhibition was shown in <i>E. coli</i>
Pyridineimidazole	  Compound 1 Compound 2	Inhibitor of the release of lipoproteins by LolCDE	Inhibition was shown in <i>E. coli</i>
G0507		Inhibitor of the release of lipoproteins by LolCDE	Inhibition was shown in <i>E. coli</i>
Compound 2 (Pyrazole)		Inhibitor of the release of lipoproteins by LolCDE	Inhibition was shown in <i>E. coli</i>

Table 1. Different compounds targeting the pathway of lipoproteins in the cell envelope of Gram-negative bacteria.

Chapter III. Intrinsic disorder in proteins

Many years ago, the structure-function relationship of a protein was the main dogma in the biochemistry field. It has been always believed that a well-defined structure is a prerequisite of a protein function, especially when it comes to enzymes. Few years ago, a new concept of unstructured proteins, or what is also known by intrinsically disordered proteins (IDPs), was introduced. IDPs are proteins that cannot fold into a unique and stable secondary or tertiary structure. The structural disorder can be in certain regions of a protein. In this case, we can talk about intrinsically disorder regions (IDRs). In this chapter, I will discuss intrinsic disorder in proteins.

3.1. Overview on intrinsically disordered proteins

Around the turn of the millennium, many studies started to talk about the existence of intrinsically disordered proteins or regions in proteins that are unstructured. Of course, this new paradigm was hardly accepted at the beginning, but it then expanded. A data base dedicated to structural disorder in proteins, DisProt, was then created ²¹⁶. Today we can count about 50 bioinformatic algorithms predictors for disorder ^{217,218}. Most of the predicting methods are based on the amino acid sequence of a protein. The characteristics of a disordered sequence will be developed below, in section 3.1.a.

Structural disorder is abundant in all species, but it is significantly higher in eukaryotic than prokaryotic proteins ^{219,220}. For instance, while only around 5% of the proteins of *E. coli* are predicted to be disordered, this number increases to 28% in *Mus musculus* ^{218,221}. Interestingly, computational analysis showed that more than one third of eukaryotic proteins have IDRs of more than 30 residues in length ²²²⁻²²⁴. The fact that

eukaryotes have more IDPs than prokaryotes was recently explained by the idea that eukaryotic proteins have more extended linker regions, and that these linkers are more disordered in eukaryotes ²²⁵. Another comparative study between *E. coli* and the yeast *S. cerevisiae* showed that IDPs are three times more abundant in yeast than in the bacterium ²²⁶. This difference was explained by the fact that disorder exists in abundant proteins unique to yeast, such as proteins related to sporulation and meiosis and in cycles regulators ²²⁶. However, we should be aware that the increased level of disorder does not always increase with the complexity of an organism; for example, the highest level of disorder is found in protists, which are unicellular eukaryotes like algae ²²⁰.

Below we will discuss more in detail the structural disorder in proteins, how it is studied and the main functions of IDPs.

3.1.a Characteristics of intrinsic disorder

IDPs are often enriched in charged (often negatively charged), alanine and structure breaking residues, like proline and glycine. These residues are called disorder promoting amino acids. In contrast, hydrophobic and aromatic residues, in addition to cysteine and asparagine, are rarely found in IDPs. This set of amino acids is called order promoting residues ²²⁷ (**Fig. 14a**). The lack of hydrophobic residues in IDPs and the high net charge ensure less driving force for protein compaction and more charges repulsion, which can explain why IDPs cannot fold into a well-defined structure. Moreover, IDPs have high conformational entropy and lack inter-residues contacts. They are therefore considered to be highly dynamics and flexible and can be called dancing proteins or protein clouds ²²⁸.

Despite the fact that IDPs are dynamic structures, they interconvert, and they can range from fully extended, random-coil-like, to more compact, pre-molten-globule-like, molten-globule-like, until they become folded (**Fig.14b**). The hydrodynamic volume increases with the unfolded state of the protein (**Fig.14b**). Interestingly, IDPs are characterized by weak but specific binding, which explains the fact that they have an adaptability in binding; they can bind different proteins where no shape complementarity, in a lock and key manner, is needed.

Now that we know that disordered regions are enriched with disorder promoting residues, the remaining question is where are these residues localized in a protein? Well, it was shown that unfolded domains are non-homogeneously distributed in a protein ²²⁹. But it seems that residues at the protein tail are more disordered than residues in the middle of the protein. However, internal intrinsically disordered regions are important for the function of the protein, as they play the role of linkers or spacers between folded domains. These linkers allow the folded domains to “move” and to adapt different positions ²²⁹. Other internal IDRs are important for interactions, and they can fold upon binding with a ligand (See section 3.1.b for more details).

Finally, IDPs can be recognized by their special biochemical behavior. They bind less sodium dodecyl sulfate (SDS) which can affect their migration in an SDS-PAGE: their apparent molecular weight is 1.2 - 1.8 higher than the actual one ²²⁷. This apparent molecular weight is also observed in size exclusion chromatography. Moreover, susceptibility to proteolysis increases with the flexibility of IDPs. However, they are resistant to high temperatures and acidic pH ²³⁰.

3.1.b. Physiological functions of IDPs

A unique feature of IDPs is that they function either in their native unfolded state, or when they become folded upon their binding to another protein. This inducing folding is called disorder-to-order transition. If the unstructured domain folds into an α -helix or a β -sheet, this domain is then called α -MoRF or β -MoRF, respectively, where MoRF stands for Molecular Recognition Feature ^{231,232} **(Fig.14c)**. These disordered segments promote protein-protein interactions. Given their importance for the protein's function, a new study describes a computational method to predict MoRFs regions in a protein ²³².

Although IDPs lack a 3D fold, they are functional, and they can interact with different types of partners, like proteins and nucleic acids. Therefore, they can perform different functions: signaling, regulation, control, interaction, inhibition, activation, protection and stabilization. For instance, in yeast most of the IDPs are in the cell nucleus and play a role in the regulation of transcription and cell signaling ²²³. They can also play the role of chaperones or dynamic lids ^{229,233,234}. Moreover, IDPs are known to facilitate many biological processes like the alternative splicing and the movement of the protein through a narrow pore or channel like the proteasome ^{235,236}.

On the other hand, folded proteins are often related to catalysis such as enzymes, translocation, binding and release of small molecules, etc. Although they are folded, some enzymes contain some flexible disordered regions and loops that facilitate the function of the enzyme ²³⁷. Interestingly, function of some folded proteins requires local or even global unfolding **(Fig.14c)**. The transition from disorder to order, or vice versa, can be passive or active **(Fig.14c)**. It is passive when it is induced by global environmental factors like the pH, temperature, redox, light and mechanical force. However, the transition is active if it is induced by the interaction with membranes,

ligand binding, protein-protein interactions, interaction with nucleic acids, post-translational modifications or release of auto inhibition ²³⁸.

Finally, one more characteristic of unfolded domains is that they can adopt different folding structures depending on their interaction partner. Therefore, the same protein might have many interacting partners. This plasticity seems to be essential on the energetic level for the cell. In fact, it is less energy consuming for the cell to use one IDP to interact with different partners, instead of producing different folded proteins, each to interact with one partner. Some examples of IDPs that can bind to around 10-100 partners are the cell cycle regulators proteins P21, P53 and the estrogen receptor²³⁹.

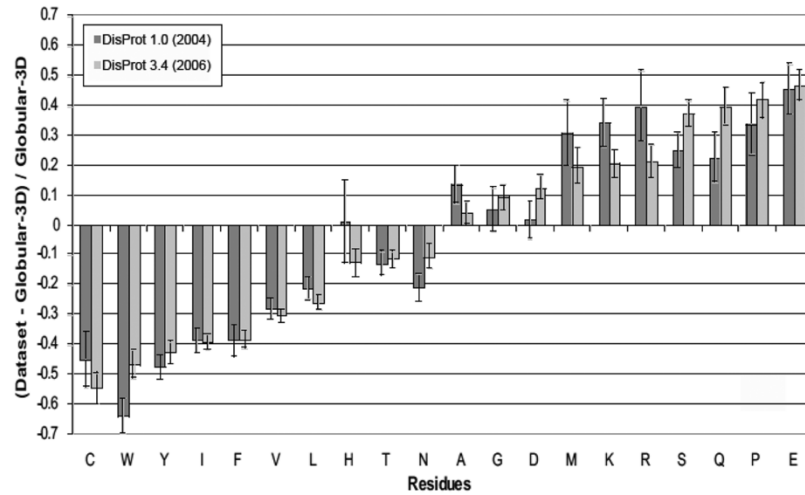
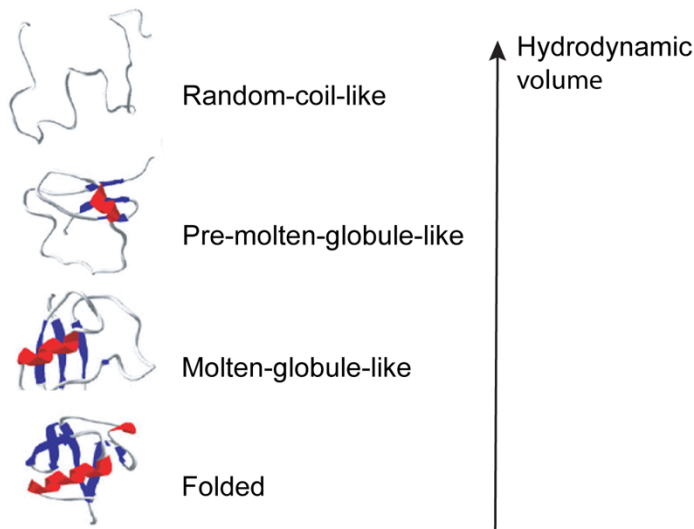
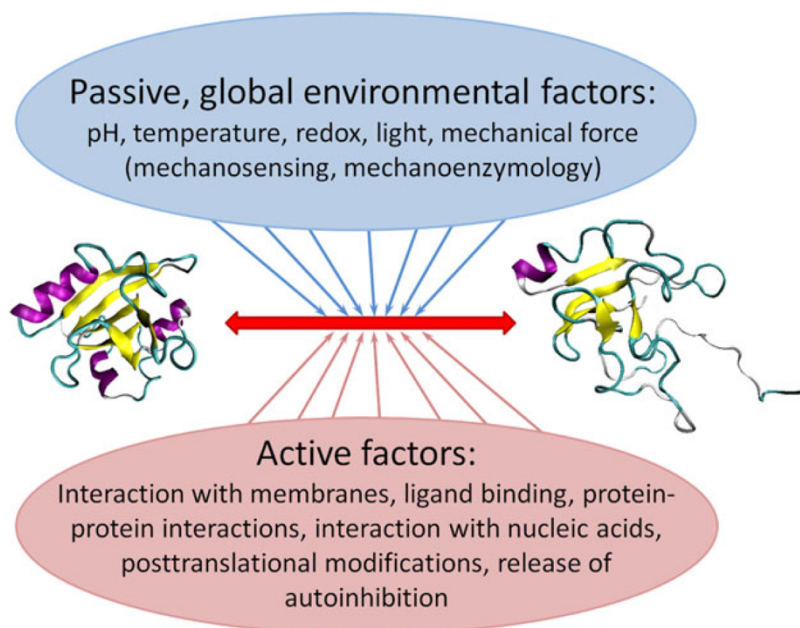
a**b****c**

Figure 14. Representation of the major characteristics of IDPs. **a.** Amino acid composition of IDPs. Disprot 1.0 and Disprot 3.4 are disorder data sets, whereas globular-3D is an order data set. IDPs are rich in disorder-promoting residues and poor in order-promoting-residues. **b.** Conformation changes of IDPs. The hydrodynamic volume of the protein increases with the flexibility degree. **c.** Factors underlying the transition of order-disorder and vice-versa. The transition can be passively induced by environmental factors, or actively induced by interaction with another partner. The images in a and b are adapted from ²²⁷ and in c from ²²⁹.

Moreover, since cell cycle regulators proteins, like P53, are part of the IDPs family, it's not surprising at all that IDPs proteins are involved in cancer ²⁴⁰⁻²⁴². Interestingly, a study showed that IDPs are more frequent in regulatory, cancer-associated and cytoskeletal proteins than in proteins involved in metabolism and biosynthesis ²⁴⁰.

3.1.c. Informative structural studies of IDPs

It is always important to reveal the structure of a protein since it gives more insight into its function, important regions and characteristics. While the crystallography has been the model technique to study a protein's structure, this method has limits when it comes to IDPs: flexible regions are invisible, giving rise to regions with missing electron density ²⁴³. Other techniques have then been developed and used to study IDPs ^{231,243}. Below we will cite some of these technics.

Structural disorder can be studied by Nuclear Magnetic Resonance (NMR) combined with another structural techniques such as the small angle X-ray scattering (SAXS) and molecular dynamics (MD) simulations ^{243,244}. However, NMR is one of the most powerful tools in providing quantitative, not only qualitative, description of IDPs. Moreover, IDPs can also be studied at the single molecule level by single-molecule

fluorescence resonance energy transfer (smFRET). Conformational changes can be visualized using atomic force microscopy (AFM), which enables us to study the energetics and dynamics of IDPs ²⁴⁵. Finally, circular dichroism (CD) is also another spectroscopic technique used to study IDPs.

Given the fact that during my thesis I was focused on *E. coli*, the next section will be dedicated to discussing IDPs in bacteria.

3.2. IDPs in bacteria

IDPs content in bacterial species varies between 18% and 28% ²²¹. Structural disorder exists in different protein types. First, it is present in proteins that can bind to DNA such as the *E. coli* single-stranded DNA (ssDNA) binding proteins (SSBs) ²⁴⁶. SSBs contain a hypervariable long disordered linker that governs the binding of the protein to the ssDNA ²⁴⁶. Second, IDPs are also present in the divisome; a family of proteins implicated in cell division ²⁴⁷. For instance, FtsZ, which scaffolds and recruit other divisome proteins, contains a C-terminal disordered linker that connects its two conserved globular domains ²⁴⁸⁻²⁵⁰. This linker is essential for the function of FtsZ as the deletion of this disorder region impairs cell division ²⁵¹. Another example is the major player in the differentiation of *Caulobacter crescentus*, the polar organizing protein (Pop)Z, which is 57% disordered ²⁵². PopZ self-assembles and is able therefore to bind to, at least, 8 different proteins ²⁵².

IDPs also exist in the bacterial “weapons” like the bacteriocins. Bacteriocins produced by bacteria are antimicrobial peptides or proteins used as toxins against a narrow spectrum of bacterial species ²⁵³. Colicins are bacteriocins specific for *E. coli* that use the OmpF porin to enter the cell ²⁵⁴. Colicin E9 has a long N-terminal disordered linker that is essential for the entry of the toxin in the cell through OmpF ²⁵⁵.

The list of bacterial IDPs goes on. The biofilm of *Bacillus subtilis* contains amyloid-like fibers whose association is mediated by the protein TapA ²⁵⁶. NMR and CD studies showed that TapA has a disordered C-terminal domain that weakly interacts with its N-terminal structured domain, and that these domains gain α -helical structures with increasing temperatures ²⁵⁷. Also, many curli proteins were shown to be disordered like CsgF, CsgE ²⁵⁸ and CsgG ¹⁵⁰. Curli are proteinaceous filaments present at the cell surface of *Enterobacteriaceae*. They facilitate the biofilm formation, cell adhesion and invasion and environmental persistence ²⁵⁹. CsgF and CsgE act as chaperone for the attachment and the secretion of the curli, while CsgG is the curli pore protein ^{260,261}. The N-terminal unfolded domain of CsgF was shown to adopt α -helical structure upon binding to CsgG, while its highly flexible C-terminal domain is responsible for the interaction with CsgB, the minor subunit of curli filaments ²⁶². Moreover, disorder seems to be important for the functionality of type III secretion system (T3SS) ²⁶³. T3SS acts as an injectosome to secrete effectors to the extracellular environment or directly into the host cell. HrpO, a protein of the T3SS of *P. syringae*, is highly disordered with coiled-coil and α -helical segments. It interacts via coiled-coil interactions with HrpE, another component of the T3SS, and this interaction is accompanied with α -helical ordering ²⁶³.

Finally, a few lipoproteins were shown to have a long N-terminal disordered flexible linker like RcsF, Pal and BamC ^{177,178,202,264-266}. The presence and function of this disordered domain in the lipoproteins of *E. coli* were the focus of my work during the thesis and will be detailed in the Results section.

In conclusion, structural disorder is widespread in many organisms: unicellular and pluricellular, prokaryote and eukaryote. The discovery of structural disorder and its conservation in different phyla grabbed the attention of the experts in the field for the last few years. These latest studies led to the discovery of more insights regarding the characteristics and the functions of structural disorder. However, the importance of disorder in many proteins remains to be elucidated. The functional role of the disordered region present at the N-terminal of many lipoproteins of *E. coli* will be discussed in the Results section of this thesis.

Section II. Results

In this section, I will present the results obtained during my thesis. As previously explained in the research objectives at the beginning of the manuscript, **the first goal** of my thesis was to elucidate the function of the linker, the intrinsically disordered region present at the N-terminus of half of OM lipoproteins. During my thesis, I was mainly interested in finding whether the N-terminal linker is important in the journey of lipoproteins in the cell envelope of *E. coli* or not. First, I wanted to evaluate the presence of the linker in the different lipoproteins of *E. coli*. Second, I wanted to define how and where does the linker affect the sorting of lipoproteins in the cell envelope. The results obtained were published in Nature Chemical Biology and will be presented in **Chapter I**. I would like to mention that Dr. Joanna Szewczyk, a previous member of the Collet lab, had already started working on the linker and obtained exciting results regarding its importance in the OM targeting of RcsF. When I joined the lab, I took the project over, confirmed key results previously obtained by Joanna, widened the search to additional *E. coli* lipoproteins including NlpD, BamC and Pal, and decided to also investigate the mechanistic importance of the linker by doing pull downs, protein purifications and looking at each step of the maturation and transport pathways. **The second goal** of my thesis was to obtain a pure, stable complex between the periplasmic chaperone LolA and a substrate lipoprotein, with the objective to crystalize it and solve its structure in order to gain further structural and molecular insights into this reaction. Preliminary results are presented in **Chapter II**.

Chapter I. Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria

The results presented in this chapter are published in Nature Chemical Biology. The paper can be found in Annex I.

Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria

Jessica El Rayes^{1,2§}, Joanna Szewczyk^{1,2§}, Michael Deghelt^{1,2}, Naemi Csoma^{1,2}, André Matagne³, Bogdan I. Iorga⁴, Seung-Hyun Cho^{1,2}, and Jean-François Collet^{1,2*}

¹WELBIO, Avenue Hippocrate 75, 1200 Brussels, Belgium.

²de Duve Institute, Université catholique de Louvain, Avenue Hippocrate 75, 1200 Brussels, Belgium.

³Centre d'ingénierie des Protéines, Institut de Chimie B6, Université de Liège, Allée de la Chimie 3, 4000 Liège, Sart Tilman, Belgium.

⁴Université Paris-Saclay, CNRS UPR 2301, Institut de Chimie des Substances Naturelles, 91198 Gif-sur-Yvette, France.

[§]Both authors contributed equally to the work

*Correspondence: jfcollet@uclouvain.be

Abstract

Gram-negative bacteria express structurally diverse lipoproteins in their envelope. Here we found that approximately half of lipoproteins destined to the *Escherichia coli* outer membrane display an intrinsically disordered linker at their N-terminus. Intrinsically disordered regions are common in proteins, but establishing their importance *in vivo* has remained challenging. As we sought to unravel how lipoproteins mature, we discovered that unstructured linkers are required for optimal trafficking by the Lol lipoprotein sorting system: linker deletion re-routes three unrelated lipoproteins to the inner membrane. Focusing on the stress sensor RcsF, we found that replacing the linker with an artificial peptide restored normal outer membrane targeting only when the peptide was of similar length and disordered. Overall, this study reveals the role played by intrinsic disorder in lipoprotein sorting, providing mechanistic insight into the biogenesis of these proteins and suggesting that evolution can select for intrinsic disorder that supports protein function.

Introduction

The cell envelope is the morphological hallmark of *Escherichia coli* and other Gram-negative bacteria. It is composed of the inner membrane, a classical phospholipid bilayer, as well as the outer membrane, an asymmetric bilayer with phospholipids in the inner leaflet and lipopolysaccharides in the outer leaflet¹. This lipid asymmetry enables the outer membrane to function as a barrier that effectively prevents the diffusion of toxic compounds in the environment into the cell. The inner and outer membranes are separated by the periplasm, a viscous compartment that contains a thin layer of peptidoglycan also known as the cell wall¹. The cell envelope is essential for growth and survival, as illustrated by the fact that several antibiotics such as the β -lactams target mechanisms of envelope assembly. Mechanisms involved in envelope biogenesis and maintenance are therefore attractive targets for novel antibacterial strategies.

Approximately one-third of *E. coli* proteins are targeted to the envelope, either as soluble proteins present in the periplasm or as proteins inserted in one of the two membranes². While inner membrane proteins cross the lipid bilayer via one or more hydrophobic α -helices, proteins inserted in the outer membrane generally adopt a β -barrel conformation³. Another important group of envelope proteins is the lipoproteins, which are globular proteins anchored to one of the two membranes by a lipid moiety⁴. Lipoproteins carry out a variety of important functions in the cell envelope: they participate in the biogenesis of the outer membrane by inserting lipopolysaccharide molecules^{5,6} and β -barrel proteins⁷, they function as stress sensors triggering signal transduction cascades when envelope integrity is altered⁸, and they control processes that are important for virulence⁹. The diverse roles played by lipoproteins in the cell envelope has drawn a lot of attention lately, revealing how crucial these proteins are in a wide range of vital processes and identifying them as attractive targets for antibiotic development.

Yet, a detailed understanding of the mechanisms involved in lipoprotein maturation and trafficking is still missing.

Lipoproteins are synthesized in the cytoplasm as precursors with an N-terminal signal peptide¹⁰. The last four C-terminal residues of this signal peptide, known as the lipobox, function as a molecular determinant of lipid modification unique to bacteria; only the cysteine at the last position of the lipobox is strictly conserved¹¹. After translocation across the inner membrane, processing of pre-prolipoproteins into mature forms takes place on the periplasmic side of the membrane. The thiol side-chain of the cysteine is first modified with a diacylglyceryl moiety by prolipoprotein diacylglyceryl transferase (Lgt)¹⁰ (**Supplementary Figure 1a**, step 1). Then, signal peptidase II (LspA) catalyzes cleavage of the signal peptide N-terminally of the lipidated cysteine before apolipoprotein N-acyltransferase (Lnt) adds a third acyl group to the N-terminal amino group of the cysteine (**Supplementary Figure 1a**, steps 2-3). Most mature lipoproteins are then transported to the outer membrane by the Lol system. Lol consists of LolCDE, an ABC transporter that extracts lipoproteins from the inner membrane and transfers them to the soluble periplasmic chaperone LolA (**Supplementary Figure 1a**, steps 4-5)¹². LolA escorts lipoproteins across the periplasm, binding their hydrophobic lipid tail, and delivers them to the outer membrane lipoprotein LolB (**Supplementary Figure 1a**, step 6). LolB finally anchors lipoproteins to the inner leaflet of the outer membrane using a mechanism that remains poorly characterized (**Supplementary Figure 1a**, step 7).

In most Gram-negative bacteria, a few lipoproteins remain in the inner membrane¹³. The current view is that inner membrane retention depends on the identity of the two residues located immediately downstream of the N-terminal cysteine on which the lipid moiety is

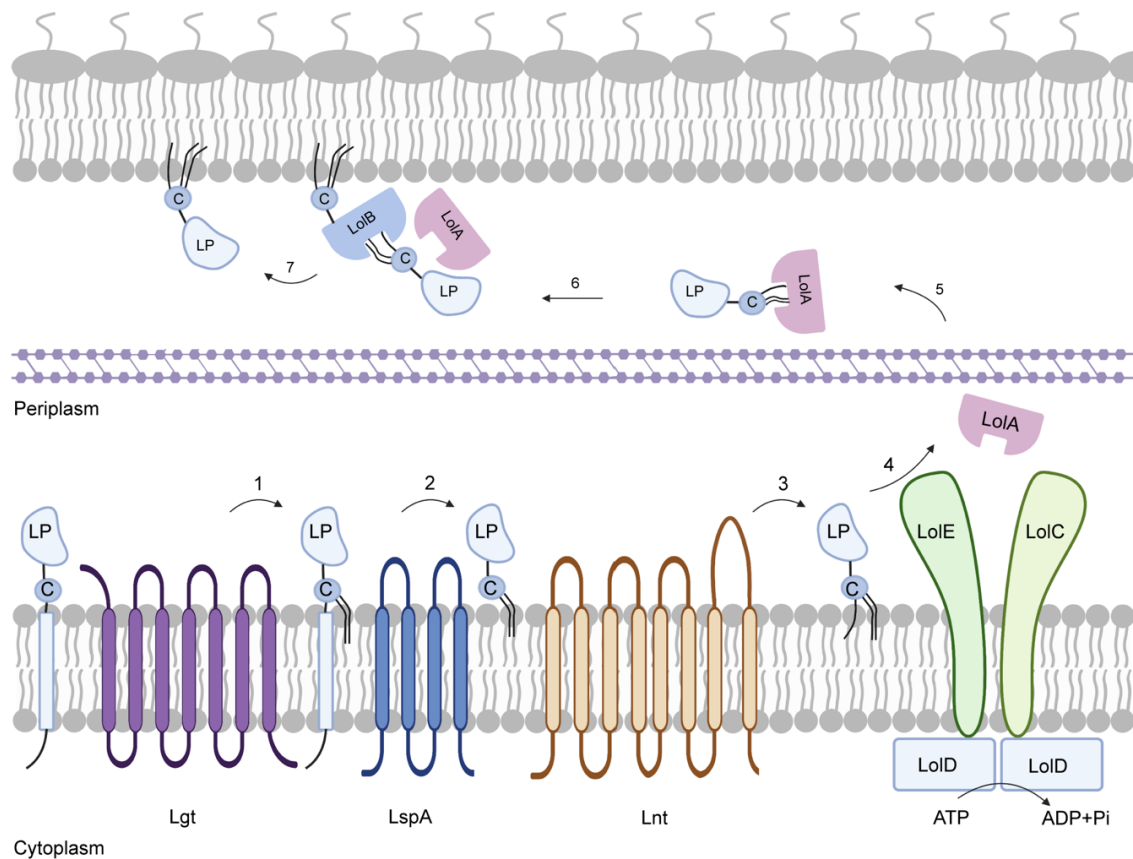
attached¹⁴; this sequence, two amino acids in length, is known as the Lol sorting signal. When lipoproteins have an aspartate at position +2 and an aspartate, glutamate, or glutamine at position +3, they remain in the inner membrane^{15,16}, possibly because strong electrostatic interactions between the +2 aspartate and membrane phospholipids prevent their interaction with LolCDE¹⁷. However, this model is largely based on data obtained in *E. coli* and variations have been described in other bacteria. For instance, in the pathogen *Pseudomonas aeruginosa*, an aspartate is rarely found at position +2 and inner membrane retention appears to be determined by residues +3 and +4^{18,19}. Surprisingly, lipoproteins are well sorted in *P. aeruginosa* cells expressing the *E. coli* LolCDE complex²⁰, despite their different Lol sorting signal. This result cannot be explained by the current model of lipoprotein sorting, underscoring that our comprehension of the precise mechanism that governs the triage of lipoproteins remains incomplete.

Excitingly, more unresolved questions regarding lipoprotein biogenesis have recently been raised. First, it was reported that a LolA-LolB-independent trafficking route to the outer membrane exists in *E. coli*²¹, but the factors involved have remained unknown. Second, although lipoproteins have traditionally been considered to be exposed to the periplasm in *E. coli* and many other bacterial models¹⁰, a series of investigations have started to challenge this view by identifying lipoproteins on the surface of *E. coli*, *Vibrio cholerae*, and *Salmonella Typhimurium*²²⁻²⁴. Overall, the field is beginning to explore a lipoprotein topological landscape that is more complex than previously assumed and raising intriguing questions about the signals that control surface targeting and exposure.

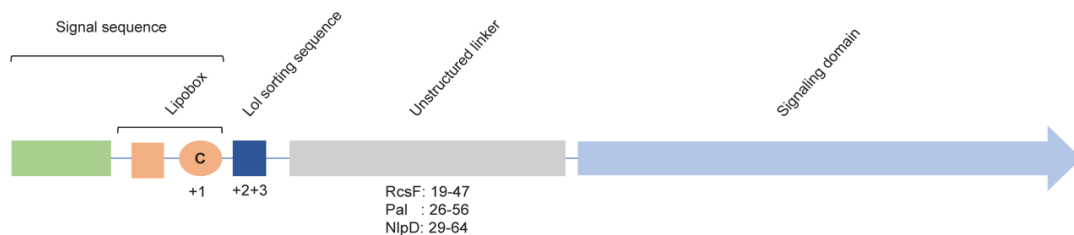
Here, stimulated by the hypothesis that crucial details of the mechanisms underlying lipoprotein maturation remained to be elucidated, we sought to identify novel molecular

determinants controlling lipoprotein biogenesis. First, we found that half of the outer membrane lipoproteins display a long and intrinsically disordered linker at their N-terminus. Intrigued by these unstructured segments, we probed their importance for the biogenesis of RcsF, NlpD, and Pal, three unrelated outer membrane lipoproteins. Unexpectedly, we found that deleting the linker altered the targeting of all three lipoproteins to the outer membrane. Focusing on RcsF, we determined that both the length and disordered character of the linker were important; we also found that the linker is required for optimal processing by Lol. Our observations reveal the unsuspected role played by protein intrinsic disorder in lipoprotein cellular localization. We also propose a model in which disordered linkers are important for targeting and/or displaying lipoproteins on the cell surface.

a



b



Supplementary Figure 1. Lipoprotein maturation and sorting in the *E. coli* cell envelope

a. After processing by Lgt (step 1), LspA (step 2), and Lnt (step 3), a new lipoprotein either remains in the inner membrane or is extracted by the LolCDE complex (step 4), depending on the residues at position +2 and +3. LolCDE transfers the lipoprotein to the periplasmic chaperone LolA (step 5), which delivers the lipoprotein to LolB (step 6). LolB, a lipoprotein itself, inserts the lipoprotein in the outer membrane using a poorly understood mechanism (step 7). **b.** Schematic of lipoprotein structural domains. The N-terminal signal sequence targets the lipoprotein to the cell envelope; the last four amino acid residues of the signal sequence form

the lipobox. The last residue of the lipobox is the invariant cysteine that undergoes lipidation. This cysteine, which is the first residue of the mature lipoprotein, is directly followed by the sorting signal, a sequence of 2 or 3 amino acids that controls the sorting of mature lipoproteins between the inner and outer membranes. The C-terminal portion of a mature lipoprotein is a globular domain. An intrinsically disordered linker separates the sorting signal from the globular domain in about half of *E. coli* lipoproteins (**Fig. 1; Extended Data Fig. 1; Supplementary Table 1**). The lengths of the deleted disordered linkers of the unrelated lipoproteins RcsF, Pal, and NlpD are indicated. LP, lipoprotein.

Results

Half of lipoproteins present N-terminal disordered segments

In an attempt to discover novel molecular determinants controlling the cellular localization of lipoproteins, we decided to systematically analyze the sequence of the lipoproteins encoded by the *E. coli* genome (strain MG1655) in search of unidentified structural features. *E. coli* encodes ~80 validated lipoproteins²⁵, of which 58 have been experimentally shown to localize in the outer membrane²⁶. Comparative modeling of existing X-ray, cryogenic electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) structures revealed that approximately half of these outer membrane lipoproteins display a long segment (>22 residues) that is predicted to be disordered at the N-terminus (**Fig. 1, Extended Data Fig. 1, Supplementary Table 1**). In contrast, only one of the 8 lipoproteins that remain in the inner membrane (DcrB; **Extended Data Fig. 1., Supplementary Table 1**) had a long, disordered linker, suggesting that disordered peptides may be important for lipoprotein sorting.

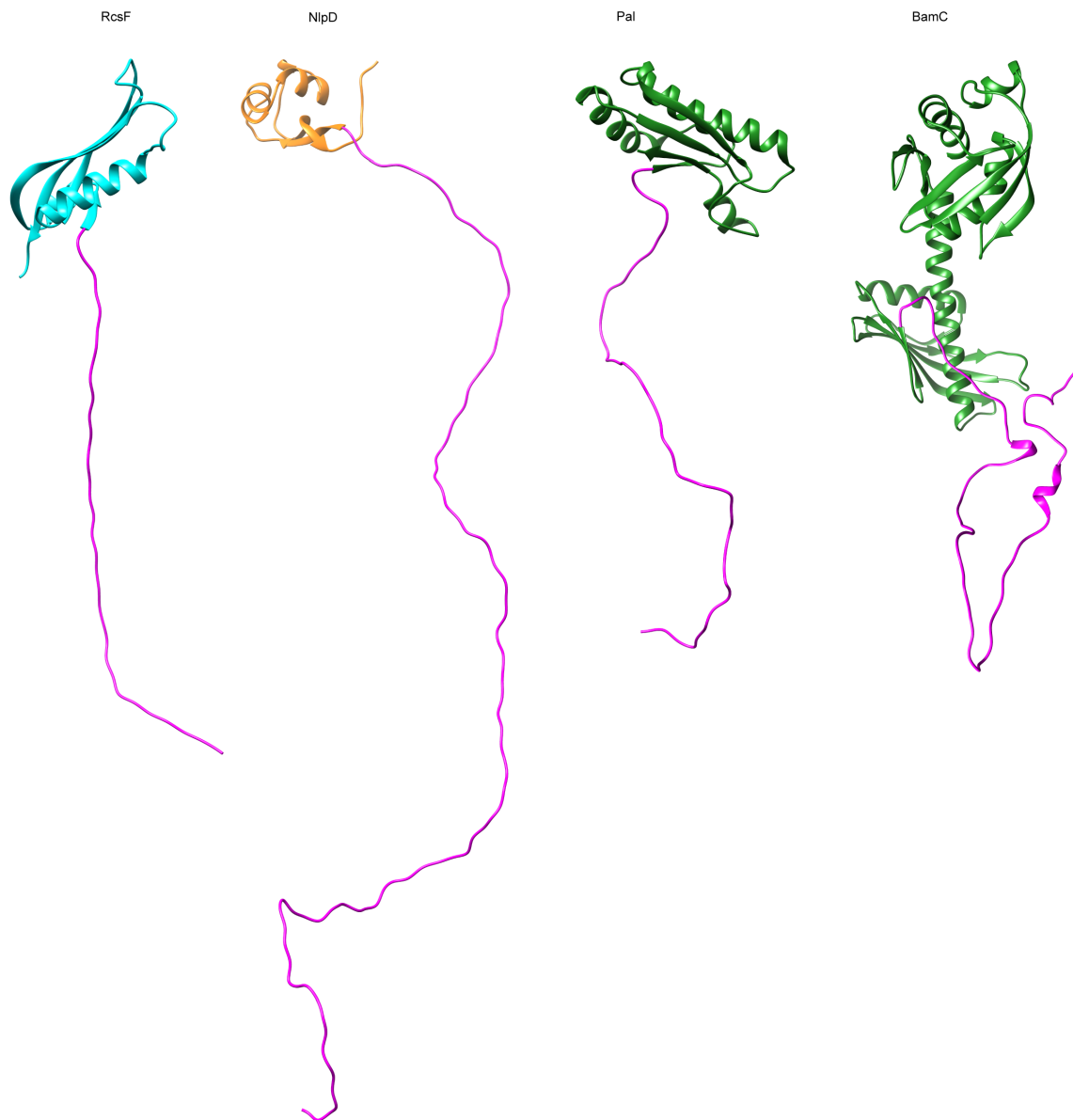
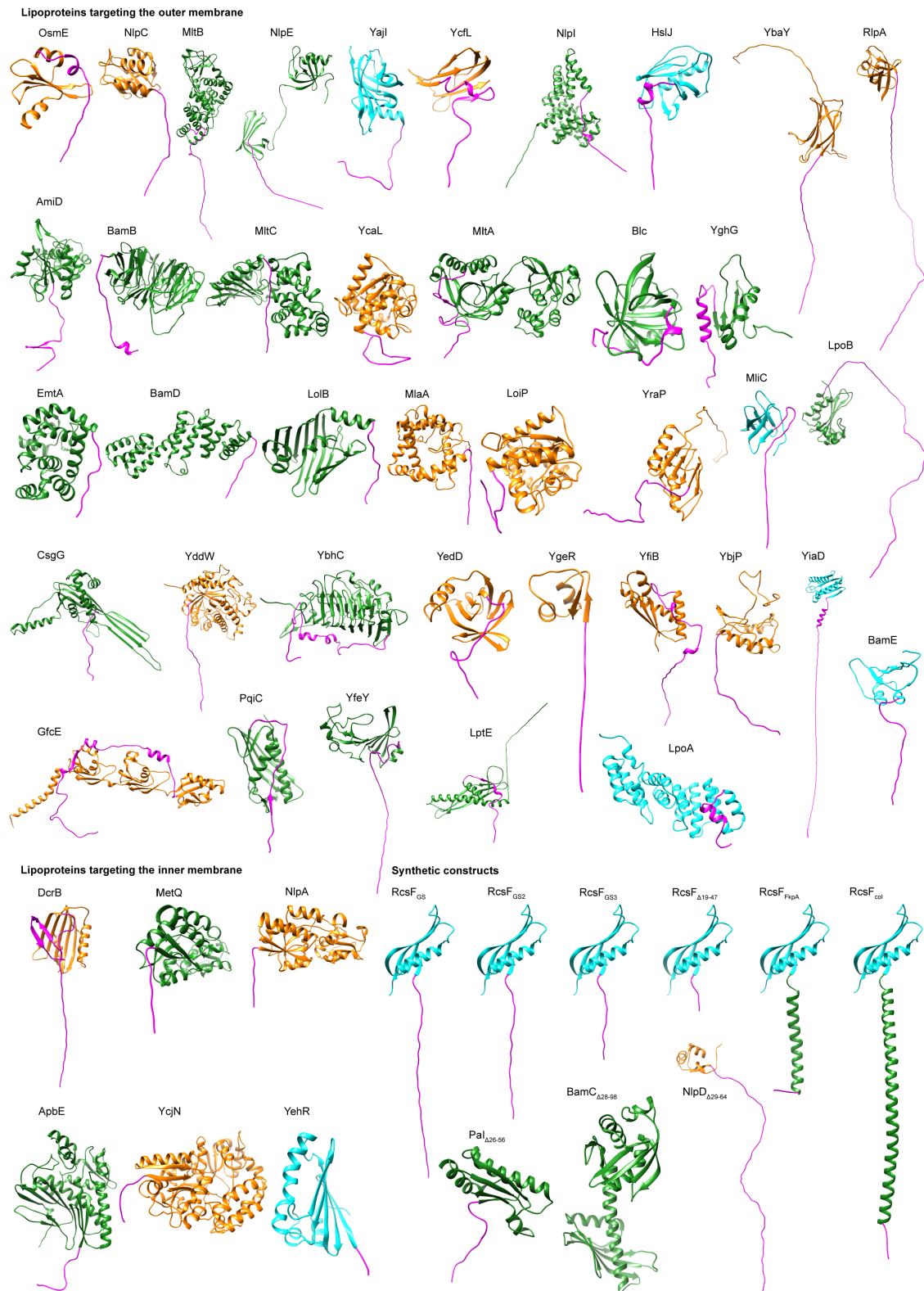


Figure 1. Structural analysis of lipoproteins reveals that half of outer membrane lipoproteins display an intrinsically disordered linker at the N-terminus. The structures of the outer membrane lipoproteins RcsF, NlpD, Pal, and BamC were generated via comparative modeling (Methods). The 39 remaining lipoproteins are shown in **Extended Data Figure 1**. X-ray and cryo-EM structures are green, NMR structures are cyan, and structures built *via* comparative modeling from the closest analog in the same PFAM group are orange. In all cases, the N-terminal linker is magenta.



Extended Data Figure 1. Structural analysis of lipoproteins reveals that half of outer membrane lipoproteins display an intrinsically disordered linker at the N-terminus. Structures were generated via comparative modeling. X-ray and cryo-EM structures are green, NMR structures are cyan, and structures built via comparative modeling from the closest analog in the same PFAM group are orange. In all cases, the N-terminal linker is magenta. Lipoproteins targeting the outer membrane: OsmE, NlpC, MltB, NlpE, YajI, YcfL, NlpI, HslJ, YbaY, RlpA, AmiD, BamB, MltC, YcaL, MltA, Blc, YghG, EmtA, BamD, LolB, MlaA, LoiP, YraP, MliC, LpoB, CsgG, YddW, YbhC, YedD, YgeR, YfiB, YbjP, YiaD, GfcE, PqiC, YfeY, LptE, LpoA, BamE. Lipoproteins targeting the IM: DcrB, MetQ, NlpA, ApbE, YcjN, YehR. Synthetic constructs: RcsF_{GS}, RcsF_{GS2}, RcsF_{GS3}, RcsF_{Δ19-47}, RcsF_{FkpA}, RcsF_{col}, Pal_{Δ26-56}, BamC_{Δ28-98}, NlpD_{Δ29-64}.

Linker deletion perturbs outer membrane targeting

Intrigued by the presence of these N-terminal disordered segments in so many outer membrane lipoproteins, we decided to investigate their functional importance. We selected three structurally unrelated lipoproteins whose function could easily be assessed: the stress sensor RcsF (which triggers the Rcs signaling cascade when damage occurs in the envelope²⁷), NlpD (which activates the periplasmic N-acetylmuramyl-L-alanine amidase AmiC, which is involved in peptidoglycan cleavage during cell division²⁸), and the peptidoglycan-binding lipoprotein Pal (which is important for outer membrane constriction during cell division²⁹).

We began by preparing truncated versions of RcsF, NlpD, and Pal devoid of their N-terminal unstructured linkers (**Supplementary Figure 1b, Extended Data Fig. 1**; RcsF_{Δ19-47}, Pal_{Δ26-56}, and NlpD_{Δ29-64}). Note that the lipidated cysteine residue (+1) and the Lol sorting signal (the amino acids at positions +2 and +3) were not altered in RcsF_{Δ19-47}, Pal_{Δ26-56}, and NlpD_{Δ29-64}, nor in any of the constructs discussed below except in RcsF_{CDD} (**Supplementary Table 2**). For Pal, although the unstructured linker spans residues 25-68 (**Fig. 1**), we used Pal_{Δ26-56} because Pal_{Δ25-68} was either degraded or not detected by the antibody. We first tested whether the

truncated lipoproteins were still correctly extracted from the inner /and transported to the outer membrane. The membrane fraction was prepared from cells expressing the three variants independently, and the outer and inner membranes were separated using sucrose density gradients (Methods). Whereas wild-type RcsF, NlpD, and Pal were mostly detected (>90%) in the outer membrane fraction, ~50% of RcsF $_{\Delta 19-47}$ and ~60% of NlpD $_{\Delta 29-64}$ were retained in the inner membrane (**Fig. 2a, Extended Data Fig. 2a**). An RcsF variant in which the Lol sorting signal was replaced by two aspartates (RcsF_{CDD}) was used as a control for inner membrane retention; as expected, RcsF_{CDD} was mostly detected in the inner membrane (**Extended Data Fig. 3**). The sorting of Pal was also affected, but to a lesser extent: 15% of Pal $_{\Delta 26-56}$ was retained in the inner membrane (**Extended Data Fig. 2b**). Notably, the expression levels of the three linker-less variants were similar (NlpD $_{\Delta 29-64}$) or lower (RcsF $_{\Delta 19-47}$; Pal $_{\Delta 26-56}$) than those of the wild-type proteins (**Extended Data Fig. 4**), indicating that accumulation in the inner membrane did not result from increased protein abundance.

We then tested the impact of linker deletion on the function of these three proteins. In cells expressing RcsF $_{\Delta 19-47}$, the Rcs system was constitutively turned on (**Fig. 2b**); when RcsF accumulates in the inner membrane, it becomes available for interaction with IgaA, its downstream Rcs partner in the inner membrane^{27,30}. Likewise, expression of NlpD $_{\Delta 29-64}$ did not rescue the chaining phenotype (**Fig. 2c**)³¹ exhibited by cells lacking both *nlpD* and *envC*, an activator of the amidases AmiA and AmiB²⁸. Finally, Pal $_{\Delta 26-56}$ partially rescued the sensitivity of the *pal* mutant to SDS-EDTA that results from increased membrane permeability³² (**Fig. 2d**). However, this observation needs to be considered with caution given that Pal $_{\Delta 26-56}$ seemed to be expressed at lower levels than wild-type Pal (**Extended Data Fig. 4**). Thus, preventing normal targeting of RcsF, NlpD and Pal to the outer membrane had functional consequences.

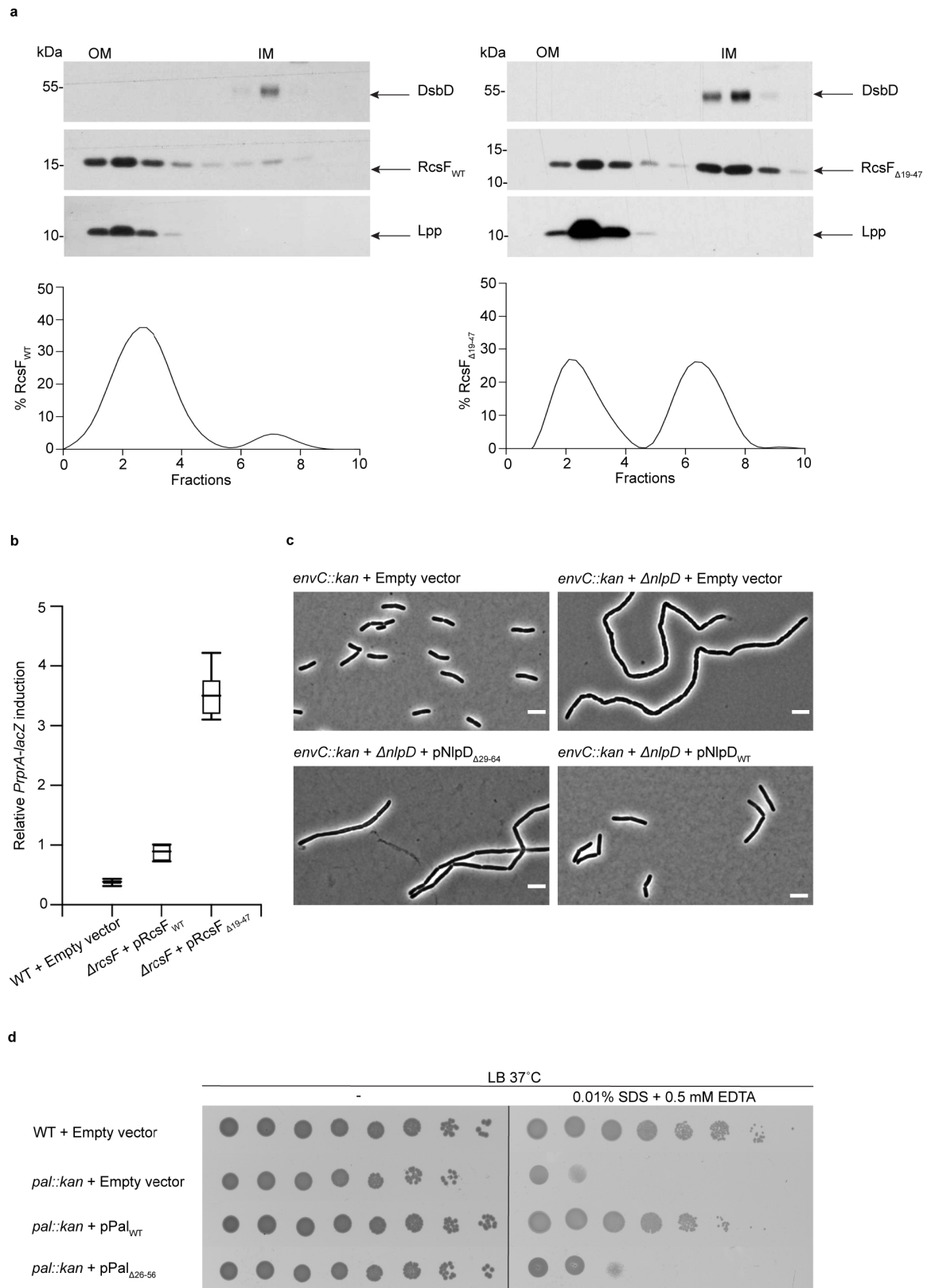
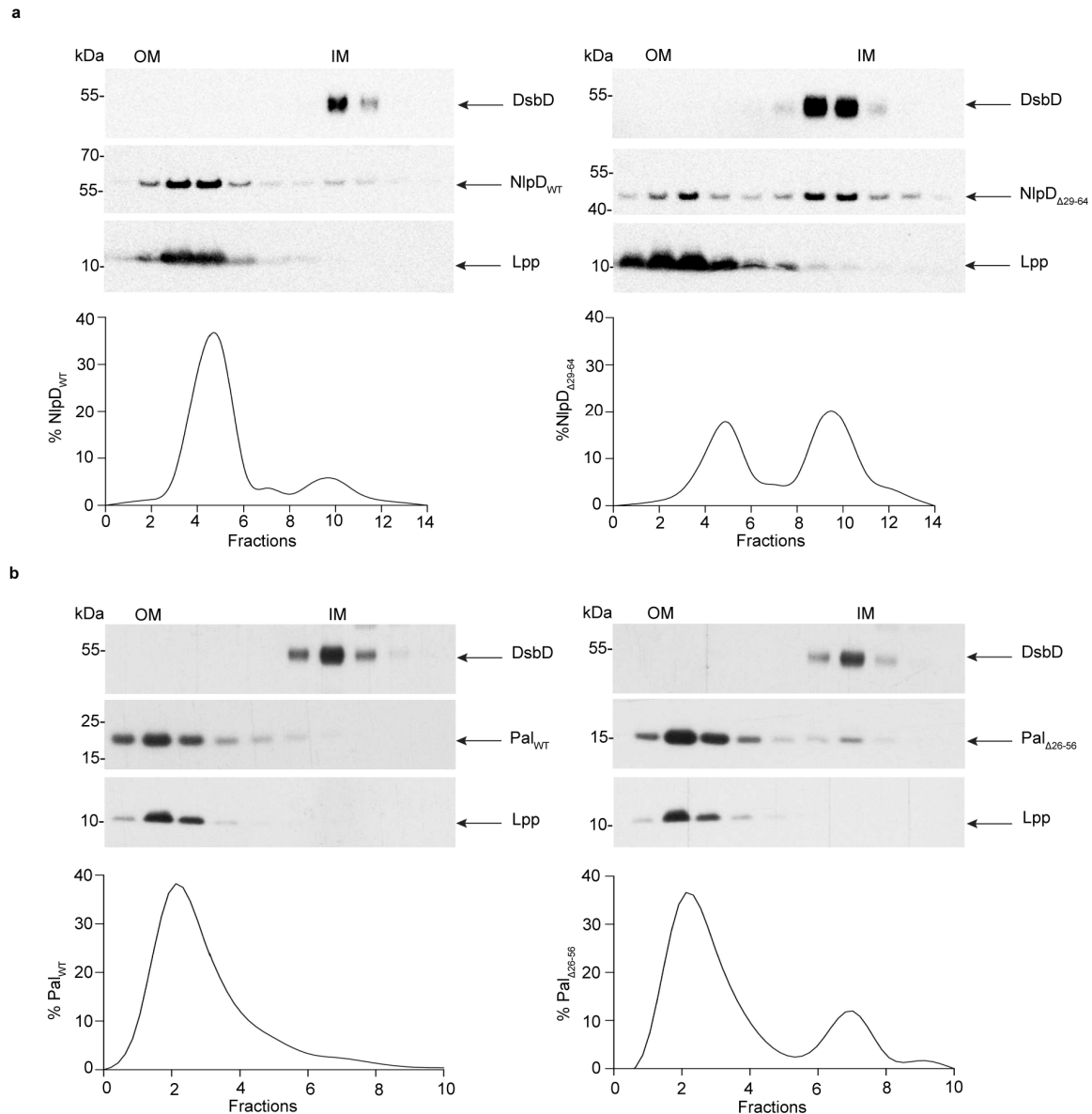


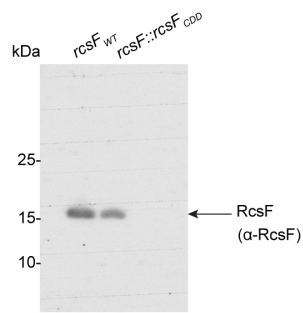
Figure 2. The N-terminal linker displayed by lipoproteins is important for outer membrane targeting. a. The outer membrane (OM) and inner membrane (IM) were separated

via centrifugation in a three-step sucrose density gradient (Methods). While RcsF_{WT} was found predominantly in the OM, RcsF_{Δ19-47} was substantially retained in the IM. Data are presented as the ratio of signal intensity in a single fraction to the total intensity in all fractions. All variants were expressed from plasmids (**Supplementary Table 4**). DsbD and Lpp were used as controls for the OM and IM, respectively. **b.** The Rcs system is constitutively active when the linker of RcsF is missing. Rcs activity was measured with a beta-galactosidase assay in a strain harboring a transcriptional *rprA::lacZ* fusion (Methods). All variants were expressed from a plasmid. Results were normalized to expression levels of the RcsF variants. Box-and-whiskers plots are shown from minimum to maximum with the central line that marks the median; first and third quartile with the whiskers are also shown; n = 6 biologically independent experiments. **c.** Phase-contrast images of the *envC::kan ΔnlpD* mutant complemented with NlpD_{WT} or NlpD_{Δ29-64}. NlpD_{Δ29-64} only partially rescues the chaining phenotype of the *envC::kan ΔnlpD* double mutant. Scale bar, 5 μm. **d.** Expression of Pal_{Δ26-56} does not rescue the sensitivity of the *pal::kan* mutant to SDS-EDTA. Cells were grown in LB medium at 37 °C until mid-log phase (OD₆₀₀ = 0.5). Tenfold serial dilutions were made in LB, plated onto LB agar or LB agar supplemented with 0.01% SDS and 0.5 mM EDTA, and incubated at 37 °C. Images in **a**, **c**, and **d** are representative of biological triplicates. Graph in **a** was created by spline analysis of curves representing a mean of three independent experiments.

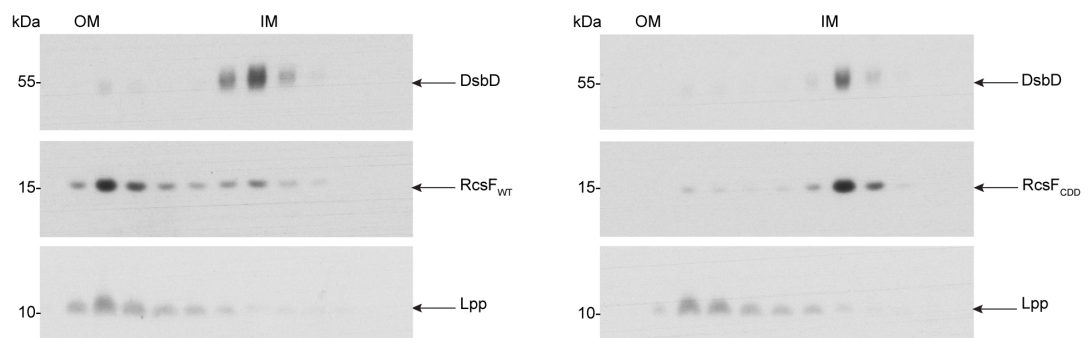


Extended Data Fig. 2. The N-terminal linker of lipoproteins is important for outer membrane targeting. The outer membrane (OM) and inner membrane (IM) were separated via centrifugation in a three-step sucrose density gradient (Methods). While **(a)** NlpD_{WT}, and **(b)** Pal_{WT} were found predominantly in the OM, NlpD_{Δ29-64}, and Pal_{Δ26-56} were substantially retained in the IM. Data are presented as the ratio of signal intensity in a single fraction to the total intensity in all fractions. All variants were expressed from plasmids (**Supplementary Table 4**). DsbD and Lpp were used as controls for the OM and IM, respectively. Images in a, and b are representative of biological triplicates. Graphs in a and b were created by spline analysis of curves representing a mean of three independent experiments.

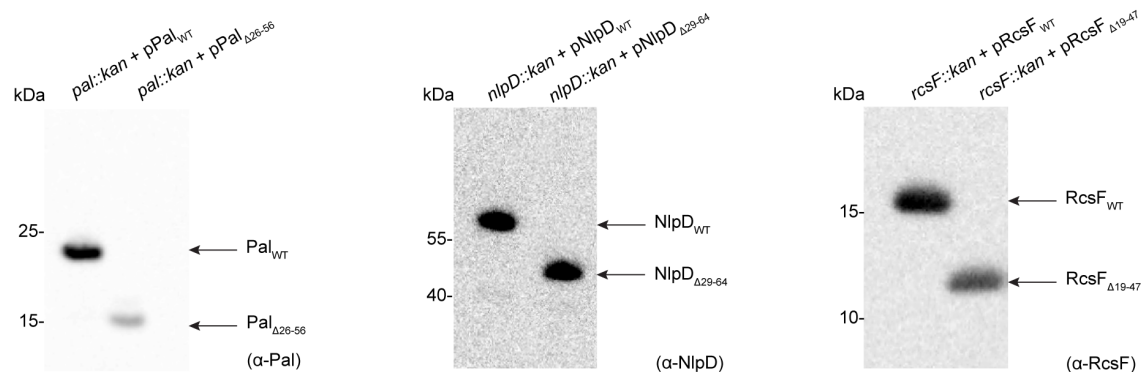
a



b



Extended Data Fig. 3. RcsF_{CDD} is retained in the inner membrane. a. Expression levels of RcsF_{WT} and RcsF_{CDD}. Cells expressing chromosomal RcsF_{WT} or RcsF_{CDD} were grown in LB until OD₆₀₀ = 0.7 and precipitated with trichloroacetic acid (Methods). Immunoblots were performed using α-RcsF. **b.** The outer (OM) and inner (IM) membranes of cells expressing chromosomal RcsF_{WT} or RcsF_{CDD} were separated by sucrose density gradient (Methods). While RcsF_{WT} was found predominantly in the OM, RcsF_{CDD} was retained in the IM. Images in a and b are representative of three independent experiments.



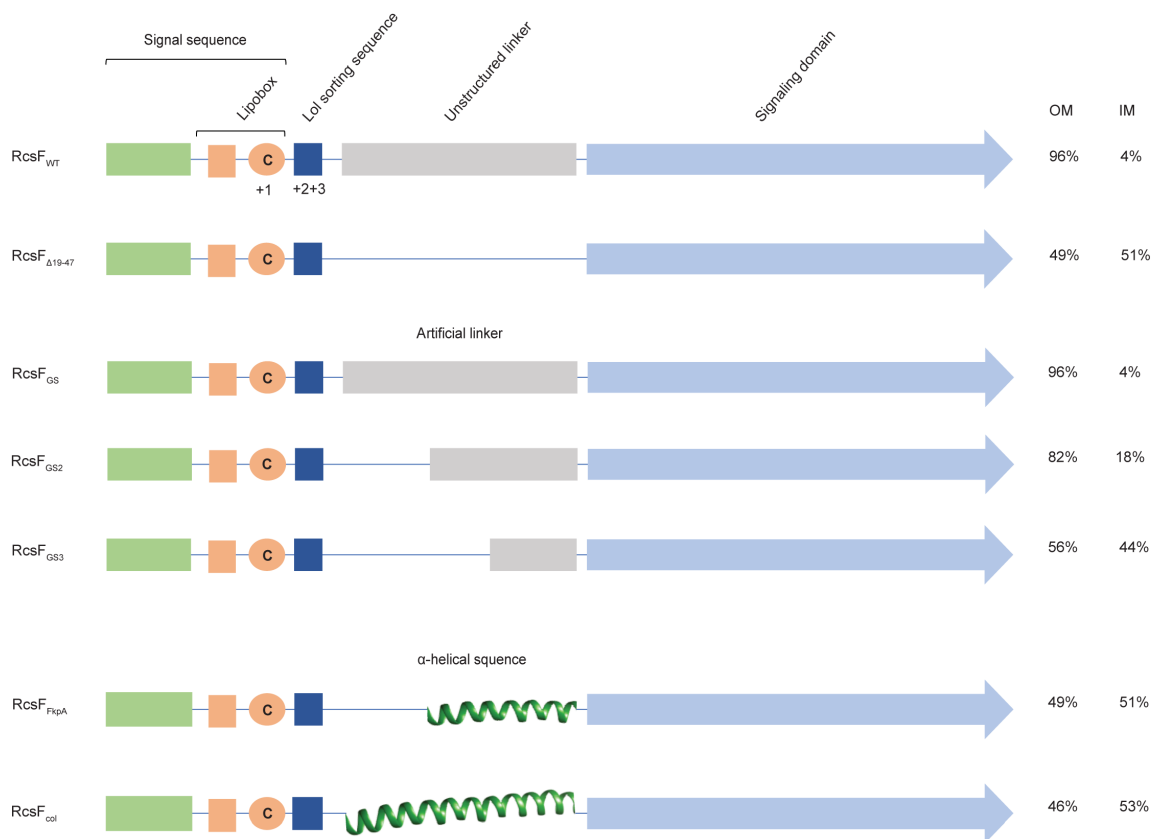
Extended Data Fig. 4. Expression levels of RcsF_{Δ19-47}, Pal_{Δ26-56}, and NlpD_{Δ29-64}. Cells were grown at 37 °C in LB until OD₆₀₀ = 0.5 and precipitated with trichloroacetic acid (Methods). Immunoblots were performed with α-RcsF, α-NlpD, and α-Pal antibodies. All images are representative of three independent experiments.

Unstructured artificial linkers do the job

The results above were surprising because they revealed that the normal targeting of RcsF, NlpD, and Pal to the outer membrane does not only require an appropriate Lol sorting signal, as proposed by the current model for lipoprotein sorting¹⁰, but also the presence of an N-terminal linker. We selected RcsF, whose accumulation in the inner membrane can be easily tracked by monitoring Rcs activity²⁷, to investigate the structural features of the linker controlling lipoprotein maturation; keeping as little as 10% of the total pool of RcsF molecules in the inner membrane is sufficient to fully activate Rcs²⁷.

We first tested whether changing the sequence of the N-terminal segment while preserving its disordered character still yielded normal targeting of the protein to the outer membrane. To that end, we prepared an RcsF variant in which the N-terminal linker was replaced by an artificial, unstructured sequence (**Supplementary Table 2, Extended Data Fig. 1, Extended Data Fig. 5**) of similar length and consisting mostly of Glycine-Serine repeats (RcsF_{GS}).

Substituting the wild-type linker with this artificial sequence was remarkably well tolerated by RcsF: RcsF_{GS} was targeted normally to the outer membrane (Fig. 3a, Extended Data Fig. 1, Supplementary Figure 2a) and did not constitutively activate the stress system (Fig. 3b). Thus, although RcsF_{GS} has an N-terminus with a completely different primary structure, it behaved like the wild-type protein.



Extended Data Fig. 5. Schematic of RcsF variants used in this study and their distributions in the outer membrane (OM) and inner membrane (IM). RcsF_{GS}, RcsF_{GS2}, and RcsF_{GS3} have linkers that are disordered and mostly consist of GS repeats. The linker of RcsF_{GS} is the same length as the linker of RcsF_{WT}. RcsF_{GS2} and RcsF_{GS3} are shorter than RcsF_{WT}. Regions of RcsF_{FkpA} and RcsF_{col} fold into alpha helices borrowed from the sequences of FkpA and colicin Ia, respectively.

We then investigated whether the N-terminal linker required a minimal length for proper targeting and function. We therefore constructed two RcsF variants with shorter, unstructured, artificial linkers (RcsF_{GS2} and RcsF_{GS3}, with linkers of 18 and 10 residues, respectively; **Supplementary Table 2, Extended Data Fig. 1, Extended Data Fig. 5**). Importantly, RcsF_{GS2} and, to a greater extent, RcsF_{GS3} did not properly localize to the outer membrane: the shorter the linker, the more RcsF remained in the inner membrane (**Fig. 3a, Extended Data Fig. 1, Supplementary Figure 2a**). Consistent with the amount of RcsF_{GS2} and RcsF_{GS3} retained in the inner membrane, Rcs activation levels were inversely related to linker length (**Fig. 3b**).

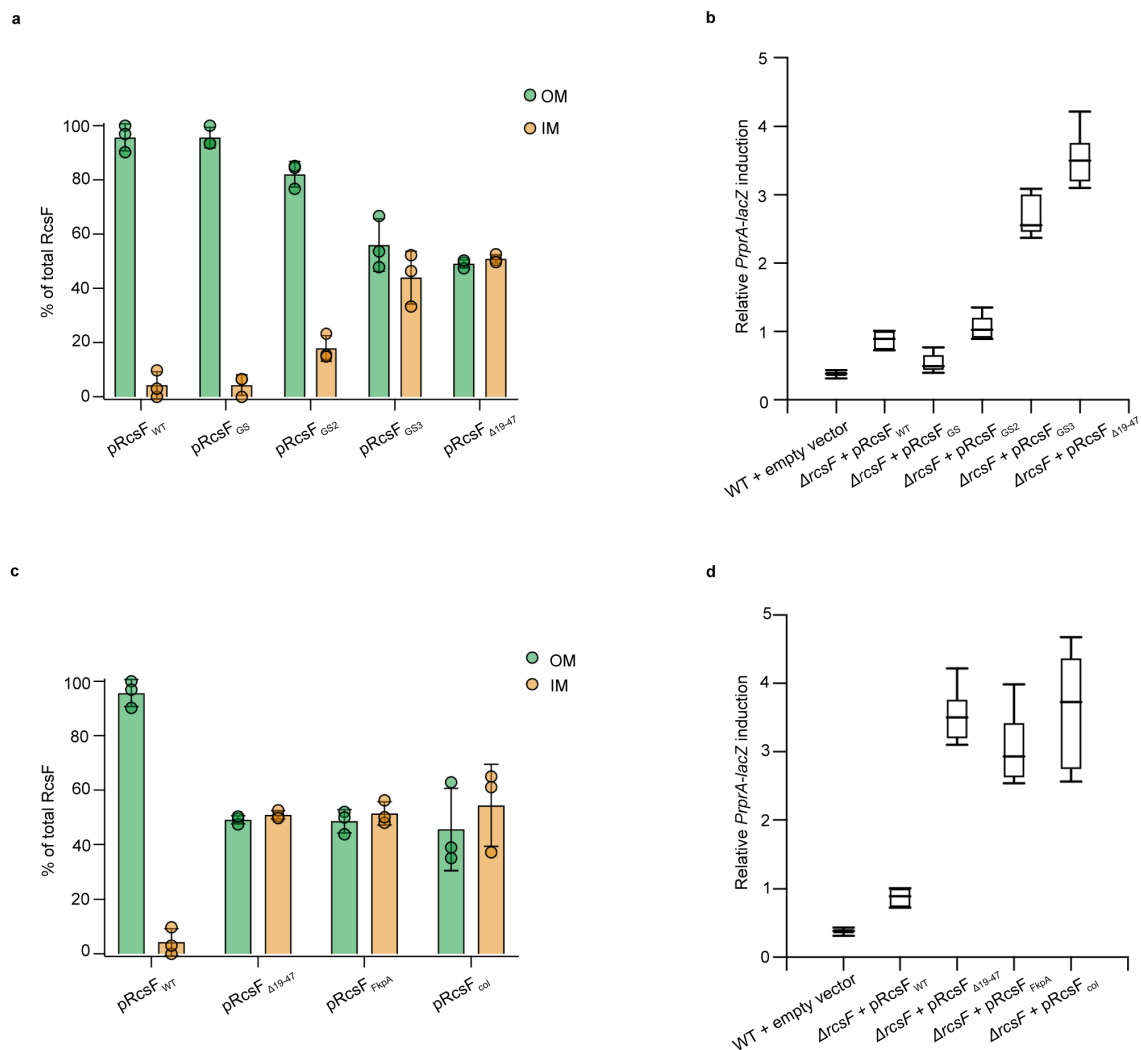
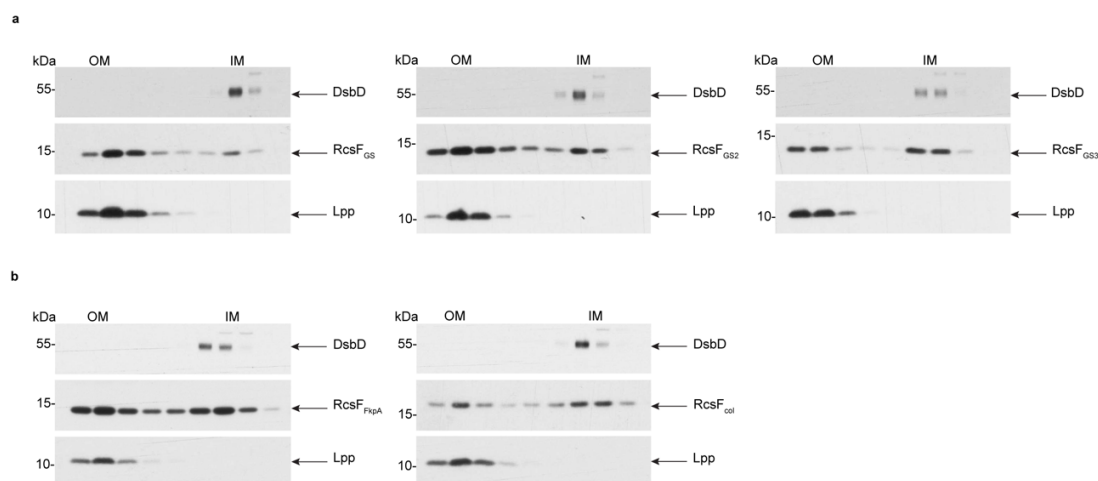


Figure 3. The length and the disordered character of the RcsF linker play key roles in RcsF targeting to the outer membrane. **a.** The outer membrane (OM) and inner membrane (IM) were separated *via* centrifugation in a three-step sucrose density gradient (Methods). DsbD and Lpp were used as controls for the OM and IM, respectively. The longer the linker, the more protein was correctly translocated to the IM. All variants were expressed from plasmids (**Supplementary Table 4**). Aligned dot-plots are shown with a line at mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. Images are representative of experiments and immunoblots performed in biological triplicate. **b.** Rcs activity was measured with a beta-galactosidase assay in a strain harboring a transcriptional *rprA::lacZ* fusion. All variants were expressed from a plasmid. Results were normalized to expression levels of RcsF variants. Box-and-whiskers plots are shown from minimum to maximum with the central line that marks the median; first and third quartile with the whiskers are also shown; $n = 6$ biologically independent experiments. Rcs activity relates to the quantity of RcsF retained in the inner membrane. **c.** RcsF mutants harboring alpha helical linkers (RcsF_{FkpA} and RcsF_{col}) were subjected to two consecutive centrifugations in sucrose density gradients. Both mutants were inefficiently translocated from the IM to the OM. All variants are expressed from plasmids. Aligned dot-plots are shown with a line at mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. Images are representative of experiments and immunoblots performed in biological triplicate. **d.** The Rcs system was constitutively active in RcsF_{FkpA} and RcsF_{col} strains; activation levels were comparable to those of RcsF $_{\Delta 19-47}$. Rcs activity was measured as in **b**. Results were normalized as in **b**. Box-and-whiskers plots are shown from minimum to maximum with the central line that marks the median; first and third quartile with the whiskers are also shown; $n = 6$ biologically independent experiments.



Supplementary Figure 2. Gels of sucrose gradients of Figure 3, showing that the length and the disordered character of the RcsF linker play key roles in RcsF targeting to the outer membrane. a, b. The outer membrane (OM) and inner membrane (IM) were separated via centrifugation in a three-step sucrose density gradient (Methods). DsbD and Lpp were used as controls for the OM and IM, respectively. All variants were expressed from plasmids (**Supplementary Table 4**). (a) The longer the linker, the more protein was correctly translocated to the IM. (b) RcsF mutants harboring alpha helical linkers (RcsF_{FkpA} and RcsF_{col}) were inefficiently translocated from the IM to the OM. Images in a and b are representative of experiments and immunoblots performed in biological triplicate.

The disordered character of the linker is crucial

Taken together, the results above demonstrated that the RcsF linker can be replaced with an artificial sequence lacking secondary structure, provided that it is of appropriate length. Next, we sought to directly probe the importance of having a disordered linker by replacing the RcsF linker with an alpha-helical segment 35 amino acids long from the periplasmic chaperone FkpA (RcsF_{FkpA}; **Supplementary Table 2, Extended Data Fig. 1, Extended Data Fig. 5**). Introducing order at the N-terminus of RcsF dramatically impacted the protein distribution between the two membranes: RcsF_{FkpA} was substantially retained in the inner membrane (**Fig. 3c, Supplementary Figure 2b**) and constitutively activated Rcs (**Fig. 3d**). As alpha-helical

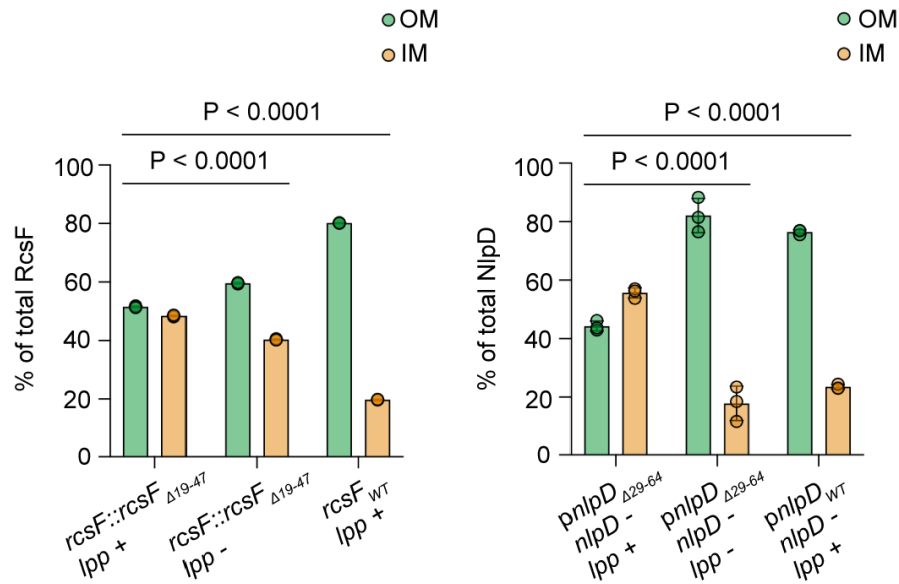
segments are considerably shorter than unstructured sequences containing a similar number of amino acids, we also prepared an RcsF variant (RcsF_{col}) with a longer alpha helix from the helical segment of colicin Ia, which is 73 amino acids in length and also predicted to remain folded in the RcsF_{col} construct (**Supplementary Table 2, Extended Data Fig. 1, Extended Data Fig. 5**). However, doubling the size of the helix had no impact, with RcsF_{col} behaving similarly to RcsF_{FkpA} (**Fig. 3c, 3d, Supplementary Figure 2b**). Together, these data demonstrate that having an N-terminal disordered linker downstream of the Lol sorting signal is required to correctly target RcsF to the outer membrane. The length of the linker is important, but the sequence is not, on the condition that the linker does not fold into a defined secondary structure.

A disordered linker is needed for optimal processing by Lol

Our finding that N-terminal disordered linkers function as molecular determinants of the targeting of lipoproteins to the outer membrane raised the question of whether these linkers work in a Lol-dependent or Lol-independent manner. To address this mechanistic question, we tested the impact of deleting *lpp* on the targeting of RcsF_{Δ19-47}. The lipoprotein Lpp, also known as the Braun lipoprotein, covalently tethers the outer membrane to the peptidoglycan and controls the size of the periplasm^{33,34}. Being expressed at ~1 million copies per cell³⁵, Lpp is numerically the most abundant protein in *E. coli*. Thus, by deleting *lpp*, we considerably decreased the load on the Lol system by removing its most abundant substrate. Remarkably, *lpp* deletion partly rescued the targeting of RcsF_{Δ19-47} to the outer membrane (**Fig. 4a**), indicating that the linker functions in a Lol-dependent manner and suggesting that accumulation of RcsF_{Δ19-47} in the inner membrane results from a decreased ability of the Lol system to process the linker-less RcsF variant. In agreement with this idea, normal targeting of NlpD_{Δ29-64} to the outer membrane was restored in cells lacking Lpp (**Fig. 4a**). Pal_{Δ26-56} could

not be tested because membrane fractionation failed with *lpp pal* double mutant cells whether or not they expressed Pal $_{\Delta 26-56}$ (the *lpp pal* double mutant is viable but highly mucoid).

a



b

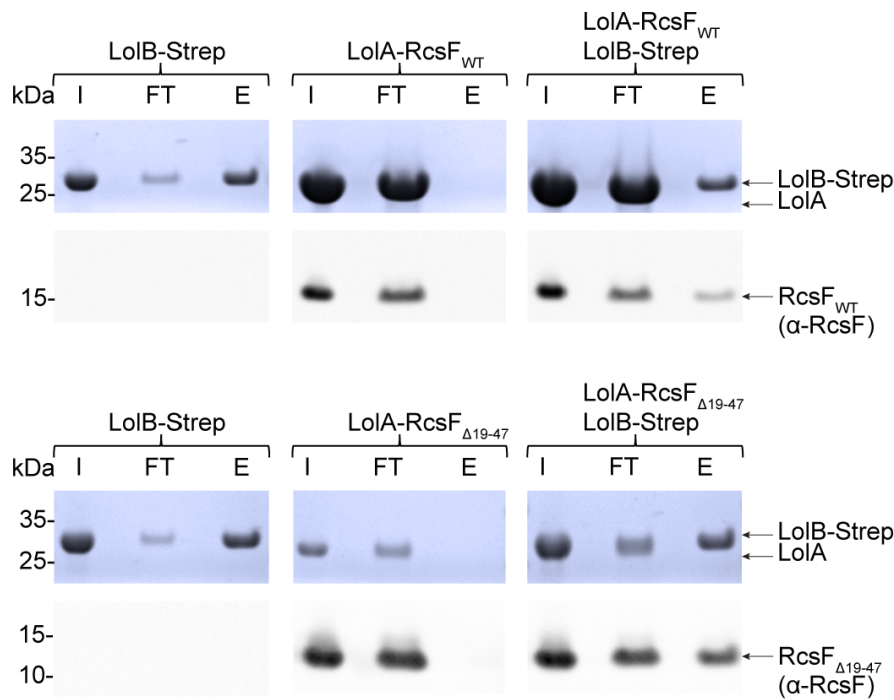
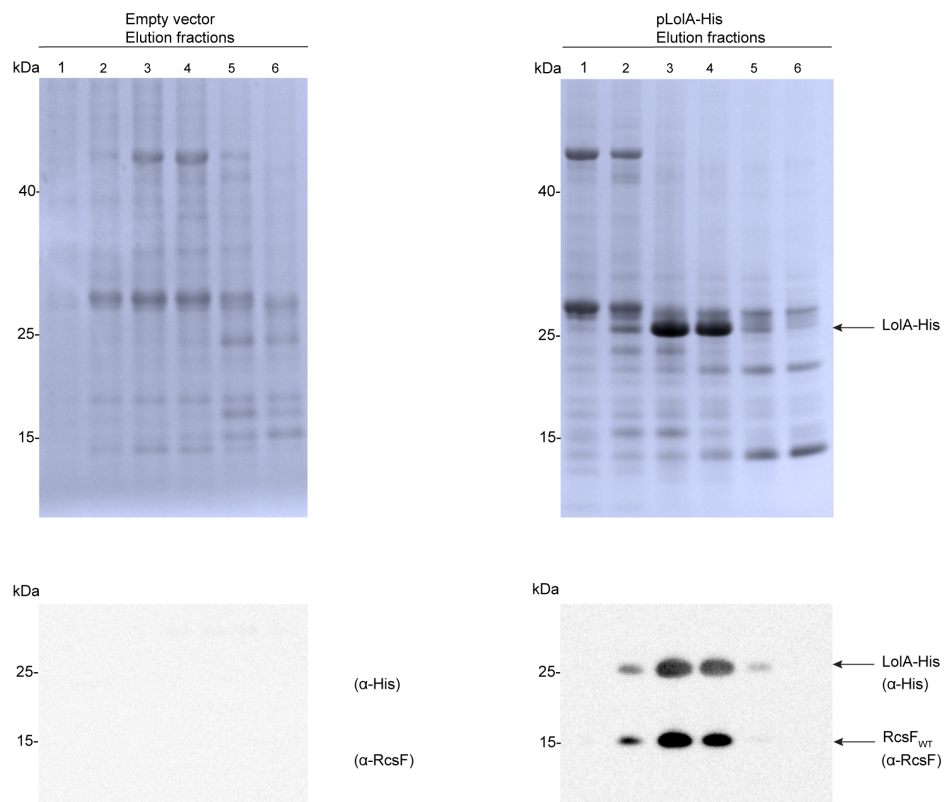


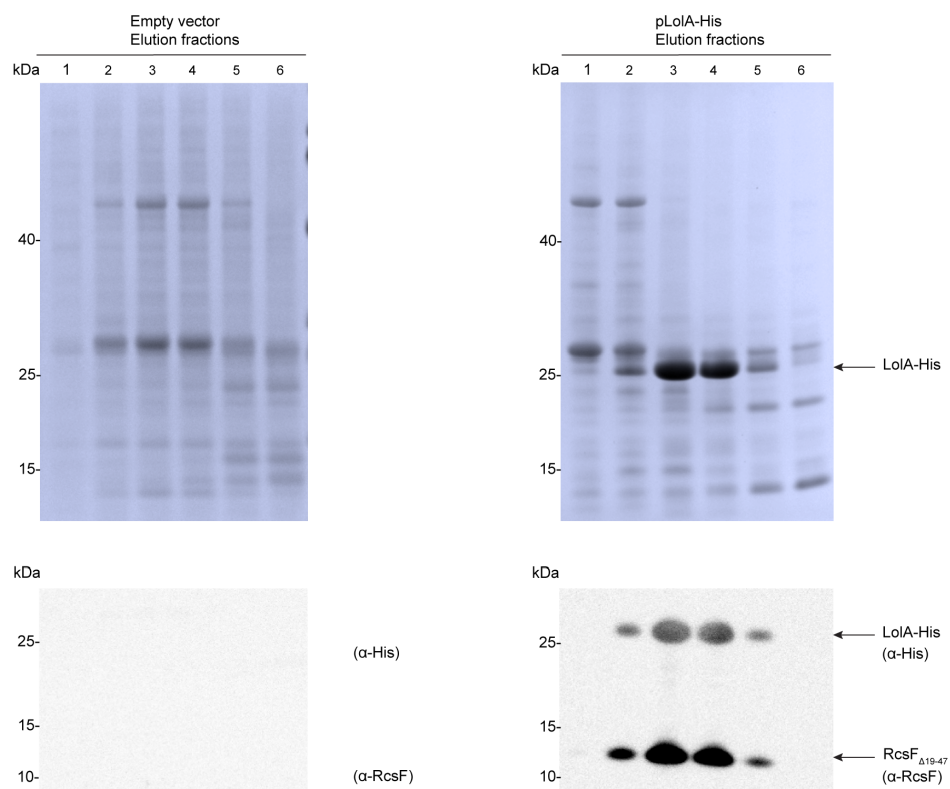
Figure 4. N-terminal disordered linkers interact with the Lol system to target lipoproteins to the outer membrane. a. Deleting *Lpp* rescues targeting of RcsF $_{\Delta 19-47}$ and NlpD $_{\Delta 29-64}$ to the outer membrane. The outer and inner membranes were separated *via* centrifugation in a sucrose density gradient (Methods). RcsF $_{\Delta 19-47}$ and NlpD $_{\Delta 29-64}$ accumulate in the inner membrane of cells expressing *Lpp*, the most abundant Lol substrate. Deletion of *lpp* improves their targeting to the outer membrane. Aligned dot-plots are shown with a line at mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. The significance of differences among bacterial strains was assessed according to analysis of variance (ANOVA), followed by the application of Tukey's multiple comparison test. No adjustments were made for multiple comparison. RcsF variants were expressed from the chromosome and NlpD variants were expressed from a plasmid. **b.** *In vitro* pull-down experiments show that RcsF_{WT} and RcsF $_{\Delta 19-47}$ are transferred from LolA to LolB. LolA-RcsF_{WT} and LolA-RcsF $_{\Delta 19-47}$ complexes were obtained by LolA-His affinity chromatography followed by size exclusion chromatography (Methods). Each complex was incubated with LolB-Strep that was previously purified *via* Strep-Tactin affinity chromatography (Methods). Both RcsF variants were eluted in complex with LolB-strep, while LolA was only present in the flow through. I, input; FT, flow through; E, eluate. $n = 3$ biologically independent experiments.

To obtain further insights into the mechanism at play here, we next monitored whether linker deletion impacted the transfer of RcsF from LolA to LolB *in vitro*. LolA with a C-terminal His-tag was expressed in the periplasm of cells expressing wild-type RcsF or RcsF $_{\Delta 19-47}$ and purified to near homogeneity via affinity chromatography (Methods; **Extended Data Fig. 6**). Both RcsF_{WT} and RcsF $_{\Delta 19-47}$ were detected in immunoblots of the fractions containing purified LolA (**Extended Data Fig. 6**), indicating that both proteins form a soluble complex with LolA and confirming that they use this chaperone for transport across the periplasm. LolB was expressed as a soluble protein in the cytoplasm and purified by taking advantage of a C-terminal Strep-tag; LolB, which has more affinity for lipoproteins than LolA³⁶, was then incubated with LolA-RcsF_{WT} or LolA-RcsF $_{\Delta 19-47}$ and pulled-down using Strep-Tactin beads (Methods). As both RcsF_{WT} and RcsF $_{\Delta 19-47}$ were detected in the LolB-containing pulled-down fractions (**Fig. 4b**), we conclude that both proteins were transferred from LolA to LolB. Thus, the linker is not required for the transfer of RcsF from LolA to LolB.

a

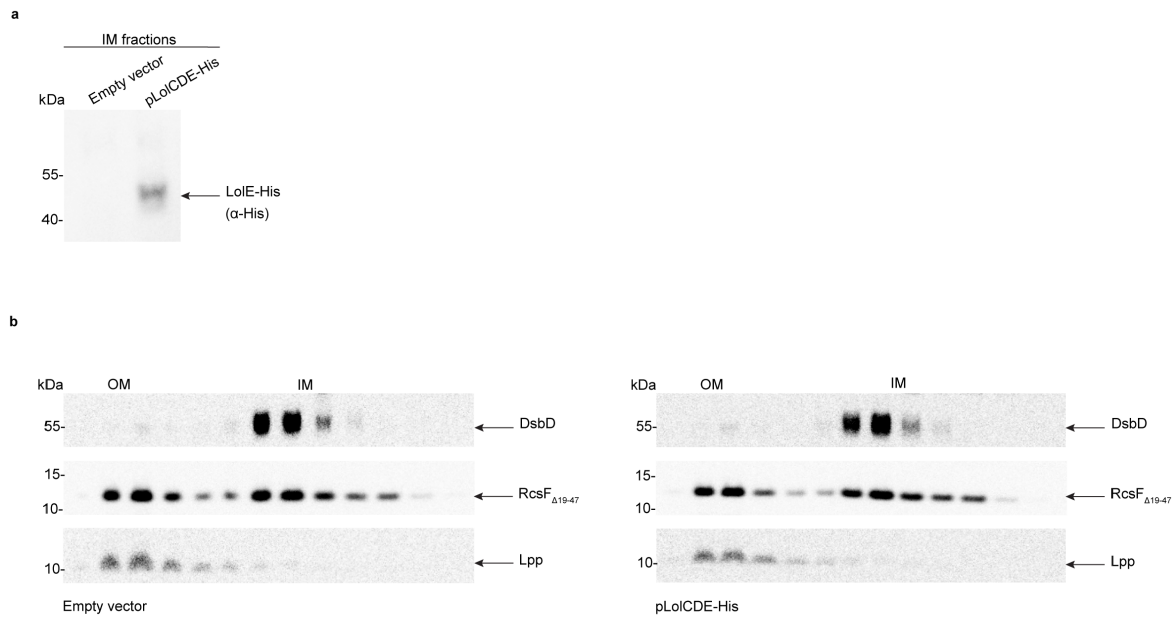


b



Extended Data Fig. 6. Complexes between LolA and RcsF_{WT} or RcsF_{Δ19-47} can be purified. Both RcsF_{WT} (a) and RcsF_{Δ19-47} (b) were eluted in complex with LolA-His *via* affinity chromatography followed by size exclusion chromatography. Elution fractions were analyzed *via* SDS-PAGE and proteins stained with Coomassie Brilliant Blue (Methods). Chromosomally-encoded RcsF variants were detected by immunoblotting fractions with α -RcsF antibodies. Images are representative of three independent experiments.

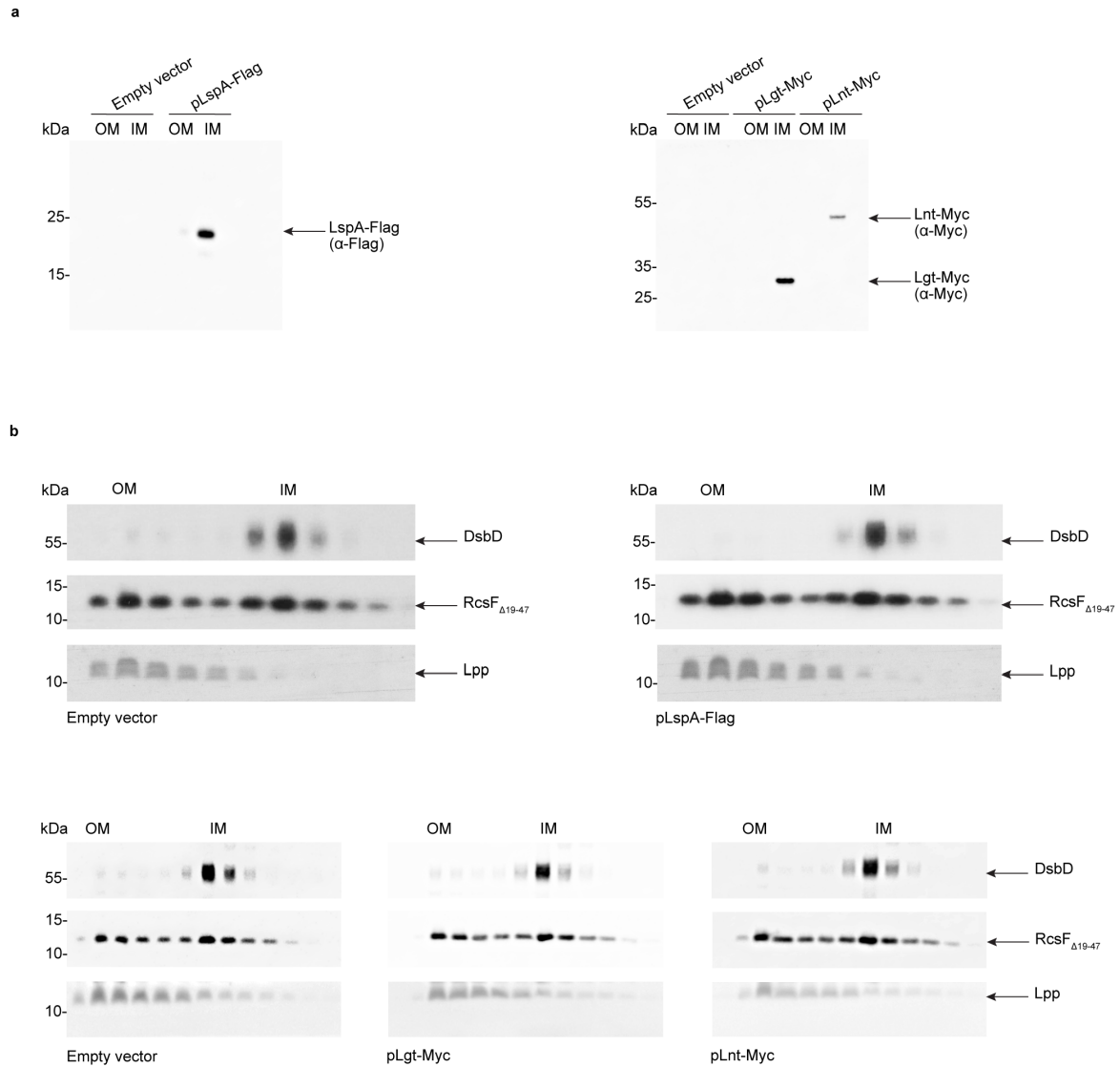
We then focused on the LolCDE ABC transporter in charge of extracting outer membrane lipoproteins and transferring them to LolA. The structure of LolCDE in complex with a lipoprotein substrate was recently reported³⁷. Although most of the lipoprotein polypeptide chain is non-visible in the structure, the triacylated N-terminal cysteine could be found inside a channel located between the plane of the membrane and the periplasmic domains of LolC and LolD. We thus hypothesized that N-terminal flexible linkers might be important to limit clashes between the lipoprotein substrate and the periplasmic domains of LolC and LolD. To test whether having a disordered linker increases the efficiency of LolCDE in lipoprotein extraction, we monitored the impact of over-expressing all LolCDE components (**Extended Data Fig. 7a**) on the targeting of RcsF_{Δ19-47} to the outer membrane. However, we found that over-expression of LolCDE had no effect (**Extended Data Fig. 7b**).



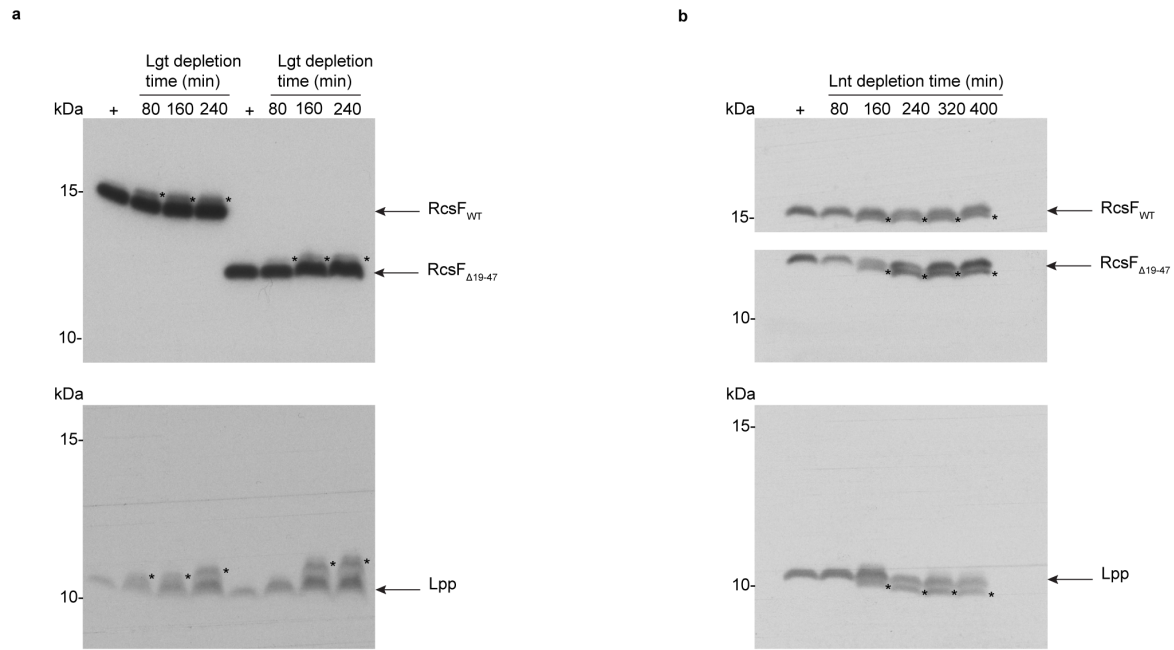
Extended Data Fig. 7. Overexpression of LolCDE does not restore targeting of RcsF $_{\Delta 19-47}$. **a.** Expression level of LolCDE-His. Cells were grown in LB plus 0.2% arabinose at 37 °C until OD₆₀₀ = 0.7 (Methods). Membrane and soluble fractions were separated with a sucrose density gradient (Methods). LolE-His was detected in the inner membrane fraction by immunoblotting with α -His (Methods). Images are representative of three independent experiments. **b.** The outer membrane (OM) and inner membrane (IM) were separated with a sucrose density gradient. Expression of LolCDE did not rescue OM targeting of RcsF $_{\Delta 19-47}$. Images are representative of experiments performed in biological triplicate.

Finally, we tested whether linker deletion impacted the processing of RcsF $_{\Delta 19-47}$ by the enzymes involved in lipoprotein maturation (Lgt, LspA, and Lnt; **Supplementary Figure 1a**). We found that overexpressing Lgt, LspA and Lnt had no impact on membrane targeting (**Extended Data Fig. 8**). Further, the immature forms of RcsF $_{\Delta 19-47}$ that accumulate when these enzymes are depleted (**Extended Data Fig. 9**) were not detected when these enzymes are expressed, indicating that deletion of the linker does not block the processing of RcsF $_{\Delta 19-47}$ by the enzymes of the maturation pathway. Thus, taken together, our results suggest that retention of RcsF $_{\Delta 19-47}$ in the inner membrane does not result from the impairment of a specific step, but rather from

less efficient processing of the truncated lipoprotein by the entire lipoprotein maturation pathway (see Discussion).



Extended Data Fig. 8. Overexpressing Lgt, LspA, and Lnt does not rescue the targeting of RcsF Δ 19-47 to the outer membrane. **a.** Expression levels of Lgt, LspA, and Lnt. Cells were grown in LB (plus 25 μ M IPTG for cells expressing LspA) at 37 °C until OD₆₀₀ = 0.7 (Methods). Outer membrane (OM) and inner membrane (IM) were separated with a sucrose density gradient (Methods). Lgt-Myc and Lnt-Myc were detected in the IM via immunoblotting with α -Myc. LspA-Flag was detected in the IM with α -Flag. **b.** Cells overexpressing Lgt, LspA, or Lnt were exposed to a sucrose density gradient (Methods). RcsF Δ 19-47 was retained in the IM in all conditions. Images in a and b are representative of three independent experiments.

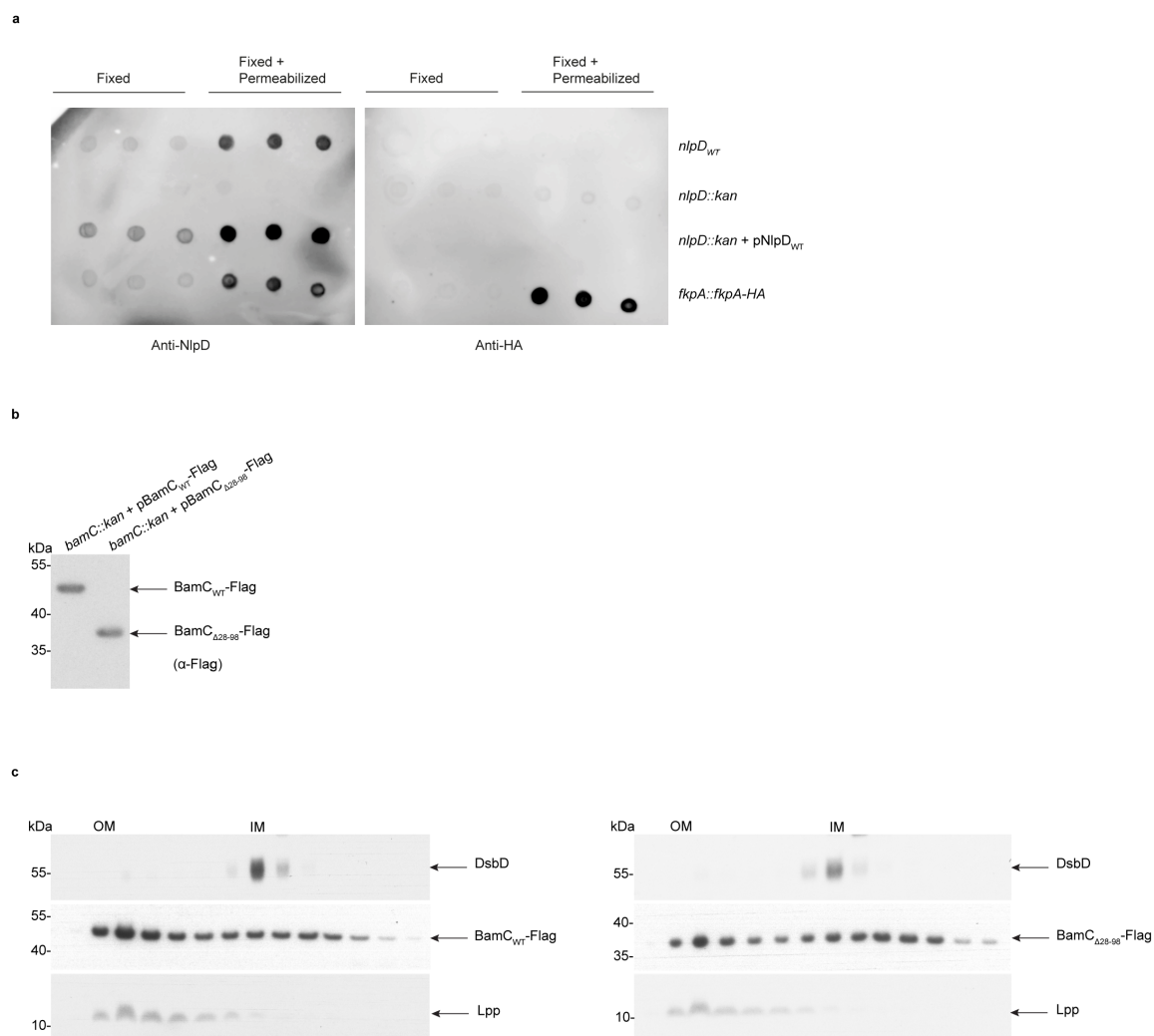


Extended Data Fig. 9. RcsF_{Δ19-47} is processed by Lgt and Lnt. **a.** Lgt replete (+) or deplete samples were obtained by growing cells in the presence or absence of L-arabinose. Overnight cultures were grown at 37°C in LB medium with 0.2 % arabinose; cells were then diluted in fresh LB medium and grown at 37°C in the absence of L-arabinose for the indicated times to deplete Lgt. **b.** Lnt replete (+) or deplete samples were obtained from cultures supplemented with L-arabinose or fucose, respectively. Overnight cultures were grown at 37°C in LB medium in the presence of 0.2 % arabinose; cells were then washed with fresh LB medium, diluted in LB supplemented with 0.2 % fucose and grown at 37°C for the indicated times to deplete Lnt. Protein samples were subjected to SDS-PAGE tricine or Tris-glycine 16% gel and probed for Lpp and RcsF by immunoblotting. PreproLpp (in the case of Lgt depletion) and apoLpp (in the case of Lnt depletion) are noted as *. The same annotation is used for RcsF variants. The prepro and apo forms of Lpp, RcsF_{WT} and RcsF_{Δ19-47} are detected only when Lgt and Lnt are depleted. This indicates that RcsF_{Δ19-47} and, as expected RcsF and Lpp, are processed by Lgt and Lnt. Images in a and b are representative of biological triplicates.

Are linkers important for surface exposure?

We noted that two of the lipoproteins under investigation here, Pal and RcsF, have been reported to be surface-exposed^{27,38,39}. A topology model has been proposed to explain how RcsF reaches the surface: the lipid moiety of RcsF is anchored in the outer leaflet of the outer membrane while the N-terminal linker is exposed on the cell surface before being threaded

through the lumen of β -barrel proteins³⁹. Thus, in this topology, the linker allows RcsF to cross the outer membrane. It was therefore tempting to speculate that N-terminal disordered linkers may be used by lipoproteins as a structural device to cross the outer membrane and reach the cell surface. In addition, the accumulation of RcsF $_{\Delta 19-47}$ in the inner membrane (**Fig. 2a**) also suggests that Lol may be using N-terminal linkers to recognize lipoproteins destined to the cell surface before their extraction from the inner membrane in order to optimize their targeting to the machinery exporting them to their final destination (BAM in the case of RcsF^{27,39,40}). To further investigate the link between disordered linkers and surface exposure, we tested whether NlpD could be detected on the cell surface, which turned out to be the case (**Extended Data Fig. 10a**). In addition, we found that the N-terminal linker of BamC, a lipoprotein with access to the cell surface^{41,42}, is important for normal outer membrane targeting (**Extended Data Fig. 10b, 10c**).



Extended Data Fig. 10. The disordered linker of BamC and NlpD, two proteins that can be detected on the cell surface, is important for targeting to the outer membrane (OM).

a. NlpD is exposed on the cell surface. Fixed cells were spotted (with or without permeabilization with lysozyme and EDTA) onto nitrocellulose membranes and probed with anti-NlpD antibodies. NlpD is detected on non-permeabilized cells, indicating that the protein is on the cell surface. An HA-tagged version of the periplasmic, soluble protein FkpA, expressed from the chromosome, was used as a permeabilization control. HA-tagged FkpA can only be detected with anti-HA antibodies following permeabilization. **b,c.** The disordered linker of BamC is important for its targeting to the outer membrane (OM). **b.** Expression levels of BamC_{WT}-Flag and BamC_{Δ28-98}-Flag. The BamC variants were expressed from a plasmid with a C-terminal Flag-tag. Cells were grown in LB plus IPTG; when OD₆₀₀ reached 0.6, cells were precipitated with trichloroacetic acid (Methods). Samples were subjected to SDS-PAGE and probed with α-Flag-tag antibodies. **c.** Cells expressing BamC_{WT}-Flag and BamC_{Δ28-98}-Flag were grown in LB plus IPTG until OD₆₀₀ reached 0.7. Cells were then subjected to fractionation; the outer (OM) and inner (IM) membranes were separated by sucrose density gradient (Methods). Deletion of the linker of BamC alters the targeting of this lipoprotein to

the OM: more BamC_{Δ28-98} is detected in the IM fractions than BamC_{WT}. Images in a, b,c are representative of biological triplicate.

Discussion

Lipoproteins are crucial for essential cellular processes such as envelope assembly and virulence. However, despite their functional importance and their potential as targets for new antibacterial therapies, we only have a vague understanding of the molecular factors that control their biogenesis. By discovering the role played by N-terminal disordered linkers in lipoprotein sorting, this study adds an important new layer to our comprehension of lipoprotein biogenesis in Gram-negative bacteria. Critically, it also indicates that the current model of lipoprotein sorting—that sorting between the two membranes is controlled by the 2 or 3 residues that are adjacent to the lipidated cysteine—needs to be revised. Lipoproteins with unstructured linkers at their N-terminus are commonly found in Gram-negative bacteria including many pathogens (see below); further work will be required to determine whether these linkers control lipoprotein targeting in organisms other than *E. coli*, laying the foundation for designing new antibiotics.

It was previously shown that both *lolA* and *lolB* (but not *lolCDE*) can be deleted under specific conditions²¹, suggesting at least one alternate route for the transport of lipoproteins across the periplasm and their delivery to the outer membrane. During this investigation, we envisaged the possibility that the linker could be required to transport lipoproteins via a yet-to-be-identified pathway independent of LolA/LolB. However, our observations that both RcsF and RcsF_{Δ19-47} were found in complex with LolA (**Extended Data Fig. 6**) and were transferred by LolA to LolB (**Fig. 4b**) does not support this hypothesis. Instead, our data clearly indicate that

lipoproteins with N-terminal linkers still depend on the Lol system for extraction from the inner membrane and transport to the outer membrane (**Supplementary Figure 1a**); they also suggest that N-terminal linkers improve lipoprotein processing by Lol (see below).

The general view is that outer membrane lipoproteins remain oriented toward the periplasm. Excitingly, a series of reports have started to challenge this paradigm by identifying surface-exposed lipoproteins in *E. coli* and other bacteria^{22,23}. By showing that four surface-exposed lipoproteins (RcsF, Pal, NlpD, BamC) have an N-terminal linker that is important for outer membrane targeting, our results also support a model in which N-terminal disordered linkers are important for displaying lipoproteins on the cell surface. Investigating whether a dedicated Lol-dependent route exists for surface-exposed lipoproteins will be the subject of future research. It is worth noting that N-terminal linkers are commonly found in lipoproteins expressed by the pathogens *Borrelia burgdorferi* and *Neisseria meningitidis*^{23,43}; lipoprotein surface exposure is common in these pathogens.

Our work also delivers crucial insights into the functional importance of disordered segments in proteins in general. Most proteins are thought to present portions that are intrinsically disordered. For instance, it is estimated that 30-50% of eukaryotic proteins contain regions that do not adopt a defined secondary structure *in vitro*⁴⁴. However, demonstrating that these unstructured regions are functionally important *in vivo* is challenging. By showing that an N-terminal disordered segment downstream of the Lol signal is required for the correct sorting of lipoproteins, our work provides direct evidence that evolution has selected intrinsic disorder by function.

In conclusion, the data reported here establish that the triage of lipoproteins between the inner and outer membranes is not solely controlled by the Lol sorting signal; additional molecular determinants, such as protein intrinsic disorder, are also involved. Our data further highlight the previously unrecognized heterogeneity of the important lipoprotein family and call for a careful evaluation of the maturation pathways of these lipoproteins.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article and its supplementary information file.

REFERENCES

1. Silhavy, T.J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**, a000414 (2010).
2. Weiner, J.H. & Li, L. Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. *Biochim Biophys Acta* **1778**, 1698-713 (2008).
3. Ricci, D.P. & Silhavy, T.J. Outer Membrane Protein Insertion by the β -barrel Assembly Machine. *EcoSal Plus* **8**, 0035-2018 (2019).
4. El Rayes, J., Rodríguez-Alonso, R. & Collet, J.F. Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles. *Curr Opin Microbiol* **61**, 25-34 (2021).
5. 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**, 2492-7 (2011).
6. Sherman, D.J. et al. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science* **359**, 798-801 (2018).
7. Malinverni, J.C. et al. YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol Microbiol* **61**, 151-64 (2006).
8. Laloux, G. & Collet, J.F. "Major Tom to ground control: how lipoproteins communicate extra-cytoplasmic stress to the decision center of the cell". *J Bacteriol*, 00216-17 (2017).
9. Kovacs-Simon, A., Titball, R.W. & Michell, S.L. Lipoproteins of bacterial pathogens. *Infect Immun* **79**, 548-61 (2011).
10. Szewczyk, J. & Collet, J.F. The Journey of Lipoproteins Through the Cell: One Birthplace, Multiple Destinations. *Adv Microb Physiol* **69**, 1-50 (2016).
11. Babu, M.M. et al. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol* **188**, 2761-73 (2006).

12. Narita, S.I. & Tokuda, H. Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1414-1423 (2017).
13. Horler, R.S., Butcher, A., Papangelopoulos, N., Ashton, P.D. & Thomas, G.H. EchoLOCATION: an in silico analysis of the subcellular locations of Escherichia coli proteins and comparison with experimentally derived locations. *Bioinformatics* **25**, 163-6 (2009).
14. Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in E. coli. *Biochim Biophys Acta* **1694**, IN1-9 (2004).
15. Gennity, J.M. & Inouye, M. The protein sequence responsible for lipoprotein membrane localization in Escherichia coli exhibits remarkable specificity. *J Biol Chem* **266**, 16458-64 (1991).
16. Terada, M., Kuroda, T., Matsuyama, S.I. & Tokuda, H. Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *J Biol Chem* **276**, 47690-4 (2001).
17. Hara, T., Matsuyama, S. & Tokuda, H. Mechanism underlying the inner membrane retention of Escherichia coli lipoproteins caused by Lol avoidance signals. *J Biol Chem* **278**, 40408-14 (2003).
18. Narita, S. & Tokuda, H. Amino acids at positions 3 and 4 determine the membrane specificity of Pseudomonas aeruginosa lipoproteins. *J Biol Chem* **282**, 13372-8 (2007).
19. Lewenza, S., Mhlanga, M.M. & Pugsley, A.P. Novel inner membrane retention signals in Pseudomonas aeruginosa lipoproteins. *J Bacteriol* **190**, 6119-25 (2008).
20. Lorenz, C., Dougherty, T.J. & Lory, S. Correct Sorting of Lipoproteins into the Inner and Outer Membranes of Pseudomonas aeruginosa by the Escherichia coli LolCDE Transport System. *mBio* **10**, 00194-19 (2019).
21. Grabowicz, M. & Silhavy, T.J. Redefining the essential trafficking pathway for outer membrane lipoproteins. *Proc Natl Acad Sci U S A* **114**, 4769-4774 (2017).
22. Wilson, M.M. & Bernstein, H.D. Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. *Trends Microbiol* **24**, 198-208 (2016).
23. Zuckert, W.R. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta* **1843**, 1509-16 (2014).
24. Pride, A.C., Herrera, C.M., Guan, Z., Giles, D.K. & Trent, M.S. The outer surface lipoprotein VolA mediates utilization of exogenous lipids by Vibrio cholerae. *MBio* **4**, e00305-13 (2013).
25. Gonnet, P., Rudd, K.E. & Lisacek, F. Fine-tuning the prediction of sequences cleaved by signal peptidase II: a curated set of proven and predicted lipoproteins of Escherichia coli K-12. *Proteomics* **4**, 1597-613 (2004).
26. Sueki, A., Stein, F., Savitski, M.M., Selkig, J. & Typas, A. Systematic Localization of Escherichia coli Membrane Proteins. *mSystems* **5**, 00808-19 (2020).
27. Cho, S.H. et al. Detecting Envelope Stress by Monitoring beta-Barrel Assembly. *Cell* **159**, 1652-64 (2014).
28. Uehara, T., Parzych, K.R., Dinh, T. & Bernhardt, T.G. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J* **29**, 1412-22 (2010).
29. Gerding, M.A., Ogata, Y., Pecora, N.D., Niki, H. & de Boer, P.A. The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-

- membrane invagination during cell constriction in *E. coli*. *Mol Microbiol* **63**, 1008-25 (2007).
30. Hussein, N.A., Cho, S.H., Laloux, G., Siam, R. & Collet, J.F. Distinct domains of *Escherichia coli* IgaA connect envelope stress sensing and down-regulation of the Rcs phosphorelay across subcellular compartments. *PLoS Genet* **14**, e1007398 (2018).
 31. Tsang, M.J., Yakhnina, A.A. & Bernhardt, T.G. NlpD links cell wall remodeling and outer membrane invagination during cytokinesis in *Escherichia coli*. *PLoS Genet* **13**, e1006888 (2017).
 32. Shrivastava, R., Jiang, X. & Chng, S.S. Outer membrane lipid homeostasis via retrograde phospholipid transport in *Escherichia coli*. *Mol Microbiol* **106**, 395-408 (2017).
 33. Cohen, E.J., Ferreira, J.L., Ladinsky, M.S., Beeby, M. & Hughes, K.T. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science* **356**, 197-200 (2017).
 34. Asmar, A.T. et al. Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS Biol* **15**, e2004303 (2017).
 35. Li, G.W., Burkhardt, D., Gross, C. & Weissman, J.S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624-35 (2014).
 36. 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**, 5877-82 (2009).
 37. Tang, X. et al. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nat Struct Mol Biol* **28**, 347-355 (2021).
 38. Michel, L.V. et al. Dual orientation of the outer membrane lipoprotein Pal in *Escherichia coli*. *Microbiology* **161**, 1251-9 (2015).
 39. Konovalova, A., Perlman, D.H., Cowles, C.E. & Silhavy, T.J. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of beta-barrel proteins. *Proc Natl Acad Sci U S A* **111**, E4350-8 (2014).
 40. Rodriguez-Alonso, R. et al. Structural insight into the formation of lipoprotein-beta-barrel complexes. *Nat Chem Biol* **16**, 1019-1025 (2020).
 41. Webb, C.T. & Lithgow, T. Identification of BamC on the Surface of *E. coli*. *Methods Mol Biol* **1329**, 215-25 (2015).
 42. Webb, C.T. et al. Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *J Mol Biol* **422**, 545-55 (2012).
 43. Brooks, C.L., Arutyunova, E. & Lemieux, M.J. The structure of lactoferrin-binding protein B from *Neisseria meningitidis* suggests roles in iron acquisition and neutralization of host defences. *Acta Crystallogr F Struct Biol Commun* **70**, 1312-7 (2014).
 44. Bardwell, J.C. & Jakob, U. Conditional disorder in chaperone action. *Trends Biochem Sci* **37**, 517-25 (2012).

ACKNOWLEDGMENTS

We thank Asma Boujtat for technical help. We are indebted to the members of the Collet laboratory and to Nassos Typas (EMBL, Heidelberg) for helpful suggestions and discussions and to Tom Silhavy (Princeton) and Nienke Buddelmeijer (Pasteur) for providing bacterial strains. J.S. was a research fellow of the FRIA and J.F.C. is an Investigator of the FRFS-WELBIO. This work was funded by the WELBIO (grant WELBIO-CR-20190-03), by grants from the F.R.S.-FNRS, from the Fédération Wallonie-Bruxelles (ARC 17/22-087), from the European Commission via the International Training Network Train2Target (721484), and from the EOS Excellence in Research Program of the FWO and FRS-FNRS (G0G0818N).

AUTHOR CONTRIBUTIONS

J.-F.C., J.E.R., J.S., and S.H.C. designed and performed the experiments. J.E.R., J.S., and S.H.C. constructed the strains and cloned the constructs. J.-F.C., J.E.R., J.S., S.H.C., and A.M. analyzed and interpreted the data. B.I.I. performed the structural analysis, M.D the microscopy analysis and N.C. the dot blot analysis. J.-F.C., J.E.R., and J.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

COMPETING INTERESTS

The authors declare no competing financial interests.

METHODS

Bacterial growth conditions

Bacterial strains used in this study are listed in **Supplementary Table 3**. Bacterial cells were cultured in Luria broth (LB) at 37 °C unless stated otherwise. The following antibiotics were added when appropriate: spectinomycin (100 µg/mL), ampicillin (200 µg/mL), chloramphenicol (25 µg/mL), and kanamycin (50 µg/mL). L-arabinose (0.2%), isopropyl-β-D-thiogalactoside (IPTG; 20 µM to 1 mM) and L-fucose (0.2%) were used for induction or repression when appropriate.

Bacterial strains and plasmids

DH300 (a derivative of *Escherichia coli* MG1655 carrying a chromosomal *rprA::lacZ* fusion at the λ attachment site⁴⁵) was used as wildtype throughout the study. All deletion mutants were obtained by transferring the corresponding alleles from the Keio collection⁴⁶ (kan^R) into DH300⁴⁵ via P1 phage transduction. Deletions were verified by PCR and the absence of the protein was verified *via* immunoblotting (when possible). If necessary, the kanamycin cassette was removed via site-specific recombination mediated by the yeast Flp recombinase with pCP20 vector⁴⁷. All strains expressing the RcsF mutants used for subcellular fractionation lacked *rscB* in order to prevent induction of Rcs.

The plasmids used in this study are listed in **Supplementary Table 4** and the primers appear in **Supplementary Table 5**. RcsF, Pal, and NlpD were expressed from the low-copy vector pAM238⁴⁸ containing the SC101 origin of replication and the *lac* promoter. To produce pSC202 for RcsF expression, *rscF* (including approximately 30 base pairs upstream of the coding sequence) was amplified by PCR from the chromosome of DH300 (primer pair

SH_RcsF(PstI)-R and SH_RcsFU-R (kpnI)-F). The amplification product was digested with KpnI and PstI and inserted into pAM238, resulting in pSC202. *nlpD* was amplified using primers JR1 and JR2 and *pal* was amplified with primers JS145 and JS146. Amplification products were digested with PstI-XbaI and KpnI-XbaI, respectively, generating pJR8 (for NlpD expression) and pJS20 (for Pal expression). To clone *rcsF*_{Δ19-47}, the nucleotides encoding the RcsF signal sequence were amplified using primers SH_RcsFU-R(kpnI)_F and SH_RcsFss-Fsg (NcoI)_R, and those encoding the RcsF signaling domain were amplified using primers SH_RcsFss-Fsg (NcoI)_R and SH_RcsF(PstI)_R. In both cases, pSC202 was used as template. Then, overlapping PCR was performed using SH_RcsFUR(kpnI)_F and SH_RcsF(PstI)_R from the two PCR products previously obtained. The final product was digested with KpnI and PstI, and ligated with pAM238 pre-digested with the same enzymes, yielding pSC201. To add a GS linker (Ser-Gly-Ser-Gly-Ser-Gly-Ala-Met) into pSC201, the primers SH_GS linker_F and SH_GS linker_R were mixed, boiled, annealed at room temperature, and ligated with pSC201 pre-digested with NcoI, generating pSC198. pSC199 was generated similarly, but using primers SH_SG linker_F and SH_SG linker_R and plasmid pSC198. pSC200 was generated using primers SH_Da linker_F and SH_SG linker_R and plasmid pSC199.

*rcsF*_{FkpA} and *rcsF*_{col} were obtained by inserting DNA sequences corresponding to helical linker fragments (FkpA Ser94-Glu125 and colicin IA Ile213-Lys282) into *rcsF*_{Δ19-47} at NcoI and RsrII restriction sites. The *fkpA* gene fragment was amplified from the *E. coli* MC4100 chromosome (JS50/JS51 primers) and the *cia* gene fragment was chemically synthesized as a gene block by Integrated DNA Technologies (IDT). The resulting plasmids were pJS18 and pJS27, respectively. pAM238 does not contain the *lacIq* repressor. Therefore, to enable expression-level regulation by IPTG, strains containing the pAM238 plasmids expressing RcsF variants

were co-transformed with pET22b, a high-copy plasmid from a different incompatibility group (pBR223 origin of replication; Novagen) containing the *lacIq* repressor. Chromosomal insertion of *rcsF*_{Δ19-47} or *rcsF*_{CDD} at the *rcsF* locus was performed via λ-Red recombineering⁴⁹ with pSIM5-Tet plasmid (a gift of D. Hughes). In the first step, the *cat-sacB* cassette was introduced and later replaced by mutant *rcsF*. The same technique was used to add an HA tag to the chromosomal *fkpA* gene.

The *pal* allele lacking the linker region (*pal*_{Δ26-56}) was created *via* overlapping PCR. The pJS20 plasmid served as template for PCR with the M13R/M13F external primers and JS152/JS153 internal primers. The truncated allele was cloned into pAM238 at the same restriction sites as the full-length allele, producing pJS24. The *nlpD* allele lacking the linker regions (*nlpD*_{Δ29-64}) was created *via* overlapping PCR. *E. coli* chromosomal *nlpD* served as template for the PCR, with JR1/JR2 as external primers and JR7/JR8 as internal primers. The truncated allele was then cloned into pAM238 at the same restriction sites as the full-length allele, producing pJR10.

To clone *bamC*_{WT} with a C-terminal Flag-tag, the chromosomal *bamC* gene was amplified by PCR using primers JR92 and JR93 and then inserted into pSC213, a modified IPTG-regulated pAM238 with a ribosome binding site and a C-terminal Flag-tag, previously digested with the restriction enzymes NcoI and XbaI, using the NEB Gibson master mix. To clone *bamC*_{Δ28-98} with a C-terminal Flag-tag using pSC213, *bamC*₁₋₂₇ and *bamC*₉₉₋₃₄₄ were amplified by PCR using the chromosomal *bamC* gene and primers JR92/JR94 and JR95/JR96, respectively. pSC213 was digested using restriction enzymes NcoI and XbaI. *bamC*_{Δ28-98} was then created by mixing PCR products (*bamC*₁₋₂₇ and *bamC*₉₉₋₃₄₄) with digested pSC213 and the NEB Gibson master mix. BamC_{WT} and BamC_{Δ28-98} were induced using 20 μM and 40 μM of IPTG, respectively (**Extended Data Figure 10**).

The chromosomal *lolCDE* operon was amplified *via* PCR using primers JS277 and JS278 (adding a C-terminal His-tag to LolE) and then inserted into pBAD33 using the restriction sites PstI and XbaI, resulting in pJR203. The expression level of LolE-His was verified *via* immunoblotting.

The sequence encoding *lolB* without its N-terminal cysteine was first amplified from the chromosome *via* PCR using primers JR50/PL387 (adding a C-terminal Strep-tag). It was then cloned into pET28a using the restriction sites XbaI and PstI, generating plasmid pJR90. *lolA* was amplified using chromosomal *lolA* as PCR template for primers JR30/JR31 (JR31 contains the sequence of a His-tag) and then cloned into pBAD18 using KpnI and XbaI, resulting in pJR48.

The genes encoding Lgt and Lnt were amplified from the chromosome with PCR primers AG389/AG403 and AG393/JR74, respectively. AG403 and JR74 also encode a Myc-tag. PCR products were cloned into pAM238 using KpnI and PstI. Expression levels of the proteins constitutively expressed from the plasmid were verified *via* immunoblotting (**Extended Data Fig. 8**). *lspA* was amplified with PCR primers JR77/JR78. The PCR product was cloned into pSC213, a modified IPTG-regulated pAM238 with a ribosome binding site and a C-terminal Flag-tag, using NcoI and XbaI. Expression of LspA-Flag was induced by adding 25 μ M IPTG.

Cell fractionation and sucrose density gradients

Cell fractionation was performed as described previously ⁵⁰ with some modifications. Four hundred milliliters of cells were grown at 37°C in LB until the optical density at 600 nm (OD₆₀₀) of the culture reached 0.7. Cells were harvested *via* centrifugation at 6,000 x g at 4 °C

for 15 min, washed with TE buffer (50 mM Tris-HCl pH 8, 1 mM EDTA), and resuspended in 20 mL of the same buffer. The washing step was skipped with the *Alpp* strains to prevent the loss of outer membrane vesicles. DNase I (1 mg, Roche), RNase A (1 mg, Thermo Scientific), and a half tablet of a protease inhibitor cocktail (cOmplete EDTA-free Protease Inhibitor Cocktail tablets; Roche) were added to cell suspensions, and cells were passed through a French pressure cell at 1,500 psi. After adding MgCl₂ to a final concentration of 2 mM, the lysate was centrifuged at 5,000 x g at 4 °C for 15 min in order to remove cell debris. Then, 16 mL of supernatant were placed on top of a two-step sucrose gradient (2.3 mL of 2.02 M sucrose in 10 mM HEPES pH 7.5 and 6.6 mL of 0.77 M sucrose in 10 mM HEPES pH 7.5). The samples were centrifuged at 180,000 x g for 3 h at 4 °C in a 55.2 Ti Beckman rotor. After centrifugation, the soluble fraction and the membrane fraction were collected. The membrane fraction was diluted four times with 10 mM HEPES pH 7.5. To separate the inner and the outer membranes, 7 mL of the diluted membrane fraction were loaded on top of a second sucrose gradient (10.5 mL of 2.02 M sucrose, 12.5 mL of 1.44 M sucrose, 7 mL of 0.77 M sucrose, always in 10 mM HEPES pH 7.5). The samples were then centrifuged at 112,000 x g for 16 h at 10 °C in a SW 28 Beckman rotor. Approximately 30 fractions of 1.5 mL were collected and odd-numbered fractions were subjected to SDS-PAGE, transferred onto a nitrocellulose membrane, and probed with specific antibodies. Graphs were created in GraphPad Prism 9 *via* spline analysis of the curves representing a mean of three independent experiments. Aligned dot-plots representing the outer and inner membrane percentages were created in GraphPad Prism 9 by measuring the area under the curve for each experiment.

Immunoblotting

Protein samples were separated *via* 10% or 4-12% SDS-PAGE (Life Technologies) and transferred onto nitrocellulose membranes (GE Healthcare Life Sciences). The membranes

were blocked with 5% skim milk in 50 mM Tris-HCl pH 7.6, 0.15 M NaCl, and 0.1% Tween 20 (TBS-T). TBS-T was used in all subsequent immunoblotting steps. The primary antibodies were diluted 5,000 to 20,000 times in 1% skim milk in TBS-T and incubated with the membrane for 1 h at room temperature. The anti-RcsF, anti-DsbD, anti-Lpp and anti-NlpD antisera were generated by our lab. These antibodies were used at 1/20,000, 1/9,000, 1/7,000 and 1/10,000, respectively. Anti-Pal, used at 1/20,000, was a gift from R. Lloubès and anti-His, used at 1/10,000, is a peroxidase-conjugated antibody (Qiagen). Anti-HA (Biolegend MMS-101R), used at 1/1,000, and anti-Flag (Sigma), used at 1/10,000, are mouse antisera while anti-Myc (Santa Cruz Biotechnology), used at 1/7,000, is a rabbit antiserum. The membranes were incubated for 1 h at room temperature with horseradish peroxidase-conjugated goat anti-rabbit IgG or anti-mouse (Sigma) at a 1:10,000 dilution. Labelled proteins were detected *via* enhanced chemiluminescence (Pierce ECL Western Blotting Substrate, Thermo Scientific) and visualized using X-ray film (Fuji) or a camera (Image Quant LAS 4000 and Vilber Fusion solo S). In order to quantify proteins levels, band intensities were measured using ImageJ version 1.46r (National Institutes of Health).

Testing the surface exposure of NlpD using dot-blots

The dot-blot method was adapted from ²⁷. Cells were grown in LB medium to an OD₆₀₀ of 0.5. Cells were directly fixed for 15 min at room temperature with 0.04 % glutaraldehyde and 2 % formaldehyde. 2 x 1 mL of cells were centrifuged for 1 min at 16,000 x g. Fixed cells were resuspended in PBS. For permeabilization, cells were incubated for 30 min at 4°C in a buffer containing lysozyme (1 mg/mL) and EDTA (10 mM). After centrifugation, they were resuspended in PBS. Cells (2 ml) were then spotted on nitrocellulose membranes and the spots were left to dry completely. The membranes were blocked for 30 min with 5 % (w/v) skim milk in TBS. Membranes were washed 5 min in TBS-Tween 20 0.05 % (TBS-T 0.05 %) and

then incubated for 1 h at room temperature with primary antibody solutions. The anti-NlpD antibody was diluted 3,000 times with 1 % skim milk in TBS. Anti-HA mouse antibody (Biolegend MMS-101R) was diluted 1,000 times in BSA 2 % in TBS. Membranes were washed 3 times 5 min in TBS-T 0.05 % and incubated for 1 h at room temperature with secondary antibody solutions: anti-rabbit-HRP (Cytiva NA934) or anti-mouse-HRP (Sigma A4416) diluted 2,000 times with skim milk 5 % in TBS. Membranes were washed three times 5 min in TBS-T 0.05 %. Labeled proteins were detected with standard ECL or Substrate HRP Immobilon (Millipore) and visualized with a CCD camera.

β -galactosidase assay

β -galactosidase activity was measured as described previously⁵¹. Briefly, cells harboring *PrprA-lacZ* at the *attB* phage lambda site on the chromosome were diluted 1:100 from overnight cultures in LB, then incubated at 37°C and harvested at OD₆₀₀ = 0.6. 20 μ L of cells were harvested and incubated with 80 μ L of permeabilization solution (60 mM Na₂HPO₄·2H₂O, 40 mM NaH₂PO₄·H₂O, 10 mM KCl, 1 mM MgSO₄·7H₂O, 50 mM β -mercaptoethanol) for 45 min at room temperature. Then, 600 μ L of substrate (1 mg/mL O-nitrophenyl- β -d-galactoside, 50 mM β -mercaptoethanol) were added. The mixture was further incubated at 30°C for 20-90 min. The reaction was then stopped by adding 700 μ L of 1 M Na₂CO₃ and absorbance was measured at 420 nm. The standardized amount of galactosidase activity was reported in Miller units. The ration of *PrprA-lacZ* induction was calculated relative to the basal level in a WT strain. Expression-level estimations were performed as follows. Cultures used for β -galactosidase activity (0.5 mL per culture) were precipitated with 10% trichloroacetic acid, washed with ice-cold acetone, and resuspended in 0.2 mL Laemmli SDS sample buffer. Samples (5 μ L) were subjected to SDS-PAGE and immunoblotted with anti-RcsF antibody. Box and whisker plots were prepared using GraphPad Prism 9.

SDS-EDTA sensitivity assay

Cells were grown in LB at 37 °C until they reached an OD₆₀₀ of 0.7. Tenfold serial dilutions were made in LB and plated on LB agar supplemented with spectinomycin (100 µg/mL) when necessary. Plates were incubated at 37 °C. To evaluate the sensitivity of the *pal* mutant, plates were supplemented with 0.01% SDS and 0.5 mM EDTA.

Microscopy image acquisition

Cells were grown in LB at 37 °C until OD₆₀₀ = 0.5. Cells growing in exponential phase were spotted onto a 1% agarose phosphate-buffered saline pad for imaging. Images were taken at 21°C. Cells were imaged on a Nikon Eclipse Ti2-E inverted fluorescence microscope with a CFI Plan Apochromat DM Lambda 100X Oil, N.A. 1.45, W.D. 0.13 mm objective. Images were collected on a Prime 95B 25 mm camera (Photometrics). We used a Cy5-4050C (32 mm) filter cube (Nikon). Image acquisition was performed with NIS-Element Advance Research version 4.5. Whole field pictures are 2048 x 2048 pixels, 1 pixel accounts for 0.11 x 0.11 µm. Exposure time for phase contrast images were taken for 70 ms with a 75 % light intensity with SOLA II shutter. Display range: 5402 – 11791.

Protein purification

JR90 cells were grown in LB supplemented with kanamycin (50 µg/mL) at 37 °C. When the culture OD₆₀₀ reached 0.5, the expression of cytoplasmic LolB-Strep was induced with 1 mM IPTG. Cells (1 L) were pelleted when they reached OD₆₀₀ = 3 and resuspended in 25 mL of buffer A (200 mM NaCl and 50 mM NaPi, pH 8) containing one tablet of cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). Cells were lysed *via* two passages through a French pressure cell at 1,500 psi. The lysate was centrifuged at 30,000 x g for 40 min at 4 °C in a JA

20 rotor and the supernatant was mixed with Strep-Tactin resin (IBA Lifesciences) previously equilibrated with buffer A. After washing the resin with 10 column volumes of buffer A, LolB-Strep was eluted with 5 column volumes of buffer A supplemented with 5 mM desthiobiotin. LolB-Strep was finally desalted using a PD10 column (GE Healthcare).

Soluble LolA-RcsF_{WT} and LolA-RcsF_{Δ19-47} complexes were purified *via* affinity chromatography as follows. Cells co-expressing LolA either with wild-type RcsF (JR47) or RcsF_{Δ19-47} (JR44) were grown in LB at 37 °C supplemented with 200 µg/mL ampicillin until an OD₆₀₀ of 0.5. Protein expression was then induced with 0.2% arabinose. Cells (1 L) were pelleted at an OD₆₀₀ of 3 and resuspended in 25 mL of buffer A containing one tablet of protease inhibitor cocktail. Cells were lysed *via* two passages through a French pressure cell at 1,500 psi. The lysate was centrifuged at 45,000 x g for 30 min at 4 °C using a 55.2 Ti Beckman rotor. To obtain the soluble fraction, the supernatant was centrifuged at 180,000 x g for 1 h at 4 °C using the same rotor. The supernatant was added to a His Trap HP column (Merck) previously equilibrated with buffer A. The column was washed with 10 column volumes of buffer A supplemented with 20 mM imidazole and LolA-His was eluted using a gradient of imidazole (from 20 mM to 300 mM). The fractions obtained were analyzed *via* SDS-PAGE; LolA was detected around 25 kDa. RcsF variants and LolA-His were detected *via* immunoblotting with an anti-RcsF and anti-His antibodies, respectively. Fractions containing LolA-RcsF variants were pooled, concentrated to 1 mL using a Vivaspin 4 Turbo concentrator (Cut-off 5 kDa; Sartorius), and purified *via* size-exclusion chromatography with a Superdex S75-10/300 column (GE Healthcare).

Pull down and transfer of RcsF variants from LolA to LolB

LolB-Strep was incubated at 30 °C for 20 min under agitation with LolA-RcsF_{WT} or with LolA-RcsF_{Δ19-47} (LolA-RcsF_{WT} and LolA-RcsF_{Δ19-47} complexes were purified as described above). The mixture was added to magnetic Strep beads (MagStrep type 3 beads, IBA Life science) previously equilibrated with buffer A and incubated for 30 min at 4 °C on a roller. After washing the beads with the same buffer, LolB-Strep was eluted with buffer A supplemented with 50 mM biotin. Samples were analyzed *via* SDS-PAGE and LolA and LolB were detected with Coomassie Brilliant Blue (Bio-Rad). RcsF was detected *via* immunoblotting with an anti-RcsF antibody.

Depletion of the maturation enzymes Lgt and Lnt

Lgt null strains (JR301 and JR302) expressing Lgt from an arabinose-inducible promoter on the pCHAP9231 plasmid⁵² were prepared using P1 phage transduction from the Lgt null strain PAP9403⁵² (gift from N. Buddelmeijer, Pasteur, France). Overnight cultures were grown at 37°C in LB medium in the presence of 0.2 % arabinose (to allow expression of Lgt) and chloramphenicol (pCHAP9231 has chloramphenicol resistance). Cells were then diluted in fresh LB medium and grown at 37°C until growth arrest was observed and OD₆₀₀ remained constant. Cells were then precipitated with 10 % trichloroacetic acid, washed with ice-cold acetone and resuspended in tricine SDS sample buffer. Samples were subjected to SDS-PAGE using tricine 16 % gel at 125 V, and immunoblotted with anti-Lpp and anti-RcsF antibodies.

The Lnt null strain PAP8504⁵³ is a kind gift from N. Buddelmeijer. In PAP8504 strain, an arabinose inducible promoter controls the expression of the chromosomal *lnt* gene. Overnight precultures were grown at 37°C in LB medium in the presence of 0.2 % arabinose to allow the expression of Lnt. Precultures were then washed three times with fresh LB medium and diluted 1:100 in LB supplemented with 0.2 % fucose. When growth arrest was observed and OD₆₀₀

remained constant, cells were precipitated as described previously. Samples were resuspended in Tris-glycine SDS sample buffer and subjected to SDS-PAGE using Tris-glycine 16 % gel at 125 V, and immunoblotted with anti-Lpp and anti-RcsF antibodies. To monitor the impact of Lnt depletion on RcsF $_{\Delta 19-47}$, plasmid pSC201 was introduced in PAP8504, yielding strain JR320.

Structural analysis of lipoproteins

When X-ray, cryo-EM, or NMR structures were available, the missing residues were completed through comparative modeling using MODELLER version 9.22⁵⁴. If no structure of the lipoprotein was available, then the most pertinent analogous structure from proteins belonging to the same PFAM group was used as template for comparative modeling. The linker was defined as the unstructured fragment from the N-terminal Cys of the mature form until the first residue with well-defined secondary structure (α -helix or β -strand) belonging to a globular domain. Short, intermediate, and long linkers had lengths of <12, 12-22, and >22 residues, respectively. Images were generated using UCSF Chimera version 1.13.1⁵⁵.

Statistical methods

The significance of differences among bacterial strains was assessed using GraphPad Prism 9 according to analysis of variance (ANOVA), followed by the application of Tukey's multiple comparison test.

45. Majdalani, N., Hernandez, D. & Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* **46**, 813-26 (2002).
46. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
47. Cherepanov, P.P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14 (1995).

48. Gil, D. & Bouche, J.P. ColE1-type vectors with fully repressible replication. *Gene* **105**, 17-22 (1991).
49. Yu, D. et al. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* **97**, 5978-83 (2000).
50. 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**, 6400-5 (2007).
51. Miller, J.C. *Experiments in Molecular Genetics*, (Cold Spring Harbor Laboratory Press, New York, 1972).
52. Pailler, J., Aucher, W., Pires, M. & Buddelmeijer, N. Phosphatidylglycerol::prolipoprotein diacylglyceryl transferase (Lgt) of *Escherichia coli* has seven transmembrane segments, and its essential residues are embedded in the membrane. *J Bacteriol* **194**, 2142-51 (2012).
53. Robichon, C., Vidal-Ingigliardi, D. & Pugsley, A.P. Depletion of apolipoprotein N-acyltransferase causes mislocalization of outer membrane lipoproteins in *Escherichia coli*. *J Biol Chem* **280**, 974-83 (2005).
54. Šali, A. & Blundell, T.L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779-815 (1993).
55. Pettersen, E.F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605-1612 (2004).

Supplementary Table 1. List of the verified lipoproteins of *E. coli* used for the structural analysis in this study

Supplementary Table 2. RcsF mutants used in this study and the amino acid sequences of their corresponding N-terminal linkers. The acylated cysteine is the first residue listed

RcsF linkers	Amino acid sequence
RcsF _{WT}	CSMLSRSPVEPVQSTAPQPKAEPAPKAPRATPV
RcsF _{CDD}	CDDLRSRSPVEPVQSTAPQPKAEPAPKAPRATPV
RcsF _{Δ19-47}	CSMGPV
RcsF _{G5}	CSMSLFDAPAMSGSGSGAMSGSGSGAMPV
RcsF _{G52}	CSMSGSGSGAMSGSGSGAMPV
RcsF _{G53}	CSMSGSGSGAMPV
RcsF _{FkpA}	CSMGSDQEIEQTLQAFEARVKSSAQAKMEKDAADNEPV
RcsF _{col}	CSMGILDTRLSELEKNGGAALAVLDAQQARLLGQQTRNDRAISEARNKLSSVTESLNTARNALTRAEQQLTQ QKPV

Supplementary Table 3. *E. coli* strains used in this study

Strains	Genotype and description	Source
DH300	<i>rprA-lacZ</i> MG1655 (<i>argF-lac</i>) U169	⁴⁵
Keio collection single mutants	<i>rcsF::kan</i> , <i>rcsB::kan</i> , <i>pal::kan</i> , <i>nlpD::kan</i> , <i>envC::kan</i> , <i>lpp::kan</i> , <i>bamC::kan</i>	⁴⁶
XL1-Blue	<i>endA1 gyrA96 (nal^R) thi-1 recA1 relA1 lac glnV44F' [::Tn10 proAB⁺ lacI^q Δ(lacZ)M15] hsdR17 (rK⁻ mK⁺)</i>	Stratagene
BL21	F- <i>ompT hsdSB</i> (rB- mB-) <i>gal dcm</i> (DE3)	Novagen
JS41	DH300 Δ <i>rcsF</i> pAM238	This study
JS265	DH300 Δ <i>rcsF</i> pJS18	This study
JS346	DH300 Δ <i>rcsF</i> <i>rcsB::kan</i> pET22b	This study
JS267	JS346 pJS18	This study
JS325	DH300 <i>pal::kan</i>	This study
JS331	JS325 pJS20	This study
JS345	JS325 pJS24	This study
JS360	DH300 Δ <i>rcsF</i> pJS27	This study
JS363	JS346 pJS27	This study
JS364	DH300 Δ <i>rcsF</i> pSC202	This study
JS372	DH300 Δ <i>rcsF</i> pSC201	This study
JS395	JS346 pSC198	This study
JS396	JS346 pSC199	This study
JS397	JS346 pSC200	This study
JS398	JS346 pSC201	This study
JS573	JS346 pSC202	This study
JS574	DH300 Δ <i>rcsF</i> pSC198	This study
JS575	DH300 Δ <i>rcsF</i> pSC199	This study
JS576	DH300 Δ <i>rcsF</i> pSC200	This study
JS639	Δ <i>rcsB</i> <i>lpp::kan</i> <i>rcsF::rcsF</i> _{Δ19-47}	This study
JS632	<i>rcsB::kan</i> <i>rcsF::rcsF</i> _{CDD}	This study

JR30	<i>nlpD::kan</i>	This study
JR31	JR30 pJR8	This study
JR32	JR30 pJR10	This study
JR2	DH300 pAM238	This study
JR88	BL21 <i>rcsF::kan</i>	This study
JR90	JR88 pET28-cytoplasmic LolB-Strep	This study
JR187	<i>rcsB::kan rcsF::rcsF_{Δ19-47}</i>	This study
JR149	<i>ΔnlpD</i>	This study
JR121	<i>ΔnlpD envC::kan</i>	This study
JR122	JR121 pJR8	This study
JR123	JR121 pJR10	This study
JR188	JR187 pAM238	This study
JR191	JR187 pAG833	This study
JR204	JR187 pJR203	This study
JR194	JR187 pBAD33	This study
JR211	JR187 pJR209	This study
JR257	JR187 pJR239	This study
JR274	JR149 <i>lpp::kan</i>	This study
JR279	JR274 pJR10	This study
JR292	JS325 pAM238	This study
JR293	JR187 pSC213	This study
JR44	<i>rcsB::kan rcsF::rcsF_{Δ19-47}</i> pJR48	This study
JR47	<i>rcsB::kan</i> pJR48	This study
JR77	<i>rcsB::kan rcsF::rcsF_{Δ19-47}</i> pBAD18	This study
JR78	<i>rcsB::kan</i> pBAD18	This study
JR311	<i>bamC::kan</i> pSC213-BamC _{WT} -Flag	This study
JR312	<i>bamC::kan</i> pSC213-BamC _{Δ28-98} -Flag	This study
PAP9403	BW25113 <i>lgt::Kanr</i> (pCHAP9231)	⁵²
JR301	<i>Δrcsb lgt::kan</i> pCHAP9231	This study

JR302	<i>Δrcsb rcsF::rcsF_{Δ19-47} lgt::kan</i> pCHAP9231	This study
PAP8504	BW25113 ybeX-(kan-rpoCter-paraB)-Int	⁵³
JR320	PAP8504 pSC201	This study
JR41	<i>Δrcsb</i>	This study
NCA99	MG1655 <i>fkpA::fkpA-HA</i>	This study

Supplementary Table 4. Plasmids used in this study

Plasmids	Features	Source
pAM238	IPTG-regulated P_{lac} , pSC101-based, spectinomycin (no <i>lacIQ</i>)	⁴⁸
pBAD18	Arabinose inducible P_{BAD} , ampicillin	⁵⁶
pBAD33	Arabinose inducible P_{BAD} , chloramphenicol	⁵⁶
pET28a	IPTG regulated T7 promoter, kanamycin	Novagen
pET22b	IPTG regulated T7 promoter, ampicillin	Novagen
pCP20	FLP ⁺ , λ cI857 ⁺ , λ PR Rep ^{ts} , ampicillin, chloramphenicol	⁴⁷
pSIM5-Tet	pSC101 plasmid, <i>repA</i> ^{ts} , tetRA, λ -Red (Gram-Beta-Exo), cI857, tetracycline	Gift from D. Hughes
pJS18	pAM238 RcsF _{FKpA} FkpA linker (S94-E125)	This study
pJS20	pAM238 Pal _{WT}	This study
pJS24	pAM238 Pal _{Δ26-56}	This study
pJS27	pAM238 RcsF _{col} Colicin Ia linker (I213-K282)	This study
pSC198	pAM238 RcsF _{GS3} (C ₁₆ S ₁₇ M ₁₈ S ₁₉ GSGSGAMG)	This study
pSC199	pAM238 RcsF _{GS2} (C ₁₆ S ₁₇ M ₁₈ S ₁₉ GSGSGAMSGSGSGAMG)	This study
pSC200	pAM238 RcsF _{GS} (C ₁₆ S ₁₇ M ₁₈ S ₁₉ LFDAPAMSGSGSGAMSGSGSGAMG)	This study
pSC201	pAM238 RcsF _{Δ19-47} (C ₁₆ S ₁₇ M ₁₈ G ₁₉ P ₂₀)	This study
pSC202	pAM238 RcsF _{WT}	This study
pJR8	pAM238 NlpD _{WT}	This study
pJR10	pAM238 NlpD _{Δ29-64} (C ₂₆ S ₂₇ D ₂₈ A ₂₉)	This study
pJR48	pBAD18 LolA-6xHis	This study
pJR90	pET28 Cytoplasmic LolB-Strep	This study
pJR203	pBAD33 LolCDE-6xHis	This study
pJR209	pAM238 Lnt-Myc	This study
pJR239	pSC213 LspA-Flag	This study
pSC213	pAM238, IPTG-regulated P_{lac} , <i>lacIQ</i> , triple Flag tag	This study
pAG833	pAM238 Lgt-Myc	This study
pJR311	pSC213 BamC _{WT} -Flag	This study

pJR312	pSC213 BamC Δ_{28-98} -Flag	This study
pCHAP9231	pBAD18s-Cm-lgtWT-c-myc2	⁵²

56. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-30 (1995).

Supplementary Table 5. Primers used in this study

Primer	Sequence 5' to 3'
JS50_FkpAlinker_fw	acatccatggggtccgaccaagagatcgaac
JS51_FkpAlinker_rv	atgtcggaccggttcgtatcagccgcgtc
JS143_Pal_-100b	cgtcttcggcaactgatgg
JS144_Pal_+100b	ttggtgcctgagcaaaagcg
JS145_Pal_fw	ACATggtaccTTAATTGAATAGTAAAGGAATC
JS146_Pal_rv	ATGTtctagaTTAgtaaaccagtaccgcac
JS152_PalNoLinker_overlapPCR_fw	tgttctccaacCAGGCTCGTCTGCAAATG
JS153_PalNoLinker_overlapPCR_rv	CAGACGAGCCTGgttgaagaacatgccgc
JS277_LolCDEHis_fw	ACATtctagaTCTTTGCTACAGCAACCAGAC
JS278_LolCDE_His_rv	ATGTctgcagTTAGTGATGGTGATGGTGATGACCctggccgctaaggactcg
JS289_lred_catSacBin_RcsF_fw	tcctgattcaatattgacgttttgatcatatcattgaggaaatactAAAATGAGACGTTGATCGGCACG
JS290_lred_catSacBin_RcsF_rev	tataggcgagcggaataacgcctatttgctcgaactggaaactgcATCAAAGGGAAAACGTGTCCA
JS291_lred_RcsF_catSacBout_fw	tcctgattcaatattgacgttttgatcatatcattgaggaaatactATGCGTGCTTTACCGATCTGTT
JS292_lred_RcsF_catSacBout_rv	tataggcgagcggaataacgcctatttgctcgaactggaaactgcTCATTCGCCGTAATGTTAAGC
JS293_junction1lred_RcsFup_fw	gcggagctgttaaaggctg
JS294_junction2lred_RcsFdown_rv	gagcaatgagatgcagttcg
JS295_junction1lred_cat-out_rv	CGGGCAAGAATGTGAATAAAGG
JS296_junction2lred_sacB-out_fw	GCTGTACCTCAAGCGAAAGG
M13R	CAGGAAACAGCTATGACCATG
M13F	TGTAAAACGACGGCCAGT
PL145_rcsF_-100b	cgtttttaccagacctggc
PL146_rcsF_+100	atatcattcaggacggcgcttgccc
PL153_rcsB_-100b	acatctgattcgtgagaagg
PL154_rcsB+100b	taatgggaatcgtaggccgg
PL168_Fw_lpp_-100	CAATTTTTTATCTAAAACCCAGCG

PL169_Rv_lpp_+100	CCAGAGCAAGGGAATATGTTACGCG
SH_Da linker_F	CATGaGcTTATTCGACGCGCCGgc
SH_Da linker_R	catggCCGGCGCGTCGAATAAgCt
SH_RcsF(PstI)_R	gagaCTGCAGtcaTTTCGCCGTAATGTTAAG
SH_RcsFUR(kpnI)_F	GAGGGTACCggtttgatcatacattg
RcsFss-Fsg (NcoI)_F	GCGGCTGTTCCATGgggccggtccgaattatac
RcsFss-Fsg (NcoI)_R	ggaccggccCCATGGAACAGCCGCTTAGCATGAG
SH_GS linker_F	CATGagtggctctggatctggtgc
SH_GS linker_R	catggcaccagatccagagccact
JR1_NlpD_fw	GAGATCTAGATTATTAACCAATTTTCCTGGGGGATAA
JR2_NlpD_rv	AGAGCTGCAGTTATCGCTGCGGCAAATAACGCA
JR7_NlpDoverlap_fw	GGCTGGCAGGCTGTTCTGACGCGCAGCAACCGCAAATTCA
JR8_NlpDoverlap_rv	TGAATTTGCGGTTGCTGCGCGTCAGAACAGCCTGCCAGCC
JR23_Fw_NlpD-98	CAGGTCAGCGTATCGTGAACATC
JR24_Rv_NlpD+100	TCATTTAAATCATGAACTTTCAGCG
JR30_Fw_LolA_-28_pBAD18	ACATGGTACCCGGGAGTGACGTAATTTGAGGAAT
JR31_Rev_LolA_His_pBAD18	ATGTTCTAGAttaatgatgatgatgatgctcgaGCTTACGTTGATCATCTACCGTGAC
JR50_Rev_cytoplasmic_LolB_nostop_StrepTag_stop	CCAACTCGAGTCACTTTTCGAACTGCGGGTGGCTCCAGCTTGCTTTCACTATCCAGTTATCCAT
JR56-Fw--100-envC	GTTGTCGCTG ATGGGTA
JR57-Rev-+100envC	AATCATCAATGACGATGGCA
JR74-Rev-Lnt-myc-tag-PstI	AAAAACTGCAGctacaggtcttctcgctaatacagttctgttcgcttgcTTACGTCGCTGACGCAGAC
JR77-Fw-NcoI-LspA	gagaCCATGGgtAGTCAATCGATCTGTTCAAC
JR78-Rev-LspA-no stop-BamHI	gagaGGATCCTTGTTTTTTCGCTCTAG
AG389_lgt_-49_Fw_KpnI	AAAAAggtaccTTCAATCGCTGTTCTCTTTC
AG393_lnt_-49_Fw_KpnI	AAAAAggtaccACCCAGCCGAAGCTGGATG
AG403_lgt_mycCT_PstI	AAAAACTGCAGctacaggtcttctcgctaatacagttctgttcgcttgcGGAAACGTGTTGCTGTGGGC
PL387-LolBwoss-Fw-NcoI	acacCCATGGcggttaccagcccaagg
Fw-catsacB-flkpa	GTTAACCTGGGGTGAGATGCCCGATCCTGGAGATATGGAAAATGAGACGTTGATCGGCACG

RV-catsacB -fkpa	TAAAAAACCGCCGCCTGGTCAGGCGGCGGTTCTTAATGCATCAAAGGGAAAAGTGTCCA
Fw-fkpa-cter-HA	GTTAACCTGGGGTGAGATGCCCCGATCCTGGAGATATGGATGAAATCACTGTTTAAAGTAAC
RV-fkpa-cter-HA + 2aa	TAAAAAACCGCCGCCTGGTCAGGCGGCGGTTCTTAATGCCTACGCATAATCCGGCACATCATA CG GATAGCTTGCTTTTTTAGCAGAATCTGCGG
JR92-Fw-BamC-FL	acaggaaacagaccatgggtGCTTACTCTGTTCAAAAGTCG
JR93-Rv-BamC-FL	tggtctttgtagtctctagaCTTGCTAAACGCAGCCTG
JR94-Rv-BamC-Linker	cggccgtgaaAGAACTACAGGCAGCGAG
JR95-Fw-BamC-Linker	ctgtagtctTTCACGGGCGATACCGCTTC
JR96-Rv-BamC-Linker	tggtctttgtagtctctagaCTTGCTAAACGCAGCCTGG
JR97-Fw--100 lgt	CCTTGTGCTA TTATTCGC
JR98-Rv-+100 lgt	GACCAAAAATGGAAAGCG
bamC EUHR F	tgaaggaactgggtcttg
bamC EDHR R	ttatccgaactacgtccg
ColicinIalinker_ geneBLOCK	acatccatggggATTCTGGACACGCGGTTGTCAGAGCTGGAAAAAATGGCGGGGCAG CCCTTGCCGTTCTTGATGCACAACAGGCCCGTCTGCTCGGGCAGCAGACACGGA ATGACAGGGCCATTCAGAGGCACGGAATAAACTCAGTTCAGTGACGGAATCG CTTAACACGGCCCGTAATGCATTAACCAGAGCTGAACAACAGCTGACGCAACA GAAAgcgggtccgacat

Chapter II. Structural insights into the LolA-RcsF complex

In this chapter, I will briefly present promising preliminary results that can lead us to the structure of LolA in complex with a lipoprotein, in this case RcsF_{Δ19-47}.

Introduction

The urgent threat of antibiotic resistance and the global shortage of novel antimicrobials force scientific researchers to find new targets, by exploring all bacterial compartments and all the possible cellular pathways, hoping to find a promising spot. We previously discussed, in chapter II of the introduction, the importance of lipoproteins for the viability of the bacteria. Being responsible of the transport of lipoproteins in the cell envelope, the Lol system is one of the attractive pathways for the design of antibiotics. One strategy to design an antibiotic is to study the structure of the target, as well as its mechanism of action¹. Although the structures of the different components of the Lol system have been solved (LolA², LolB², and LolCDE³), none of them encompasses a lipoprotein in complex. It is true that the structure of LolA, LolB or LolCDE might be enough to design possible antimicrobials, however there is no doubt that if the structure of one the constituents of the Lol system exists in complex with a lipoprotein, this new structure will give more insights into the interaction between a Lol component and lipoproteins. It might also lead to answer the remaining questions regarding the mechanistic details of the interaction between Lol and lipoproteins. For instance, we could have more insights into the mechanism used by LolB to insert lipoproteins in the OM. We could also answer the question regarding how many acyl chains bind inside the cavity of LolA and LolB: are there three acyl chains present inside of this cavity, two or only one? In this chapter, I will present preliminary results that show a promising way into the determination of the crystal structure of LolA in complex with the variant of RcsF that lacks its disorder linker (RcsF_{Δ19-47}).

Results

RcsF_{WT} and RcsF_{Δ19-47} co-elute with the periplasmic chaperone LolA

As previously shown in chapter I of the results section, RcsF_{WT} and the linker-less variant, RcsF_{Δ19-47}, can both co-elute with the periplasmic chaperone LolA (**Extended Data Fig.6**). Both complexes LolA-RcsF_{WT} and LolA-RcsF_{Δ19-47} were obtained after a histidine affinity purification and a size-exclusion chromatography. However, we noticed that the LolA-RcsF_{Δ19-47} complex was more stable *in vitro* than LolA-RcsF_{WT}. In fact, the purified LolA-RcsF_{WT} complex aggregates after being conserved at -80°C, which is not the case of the LolA-RcsF_{Δ19-47} complex. This could be explained by the high flexibility of the N-terminal linker of RcsF_{WT} that might destabilize the purified complex *in vitro*. Indeed, IDPs are very hard to crystallize for instance, due to the lack of their conformational stability *in vitro*, which constitutes a barrier against the formation of crystals^{4,5}.

Towards the structure determination of LolA-RcsF_{Δ19-47}

Based on the co-elution results, we decided to take advantage of the *in vitro* stability of the LolA-RcsF_{Δ19-47} complex and try to solve the structure of a liganded LolA. To optimize the purification, the nickel affinity purification was followed up by a second affinity chromatography step by taking advantage of a Strep-tag fused to the C-terminus of RcsF_{Δ19-47}. Therefore, we first pulled on the histidine tag of LolA (**Fig.1a**) and then on the Strep-tag of RcsF_{Δ19-47} (**Fig.1b**). The elution fractions obtained were finally subjected to size exclusion chromatography (**Fig.1c**). The complex obtained after the three purification steps is stable and does not form aggregates, which opens

the way to the resolution of the crystal structure of LolA-RcsF $_{\Delta 19-47}$. Trials for obtaining crystals from this complex are still ongoing.

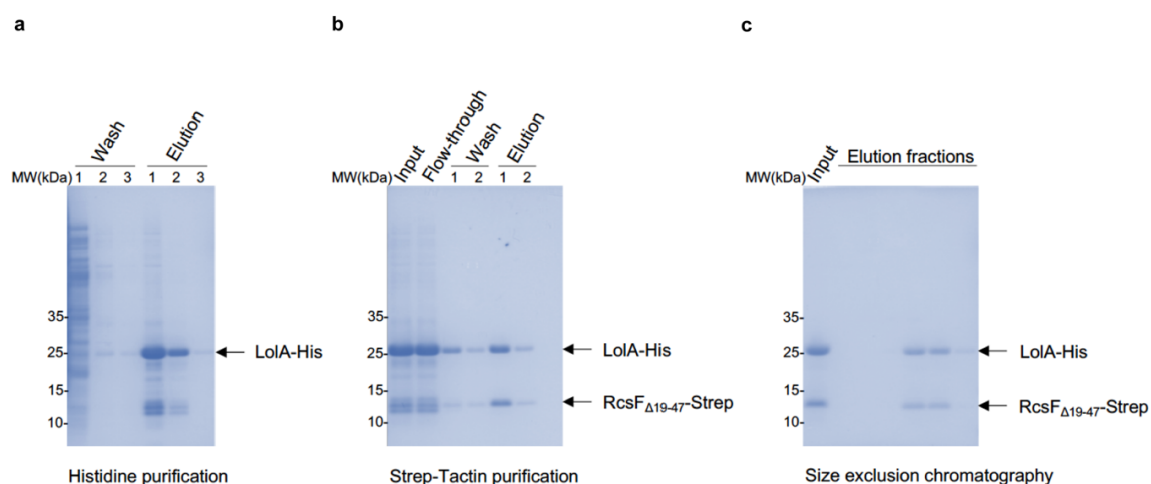


Figure 1. Three steps purification of the LolA-RcsF $_{\Delta 19-47}$ complex. The periplasmic complex LolA-RcsF $_{\Delta 19-47}$ was purified by affinity chromatography using the His and the Strep tags of LolA and RcsF $_{\Delta 19-47}$, respectively (a,b). The complex was then subjected to size-exclusion chromatography (c).

Discussion

The Lol system has been studied in detail by many groups for more than thirty years. It is still a spotlight for research, due to its importance for bacterial survival. However, so far, none of the available structures show a lipoprotein bound to a Lol component. This is an obvious indicator that purifying of a Lol-lipoprotein complex and solving its structure is challenging. Few years ago, the group of Tokuda had to use around 12 liters of cell cultures, co-expressing LolA and the lipoprotein Pal, to obtain a pure complex after size-exclusion chromatography⁶. A Flag-tag and a Histidine-tag were added to LolA and Pal, respectively. However, when the complex was purified by a

two-steps purification (first pulling on LolA and then on Pal), the yield of LolA-Pal obtained was very low (100 ng/L of culture) ⁶. Therefore, the complex was purified using a one-step His-tag affinity purification followed by a size-exclusion chromatography⁶. Next the authors tried to increase the amount of the periplasmic LolA-Pal complex by depleting LolB and deleting the most abundant lipoprotein, Lpp. By doing this, they managed to increase the amount of periplasmic Pal by about four times. This clearly highlights how fast and efficient is the transfer of lipoproteins from the inner to the outer membranes in the cell envelope by the Lol system. It also proves how difficult it is to purify the periplasmic chaperone LolA in complex with a lipoprotein from wild-type cells.

Here we were able to obtain sufficient amount of LolA in complex with the linker-less RcsF variant, RcsF_{Δ19-47} (**Fig.1**). The LolA-RcsF_{Δ19-47} complex was obtained from three liters of culture. Moreover, the complex can be kept at -80°C for months without forming aggregates, even after several freeze-thaw cycles. Therefore, our results open the way to the determination of the structure of LolA in complex with a lipoprotein. It is indeed an important step to purify a stable high-yielding *in vitro* complex between LolA and a lipoprotein. Now, we are trying different crystallization and purifications conditions to obtain promising crystals. We hope in the future to finally solve the structure of the liganded LolA.

The structure of LolA revealed the importance of the arginine residue at position 43 (R43) at the lid of LolA⁷. Mutating this residue into leucine leads to an increase of the hydrophobic interaction between LolA and the lipoprotein. The transfer of lipoproteins from LolA to LolB occurs in an energy independent manner and is due to the increased

affinity of lipoproteins to LolB⁸. The LolA(R43L) mutant accepts lipoproteins from LolCDE but fails to transfer them to LolB, because of the increased affinity between LolA(R43L) and the acyl chains of the lipoprotein⁹. These data suggest that another way to copurify the lipoprotein in complex with LolA, at high yields, would be to express the R43L mutant of LolA in the periplasm.

Given the bacterial antibiotic resistance crisis that the world is facing, finding new antimicrobials is urgently required. Thus, determining the structure of LolA bound to a lipoprotein is essential, because this structure will help identifying all the key residues important for the interaction of LolA with a substrate.

Previous studies showed two residues that are important for LolA's activity. First, the arginine at position 43 is a key residue for the opening and the closure of LolA's lid ². In fact, after the release of a lipoprotein, the lid of LolA is closed by hydrogen bonds formed between R43 and residues of the $\alpha 1$ and $\alpha 2$ helices of LolA ^{2,10}. Second, substitution of phenylalanine at position 47 by polar residues inhibits the ability of LolA to accept lipoproteins from LolCDE⁷. However, more residues of LolA definitely affect its activity and the movement of its lid. Therefore, these residues influence the binding and release of lipoproteins by LolA. Thus, by solving the structure of a LolA in complex with RcsF, we hope to identify all these critical regions and residues. We expect that this structure will also be useful to design new antimicrobial that can potentially bind the cavity of LolA and block the interaction of this periplasmic chaperone with lipoproteins.

Finally, beside the role of the structural determination in the design of new antimicrobials, solving the structure of LolA in complex with a substrate can answer

the remaining questions regarding the binding of lipoproteins. For instance, it is still not clear how many acyl chains bind the cavity of LolA. According to the structure of LolA, one acyl chain can fit inside the hydrophobic cavity of the periplasmic chaperone². However, simulation studies showed that the cavity can get enlarged and accommodate the three acyl chains ¹¹. Thus, solving the structure of liganded-LolA will definitely answer the question.

Methods

Bacterial growth conditions

Bacterial strains used in this study are listed in **Table 1**. Bacterial cells were cultured in Luria broth (LB) at 37 °C. The ampicillin (200 µg/mL) antibiotic was added, and L-arabinose (0.2%) was used for induction when appropriate.

Bacterial strains and plasmids

DH300 (a derivative of *Escherichia coli* MG1655 carrying a chromosomal *rprA::lacZ* fusion at the λ attachment site¹²) was used as wildtype throughout the study. All deletion mutants were obtained by transferring the corresponding alleles from the Keio collection¹³ (kan^R) into DH300¹² via P1 phage transduction. Deletions were verified by PCR and the absence of the protein was verified *via* immunoblotting (when possible). If necessary, the kanamycin cassette was removed via site-specific recombination mediated by the yeast Flp recombinase with pCP20 vector¹⁴. All strains expressing the RcsF mutant lacked *rcsB* in order to prevent induction of Rcs.

The plasmids and the primers used in this study are listed in **Table 2** and **3**, respectively.

To generate a pBAD18 vector expressing LolA-His and RcsF $_{\Delta 19-47}$ -Strep, cloning was performed as follow. First, *lolA* was amplified using chromosomal *lolA* as PCR template for primers JR30/JR31 (JR31 contains the sequence of a His-tag) and then cloned into pBAD18 using KpnI and XbaI, resulting in pJR48. Chromosomal *rcsF* $_{\Delta 19-47}$ was then amplified by PCR using primers JR53/JR54 (JR54 contains the sequence of a Strep-tag) and then cloned into pJR48, downstream of *lolA-his*, using XbaI and Sall, resulting in pJR127.

Protein purification

Soluble LolA-RcsF $_{\Delta 19-47}$ complex was purified *via* affinity chromatography as follows. Cells co-expressing LolA with RcsF $_{\Delta 19-47}$ (JR147) were grown in LB at 37°C supplemented with 200 µg/mL ampicillin until an OD₆₀₀ of 0.5. Protein expression was then induced with 0.2% arabinose. Cells were pelleted at an OD₆₀₀ of 3 and resuspended in buffer A (200 mM NaCl and 50 mM NaPi at pH 8) containing a tablet of protease inhibitor cocktail. Cells were lysed *via* two passages through a French pressure cell at 1,500 psi. The lysate was centrifuged at 45,000 x g for 30 min at 4 °C using a 55.2 Ti Beckman rotor. To obtain the soluble fraction, the supernatant was centrifuged at 180,000 x g for 1 h at 4 °C using the same rotor. The supernatant was added to a His Trap HP column previously equilibrated with buffer A. The column was washed three times with 10 column volumes of buffer A supplemented with 20 mM imidazole and LolA-His was eluted using the buffer A supplemented with 300 mM of imidazole. The fractions obtained were analyzed *via* SDS-PAGE; LolA was detected around 25 kDa and RcsF $_{\Delta 19-47}$ between 10 and 15 kDa. Elution fractions containing LolA-RcsF $_{\Delta 19-47}$ were mixed with Strep-Tactin resin previously equilibrated with buffer A. After washing the resin with 10 column volumes of buffer A, RcsF $_{\Delta 19-47}$ -Strep was eluted with 5 column volumes of buffer A supplemented with 2.5 mM desthiobiotin. Eluted fractions were then concentrated to 1 mL using a Vivaspin 4 Turbo concentrator and purified *via* size-exclusion chromatography with a Superdex S75-10/300 column.

Table 1. *E. coli* strains used in this study

Strain	Genotype and description	Source
DH300	<i>rprA-lacZ</i> MG1655 (<i>argF-lac</i>) U169	¹²
Keio collection mutants	<i>rcsB::kan</i> , <i>rscF::kan</i>	¹³
XL1-blue	<i>endA1 gyrA96 (nalR) thi-1 recA1 relA1 lac glnV44F' [::Tn10 proAB+ lacIq D(lacZ)M15] hsdR17 (rK- mK+)</i>	Stratagene
NH295	DH300 $\Delta rcsB \Delta rcsF$	This study
JR147	NH295 pJR127	This study

Table 2. Plasmids used in this study

Plasmid	Features	Source
pBAD18	Arabinose inducible P _{BAD} , ampicillin	¹⁵
pJR48	pBAD18 LolA-6xHis	This study
pJR127	pBAD18 LolA-6xHis- RcsF Δ_{19-47} -Strep	This study

Table 3. Primers used in this study

Primer	Sequence 5' to 3'
JR30_Fw_LolA_28_pBAD18	ACATGGTACCCGGGAGTGACGTAATTTGAGGAAT
JR31_Rev_LolA_His_pBAD18	ATGTTCTAGAttaatgatgatgatgatgatgctcgaGCTTACGTTG ATCATCTACCGTGAC
JR53-Fw-RcsF-XbaI	TTGGTCTAGACGTTTTGATCATAATTGAG
JR54-Rv-RcsF-streptagged-Sall	TTGGGTCGACTCACTTTTCGAACTGCGGGTGGCTCC AGCTTGCTTTCCCGTAATGTTAAGCG

References

1. El Arnaout, T. & Soulimane, T. Targeting Lipoprotein Biogenesis: Considerations towards Antimicrobials. *Trends Biochem Sci* **44**, 701-715 (2019).
2. Takeda, K. et al. Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *Embo j* **22**, 3199-209 (2003).
3. Tang, X. et al. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nat Struct Mol Biol* **28**, 347-355 (2021).
4. Na, J.H., Lee, W.K. & Yu, Y.G. How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. *Int J Mol Sci* **19**(2018).
5. Ota, M. et al. An assignment of intrinsically disordered regions of proteins based on NMR structures. *J Struct Biol* **181**, 29-36 (2013).
6. Watanabe, S., Oguchi, Y., Yokota, N. & Tokuda, H. Large-scale preparation of the homogeneous LolA lipoprotein complex and efficient in vitro transfer of lipoproteins to the outer membrane in a LolB-dependent manner. *Protein Sci* **16**, 2741-9 (2007).
7. Taniguchi, N., Matsuyama, S. & Tokuda, H. Mechanisms underlying energy-independent transfer of lipoproteins from LolA to LolB, which have similar unclosed {beta}-barrel structures. *J Biol Chem* **280**, 34481-8 (2005).
8. 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**, 5877-82 (2009).
9. Miyamoto, A., Matsuyama, S. & Tokuda, H. Mutant of LolA, a lipoprotein-specific molecular chaperone of Escherichia coli, defective in the transfer of lipoproteins to LolB. *Biochem Biophys Res Commun* **287**, 1125-8 (2001).
10. Oguchi, Y. et al. Opening and closing of the hydrophobic cavity of LolA coupled to lipoprotein binding and release. *J Biol Chem* **283**, 25414-20 (2008).
11. Boags, A., Samsudin, F. & Khalid, S. Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA. *Sci Rep* **9**, 3717 (2019).
12. Majdalani, N., Hernandez, D. & Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol Microbiol* **46**, 813-26 (2002).
13. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
14. Cherepanov, P.P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Fip-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9-14 (1995).
15. Guzman, L.M., Belin, D., Carson, M.J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**, 4121-30 (1995).

Section III. General Discussion and perspectives

By the time I joined the lab, the journey of lipoproteins in the cell envelope of Gram-negative bacteria had already been intensely studied ²⁶⁷. It was considered that the key factors controlling the sorting of lipoproteins between the two membranes had been identified in *E. coli* and that Lol was the only system responsible for lipoprotein transport between the inner and the outer membranes. However, recent studies have challenged this view and suggested the presence of distinct factors that affect the sorting and transport of lipoproteins. In this section, I will first discuss the results obtained during my thesis, calling for a revision of the model of lipoprotein sorting. Then, I will talk about the different structural studies of the Lol components in complex with a lipoprotein.

Disorder in the journey of lipoproteins

Before I joined to lab, Joanna Szewczyk was already working on the lipoprotein RcsF and tried to reveal the role of its linker, the N-terminal intrinsically disordered region. By deleting the linker of RcsF (residues 19-47), she showed that this unstructured domain is important for the correct OM targeting of RcsF. In fact, around 90% of RcsF_{WT} are present in the OM while 50% of RcsF_{Δ19-47} are retained in the IM (**Fig. 15 and Fig.1 of Chap. I of the Results section**). She was also able to show that the length, but not the sequence, of the linker is important on the condition that this linker stays unfolded. However, this was a side project for Joanna, and she did not have time to finish the story. Intrigued by these exciting results, I took over the project and tried to reveal the role of unstructured linkers in lipoprotein sorting.

The first thing I wanted to know was if the presence and the effect of the linker are specific to RcsF or applicable to other lipoproteins. With the help of Dr. Bogdan I. Iorga,

we showed that half of *E. coli* OM lipoproteins have a long, disordered N-terminal linker (**Fig. 1, Extended Data Fig.1 and Supplementary Table 1 of Chap. I of the Results section**). This unexpected result pushed me to test the effect of linker deletion on additional lipoproteins. I was then able to show that deleting the linker of NlpD, Pal and BamC affects their OM targeting (**Extended Data Fig.2 & 10c of Chap. I of the Results section**). It is quite amazing how different lipoproteins share the presence of an unfolded N-terminus domain that is not conserved, neither in terms of amino acid sequence nor in length. The only common feature is the disorder of the linker. Despite being different, the linkers of RcsF, NlpD, Pal and BamC affect the OM targeting of these lipoproteins which leads to functional consequences.

The importance of the linker in OM targeting, shown during my thesis, fits well with the fact that in some bacteria, outer membrane targeting and surface exposure do not only depend on a sorting signal. For instance, in *Borrelia sp.*, the presence of an N-terminal disordered region is required for lipoproteins to reach the cell surface ²⁶⁸.

The linker functions in a Lol-dependent manner

The most challenging part of my project was to reveal the mechanism used by the linker to affect the OM targeting of lipoproteins. The first question that came to my mind was if the linker works in a Lol dependent manner or not.

To answer this question, I expressed an RcsF variant with two aspartates at positions +2 +3 (RcsF_{CDD}), as the aspartate is a Lol avoidance residue ²⁶⁹. Although the linker is present in RcsF_{CDD}, this RcsF variant is highly retained in the IM of *E. coli* (**Extended Data Fig.3 of Chap. I of the Results section**). This result clearly

shows that, despite having an N-terminal linker, the sorting of RcsF still follows the +2 rule of Lol.

Based on this result, I then questioned if the linker affects the interaction between the lipoprotein and the components of the Lol system. However, RcsF_{WT} and RcsF_{Δ19-47}, both co-purifies with the periplasmic chaperone LolA (**Extended Data Fig.6 of Chap. I of the Results section**) and can be transferred *in vitro* to the soluble variant of LolB (**Fig.4b of Chap. I of the Results section**). Moreover, RcsF_{Δ19-47} is still retained in the IM even when the LolCDE complex or one of the three maturation enzymes (Lgt, LspA and Lnt) is overexpressed (**Extended Data Fig.7 & 8 of Chap. I of the Results section**). Altogether, these results indicate that the absence of the linker does not prevent the interaction of RcsF with LolA, LolB, LolCDE or the maturation enzymes. However, one thing is sure; the linker is important for efficient OM targeting of RcsF, NlpD, Pal and BamC.

Based on the results I obtained, I showed that the linker is required for correct OM lipoprotein targeting in a Lol dependent manner.

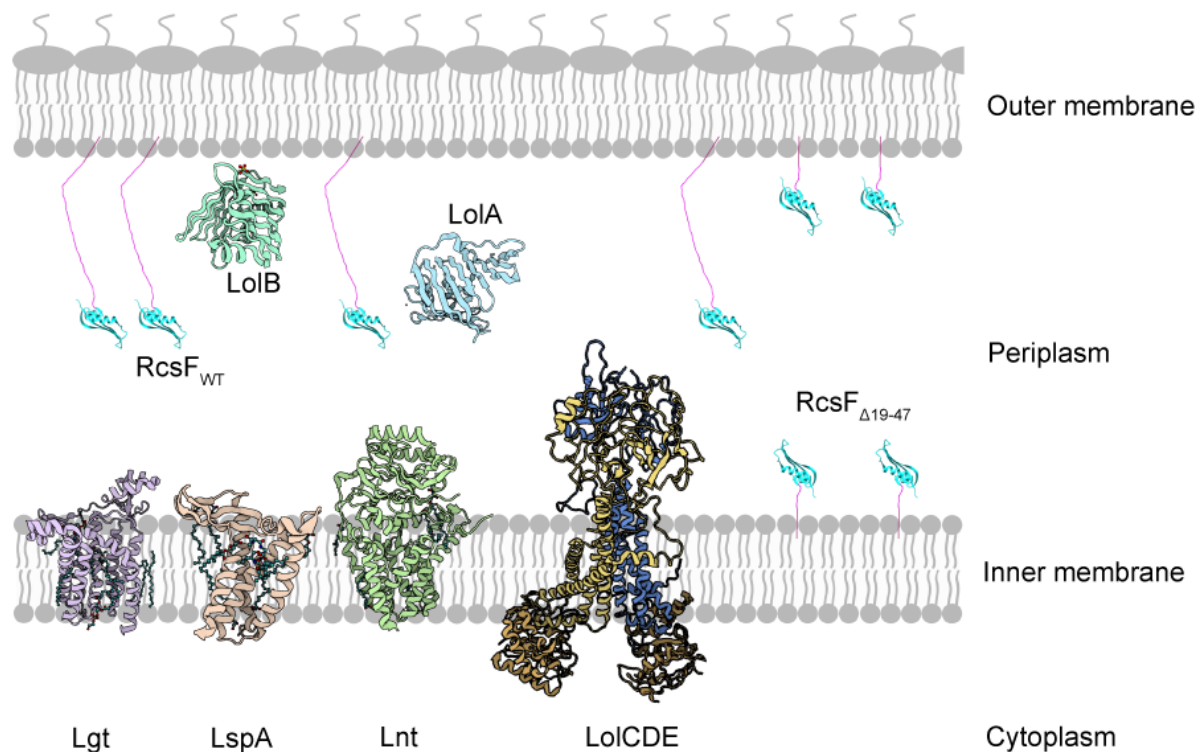


Figure 15. Recapitulative model for the importance of the linker in the sorting of lipoproteins in the cell envelope of *E. coli*. The full-length RcsF, RcsF_{WT}, is correctly targeted to the OM, while the RcsF variant devoid of its linker, RcsF_{Δ19-47}, is partially retained in the IM.

The linker ensures the efficiency of lipoprotein biogenesis

Lpp is present at ~1 million copies per cell¹⁸⁵, and its OM targeting is a must as the accumulation of Lpp in the IM is toxic for the cell²⁰⁶. Therefore, I thought that Lol could be giving the privilege for OM targeting of Lpp at the expense of other lipoproteins. To test this, I checked the sorting of lipoproteins devoid of their linker in Δlpp cells. Surprisingly, deleting *lpp* rescues the OM targeting of the linker-less lipoproteins, RcsF_{Δ19-47} and NlpD_{Δ29-64} (**Fig.4a of Chap. I of the Results section**). This result indicates that the accumulation of lipoproteins devoid of their linker in the inner membrane comes from a decreased ability of Lol to process these variants of lipoproteins. It also

means that, in *E. coli*, the role of the linker is to optimize the processing of lipoproteins in the cell envelope in the presence of the highly abundant lipoprotein Lpp.

Why do some lipoproteins need an N-terminal linker?

After showing that the linker is important for efficient OM targeting of RcsF, I was intrigued by the presence of the linker at the N-terminus of some OM lipoproteins, not all of them. The evident question to me was why do some lipoproteins need a linker? What is the common feature between these lipoproteins?

While I was trying to find what was in common between the lipoproteins that have a linker, I noticed that RcsF^{142,143}, Pal¹⁴¹ and BamC^{147,148} are all surface-exposed. Therefore, I wondered if the linker of these lipoproteins is important for their surface exposure, and for Lol to recognize surface-exposed lipoproteins. I then showed that NlpD is also exposed on the cell surface of *E. coli* (**Extended Data Fig.10 of Chap. I of the Results section**). The fact that the four lipoproteins are surface-exposed and have a disordered linker whose deletion retains the lipoprotein in the IM suggests that the N-terminus linker might be important for the Lol to recognize surface-exposed lipoproteins and to optimize their targeting to the OM. This idea of having a linker that is important for the surface exposure of lipoproteins is not new. In fact, in *Borrelia sp.*, the presence of a disordered N-terminal sequence is required for the lipoproteins to reach the surface^{108,268}.

More experiments need to be done to validate the relationship between the presence of N-terminal linkers and surface exposure. One key experiment would be to detect the surface exposure of lipoproteins devoid of their linkers such as RcsF_{Δ19-47},

NlpD Δ_{29-64} , Pal Δ_{26-56} and BamC Δ_{28-98} . Another experiment that could be done would be to add the N-terminal linker of RcsF_{WT} to an outer membrane lipoprotein that is not on the cell surface, and to evaluate if this linker addition would send the tested lipoprotein to the cell surface.

Finally, little is known about how lipoproteins reach the cell surface. Is it possible that the linker facilitates the interaction with possible transporters like BamA, OMPs and SLAMs? Could it be that LolB interacts with these transporters, and that the linker of the lipoprotein facilitates this interaction? Is it possible that unidentified flippases exist in the OM and are responsible of the surface exposure of lipoproteins? Could the disordered linker be important for lipoproteins to be recognized by these flippases and to interact with them? It is clear that many questions remain to be answered. Hopefully, future studies will clarify the mechanistic details about the whole lipoprotein journey in the cell envelope.

To sum up, the data obtained during my thesis shows that, in *E. coli*, a disordered region at the N-terminus of half of OM lipoproteins is important for efficient OM targeting and calls for a revision of the lipoprotein sorting model. Although I could not specify at which step the linker interfere with the journey of the lipoproteins in the cell envelope, and how, the results obtained are of a great interest as they show that novel molecular determinants, the linker for example, exist in the cell in addition to the Lol system, and that the current understanding of lipoprotein transport by Lol is incomplete.

Where could the N-terminal linker be interfering?

As discussed above, during my thesis, I showed the importance of the linker in ensuring an efficient transport of the lipoproteins in the cell envelope of *E. coli*. However, I could not determine at which step, the linker is affecting the lipoprotein's pathway. Despite the difficulty in showing this, we can speculate by proposing different hypotheses.

First, as we proposed in the paper (*Chapter I of the results section*), the linker could be important to optimize the different steps in the journey of lipoproteins. Since these steps are successive and dependent on each other, we can imagine the pathway of lipoproteins as a corridor with several doors parallel to each other (**Fig. 16a**). You cannot arrive to the end of this corridor unless you open and pass all the doors. If you force the opening of one door, this will not be enough to arrive to the end. The same model applies here; the overexpression of one of the different maturation enzymes and the Lol components was not enough to observe an enhancement in the OM targeting of RcsF $_{\Delta 19-47}$. It might have been the case if we could overexpress all the proteins (Lgt, Lsp, Lnt, LolCDE, LolA and LolB), which is technically quite impossible.

Another hypothesis where the linker could be interfering is in the interaction between the lipoprotein and LolCDE. Recently, the cryo-EM structure of LolCDE complex was solved^{115,116} and shows the first ten N-terminal residues of the lipoprotein, potentially Lpp, inside of the central channel of LolCE, with the acyl chains pointing down and the ten residues pointing up (**Fig.8b, c**). In this model, the authors propose that the globular domain of the lipoprotein interacts with the periplasmic domain of LolE (**Fig.8c**). However, the *in vivo* and *in vitro* crosslinking between LolE and Pal shows

that the N-terminal residues of Pal, at positions 4 and 5, interact with LolE, while no crosslinking is detected between LolE and residues in the globular domain of Pal, at positions 100 and 147¹⁰¹. Moreover, LolCDE is important for the IM extraction of different lipoproteins. However, lipoproteins have different globular domain and so it is hard to imagine that LolE can interact with the globular domain of the different OM lipoproteins of *E. coli*. Thus, having a disordered linker at the N-terminus of the lipoprotein might facilitate the interaction of the lipoprotein with LolCDE, by pushing the globular domain of the lipoprotein away from LolE (**Fig. 16b**). This would prevent the steric hindrance between the globular domain of the lipoprotein and the periplasmic domain of LolE, thus accelerating the extraction of the lipoprotein from the IM. We could then imagine that the linker is interacting with LolE and that this interaction is mediated by hydrogen bonds, polar and electrostatic interactions.

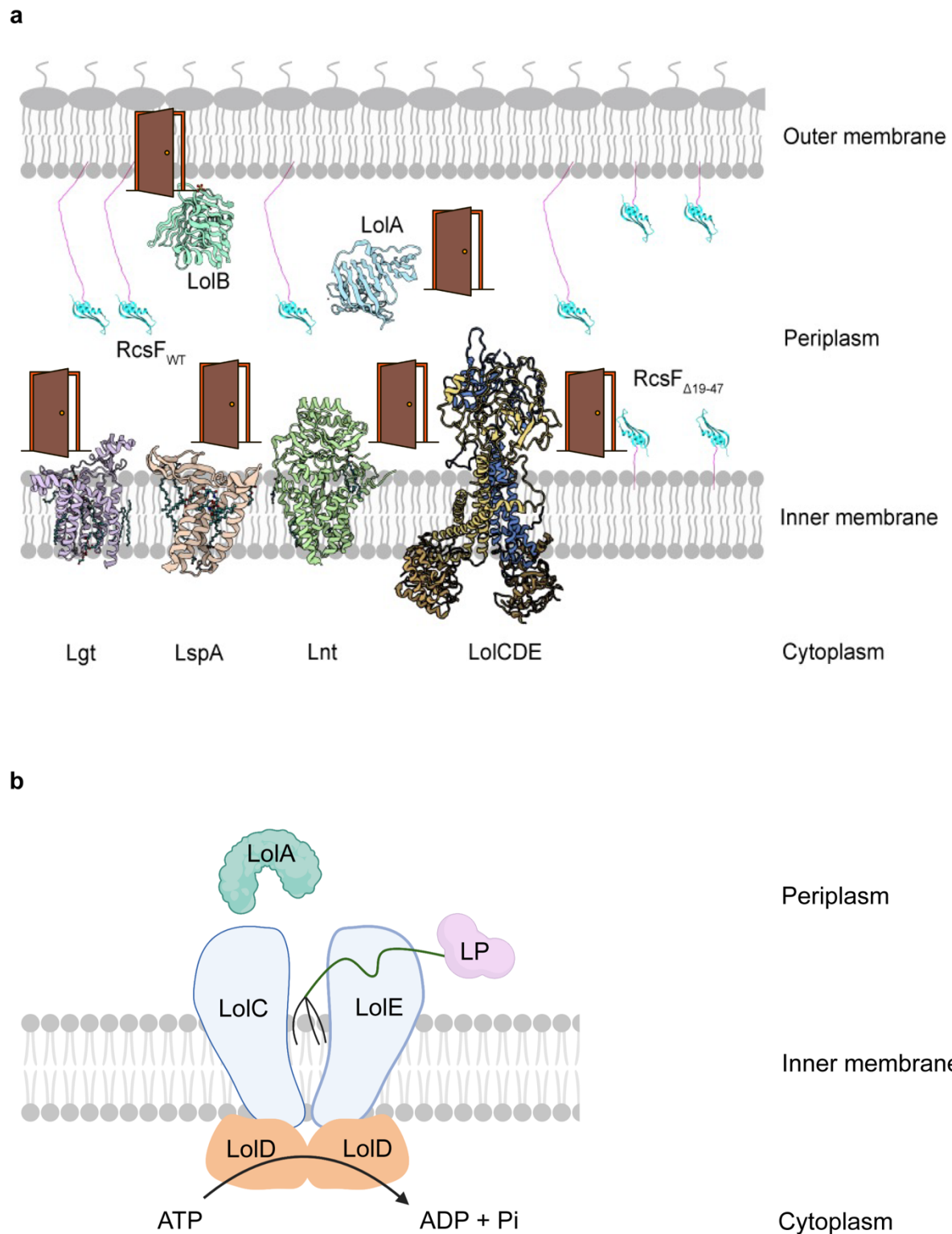


Figure 16. Proposed models where the N-terminal disordered linker might be interfering with the lipoprotein's pathway. a. The linker might be important for an efficient interaction with each component of the lipoprotein pathway. **b.** The linker

could facilitate the interaction with LolCDE by avoiding the steric hindrance between LolE and the lipoprotein. LP: Lipoprotein.

What about other bacteria?

At the end, it is important to mention that the results I got during my thesis were obtained in *E. coli*. It will be very interesting to see if N-terminal disorder exists in the lipoproteins of other Gram-negative bacteria, and if it also impacts OM targeting of lipoproteins, the way it does in *E. coli*. During my thesis, I showed that the linker is important for an efficient transport of lipoproteins, potentially the surface-exposed ones, in the presence of the abundant Lpp, in *E. coli*. A bacterium in which disorder has been shown to play a role is *Borrelia*. In fact, a disordered N-terminal region is essential for lipoproteins to reach the surface of *Borrelia*^{109,268} where most of these lipoproteins act as virulence factors^{270,271}. Thus, reaching the surface seems to be essential for the survival of *Borrelia*. However, the details of the mechanism are unknown.

Light at the end of the tunnel: towards the structural determination of the Lol components liganded to a lipoprotein

In the last few years, many research groups doubled the effort to determine the structure of a Lol component in complex with a lipoprotein. Although recent studies showed promising results ^{114,115}, the resolution of such structure has not been completed yet.

The recent cryo-EM structure of *E. coli* LolCDE was obtained in complex with a triacyl-cysteine and a peptide of ten amino acids of a lipoprotein, potentially Lpp ¹¹⁵. The decapeptide of the lipoprotein appears to be bound in the V-shape central channel of LolCDE above the IM (**Fig. 8**). The authors propose a model of lipoprotein extraction by LolCDE where the globular domain of the lipoprotein interacts with the periplasmic domain of LolE (**See section 2.2.a. of chapter II of the Introduction for more details about the extraction model**). Although the globular domain of the lipoprotein was not detected in the structure, this study is of a great interest, as it opens the way toward the determination of complete mechanistic details on lipoprotein extraction by LolCDE. However, many questions remain unanswered. Is it possible that all lipoproteins have the same interaction mechanism with LolCDE? Where are the exact interaction regions between LolE and the lipoprotein? Is the lipoprotein well folded at this stage? Future investigations are definitely required.

During my thesis, I managed to purify a stable periplasmic complex between the chaperone LolA and a lipoprotein, in this case RcsF_{Δ19-47} (**Fig.1 of Chap. II of the Results section**). For some, this might seem familiar as any other protein purification. However, this result is an important step in the lipoprotein field, as the purification of

an *in vitro* large-scale liganded LolA has always been very challenging ²⁷². Moreover, LolA-lipoproteins complexes are usually obtained after their release from spheroplasts and addition of previously purified LolA in the medium ²⁷³. Here, we show that a stable *in vitro* periplasmic complex between LolA and RcsF_{Δ19-47} can be obtained after three purification steps (**Chap II of the Results section**). We hope in the future to be able to determine the crystal structure of the LolA- RcsF_{Δ19-47} complex. The resolution of this structure will definitely answer many remaining questions. So far, the molecular interaction details between LolA and the lipid moiety of a lipoprotein are still ambiguous. In fact, it is still unknown whether LolA can bind one ¹¹⁷ or the three ¹²² acyl chains that are attached to the cysteine of a lipoprotein. Moreover, the structure of LolA of *P. aeruginosa* presents hydrophobic surface patches that were shown to be important for the binding of the acyl chains of lipoproteins ²⁷⁴. Thus, resolving the structure of a LolA-lipoprotein complex might help in understanding the exact mechanism of interaction between the periplasmic chaperone and lipoproteins. Hence, the purification results obtained during my thesis, and discussed in chapter II of the Results section, are a first step that might lead us to the bottom of this mystery. Finally, let's not forget one last important structure to be solved: LolB-lipoprotein. The determination of the structure of LolB in complex with a lipoprotein can give us more insights into the mechanism used by LolB to anchor lipoproteins in the OM.

In the end, despite all the remaining questions in the lipoprotein field, hopefully future studies will elucidate the molecular details regarding the sorting, the transport and the surface exposure of lipoproteins. Defining the exact mechanisms of interaction between a lipoprotein and Lol would help in the design of new antimicrobials that block

these interactions. Such discoveries are essential in fighting the antibiotic resistance crisis.

Section IV. Conclusion

By the time I joined the Collet lab, lipoproteins in the cell envelope of *E. coli* were already under the spotlight. Different mechanistic aspects of the stress sensor, RcsF, were being studied by many lab members. The importance of the intrinsically disordered region in lipoprotein sorting was a side project for Joanna Szewczyk. I therefore took over the project and tried to finish the story.

To sum up, lipoproteins are proteins covalently linked to a lipid moiety that anchors them to the IM or the OM. They consist of an essential group of proteins important for the cell viability, as they fulfil a variety of different functions in the cell envelope. During my PhD thesis, I was able to clarify the importance of the intrinsically disordered linker present at the N-terminal of half OM lipoproteins; half of the OM lipoproteins have a long linker that consists of more than 22 residues. I showed that the presence of this linker is crucial for the lipoprotein to efficiently reach the OM by the Lol machinery. The linker-less mutant of a lipoprotein is retained in the IM and partially reaches the OM. Moreover, the length and the disorder character of the linker are important for the correct OM targeting of the lipoprotein. I was also able to show that lipoproteins, with or without the linker, use the Lol system to travel throughout the cell envelope. However, it seems that the presence of the most abundant lipoprotein, Lpp, has a significant role in the story. In fact, deleting *lpp* restores the OM targeting of linker-less lipoproteins. This result shows that the disorder linker optimizes the processing of lipoproteins by the Lol system. Finally, we proposed that the linker could be used by OM lipoproteins, such as RcsF, BamC, NlpD and Pal, to reach the cell surface.

Finally, I am very happy that I could participate in a better understanding of the lipoprotein pathway, and to bring new insights into this field.

At the end of this journey, I cannot help but express my gratitude for the Collet lab. I am very happy with my PhD adventure and the scientific person I became. I learned patience, resilience and persistence. The road was not easy at all, but it was worth the hard work, and as Erol Ozan once said: “Some beautiful paths cannot be discovered without getting lost”.

Bibliography

1. Silhavy, T.J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb Perspect Biol* **2**, a000414 (2010).
2. Weiner, J.H. & Li, L. Proteome of the Escherichia coli envelope and technological challenges in membrane proteome analysis. *Biochim Biophys Acta* **1778**, 1698-713 (2008).
3. Ricci, D.P. & Silhavy, T.J. Outer Membrane Protein Insertion by the β -barrel Assembly Machine. *EcoSal Plus* **8**, 0035-2018 (2019).
4. Braun, V. & Hantke, K. Lipoproteins: Structure, Function, Biosynthesis. *Subcell Biochem* **92**, 39-77 (2019).
5. Rosadini, C.V. & Kagan, J.C. Early innate immune responses to bacterial LPS. *Curr Opin Immunol* **44**, 14-19 (2017).
6. Raetz, C.R. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635-700 (2002).
7. Bertani, B. & Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **8**(2018).
8. Stevenson, G. et al. Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster. *J Bacteriol* **176**, 4144-56 (1994).
9. Meredith, T.C. et al. Modification of lipopolysaccharide with colanic acid (M-antigen) repeats in Escherichia coli. *J Biol Chem* **282**, 7790-8 (2007).
10. Wilson, A. & Ruiz, N. Transport of lipopolysaccharides and phospholipids to the outer membrane. *Curr Opin Microbiol* **60**, 51-57 (2021).
11. Miller, S.I. & Salama, N.R. The gram-negative bacterial periplasm: Size matters. *PLoS Biol* **16**, e2004935 (2018).
12. Arts, I.S., Gennaris, A. & Collet, J.F. Reducing systems protecting the bacterial cell envelope from oxidative damage. *FEBS Lett* **589**, 1559-1568 (2015).
13. Goemans, C., Denoncin, K. & Collet, J.F. Folding mechanisms of periplasmic proteins. *Biochim Biophys Acta* **1843**, 1517-28 (2014).
14. Vollmer, W. & Holtje, J.V. The architecture of the murein (peptidoglycan) in gram-negative bacteria: vertical scaffold or horizontal layer(s)? *J Bacteriol* **186**, 5978-87 (2004).
15. Egan, A.J.F., Errington, J. & Vollmer, W. Regulation of peptidoglycan synthesis and remodelling. *Nat Rev Microbiol* **18**, 446-460 (2020).
16. Braun, V. Covalent lipoprotein from the outer membrane of Escherichia coli. *Biochim Biophys Acta* **415**, 335-77 (1975).
17. Parsons, L.M., Lin, F. & Orban, J. Peptidoglycan recognition by Pal, an outer membrane lipoprotein. *Biochemistry* **45**, 2122-8 (2006).
18. Lazzaroni, J.C. & Portalier, R. The excC gene of Escherichia coli K-12 required for cell envelope integrity encodes the peptidoglycan-associated lipoprotein (PAL). *Mol Microbiol* **6**, 735-42 (1992).
19. Luirink, J., Yu, Z., Wagner, S. & de Gier, J.W. Biogenesis of inner membrane proteins in Escherichia coli. *Biochim Biophys Acta* **1817**, 965-76 (2012).
20. Natale, P., Brüser, T. & Driessen, A.J. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. *Biochim Biophys Acta* **1778**, 1735-56 (2008).
21. Tsirigotaki, A., De Geyter, J., Šoštaric, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat Rev Microbiol* **15**, 21-36 (2017).
22. Paetzel, M. Bacterial Signal Peptidases. *Subcell Biochem* **92**, 187-219 (2019).

23. Steinberg, R., Knüpfner, L., Origi, A., Asti, R. & Koch, H.G. Co-translational protein targeting in bacteria. *FEMS Microbiol Lett* **365**(2018).
24. Frain, K.M., Robinson, C. & van Dijl, J.M. Transport of Folded Proteins by the Tat System. *Protein J* **38**, 377-388 (2019).
25. Palmer, T. & Stansfeld, P.J. Targeting of proteins to the twin-arginine translocation pathway. *Mol Microbiol* **113**, 861-871 (2020).
26. Cline, K. Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat). *J Biol Chem* **290**, 16530-8 (2015).
27. Lycklama, A.N.J.A. & Driessen, A.J. The bacterial Sec-translocase: structure and mechanism. *Philos Trans R Soc Lond B Biol Sci* **367**, 1016-28 (2012).
28. Amaral, L., Martins, A., Spengler, G. & Molnar, J. Efflux pumps of Gram-negative bacteria: what they do, how they do it, with what and how to deal with them. *Front Pharmacol* **4**, 168 (2014).
29. Costa, T.R. et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**, 343-59 (2015).
30. Tomasek, D. & Kahne, D. The assembly of β -barrel outer membrane proteins. *Curr Opin Microbiol* **60**, 16-23 (2021).
31. Masi, M., Winterhalter, M. & Pagès, J.M. Outer Membrane Porins. *Subcell Biochem* **92**, 79-123 (2019).
32. Baslé, A., Rummel, G., Storici, P., Rosenbusch, J.P. & Schirmer, T. Crystal structure of osmoporin OmpC from E. coli at 2.0 Å. *J Mol Biol* **362**, 933-42 (2006).
33. Yamashita, E., Zhalnina, M.V., Zakharov, S.D., Sharma, O. & Cramer, W.A. Crystal structures of the OmpF porin: function in a colicin translocon. *Embo j* **27**, 2171-80 (2008).
34. Pautsch, A. & Schulz, G.E. Structure of the outer membrane protein A transmembrane domain. *Nat Struct Biol* **5**, 1013-7 (1998).
35. Koebnik, R. Proposal for a peptidoglycan-associating alpha-helical motif in the C-terminal regions of some bacterial cell-surface proteins. *Mol Microbiol* **16**, 1269-70 (1995).
36. Samsudin, F., Boags, A., Piggot, T.J. & Khalid, S. Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. *Biophys J* **113**, 1496-1504 (2017).
37. Stathopoulos, C. An alternative topological model for Escherichia coli OmpA. *Protein Sci* **5**, 170-3 (1996).
38. Sugawara, E. & Nikaido, H. Pore-forming activity of OmpA protein of Escherichia coli. *J Biol Chem* **267**, 2507-11 (1992).
39. Sugawara, E. & Nikaido, H. OmpA protein of Escherichia coli outer membrane occurs in open and closed channel forms. *J Biol Chem* **269**, 17981-7 (1994).
40. Tokuda, H. et al. Chapter 4: Bacterial lipoproteins: biogenesis, virulence/pathogenicity and trafficking. in *Bacterial membranes: Structural and molecular biology*. (pp. 133-178) (eds. Remaut, H. & Fronzes, R.) (Caister Academic Press, Norfolk, UK, 2013).
41. Qin, A. et al. FipB, an essential virulence factor of Francisella tularensis subsp. tularensis, has dual roles in disulfide bond formation. *J Bacteriol* **196**, 3571-81 (2014).
42. Liu, J. et al. Outer Membrane Lipoprotein Lip40 Modulates Adherence, Colonization, and Virulence of Actinobacillus pleuropneumoniae. *Front Microbiol* **9**, 1472 (2018).
43. Andersson, J.A. et al. Identification of New Virulence Factors and Vaccine Candidates for Yersinia pestis. *Front Cell Infect Microbiol* **7**, 448 (2017).

44. Seib, K.L., Scarselli, M., Comanducci, M., Toneatto, D. & Masiagnani, V. Neisseria meningitidis factor H-binding protein fHbp: a key virulence factor and vaccine antigen. *Expert Rev Vaccines* **14**, 841-59 (2015).
45. Moens, S. & Vanderleyden, J. Functions of bacterial flagella. *Crit Rev Microbiol* **22**, 67-100 (1996).
46. Lukaszczyk, M., Pradhan, B. & Remaut, H. The Biosynthesis and Structures of Bacterial Pili. *Subcell Biochem* **92**, 369-413 (2019).
47. Delhay, A., Collet, J.F. & Laloux, G. A Fly on the Wall: How Stress Response Systems Can Sense and Respond to Damage to Peptidoglycan. *Front Cell Infect Microbiol* **9**, 380 (2019).
48. Mitchell, A.M. & Silhavy, T.J. Envelope stress responses: balancing damage repair and toxicity. *Nat Rev Microbiol* **17**, 417-428 (2019).
49. Hayashi, S. & Wu, H.C. Lipoproteins in bacteria. *J Bioenerg Biomembr* **22**, 451-71 (1990).
50. Babu, M.M. et al. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol* **188**, 2761-73 (2006).
51. Crane, J.M. & Randall, L.L. The Sec System: Protein Export in Escherichia coli. *EcoSal Plus* **7**(2017).
52. 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**, 2511-6 (1992).
53. Watanabe, T., Hayashi, S. & Wu, H.C. Synthesis and export of the outer membrane lipoprotein in Escherichia coli mutants defective in generalized protein export. *J Bacteriol* **170**, 4001-7 (1988).
54. Froderberg, L., Houben, E.N., Baars, L., Lührink, J. & de Gier, J.W. Targeting and translocation of two lipoproteins in Escherichia coli via the SRP/Sec/YidC pathway. *J Biol Chem* **279**, 31026-32 (2004).
55. Tsukazaki, T. Structural Basis of the Sec Translocon and YidC Revealed Through X-ray Crystallography. *Protein J* **38**, 249-261 (2019).
56. Palmer, T. & Berks, B.C. The twin-arginine translocation (Tat) protein export pathway. *Nat Rev Microbiol* **10**, 483-96 (2012).
57. Shruthi, H., Babu, M.M. & Sankaran, K. TAT-pathway-dependent lipoproteins as a niche-based adaptation in prokaryotes. *J Mol Evol* **70**, 359-70 (2010).
58. Shruthi, H., Anand, P., Murugan, V. & Sankaran, K. Twin arginine translocase pathway and fast-folding lipoprotein biosynthesis in E. coli: interesting implications and applications. *Mol Biosyst* **6**, 999-1007 (2010).
59. Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, 239-59 (2011).
60. Kovacs-Simon, A., Titball, R.W. & Michell, S.L. Lipoproteins of bacterial pathogens. *Infect Immun* **79**, 548-61 (2011).
61. Buddelmeijer, N. The molecular mechanism of bacterial lipoprotein modification--how, when and why? *FEMS Microbiol Rev* **39**, 246-61 (2015).
62. Qi, H.Y., Sankaran, K., Gan, K. & Wu, H.C. Structure-function relationship of bacterial prolipoprotein diacylglyceryl transferase: functionally significant conserved regions. *J Bacteriol* **177**, 6820-4 (1995).
63. Mao, G. et al. Crystal structure of E. coli lipoprotein diacylglyceryl transferase. *Nat Commun* **7**, 10198 (2016).

64. Lai, S.H., Philbrick, W.M. & Wu, H.C. Acyl moieties in phospholipids are the precursors for the fatty acids in murein lipoprotein of *Escherichia coli*. *J Biol Chem* **255**, 5384-7 (1980).
65. Sankaran, K. & Wu, H.C. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J Biol Chem* **269**, 19701-6 (1994).
66. Tokunaga, M., Tokunaga, H. & Wu, H.C. Post-translational modification and processing of *Escherichia coli* prolipoprotein in vitro. *Proc Natl Acad Sci U S A* **79**, 2255-9 (1982).
67. Pailler, J., Aucher, W., Pires, M. & Buddelmeijer, N. Phosphatidylglycerol::prolipoprotein diacylglyceryl transferase (Lgt) of *Escherichia coli* has seven transmembrane segments, and its essential residues are embedded in the membrane. *J Bacteriol* **194**, 2142-51 (2012).
68. Singh, W. et al. Mechanism of Phosphatidylglycerol Activation Catalyzed by Prolipoprotein Diacylglyceryl Transferase. *J Phys Chem B* **123**, 7092-7102 (2019).
69. Sangith, N., Kumar, S. & Sankaran, K. Evidence to Suggest Bacterial Lipoprotein Diacylglyceryl Transferase (Lgt) is a Weakly Associated Inner Membrane Protein. *J Membr Biol* **252**, 563-575 (2019).
70. Vogeley, L. et al. Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin. *Science* **351**, 876-80 (2016).
71. Olatunji, S. et al. Structures of lipoprotein signal peptidase II from *Staphylococcus aureus* complexed with antibiotics globomycin and myxovirescin. *Nat Commun* **11**, 140 (2020).
72. Inukai, M., Nakajima, M., Osawa, M., Haneishi, T. & Arai, M. Globomycin, a new peptide antibiotic with spheroplast-forming activity. II. Isolation and physico-chemical and biological characterization. *J Antibiot (Tokyo)* **31**, 421-5 (1978).
73. Hussain, M., Ichihara, S. & Mizushima, S. Accumulation of glyceride-containing precursor of the outer membrane lipoprotein in the cytoplasmic membrane of *Escherichia coli* treated with globomycin. *J Biol Chem* **255**, 3707-12 (1980).
74. Dev, I.K., Harvey, R.J. & Ray, P.H. Inhibition of prolipoprotein signal peptidase by globomycin. *J Biol Chem* **260**, 5891-4 (1985).
75. Xiao, Y., Gerth, K., Müller, R. & Wall, D. Myxobacterium-produced antibiotic TA (myxovirescin) inhibits type II signal peptidase. *Antimicrob Agents Chemother* **56**, 2014-21 (2012).
76. Sangith, N. & Sankaran, K. Stringency of bacterial prolipoprotein signal peptidase (LspA) in recognition of signal peptides - Structure-function correlation. *Biochem Biophys Res Commun* **488**, 413-417 (2017).
77. Pace, H.C. & Brenner, C. The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**, Reviews0001 (2001).
78. Noland, C.L. et al. Structural insights into lipoprotein N-acylation by *Escherichia coli* apolipoprotein N-acyltransferase. *Proc Natl Acad Sci U S A* **114**, E6044-e6053 (2017).
79. Wiktor, M. et al. Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis. *Nat Commun* **8**, 15952 (2017).
80. Wiseman, B. & Högbom, M. Conformational changes in Apolipoprotein N-acyltransferase (Lnt). *Sci Rep* **10**, 639 (2020).
81. Lu, G. et al. Crystal structure of *E. coli* apolipoprotein N-acyl transferase. *Nat Commun* **8**, 15948 (2017).

82. Vidal-Ingigliardi, D., Lewenza, S. & Buddelmeijer, N. Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family. *J Bacteriol* **189**, 4456-64 (2007).
83. Hillmann, F., Argentini, M. & Buddelmeijer, N. Kinetics and phospholipid specificity of apolipoprotein N-acyltransferase. *J Biol Chem* **286**, 27936-46 (2011).
84. Cheng, W., Doyle, D.A. & El Arnaout, T. The N-acyltransferase Lnt: Structure-function insights from recent simultaneous studies. *Int J Biol Macromol* **117**, 870-877 (2018).
85. Buddelmeijer, N. & Young, R. The essential Escherichia coli apolipoprotein N-acyltransferase (Lnt) exists as an extracytoplasmic thioester acyl-enzyme intermediate. *Biochemistry* **49**, 341-6 (2010).
86. El Arnaout, T. & Soulimane, T. Targeting Lipoprotein Biogenesis: Considerations towards Antimicrobials. *Trends Biochem Sci* **44**, 701-715 (2019).
87. Legood, S., Boneca, I.G. & Buddelmeijer, N. Mode of action of lipoprotein modification enzymes-Novel antibacterial targets. *Mol Microbiol* (2020).
88. Nozeret, K., Boucharlat, A., Agou, F. & Buddelmeijer, N. A sensitive fluorescence-based assay to monitor enzymatic activity of the essential integral membrane protein Apolipoprotein N-acyltransferase (Lnt). *Sci Rep* **9**, 15978 (2019).
89. Kitamura, S., Owensby, A., Wall, D. & Wolan, D.W. Lipoprotein Signal Peptidase Inhibitors with Antibiotic Properties Identified through Design of a Robust In Vitro HT Platform. *Cell Chem Biol* **25**, 301-308.e12 (2018).
90. Horler, R.S., Butcher, A., Papangelopoulos, N., Ashton, P.D. & Thomas, G.H. EchoLOCATION: an in silico analysis of the subcellular locations of Escherichia coli proteins and comparison with experimentally derived locations. *Bioinformatics* **25**, 163-6 (2009).
91. Tokuda, H. Biogenesis of outer membranes in Gram-negative bacteria. *Biosci Biotechnol Biochem* **73**, 465-73 (2009).
92. Szewczyk, J. & Collet, J.F. The Journey of Lipoproteins Through the Cell: One Birthplace, Multiple Destinations. *Adv Microb Physiol* **69**, 1-50 (2016).
93. Ito, Y., Kanamaru, K., Taniguchi, N., Miyamoto, S. & Tokuda, H. A novel ligand bound ABC transporter, LolCDE, provides insights into the molecular mechanisms underlying membrane detachment of bacterial lipoproteins. *Mol Microbiol* **62**, 1064-75 (2006).
94. Taniguchi, N., Matsuyama, S. & Tokuda, H. Mechanisms underlying energy-independent transfer of lipoproteins from LolA to LolB, which have similar unclosed {beta}-barrel structures. *J Biol Chem* **280**, 34481-8 (2005).
95. Karow, M. & Georgopoulos, C. The essential Escherichia coli msbA gene, a multicopy suppressor of null mutations in the htrB gene, is related to the universally conserved family of ATP-dependent translocators. *Mol Microbiol* **7**, 69-79 (1993).
96. Ruiz, N., Gronenberg, L.S., Kahne, D. & Silhavy, T.J. 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**, 5537-42 (2008).
97. Narita, S., Tanaka, K., Matsuyama, S. & Tokuda, H. 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**, 1417-22 (2002).
98. Rees, D.C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat Rev Mol Cell Biol* **10**, 218-27 (2009).
99. Thomas, C. & Tampé, R. Structural and Mechanistic Principles of ABC Transporters. *Annu Rev Biochem* **89**, 605-636 (2020).

100. Yakushi, T., Masuda, K., Narita, S., Matsuyama, S. & Tokuda, H. A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nat Cell Biol* **2**, 212-8 (2000).
101. Mizutani, M. et al. Functional differentiation of structurally similar membrane subunits of the ABC transporter LolCDE complex. *FEBS Lett* **587**, 23-9 (2013).
102. Kaplan, E., Greene, N.P., Crow, A. & Koronakis, V. Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain. *Proc Natl Acad Sci U S A* **115**, E7389-e7397 (2018).
103. Terada, M., Kuroda, T., Matsuyama, S.I. & Tokuda, H. Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *J Biol Chem* **276**, 47690-4 (2001).
104. Gennity, J.M. & Inouye, M. The protein sequence responsible for lipoprotein membrane localization in Escherichia coli exhibits remarkable specificity. *J Biol Chem* **266**, 16458-64 (1991).
105. Hara, T., Matsuyama, S. & Tokuda, H. Mechanism underlying the inner membrane retention of Escherichia coli lipoproteins caused by Lol avoidance signals. *J Biol Chem* **278**, 40408-14 (2003).
106. Lewenza, S., Mhlanga, M.M. & Pugsley, A.P. Novel inner membrane retention signals in Pseudomonas aeruginosa lipoproteins. *J Bacteriol* **190**, 6119-25 (2008).
107. Narita, S. & Tokuda, H. Amino acids at positions 3 and 4 determine the membrane specificity of Pseudomonas aeruginosa lipoproteins. *J Biol Chem* **282**, 13372-8 (2007).
108. Schulze, R.J. & Zuckert, W.R. Borrelia burgdorferi lipoproteins are secreted to the outer surface by default. *Mol Microbiol* **59**, 1473-84 (2006).
109. Schulze, R.J., Chen, S., Kumru, O.S. & Zuckert, W.R. 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**, 1266-78 (2010).
110. Lorenz, C., Dougherty, T.J. & Lory, S. Correct Sorting of Lipoproteins into the Inner and Outer Membranes of Pseudomonas aeruginosa by the Escherichia coli LolCDE Transport System. *mBio* **10**, 00194-19 (2019).
111. Miyamoto, S. & Tokuda, H. Diverse effects of phospholipids on lipoprotein sorting and ATP hydrolysis by the ABC transporter LolCDE complex. *Biochim Biophys Acta* **1768**, 1848-54 (2007).
112. 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**, 5877-82 (2009).
113. Crow, A., Greene, N.P., Kaplan, E. & Koronakis, V. Structure and mechanotransmission mechanism of the MacB ABC transporter superfamily. *Proc Natl Acad Sci U S A* **114**, 12572-12577 (2017).
114. Rao, S. et al. Characterizing Membrane Association and Periplasmic Transfer of Bacterial Lipoproteins through Molecular Dynamics Simulations. *Structure* **28**, 475-487.e3 (2020).
115. Tang, X. et al. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nat Struct Mol Biol* **28**, 347-355 (2021).
116. Sharma, S. et al. Mechanism of LolCDE as a molecular extruder of bacterial triacylated lipoproteins. *Nat Commun* **12**, 4687 (2021).

117. Takeda, K. et al. Crystal structures of bacterial lipoprotein localization factors, LolA and LolB. *Embo j* **22**, 3199-209 (2003).
118. Okuda, S., Watanabe, S. & Tokuda, H. A short helix in the C-terminal region of LolA is important for the specific membrane localization of lipoproteins. *FEBS Lett* **582**, 2247-51 (2008).
119. 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**, 10530-9 (2014).
120. Tsukahara, J., Mukaiyama, K., Okuda, S., Narita, S. & Tokuda, H. Dissection of LolB function--lipoprotein binding, membrane targeting and incorporation of lipoproteins into lipid bilayers. *Febs j* **276**, 4496-504 (2009).
121. Nakada, S. et al. Structural investigation of the interaction between LolA and LolB using NMR. *J Biol Chem* **284**, 24634-43 (2009).
122. Boags, A., Samsudin, F. & Khalid, S. Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA. *Sci Rep* **9**, 3717 (2019).
123. Grabowicz, M. Lipoprotein Transport: Greasing the Machines of Outer Membrane Biogenesis: Re-Examining Lipoprotein Transport Mechanisms Among Diverse Gram-Negative Bacteria While Exploring New Discoveries and Questions. *Bioessays* **40**, e1700187 (2018).
124. Grabowicz, M. & Silhavy, T.J. Redefining the essential trafficking pathway for outer membrane lipoproteins. *Proc Natl Acad Sci U S A* **114**, 4769-4774 (2017).
125. Grabowicz, M. Lipoproteins and Their Trafficking to the Outer Membrane. *EcoSal Plus* **8**(2019).
126. Gwin, C.M. et al. The apolipoprotein N-acyl transferase Lnt is dispensable for growth in *Acinetobacter* species. *Microbiology (Reading)* **164**, 1547-1556 (2018).
127. McClain, M.S., Voss, B.J. & Cover, T.L. Lipoprotein Processing and Sorting in *Helicobacter pylori*. *mBio* **11**(2020).
128. LoVullo, E.D., Wright, L.F., Isabella, V., Huntley, J.F. & Pavelka, M.S., Jr. Revisiting the Gram-negative lipoprotein paradigm. *J Bacteriol* **197**, 1705-15 (2015).
129. Narita, S.I. & Tokuda, H. Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1414-1423 (2017).
130. Konovalova, A. & Silhavy, T.J. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos Trans R Soc Lond B Biol Sci* **370**(2015).
131. Principato, S., Pizza, M. & Rappuoli, R. Meningococcal factor H binding protein as immune evasion factor and vaccine antigen. *FEBS Lett* **594**, 2657-2669 (2020).
132. Wilson, M.M. & Bernstein, H.D. Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. *Trends Microbiol* **24**, 198-208 (2016).
133. Dowdell, A.S. et al. Comprehensive Spatial Analysis of the *Borrelia burgdorferi* Lipoproteome Reveals a Compartmentalization Bias toward the Bacterial Surface. *J Bacteriol* **199**(2017).
134. Zuckert, W.R. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta* **1843**, 1509-16 (2014).
135. Zielke, R.A. et al. SliC is a surface-displayed lipoprotein that is required for the anti-lysozyme strategy during *Neisseria gonorrhoeae* infection. *PLoS Pathog* **14**, e1007081 (2018).

136. Hooda, Y., Shin, H.E., Bateman, T.J. & Moraes, T.F. Neisserial surface lipoproteins: structure, function and biogenesis. *Pathog Dis* **75**(2017).
137. Lauber, F., Cornelis, G.R. & Renzi, F. Identification of a New Lipoprotein Export Signal in Gram-Negative Bacteria. *mBio* **7**(2016).
138. Hirashima, A., Wu, H.C., Venkateswaran, P.S. & Inouye, M. Two forms of a structural lipoprotein in the envelope of Escherichia coli. Further characterization of the free form. *J Biol Chem* **248**, 5654-9 (1973).
139. Cowles, C.E., Li, Y., Semmelhack, M.F., Cristea, I.M. & Silhavy, T.J. The free and bound forms of Lpp occupy distinct subcellular locations in Escherichia coli. *Mol Microbiol* **79**, 1168-81 (2011).
140. Cascales, E., Bernadac, A., Gavioli, M., Lazzaroni, J.C. & Lloubes, R. Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. *J Bacteriol* **184**, 754-9 (2002).
141. Michel, L.V. et al. Dual orientation of the outer membrane lipoprotein Pal in Escherichia coli. *Microbiology* **161**, 1251-9 (2015).
142. Cho, S.H. et al. Detecting Envelope Stress by Monitoring beta-Barrel Assembly. *Cell* **159**, 1652-64 (2014).
143. Konovalova, A., Perlman, D.H., Cowles, C.E. & Silhavy, T.J. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of beta-barrel proteins. *Proc Natl Acad Sci U S A* **111**, E4350-8 (2014).
144. Konovalova, A., Mitchell, A.M. & Silhavy, T.J. A lipoprotein/ β -barrel complex monitors lipopolysaccharide integrity transducing information across the outer membrane. *Elife* **5**(2016).
145. Rodriguez-Alonso, R. et al. Structural insight into the formation of lipoprotein-beta-barrel complexes. *Nat Chem Biol* **16**, 1019-1025 (2020).
146. Dekoninck, K. et al. Defining the function of OmpA in the Rcs stress response. *Elife* **9**(2020).
147. Webb, C.T. & Lithgow, T. Identification of BamC on the Surface of E. coli. *Methods Mol Biol* **1329**, 215-25 (2015).
148. Webb, C.T. et al. Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *J Mol Biol* **422**, 545-55 (2012).
149. Lenhart, T.R. & Akins, D.R. Borrelia burgdorferi locus BB0795 encodes a Bama orthologue required for growth and efficient localization of outer membrane proteins. *Mol Microbiol* **75**, 692-709 (2010).
150. Goyal, P. et al. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature* **516**, 250-3 (2014).
151. Dong, C. et al. Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. *Nature* **444**, 226-9 (2006).
152. Dunstan, R.A. et al. Assembly of the secretion pores GspD, Wza and CsgG into bacterial outer membranes does not require the Omp85 proteins BamA or TamA. *Mol Microbiol* **97**, 616-29 (2015).
153. Hooda, Y. & Moraes, T.F. Translocation of lipoproteins to the surface of gram negative bacteria. *Curr Opin Struct Biol* **51**, 73-79 (2018).
154. Hooda, Y., Lai, C.C.L. & Moraes, T.F. Identification of a Large Family of Slam-Dependent Surface Lipoproteins in Gram-Negative Bacteria. *Front Cell Infect Microbiol* **7**, 207 (2017).

155. Hooda, Y. et al. Slam is an outer membrane protein that is required for the surface display of lipitated virulence factors in *Neisseria*. *Nature Microbiology* **1**, 16009 (2016).
156. Nicchi, S. et al. Decorating the surface of *Escherichia coli* with bacterial lipoproteins: a comparative analysis of different display systems. *Microb Cell Fact* **20**, 33 (2021).
157. Pena, R.T. et al. Relationship Between Quorum Sensing and Secretion Systems. *Front Microbiol* **10**, 1100 (2019).
158. Palmer, T., Finney, A.J., Saha, C.K., Atkinson, G.C. & Sargent, F. A holin/peptidoglycan hydrolase-dependent protein secretion system. *Mol Microbiol* **115**, 345-355 (2021).
159. Grossman, A.S., Mauer, T.J., Forest, K.T. & Goodrich-Blair, H. A Widespread Bacterial Secretion System with Diverse Substrates. *mBio* **12**, e0195621 (2021).
160. Francetić, O. & Pugsley, A.P. Towards the identification of type II secretion signals in a nonacylated variant of pullulanase from *Klebsiella oxytoca*. *J Bacteriol* **187**, 7045-55 (2005).
161. East, A. et al. Structural Basis of Pullulanase Membrane Binding and Secretion Revealed by X-Ray Crystallography, Molecular Dynamics and Biochemical Analysis. *Structure* **24**, 92-104 (2016).
162. Pugsley, A.P., Kornacker, M.G. & Ryter, A. Analysis of the subcellular location of pullulanase produced by *Escherichia coli* carrying the *pulA* gene from *Klebsiella pneumoniae* strain UNF5023. *Mol Microbiol* **4**, 59-72 (1990).
163. Green, E.R. & Mecsas, J. Bacterial Secretion Systems: An Overview. *Microbiol Spectr* **4**(2016).
164. Roussel-Jazede, V., Grijpstra, J., van Dam, V., Tommassen, J. & van Ulsen, P. Lipidation of the autotransporter NalP of *Neisseria meningitidis* is required for its function in the release of cell-surface-exposed proteins. *Microbiology* **159**, 286-95 (2013).
165. Icke, C. et al. Glycine acylation and trafficking of a new class of bacterial lipoprotein by a composite secretion system. *Elife* **10**(2021).
166. Majdalani, N. & Gottesman, S. The Rcs phosphorelay: a complex signal transduction system. *Annu Rev Microbiol* **59**, 379-405 (2005).
167. 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**, 2065-74 (2008).
168. Torres-Cabassa, A.S. & Gottesman, S. Capsule synthesis in *Escherichia coli* K-12 is regulated by proteolysis. *J Bacteriol* **169**, 981-9 (1987).
169. Wall, E., Majdalani, N. & Gottesman, S. The Complex Rcs Regulatory Cascade. *Annu Rev Microbiol* **72**, 111-139 (2018).
170. Hussein, N.A., Cho, S.H., Laloux, G., Siam, R. & Collet, J.F. Distinct domains of *Escherichia coli* IgaA connect envelope stress sensing and down-regulation of the Rcs phosphorelay across subcellular compartments. *PLoS Genet* **14**, e1007398 (2018).
171. Cano, D.A., Dominguez-Bernal, G., Tierrez, A., Garcia-Del Portillo, F. & Casadesus, J. Regulation of capsule synthesis and cell motility in *Salmonella enterica* by the essential gene *igaA*. *Genetics* **162**, 1513-23 (2002).
172. Gottesman, S., Trisler, P. & Torres-Cabassa, A. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. *J Bacteriol* **162**, 1111-9 (1985).

173. Aslund, F., Zheng, M., Beckwith, J. & Storz, G. Regulation of the OxyR transcription factor by hydrogen peroxide and the cellular thiol-disulfide status. *Proc Natl Acad Sci U S A* **96**, 6161-5 (1999).
174. Ferrieres, L. & Clarke, D.J. The RcsC sensor kinase is required for normal biofilm formation in Escherichia coli K-12 and controls the expression of a regulon in response to growth on a solid surface. *Mol Microbiol* **50**, 1665-82 (2003).
175. Hagiwara, D. et al. Genome-wide analyses revealing a signaling network of the RcsC-YojN-RcsB phosphorelay system in Escherichia coli. *J Bacteriol* **185**, 5735-46 (2003).
176. Takeda, S., Fujisawa, Y., Matsubara, M., Aiba, H. & Mizuno, T. A novel feature of the multistep phosphorelay in Escherichia coli: a revised model of the RcsC --> YojN --> RcsB signalling pathway implicated in capsular synthesis and swarming behaviour. *Mol Microbiol* **40**, 440-50 (2001).
177. Rogov, V.V., Rogova, N.Y., Bernhard, F., Lohr, F. & Dotsch, V. A disulfide bridge network within the soluble periplasmic domain determines structure and function of the outer membrane protein RCSF. *J Biol Chem* **286**, 18775-83 (2011).
178. Leverrier, P. et al. Crystal structure of the outer membrane protein RcsF, a new substrate for the periplasmic protein-disulfide isomerase DsbC. *J Biol Chem* **286**, 16734-42 (2011).
179. Masi, M., Pinet, E. & Pagès, J.M. Complex Response of the CpxAR Two-Component System to β -Lactams on Antibiotic Resistance and Envelope Homeostasis in Enterobacteriaceae. *Antimicrob Agents Chemother* **64**(2020).
180. Delhay, A., Collet, J.F. & Laloux, G. Fine-Tuning of the Cpx Envelope Stress Response Is Required for Cell Wall Homeostasis in Escherichia coli. *MBio* **7**, e00047-16 (2016).
181. Laloux, G. & Collet, J.F. "Major Tom to ground control: how lipoproteins communicate extra-cytoplasmic stress to the decision center of the cell". *J Bacteriol*, 00216-17 (2017).
182. Delhay, A., Laloux, G. & Collet, J.F. The Lipoprotein NlpE Is a Cpx Sensor That Serves as a Sentinel for Protein Sorting and Folding Defects in the Escherichia coli Envelope. *J Bacteriol* **201**(2019).
183. Raivio, T.L., Leblanc, S.K. & Price, N.L. The Escherichia coli Cpx envelope stress response regulates genes of diverse function that impact antibiotic resistance and membrane integrity. *J Bacteriol* **195**, 2755-67 (2013).
184. May, K.L., Lehman, K.M., Mitchell, A.M. & Grabowicz, M. A Stress Response Monitoring Lipoprotein Trafficking to the Outer Membrane. *mBio* **10**(2019).
185. Li, G.W., Burkhardt, D., Gross, C. & Weissman, J.S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624-35 (2014).
186. Hoekstra, D., van der Laan, J.W., de Leij, L. & Witholt, B. Release of outer membrane fragments from normally growing Escherichia coli. *Biochim Biophys Acta* **455**, 889-99 (1976).
187. Asmar, A.T. et al. Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS Biol* **15**, e2004303 (2017).
188. Winkle, M. et al. DpaA Detaches Braun's Lipoprotein from Peptidoglycan. *mBio* **12**(2021).
189. Cohen, E.J., Ferreira, J.L., Ladinsky, M.S., Beeby, M. & Hughes, K.T. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science* **356**, 197-200 (2017).

190. Godessart, P. et al. β -Barrels covalently link peptidoglycan and the outer membrane in the α -proteobacterium *Brucella abortus*. *Nat Microbiol* **6**, 27-33 (2021).
191. Sandoz, K.M. et al. β -Barrel proteins tether the outer membrane in many Gram-negative bacteria. *Nat Microbiol* **6**, 19-26 (2021).
192. Szczepaniak, J., Press, C. & Kleanthous, C. The multifarious roles of Tol-Pal in Gram-negative bacteria. *FEMS Microbiol Rev* **44**, 490-506 (2020).
193. Gerding, M.A., Ogata, Y., Pecora, N.D., Niki, H. & de Boer, P.A. The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol Microbiol* **63**, 1008-25 (2007).
194. Petiti, M. et al. Tol Energy-Driven Localization of Pal and Anchoring to the Peptidoglycan Promote Outer-Membrane Constriction. *J Mol Biol* **431**, 3275-3288 (2019).
195. Szczepaniak, J. et al. The lipoprotein Pal stabilises the bacterial outer membrane during constriction by a mobilisation-and-capture mechanism. *Nat Commun* **11**, 1305 (2020).
196. Heidrich, C. et al. Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*. *Mol Microbiol* **41**, 167-78 (2001).
197. Uehara, T., Parzych, K.R., Dinh, T. & Bernhardt, T.G. Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis. *EMBO J* **29**, 1412-22 (2010).
198. Tsang, M.J., Yakhnina, A.A. & Bernhardt, T.G. NlpD links cell wall remodeling and outer membrane invagination during cytokinesis in *Escherichia coli*. *PLoS Genet* **13**, e1006888 (2017).
199. Bryant, J.A. et al. Structure of dual BON-domain protein DolP identifies phospholipid binding as a new mechanism for protein localisation. *Elife* **9**(2020).
200. Ranava, D. et al. Lipoprotein DolP supports proper folding of BamA in the bacterial outer membrane promoting fitness upon envelope stress. *Elife* **10**(2021).
201. Kim, K.H., Aulakh, S. & Paetzel, M. Crystal structure of beta-barrel assembly machinery BamCD protein complex. *J Biol Chem* **286**, 39116-21 (2011).
202. Warner, L.R. et al. Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set. *J Mol Biol* **411**, 83-95 (2011).
203. Kim, K.H., Aulakh, S., Tan, W. & Paetzel, M. Crystallographic analysis of the C-terminal domain of the *Escherichia coli* lipoprotein BamC. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 1350-8 (2011).
204. Diao, J. et al. Inhibition of *Escherichia coli* Lipoprotein Diacylglyceryl Transferase Is Insensitive to Resistance Caused by Deletion of Braun's Lipoprotein. *J Bacteriol* **203**, e0014921 (2021).
205. Garland, K. et al. Optimization of globomycin analogs as novel gram-negative antibiotics. *Bioorg Med Chem Lett* **30**, 127419 (2020).
206. TOSHIHARU YAKUSHI, T.T., SHIN-ICHI MATSUYAMA, HAJIME TOKUDA. Lethality of the Covalent Linkage between Mislocalized Major Outer Membrane Lipoprotein and the Peptidoglycan of *Escherichia coli*. *Journal of Bacteriology* **179**, 2857-2862 (1997).
207. Lehman, K.M. & Grabowicz, M. Countering Gram-Negative Antibiotic Resistance: Recent Progress in Disrupting the Outer Membrane with Novel Therapeutics. *Antibiotics (Basel)* **8**(2019).

208. Sarah M. McLeod, P.R.F., Kathleen MacCormack, Robert E. McLaughlin, James D. Whiteaker, Shin-ichiro Narita, Makiko Mori, Hajime Tokuda, Alita A. Millera. Small-Molecule Inhibitors of Gram-Negative Lipoprotein Trafficking Discovered by Phenotypic Screening. *Journal of Bacteriology* **197**, 1075-1082 (2015).
209. Nickerson, N.N. et al. A Novel Inhibitor of the LolCDE ABC Transporter Essential for Lipoprotein Trafficking in Gram-Negative Bacteria. *Antimicrob Agents Chemother* **62**(2018).
210. Nayar, A.S. et al. Novel antibacterial targets and compounds revealed by a high-throughput cell wall reporter assay. *J Bacteriol* **197**, 1726-34 (2015).
211. Bean, G.J. et al. A22 disrupts the bacterial actin cytoskeleton by directly binding and inducing a low-affinity state in MreB. *Biochemistry* **48**, 4852-7 (2009).
212. Pathania, R. et al. Chemical genomics in Escherichia coli identifies an inhibitor of bacterial lipoprotein targeting. *Nat Chem Biol* **5**, 849-56 (2009).
213. Barker, C.A. et al. Degradation of MAC13243 and studies of the interaction of resulting thiourea compounds with the lipoprotein targeting chaperone LolA. *Bioorg Med Chem Lett* **23**, 2426-31 (2013).
214. Muheim, C. et al. Increasing the permeability of Escherichia coli using MAC13243. *Sci Rep* **7**, 17629 (2017).
215. Buss, J.A. et al. Pathway-Directed Screen for Inhibitors of the Bacterial Cell Elongation Machinery. *Antimicrob Agents Chemother* **63**(2019).
216. Sickmeier, M. et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res* **35**, D786-93 (2007).
217. He, B. et al. Predicting intrinsic disorder in proteins: an overview. *Cell Res* **19**, 929-49 (2009).
218. Oldfield, C.J. et al. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989-2000 (2005).
219. Peng, Z. et al. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137-51 (2015).
220. Pancsa, R. & Tompa, P. Structural disorder in eukaryotes. *PLoS One* **7**, e34687 (2012).
221. Xue, B., Dunker, A.K. & Uversky, V.N. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn* **30**, 137-49 (2012).
222. Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C. & Brown, C.J. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **11**, 161-71 (2000).
223. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-45 (2004).
224. Galea, C.A. et al. Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *J Proteome Res* **8**, 211-26 (2009).
225. Basile, W., Salvatore, M., Bassot, C. & Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput Biol* **15**, e1007186 (2019).
226. Tompa, P., Dosztanyi, Z. & Simon, I. Prevalent structural disorder in E. coli and S. cerevisiae proteomes. *J Proteome Res* **5**, 1996-2000 (2006).
227. Habchi, J., Tompa, P., Longhi, S. & Uversky, V.N. Introducing protein intrinsic disorder. *Chem Rev* **114**, 6561-88 (2014).

228. Mishra, P.M., Verma, N.C., Rao, C., Uversky, V.N. & Nandi, C.K. Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis. *Prog Mol Biol Transl Sci* **174**, 1-78 (2020).
229. Uversky, V.N. Functional roles of transiently and intrinsically disordered regions within proteins. *Febs j* **282**, 1182-9 (2015).
230. Uversky, V.N. What does it mean to be natively unfolded? *Eur J Biochem* **269**, 2-12 (2002).
231. Oldfield, C.J. & Dunker, A.K. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**, 553-84 (2014).
232. He, H., Zhao, J. & Sun, G. Computational prediction of MoRFs based on protein sequences and minimax probability machine. *BMC Bioinformatics* **20**, 529 (2019).
233. Ivanyi-Nagy, R., Davidovic, L., Khandjian, E.W. & Darlix, J.L. Disordered RNA chaperone proteins: from functions to disease. *Cell Mol Life Sci* **62**, 1409-17 (2005).
234. Dunker, A.K. & Obradovic, Z. The protein trinity--linking function and disorder. *Nat Biotechnol* **19**, 805-6 (2001).
235. Romero, P.R. et al. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci U S A* **103**, 8390-5 (2006).
236. Inobe, T., Fishbain, S., Prakash, S. & Matouschek, A. Defining the geometry of the two-component proteasome degron. *Nat Chem Biol* **7**, 161-7 (2011).
237. McElheny, D., Schnell, J.R., Lansing, J.C., Dyson, H.J. & Wright, P.E. Defining the role of active-site loop fluctuations in dihydrofolate reductase catalysis. *Proc Natl Acad Sci U S A* **102**, 5032-7 (2005).
238. Jakob, U., Kriwacki, R. & Uversky, V.N. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem Rev* **114**, 6779-805 (2014).
239. Uversky, V.N. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* **43**, 1090-103 (2011).
240. Iakoucheva, L.M., Brown, C.J., Lawson, J.D., Obradović, Z. & Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol* **323**, 573-84 (2002).
241. Lee, H. et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J Biol Chem* **275**, 29426-32 (2000).
242. Adkins, J.N. & Lumb, K.J. Intrinsic structural disorder and sequence features of the cell cycle inhibitor p57Kip2. *Proteins* **46**, 1-7 (2002).
243. Uversky, V.N. & Dunker, A.K. Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal Chem* **84**, 2096-104 (2012).
244. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **37**, 509-16 (2012).
245. Kodera, N., Yamamoto, D., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. *Nature* **468**, 72-6 (2010).
246. Cohan, M.C. & Pappu, R.V. Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria. *Trends Biochem Sci* **45**, 668-680 (2020).
247. den Blaauwen, T., Hamoen, L.W. & Levin, P.A. The divisome at 25: the road ahead. *Curr Opin Microbiol* **36**, 85-94 (2017).

248. Gardner, K.A., Moore, D.A. & Erickson, H.P. The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide. *Mol Microbiol* **89**, 264-75 (2013).
249. Löwe, J. & Amos, L.A. Crystal structure of the bacterial cell-division protein FtsZ. *Nature* **391**, 203-6 (1998).
250. Bi, E.F. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161-4 (1991).
251. Buske, P.J. & Levin, P.A. A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo. *Mol Microbiol* **89**, 249-63 (2013).
252. Holmes, J.A. et al. Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc Natl Acad Sci U S A* **113**, 12490-12495 (2016).
253. Kumariya, R. et al. Bacteriocins: Classification, synthesis, mechanism of action and resistance development in food spoilage causing bacteria. *Microb Pathog* **128**, 171-177 (2019).
254. Housden, N.G. et al. Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **340**, 1570-4 (2013).
255. Housden, N.G. et al. Directional Porin Binding of Intrinsically Disordered Protein Sequences Promotes Colicin Epitope Display in the Bacterial Periplasm. *Biochemistry* **57**, 4374-4381 (2018).
256. Earl, C. et al. The majority of the matrix protein TapA is dispensable for Bacillus subtilis colony biofilm architecture. *Mol Microbiol* **114**, 920-933 (2020).
257. Abbasi, R. et al. The Bacterial Extracellular Matrix Protein TapA Is a Two-Domain Partially Disordered Protein. *Chembiochem* **20**, 355-359 (2019).
258. Green, A. et al. Are the curli proteins CsgE and CsgF intrinsically disordered? *Intrinsically Disord Proteins* **4**, e1130675 (2016).
259. Barnhart, M.M. & Chapman, M.R. Curli biogenesis and function. *Annu Rev Microbiol* **60**, 131-47 (2006).
260. Evans, M.L. & Chapman, M.R. Curli biogenesis: order out of disorder. *Biochim Biophys Acta* **1843**, 1551-8 (2014).
261. Zhang, M., Shi, H., Zhang, X., Zhang, X. & Huang, Y. Cryo-EM structure of the nonameric CsgG-CsgF complex and its implications for controlling curli biogenesis in Enterobacteriaceae. *PLoS Biol* **18**, e3000748 (2020).
262. Yan, Z., Yin, M., Chen, J. & Li, X. Assembly and substrate recognition of curli biogenesis system. *Nat Commun* **11**, 241 (2020).
263. Gazi, A.D. et al. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. *J Biol Chem* **283**, 34062-8 (2008).
264. Abergel, C., Walburger, A., Chenivresse, S. & Lazdunski, C. Crystallization and preliminary crystallographic study of the peptidoglycan-associated lipoprotein from Escherichia coli. *Acta Crystallogr D Biol Crystallogr* **57**, 317-9 (2001).
265. Clavel, T., Germon, P., Vianney, A., Portalier, R. & Lazzaroni, J.C. TolB protein of Escherichia coli K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Mol Microbiol* **29**, 359-67 (1998).
266. Cascales, E. & Lloubes, R. Deletion analyses of the peptidoglycan-associated lipoprotein Pal reveals three independent binding sequences including a TolA box. *Mol Microbiol* **51**, 873-85 (2004).

267. El Rayes, J., Rodríguez-Alonso, R. & Collet, J.F. Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles. *Curr Opin Microbiol* **61**, 25-34 (2021).
268. Chen, S. & Zuckert, W.R. Probing the *Borrelia burgdorferi* surface lipoprotein secretion pathway using a conditionally folding protein domain. *J Bacteriol* **193**, 6724-32 (2011).
269. Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim Biophys Acta* **1694**, IN1-9 (2004).
270. von Lackum, K. et al. *Borrelia burgdorferi* regulates expression of complement regulator-acquiring surface protein 1 during the mammal-tick infection cycle. *Infect Immun* **73**, 7398-405 (2005).
271. Pulzova, L. & Bhide, M. Outer surface proteins of *Borrelia*: peerless immune evasion tools. *Curr Protein Pept Sci* **15**, 75-88 (2014).
272. Watanabe, S., Oguchi, Y., Yokota, N. & Tokuda, H. Large-scale preparation of the homogeneous LolA lipoprotein complex and efficient in vitro transfer of lipoproteins to the outer membrane in a LolB-dependent manner. *Protein Sci* **16**, 2741-9 (2007).
273. Narita, S. & Tokuda, H. Sorting of bacterial lipoproteins to the outer membrane by the Lol system. *Methods Mol Biol* **619**, 117-29 (2010).
274. Remans, K. et al. Hydrophobic surface patches on LolA of *Pseudomonas aeruginosa* are essential for lipoprotein binding. *J Mol Biol* **401**, 921-30 (2010).

Annexes

I. Published paper: Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria



Disorder is a critical component of lipoprotein sorting in Gram-negative bacteria

Jessica El Rayes^{1,2,5}, Joanna Szewczyk^{1,2,5}, Michaël Deghelt^{1,2}, Naemi Csoma^{1,2}, André Matagne³, Bogdan I. Iorga⁴, Seung-Hyun Cho^{1,2} and Jean-François Collet^{1,2}✉

Gram-negative bacteria express structurally diverse lipoproteins in their cell envelope. Here, we find that approximately half of lipoproteins destined to the *Escherichia coli* outer membrane display an intrinsically disordered linker at their N terminus. Intrinsically disordered regions are common in proteins, but establishing their importance in vivo has remained challenging. As we sought to unravel how lipoproteins mature, we discovered that unstructured linkers are required for optimal trafficking by the Lol lipoprotein sorting system, whereby linker deletion re-routes three unrelated lipoproteins to the inner membrane. Focusing on the stress sensor RcsF, we found that replacing the linker with an artificial peptide restored normal outer-membrane targeting only when the peptide was of similar length and disordered. Overall, this study reveals the role played by intrinsic disorder in lipoprotein sorting, providing mechanistic insight into the biogenesis of these proteins and suggesting that evolution can select for intrinsic disorder that supports protein function.

The cell envelope is the morphological hallmark of *Escherichia coli* and other Gram-negative bacteria. It is composed of the inner membrane (a classical phospholipid bilayer) and the outer membrane (an asymmetric bilayer with phospholipids in the inner leaflet and lipopolysaccharides in the outer leaflet)¹. This lipid asymmetry enables the outer membrane to function as a barrier that effectively prevents the diffusion of toxic compounds from the environment and into the cell. The inner and outer membranes are separated by the periplasm, a viscous compartment that contains a thin layer of peptidoglycan, also known as the cell wall¹. The cell envelope is essential for growth and survival, as illustrated by the fact that several antibiotics, such as the β -lactams, target mechanisms of envelope assembly. Mechanisms involved in envelope biogenesis and maintenance are therefore attractive targets for novel antibacterial strategies.

Approximately one-third of *E. coli* proteins are targeted to the envelope, either as soluble proteins present in the periplasm or as proteins inserted in one of the two membranes². Although inner-membrane proteins cross the lipid bilayer via one or more hydrophobic α -helices, proteins inserted in the outer membrane generally adopt a β -barrel conformation³. Another important group of envelope proteins comprises the lipoproteins, which are globular proteins anchored to one of the two membranes by a lipid moiety⁴. Lipoproteins carry out a variety of important functions in the cell envelope: they participate in biogenesis of the outer membrane by inserting lipopolysaccharide molecules^{5,6} and β -barrel proteins⁷, they function as stress sensors triggering signal transduction cascades when envelope integrity is altered⁸, and they control processes that are important for virulence⁹. The diverse roles played by lipoproteins in the cell envelope have drawn a lot of attention, revealing how crucial these proteins are in a wide range of vital processes and identifying them as attractive targets for antibiotic development. Yet, a detailed understanding of the mechanisms involved in lipoprotein maturation and trafficking is still missing.

Lipoproteins are synthesized in the cytoplasm as precursors with an N-terminal signal peptide¹⁰. The last four C-terminal residues of this signal peptide, known as the lipobox, function as a molecular determinant of lipid modification unique to bacteria; only the cysteine at the last position of the lipobox is strictly conserved¹¹. After translocation across the inner membrane, processing of pre-lipoproteins into mature forms takes place on the periplasmic side of the membrane. The thiol side chain of the cysteine is first modified with a diacylglycerol moiety by prolipoprotein diacylglycerol transferase (Lgt)¹⁰ (Supplementary Fig. 1a, step 1), then signal peptidase II (LspA) catalyzes cleavage of the signal peptide N-terminally of the lipidated cysteine before apolipoprotein N-acyltransferase (Lnt) adds a third acyl group to the N-terminal amino group of the cysteine (Supplementary Fig. 1a, steps 2–3). Most mature lipoproteins are then transported to the outer membrane by the Lol system. Lol consists of LolCDE, an ABC transporter that extracts lipoproteins from the inner membrane and transfers them to the soluble periplasmic chaperone LolA (Supplementary Fig. 1a, steps 4–5)¹². LolA escorts lipoproteins across the periplasm, binding their hydrophobic lipid tail, and delivers them to the outer-membrane lipoprotein LolB (Supplementary Fig. 1a, step 6). LolB finally anchors lipoproteins to the inner leaflet of the outer membrane using a mechanism that remains poorly characterized (Supplementary Fig. 1a, step 7).

In most Gram-negative bacteria, a few lipoproteins remain in the inner membrane¹³. The current view is that inner-membrane retention depends on the identity of the two residues located immediately downstream of the N-terminal cysteine on which the lipid moiety is attached¹⁴; this sequence, two amino acids in length, is known as the Lol sorting signal. When lipoproteins have an aspartate at position +2 and an aspartate, glutamate or glutamine at position +3, they remain in the inner membrane^{15,16}, possibly because strong electrostatic interactions between the +2 aspartate and membrane phospholipids prevent their interaction with LolCDE¹⁷. However, this model is largely based on data obtained in *E. coli* and variations

¹WELBIO, Brussels, Belgium. ²de Duve Institute, Université catholique de Louvain, Brussels, Belgium. ³Centre d'ingénierie des Protéines, Institut de Chimie B6, Université de Liège, Liège, Belgium. ⁴Université Paris-Saclay, CNRS UPR 2301, Institut de Chimie des Substances Naturelles, Gif-sur-Yvette, France.

⁵These authors contributed equally: Jessica El Rayes, Joanna Szewczyk. ✉e-mail: jfcollet@uclouvain.be

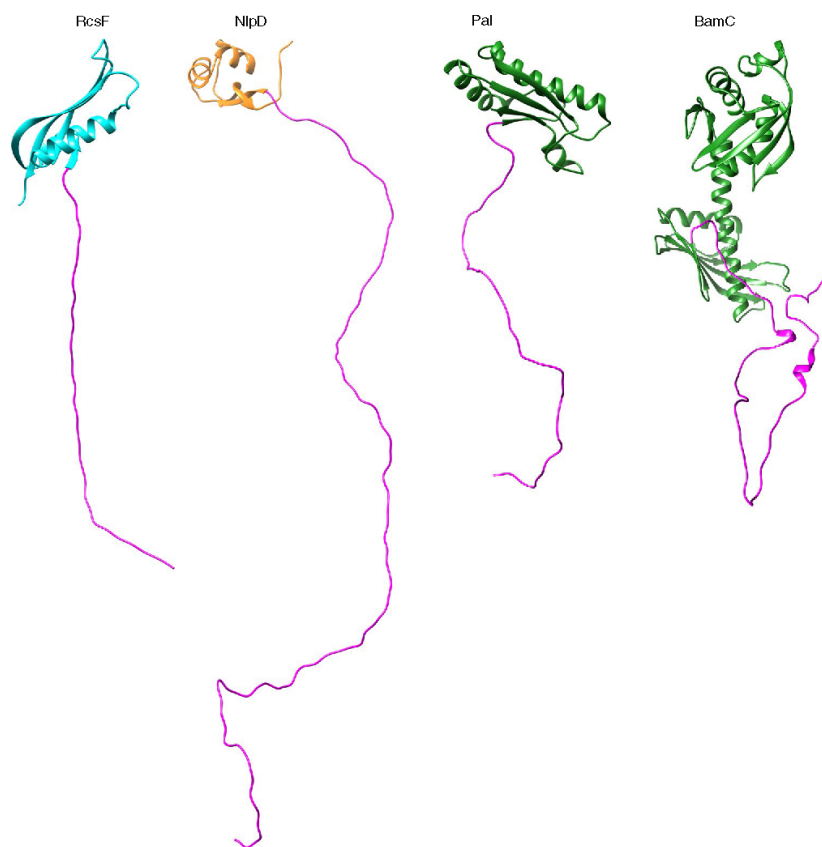


Fig. 1 | Structural analysis of lipoproteins reveals that half of outer-membrane lipoproteins display an intrinsically disordered linker at the N terminus. The structures of the outer-membrane lipoproteins RcsF, NlpD, Pal and BamC were generated by comparative modeling (Methods). The 39 remaining lipoproteins are shown in Extended Data Fig. 1. X-ray and cryo-EM structures are green, NMR structures are cyan, and structures built by comparative modeling from the closest analog in the same PFAM group are orange. In all cases, the N-terminal linker is magenta.

have been described in other bacteria. For example, in the pathogen *Pseudomonas aeruginosa*, an aspartate is rarely found at position +2, and inner-membrane retention appears to be determined by residues +3 and +4^{18,19}. Surprisingly, lipoproteins are well sorted in *P. aeruginosa* cells expressing the *E. coli* LolCDE complex²⁰, despite their different Lol sorting signal. This result cannot be explained by the current model of lipoprotein sorting, underscoring that our comprehension of the precise mechanism that governs the triage of lipoproteins remains incomplete.

Excitingly, more unresolved questions regarding lipoprotein biogenesis have recently been raised. First, it was reported that a LolA-LolB-independent trafficking route to the outer membrane exists in *E. coli*²¹, but the factors involved have remained unknown. Second, although lipoproteins have traditionally been considered to be exposed to the periplasm in *E. coli* and many other bacterial models¹⁰, a series of investigations have started to challenge this view by identifying lipoproteins on the surface of *E. coli*, *Vibrio cholerae* and *Salmonella typhimurium*^{22–24}. Overall, the field is beginning to explore a lipoprotein topological landscape that is more complex than previously assumed, raising intriguing questions about the signals that control surface targeting and exposure.

Stimulated by the hypothesis that crucial details of the mechanisms underlying lipoprotein maturation remained to be elucidated,

we sought to identify molecular determinants controlling lipoprotein biogenesis. First, we found that half of the outer-membrane lipoproteins display a long and intrinsically disordered linker at their N terminus. Intrigued by these unstructured segments, we probed their importance for the biogenesis of RcsF, NlpD and Pal, three unrelated outer-membrane lipoproteins. Unexpectedly, we found that deleting the linker altered the targeting of all three lipoproteins to the outer membrane. Focusing on RcsF, we determined that both the length and disordered character of the linker were important. We also found that the linker is required for optimal processing by Lol. Our observations reveal the unsuspected role played by protein intrinsic disorder in lipoprotein cellular localization. We propose a model in which disordered linkers are important for targeting and/or displaying lipoproteins on the cell surface.

Results

Half of lipoproteins present N-terminal disordered segments. In an attempt to discover novel molecular determinants controlling the cellular localization of lipoproteins, we decided to systematically analyze the sequence of the lipoproteins encoded by the *E. coli* genome (strain MG1655) in search of unidentified structural features. *Escherichia coli* encodes approximately 80 validated lipoproteins²⁵, of which 58 have been shown experimentally to localize

in the outer membrane²⁶. Comparative modeling of existing X-ray, cryo-EM and NMR structures revealed that approximately half of these outer-membrane lipoproteins display a long segment (>22 residues) that is predicted to be disordered at the N terminus (Fig. 1, Extended Data Fig. 1 and Supplementary Table 1). By contrast, only one of the eight lipoproteins that remain in the inner membrane (DcrB; Extended Data Fig. 1 and Supplementary Table 1) had a long, disordered linker, suggesting that disordered peptides may be important for lipoprotein sorting.

Linker deletion perturbs outer-membrane targeting. Intrigued by the presence of these N-terminal disordered segments in so many outer-membrane lipoproteins, we decided to investigate their functional importance. We selected three structurally unrelated lipoproteins whose function could easily be assessed: the stress sensor RcsF (which triggers the Rcs signaling cascade when damage occurs in the envelope²⁷), NlpD (which activates the periplasmic *N*-acetylmuramyl-L-alanine amidase AmiC, which is involved in peptidoglycan cleavage during cell division²⁸) and the peptidoglycan-binding lipoprotein Pal (which is important for outer-membrane constriction during cell division²⁹).

We began by preparing truncated versions of RcsF, NlpD and Pal, devoid of their N-terminal unstructured linkers (Supplementary Fig. 1b and Extended Data Fig. 1; RcsF_{Δ19–47}, Pal_{Δ26–56} and NlpD_{Δ29–64}). Note that the lipidated cysteine residue (+1) and the Lol sorting signal (the amino acids at positions +2 and +3) were not altered in RcsF_{Δ19–47}, Pal_{Δ26–56} and NlpD_{Δ29–64}, nor in any of the constructs discussed below except in RcsF_{CDP} (Supplementary Table 2). For Pal, although the unstructured linker spans residues 25–68 (Fig. 1), we used Pal_{Δ26–56} because Pal_{Δ25–68} was either degraded or not detected by the antibody. We first tested whether the truncated lipoproteins were still correctly extracted from the inner membrane and transported to the outer membrane. The membrane fraction was prepared from cells expressing the three variants independently, and the outer and inner membranes were separated using sucrose density gradients (Methods). Whereas wild-type RcsF, NlpD and Pal were mostly detected (>90%) in the outer-membrane fraction, ~50% of RcsF_{Δ19–47} and ~60% of NlpD_{Δ29–64} were retained in the inner membrane (Fig. 2a and Extended Data Fig. 2a). An RcsF variant in which the Lol sorting signal was replaced by two aspartates (RcsF_{CDP}) was used as a control for inner-membrane retention. As expected, RcsF_{CDP} was mostly detected in the inner membrane (Extended Data Fig. 3). The sorting of Pal was also affected, but to a lesser extent, with 15% of Pal_{Δ26–56} retained in the inner membrane (Extended Data Fig. 2b). Notably, the expression levels of the three linker-less variants were similar (NlpD_{Δ29–64}) or lower (RcsF_{Δ19–47}, Pal_{Δ26–56}) than those of the wild-type proteins (Extended Data Fig. 4), indicating that accumulation in the inner membrane did not result from increased protein abundance.

We then tested the impact of linker deletion on the function of these three proteins. In cells expressing RcsF_{Δ19–47}, the Rcs system

was constitutively turned on (Fig. 2b). When RcsF accumulates in the inner membrane, it becomes available for interaction with IgaA, its downstream Rcs partner in the inner membrane^{27,30}. Similarly, expression of NlpD_{Δ29–64} did not rescue the chaining phenotype (Fig. 2c)³¹ exhibited by cells lacking both *nlpD* and *envC*, an activator of amidases AmiA and AmiB²⁸. Finally, Pal_{Δ26–56} partially rescued the sensitivity of the *pal* mutant to SDS-EDTA that results from increased membrane permeability³² (Fig. 2d). However, this observation needs to be considered with caution given that Pal_{Δ26–56} seemed to be expressed at lower levels than wild-type Pal (Extended Data Fig. 4). Thus, preventing normal targeting of RcsF, NlpD and Pal to the outer membrane has functional consequences.

Unstructured artificial linkers do the job. The above results were surprising because they revealed that the normal targeting of RcsF, NlpD and Pal to the outer membrane not only requires an appropriate Lol sorting signal, as proposed by the current model for lipoprotein sorting¹⁰, but also the presence of an N-terminal linker. We selected RcsF, the accumulation of which in the inner membrane can be tracked easily by monitoring Rcs activity²⁷, to investigate the structural features of the linker controlling lipoprotein maturation. Keeping as little as 10% of the total pool of RcsF molecules in the inner membrane is sufficient to fully activate Rcs²⁷.

We first tested whether changing the sequence of the N-terminal segment while preserving its disordered character still yielded normal targeting of the protein to the outer membrane. To that end, we prepared an RcsF variant in which the N-terminal linker was replaced by an artificial, unstructured sequence (Supplementary Table 2 and Extended Data Figs. 1 and 5) of similar length and consisting mostly of glycine-serine repeats (RcsF_{GS}). Substituting the wild-type linker with this artificial sequence was remarkably well tolerated by RcsF: RcsF_{GS} was targeted normally to the outer membrane (Fig. 3a, Extended Data Fig. 1 and Supplementary Fig. 2a) and did not constitutively activate the stress system (Fig. 3b). Thus, although RcsF_{GS} has an N terminus with a completely different primary structure, it behaved like the wild-type protein.

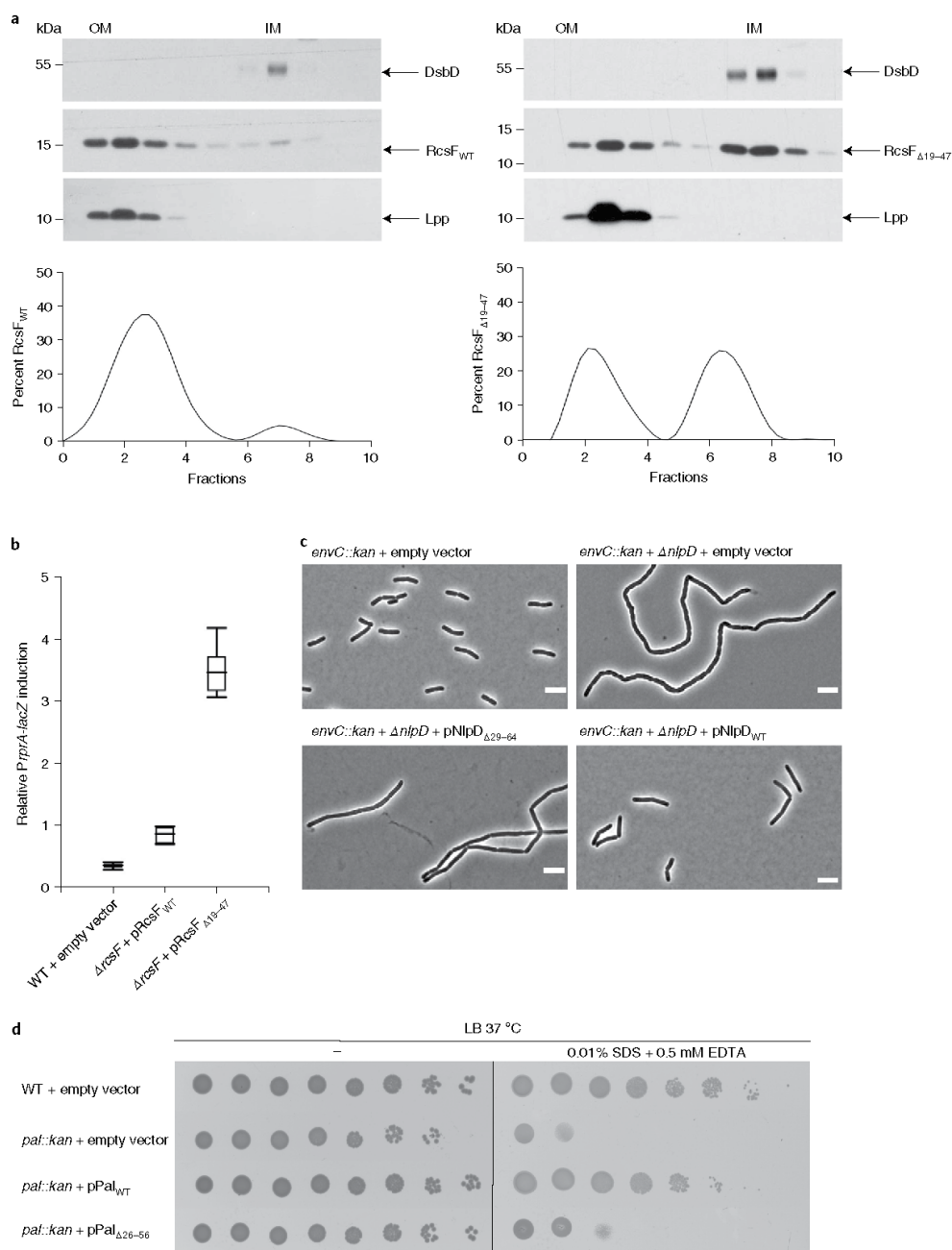
We then investigated whether the N-terminal linker required a minimal length for proper targeting and function. We therefore constructed two RcsF variants with shorter, unstructured, artificial linkers (RcsF_{GS2} and RcsF_{GS3}, with linkers of 18 and 10 residues, respectively; Supplementary Table 2 and Extended Data Figs. 1 and 5). Importantly, RcsF_{GS2} and, to a greater extent, RcsF_{GS3} did not properly localize to the outer membrane: the shorter the linker, the more RcsF remained in the inner membrane (Fig. 3a, Extended Data Fig. 1 and Supplementary Fig. 2a). Consistent with the amount of RcsF_{GS2} and RcsF_{GS3} retained in the inner membrane, Rcs activation levels were inversely related to linker length (Fig. 3b).

The disordered character of the linker is crucial. Taken together, the results above demonstrated that the RcsF linker can be replaced with an artificial sequence lacking secondary structure, provided

Fig. 2 | The N-terminal linker displayed by lipoproteins is important for outer-membrane targeting. **a**, The outer membrane (OM) and inner membrane (IM) were separated by centrifugation in a three-step sucrose density gradient (Methods). Although RcsF_{WT} was found predominantly in the outer membrane, RcsF_{Δ19–47} was substantially retained in the inner membrane. Data are presented as the ratio of signal intensity in a single fraction to the total intensity in all fractions. All variants were expressed from plasmids (Supplementary Table 4). DsbD and Lpp were used as controls for the outer membrane and inner membrane, respectively. **b**, The Rcs system is constitutively active when the linker of RcsF is missing. Rcs activity was measured with a β-galactosidase assay in a strain harboring a transcriptional *rprA::lacZ* fusion (Methods). All variants were expressed from a plasmid. Results were normalized to expression levels of the RcsF variants. Box-and-whisker plots are shown from minimum to maximum, with the central line marking the median and the first and third quartiles shown by whiskers; *n* = 6 biologically independent experiments. **c**, Phase-contrast images of the *envC::kan ΔnlpD* mutant complemented with NlpD_{WT} or NlpD_{Δ29–64}. NlpD_{Δ29–64} only partially rescues the chaining phenotype of the *envC::kan ΔnlpD* double mutant. Scale bars, 5 μm. **d**, Expression of Pal_{Δ26–56} does not rescue the sensitivity of the *pal::kan* mutant to SDS-EDTA. Cells were grown in LB medium at 37 °C until mid-log phase (OD₆₀₀ = 0.5). Tenfold serial dilutions were made in LB medium, plated onto LB agar or LB agar supplemented with 0.01% SDS and 0.5 mM EDTA, and incubated at 37 °C. Images in **a**, **c** and **d** are representative of biological triplicates. The graphs in **a** were created by spline analysis of curves representing a mean of three independent experiments.

that it is of appropriate length. Next, we sought to directly probe the importance of having a disordered linker by replacing the RcsF linker with an α -helical segment, 35 amino acids long, from the periplasmic chaperone FkpA (RcsF_{FkpA}; Supplementary Table 2 and Extended Data Figs. 1 and 5). Introducing order at the N terminus of RcsF dramatically impacted the protein distribution between the two membranes: RcsF_{FkpA} was substantially retained in the inner membrane (Fig. 3c and Supplementary Fig. 2b) and constitutively

activated Rcs (Fig. 3d). As α -helical segments are considerably shorter than unstructured sequences containing a similar number of amino acids, we also prepared an RcsF variant (RcsF_{col}) with a longer α -helix from the helical segment of colicin Ia, which is 73 amino acids in length and also predicted to remain folded in the RcsF_{col} construct (Supplementary Table 2 and Extended Data Figs. 1 and 5). However, doubling the size of the helix had no impact, with RcsF_{col} behaving similarly to RcsF_{FkpA} (Fig. 3c,d



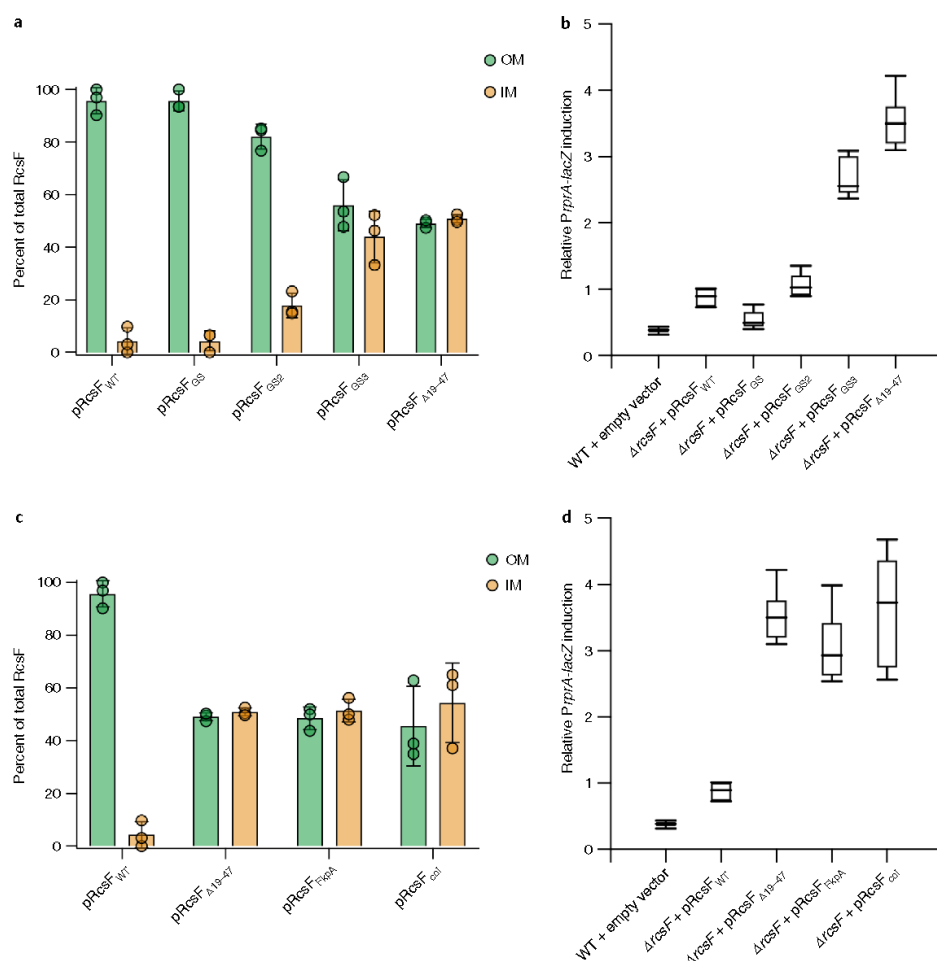


Fig. 3 | The length and the disordered character of the RcsF linker play key roles in RcsF targeting to the outer membrane. **a**, The outer and inner membranes were separated by centrifugation in a three-step sucrose density gradient (Methods). DsbD and Lpp were used as controls for the outer and inner membrane, respectively. The longer the linker, the more protein was correctly translocated to the inner membrane. All variants were expressed from plasmids (Supplementary Table 4). Aligned dot plots are shown with mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. Images are representative of experiments and immunoblots performed in biological triplicate. **b**, Rcs activity was measured with a β -galactosidase assay in a strain harboring a transcriptional *rprA::lacZ* fusion. All variants were expressed from a plasmid. Results were normalized to expression levels of RcsF variants. Box-and-whisker plots are shown from minimum to maximum, with the central line marking the median and the first and third quartiles shown as whiskers; $n = 6$ biologically independent experiments. Rcs activity relates to the quantity of RcsF retained in the inner membrane. **c**, RcsF mutants harboring α -helical linkers (RcsF_{RlpA} and RcsF_{col}) were subjected to two consecutive centrifugations in sucrose density gradients. Both mutants were inefficiently translocated from the inner to the outer membrane. All variants are expressed from plasmids. Aligned dot plots are shown with mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. Images are representative of experiments and immunoblots were performed in biological triplicate. **d**, The Rcs system was constitutively active in RcsF_{RlpA} and RcsF_{col} strains. Activation levels were comparable to those of RcsF _{Δ 19-47}. Rcs activity was measured as in **b**. Results were normalized as in **b**. Box-and-whisker plots are shown from minimum to maximum with the central line marking the median and the first and third quartiles shown as whiskers; $n = 6$ biologically independent experiments.

and Supplementary Fig. 2b). Together, these data demonstrate that having an N-terminal disordered linker downstream of the Lol sorting signal is required to correctly target RcsF to the outer membrane. The length of the linker is important, but the sequence is not, on the condition that the linker does not fold into a defined secondary structure.

NATURE CHEMICAL BIOLOGY | www.nature.com/naturechemicalbiology

A disordered linker is needed for optimal processing by Lol. Our finding that N-terminal disordered linkers function as molecular determinants of the targeting of lipoproteins to the outer membrane raised the question of whether these linkers work in a Lol-dependent or Lol-independent manner. To address this mechanistic question, we tested the impact of deleting *lpp* on the targeting

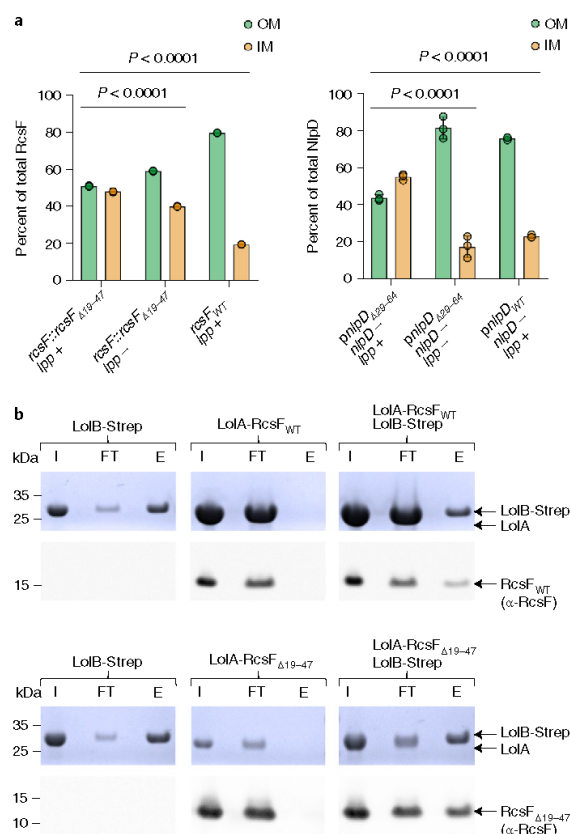


Fig. 4 | N-terminal disordered linkers interact with the Lol system to target lipoproteins to the outer membrane. **a**, Deleting *Lpp* rescues targeting of RcsF_{Δ19-47} and NlpD_{Δ29-64} to the outer membrane. The outer and inner membranes were separated by centrifugation in a sucrose density gradient (Methods). RcsF_{Δ19-47} and NlpD_{Δ29-64} accumulate in the inner membrane of cells expressing *Lpp*, the most abundant Lol substrate. Deletion of *lpp* improves their targeting to the outer membrane. Aligned dot plots are shown with mean $\pm$ standard deviation of $n = 3$ biologically independent experiments. The significance of differences among bacterial strains was assessed according to analysis of variance (ANOVA), followed by the application of Tukey's multiple comparison test. No adjustments were made for multiple comparison. RcsF variants were expressed from the chromosome and NlpD variants were expressed from a plasmid. **b**, In vitro pull-down experiments show that RcsF_{WT} and RcsF_{Δ19-47} are transferred from LolA to LolB. LolA-RcsF_{WT} and LolA-RcsF_{Δ19-47} complexes were obtained by LolA-His affinity chromatography followed by size exclusion chromatography (Methods). Each complex was incubated with LolB-Strep that had been previously purified via Strep-Tactin affinity chromatography (Methods). Both RcsF variants were eluted in complex with LolB-strep, whereas LolA was only present in the flow through. I, input; FT, flow through; E, eluate. $n = 3$ biologically independent experiments.

of RcsF_{Δ19-47}. The lipoprotein *Lpp*, also known as the Braun lipoprotein, covalently tethers the outer membrane to the peptidoglycan and controls the size of the periplasm^{33,34}. Being expressed at ~ 1 million copies per cell³⁵, *Lpp* is numerically the most abundant protein in *E. coli*. Thus, by deleting *lpp*, we considerably decreased the load on the Lol system by removing its most abundant substrate. Remarkably, *lpp* deletion partly rescued the targeting of RcsF_{Δ19-47}

to the outer membrane (Fig. 4a), indicating that the linker functions in a Lol-dependent manner and suggesting that accumulation of RcsF_{Δ19-47} in the inner membrane results from a decreased ability of the Lol system to process the linker-less RcsF variant. In agreement with this idea, normal targeting of NlpD_{Δ29-64} to the outer membrane was restored in cells lacking *Lpp* (Fig. 4a). Pal_{Δ26-56} could not be tested because membrane fractionation failed with *lpp pal* double-mutant cells whether or not they expressed Pal_{Δ26-56} (the *lpp pal* double mutant is viable but highly mucoid).

To obtain further insights into the mechanism at play here, we next monitored whether linker deletion impacted the transfer of RcsF from LolA to LolB in vitro. LolA with a C-terminal His-tag was expressed in the periplasm of cells expressing wild-type RcsF or RcsF_{Δ19-47} and purified by near homogeneity by affinity chromatography (Methods and Extended Data Fig. 6). Both RcsF_{WT} and RcsF_{Δ19-47} were detected in immunoblots of the fractions containing purified LolA (Extended Data Fig. 6), indicating that both proteins form a soluble complex with LolA and confirming that they use this chaperone for transport across the periplasm. LolB was expressed as a soluble protein in the cytoplasm and purified by taking advantage of a C-terminal Strep-tag. LolB, which has more affinity for lipoproteins than LolA³⁶, was then incubated with LolA-RcsF_{WT} or LolA-RcsF_{Δ19-47} and pulled down using Strep-Tactin beads (Methods). As both RcsF_{WT} and RcsF_{Δ19-47} were detected in the LolB-containing pulled-down fractions (Fig. 4b), we conclude that both proteins were transferred from LolA to LolB. Thus, the linker is not required for the transfer of RcsF from LolA to LolB.

We then focused on the LolCDE ABC transporter in charge of extracting outer-membrane lipoproteins and transferring them to LolA. The structure of LolCDE in complex with a lipoprotein substrate was recently reported³⁷. Although most of the lipoprotein polypeptide chain is non-visible in the structure, the triacylated N-terminal cysteine could be found inside a channel located between the plane of the membrane and the periplasmic domains of LolC and LolD. We thus hypothesized that N-terminal flexible linkers might be important to limit clashes between the lipoprotein substrate and the periplasmic domains of LolC and LolD. To test whether having a disordered linker increases the efficiency of LolCDE in lipoprotein extraction, we monitored the impact of overexpressing all LolCDE components (Extended Data Fig. 7a) on the targeting of RcsF_{Δ19-47} to the outer membrane. However, we found that overexpression of LolCDE had no effect (Extended Data Fig. 7b).

Finally, we tested whether linker deletion impacted the processing of RcsF_{Δ19-47} by the enzymes involved in lipoprotein maturation (Lgt, LspA and Lnt; Supplementary Fig. 1a). We found that overexpressing Lgt, LspA and Lnt had no impact on membrane targeting (Extended Data Fig. 8). Furthermore, the immature forms of RcsF_{Δ19-47} that accumulate when these enzymes are depleted (Extended Data Fig. 9) were not detected when these enzymes are expressed, indicating that deletion of the linker does not block the processing of RcsF_{Δ19-47} by the enzymes of the maturation pathway. Thus, taken together, our results suggest that retention of RcsF_{Δ19-47} in the inner membrane does not result from the impairment of a specific step, but rather from less efficient processing of the truncated lipoprotein by the entire lipoprotein maturation pathway (Discussion).

Are linkers important for surface exposure? We noted that two of the lipoproteins under investigation here, Pal and RcsF, have been reported to be surface-exposed^{27,38,39}. A topology model has been proposed to explain how RcsF reaches the surface, where the lipid moiety of RcsF is anchored in the outer leaflet of the outer membrane while the N-terminal linker is exposed on the cell surface before being threaded through the lumen of β -barrel proteins³⁹. Thus, in this topology, the linker allows RcsF to cross the outer

membrane. It was therefore tempting to speculate that N-terminal disordered linkers may be used by lipoproteins as a structural device to cross the outer membrane and reach the cell surface. In addition, the accumulation of RcsF $_{\Delta 19-47}$ in the inner membrane (Fig. 2a) also suggests that Lol may be using N-terminal linkers to recognize lipoproteins destined to the cell surface before their extraction from the inner membrane to optimize their targeting to the machinery exporting them to their final destination (β -Barrel Assembly Machinery (BAM) in the case of RcsF 27,39,40). To further investigate the link between disordered linkers and surface exposure, we tested whether NlpD could be detected on the cell surface, which turned out to be the case (Extended Data Fig. 10a). In addition, we found that the N-terminal linker of BamC, a lipoprotein with access to the cell surface 41,42 , is important for normal outer-membrane targeting (Extended Data Fig. 10b,c).

Discussion

Lipoproteins are crucial for essential cellular processes such as envelope assembly and virulence. However, despite their functional importance and their potential as targets for new antibacterial therapies, we only have a vague understanding of the molecular factors that control their biogenesis. By discovering the role played by N-terminal disordered linkers in lipoprotein sorting, this study adds an important new layer to our comprehension of lipoprotein biogenesis in Gram-negative bacteria. Critically, it also indicates that the current model of lipoprotein sorting—that sorting between the two membranes is controlled by the two or three residues that are adjacent to the lipidated cysteine—needs to be revised. Lipoproteins with unstructured linkers at their N terminus are commonly found in Gram-negative bacteria, including many pathogens (see below), and further work will be required to determine whether these linkers control lipoprotein targeting in organisms other than *E. coli*, laying the foundation for designing new antibiotics.

It was previously shown that both *lolA* and *lolB* (but not *lolCDE*) can be deleted under specific conditions 21 , suggesting at least one alternate route for the transport of lipoproteins across the periplasm and their delivery to the outer membrane. During this investigation, we envisaged the possibility that the linker could be required to transport lipoproteins via a yet-to-be-identified pathway independent of LolA/LolB. However, our observations that both RcsF and RcsF $_{\Delta 19-47}$ were found in complex with LolA (Extended Data Fig. 6) and were transferred by LolA to LolB (Fig. 4b) do not support this hypothesis. Instead, our data clearly indicate that lipoproteins with N-terminal linkers still depend on the Lol system for extraction from the inner membrane and transport to the outer membrane (Supplementary Fig. 1a). They also suggest that N-terminal linkers improve lipoprotein processing by Lol (see below).

The general view is that outer-membrane lipoproteins remain oriented toward the periplasm. Excitingly, a series of reports have started to challenge this paradigm by identifying surface-exposed lipoproteins in *E. coli* and other bacteria 22,23 . By showing that four surface-exposed lipoproteins (RcsF, Pal, NlpD and BamC) have an N-terminal linker that is important for outer-membrane targeting, our results also support a model in which N-terminal disordered linkers are important for displaying lipoproteins on the cell surface. Investigating whether a dedicated Lol-dependent route exists for surface-exposed lipoproteins will be the subject of future research. It is worth noting that N-terminal linkers are commonly found in lipoproteins expressed by the pathogens *Borrelia burgdorferi* and *Neisseria meningitidis* 23,43 , and lipoprotein surface exposure is common in these pathogens.

Our work also delivers crucial insights into the functional importance of disordered segments in proteins in general. Most proteins are thought to present portions that are intrinsically disordered. For example, it is estimated that 30–50% of eukaryotic proteins contain regions that do not adopt a defined secondary structure

in vitro 44 . However, demonstrating that these unstructured regions are functionally important in vivo is challenging. By showing that an N-terminal disordered segment downstream of the Lol signal is required for the correct sorting of lipoproteins, our work provides direct evidence that evolution has selected intrinsic disorder by function.

In conclusion, the data reported here establish that the triage of lipoproteins between the inner and outer membranes are not solely controlled by the Lol sorting signal; additional molecular determinants, such as protein intrinsic disorder, are also involved. Our data further highlight the previously unrecognized heterogeneity of the important lipoprotein family and call for a careful evaluation of the maturation pathways of these lipoproteins.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-021-00845-z>.

Received: 9 December 2020; Accepted: 29 June 2021;

Published online: 29 July 2021

References

- Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, a000414 (2010).
- Weiner, J. H. & Li, L. Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis. *Biochim. Biophys. Acta* **1778**, 1698–1713 (2008).
- Ricci, D. P. & Silhavy, T. J. Outer membrane protein insertion by the β -barrel assembly machine. *EcoSal Plus* <https://doi.org/10.1128/ecosalplus.ESP-0035-2018> (2019).
- El Rayes, J., Rodríguez-Alonso, R. & Collet, J. F. Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles. *Curr. Opin. Microbiol.* **61**, 25–34 (2021).
- Chimalakonda, G. et al. Lipoprotein LptE is required for the assembly of LptD by the β -barrel assembly machine in the outer membrane of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **108**, 2492–2497 (2011).
- Sherman, D. J. et al. Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge. *Science* **359**, 798–801 (2018).
- Malinverni, J. C. et al. YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol. Microbiol.* **61**, 151–164 (2006).
- Laloux, G. & Collet, J. F. Major Tom to ground control: how lipoproteins communicate extra-cytoplasmic stress to the decision center of the cell. *J. Bacteriol.* <https://doi.org/10.1128/jb.00216-17> (2017).
- Kovacs-Simon, A., Titball, R. W. & Michell, S. L. Lipoproteins of bacterial pathogens. *Infect. Immun.* **79**, 548–561 (2011).
- Szewczyk, J. & Collet, J. F. The journey of lipoproteins through the cell: one birthplace, multiple destinations. *Adv. Micro. Physiol.* **69**, 1–50 (2016).
- Babu, M. M. et al. A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J. Bacteriol.* **188**, 2761–2773 (2006).
- Narita, S. I. & Tokuda, H. Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim. Biophys. Acta Mol. Cell. Biol. Lipids* **1862**, 1414–1423 (2017).
- Horler, R. S., Butcher, A., Papangelopoulos, N., Ashton, P. D. & Thomas, G. H. EchoLOCATION: an in silico analysis of the subcellular locations of *Escherichia coli* proteins and comparison with experimentally derived locations. *Bioinformatics* **25**, 163–166 (2009).
- Tokuda, H. & Matsuyama, S. Sorting of lipoproteins to the outer membrane in *E. coli*. *Biochim. Biophys. Acta* **1694**, IN1–9 (2004).
- Gennity, J. M. & Inouye, M. The protein sequence responsible for lipoprotein membrane localization in *Escherichia coli* exhibits remarkable specificity. *J. Biol. Chem.* **266**, 16458–16464 (1991).
- Terada, M., Kuroda, T., Matsuyama, S. I. & Tokuda, H. Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of *Escherichia coli*. *J. Biol. Chem.* **276**, 47690–47694 (2001).
- Hara, T., Matsuyama, S. & Tokuda, H. Mechanism underlying the inner membrane retention of *Escherichia coli* lipoproteins caused by Lol avoidance signals. *J. Biol. Chem.* **278**, 40408–40414 (2003).

18. Narita, S. & Tokuda, H. Amino acids at positions 3 and 4 determine the membrane specificity of *Pseudomonas aeruginosa* lipoproteins. *J. Biol. Chem.* **282**, 13372–13378 (2007).
19. Lewenza, S., Mhlanga, M. M. & Pugsley, A. P. Novel inner membrane retention signals in *Pseudomonas aeruginosa* lipoproteins. *J. Bacteriol.* **190**, 6119–6125 (2008).
20. Lorenz, C., Dougherty, T. J. & Lory, S. Correct sorting of lipoproteins into the inner and outer membranes of *Pseudomonas aeruginosa* by the *Escherichia coli* LolCDE transport system. *mBio* **10**, e00194-19 (2019).
21. Grabowicz, M. & Silhavy, T. J. Redefining the essential trafficking pathway for outer membrane lipoproteins. *Proc. Natl Acad. Sci. USA* **114**, 4769–4774 (2017).
22. Wilson, M. M. & Bernstein, H. D. Surface-exposed lipoproteins: an emerging secretion phenomenon in Gram-negative bacteria. *Trends Microbiol.* **24**, 198–208 (2016).
23. Zuckert, W. R. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim. Biophys. Acta* **1843**, 1509–1516 (2014).
24. Pride, A. C., Herrera, C. M., Guan, Z., Giles, D. K. & Trent, M. S. The outer surface lipoprotein VolA mediates utilization of exogenous lipids by *Vibrio cholerae*. *mBio* **4**, e00305-13 (2013).
25. Gonnet, P., Rudd, K. E. & Lisacek, F. Fine-tuning the prediction of sequences cleaved by signal peptidase II: a curated set of proven and predicted lipoproteins of *Escherichia coli* K-12. *Proteomics* **4**, 1597–1613 (2004).
26. Sueki, A., Stein, E., Savitski, M. M., Selkrig, J. & Typas, A. Systematic localization of *Escherichia coli* membrane proteins. *mSystems* **5**, 00808-19 (2020).
27. Cho, S. H. et al. Detecting envelope stress by monitoring β -barrel assembly. *Cell* **159**, 1652–1664 (2014).
28. Uehara, T., Parzych, K. R., Dinh, T. & Bernhardt, T. G. Daughter cell separation is controlled by cytotkinetic ring-activated cell wall hydrolysis. *EMBO J.* **29**, 1412–1422 (2010).
29. Gerding, M. A., Ogata, Y., Pecora, N. D., Niki, H. & de Boer, P. A. The trans-envelope Tol–Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol. Microbiol.* **63**, 1008–1025 (2007).
30. Hussein, N. A., Cho, S. H., Laloux, G., Siam, R. & Collet, J. F. Distinct domains of *Escherichia coli* LgaA connect envelope stress sensing and down-regulation of the Rcs phosphorelay across subcellular compartments. *PLoS Genet.* **14**, e1007398 (2018).
31. Tsang, M. J., Yakhnina, A. A. & Bernhardt, T. G. NlpD links cell wall remodeling and outer membrane invagination during cytokinesis in *Escherichia coli*. *PLoS Genet.* **13**, e1006888 (2017).
32. Shrivastava, R., Jiang, X. & Chng, S. S. Outer membrane lipid homeostasis via retrograde phospholipid transport in *Escherichia coli*. *Mol. Microbiol.* **106**, 395–408 (2017).
33. Cohen, E. J., Ferreira, J. L., Ladinsky, M. S., Beeby, M. & Hughes, K. T. Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane. *Science* **356**, 197–200 (2017).
34. Asmar, A. T. et al. Communication across the bacterial cell envelope depends on the size of the periplasm. *PLoS Biol.* **15**, e2004303 (2017).
35. Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
36. 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. USA* **106**, 5877–5882 (2009).
37. Tang, X. et al. Structural basis for bacterial lipoprotein relocation by the transporter LolCDE. *Nat. Struct. Mol. Biol.* **28**, 347–355 (2021).
38. Michel, L. V. et al. Dual orientation of the outer membrane lipoprotein Pal in *Escherichia coli*. *Microbiology* **161**, 1251–1259 (2015).
39. Kononova, A., Perlman, D. H., Cowles, C. E. & Silhavy, T. J. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins. *Proc. Natl Acad. Sci. USA* **111**, E4350–E4358 (2014).
40. Rodríguez-Alonso, R. et al. Structural insight into the formation of lipoprotein– β -barrel complexes. *Nat. Chem. Biol.* **16**, 1019–1025 (2020).
41. Webb, C. T. & Lithgow, T. Identification of BamC on the surface of *E. coli*. *Methods Mol. Biol.* **1329**, 215–225 (2015).
42. Webb, C. T. et al. Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *J. Mol. Biol.* **422**, 545–555 (2012).
43. Brooks, C. L., Arutyunova, E. & Lemieux, M. J. The structure of lactoferrin-binding protein B from *Neisseria meningitidis* suggests roles in iron acquisition and neutralization of host defences. *Acta Crystallogr. F Struct. Biol. Commun.* **70**, 1312–1317 (2014).
44. Bardwell, J. C. & Jakob, U. Conditional disorder in chaperone action. *Trends Biochem. Sci.* **37**, 517–525 (2012).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Bacterial growth conditions. The bacterial strains used in this study are listed in Supplementary Table 3. Bacterial cells were cultured in Luria broth (LB) at 37 °C unless stated otherwise. The following antibiotics were added when appropriate: spectinomycin (100 µg ml⁻¹), ampicillin (200 µg ml⁻¹), chloramphenicol (25 µg ml⁻¹) and kanamycin (50 µg ml⁻¹). L-arabinose (0.2%), isopropyl-β-D-thiogalactoside (IPTG; 20 µM to 1 mM) and L-fucose (0.2%) were used for induction or repression when appropriate.

Bacterial strains and plasmids. DH300 (a derivative of *E. coli* MG1655 carrying a chromosomal *rprA::lacZ* fusion at the λ attachment site⁴⁵) was used as wild type throughout the study. All deletion mutants were obtained by transferring the corresponding alleles from the Keio collection⁴⁶ (kan^r) into DH300⁴⁷ via P1 phage transduction. Deletions were verified by PCR and the absence of the protein was verified by immunoblotting (when possible). If necessary, the kanamycin cassette was removed via site-specific recombination mediated by the yeast Flp recombinase with pCP20 vector⁴⁷. All strains expressing the RcsF mutants used for subcellular fractionation lacked *rscB* to prevent induction of Rcs.

The plasmids used in this study are listed in Supplementary Table 4 and the primers in Supplementary Table 5. RcsF, Pal and NlpD were expressed from the low-copy vector pAM238⁴⁸ containing the SC101 origin of replication and the *lac* promoter. To produce pSC202 for RcsF expression, *rscF* (including ~30 base pairs upstream of the coding sequence) was amplified by PCR from the chromosome of DH300 (primer pair SH_RcsF(PstI)-R and SH_RcsFU-R (KpnI)-F). The amplification product was digested with KpnI and PstI and inserted into pAM238, resulting in pSC202. *nlpD* was amplified using primers JR1 and JR2 and *pal* was amplified with primers JS145 and JS146. Amplification products were digested with PstI-XbaI and KpnI-XbaI, respectively, generating pJR8 (for NlpD expression) and pJS20 (for Pal expression). To clone *rscF*_{Δ19-47}, the nucleotides encoding the RcsF signal sequence were amplified using primers SH_RcsFU-R (KpnI)-F and SH_RcsFss-Fsg (NcoI)-R, and those encoding the RcsF signaling domain were amplified using primers SH_RcsFss-Fsg (NcoI)-R and SH_RcsF(PstI)-R. In both cases, pSC202 was used as template. Overlapping PCR was then performed using SH_RcsFUR(kpnI)-F and SH_RcsF(PstI)-R from the two PCR products previously obtained. The final product was digested with KpnI and PstI, and ligated with pAM238 pre-digested with the same enzymes, yielding pSC201. To add a GS linker (Ser-Gly-Ser-Gly-Ser-Gly-Ala-Met) into pSC201, primers SH_GS linker_F and SH_GS linker_R were mixed, boiled, annealed at room temperature, and ligated with pSC201 pre-digested with NcoI, generating pSC198. pSC199 was generated similarly, but using primers SH_GS linker_F and SH_GS linker_R and plasmid pSC198. pSC200 was generated using primers SH_Da linker_F and SH_GS linker_R and plasmid pSC199.

*rscF*_{Δ19-47} and *rscF*_{Δ1} were obtained by inserting DNA sequences corresponding to helical linker fragments (FkpA Ser94-Glu125 and colicin IA Ile213-Lys282) into *rscF*_{Δ19-47} at NcoI and RsrII restriction sites. The *fkpA* gene fragment was amplified from the *E. coli* MC4100 chromosome (JS50/JS51 primers) and the *cia* gene fragment was chemically synthesized as a gene block by Integrated DNA Technologies (IDT). The resulting plasmids were pJS18 and pJS27, respectively. pAM238 does not contain the *lacIq* repressor. Therefore, to enable expression-level regulation by IPTG, strains containing the pAM238 plasmids expressing RcsF variants were co-transformed with pET22b, a high-copy plasmid from a different incompatibility group (pBR223 origin of replication; Novagen) containing the *lacIq* repressor. Chromosomal insertion of *rscF*_{Δ19-47} or *rscF*_{Δ1} at the *rscF* locus was performed via λ-Red recombineering⁴⁹ with pSIM5-Tet plasmid (a gift of D. Hughes). In the first step, the *cat-sacB* cassette was introduced and later replaced by mutant *rscF*. The same technique was used to add a human influenza hemagglutinin (HA) tag to the chromosomal *fkpA* gene.

The *pal* allele lacking the linker region (*pal*_{Δ26-54}) was created using overlapping PCR. The pJS20 plasmid served as template for PCR with the M13R/M13F external primers and JS152/JS153 internal primers. The truncated allele was cloned into pAM238 at the same restriction sites as the full-length allele, producing pJS24. The *nlpD* allele lacking the linker regions (*nlpD*_{Δ29-44}) was created by overlapping PCR. *Escherichia coli* chromosomal *nlpD* served as template for the PCR, with JR1/JR2 as external primers and JR7/JR8 as internal primers. The truncated allele was then cloned into pAM238 at the same restriction sites as the full-length allele, producing pJR10.

To clone *bamC*_{WT} with a C-terminal Flag-tag, the chromosomal *bamC* gene was amplified by PCR using primers JR92 and JR93 and then inserted into pSC213, a modified IPTG-regulated pAM238 with a ribosome binding site and a C-terminal Flag-tag, previously digested with the restriction enzymes NcoI and XbaI, using NEB Gibson Master Mix. To clone *bamC*_{Δ28-38} with a C-terminal Flag-tag using pSC213, *bamC*_{Δ1-27} and *bamC*_{Δ99-344} were amplified by PCR using the chromosomal *bamC* gene and primers JR92/JR94 and JR95/JR96, respectively. pSC213 was digested using restriction enzymes NcoI and XbaI. *bamC*_{Δ28-38} was then created by mixing PCR products (*bamC*_{Δ1-27} and *bamC*_{Δ99-344}) with digested pSC213 and NEB Gibson Master Mix. *BamC*_{WT} and *BamC*_{Δ28-38} were induced using 20 µM and 40 µM of IPTG, respectively (Extended Data Fig. 10).

The chromosomal *loIEDE* operon was amplified via PCR using primers JS277 and JS278 (adding a C-terminal His-tag to LoIE) and then inserted into pBAD33

using the restriction sites PstI and XbaI, resulting in pJR203. The expression level of LoIE-His was verified by immunoblotting.

The sequence encoding *loIB* without its N-terminal cysteine was first amplified from the chromosome via PCR using primers JR50/PL387 (adding a C-terminal Strep-tag). It was then cloned into pET28a using the restriction sites XbaI and PstI, generating plasmid pJR90. *loIA* was amplified using chromosomal *loIA* as PCR template for primers JR30/JR31 (JR31 contains the sequence of a His-tag) and then cloned into pBAD18 using KpnI and XbaI, resulting in pJR48.

The genes encoding Lgt and Lnt were amplified from the chromosome with PCR primers AG389/AG403 and AG393/JR74, respectively. AG403 and JR74 also encode a Myc-tag. PCR products were cloned into pAM238 using KpnI and PstI. Expression levels of the proteins constitutively expressed from the plasmid were verified via immunoblotting (Extended Data Fig. 8). *lspA* was amplified with PCR primers JR77/JR78. The PCR product was cloned into pSC213, a modified IPTG-regulated pAM238 with a ribosome binding site and a C-terminal Flag-tag, using NcoI and XbaI. Expression of LspA-Flag was induced by adding 25 µM IPTG.

Cell fractionation and sucrose density gradients. Cell fractionation was performed as described previously⁵⁰ with some modifications. Cells (400 ml) were grown at 37 °C in LB medium until the optical density at 600 nm (OD₆₀₀) of the culture reached 0.7. Cells were collected by centrifugation at 6,000g at 4 °C for 15 min, washed with TE buffer (50 mM Tris-HCl pH 8, 1 mM EDTA) and resuspended in 20 ml of the same buffer. The washing step was skipped with the *Δlpp* strains to prevent the loss of outer-membrane vesicles. DNase I (1 mg, Roche), RNase A (1 mg, Thermo Scientific) and a half tablet of a protease inhibitor cocktail (cComplete EDTA-free Protease Inhibitor Cocktail tablets; Roche) were added to cell suspensions, and the cells were passed through a French pressure cell at 1,500 p.s.i. After adding MgCl₂ to a final concentration of 2 mM, the lysate was centrifuged at 5,000g at 4 °C for 15 min to remove cell debris. Then, 16 ml of supernatant was placed on top of a two-step sucrose gradient (2.3 ml of 2.02 M sucrose in 10 mM HEPES pH 7.5 and 6.6 ml of 0.77 M sucrose in 10 mM HEPES pH 7.5). The samples were centrifuged at 180,000g for 3 h at 4 °C in a 55.2 Ti Beckman rotor. After centrifugation, the soluble fraction and the membrane fraction were collected. The membrane fraction was diluted four times with 10 mM HEPES pH 7.5. To separate the inner and outer membranes, 7 ml of the diluted membrane fraction was loaded on top of a second sucrose gradient (10.5 ml of 2.02 M sucrose, 12.5 ml of 1.44 M sucrose, 7 ml of 0.77 M sucrose, always in 10 mM HEPES pH 7.5). The samples were then centrifuged at 112,000g for 16 h at 10 °C in a SW 28 Beckman rotor. Approximately 30 fractions of 1.5 ml were collected and odd-numbered fractions were subjected to SDS-PAGE, transferred onto a nitrocellulose membrane and probed with specific antibodies. Graphs were created in GraphPad Prism 9 via spline analysis of the curves representing a mean of three independent experiments. Aligned dot plots representing the outer and inner membrane percentages were created in GraphPad Prism 9 by measuring the area under the curve for each experiment.

Immunoblotting. Protein samples were separated via 10% or 4–12% SDS-PAGE (Life Technologies) and transferred onto nitrocellulose membranes (GE Healthcare Life Sciences). The membranes were blocked with 5% skimmed milk in 50 mM Tris-HCl pH 7.6, 0.15 M NaCl and 0.1% Tween 20 (TBS-T). TBS-T was used in all subsequent immunoblotting steps. The primary antibodies were diluted 5,000 to 20,000 times in 1% skimmed milk in TBS-T and incubated with the membrane for 1 h at room temperature. The anti-RcsF, anti-DsbD, anti-Lpp and anti-NlpD antisera were generated by our laboratory. These antibodies were used at 1/20,000, 1/9,000, 1/7,000 and 1/10,000, respectively. Anti-Pal, used at 1/20,000, was a gift from R. Lloubès and anti-His, used at 1/10,000, is a peroxidase-conjugated antibody (Qiagen). Anti-HA (Biolegend MMS-101R), used at 1/1,000, and anti-Flag (Sigma), used at 1/10,000, are mouse antisera, whereas anti-Myc (Santa Cruz Biotechnology), used at 1/7,000, is a rabbit antiserum. The membranes were incubated for 1 h at room temperature with horseradish peroxidase (HRP)-conjugated goat anti-rabbit immunoglobulin-G or anti-mouse (Sigma) at a 1:10,000 dilution. Labeled proteins were detected via enhanced chemiluminescence (Pierce enhanced chemiluminescence (ECL) western blotting substrate, Thermo Scientific) and visualized using X-ray film (Fuji) or a camera (Image Quant LAS 4000 and Vilber Fusion solo S). To quantify proteins levels, band intensities were measured using ImageJ version 1.46r (National Institutes of Health).

Testing the surface exposure of NlpD using dot blots. The dot-blot method was adapted from ref. 27. Cells were grown in LB medium to an OD₆₀₀ of 0.5. Cells were directly fixed for 15 min at room temperature with 0.04% glutaraldehyde and 2% formaldehyde. 2 × 1 ml of cells were centrifuged for 1 min at 16,000g. Fixed cells were resuspended in PBS. For permeabilization, cells were incubated for 30 min at 4 °C in a buffer containing lysozyme (1 mg ml⁻¹) and EDTA (10 mM). After centrifugation, they were resuspended in PBS. Cells (2 ml) were then spotted on nitrocellulose membranes and the spots were left to dry completely. The membranes were blocked for 30 min with 5% (wt/vol) skimmed milk in TBS. Membranes were washed for 5 min in TBS-Tween 20 0.05% (TBS-T 0.05%) and then incubated for 1 h at room temperature with primary antibody solutions. The anti-NlpD antibody was diluted 3,000 times with 1% skimmed milk in TBS. Anti-HA mouse antibody

(Biolegend MMS-101R) was diluted 1,000 times in BSA 2% in TBS. Membranes were washed three times for 5 min in TBS-T 0.05% and incubated for 1 h at room temperature with secondary antibody solutions: anti-rabbit-HRP (Cytiva NA934) or anti-mouse-HRP (Sigma A4416) diluted 2,000 times with skimmed milk 5% in TBS. Membranes were washed three times for 5 min in TBS-T 0.05%. Labeled proteins were detected with standard ECL or Immobilon HRP substrate (Millipore) and visualized with a charge-coupled device camera.

β -galactosidase assay. β -galactosidase activity was measured as described previously²¹. Briefly, cells harboring *PrpA-lacZ* at the *attB* phage lambda site on the chromosome were diluted 1:100 from overnight cultures in LB medium, then incubated at 37°C and collected at OD₆₀₀ = 0.6. Cells (20 μ l) were collected and incubated with 80 μ l of permeabilization solution (60 mM Na₂HPO₄·2H₂O, 40 mM NaH₂PO₄·H₂O, 10 mM KCl, 1 mM MgSO₄·7H₂O, 50 mM β -mercaptoethanol) for 45 min at room temperature. Then, 600 μ l of substrate (1 mg ml⁻¹ *o*-nitrophenyl- β -D-galactoside, 50 mM β -mercaptoethanol) were added. The mixture was further incubated at 30°C for 20–90 min. The reaction was then stopped by adding 700 μ l of 1 M Na₂CO₃ and absorbance was measured at 420 nm. The standardized amount of galactosidase activity was reported in Miller units. The ratio of *PrpA-lacZ* induction was calculated relative to the basal level in a wild-type strain. Expression-level estimations were performed as follows. Cultures used for β -galactosidase activity (0.5 ml per culture) were precipitated with 10% trichloroacetic acid, washed with ice-cold acetone and resuspended in 0.2 ml of Laemmli SDS sample buffer. Samples (5 μ l) were subjected to SDS-PAGE and immunoblotted with anti-RcsF antibody. Box-and-whisker plots were prepared using GraphPad Prism 9.

SDS-EDTA sensitivity assay. Cells were grown in LB medium at 37°C until they reached an OD₆₀₀ of 0.7. Tenfold serial dilutions were made in LB medium and plated on LB agar supplemented with spectinomycin (100 μ g ml⁻¹) when necessary. Plates were incubated at 37°C. To evaluate the sensitivity of the *pal* mutant, plates were supplemented with 0.01% SDS and 0.5 mM EDTA.

Microscopy image acquisition. Cells were grown in LB medium at 37°C until an OD₆₀₀ of 0.5. Cells growing in exponential phase were spotted onto a 1% agarose PBS pad for imaging. Images were taken at 21°C. Cells were imaged on a Nikon Eclipse Ti2-E inverted fluorescence microscope with a CFI Plan Apochromat DM Lambda $\times$ 100 oil, 1.45-numerical-aperture, 0.13-mm-working-distance objective. Images were collected on a Prime 95B 25-mm camera (Photometrics). We used a Cy5-4050C (32mm) filter cube (Nikon). Image acquisition was performed with NIS-Element Advance Research version 4.5. Whole-field pictures had 2,048 $\times$ 2,048 pixels, where 1 pixel accounts for 0.11 $\times$ 0.11 μ m. The exposure time for phase contrast images was 70 ms with 75% light intensity, using a SOLA II shutter. The display range was 5,402–11,791.

Protein purification. JR90 cells were grown in LB medium supplemented with kanamycin (50 μ g ml⁻¹) at 37°C. When the culture OD₆₀₀ reached 0.5, the expression of cytoplasmic LolB-Strep was induced with 1 mM IPTG. Cells (1 l) were pelleted when they reached an OD₆₀₀ of 3 and resuspended in 25 ml of buffer A (200 mM NaCl and 50 mM NaPi, pH 8) containing one tablet of cComplete EDTA-free Protease Inhibitor Cocktail (Roche). Cells were lysed via two passages through a French pressure cell at 1,500 p.s.i. The lysate was centrifuged at 30,000g for 40 min at 4°C in a JA 20 rotor and the supernatant was mixed with Strep-Tactin resin (IBA Lifesciences) previously equilibrated with buffer A. After washing the resin with 10 column volumes of buffer A, LolB-Strep was eluted with five column volumes of buffer A supplemented with 5 mM desthiobiotin. LolB-Strep was finally desalted using a PD10 column (GE Healthcare).

Soluble LolA-RcsF_{WT} and LolA-RcsF _{Δ 19–47} complexes were purified by affinity chromatography as follows. Cells co-expressing LolA either with wild-type RcsF (JR47) or RcsF _{Δ 19–47} (JR44) were grown in LB medium at 37°C supplemented with 200 μ g ml⁻¹ ampicillin until reaching an OD₆₀₀ of 0.5. Protein expression was then induced with 0.2% arabinose. Cells (1 l) were pelleted at an OD₆₀₀ of 3 and resuspended in 25 ml of buffer A containing one tablet of protease inhibitor cocktail. Cells were lysed by two passages through a French pressure cell at 1,500 p.s.i. The lysate was centrifuged at 45,000g for 30 min at 4°C using a 55.2 Ti Beckman rotor. To obtain the soluble fraction, the supernatant was centrifuged at 180,000g for 1 h at 4°C using the same rotor. The supernatant was added to a His Trap HP column (Merck) previously equilibrated with buffer A. The column was washed with 10 column volumes of buffer A supplemented with 20 mM imidazole and LolA-His was eluted using a gradient of imidazole (from 20 mM to 300 mM). The fractions obtained were analyzed using SDS-PAGE, and LolA was detected around 25 kDa. RcsF variants and LolA-His were detected by immunoblotting with anti-RcsF and anti-His antibodies, respectively. Fractions containing LolA-RcsF variants were pooled, concentrated to 1 ml using a Vivaspin 4 Turbo concentrator (cutoff of 5 kDa, Sartorius) and purified using size-exclusion chromatography with a Superdex S75–10/300 column (GE Healthcare).

Pulldown and transfer of RcsF variants from LolA to LolB. LolB-Strep was incubated at 30°C for 20 min under agitation with LolA-RcsF_{WT} or with

LolA-RcsF _{Δ 19–47} (LolA-RcsF_{WT} and LolA-RcsF _{Δ 19–47} complexes were purified as described above). The mixture was added to magnetic Strep beads (MagStrep type 3 beads, IBA Life Science) previously equilibrated with buffer A and incubated for 30 min at 4°C on a roller. After washing the beads with the same buffer, LolB-Strep was eluted with buffer A supplemented with 50 mM biotin. Samples were analyzed via SDS-PAGE and LolA and LolB were detected with Coomassie Brilliant Blue (Bio-Rad). RcsF was detected by immunoblotting with an anti-RcsF antibody.

Depletion of the maturation enzymes Lgt and Lnt. Lgt null strains (JR301 and JR302) expressing Lgt from an arabinose-inducible promoter on the pCHAP9231 plasmid⁵² were prepared using P1 phage transduction from the Lgt null strain PAP9403⁵³ (gift from N. Buddelmeijer). Overnight cultures were grown at 37°C in LB medium in the presence of 0.2% arabinose (to allow expression of Lgt) and chloramphenicol (pCHAP9231 has chloramphenicol resistance). Cells were then diluted in fresh LB medium and grown at 37°C until growth arrest was observed and OD₆₀₀ remained constant. Cells were then precipitated with 10% trichloroacetic acid, washed with ice-cold acetone and resuspended in tricine SDS sample buffer. Samples were subjected to SDS-PAGE using tricine 16% gel at 125 V, and immunoblotted with anti-Lpp and anti-RcsF antibodies.

The Lnt null strain PAP8504⁵³ was a kind gift from N. Buddelmeijer. In the PAP8504 strain, an arabinose-inducible promoter controls the expression of the chromosomal *lnt* gene. Overnight precultures were grown at 37°C in LB medium in the presence of 0.2% arabinose to allow the expression of Lnt. Precultures were then washed three times with fresh LB medium and diluted 1:100 in LB medium supplemented with 0.2% fucose. When growth arrest was observed and OD₆₀₀ remained constant, cells were precipitated as described previously. Samples were resuspended in Tris-glycine SDS sample buffer and subjected to SDS-PAGE using Tris-glycine 16% gel at 125 V, and immunoblotted with anti-Lpp and anti-RcsF antibodies. To monitor the impact of Lnt depletion on RcsF _{Δ 19–47} plasmid pSC201 was introduced in PAP8504, yielding strain JR320.

Structural analysis of lipoproteins. When X-ray, cryo-EM or NMR structures were available, the missing residues were completed through comparative modeling using MODELLER version 9.22⁵⁴. If no structure of the lipoprotein was available, then the most pertinent analogous structure from proteins belonging to the same PFAM group was used as template for comparative modeling. The linker was defined as the unstructured fragment from the N-terminal Cys of the mature form until the first residue with well-defined secondary structure (α -helix or β -strand) belonging to a globular domain. Short, intermediate and long linkers had lengths of <12, 12–22 and >22 residues, respectively. Images were generated using UCSF Chimera version 1.13.1⁵⁵.

Statistical methods. The significance of differences among bacterial strains was assessed using GraphPad Prism 9 according to analysis of variance (ANOVA), followed by the application of Tukey's multiple comparison test.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data availability

Source data are provided with this paper. All other data generated or analyzed during this study are included in this paper and its Supplementary Information file.

References

- Majdalani, N., Hernandez, D. & Gottesman, S. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Mol. Microbiol.* **46**, 813–826 (2002).
- Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
- Cherepanov, P. P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14 (1995).
- Gil, D. & Bouche, J. P. ColE1-type vectors with fully repressible replication. *Gene* **105**, 17–22 (1991).
- Yu, D. et al. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **97**, 5978–5983 (2000).
- 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. USA* **104**, 6400–6405 (2007).
- Miller, J. C. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, 1972).
- Pailler, J., Aucher, W., Pires, M. & Buddelmeijer, N. Phosphatidylglycerol:lipid protein diacylglycerol transferase (Lgt) of *Escherichia coli* has seven transmembrane segments, and its essential residues are embedded in the membrane. *J. Bacteriol.* **194**, 2142–2151 (2012).
- Robichon, C., Vidal-Ingigliardi, D. & Pugsley, A. P. Depletion of apolipoprotein N-acyltransferase causes mislocalization of outer membrane lipoproteins in *Escherichia coli*. *J. Biol. Chem.* **280**, 974–983 (2005).

54. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
55. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

We thank A. Boujtat for technical help. We are indebted to the members of the Collet laboratory and to Nassos Typas (EMBL, Heidelberg) for helpful suggestions and discussions and to T. Silhavy (Princeton) and N. Buddelmeijer (Pasteur) for providing bacterial strains. J.S. was a research fellow of FRIA and J.-E.C. is an Investigator of FRFS-WELBIO. This work was funded by WELBIO (grant no. WELBIO-CR-20190-03), by grants from F.R.S.-FNRS, from Fédération Wallonie-Bruxelles (ARC 17/22-087), from the European Commission via the International Training Network Train2Target (721484) and from the EOS Excellence in Research Program of FWO and F.R.S.-FNRS (G0G0818N).

Author contributions

J.-E.C., J.E.R., J.S. and S.-H.C. designed and performed the experiments. J.E.R., J.S. and S.-H.C. constructed the strains and cloned the constructs. J.-E.C., J.E.R., J.S., S.-H.C.

and A.M. analyzed and interpreted the data. B.I.I. performed the structural analysis, M.D. the microscopy analysis and N.C. the dot-blot analysis. J.-E.C., J.E.R. and J.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41589-021-00845-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41589-021-00845-z>.

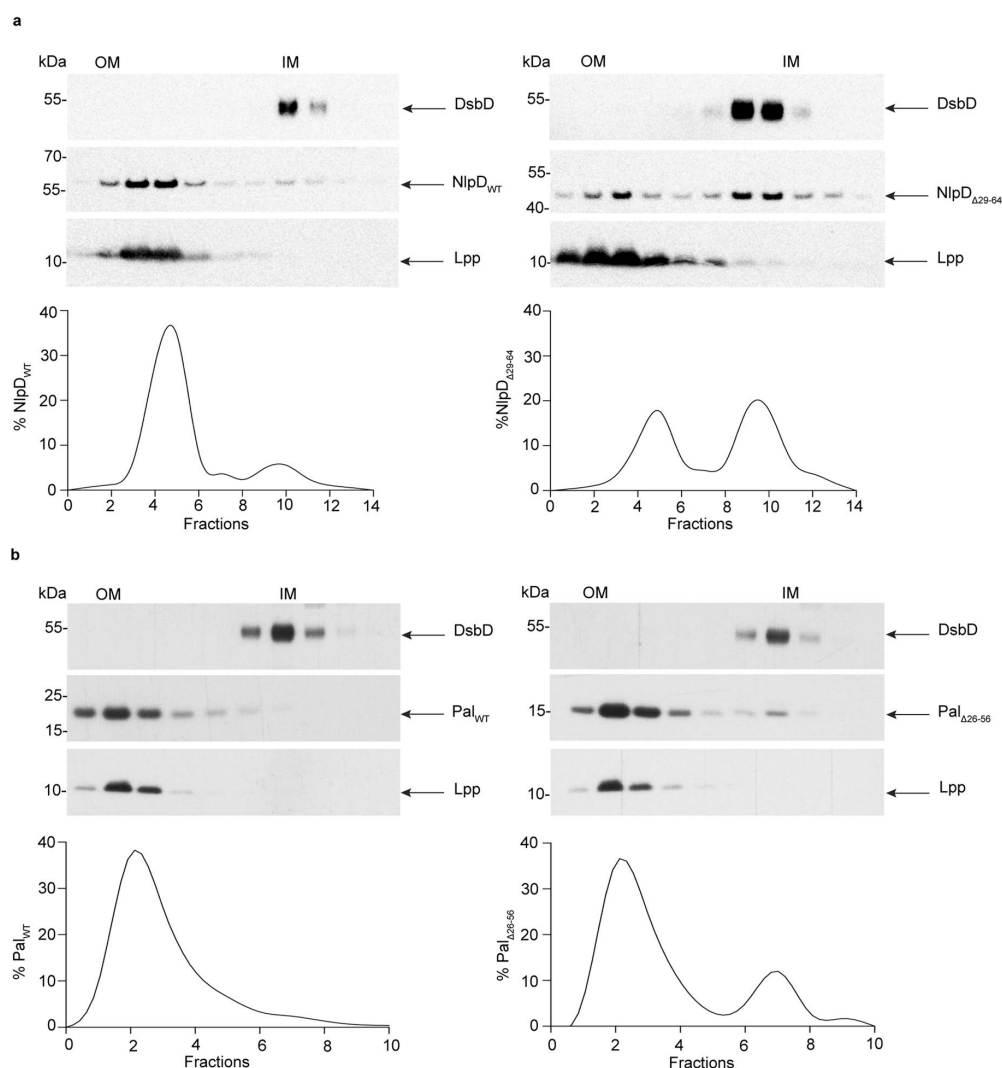
Correspondence and requests for materials should be addressed to J.-E.C.

Peer review information *Nature Chemical Biology* thanks Vladimir Uversky and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

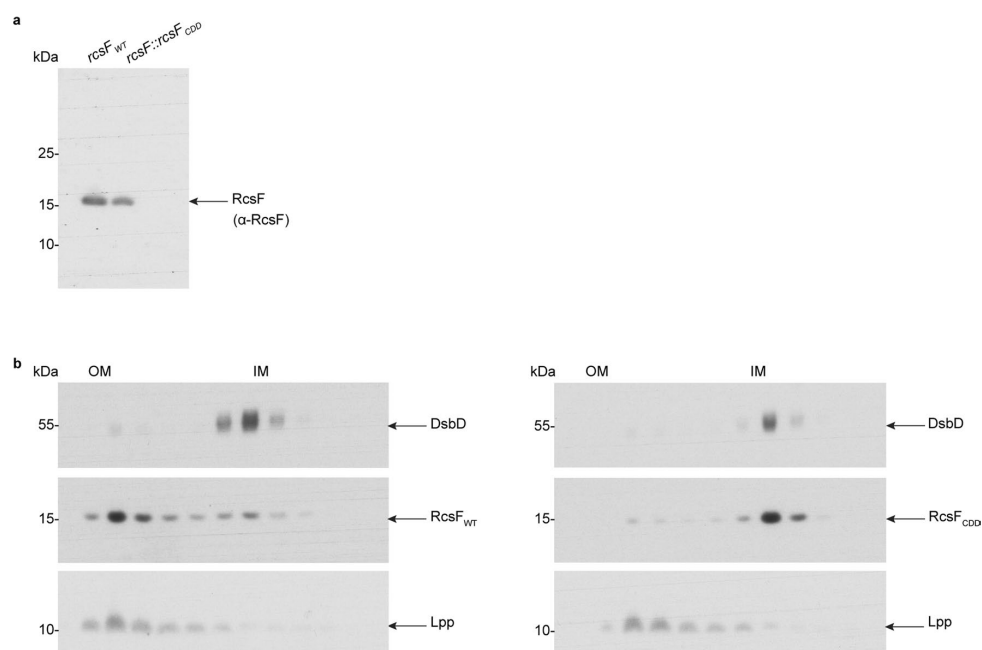
Reprints and permissions information is available at www.nature.com/reprints.



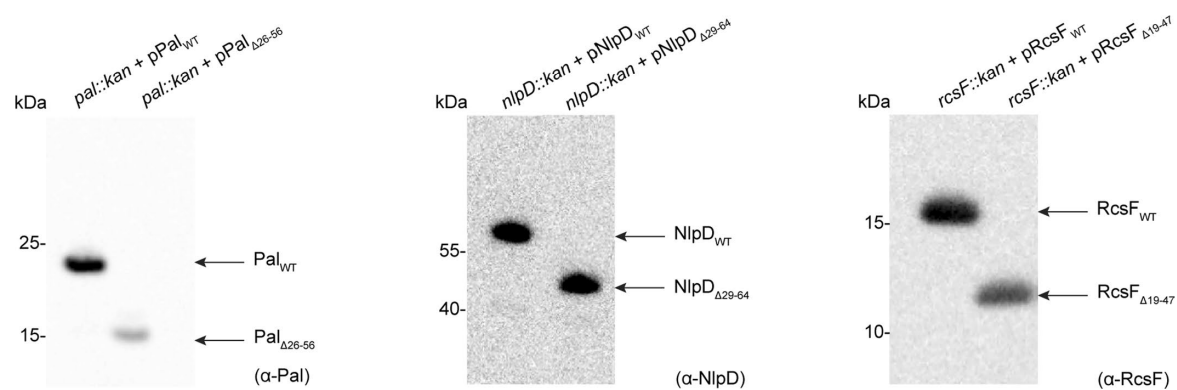
Extended Data Fig. 1 | Structural analysis of lipoproteins reveals that half of outer membrane lipoproteins display an intrinsically disordered linker at the N-terminus. Structures were generated via comparative modeling. X-ray and cryo-EM structures are green, NMR structures are cyan, and structures built via comparative modeling from the closest analog in the same PFAM group are orange. In all cases, the N-terminal linker is magenta. Lipoproteins targeting the outer membrane: OsmE, NlpC, MltB, NlpE, YajI, YcfL, NlpI, HslJ, YbaY, RipA, AmiD, BamB, MltC, YcaL, MltA, Blic, YghG, EmtA, BamD, LolB, MiaA, LolP, YraP, MliC, LpoB, CsgG, YddW, YbhC, YedD, YgeR, YfiB, YbjP, YiaD, GfcE, PqiC, YfeY, LptE, LpoA, BamE. Lipoproteins targeting the IM: DcrB, MetQ, NlpA, ApbE, YqjN, YehR. Synthetic constructs: RcsF_{G5}, RcsF_{G52}, RcsF_{G53}, RcsF_{Δ19-47}, RcsF_{FkpA}, RcsF_{col}, Pal_{Δ26-56}, BamC_{Δ29-68}, NlpD_{Δ29-64}.



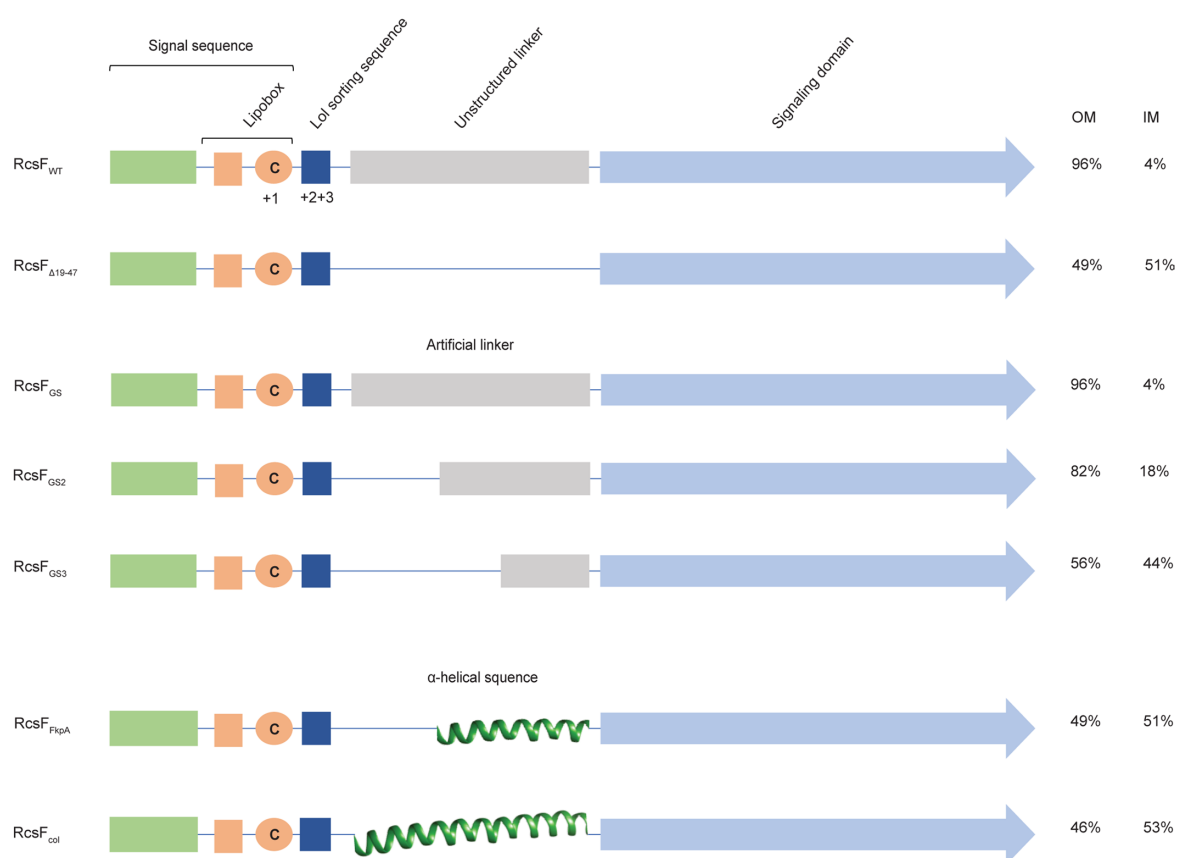
Extended Data Fig. 2 | The N-terminal linker of lipoproteins is important for outer membrane targeting. The outer membrane (OM) and inner membrane (IM) were separated via centrifugation in a three-step sucrose density gradient (Methods). While (a) NlpD_{WT} and (b) Pal_{WT} were found predominantly in the OM, NlpD_{Δ29-64} and Pal_{Δ26-56} were substantially retained in the IM. Data are presented as the ratio of signal intensity in a single fraction to the total intensity in all fractions. All variants were expressed from plasmids (Supplementary Table 4). DsbD and Lpp were used as controls for the OM and IM, respectively. Images in a, and b are representative of biological triplicates. Graphs in a and b were created by spline analysis of curves representing a mean of three independent experiments.



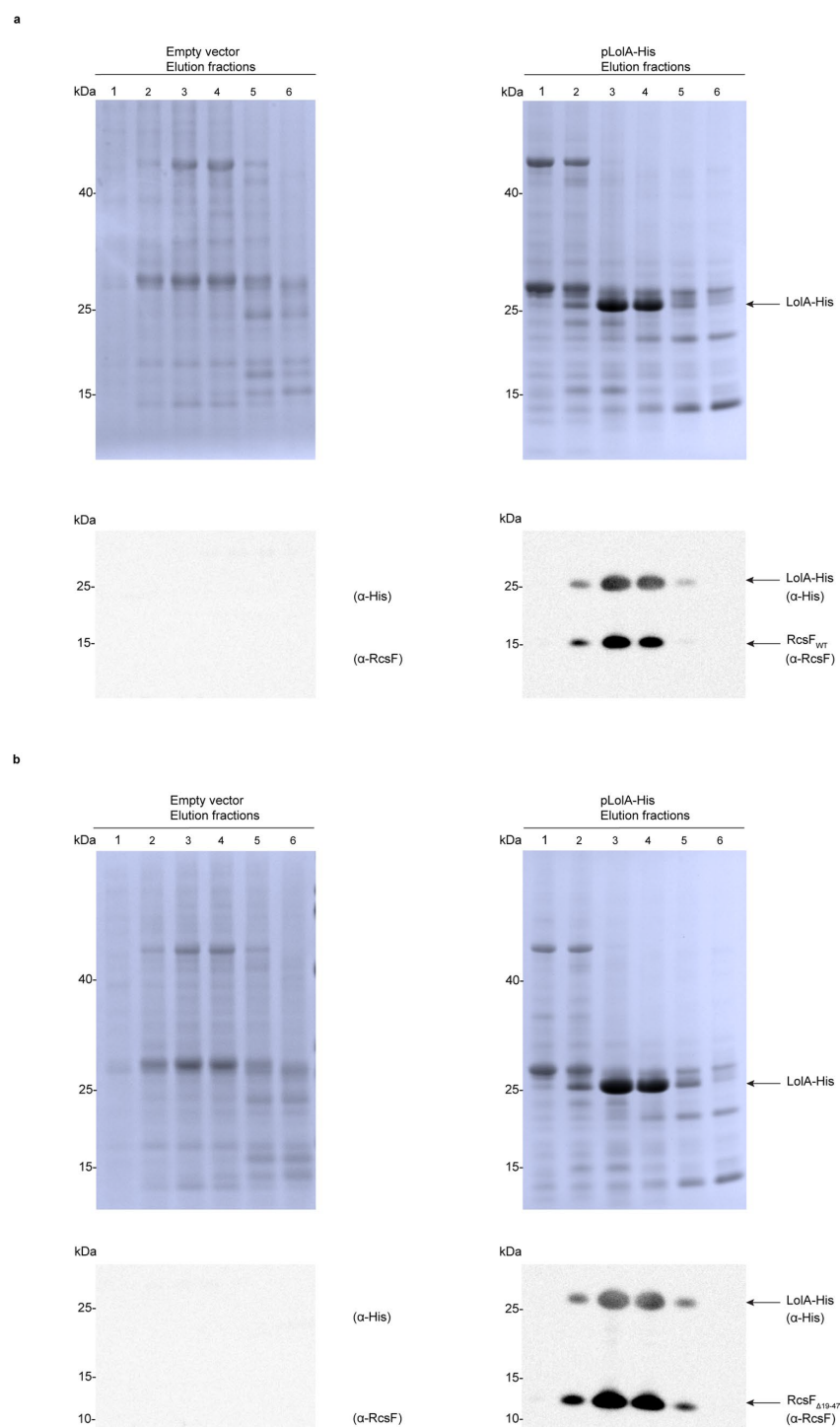
Extended Data Fig. 3 | RcsF_{CDD} is retained in the inner membrane. **a**, Expression levels of RcsF_{WT} and RcsF_{CDD}. Cells expressing chromosomal RcsF_{WT} or RcsF_{CDD} were grown in LB until OD₆₀₀ = 0.7 and precipitated with trichloroacetic acid (Methods). Immunoblots were performed using α-RcsF. **b**, The outer (OM) and inner (IM) membranes of cells expressing chromosomal RcsF_{WT} or RcsF_{CDD} were separated by sucrose density gradient (Methods). While RcsF_{WT} was found predominantly in the OM, RcsF_{CDD} was retained in the IM. Images in a and b are representative of three independent experiments.



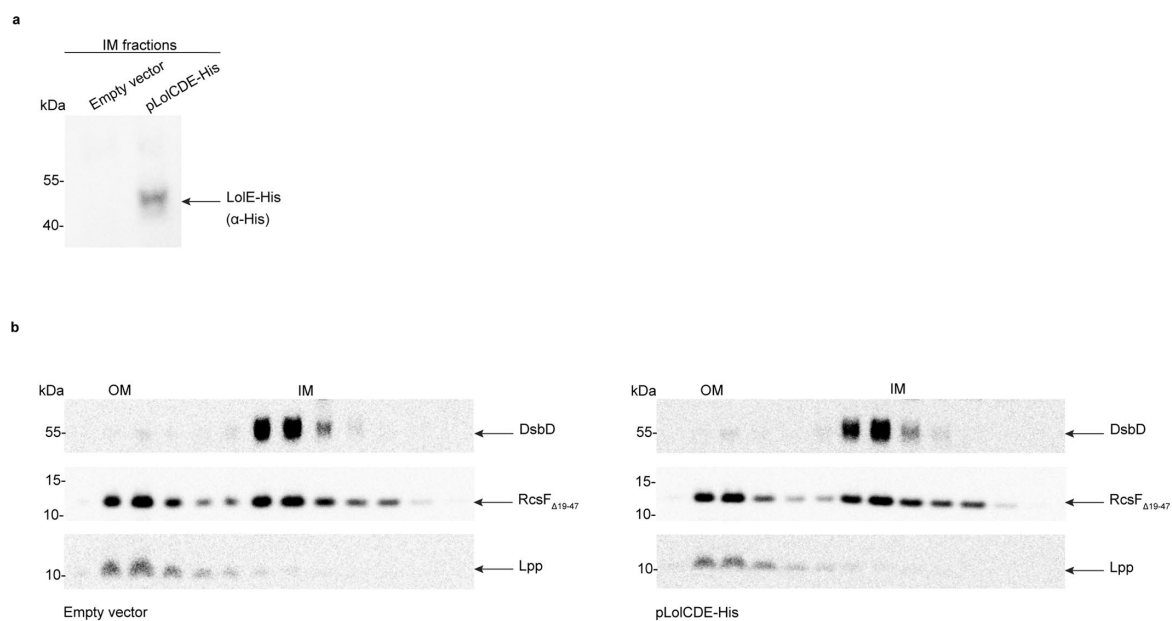
Extended Data Fig. 4 | Expression levels of RcsF_{Δ19-47}, Pal_{Δ26-56}, and NlpD_{Δ29-64}. Cells were grown at 37 °C in LB until OD₆₀₀ = 0.5 and precipitated with trichloroacetic acid (Methods). Immunoblots were performed with α-RcsF, α-NlpD, and α-Pal antibodies. All images are representative of three independent experiments.



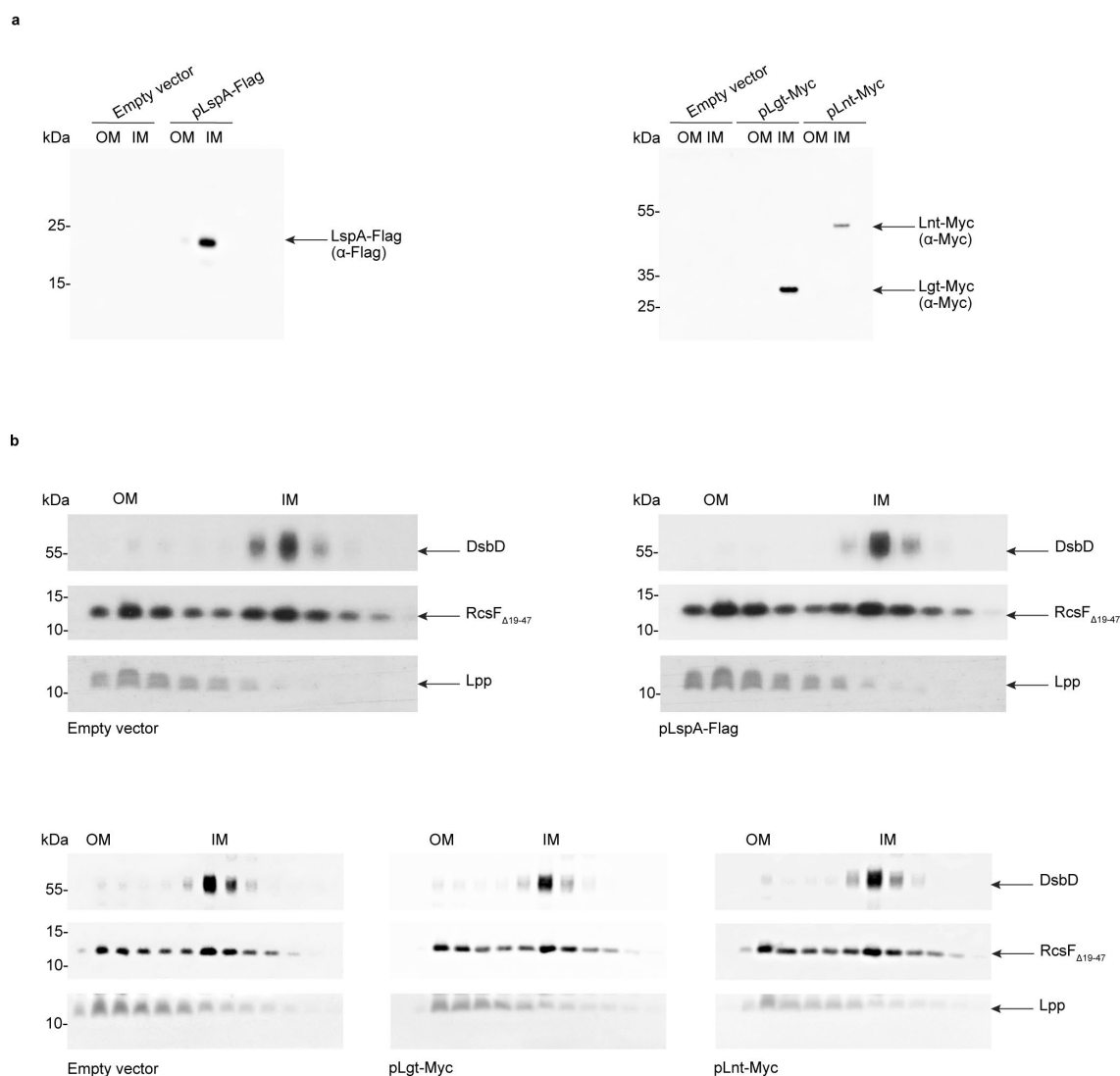
Extended Data Fig. 5 | Schematic of RcsF variants used in this study and their distributions in the outer membrane (OM) and inner membrane (IM). RcsF_{GS}, RcsF_{GS2}, and RcsF_{GS3} have linkers that are disordered and mostly consist of GS repeats. The linker of RcsF_{GS} is the same length as the linker of RcsF_{WT}. RcsF_{GS2} and RcsF_{GS3} are shorter than RcsF_{WT}. Regions of RcsF_{FkpA} and RcsF_{col} fold into α helices borrowed from the sequences of FkpA and colicin Ia, respectively.

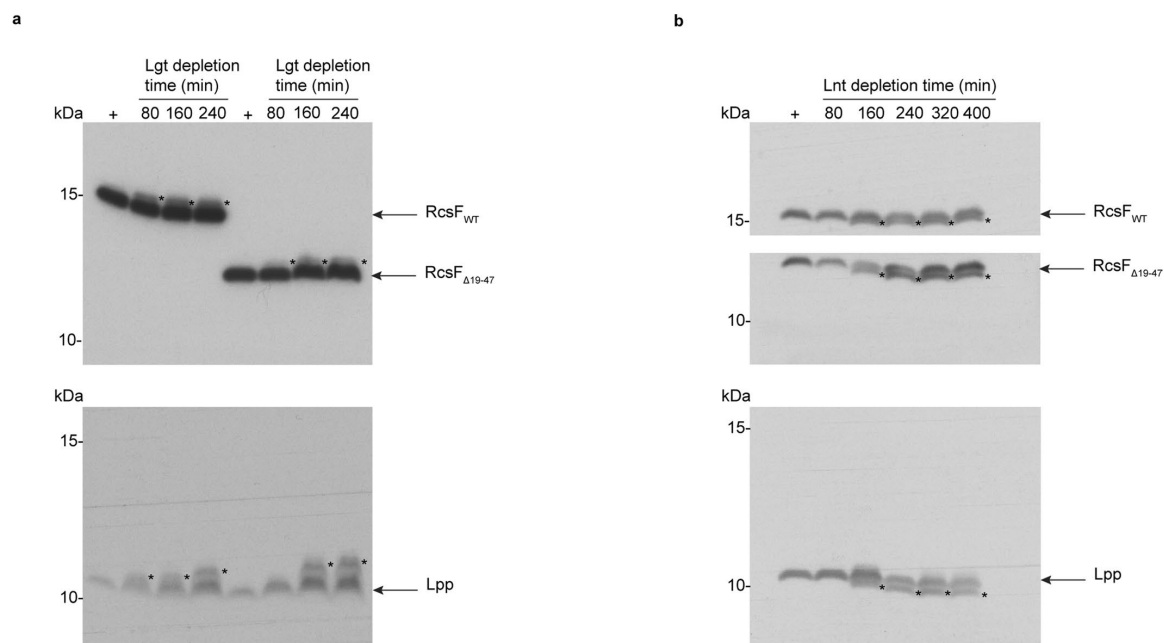


Extended Data Fig. 6 | Complexes between LolA and RcsF_{WT} or RcsF_{Δ19-47} can be purified. Both RcsF_{WT} (**a**) and RcsF_{Δ19-47} (**b**) were eluted in complex with LolA-His via affinity chromatography followed by size exclusion chromatography. Elution fractions were analyzed via SDS-PAGE and proteins stained with Coomassie Brilliant Blue (Methods). Chromosomally-encoded RcsF variants were detected by immunoblotting fractions with α-RcsF antibodies. Images are representative of three independent experiments.

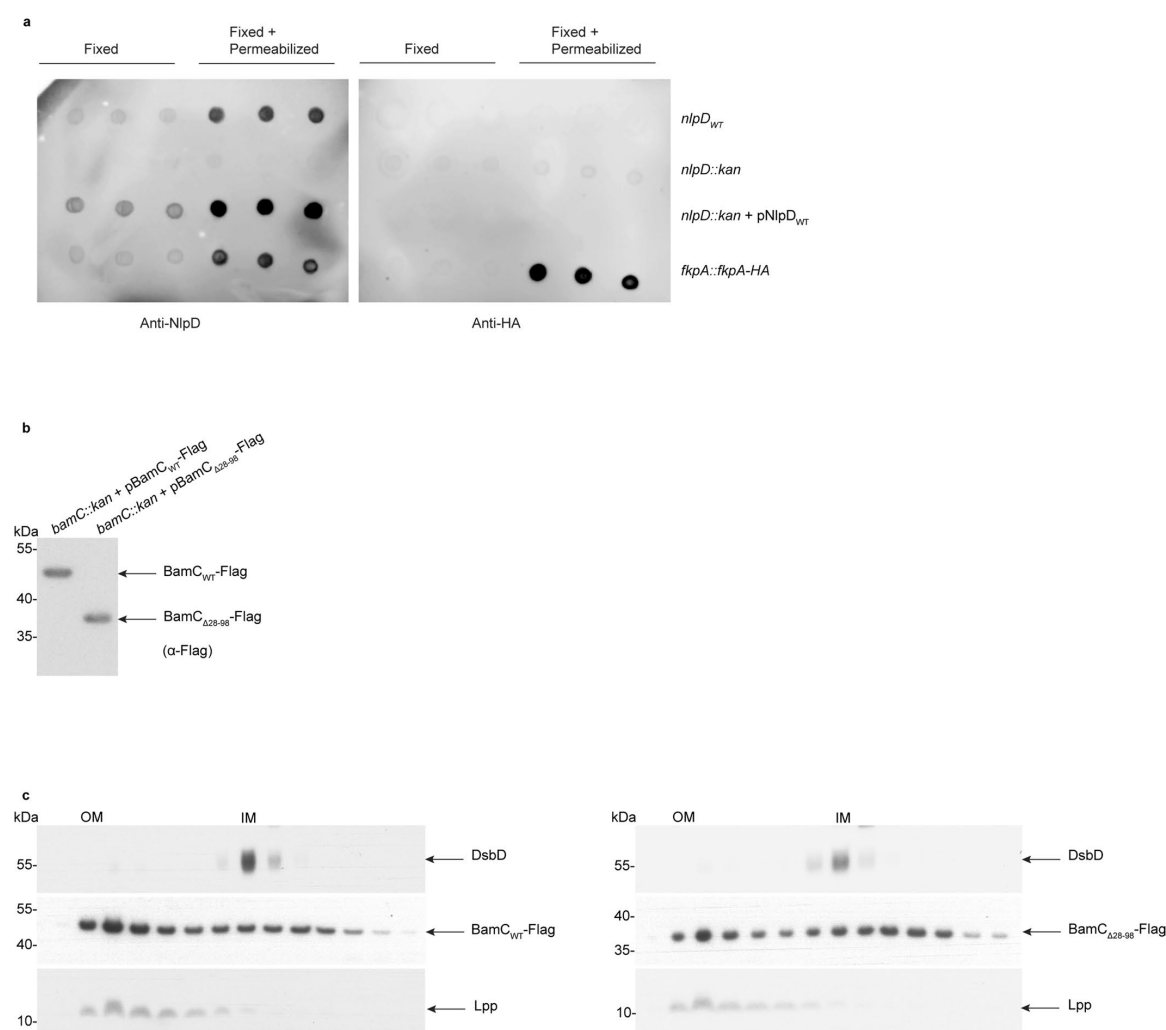


Extended Data Fig. 7 | Overexpression of LoICDE does not restore targeting of RcsF $_{\Delta 19-47}$. **a**, Expression level of LoICDE-His. Cells were grown in LB plus 0.2% arabinose at 37 °C until OD₆₀₀ = 0.7 (Methods). Membrane and soluble fractions were separated with a sucrose density gradient (Methods). LoIE-His was detected in the inner membrane fraction by immunoblotting with α-His (Methods). Images are representative of three independent experiments. **b**, The outer membrane (OM) and inner membrane (IM) were separated with a sucrose density gradient. Expression of LoICDE did not rescue OM targeting of RcsF $_{\Delta 19-47}$. Images are representative of experiments performed in biological triplicate.





Extended Data Fig. 9 | RcsF_{Δ19-47} is processed by Lgt and Lnt. **a**, Lgt replete (+) or deplete samples were obtained by growing cells in the presence or absence of L-arabinose. Overnight cultures were grown at 37 °C in LB medium with 0.2 % arabinose; cells were then diluted in fresh LB medium and grown at 37 °C in the absence of L-arabinose for the indicated times to deplete Lgt. **b**, Lnt replete (+) or deplete samples were obtained from cultures supplemented with L-arabinose or fucose, respectively. Overnight cultures were grown at 37 °C in LB medium in the presence of 0.2 % arabinose; cells were then washed with fresh LB medium, diluted in LB supplemented with 0.2 % fucose and grown at 37 °C for the indicated times to deplete Lnt. Protein samples were subjected to SDS-PAGE tricine or Tris-glycine 16% gel and probed for Lpp and RcsF by immunoblotting. PreproLpp (in the case of Lgt depletion) and apoLpp (in the case of Lnt depletion) are noted as *. The same annotation is used for RcsF variants. The prepro and apo forms of Lpp, RcsF_{WT} and RcsF_{Δ19-47} are detected only when Lgt and Lnt are depleted. This indicates that RcsF_{Δ19-47} and, as expected RcsF and Lpp, are processed by Lgt and Lnt. Images in a and b are representative of biological triplicates.



Extended Data Fig. 10 | The disordered linker of BamC and NlpD, two proteins that can be detected on the cell surface, is important for targeting to the outer membrane (OM). **a**, NlpD is exposed on the cell surface. Fixed cells were spotted (with or without permeabilization with lysozyme and EDTA) onto nitrocellulose membranes and probed with anti-NlpD antibodies. NlpD is detected on non-permeabilized cells, indicating that the protein is on the cell surface. An HA-tagged version of the periplasmic, soluble protein FkpA, expressed from the chromosome, was used as a permeabilization control. HA-tagged FkpA can only be detected with anti-HA antibodies following permeabilization. **b,c**, The disordered linker of BamC is important for its targeting to the outer membrane (OM). **b**, Expression levels of BamC_{WT}-Flag and BamC_{Δ28-98}-Flag. The BamC variants were expressed from a plasmid with a C-terminal Flag-tag. Cells were grown in LB plus IPTG; when OD₆₀₀ reached 0.6, cells were precipitated with trichloroacetic acid (Methods). Samples were subjected to SDS-PAGE and probed with α-Flag-tag antibodies. **c**, Cells expressing BamC_{WT}-Flag and BamC_{Δ28-98}-Flag were grown in LB plus IPTG until OD₆₀₀ reached 0.7. Cells were then subjected to fractionation; the outer (OM) and inner (IM) membranes were separated by sucrose density gradient (Methods). Deletion of the linker of BamC alters the targeting of this lipoprotein to the OM: more BamC_{Δ28-98} is detected in the IM fractions than BamC_{WT}. Images in a, b, c are representative of biological triplicate.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a Confirmed
- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
 - ☒ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
 - ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
 - ☒ ☐ A description of all covariates tested
 - ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
 - ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
 - ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
 - ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
 - ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
 - ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection NIS-Element Advance Research version 4.5

Data analysis GraphPad Prism 9, ImageJ version 1.46r, MODELLER version 9.22, UCSF Chimera version 1.13.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analyzed during this study are included in this published article and its supplementary information file.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	At least 3 biologically independent experiments were performed, which ensures that the averaged results are more accurate than a result from a single experiment. For statistical analyses, each sample had three independent replicates which enabled us to perform the one-way ANOVA and to compare between each two samples by applying Tukey's multiple comparison test.
Data exclusions	No data were excluded from the analyses.
Replication	At least 3 biologically independent experiments were performed. All attempts at replication were successful.
Randomization	This is not relevant to our study given the type of experiments that were carried out. All experiments were performed on bacterial strains and replicates were performed on the same strain.
Blinding	Blinding was not relevant to our study given the type of experiments that were carried out. The results and discussion were based on empirical evidence and data obtained from the performed experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your research. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The polyclonal anti-RcsF (Cho et al., 2014), anti-Lpp (Asmar et al., 2017), anti-DsbD (this study), anti-NlpD (this study) rabbit antisera were generated by our lab. Anti-Pal was a gift the R. Llobès lab (Marseille; Cascales et al., 2002, J. Bact.). The monoclonal anti-His antibody is a peroxidase-conjugated antibody (Qiagen-166035405). Anti-HA (Biolegend MMS-101R) and anti-Flag (Sigma-F1804) are mouse antisera while anti-Myc (Santa Cruz Biotechnology-A0907) is a rabbit antiserum. Both anti-rabbit (A0545) and anti-mouse (A5278) HRP are from Sigma and were used at 1/10000 in all western blots. As for the dot blot experiment, the anti-rabbit-HRP is from Cytiva (NA934) and the anti-mouse-HRP is from Sigma (A4416), and both were used at 1/2000.
Validation	Antibody specificity was confirmed by comparing the immunoblot of the wild-type strain with that of a deletion mutant strain (Δ nlpD, Δ rscF, Δ dsbD, Δ lpp) using each corresponding antibody. Data are available upon request.

II. Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles

Lipoproteins in Gram-negative bacteria: new insights into their biogenesis, subcellular targeting and functional roles[☆]

Jessica El Rayes^{1,2,3}, Raquel Rodríguez-Alonso^{1,2,3} and Jean-François Collet^{1,2}



Bacterial lipoproteins are globular proteins anchored to a membrane by a lipid moiety. By discovering new functions carried out by lipoproteins, recent research has highlighted the crucial roles played by these proteins in the cell envelope of Gram-negative bacteria. Here, after discussing the wide range of activities carried out by lipoproteins in the model bacterium *Escherichia coli*, we review new insights into the essential mechanisms involved in lipoprotein maturation, sorting and targeting to their final destination. A special attention will also be given to the recent identification of lipoproteins on the surface of *E. coli* and of other bacteria. The renewed interest in lipoproteins is driven by the need to identify novel targets for antibiotic development.

Addresses

¹ WELBIO, Avenue Hippocrate 75, 1200 Brussels, Belgium

² de Duve Institute, Université catholique de Louvain, Avenue Hippocrate 75, 1200 Brussels, Belgium

Corresponding author: Collet, Jean-François (jfcollet@uclouvain.be)

³ Both authors contributed equally to the work.

Current Opinion in Microbiology 2021, 61:25–34

This review comes from a themed issue on **Bacterial cell wall synthesis (June 2021)**

Edited by Jean-François Collet and Angelika Gründling

<https://doi.org/10.1016/j.mib.2021.02.003>

1369-5274/© 2021 Elsevier Ltd. All rights reserved.

Introduction

The cell envelope is the morphological hallmark of Gram-negative bacteria. It is composed of two concentric membranes: the inner membrane, which is in contact with the cytoplasm, and the outer membrane, which constitutes the interface with the environment [1]. The inner and outer membranes are separated by the periplasm, a compartment that represents 10–20% of the total cell volume and contains a thin layer of peptidoglycan [1]. The

envelope is required for growth and survival, maintaining cell shape and providing osmotic protection to cells.

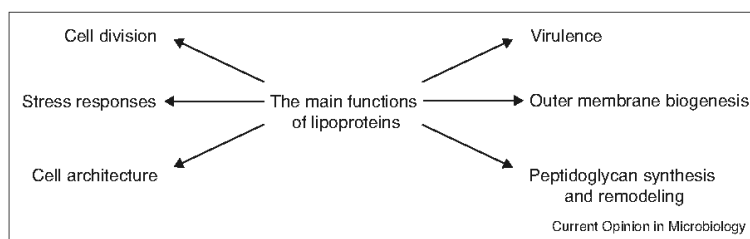
In the model bacterium *Escherichia coli*, about one-third of all proteins are targeted to the cell envelope [2]. Soluble proteins are present in the periplasm where they participate in a variety of processes including peptidoglycan assembly, disulfide bond formation and nutrient import. Integral membrane proteins are present in both the inner and outer membranes; in the outer membrane, they typically adopt a β -barrel conformation [3]. Another important group of envelope proteins is the lipoproteins, which are globular proteins anchored to a membrane by a lipid moiety [4]. Lipoproteins will be the focus of this article.

Lipoproteins exert diverse functions in the cell envelope

A bacterium like *E. coli* can produce and contains ~80 different lipoproteins in its cell envelope [5–7]. These proteins fulfill a variety of important functions (Figure 1). Whereas some, like the outer membrane stress sensors NlpE and RcsF, monitor envelope integrity [8], others play a role in virulence or in envelope biogenesis. For instance, the essential lipoproteins BamD and LptE are required for the assembly of the outer membrane [9,10]: BamD is one of the accessory lipoproteins of the β -barrel assembly machinery (BAM) [3,11] (see the review by Tomasek and Kahne in this issue), while LptE is involved in the trans-envelope complex that transports lipopolysaccharide molecules across the periplasm and inserts them into the outer membrane [12] (see the review by Wilson and Ruiz in this issue). DolP (YraP) is another example of a lipoprotein that is important for the integrity of the outer membrane [13] (Ranava *et al.*, bioRxiv doi: 10.1101/2020.09.23.308296), but its exact function remains to be established. The structure of DolP was recently reported, revealing that the protein is composed of two BON (Bacterial OsmY and Nodulation) domains [13]. Interestingly, the C-terminal BON domain binds anionic phospholipids, an interaction which is important for the function of DolP and its localization to the site of cell division [13].

[☆] Given his role as Guest Editor, Jean-François Collet had no involvement in the peer-review of this article and has no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to Angelika Gründling.

Figure 1



Lipoproteins are involved in a wide range of vital processes in Gram-negative bacteria.

Lipoproteins also play important roles in peptidoglycan synthesis and remodeling. Examples include LpoA and LpoB, two outer membrane lipoproteins that regulate the transpeptidase activity of the peptidoglycan synthases PBP1A and PBP1B, respectively [14,15] (see the review by Pazos and Vollmer in this issue), and NlpD, which controls the activity of AmiC, one of the amidases that contribute to septum cleavage during cell division [16,17]. The list of lipoproteins involved in peptidoglycan remodeling keeps growing: the outer membrane lipoprotein YddW (renamed DigH) was recently identified as a divisome-localized glycosyl hydrolase able to cleave denuded peptide-free glycan strands [18], while NlpI was reported to scaffold peptidoglycan hydrolases and to contribute to complexes active in peptidoglycan synthesis during cell elongation and division [19].

Lipoproteins also are important for controlling the architecture of the cell envelope, for instance by connecting the outer membrane to the peptidoglycan. To that respect, we will now briefly discuss recent studies that brought new insights into two of the most studied *E. coli* lipoproteins, Lpp and Pal.

Lpp, the numerically most abundant protein in *E. coli* with 1 million copies per cell [20], functions as an α -helical tether covalently anchoring the outer membrane to the peptidoglycan: in its absence, the outer membrane pulls away from the cell body and forms blebs [21–23]. Lpp was also shown to dictate the size of the periplasm: increasing the length of Lpp augments the distance between the inner and outer membranes [22,23], which has functional consequences. First, it impacts the size of structures such as the flagellum rod spanning the inter-membrane space [22]. Second, it renders cells unable to detect envelope alterations and to trigger the Rcs stress system as a response [23]. Given the crucial role played by Lpp in *E. coli*, it was intriguing that this protein is only present in a small number of γ -proteobacteria. Recent work addressed this problem by discovering that in several Gram-negative species, the outer membrane is covalently attached to the peptidoglycan by β -barrel proteins

and, in *Coxiella burnetii*, by the lipoprotein LimB [24[•],25^{••}].

New insights have also been reported on Pal, the outer membrane component of the highly conserved Tol-Pal system [26]. The core of Tol-Pal consists of five proteins, 3 in the inner membrane (TolA, TolQ and TolR), one in the periplasm (TolB) and Pal. Like Lpp, Pal links the outer membrane to the peptidoglycan by binding to free diaminopimelic acid residue in the peptidoglycan stem peptides [27]; this interaction is however non-covalent. Tol-Pal is important in cell division, for maintaining OM integrity and for the uptake of colicins and of filamentous phage DNA [26]. In cell division, Tol-Pal stabilizes the outer membrane-peptidoglycan connection during constriction. Interestingly, the de Boer laboratory reported that Tol-Pal proteins, which are dispersed along the cell envelope in non-dividing cells, are dynamically recruited to the divisome near the time of constriction initiation [28]. The relocation of Pal at the division site, which was recently confirmed using functional, chromosomally-encoded proteins fused to fluorescent markers [29,30[•]], was puzzling given that the interaction of Pal with the cell wall limits its ability to laterally diffuse. De Boer and colleagues speculated that TolB, which binds Pal in a mutually exclusive manner with peptidoglycan, was involved in Pal's relocation by forming mobile TolB-Pal complexes [28]. Recent work experimentally confirmed this hypothesis and proposed an elegant mechanism [30[•]] for the mobilization of Pal by TolB and its deposition at the site of constriction in a process fueled by the proton motor force [29,30[•]].

The journey of lipoproteins in the envelope

The maturation of lipoproteins

Lipoproteins are synthesized in the cytoplasm as precursors. They display an N-terminal signal peptide that is ~25 amino acids long and generally presents a short, positively charged region followed by a longer stretch of hydrophobic residues [31,32]. The last four residues of the signal peptide form a conserved [LVII][ASTVI][GAS]C motif referred to as the lipobox [32]. The

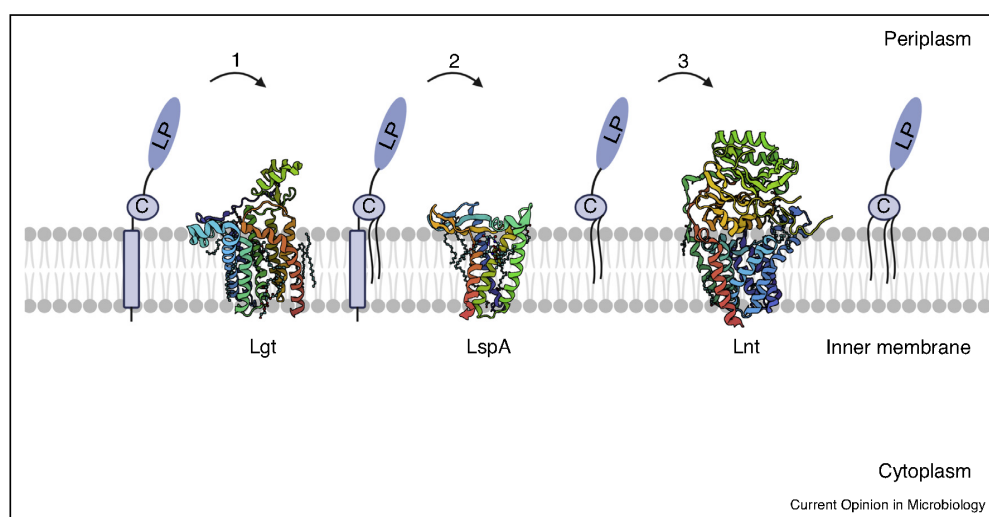
lipobox functions as a molecular determinant of lipid modification which is unique to bacteria; it can be used to predict genes encoding lipoproteins in bacterial genomes [32]. Only the cysteine at the last position of the lipobox is strictly conserved: this cysteine will later become the first residue of the mature protein.

The vast majority of lipoproteins cross the inner membrane in an unfolded state using the general secretory system (Sec) [33]. A few, however, cross the inner membrane in a folded state through the twin-arginine translocation (TAT) machinery [34]. After translocation, the lipoprotein remains anchored to the inner membrane by its N-terminal signal peptide and undergoes several maturation steps. Three inner membrane enzymes participate in this process [35] (Figure 2). The first enzyme is the prolipoprotein diacylglyceryl transferase (Lgt), which catalyzes the covalent attachment of a diacylglyceryl moiety to the cysteine of the lipobox by a thioether linkage. After addition of the diacylglyceryl moiety, the signal peptide is cleaved by the signal peptidase II (LspA), which generates an apolipoprotein, with a cysteine as first residue. The third and last step of lipoprotein maturation consists in acylating the N-terminal amino group of the cysteine, yielding a mature lipoprotein with three acyl chains. This reaction is catalyzed by the apolipoprotein N-acyltransferase (Lnt). Lgt, LspA and Lnt are conserved and essential in the majority of Gram-negative organisms, including pathogenic species [35–38]; they are

therefore attractive targets for the design of new antimicrobials.

In the last few years, important achievements have been made in the understanding of the catalytic mechanisms of Lgt, LspA and Lnt. First, the crystal structure of *E. coli* Lgt was solved (Figure 2) in complex with either phosphatidylglycerol or palmitic acid [39], allowing potential mechanisms for the transacylation reaction between phosphatidylglycerol and the prelipoprotein to be proposed [39,40]. Second, the structure of LspA from two different bacteria, the Gram-negative *Pseudomonas aeruginosa* (Figure 2) and the Gram-positive organism *Staphylococcus aureus* were reported [41,42]. In both organisms, LspA appears as a monomer with four transmembrane α -helices and a periplasmic domain with an amphiphilic β -sheet that rests on the membrane. The structures were solved in complex with antibiotics (globomycin for *Pa*LspA and *Sa*LspA, and myxovirescin for *Sa*LspA), which were found to block access to a region identified as the catalytic site. The active site contains two aspartate residues proposed to represent a catalytic dyad with a general acid–base role; the catalytic aspartates are located in the transmembrane domain and point towards the outer surface of the membrane. Finally, recent publications reported the crystal structures of *E. coli* [43,44–46] and *P. aeruginosa* Lnt [45]. Both enzymes are remarkably similar and consist of a transmembrane domain with eight α -helices providing support to a

Figure 2



Three inner membrane proteins catalyze lipoprotein maturation. Following translocation of the prelipoprotein across the inner membrane, Lgt catalyzes the addition of diacylglyceryl to the thiol group of the cysteine residue in the lipobox, generating “prolipoprotein” (1). The signal peptide is then cleaved by the peptidase LspA (2), giving rise to “apolipoprotein”. The last step of the maturation process, the acylation of the amino group of the now N-terminal cysteine, is catalyzed by Lnt (3). The crystal structures of the three enzymes are shown: Lgt in complex with phosphatidylglycerol (pdb: 5AZC), LspA in complex with globomycin (pdb: 5DIR) and Lnt (pdb: 5VRG). LP: Lipoprotein; C: Cysteine.

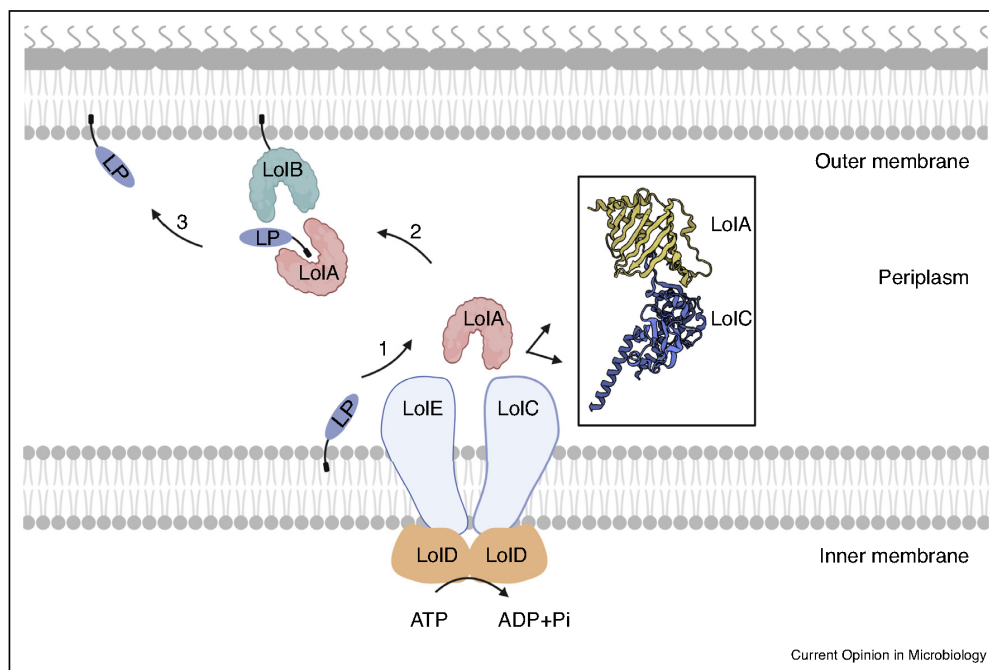
periplasmic nitrilase-like domain (Figure 2). A catalytic triad consisting of a cysteine, a glutamate and a lysine is present in the periplasmic domain, residing slightly above the membrane surface. The available biochemical and structural data support a two-steps ping-pong mechanism for the N-acyltransferase reaction [46–48]. First, an acyl chain is transferred from a phospholipid substrate (preferably phosphatidylethanolamine) to form a thioester linkage with the catalytic cysteine of Lnt. In a second step, the acyl chain is transferred from the cysteine to the α -amino group of the N-terminal cysteine of the lipoprotein. Altogether, the high-resolution structures of Lgt, LspA and Lnt set the stage to the identification of novel antibiotics targeting lipoprotein maturation [49*,50*]; novel fluorescent-based assays that were developed to monitor the activity of Lnt and LspA *in vitro* will facilitate the search for specific inhibitors of these enzymes [51,52].

Dispatching lipoproteins to their final destination

Once matured, lipoproteins either remain in the inner membrane or are exported to the outer membrane. In *E. coli*, most lipoproteins (~90%) are targeted to the outer membrane [53,54]. Lipoprotein extraction from the inner membrane and transport to the outer membrane is carried

out by the Lol system (Figure 3). Lol consists of the inner membrane ABC transporter LolCDE, which extracts lipoproteins from the inner membrane at the expense of ATP hydrolysis and transfers them to the soluble periplasmic chaperone LolA [55]. LolA then shuttles the lipoprotein across the periplasm for delivery to LolB, an outer membrane lipoprotein in charge of outer membrane insertion [55]. In *E. coli*, the current view is that Lol uses the residue adjacent to the lipidated cysteine to sort inner and outer membrane lipoproteins. If the residue at position +2 is an aspartate, the lipoprotein remains in the inner membrane [56,57], most likely because of strong electrostatic interactions between this residue and the headgroups of anionic phospholipids that prevent lipoproteins from interacting with LolCDE [58]. Noteworthy, the residue at position +3 also plays a role: aspartate, glutamate and glutamine strengthen inner membrane retention, whereas histidine and lysine weaken it [56,57]. Different mechanisms have been described in other bacteria. In *P. aeruginosa* for instance, an aspartate is rarely found at position +2 and inner membrane retention appears to be determined by residues at position +3 and +4 [59,60], while in *Borrelia burgdorferi*, the causative agent of Lyme disease in humans, lipoproteins are

Figure 3



The Lol system transports lipoproteins to the outer membrane. Lipoproteins which are destined to the outer membrane are extracted from the inner membrane by LolCDE, an ABC transporter, and delivered to the periplasmic chaperone LolA (1). The mouth-to-mouth interaction between LolA and LolB delivers lipoproteins to LolB (2), which finally inserts lipoproteins in the outer membrane by a poorly known mechanism (3). The crystal structure of LolA bound to the periplasmic domain of LolC (pdb: 6F3Z) is shown. LP: Lipoprotein.

secreted to the outer membrane by default [61]. Interestingly, it was recently shown that lipoproteins are correctly sorted in *P. aeruginosa* cells expressing the *E. coli* LolCDE complex, despite differences in the retention signal between these two species [62^{*}]. To explain this surprising result, a more general mechanism was proposed in which inner membrane retention results from the ability of the N-terminal domain of inner membrane lipoproteins to adopt a conformation favoring strong interactions with the phospholipid bilayer [62^{*}]. Finally, in a recent study, we found that approximately half of lipoproteins destined to the *E. coli* outer membrane display an intrinsically disordered linker at their N-terminus and that these linkers are required for optimal trafficking by the Lol system and normal targeting to the outer membrane (IMN El Rayes & Szewczyk et al, bioRxiv doi: 10.1101/2021.01.05.425367). Clearly, further work remains to be done to fully comprehend how Gram-negative bacteria control the triage of lipoproteins between their two membranes.

When lipoproteins do not have an inner membrane retention signal, they are extracted from the membrane by LolCDE. LolCDE comprises a heterodimer of the integral membrane proteins LolC and LolE and a homodimer of the peripheral membrane protein LolD [63,64]. LolD, present on the cytoplasmic side of the inner membrane, has a nucleotide binding domain that is responsible for ATP hydrolysis, thus providing the energy necessary for lipoprotein extraction. The mechanism by which LolCDE couples ATP hydrolysis, lipoprotein extraction and delivery to LolA remains unclear. Current data support a model in which lipoprotein substrates bind to LolE, before being passed on to LolA, while LolC would act as a scaffold for LolA [65,66^{**}]. On the basis of molecular dynamics simulations, LolC was also recently proposed to facilitate lipoprotein transfer to LolA by destabilizing the inner membrane [67^{*}].

The structure of a complex between LolA and the periplasmic domain of LolC has been solved [66^{**}], but the structure of the full LolCDE complex remains unknown. However, a model has been built on the basis of the structure of MacB, an antibiotic transporter that is homologous to LolCDE [66^{**}]. This model shows LolCDE in two conformations, corresponding to the nucleotide-free state and the ATP-bound state [66^{**}]. In both conformations, LolA is located on top of LolCDE, ~60 Å away from the inner membrane, thus suggesting that lipoproteins, with their hydrophobic lipid moiety, travel a relatively long distance before reaching LolA. The molecular details of this “migration” as well as the potential role of LolC in membrane deformation suggested by molecular dynamics simulations [67^{*}] remain to be determined. It is interesting to note that some Gram-negative bacteria express a homodimer of LolF proteins instead of a heterodimer of LolC and LolE. Bacteria that express LolF,

such as *Francisella tularensis*, *Neisseria meningitidis*, *Acinetobacter species* and *Helicobacter pylori* can tolerate Lnt deletion [68–71].

ATP hydrolysis by LolD causes lipoproteins to be released and delivered to the periplasmic chaperone LolA [65]. LolA folds into an incomplete β -barrel forming a hydrophobic cavity that binds the lipid moiety of substrate lipoproteins, thereby protecting the acyl chains from the hydrophilic periplasm. The hydrophobic cavity of LolA was considered large enough to only accommodate one of the acyl chains of the lipoprotein substrates [72], but recent molecular dynamics simulations questioned this idea, suggesting that it could in fact fit up to three [73]. After crossing the periplasm, LolA delivers its lipoprotein substrates to LolB by a mouth to mouth interaction [74]. LolB, which is a structural homolog of LolA, then anchors the lipoprotein into the inner leaflet of the outer membrane via an unknown mechanism [75]. LolA is well conserved in Gram-negative bacteria, but LolB is only present in γ - and β -proteobacteria [35,76], which indicates that other mechanisms mediating lipoprotein insertion into the outer membrane remain to be identified. In line with this idea, it was recently reported that LolA and LolB, but not LolCDE, could be deleted under certain specific conditions (including *lpp* deletion) in *E. coli*, which suggests that at least one alternative trafficking route to the outer membrane exists in this bacterium [77].

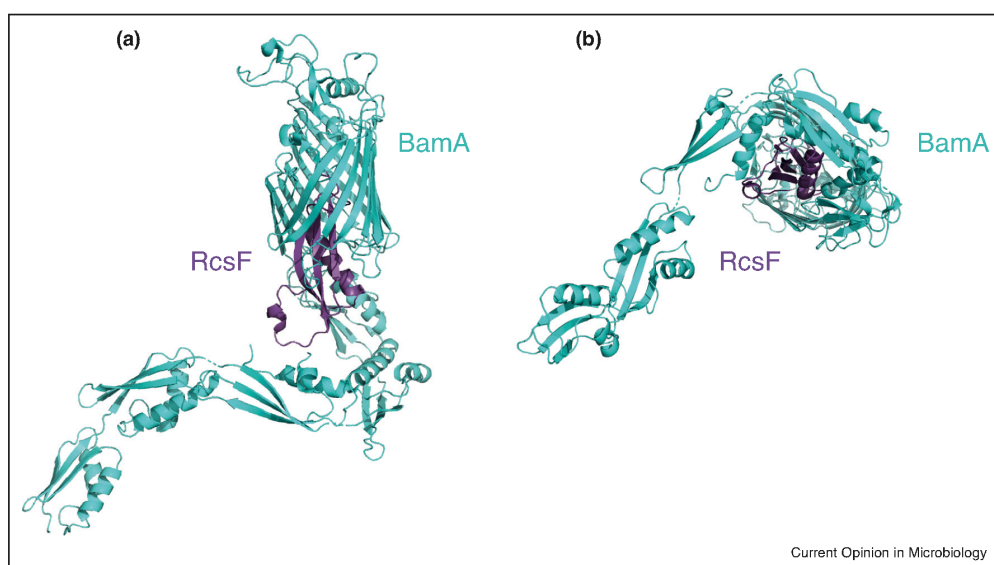
Monitoring lipoprotein trafficking

Given the crucial function played by many outer membrane lipoproteins, bacteria have evolved molecular systems to monitor their trafficking across the cell envelope. In *E. coli*, failure to target outer membrane lipoproteins to their final destination triggers two envelope stress responses, Rcs and Cpx [77–81]. While Rcs copes with outer membrane and peptidoglycan defects [82^{*}], Cpx responds to a broad range of perturbations, including protein misfolding and inner membrane stress [83]. Both Rcs and Cpx can be triggered by an outer membrane lipoprotein, RcsF and NlpE, respectively [8]. When perturbations occur in lipoprotein trafficking, RcsF and NlpE accumulate in the inner membrane, where they can reach their downstream partner (IgaA for RcsF [80,81] and CpxA for NlpE [78,79]), thereby activating their respective signaling cascade. Although Rcs activation leads to higher *lolA* expression [84], which should help fix the damage, it does not provide a fitness advantage when trafficking is perturbed, in contrast to Cpx activation which is beneficial to mutants with defective lipoprotein biogenesis [77–79].

Surface exposed lipoproteins and lipoprotein-OMP complexes

Some lipoproteins, after reaching the outer membrane, cross the asymmetric lipid bilayer and reach the cell

Figure 4



The structure of a complex between the lipoprotein RcsF and BamA suggests a push-and-pull mechanism for the export of lipoproteins to the cell surface. Ribbon diagram of the BamA-RcsF complex in side (a) and periplasmic (b) view. BamA, cyan; RcsF, purple. RcsF is located deep inside the lumen of the BamA β -barrel.

surface, using poorly characterized mechanisms [37,85]. Surface-exposed lipoproteins, which are attractive candidates for vaccine development [86] and often are required for virulence [87], are common in certain bacterial species such as *B. burgdorferi*, where as many as two-thirds of all lipoproteins are surface-exposed [88]. Other bacteria commonly displaying lipoproteins on their surface include species from the *Bacteroides* and *Neisseria* genus [89–91]. In *Neisseria*, surface exposure involves a Surface Lipoprotein Assembly Modulator (SLAM) [92], an outer membrane protein with an N-terminal domain and a 14-stranded β -barrel which is required for virulence [91]. In contrast, only a handful lipoproteins, including Lpp and Pal, have been identified on the *E. coli* cell surface [80,93–99]. Caution is required when considering lipoprotein surface exposure: unambiguously determining whether a lipoprotein crosses the outer membrane or not is technically challenging due to the high yields of false positive. The only *E. coli* lipoprotein that has been shown to reach the cell surface by different groups and complementary techniques is the stress sensor RcsF [80,96,100]. Therefore, RcsF has become a paradigm to investigate the mechanisms of lipoprotein surface exposure in Enterobacteriaceae. According to the current model, RcsF is exported to the cell surface by the β -barrel assembly machinery (BAM) via the assembly of complexes with β -barrel proteins. The crystal structure of a complex between RcsF and BamA, the core BAM component, has recently been solved [101**]. In this complex, which is

considered to be an intermediate in the formation of the RcsF- β -barrel complexes, RcsF binds inside the lumen of the BamA β -barrel (Figure 4), interacting with regions proposed to undergo a lateral opening during β -barrel insertion. This structure suggests a push-and-pull mechanism for the transfer of RcsF to its β -barrel partners [101**] and provides a molecular explanation to the observation that RcsF accumulation on BamA lethally blocks β -barrel assembly [102*,103*]. Interestingly, RcsF uses its interaction with BamA to detect stress in the cell envelope: when damage occurs in the peptidoglycan or the outer membrane, newly synthesized RcsF molecules fail to interact with BamA, they are not exported to the surface and remain exposed to the periplasm, which allows them to trigger the Rcs signaling cascade by reaching IgaA [80,81].

The abundant β -barrels OmpA, OmpC and OmpF were identified as potential vehicles for RcsF surface-exposure [80,96]. However, by dissecting the OmpA-RcsF interaction *in vivo* and *in vitro*, we found that RcsF interacts with the C-terminal, periplasmic domain of OmpA, not with the N-terminal β -barrel. We therefore concluded that OmpA does not mediate the surface exposure of RcsF and predicted that only OmpC and OmpF (which form large β -barrels) serve as vehicles for surface exposure [104].

Conclusions

The diverse roles played by lipoproteins in the cell envelope have been drawing a lot of attention lately,

revealing how crucial these proteins are in a wide range of vital processes. However, this is only the tip of the iceberg: even in the well-studied model organism *E. coli*, the function of most lipoproteins remains to be discovered. Future research will also establish whether lipoproteins commonly decorate the bacterial surface, which could change our view of the cell envelope, and determine whether surface-exposed lipoproteins use BAM as a general machinery to cross the outer membrane or if other machineries, like SLAM [105], are involved. Finally, future research will also identify potential alternative trafficking routes across the periplasm and address how lipoproteins are inserted in the outer membrane of bacteria missing LolB. Addressing these questions will open the way to novel antibiotics by providing both new molecular targets and crucial mechanistic information.

Acknowledgments

We thank Dr Antoine Delhaye for helping with the figures. This work was supported, in part, by grants from the FRFS-WELBIO (WELBIO-CR-2019C-03), from the EOS Excellence in Research Program of the FWO and FRF-FNRS (G0G0818N), from the Fédération Wallonie-Bruxelles (ARC 17/22-087), and from the European Commission via the International Training Network Train2Target (721484).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Silhavy TJ, Kahne D, Walker S: **The bacterial cell envelope.** *Cold Spring Harb Perspect Biol* 2010, **2**:a000414.
2. Weiner JH, Li L: **Proteome of the *Escherichia coli* envelope and technological challenges in membrane proteome analysis.** *Biochim Biophys Acta* 2008, **1778**:1698-1713.
3. Ricci DP, Silhavy TJ: **Outer membrane protein insertion by the β -barrel assembly machine.** *EcoSal Plus* 2019, **8**.
4. Braun V, Hantke K: **Lipoproteins: structure, function, biosynthesis.** *Subcell Biochem* 2019, **92**:39-77.
5. Gonnet P, Rudd KE, Lisacek F: **Fine-tuning the prediction of sequences cleaved by signal peptidase II: a curated set of proven and predicted lipoproteins of *Escherichia coli* K-12.** *Proteomics* 2004, **4**:1597-1613.
6. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A: **Prediction of lipoprotein signal peptides in Gram-negative bacteria.** *Protein Sci* 2003, **12**:1652-1662.
7. Brokx SJ, Ellison M, Locke T, Bottorff D, Frost L, Weiner JH: **Genome-wide analysis of lipoprotein expression in *Escherichia coli* MG1655.** *J Bacteriol* 2004, **186**:3254-3258.
8. Laloux G, Collet JF: **Major Tom to ground control: how lipoproteins communicate extra-cytoplasmic stress to the decision center of the cell.** *J Bacteriol* 2017.
9. Chimalakonda G, Ruiz N, Chng SS, Garner RA, Kahne D, Silhavy TJ: **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* 2011, **108**:2492-2497.
10. Malinverni JC, Werner J, Kim S, Sklar JG, Kahne D, Misra R, Silhavy TJ: **YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*.** *Mol Microbiol* 2006, **61**:151-164.
11. Wu T, Malinverni J, Ruiz N, Kim S, Silhavy TJ, Kahne D: **Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*.** *Cell* 2005, **121**:235-245.
12. Sherman DJ, Xie R, Taylor RJ, George AH, Okuda S, Foster PJ, Needleman DJ, Kahne D: **Lipopolysaccharide is transported to the cell surface by a membrane-to-membrane protein bridge.** *Science* 2018, **359**:798-801.
13. Bryant JA, Morris FC, Knowles TJ, Maderbocus R, Heinz E, Boelter G, Alodaini D, Colyer A, Wotherspoon PJ, Staunton KA et al.: **Structure of dual BON-domain protein DolP identifies phospholipid binding as a new mechanism for protein localisation.** *Elife* 2020, **9**.
14. Paradis-Bleau C, Markovski M, Uehara T, Lupoli TJ, Walker S, Kahne DE, Bernhardt TG: **Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases.** *Cell* 2010, **143**:1110-1120.
15. Typas A, Banzhaf M, van den Berg van Saparoea B, Verheul J, Biboy J, Nichols RJ, Zietek M, Beilharz K, Kannenberg K, von Rechenberg M et al.: **Regulation of peptidoglycan synthesis by outer-membrane proteins.** *Cell* 2010, **143**:1097-1109.
16. Heidrich C, Templin MF, Ursinus A, Merdanovic M, Berger J, Schwarz H, de Pedro MA, Hölte JV: **Involvement of N-acetylmuramyl-L-alanine amidases in cell separation and antibiotic-induced autolysis of *Escherichia coli*.** *Mol Microbiol* 2001, **41**:167-178.
17. Uehara T, Parzych KR, Dinh T, Bernhardt TG: **Daughter cell separation is controlled by cytokinetic ring-activated cell wall hydrolysis.** *EMBO J* 2010, **29**:1412-1422.
18. Yakhnina AA, Bernhardt TG: **The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division.** *Proc Natl Acad Sci U S A* 2020, **117**:6777-6783.
19. Banzhaf M, Yau HC, Verheul J, Lodge A, Kritikos G, Mateus A, Cordier B, Hov AK, Stein F, Wartel M et al.: **Outer membrane lipoprotein NlpI scaffolds peptidoglycan hydrolases within multi-enzyme complexes in *Escherichia coli*.** *EMBO J* 2020, **39**:e102246.
20. Li GW, Burkhardt D, Gross C, Weissman JS: **Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources.** *Cell* 2014, **157**:624-635.
21. Hoekstra D, van der Laan JW, de Leij L, Witholt B: **Release of outer membrane fragments from normally growing *Escherichia coli*.** *Biochim Biophys Acta* 1976, **455**:889-899.
22. Cohen EJ, Ferreira JL, Ladinsky MS, Beeby M, Hughes KT: **Nanoscale-length control of the flagellar driveshaft requires hitting the tethered outer membrane.** *Science* 2017, **356**:197-200.
23. Asmar AT, Ferreira JL, Cohen EJ, Cho SH, Beeby M, Hughes KT, Collet JF: **Communication across the bacterial cell envelope depends on the size of the periplasm.** *PLoS Biol* 2017, **15**:e2004303.
24. Sandoz KM, Moore RA, Beare PA, Patel AV, Smith RE, Bem M, Hwang H, Cooper CJ, Priola SA, Parks JM et al.: **β -Barrel proteins tether the outer membrane in many Gram-negative bacteria.** *Nat Microbiol* 2021, **6**:19-26.
25. Godessart P, Lannoy A, Dieu M, Van der Verren SE, Soumillion P, Collet JF, Remaut H, Renard P, De Bolle X: **β -Barrels covalently link peptidoglycan and the outer membrane in the α -proteobacterium *Brucella abortus*.** *Nat Microbiol* 2021, **6**:27-33.
26. Szczepaniak J, Press C, Kleanthous C: **The multifarious roles of Tol-Pal in Gram-negative bacteria.** *FEMS Microbiol Rev* 2020, **44**:490-506.
27. Parsons LM, Lin F, Orban J: **Peptidoglycan recognition by Pal, an outer membrane lipoprotein.** *Biochemistry* 2006, **45**:2122-2128.

In this paper, the authors report that β -barrel proteins covalently tether the outer membrane to the peptidoglycan in several Gram-negative bacteria. They also report the identification of the lipoprotein LimB as a covalent tether in *Coxiella burnetii*.

In this paper, the authors report that β -barrel proteins covalently tether the outer membrane to the peptidoglycan in alpha-proteobacteria.

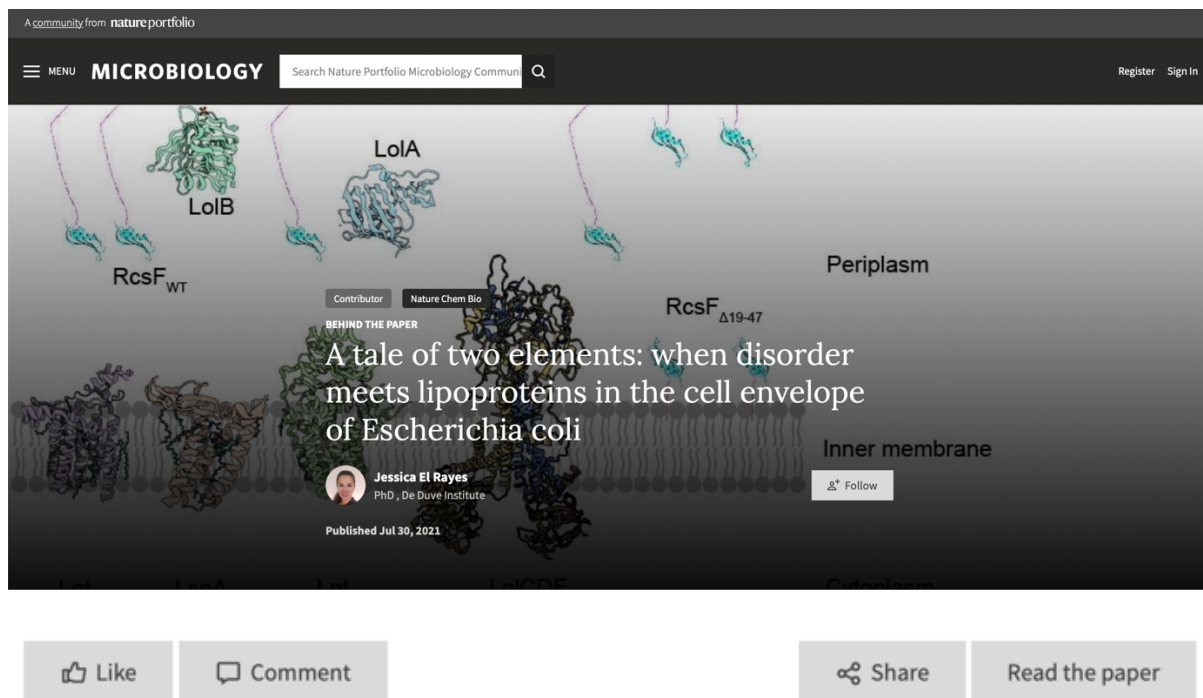
28. Gerding MA, Ogata Y, Pecora ND, Niki H, de Boer PA: **The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*.** *Mol Microbiol* 2007, **63**:1008-1025.
 29. Petiti M, Serrano B, Faure L, Lloubes R, Mignot T, Duch D: **Tol energy-driven localization of Pal and anchoring to the peptidoglycan promote outer-membrane constriction.** *J Mol Biol* 2019, **431**:3275-3288.
 30. Szczepaniak J, Holmes P, Rajasekar K, Kaminska R, Samsudin F, Inns PG, Rassam P, Khalid S, Murray SM, Redfield C *et al.*: **The lipoprotein Pal stabilises the bacterial outer membrane during constriction by a mobilisation-and-capture mechanism.** *Nat Commun* 2020, **11**:1305.
- This paper proposes an elegant mechanism for the mobilization of Pal by TolB.
31. Hayashi S, Wu HC: **Lipoproteins in bacteria.** *J Bioenerg Biomembr* 1990, **22**:451-471.
 32. Babu MM, Priya ML, Selvan AT, Madera M, Gough J, Aravind L, Sankaran K: **A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins.** *J Bacteriol* 2006, **188**:2761-2773.
 33. Crane JM, Randall LL: **The Sec system: protein export in *Escherichia coli*.** *EcoSal Plus* 2017, **7**.
 34. Palmer T, Berks BC: **The twin-arginine translocation (Tat) protein export pathway.** *Nat Rev Microbiol* 2012, **10**:483-496.
 35. Okuda S, Tokuda H: **Lipoprotein sorting in bacteria.** *Annu Rev Microbiol* 2011, **65**:239-259.
 36. Kovacs-Simon A, Titball RW, Michell SL: **Lipoproteins of bacterial pathogens.** *Infect Immun* 2011, **79**:548-561.
 37. Buddelmeijer N: **The molecular mechanism of bacterial lipoprotein modification—how, when and why?** *FEMS Microbiol Rev* 2015, **39**:246-261.
 38. Qi HY, Sankaran K, Gan K, Wu HC: **Structure-function relationship of bacterial prolipoprotein diacylglycerol transferase: functionally significant conserved regions.** *J Bacteriol* 1995, **177**:6820-6824.
 39. Mao G, Zhao Y, Kang X, Li Z, Zhang Y, Wang X, Sun F, Sankaran K, Zhang XC: **Crystal structure of *E. coli* lipoprotein diacylglycerol transferase.** *Nat Commun* 2016, **7**:10198.
 40. Singh W, Bilal M, McClory J, Dourado D, Quinn D, Moody TS, Sutcliffe I, Huang M: **Mechanism of phosphatidylglycerol activation catalyzed by prolipoprotein diacylglycerol transferase.** *J Phys Chem B* 2019, **123**:7092-7102.
 41. Vogeley L, El Arnaout T, Bailey J, Stansfeld PJ, Boland C, Caffrey M: **Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin.** *Science* 2016, **351**:876-880.
 42. Olatunji S, Yu X, Bailey J, Huang CY, Zapotoczna M, Bowen K, Remškar M, Müller R, Scanlan EM, Geoghegan JA *et al.*: **Structures of lipoprotein signal peptidase II from *Staphylococcus aureus* complexed with antibiotics globomycin and myxovirescin.** *Nat Commun* 2020, **11**:140.
- This paper describes the crystal structure of LspA of *S. aureus* in complex with the antibiotics globomycin and myxovirescin (pdb: 6RYP). Both antibiotics inhibit LspA as non-cleavable tetrahedral intermediate analogs and block the catalytic dyad of LspA.
43. Wiseman B, Högbom M: **Conformational changes in Apolipoprotein N-acyltransferase (Lnt).** *Sci Rep* 2020, **10**:639.
- This paper reports two crystal forms of Lnt from *E. coli*. Analysis of the structures suggest that the movement of essential loops and residues triggered by substrate binding could control the interaction between Lnt and the incoming substrate apolipoprotein.
44. Lu G, Xu Y, Zhang K, Xiong Y, Li H, Cui L, Wang X, Lou J, Zhai Y, Sun F *et al.*: **Crystal structure of *E. coli* apolipoprotein N-acyl transferase.** *Nat Commun* 2017, **8**:15948.
 45. Wiktor M, Weichert D, Howe N, Huang CY, Olieric V, Boland C, Bailey J, Vogeley L, Stansfeld PJ, Buddelmeijer N *et al.*: **Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis.** *Nat Commun* 2017, **8**:15952.
 46. Noland CL, Kattke MD, Diao J, Gloor SL, Pantua H, Reichelt M, Katakam AK, Yan D, Kang J, Zilberleyb I *et al.*: **Structural insights into lipoprotein N-acylation by *Escherichia coli* apolipoprotein N-acyltransferase.** *Proc Natl Acad Sci U S A* 2017, **114**:E6044-E6053.
 47. Hillmann F, Argentini M, Buddelmeijer N: **Kinetics and phospholipid specificity of apolipoprotein N-acyltransferase.** *J Biol Chem* 2011, **286**:27936-27946.
 48. Cheng W, Doyle DA, El Arnaout T: **The N-acyltransferase Lnt: structure-function insights from recent simultaneous studies.** *Int J Biol Macromol* 2018, **117**:870-877.
 49. El Arnaout T, Soulimane T: **Targeting lipoprotein biogenesis: considerations towards antimicrobials.** *Trends Biochem Sci* 2019, **44**:701-715.
- This review describes very carefully the characteristics of the three maturation enzymes: Lgt, LspA and Lnt as a great target for the design of new antibiotics.
50. Legood S, Boneca IG, Buddelmeijer N: **Mode of action of lipoprotein modification enzymes—Novel antibacterial targets.** *Mol Microbiol* 2020.
- In this review the authors discuss the recent literature on the three maturation enzymes and the use of *in vitro* assays to screen for novel inhibitors.
51. Nozeret K, Boucharlat A, Agou F, Buddelmeijer N: **A sensitive fluorescence-based assay to monitor enzymatic activity of the essential integral membrane protein Apolipoprotein N-acyltransferase (Lnt).** *Sci Rep* 2019, **9**:15978.
 52. Kitamura S, Owensby A, Wall D, Wolan DW: **Lipoprotein signal peptidase inhibitors with antibiotic properties identified through design of a robust *in vitro* HT platform.** *Cell Chem Biol* 2018, **25**:301-308.e312.
 53. Horler RS, Butcher A, Papangelopoulos N, Ashton PD, Thomas GH: **EchoLOCATION: an *in silico* analysis of the subcellular locations of *Escherichia coli* proteins and comparison with experimentally derived locations.** *Bioinformatics* 2009, **25**:163-166.
 54. Tokuda H: **Biogenesis of outer membranes in Gram-negative bacteria.** *Biosci Biotechnol Biochem* 2009, **73**:465-473.
 55. Narita S, Tokuda H: **Biogenesis and membrane targeting of lipoproteins.** *EcoSal Plus* 2010, **4**.
 56. Terada M, Kuroda T, Matsuyama SI, Tokuda H: **Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of *Escherichia coli*.** *J Biol Chem* 2001, **276**:47690-47694.
 57. Gennity JM, Inouye M: **The protein sequence responsible for lipoprotein membrane localization in *Escherichia coli* exhibits remarkable specificity.** *J Biol Chem* 1991, **266**:16458-16464.
 58. Hara T, Matsuyama S, Tokuda H: **Mechanism underlying the inner membrane retention of *Escherichia coli* lipoproteins caused by Lol avoidance signals.** *J Biol Chem* 2003, **278**:40408-40414.
 59. Lewenza S, Mhlanga MM, Pugsley AP: **Novel inner membrane retention signals in *Pseudomonas aeruginosa* lipoproteins.** *J Bacteriol* 2008, **190**:6119-6125.
 60. Narita S, Tokuda H: **Amino acids at positions 3 and 4 determine the membrane specificity of *Pseudomonas aeruginosa* lipoproteins.** *J Biol Chem* 2007, **282**:13372-13378.
 61. Schulze RJ, Zuckert WR: **Borrelia burgdorferi lipoproteins are secreted to the outer surface by default.** *Mol Microbiol* 2006, **59**:1473-1484.
 62. Lorenz C, Dougherty TJ, Lory S: **Correct sorting of lipoproteins into the inner and outer membranes of *Pseudomonas aeruginosa* by the *Escherichia coli* LolCDE transport system.** *mBio* 2019, **10**.
- This paper demonstrates that LolCDE of *E. coli* and *P. aeruginosa* are interchangeable. The authors proposed a general sorting model based on

the ability of the N-terminal domain to adopt a conformation that facilitates strong interactions with the inner membrane.

63. Yakushi T, Masuda K, Narita S, Matsuyama S, Tokuda H: **A new ABC transporter mediating the detachment of lipid-modified proteins from membranes.** *Nat Cell Biol* 2000, **2**:212-218.
64. Szewczyk J, Collet JF: **The journey of lipoproteins through the cell: one birthplace, multiple destinations.** *Adv Microb Physiol* 2016, **69**:1-50.
65. Mizutani M, Mukaiyama K, Xiao J, Mori M, Satou R, Narita S, Okuda S, Tokuda H: **Functional differentiation of structurally similar membrane subunits of the ABC transporter LolCDE complex.** *FEBS Lett* 2013, **587**:23-29.
66. Kaplan E, Greene NP, Crow A, Koronakis V: **Insights into bacterial lipoprotein trafficking from a structure of LolA bound to the LolC periplasmic domain.** *Proc Natl Acad Sci U S A* 2018, **115**:E7389-E7397.
- The crystal structure of LolA in complex with the periplasmic domain of LolC is described in this paper, revealing the zone of interaction between these two proteins. Moreover, the authors built a model for the LolCDE complex based on the homology with the crystal structure of MacB. In this model, LolA is located 60 Å away from the inner membrane, implicating mechanotransmission in lipoprotein extraction and delivery to LolA.
67. Rao S, Bates GT, Matthews CR, Newport TD, Vickery ON, Stansfeld PJ: **Characterizing membrane association and periplasmic transfer of bacterial lipoproteins through molecular dynamics simulations.** *Structure* 2020, **28**:475-487. e473
- In this study, the authors used molecular dynamics simulations to model the interactions of 81 *E. coli* lipoproteins with the lipid membrane. They also show that LolC facilitates the transfer of the lipoprotein to LolA by destabilizing the membrane.
68. Grabowicz M: **Lipoproteins and their trafficking to the outer membrane.** *EcoSal Plus* 2019, **8**.
69. Gwin CM, Prakash N, Christian Belisario J, Haider L, Rosen ML, Martinez LR, Rigel NW: **The apolipoprotein N-acyl transferase Lnt is dispensable for growth in Acinetobacter species.** *Microbiology (Reading)* 2018, **164**:1547-1556.
70. McClain MS, Voss BJ, Cover TL: **Lipoprotein processing and sorting in *Helicobacter pylori*.** *mBio* 2020, **11**.
71. LoVullo ED, Wright LF, Isabella V, Huntley JF, Pavelka MS Jr: **Revisiting the Gram-negative lipoprotein paradigm.** *J Bacteriol* 2015, **197**:1705-1715.
72. Takeda K, Miyatake H, Yokota N, Matsuyama S, Tokuda H, Miki K: **Crystal structures of bacterial lipoprotein localization factors, LolA and LolB.** *EMBO J* 2003, **22**:3199-3209.
73. Boags A, Samsudin F, Khalid S: **Details of hydrophobic entanglement between small molecules and Braun's lipoprotein within the cavity of the bacterial chaperone LolA.** *Sci Rep* 2019, **9**:3717.
74. 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* 2009, **106**:5877-5882.
75. Narita SI, Tokuda H: **Bacterial lipoproteins: biogenesis, sorting and quality control.** *Biochim Biophys Acta Mol Cell Biol Lipids* 2017, **1862**:1414-1423.
76. Grabowicz M: **Lipoprotein transport: greasing the machines of outer membrane biogenesis: re-examining lipoprotein transport mechanisms among diverse gram-negative bacteria while exploring new discoveries and questions.** *Bioessays* 2018, **40**:e1700187.
77. Grabowicz M, Silhavy TJ: **Redefining the essential trafficking pathway for outer membrane lipoproteins.** *Proc Natl Acad Sci U S A* 2017, **114**:4769-4774.
78. Delhay A, Laloux G, Collet JF: **The lipoprotein NlpE is a Cpx sensor that serves as a sentinel for protein sorting and folding defects in the *Escherichia coli* envelope.** *J Bacteriol* 2019, **201**.
79. May KL, Lehman KM, Mitchell AM, Grabowicz M: **A stress response monitoring lipoprotein trafficking to the outer membrane.** *mBio* 2019, **10**.
80. Cho SH, Szewczyk J, Pesavento C, Zietek M, Banzhaf M, Roszczenko P, Asmar A, Laloux G, Hov AK, Leverrier P et al.: **Detecting envelope stress by monitoring beta-barrel assembly.** *Cell* 2014, **159**:1652-1664.
81. Hussein NA, Cho SH, Laloux G, Siam R, Collet JF: **Distinct domains of *Escherichia coli* IgA connect envelope stress sensing and down-regulation of the Rcs phosphorelay across subcellular compartments.** *PLoS Genet* 2018, **14**:e1007398.
82. Wall E, Majdalani N, Gottesman S: **The complex Rcs regulatory cascade.** *Annu Rev Microbiol* 2018, **72**:111-139.
- A comprehensive review on Rcs, one of the most complex stress sensing system in bacteria.
83. Delhay A, Collet JF, Laloux G: **A fly on the wall: how stress response systems can sense and respond to damage to peptidoglycan.** *Front Cell Infect Microbiol* 2019, **9**:380.
84. Tao K, Narita S, Tokuda H: **Defective lipoprotein sorting induces lolA expression through the Rcs stress response phosphorelay system.** *J Bacteriol* 2012, **194**:3643-3650.
85. Konovalova A, Silhavy TJ: **Outer membrane lipoprotein biogenesis: Lol is not the end.** *Philos Trans R Soc Lond B Biol Sci* 2015, **370**.
86. Principato S, Pizzo M, Rappuoli R: **Meningococcal factor H binding protein as immune evasion factor and vaccine antigen.** *FEBS Lett* 2020, **594**:2657-2669.
87. Wilson MM, Bernstein HD: **Surface-exposed lipoproteins: an emerging secretion phenomenon in gram-negative bacteria.** *Trends Microbiol* 2016, **24**:198-208.
88. Dowdell AS, Murphy MD, Azodi C, Swanson SK, Florens L, Chen S, Zuckert WR: **Comprehensive spatial analysis of the *Borrelia burgdorferi* lipoproteome reveals a compartmentalization bias toward the bacterial surface.** *J Bacteriol* 2017, **199**.
89. Zuckert WR: **Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond.** *Biochim Biophys Acta* 2014, **1843**:1509-1516.
90. Zielke RA, Le Van A, Baarda BI, Herrera MF, Acosta CJ, Jerse AE, Sikora AE: **SlitC is a surface-displayed lipoprotein that is required for the anti-lysozyme strategy during *Neisseria gonorrhoeae* infection.** *PLoS Pathog* 2018, **14**:e1007081.
91. Hooda Y, Shin HE, Bateman TJ, Moraes TF: **Neisserial surface lipoproteins: structure, function and biogenesis.** *Pathog Dis* 2017, **75**.
92. Hooda Y, Lai CC-L, Judd A, Buckwalter CM, Shin HE, Gray-Owen SD, Moraes TF: **Slam is an outer membrane protein that is required for the surface display of lipidated virulence factors in *Neisseria*.** *Nat Microbiol* 2016, **1**:16009.
93. Cowles CE, Li Y, Semmelhack MF, Cristea IM, Silhavy TJ: **The free and bound forms of Lpp occupy distinct subcellular locations in *Escherichia coli*.** *Mol Microbiol* 2011, **79**:1168-1181.
94. Dong C, Beis K, Nesper J, Brunkan-Lamontagne AL, Clarke BR, Whitfield C, Naismith JH: **Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein.** *Nature* 2006, **444**:226-229.
95. Webb CT, Selkig J, Perry AJ, Noinaj N, Buchanan SK, Lithgow T: **Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC.** *J Mol Biol* 2012, **422**:545-555.
96. Konovalova A, Perlman DH, Cowles CE, Silhavy TJ: **Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of beta-barrel proteins.** *Proc Natl Acad Sci U S A* 2014, **111**:E4350-E4358.
97. Goyal P, Krasteva PV, Van Gerven N, Gubellini F, Van den Broeck I, Troupiotis-Tsailaki A, Jonckheere W, Pehau-Arnaudet G, Pinkner JS, Chapman MR et al.: **Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG.** *Nature* 2014, **516**:250-253.

98. Arnold MF, Caro-Hernandez P, Tan K, Runti G, Wehmeier S, Scoocchi M, Doemler WT, Walker GC, Ferguson GP: **Enteric YaiW is a surface-exposed outer membrane lipoprotein that affects sensitivity to an antimicrobial peptide.** *J Bacteriol* 2014, **196**:436-444.
99. Michel LV, Shaw J, MacPherson V, Barnard D, Bettinger J, D'Arcy B, Surendran N, Hellman J, Pichichero ME: **Dual orientation of the outer membrane lipoprotein Pal in *Escherichia coli*.** *Microbiology* 2015, **161**:1251-1259.
100. Konovalova A, Mitchell AM, Silhavy TJ: **A lipoprotein/ β -barrel complex monitors lipopolysaccharide integrity transducing information across the outer membrane.** *Elife* 2016, **5**.
101. Rodriguez-Alonso R, Letoquart J, Nguyen VS, Louis G, Calabrese AN, Iorga BI, Radford SE, Cho SH, Remaut H, Collet JF: **Structural insight into the formation of lipoprotein-beta-barrel complexes.** *Nat Chem Biol* 2020, **16**:1019-1025.
This paper is the first to describe the crystal structure of the outer membrane β -barrel protein BamA, the key component of the BAM machinery, in complex with a lipoprotein substrate, in this case the stress sensor RcsF. RcsF is lodged deep within the lumen of BamA. A push-and-pull mechanism is proposed for the export of RcsF to the cell surface.
102. Tata M, Konovalova A: **Improper coordination of BamA and BamD results in Bam complex jamming by a lipoprotein substrate.** *MBio* 2019, **10**.
In this paper, the authors show that in the absence of BamE, one of the accessory lipoproteins of BAM, RcsF accumulates on BamA blocking the activity of the BAM complex. They also propose that BamE plays a role in coordinating the assembly of RcsF- β -barrel complexes.
103. Hart EM, Gupta M, Wuhr M, Silhavy TJ: **The synthetic phenotype of DeltabamB DeltabamE double mutants results from a lethal jamming of the Bam complex by the lipoprotein RcsF.** *MBio* 2019, **10**.
In this paper, the authors found that the lethal phenotype of the mutant lacking BamB and BamE is caused by the jamming of BAM by the lipoprotein RcsF. They show that RcsF deletion suppresses the lethality of the double mutant and restores global β -barrel levels.
104. Dekoninck K, L  toquart J, Laguri C, Demange P, Bevernaegie R, Simorre JP, Dehu O, Iorga BI, Elias B, Cho SH *et al.*: **Defining the function of OmpA in the Rcs stress response.** *Elife* 2020, **9**.
105. Hooda Y, Lai CCL, Moraes TF: **Identification of a large family of slam-dependent surface lipoproteins in gram-negative bacteria.** *Front Cell Infect Microbiol* 2017, **7**:207.

III. A tale of two elements: when disorder meets lipoproteins in the cell envelope of *Escherichia coli*



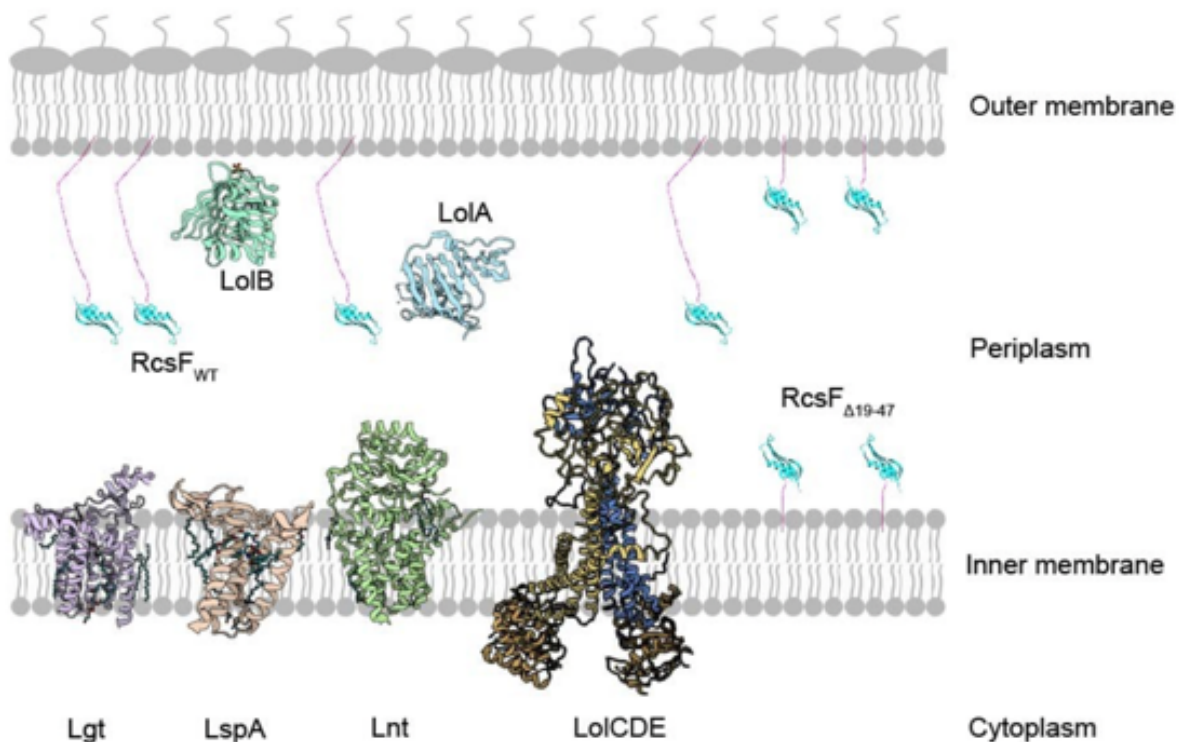
In Gram-negative bacteria, many proteins are anchored to the inner or the outer membrane via a lipid moiety. This group of proteins is called lipoproteins. They are important for the viability of the cell as they fulfil a variety of different functions¹. Some are stress sensors, others participate in the biogenesis and the integrity of the cell envelope, while some contribute to the bacterial virulence.

With the help of our collaborator Dr. Bogdan I. Iorga, we were able to show that half of the outer membrane lipoproteins of *E. coli* display a long N-terminal intrinsically disordered region composed of more than 22 residues. Intrigued by the presence of these unstructured linkers, we decided to study their role in the targeting of the lipoproteins of *E. coli*.

The current paradigm is that the sorting of lipoproteins between the inner and outer membranes depends on the identity of two amino acids (the “Lol sorting signal”) located at the N-terminus of the protein, next to the lipidated cysteine residue². Our results challenge this widely accepted model by showing that correct trafficking by the Lol lipoprotein sorting system and targeting to the outer membrane also depend on the N-terminal intrinsically disordered linkers. We determined that the length of the linker is important, but the sequence is not, on the condition that it cannot fold into a defined secondary structure. Focusing on three structurally unrelated outer

membrane lipoproteins (RcsF, Pal and NlpD), we showed the functional relevance of our findings. Additionally, we demonstrated that N-terminal disordered segments are required for optimal processing of lipoproteins by the Lol system: decreasing the load of the Lol system by deleting the gene encoding Lpp, the most abundant lipoprotein in *E. coli*, rescues outer membrane targeting of lipoproteins variants devoid of their N-terminal linker. Finally, our results also suggest that unstructured linkers might be used by lipoproteins to cross the outer membrane and reach the cell surface.

Altogether, our results provide important mechanistic insights into the biogenesis of lipoproteins, which are attracting a lot of attention due to their crucial cellular functions and potential as targets for new antibacterials.



Excitingly, our work also provides direct evidence that evolution can select protein intrinsic disorder by function. Most proteins present segments that are intrinsically disordered (it is estimated that 30-50% of eukaryotic proteins contain regions that do not adopt a defined secondary structure *in vitro*³). However, demonstrating that these unstructured regions are functionally important *in vivo* has been challenging. By showing that displaying an N-terminal disordered segment downstream of the Lol signal is required for the correct sorting of some lipoproteins, the *current* investigation provides crucial functional evidence that disorder matters.

To find out about our results and the role of disorder in lipoprotein processing, you can find our full publication by using the following link:

<https://www.nature.com/articles/s41589-021-00845-z>

1. Szewczyk, J. & Collet, J.F. The Journey of Lipoproteins Through the Cell: One Birthplace, Multiple Destinations. *Adv Microb Physiol* **69**, 1-50 (2016).
2. Okuda, S. & Tokuda, H. Lipoprotein sorting in bacteria. *Annu Rev Microbiol* **65**, 239-59 (2011).
3. Bardwell, J.C. & Jakob, U. Conditional disorder in chaperone action. *Trends Biochem Sci* **37**, 517-25 (2012).



Jessica El Rayes

PhD , De Duve Institute

 Follow