Uncovering the formation of the Gram-negative Bacterial Cell Envelope, and peptidoglycan synthesis

By Christopher Laurie Barnett Graham

A thesis submitted in partial fulfillment for the degree of

Doctor of Philosophy in Life Sciences

University of Warwick, Department of Life Sciences

March 2023

Thesis Contents

Attached Papers – After References

Figures and Tables

Abbreviations

Antimicrobial Resistance - AMR

Peptidoglycan - PG

Outer membrane - OM

Inner membrane - IM

Lipopolysaccharide - LPS

Sporulation, Elongation and Division complexes - SEDS

n-Dodecyl-B-D-Maltoside - DDM

N,N-Dimethyl-1-Dodecanamine-N-Oxide - LDAO

Cryo-Electron microscopy - Cryo-EM

Penicillin binding protein - PBP

Double Electron-electron resonance - DEER

Mammalian cell entry - MCE

A bacterium with two membrane layers - Diderm

Diderm - Gram-negative

Diethylaminoethyl beads - DEAE bead

S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate-MTSL

Acknowledgements

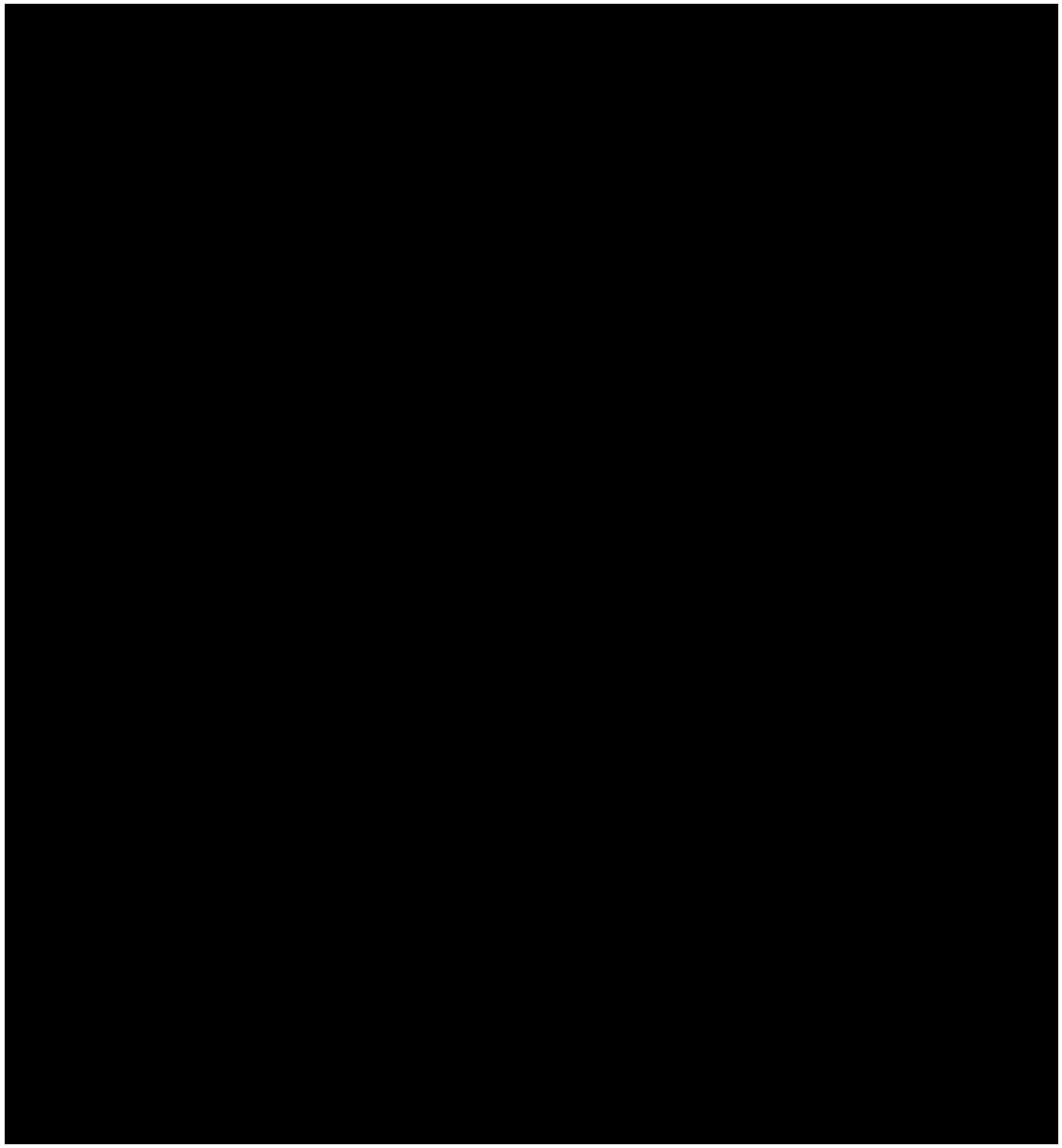
In reality, I started this PhD many years ago. My family would inspire me to learn about the natural world. I lived in a zoo and shared ownership of pigs, tortoises and emus etc, and I also read textbooks. It was the trust in me with the farm, and the care or thought into inspiring, taking me to the park to feed the ducks, watch the birds, draw or letting me use the work computer and play Lego rock raiders as a toddler that made this PhD. It was the idea that there is nothing you cannot do if you try, from the handy work of my father and his schemes, to the dream realised by my mother just as I started my PhD proper that has shaped it. Science is of course trying to make stories from the data produced by the natural world. Here this science and PhD focuses on tiny bacteria and how they grow.

But I changed a lot over the PhD, masters and placements before it. I would like to thank Prof David Roper, Julie Todd, Dr Hector Newman, Prof Marc Santolini, Thomas Landrain, Prof Liz Sockett, Dr David Edwards, Francessca Gillet, Dr Andrew Lovering, Dr Manuel Banzhaf, Dr Jack Bryant, Joseph Parker, Morgane Opoix, Camille Masselot, Luca Haenel, Dr Emily Skates, Dr Tyler Baverstock, Dr Alex Murray, Heather Scott, Amber Vjestica, Oscar Garcia, Elliot Lawton and Danielle Graham MSc, who have all helped me finish this PhD one way or another. I would hope to consider you truly to have been a friend, and will forever be grateful. Id also like to thank my family who have lived at Wain lea farm and Congleton for raising me in the way I was raised, so as to be able to complete the following.

For practical help during the PhD I'd like to thank Francessca Gillet, Nic Briggs, Dr Megan Sambrook, Katie Graham (no blood relation) Prof David Roper (he would like to be referred to as the post doctoral Tea boy), Prof Phillip Stansfeld as well as Cleidiane from the proteomics facility at the University of Warwick

Id also like to thank my funders the BBSRC, The University of Warwick, Antibiotic Research UK, NERC, MITACs and the Microbiology Society.

Thank you to Damien Rice, and Prof J.R.R.Tolkien for their artistic contributions.



When life gives you lemons, you just eat them, rinds and all.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy in Biology and Biochemistry. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Stability of RodA-PBP2 4.7A complex – Chapter 2.3 RMSD calculations and simulations were performed by Dr Owen Vickery
Double Electron Electron Resonance work – Chapter 2.7 'Flexibility of the RodA-PBP2 complex interrogated by DEER' – These measurements were made by Dr Meagan Dulphrisne of The University of Virginia, with my consultation
Final Cryo-EM structure of RodA-PBP2, partly shown in Chapter 2.3.1, 2.9.1 and Chapter 2.9.2. The final cryo-EM and its data was collected by Dr Rie Nygaard.
Dynamic ligand binding simulations (Different from Autodock Vina work) – shown in Chapter 2.8. This work was produced by Dr Jonny Colburn, on behalf of myself and Philip Stansfield.
The binding of Lipid XX to RodA-PBP2 – shown in Chapter 2.9. This work was done exclusively by Prof Philip Stansfeld,
The final purification of MltC, D, E and F – shown in Chapter 3.3.1 The proteins used in this gel and in the final FITC assay performed together with Francesca Gillet Chapter 3.9, were made by Francesca Gillet.
Michael Alao helped perform the microbial genomics assay, as well as provided the method – shown in Chapter 4 'Chemical genomics assays' and helped provide the data analysed by myself in Figure 4.4.1
Dr Harvey Hanjeong and Katie Graham performed sucrose selection using my suicide vectors, to create the fluorescent Pseudomonas aeruginosa gene knockins assayed and imaged in Chapter 4.
Collaborative work included all of the above collaborations, as well as working with Prof Lori Burrows on the MCE domain containing protein fluorescent localisation, whom performed sucrose gradient selection for creating the strains imaged, and Dr Manuel Banzhaf on

supervision of interrogating the lipid dye. I was in all cases, except those stated above the lead in these collaborations and the data collected is my own.

Parts of this thesis have been published by the author:

Networks of peptidoglycan interaction - Chapter 1.1 - Graham CLB, Newman H, Gillett FN, Smart K, Briggs N, Banzhaf M, Roper DI. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-negative Bacteria. *International Journal of Molecular Sciences*. 2021; 22(23):12831. https://doi.org/10.3390/ijms222312831

Various RodA-PBP2 experiments -Chapter 2 Cryo-EM structure of the Peptidoglycan Elongasome complex, and Lipid II biosynthesis mechanism, R Nygaard, CLB Graham et al, (Accepted) - *Nature Communications 2023*)

GT-C Protein family work – Chapter 2.5.2 Ashraf, Khuram U., et al. "Structural basis of lipopolysaccharide maturation by the O-antigen ligase." *Nature* 604.7905 (2022): 371-376.

MCE domain containing protein work – Chapter 4 '*Pseudomonas* MCE domaining proteins PqiB and MlaD do not coordinate with division, CLB Graham et al 2023 (Submitting to *Microbiology*)

Phosphatidyl choline mimic work – Chapter 4 - 'Phosphatidyl choline mimic propargyl-choline for membrane staining and phospholipid tracking in Pseudomonas *aeruginosa*'. CLB Graham et al 2023 (*Access Microbiology*)

The above papers in their submitted and published forms are attached after the references section of the PhD. Chapter 1.1 is a direct copy of the published work completed during my PhD, used as the peptidoglycan introduction and review in the thesis. For other papers written during the PhD and not included in the thesis, specifically on pedagogy and sociology please refer to Appendix 5 and the papers attached at the back of the thesis.

Researcher Training

During the PhD, I completed Graduate Teacher Assistant Training. I also was mentored by my supervisor, and Dr Manuel Banzhaf in grant writing, together writing and winning four independent grants:

- **MITACs NERC Doctoral Exchange Grant (2019/2020) £12,500.** Awarded by NERC and MITACs – Principle applicant. Worked on the Outer membrane lipid homeostasis MCE proteins in Pseudomonas Aeruginosa, leading to further funding with ANTRUK.
- **ANTRUK Grant (2021/2022) £25,000** Awarded by Antibiotic research UK for work on the OM-PG balance - Principle applicant. Work on the outer membrane lipid homeostasis proteins in P.*aeruginosa* and E.*coli.* Currently writing paper on outcomes
- **Travel Award (2022) £500** Awarded by Biochemistry Society, used for travel to Gordon Research Conference 2022, in addition to ANTRUK funding.
- **Warwick Business Impact Award 2022 (£2000)** Awarded by University- To host Innovations for AMR conference, co-application with Cytecom and Professor David Roper.

Other than the informal grant writing training and laboratory training by my supervisor, I took part in the MIBTP's training programme and events.

Thesis Abstract

This thesis is about how bacteria grow, especially how they grow the cell wall that keeps them rigid, and their cell envelope. Specifically, this thesis is on understanding the formation of the Gram negative version of the bacterial cell envelope, which constitutes a great deal of pathogenic species. Gram-negative bacteria are a sub-class of bacteria within the phylogeny of prokaryotes (1). Bacteria like all living organisms are formed partly of a DNA core or chromatin, with central processes that allow for propagation of this chromatin, encapsulated by a cellular envelope. The cell envelope in Gram-negative is a layer of protection including a cell wall with fatty membranes on either side, including especially an outer membrane(2). My PhD project on these bacteria has explored the coordination and creation of this cell envelope, including the creation and destruction of bacteria peptidoglycan or glycan sugar mesh which gives this envelope structure (Chapters 1-3). In addition to this it looks at the lipid/fatty part of the cell envelope, and postulates parts of its formation (Chapter 4). Phospholipid layer biosynthesis systems create the fatty membranes that give the envelope flexibility and its ability to separate the ions and other small molecules within the organism from the environment, this is essential for life. The aim of this thesis is to improve understanding of Gram-negative (aka diderm) bacterial cell growth and division with a notion to use this understanding for future antibiotic treatment. Our central hypothesis is that peptidoglycan (PG) and outer membrane (OM) cell envelope metabolism systems must be linked. This combined study is essential, as Gram-negative bacteria make up some of our most relevant pathogens, with more deaths caused by these organisms and their resistance to treatment, than AIDs and Malaria (3). Understanding the cell envelope will allow us to combat these organisms.

The starting point of this thesis came from our studies working on expanding the knowledge of the current core of the 'elongasome' responsible for cell growth; RodA and PBP2 (4,5), which have been shown to together increase the chain length of glycan strands, required for longitudinal cell wall extension (6). I started with an aim to link this core complex with hypothesized partners LpoA, PBP1a and other proteins with potential regulatory or enzymatic roles. However upon bioinformatic analysis (2)(Chapter 1) I realised this is more complex and likely involves systems which create the entire cell envelope, and thus the scope of the PhD initially imagined as Chapter 2's topic on RodA-PBP2 was widened beyond the initial RodA-PBP2 complex. The Chapter 1 background along with methods established in Chapter 2, (which use *in vitro* assays of the RodA-PBP2 system to create polymerised lipid II, and find the mechanism of this enzymes action) are then used to create a new assay to uncover function of the cell wall breakdown enzymes the lytic transglycosylases in Chapter 3. In Chapter 4 I took the understanding gained from the interaction of the SEDS machinery with other proteins, and the connection to outer membrane proteins uncovered in Chapter 1, 2 and 3 to postulate and see under the microscope the localisation of outer membrane phospholipid insertion (7,8) and biosynthetic machines (MlaD/PqiB/YebT/etc) in the cell envelope (10), and use a phosphatidylcholine mimic only usable in *Pseudomonas aeruginosa* in Chapter 4 to study lipid movement and localisation. This was in the hopes of finding new patterns in outer membrane and cell envelope laydown, which may follow similar trends to that of peptidoglycan laydown established in Chapter 1 and 2. I summarise this work, and the linked papers in each chapter's conclusion and my thesis's impact in the 'Thesis conclusion'.(9) Finally, whilst doing my PhD which explores the network of proteins in Gram-negative cell envelopes, I also wrote three papers on creating communities in science, which acknowledge the thesis, and hypothesise new tools for combatting the future of AMR internationally in a changing world.

Chapter 1: The Diderm Cell Envelope

Summary:

The initial search for a central thesis on the cell envelope required good literature knowledge, indeed before we started this research, there was a dream of finding all the interaction partners of RodA and PBP2 in 2019. I had also previously worked on the Gram-negative predatory bacteria *Bdellovibrio bacteriovorus* and its cell envelope(11), together this brought the focus of the PhD onto the cell envelope's outer membrane being created in tandem with the peptidoglycan layer, as a central thesis connecting two topics "peptidoglycan" and "cell membranes synthesis", often studied separately.

In order to understand the protein-protein interactions required to create peptidoglycan, in this chapter and as a peer reviewed and published piece of work, I wrote a meta-review on all peptidoglycan related proteins in the context of a cell envelope system (Chapter 1.1) (2). Further to this review on which my PhD has grown, there is a second part to this chapter which gives a background to the lipid centered cell envelope aspect of my PhD, in which I describe the outer membrane creation systems (Chapter 1.2). However, the impact of such work, and the work in all my thesis can only be understood in the context of 'antimicrobial resistance' and bacteria as a pathogen.

Antimicrobial resistance

The importance of this PhD and its research is that it studies the mechanisms of forming the cell envelope required for bacterial growth. Bacteria, and more specifically Gram-negative/Diderm bacteria can be pathogenic, or opportunistic in their lifestyles, and these traits often lead to human suffering and death. Bacteria often use human and animal bodies as an environment they can use for nutrition and growth, synergistically or in a more harmful way. Sometimes this leads to widespread pandemics, and before the invention of antibiotics in the early 20th century, it caused frequent global catastrophe, and a lowered life expectancy, with not only pandemics, but superficial cuts and tissue surgery for other unrelated health issues having a high mortality rate due to bacterial invasion. Antibiotics kill bacteria responsible for causing these illnesses and have reduced the impact of these opportunistic lifeforms.(12)

However, the age of antibiotics effectiveness as a drug is ending. Bacteria are capable of adapting to antibiotics, and a 'one-health' meaning a wholistic chronic issue of some regions having cross-contamination of water sources with antibiotics, as well as antibiotic overuse, has led to resistance in environmental bacteria. This resistance has been spread by plasmids, as well as opportunistic DNA scavenging. Sometimes these resistant bacteria and mutations, often in modifications to the pathways of the cell envelope formation are obtained by our pathogens. We have not uncovered enough antibiotics in recent years to keep up with this rise in resistance which will continue to occur, especially to Gram-negative bacteria which already have an intrinsic resistance to many antibiotics. Therefore, we have entered a new age, which requires novel solutions and understanding, both from a policy standpoint and from a research centric antibiotic discovery standpoint. (12)

Whilst changes and study of antimicrobial stewardship policy, as well as education and antimicrobial resistance (AMR) tracking, can lead to a positive change(13). It is the antibiotics themselves and new mechanistic understanding of how antibiotic resistance occurs that can be used in our fight against infections that already exist, and with it fight the inevitable pathogenically relevant resistance that will arise despite the best of planning and stewardship programmes.

Antibiotics can be found in a range of forms, from that of live predatory bacteria able to target specific species(14), to bacteriophage capable of lysing cells on mass(15), to discovery of compounds produced by other organisms. However, one of the most potent and famous of

antibiotics are the class of antibiotics which affect the cell wall, or peptidoglycan such as “Penicillin”(16).

Penicillin works by binding proteins involved in the transpeptidation or crosslinking of the cell wall (Chapter 1.1), which we will discuss in greater detail throughout my PhD. These proteins are known as the “Penicillin binding proteins” or PBPs. Their function, and role in the cell envelope, is part of crosslinking the cell wall joining peptide side chains emanating from the glycan sugar polymer backbone into a mesh which maintains cell turgor pressure and shape. Since the discovery of penicillin, new antibiotics have been found which effect glycosyltransferases, or addition of sugars to one another in a chain in addition to the breakdown of peptidoglycan itself by lytic enzymes(17), as well as interfering in the steps prior to this. The PBPs, and their protein partners which catalyse glycosyltransferase the SEDS (Sporulation, elongation, and division) (2) complexes are the central research piece of the thesis.

In addition to antibiotics which effect the cell wall, the lipid bilayer component of the cell envelope (Chapter 1.2), can also be challenged. Nisin, a cytoplasmic membrane agonist and lipid II binding molecule (linking cell wall formation and membrane integrity) is one example and is effective against Gram-positive bacteria and has yet to be overcome by natural resistance (18). However other drugs, such as MRL-494 can now target the outer membrane itself, specifically targeting Gram-negative/diderm bacteria(19). The use of a combination of compounds or therapies such as these (or discovery of more compounds similar to them) could be the future of antibiotic therapies, therefore understanding the interplay and mechanisms which govern the formation of cell envelope layers is important.

Chapter 1.1 A dynamic network of proteins facilitate cell envelope biogenesis in Gram-negative bacteria.

As Published in:

Graham CLB, Newman H, Gillett FN, Smart K, Briggs N, Banzhaf M, Roper DI. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-negative Bacteria. *International Journal of Molecular Sciences*. 2021; 22(23):12831. https://doi.org/10.3390/ijms222312831

Bacteria must maintain the ability to modify and repair the peptidoglycan layer without jeopardizing its essential functions in cell shape, cellular integrity, and intermolecular interactions. A range of new experimental techniques is bringing an advanced understanding of how bacteria regulate and achieve peptidoglycan synthesis particularly in respect of the central role played by complexes of Sporulation, Elongation or Division (SEDS) and class B penicillin-binding proteins required for cell division and cell shape. In this review we highlight relationships implicated by a bioinformatic approach between outer membrane, cytoskeletal components, periplasmic control proteins and cell elongation/division proteins to provide further perspective on the interactions of these cell division and cell shape complexes. I detail the network of protein interactions that assist in the formation of peptidoglycan and highlight the increasingly dynamic and connected set of protein machinery and macrostructures that assist in creating the cell envelope layers in Gram-negative bacteria.

Peptidoglycan in Gram-negatives

Peptidoglycan plays a vital role in the maintenance of cell envelope integrity in bacteria generally, in Gram-negative bacteria it acts as a stabilising structure which is attached to both the inner and

outer membranes lipid bilayers (20). Peptidoglycan is formed from a repeating beta-1-4-linked N-acetylmuramic acid-N-acetylglucosamine disaccharide glycan polymer (MurNAc-GlcNAc) with peptide side chains. The amino acid chains extend from the N-acetylmuramic sugar and crosslink to adjacent glycan polymer to create a macroscopic mesh-like structure (Figure 1.1.1) (21). The cell constantly modifies this mesh-like macromolecule with a set of hydrolases and transferases to allow for cell expansion, shape changes and septation. A recent review covers these modifications and the proteins involved in detail (22).

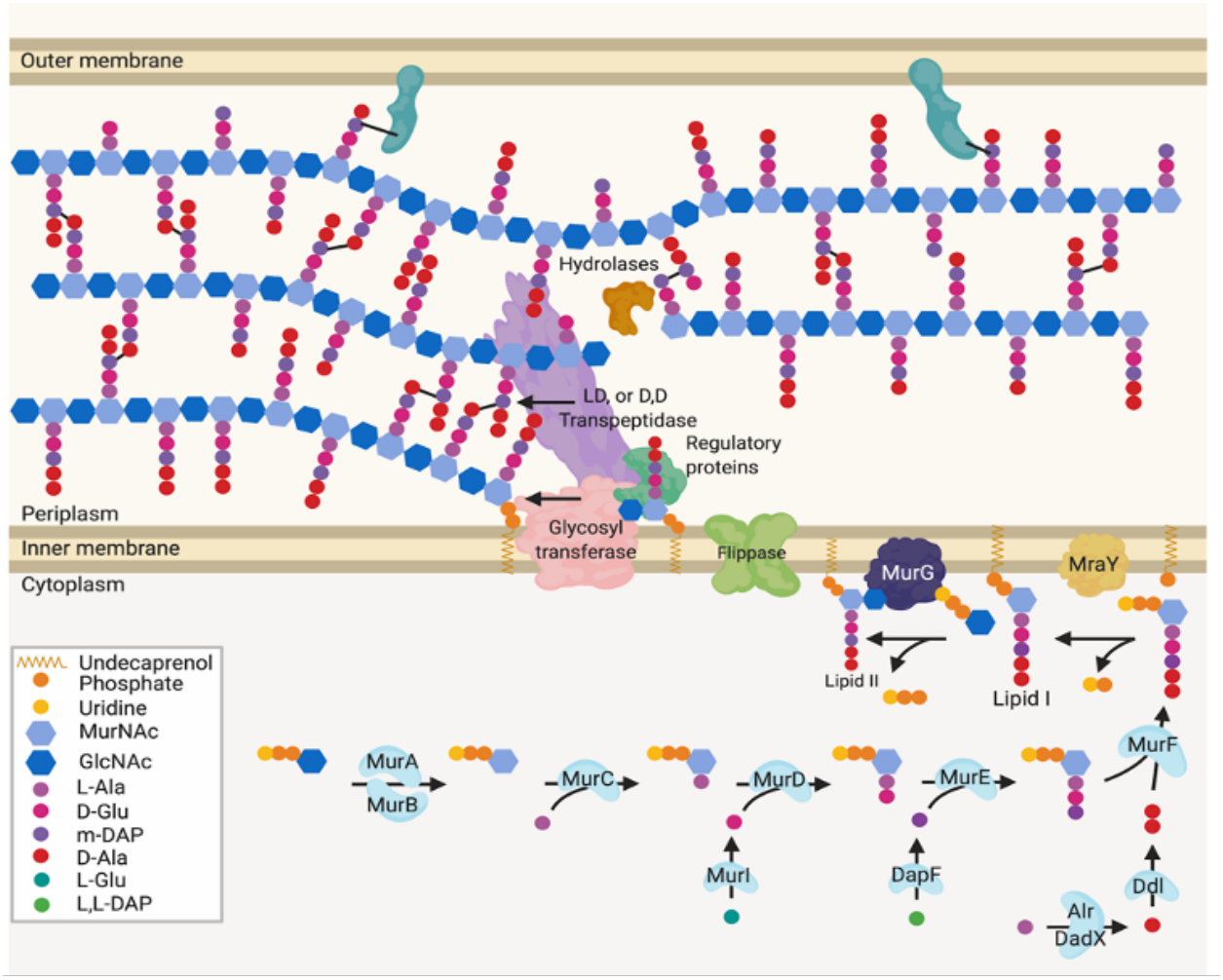


Figure 1.1.1. Generalised peptidoglycan synthesis and insertion pathway.

Lipid II is the peptidoglycan building block precursor. Its precursor is synthesized in the cytoplasm by sequential enzymatic steps then attached to undecaprenyl phosphate in the inner membrane (23,24). The newly formed Lipid II is then flipped across the inner membrane and polymerized into glycan chains by the glycosyltransferase (GT) action of class A bifunctional penicillin binding proteins (PBPs), Sporulation,

Elongation or Division proteins (SEDS) in complex with class B monofunctional PBPs or monofunctional glycosyltransferases (25).

Cell wall modifying enzymes and complexes have altered localisation during growth which is essential for specialized peptidoglycan biosynthesis.

The location of the enzymes required for the synthesis of peptidoglycan and its later modification (Figure 1.1.1) can vary depending on cellular events and conditions (26,27). The proteins and complexes involved are also dynamic, as many of their locations have been shown to change during the cell cycle. Studies using fluorescent gene fusions within the chromosome and peptidoglycan protein tracking approaches (4,14,28) now provide indications of coordinated peptidoglycan protein complex movement during the cell cycle (29,30) (Figure 1.1.2A, B, C). Localisation of these complexes presumably ensures that peptidoglycan is synthesized at particular regions for overall or highly specialized growth situations such as: cell division (Fig1.1.2D); cell curvature; (Fig1.1.2G) polar growth and maintenance (Fig1.1.2E); and flagella associated regions (Fig1.1.2F).

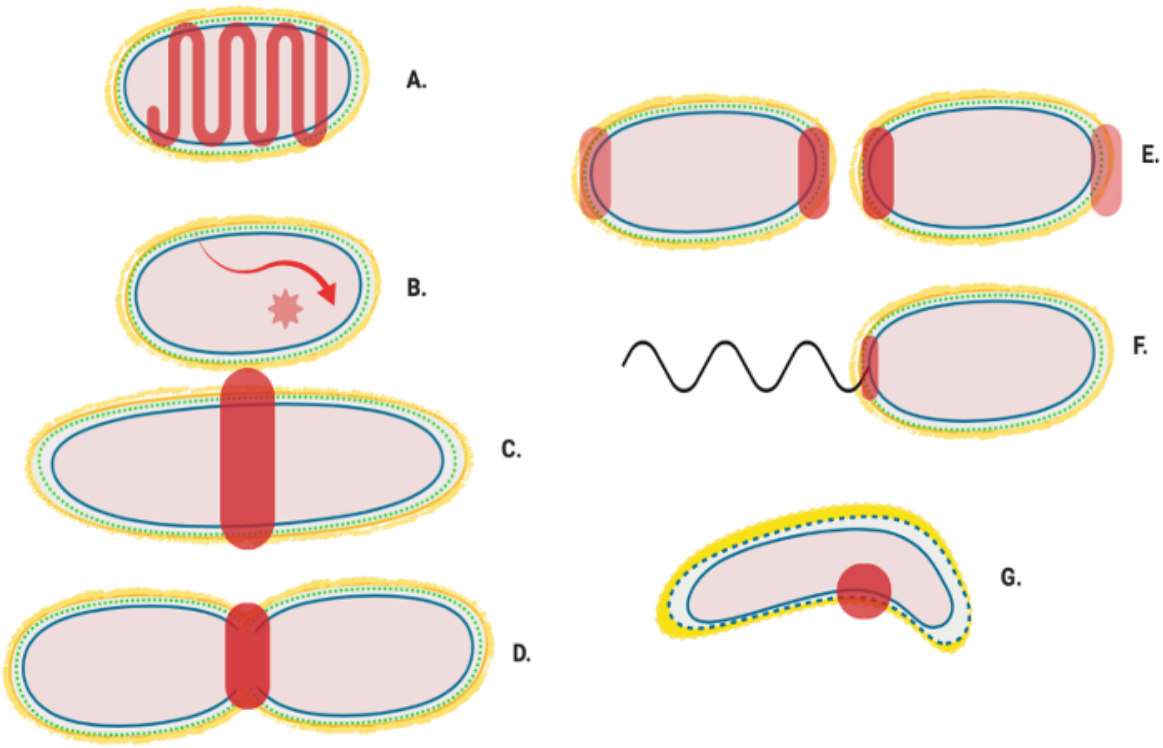


Figure 1.1.2 Generalised localisation of peptidoglycan modifying proteins.

Localisation regions of known and potential peptidoglycan modifying enzymes in Gram-negative bacteria. Localisation sites are highlighted in red. A Helical and MreB associated Elongasome B. Free diffusion (unlocalised) (29)C. Pre-septal machinery, D. Division machinery, E. Post septal polar machinery and polar growth, (31–33), F. Flagella peptidoglycan modification machinery (34), and G. Shape determining pinpoint/seam (35–37).

Regulation of peptidoglycan modifying enzymes by their interacting partners

In order to achieve such diversity of in the form and location of peptidoglycan, its synthesis and subsequent modification must be highly coordinated. Peptidoglycan is a complex three-dimensional molecule with an architecture and chemistry which is dependent on species and localisation (38,39). The cell wall responds (via modified synthetic pathways) to antibiotic challenge and changing osmotic conditions (21,40) (Fig 1.1.3).

Specialisation of peptidoglycan has been postulated to be driven by pathways which are regulated by local enzyme concentrations and protein:protein interactions.

Integral peptidoglycan synthesis machinery such as RodA/FtsW and PBP2/PBP3 have been shown to have non-enzymatic regulatory partners such MreC/MreD (40) and FtsN/L/Q, (41) but also enzymatic regulatory pairs exist e.g. PBP1A-PBP2 , and RodA-PBP2. This network of peptidoglycan altering enzymes and regulators is far from being understood structurally or functionally (2).

Method used to visualise PG synthesis networks for this meta-review.

To visualise the interactions of the genes and proteins relevant to peptidoglycan synthesis and allow a full informative meta-analysis, we have performed a network analysis of relevant genes using contemporary bioinformatic approaches (42).

Genetic and protein interactions confirmed by the literature

Peptidoglycan modifying and related genes, as listed in cell division, peptidoglycan, and peptidoglycan related papers on *E. coli* were collated to create a peptidoglycan relevant gene list (26,41,43). I submitted this list of genes as a joint submission to the gene data trawling engine "STRING" (STRING) to create an interaction map centred around our listed proteins' data. Genes that interacted with this initial list, or were not in our initial list and given a combined "STRING" score of 0.7 (determined by co-occurrence data among species, gene neighborhood scores and/or experimental data) (44–46) were then added and this new list was inputted as a meta-submission.

Those genes which after meta-submission were found to have compound interaction scores with other listed genes > 0.9 were used in our final literature analysis (Figure 3 and 4). Additional annotations of protein interactions were added from the literature. This has created a comprehensive picture of the current literature representing peptidoglycan synthesis and modification in Gram-negative bacteria (Figure 1.1.4).

Proposed genetic interactions

In addition to the confirmed interactions so far uncovered by the literature as shown in Figure 3, we show the predicted network used to create Figure 1.1.3. In Figure 1.1.4 unconfirmed STRING determined interactions of 0.7 or higher are shown in addition to known interactions. This more expansive and connected network represented in Figure 4 was constructed using a database of known and predicted gene-gene/protein-protein interactions derived from direct (physical) and indirect (functional) genetic associations, along with interactions aggregated from other (primary) databases (44,45). I have grouped proteins currently linked to specific peptidoglycan assembly machinery and their cellular locations into their respective groups through layering of background colour. This creates a gene-gene interaction pattern network, contextualised by the large protein assemblies they associate, such as the “divisome” and the “elongasome", as well as other potential environment dependent complexes.

Most peptidoglycan synthases and modifiers are members of multiple local complexes as predicted by genetic interactions and confirmed by literature

Our network analysis (Figure 1.1.3, 1.1.4 and SI Figure 3) suggests that the overlapping protein complex localisations of Figure 3 such as those related to cell stress, the “divisome“ and “elongasome” involve some of the same proteins, shared among complexes. Our analysis of Figure 1.1.3, which focuses on known literature-verified interactions, indicates that relevant proteins interact with multiple complexes, often with coordination at the membrane and outer membrane; between cytoplasmic and periplasmic processes of peptidoglycan formation and across the peptidoglycan barrier within the periplasm. These interactions involve lytic, regulator, anchoring, cytoskeletal, and anabolic enzymes, often acting on the same proteins.

The central message revealed by the networks suggests that not only does regulation of proteins occur in complexes, but also a degree of protein exchange and sharing occurs among them to enable a range of possible complexes. The roles, an interaction table and activity of each protein

in the network below is discussed in this chapter's supplementary data. A datafile of the literature in context of this interaction matrix is available online in the published article (2).

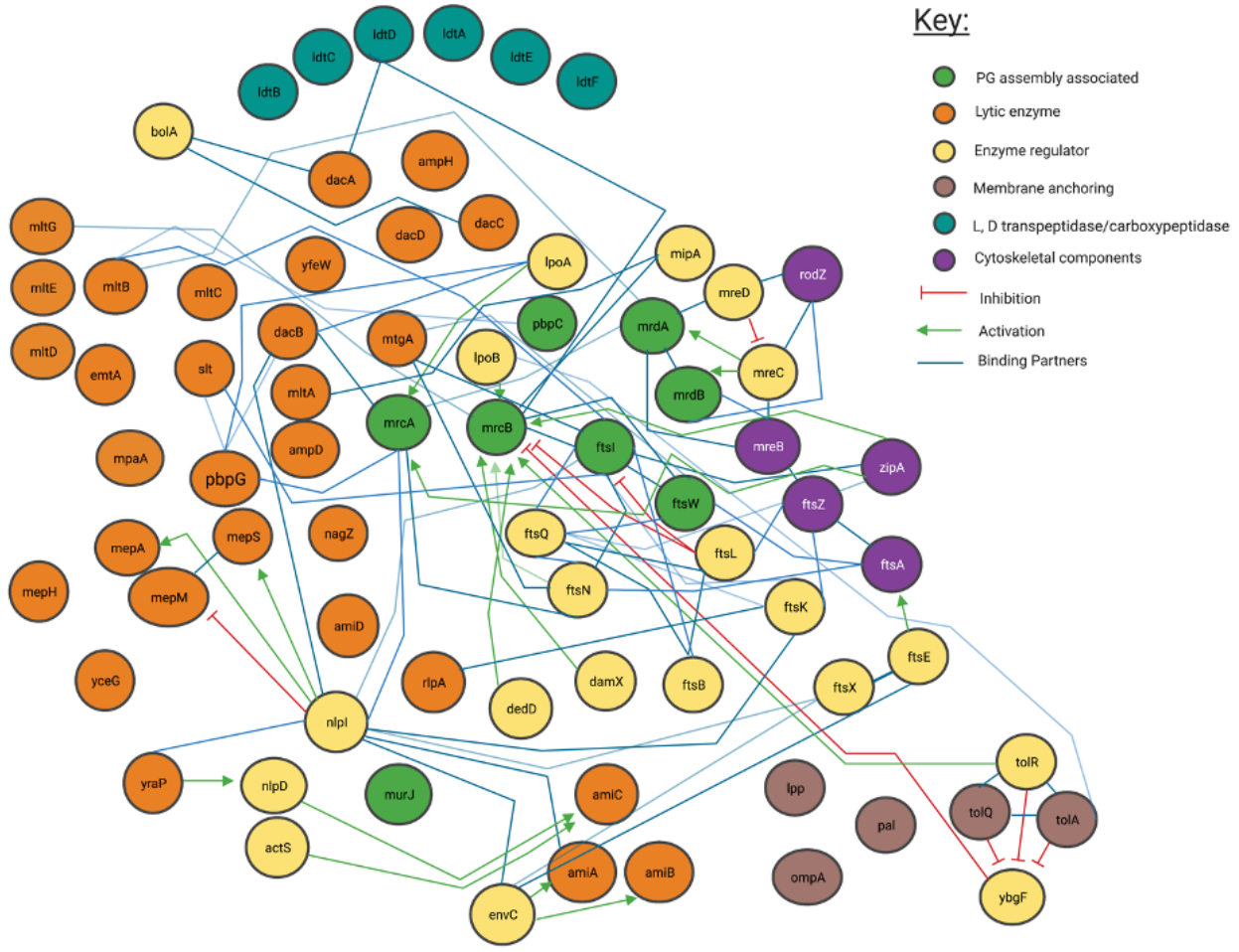


Figure 1.1.3. Interaction network of peptidoglycan modifying enzymes and their partners from Literature inspection.

Interaction map of peptidoglycan-associated proteins sorted by enzymatic action. Network structure determined by STRING, with manual addition of interactions through literature associated with each protein. Reference matrix available in supplementary data.

The networks of literature and genetic interaction in Figure 1.1.3 created through use of an interaction matrix of the literature (Supplementary data "Network.csv") confirms already widely held theories, that the differences in central peptidoglycan formation units or "asomes" are often based upon changes in accessory proteins and there is rarely a fixed static complex or "asome". Figure 1.1.4 indicates complex cross-reactivity most clearly due to the additional interactions

predicted by data searching, however with proteins such as “PBP1b/mrcB” of Figure 1.1.3 and shown by the literature to already be sustaining more interactions than viable in a single interaction state these interactions are likely timed or in transience. This compliments several proteins theorised to be within multiple complexes as indicated by Figure 1.1.4. These figures show the overwhelming complexity of the PG protein network, as a collection of many complexes.

The pattern of protein-re-use and interdependence however is not constant, for example the mono-glycosyltransferases FtsW and RodA have already been shown to be reliant on their partner class B PBP interactions for function (41,47,48). These dependent glycosyltransferase proteins form a complex with specific class B PBPs such as *mrdA*/PBP2 for RodA and *ftsI*/PBP3 to FtsW (Figure 1.1.2) and in doing so create glycosyltransferase and transpeptidase peptidoglycan machines (4,6,49); known as Sporulation, Elongation and Division (SEDS) complexes which have alternative activity dependent on additional cytoskeletal and regulatory proteins. In *E. coli,* simplistically this includes the RodA-PBP2 (*mrdB-mrdA*) complex for elongation and the FtsW-PBP3 (*ftsW-ftsI*) complex for cell division. These complexes act as nodes, displayed in Figure 1.1.3 and 1.1.4 creating the basis of the dependent complexes of the elongasomes and divisomes, which too interact with one another, with diverse interactions among proteins with which they interact such as LpoB, but also direct interactions such as those between PBP3 and PBP2. Existing alongside these functionally relevant dependent multimers, bifunctional transpeptidases and glycosylase class A PBPs such as PBP1B and PBP1a interact too, but have also been shown to form their own complexes and act independently of these SEDS machine nodes to modify and synthesis peptidoglycan, introducing further complexity(50).

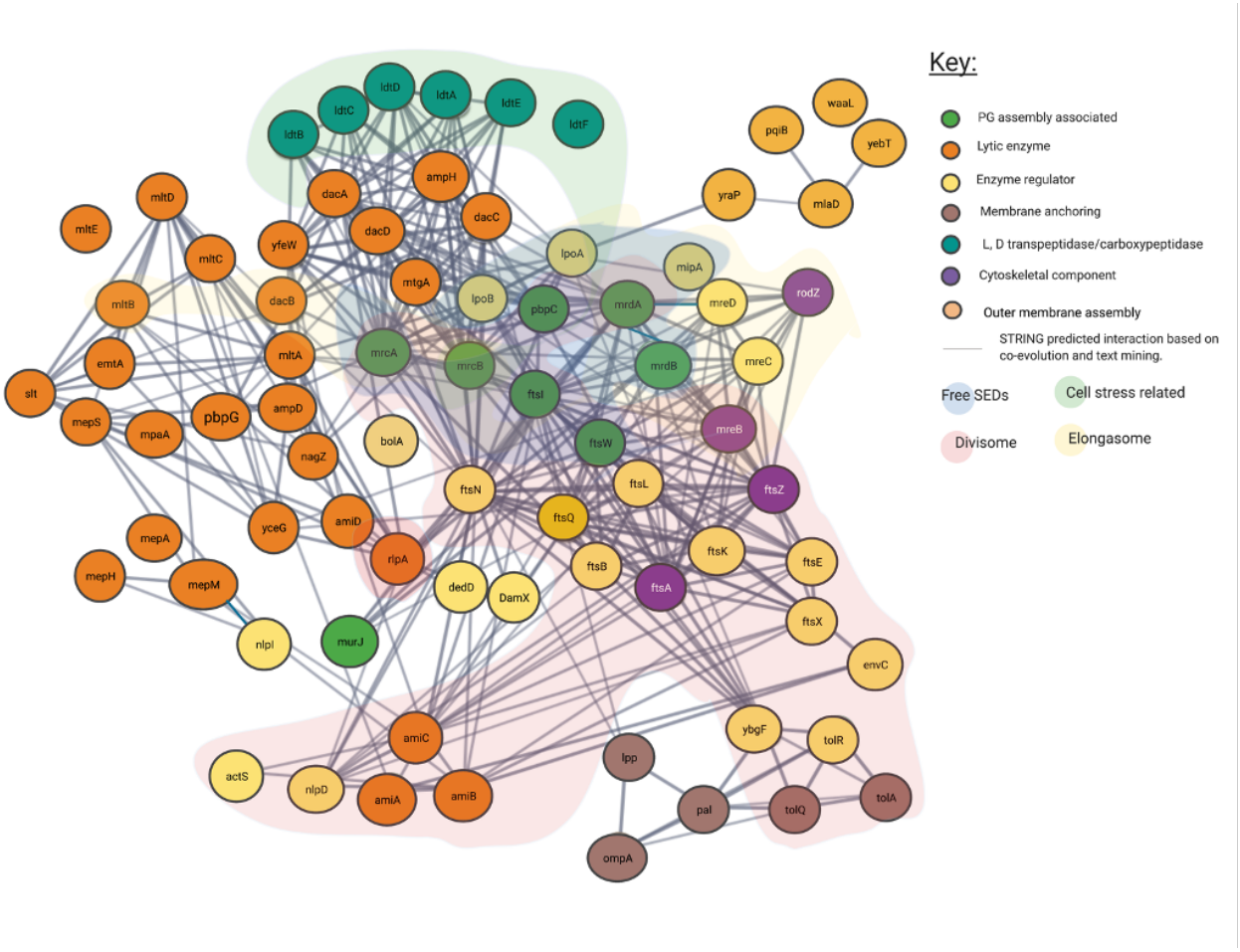


Figure 1.1.4. Interaction network of peptidoglycan associated proteins using STRING

Interaction map of peptidoglycan associated proteins, sorted by enzymatic action. Network structure determined by STRING, with systematically added manual additions. SEDS complex interaction type indicated by area colouring and arrows. Interactive network link: https://version-11-5.string-db.org/cgi/network?taskId=bIzLkBRoqjLb&sessionId=bBi0rwtoih3p

Figure 1.1.4 indicates all the theorised interactions of the complex machineries creating peptidoglycan. To understand the general mechanism for peptidoglycan synthesis and modification across species in the context of all steps involved in creation, and contextualise the networks we show, a specific example can be called upon. One of the core peptidoglycan biosynthesis and modifying complexes is the elongasome; a biosynthetic complex of peptidoglycan manufacturing machinery spanning the periplasm and inner membrane

All the complexes shown, including the Class A PBP centric ones, contain a host of additional regulatory, structural and enzymatically essential proteins that are shared among them, with

interactions that span across “asomes”, each interaction determining their specific overall activity dependent on interacting protein concentration, local substrates and their overall lipid environment (Figure 1.1.4). Complexes of proteins such as the web of potential interactions shown in Figure 1.1.3 and Figure 1.1.4 drive cell envelope synthesis. This review attempts to explain these interactions and their importance through stories presented by the current literature and investigates the specific nodes to which they centre.

Cytoskeletal proteins create nodes of complex formation

As shown in Figure 1.1.4, some of the peptidoglycan modifying enzymes can be directly linked to cytoskeletal proteins. These cytoskeletal proteins also act as molecular treadmills (29,30,51)(Figure 1.1.4). Typically, two treadmills run around the circumference of the Gram-negative cell, one for division and one for elongation made of FtsZ and MreB filaments respectively (51). FtsZ is a cytoskeletal component (52), that has been shown to localize and move along the cell circumference during cell division in individual strands that make up a macromolecular “Z ring” (22). Soon after the discovery of FtsZ as a division-essential component, it was shown to interact with the peptidoglycan synthesis machinery of FtsW and FtsI/PBP3 and the regulatory proteins FtsQ, B and L (31,32). Since then it has been associated with many other penicillin binding proteins including PBP1b (29,30)(Figure 1.1.4). It is thought that the polymerization of FtsZ is responsible for the directional movement of the SEDS complexes during division, constriction and septation (51,53). As a result, this is an important cell division protein; an antagonist of FtsZ polymerisation, viriditoxin causes cell division defects (54). In Archaea, its analogous FtsZ and often multiple of it’s paralogues are integral as division orchestrating proteins associated with pseudomurein laydown, causing cell division defects if not genetically present (55). Similarly, MreB is another cytoskeletal component implicated in cell shape (56)shown to co-localise with the elongasome associated proteins during cell growth. It has been shown to bind the Mur ligases which produce the lipid II precursor, the MreB polymerization antagonist A22 too causes cell morphology defects, highlighting the importance of both MreB and FtsZ as shape determining proteins (57).

Though FtsZ and MreB are important for correct cell growth and cell division, the literature has shown some of the peptidoglycan modifying machinery may only transiently attach to FtsZ and MreB treadmills, although it is not always necessary for their function (30,56). Recent models of transient FtsI-FtsZ interactions by “Brownian-ratcheting” would suggest the peptidoglycan

production and modification complexes move in and out of interaction with cytoskeletal components by a transient system of attachment to a cytoskeletal component, followed by protein wandering, allowing the peptidoglycan altering complex speeds to differ dependent on degree of cytoskeletal attachment (58). This “Brownian-ratcheting” model hypothesizes a zone of active peptidoglycan producing, slower moving complexes near the faster moving FtsZ rings or MreB filaments that transport inactive complexes in a dynamic equilibrium of interaction with the cytoskeletal nodes .

The “Elongasome” is a collection of multiple complexes

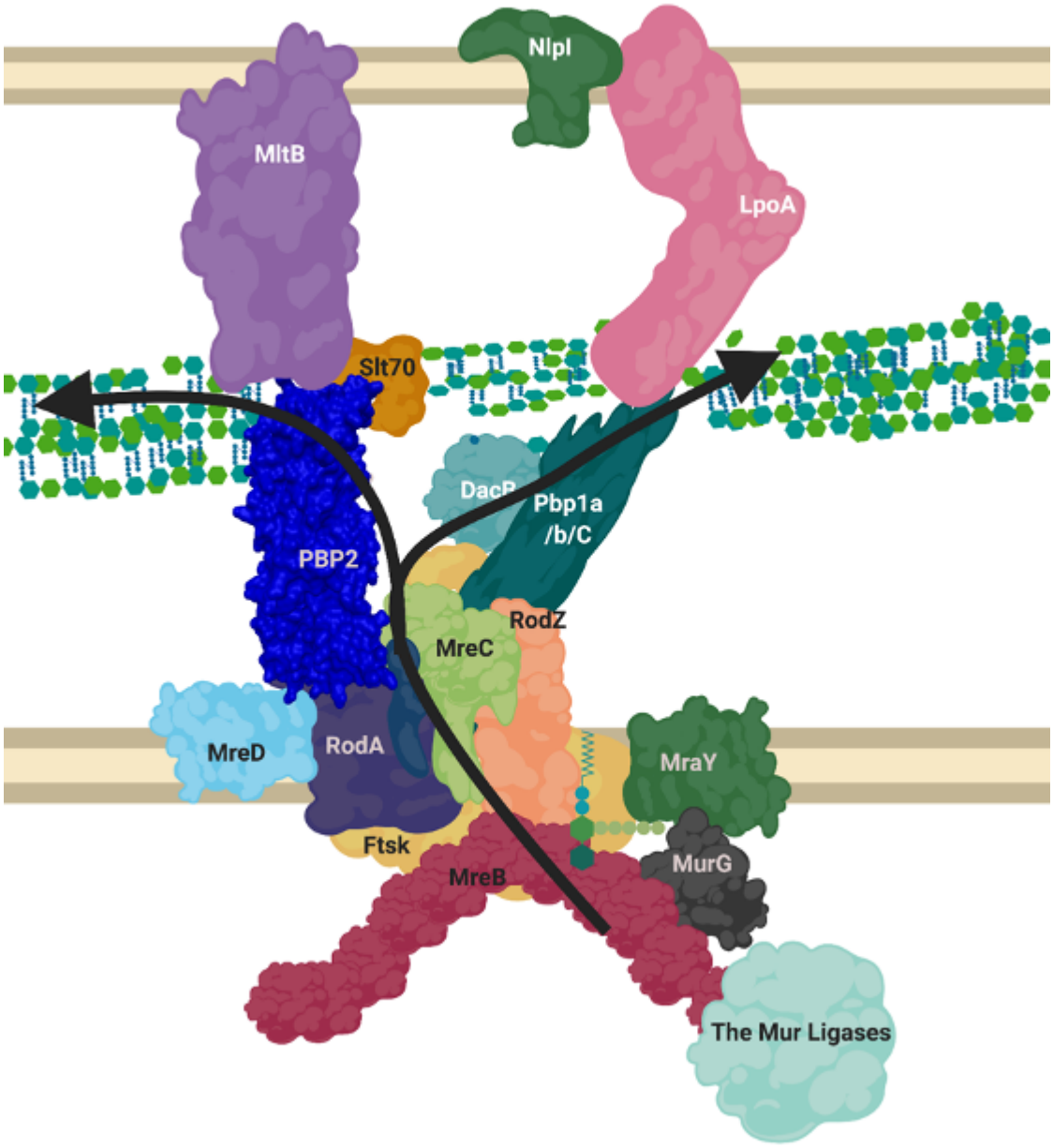


Figure 1.1.5. Proposed layout of the RodA orientated elongasome complex

A sketch of peptidoglycan insertion by a proposed formation of the elongasome complex. MreB and C sequester enzymes to the elongasome complex, including RodZ. MurG transforms Lipid I to Lipid II, MurJ/FtsW/RodA flip this into the periplasm. The PBP1a/b/c and/or PBP2-RodA complexes transglycosylate the lipid II into the nascent strand. DacB may remove the terminal D-Alanine from

pentapeptide and a transpeptidation reaction occurs through catalysis from transpeptidases PBP2 or PBP1a/b/C. Lytic transglycosylases MltB or Slt70 open new nascent strands for modification as machinery moves forward. LpoA/B along with other regulatory mechanisms listed in SI Table 3 control PBP activities. This diagram disregards the dynamic nature of the SEDS complexes and MreB ratcheting. Figure created in BioRender.

The elongasome complex (Figure 1.1.5) contains the components of generalised peptidoglycan synthesis (Figure 1), but we postulate as have others, on the basis of our bioinformatic analysis and the literature (Figure 3), that the complex also contains class C PBP D,D-carboxypeptidases and lytic transglycosylases to modify peptidoglycan structure and prime it for attachment with new peptidoglycan (5). This model contains the core monofunctional class B transpeptidase PBP2 which inserts its single transmembrane helix into the seven transmembrane helices of RodA, activating it as a glycosyltransferase (6) .This central peptidoglycan biosynthetic capability of the core of RodA-PBP2 is then augmented by the bifunctional Class A PBP, PBP1a also, which associates with the core complex (59).

The scaffolding and regulatory proteins of this elongasome, RodZ and MreC respectively, both interact with cytoplasmic MreB, as well as binding the transpeptidase PBP2 (60). It has also been shown that MreC, anchored in the membrane with a single transmembrane helix, regulates the crosslinking transpeptidase activity of PBP2, and glycosyltransferase activity of RodA via an interaction of its own periplasmic globular domain with the pedestal domain of PBP2. MreC may also have a role in binding to PBP1a among other components (60), this is especially interesting as a recent paper shows that *P. aeruginosa* MreC forms large self-storage filaments within the periplasm to likely regulate MreC concentration in the membrane (61). The integral membrane protein MreD has been shown to act as a coordinating partner to MreC in its interaction with RodA and PBP2, with an antagonizing effect, however the interaction interface and the regulatory mechanism they perform itself is not yet known, ribosomal studies suggest MreD levels are half that of MreC, indicating MreC's storage filaments are likely integral to proper regulation of this system among others possible (40).

The elongasome is transiently linked to the cytoplasmic MreB cytoskeleton of Gram-negative bacterial cells (31)(Figure 1.1.5). The "Brownian-ratcheting" mechanism of FtsZ (see above) could also apply to MreB interaction, considering the similarity of FtsZ and MreB as cytoskeletal protein homologues. This would suggest that the elongasome complex may instead move in and out of interaction with MreB, rather than remaining always associated (58). This model would be in agreement with the RodA-PBP2 complex moving along the cell circumferentially alongside MreB bi-modally active and inactive, at different speeds and with alternative partners (29).

Beyond the cytoskeletal interactors, regulation of this elongation apparatus has been shown to require the outer membrane regulatory lipoproteins LpoA and LpoB (42)(Figure 1.1.3). LpoA and LpoB span the periplasm to make contact with PBP1a (62) and PBP1b respectively and form synthetically lethal pairings upon genetic deletion, underlying their essential regulatory role (42). LpoA stimulates the transpeptidase activity of PBP1a specifically, which in turn upregulates its glycosyltransferase activity and peptidoglycan production (63) and by contrast LpoB has been shown to stimulate both PBP1b transglycosylase and transpeptidase activity(64) Recent analysis of the kinetics of the related PBP1b-LpoB pairing required for cell division, shows that LpoB is an effective on/off kinetic switch for peptidoglycan transpeptidation by PBP1b (42,64). Clearly therefore, the elongasome contains multiple overlapping and seeming duplicate activities and interactions, but this almost certainly belies the complex network of interactions required for optimal peptidoglycan biosynthesis. One interpretation of this complexity is that the central RodA-PBP2 complex is required for production of a peptidoglycan scaffold for elongation which the PBP1a-LpoA pairing (connecting inner membrane-based synthesis with outer membrane control) then overlays with additional glycan stands and crosslinks, required to produce a complete layered structure (65).

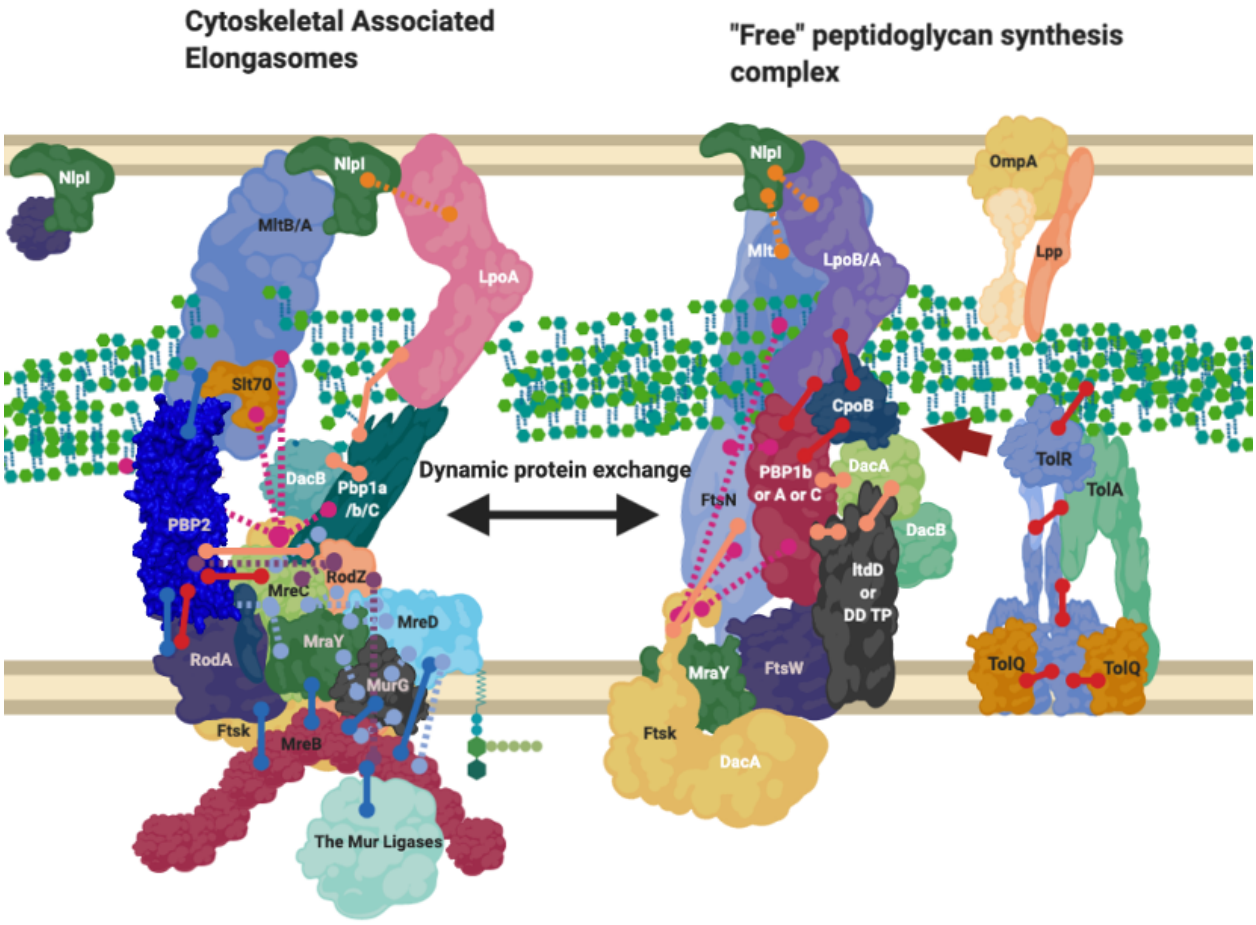


Figure 1.1.6. The Elongasome complexes - Interactions of the elongasome proteins

The elongasome has been shown to be a series of complexes which are either attached to or not attached to the MreB filament. Localisation-dependent studies have shown the enzymes involved co-localise at the MreB filament [8,9], and each interaction is cited. This diagram is one of many possible configurations based on the current data, with incorporation of the flippase not mentioned. Proteins have been shown to be shared between complexes. Diagram rendered using BioRender. Connections: Purple (66), Red (6,28,40), Cyan (67), Orange (68), Pink (69), Peach (28,48,59,62,70,71), Blue (4,72–75).

NlpI acts as a facilitator of PBP nucleation and complex interaction.

The literature has shown peptidoglycan associated enzymes interact with a great deal of enzymatically inactive structural proteins (Figure 1.1.3). A recent paper (68), has shown there to be an outer membrane protein with a large number of protein-protein interactions, dominating our interaction networks called NlpI. It is postulated to act as a scaffold for peptidoglycan associated proteins and is required for their formation and control. Microscale thermophoresis, pull-down and bacterial two hybrid studies have shown that NlpI can form trimeric complexes

with PBPs, for example MepS-NlpI-DacA, MepS-NlpI-PBP7 and LpoA/PBP1a/NlpI among many others (68). NlpI regulates a set of peptidoglycan hydrolases, as well as being able to form a trimer with PBP1A and LpoA(68). Its absence leads to increased vesicle creation (76) suggesting its importance to the cell envelope, and Banzhaf et al. concluded NlpI may facilitate the interaction and/or change the composition of the peptidoglycan editing complexes, controlling the potentially harmful hydrolases and facilitating regulation of other proteins (68).

NlpI's dispersal around the cell indicates it is likely involved in many of the complexes responsible for creating peptidoglycan, including the divisome and elongasome, and possible intermediary complexes that likely exist between those in turn (26). As a result, its abundance across the entirety of the cell and regulatory ability suggests it may be part of the system of dynamic protein complex formation this review focuses upon (Figure 1.1.3, 1.1.4) and worth noting, however its role beyond this is not well known (26,77).

The Divisome is a series of complexes controlled by cytoplasmic events

The reasoning behind the complex series of interactions in Figure 3 and 4 can be more fully understood in the context of the cell cycle, as not all interactions must occur simultaneously, but rather on a cell stage basis. The divisome is responsible for the division of cells, it is a peptidoglycan modifying complex that has been studied for some time and exists as a series of complex protein-protein interactions (Figure 1.1.3), but these have been shown to occur at intermediate stages and to be dynamic (65). The divisome's function is similar to that of the elongasome's with analogous flippase, transglycosylase and transpeptidase partners, which have been shown to be dependent upon a cascade of interactions (65). The divisome proteins that modify peptidoglycan such as PBP3 and PBP1b are not always present in the divisomes central transglycosylase FtsW, as they change their cellular localisation dependent on the cell cycle and by their interactors (Figure 1.1.6) (3). There is a higher concentration of peptidoglycan synthesis and hydrolysis enzymes at the septa during cell division, in a series of stages and cascades, suggesting a dynamic system much like the elongasome system, where protein composition changes over time as need and function changes (67,72,75).

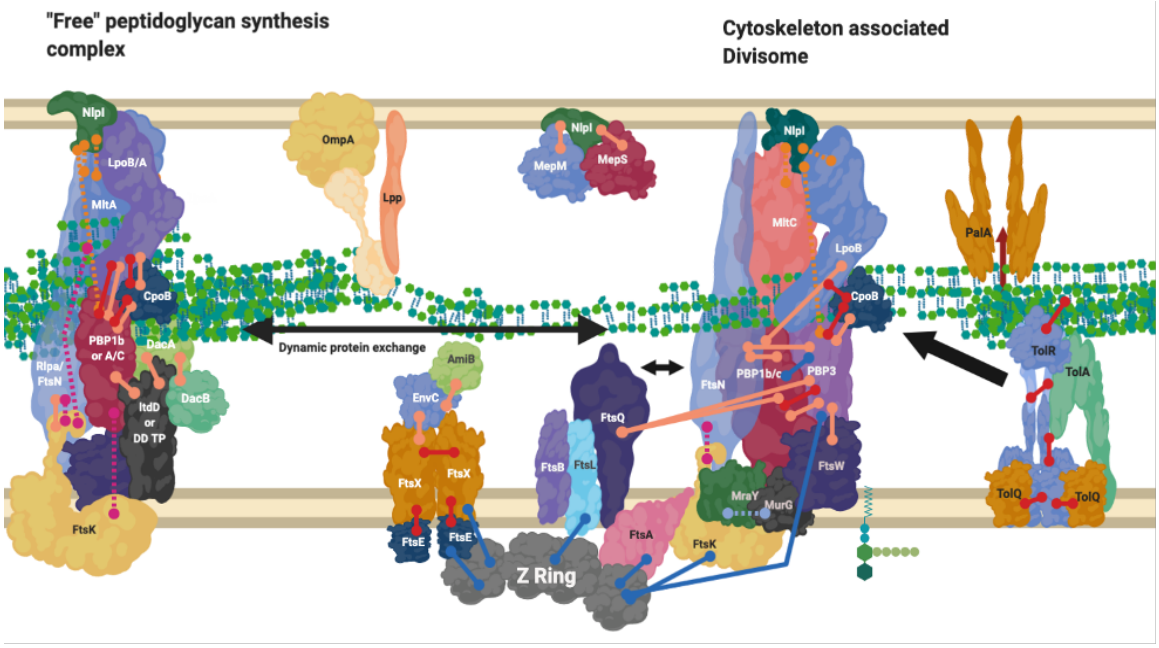


Figure 1.1.7. Divisome interaction network at the mature divisome

A diagram representing the interactions of the divisome proteins. Certain proteins within the divisome interact directly with the FtsZ filament. Localisation dependent studies have shown the enzymes co-localise at the FtsZ filament or in the pre-septal region (PIPs). Models suggest interactions are only brief (58). This diagram is one of many possible configurations based on the current data. Connections: Purple (66), Red (6,28,40), Cyan (67), Orange (68), Pink (69), Peach (28,48,59,62,70,71), Blue (4,72–75).

Networks of interaction presented in Figures 1.1.3, 1.1.4 and 1.1.7 make clear the abundance of protein interactions possible. Division must account for osmotic conditions, cytoplasmic events, antibiotic challenge, and periplasmic protein complexes, whilst also maintaining the stability of cell envelope layers to prevent cell lysis, and finally allow septation.

Proteins interchange between complexes, and complexes interact.

Throughout this review, it is mentioned that proteins can exist in more than one complex (Figure 1.1.4). Despite the notion that PBP2 and PBP1a are normally associated with the elongasomes as discussed above, they have also been found at the division site, division proteins and have been shown to interact with the division related cytoskeletal, protein FtsZ (75,78)(Figure 1.1.3). A hypothesis involving a balance between the elongasome and the free-floating or unbound elongasome was investigated in the Gram-positive species, *Bacillus subtilis* (79), which found that the homologue of PBP1a is able to move independently of MreC homologue or RodZ in the cell (80). In the same study, the quantity of MreC and PBP1a also determined the cell width in *B.*

subtilis, suggesting this balance of two systems: one elongasome free of cytoskeletal proteins and RodZ, which can diffuse across the circumference of the cell to allow radial expansion, and one which interacts transiently with the cytoskeleton of *E. coli* and dominates the elongation process and plays a part in morphology determination in *B. subtilis*, which could also be indicative of Gram-negative systems (Figure 4)(75,78,81) .

PBP1b (encoded by the mrcB gene) is a bifunctional glycosyltransferase and transpeptidase enzyme that interacts and plays part in the regulation of the divisome, it dominates Figure 1.1.3 and 1.1.4 as a node with high levels of interaction, beyond which is reasonable to exist at any one time simultaneously (28). In contrast to PBP1a (encoded by the mrcA gene), PBP1b has been postulated to have division complex roles as well as a wandering role (4). PBP1b and its partner activator LpoB have been shown to be essential to peptidoglycan rebuilding in peptidoglycan-deficient spheroplasts of *E. coli* (50). Their essentiality outside of division processes to create new peptidoglycan in spheroplasts, suggests that PBP1b must play a major role in the creation of new peptidoglycan, which in wildtype cells (*E. coli*) is carried out 70-80% by the bifunctional PBPs such as PBP1b and PBP1a which have roles in the elongasome and divisome (4). This dependence suggests either; the cytoskeleton-bound or free "elongasome" for cell growth including PBP2 and RodA involves PBP1b more than just transiently, or that more than the static model of the elongation machinery RodA-PBP2 exists and PBP1b has a separate role.

A "free" diffused PBP1b, and PBP2 have been observed independent of MreB/FtsZ systems by fluorescent localisation (4,29). The interactions shown in Figures 1.1.3 and 1.1.4 suggest many possible complexes, that vary in their composition, position, and association with the cytoskeleton by PBP1b, PBP1a and PBP2.

The complexes that contain these interactive proteins may also interact. The elongation and division machineries share common protein components and interactors (Figure 1.1.3, 1.1.4 and 1.1.7) with the elongation machinery associated PBP2 even transiently localising to the Z-ring during cell division (75). PBP2, a protein known to be integral to the elongation machine, has also been shown to interact with PBP3 (*ftsI*), with PBP2 knockout studies revealing division defects. In addition to division and elongation related localisation, the peptidoglycan synthase proteins PBP2, PBP1a, PBP1b and FtsW have also been shown to localise diffusely around the cell (4,28,30), moving independently to the cytoskeletal-associated elongation and division complexes.

During the midstage of division, MreB and FtsZ appear to co-localise at the Z-ring whilst treadmilling. It has been postulated that enzyme exchange of these proteins between the divisome and elongasome may occur through an interaction with the cytoskeletal components MreB and

FtsZ (75). This would support the Brownian ratchet model theory of cytoskeletal protein control, citing a transient interaction rather than permanent interaction of the PG machinery with FtsZ and MreB(58), allowing for the exchange of proteins between cytoskeletons more easily.

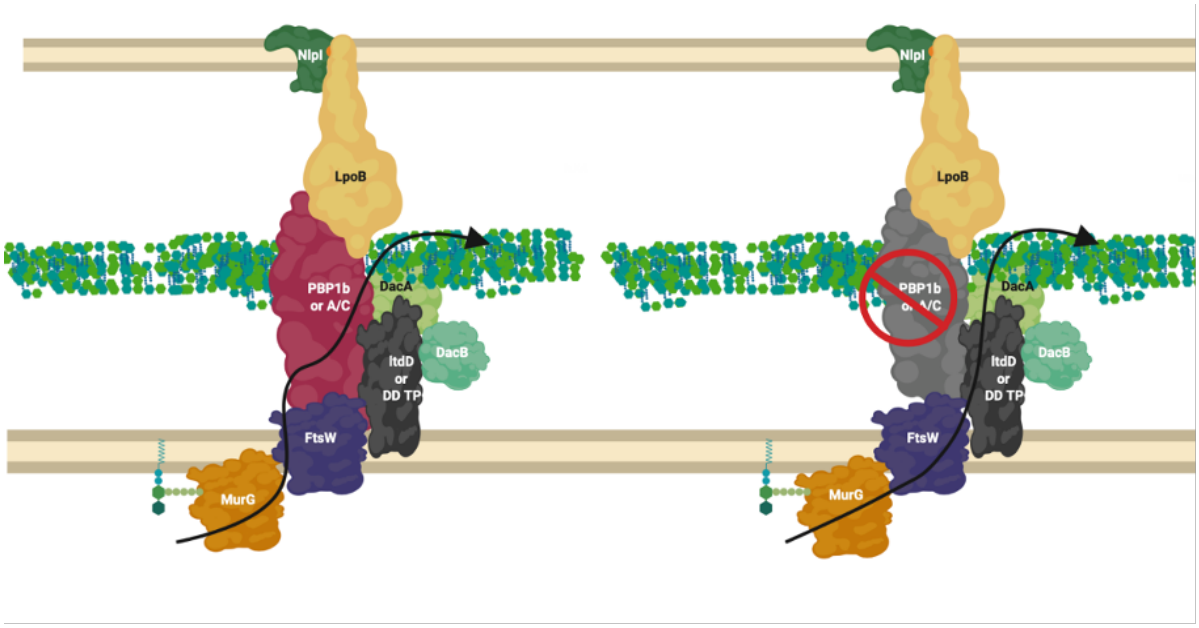
PBP1b and PBP2 localize to the septum adjacent to the Z-ring during division and become delocalized from the septum in an *mreB* knockout strain (31). Mutation of the FtsZ-interacting residue of MreB similarly delocalises PBP1b and PBP2 from the FtsZ rings, despite successful MreB and Z-ring formation. The unused Z-rings remain as "locked" stripes of unsuccessful division sites, and cells containing these Z-rings stripes become filamentous cells. These "locked" Z-rings fail to incorporate fluorescent single D-amino acid probes such as HADA denoting new peptidoglycan biosynthesis, and thus do not actively synthesise peptidoglycan, while the elongation enzymes along the rest of the cell remain functional and successfully incorporate HADA throughout the rest of the cell (31). This may be due to the absence of the PBP1b-FtsN interaction which would normally interact with MreB, and transition to divisome interaction FtsZ to relieve the FtsQLB inhibition of PBP3 (41,48), without this MreB-FtsZ transient reaction this inhibition remains in place. The cytoskeletal component amino acid knockouts described above, in conjunction with a known lethal PBP2 knockout phenotype, show the necessity of elongation enzymes such as PBP2 to also be required for division and highlights dynamic interchange between complexes (32).

Alternate protein complexes exist, containing 3-3 crosslinking L, D transpeptidases as an alternative to 3-4 crosslinking PBPs important for antibiotic resistance

Figure 1.1.3 and 1.1.4 and the literature they represent indicates enzymes not yet confirmed to be integral machineries to be involved in interactions in multiple complexes. There is good evidence that during extended growth, osmotic cell stress and some instances of β-lactam challenge, 3-3 crosslinking increases in the peptidoglycan of many Gram-negative bacteria (82). This form of crosslinking is catalysed by L, D-transpeptidases (Supplementary Figure 2). The literature, and genetic predictions indicate these proteins interact with many other PBP related proteins (Figures 1.1.3 and 1.1.4).

The L, D-transpeptidase LdtD, has recently been shown to interact with peptidoglycan endopeptidase DacA and bifunctional synthase PBP1b by microscale thermophoresis (70). Following PBP1b inhibition by β-lactams, LdtD compensates for the loss of 3-4 crosslinking by 3-3 crosslinking, enabling cell survival in the presence of β-lactams. In this situation LdtD could compensate for part of PBP1b's normal bifunctional role as a transpeptidase (when in complex

with the transglycosylase FtsW and PBP3) by replacing the PBP catalysed 3-4 transpeptidase activity with 3-3 crosslinking, using the PBP1b and FtsW transglycosylase glycan chain products as substrates. In this scenario the D, D-carboxypeptidase of DacA, shown to be essential for β-lactam resistance mediated by LdtD (63,74), modifies the pentapeptide by removal of the terminal amino acid to provide a suitable tetra peptide substrate to LdtD. This is necessary because LdtD requires a tetrapeptide as a substrate (63,74). Depending upon the availability of suitable substrates and some environmental conditions, PBP1b will also generate tetrapeptide products which could become substrates for the Ldt's (57,74). It is likely this isolated complex is only one of many complexes incorporating the non-canonical peptidoglycan crosslinkers L, D-transpeptidases (Figure 1.1.8).



(a) (b)

Figure 1.1.8. LdtD in complex with PBP1b and DacA

(a). PBP1b active complex performs 3-4 crosslinking (b). On β-lactam challenge PBP1b activity is reduced, allowing LtdD, in complex with DacA and DacB, to take over crosslinking and allow cell viability (70,83). The complex represented above is one of many possible complexes. The flippase is represented by FtsW due to its potential bifunctional role for simplification but likely involves MurJ, this is one of many possible configurations. MurG ligase has been shown to interact with FtsW (71,84).

An elongasome complex synthesises peptidoglycan in curved Gram-negative bacteria.

Beyond the simple models presented as elongation or division mechanisms, many species of Gram-negative bacteria are more morphologically complex than just rods or spheres (85) and the complex of proteins presented for *E. coli* (Figure 1.1.4). These species require the peptidoglycan machinery to be altered compared to these classical exemplar species. *Campylobacter jejuni* and *Helicobacter pylori* have been shown to contain hydrolytic L, D carboxypeptidase proteins essential to cell curvature (85) . The most well studied of these systems is *Helicobacter pylori's* elongasome and the conserved shape determinant (Csd) protein family. Knockouts of Csd6, CcmA and Csd5 all lead to curvature loss, along with peptidoglycan peptidases Csd1,2,3 4 and 6 (37,86,87)(Figure 1.1.9).

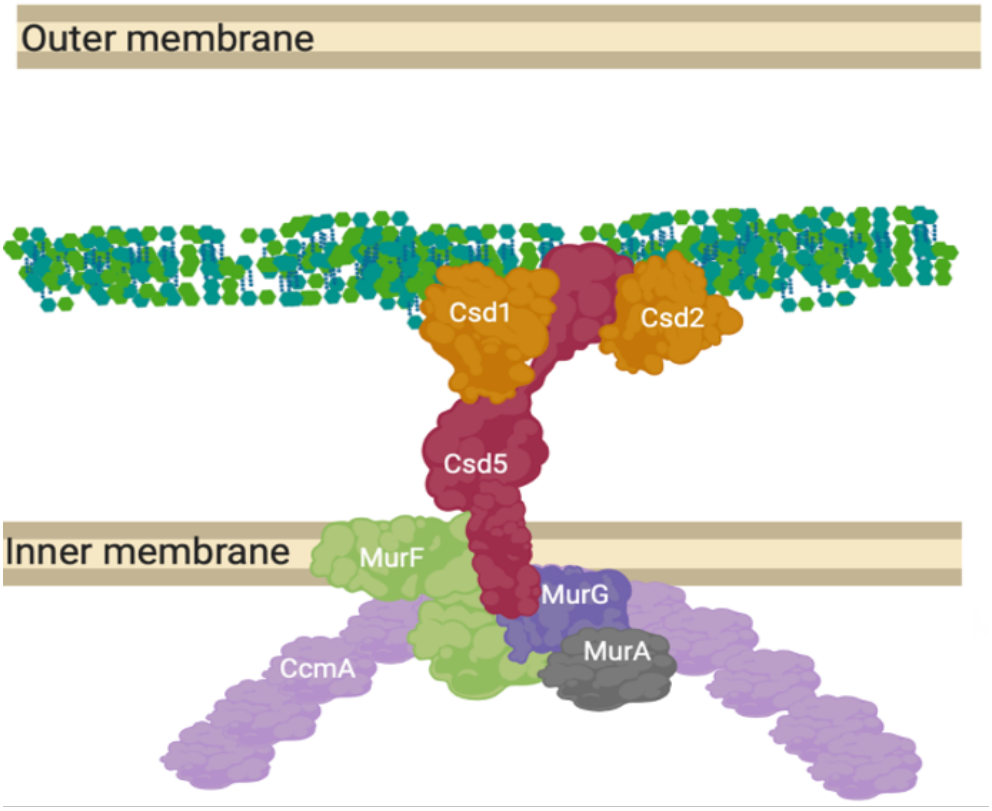


Figure 1.1.9. CcmA curvature promoting complex in Helicobacter pylori. The prospective anchored "Shapeosome" of Helicobacter pylori. Created using Biorender , hypothetical configuration (37,86,87).

The Shapeosome is a non-canonical peptidoglycan synthesis complex that facilitates cell shape of curved bacteria, associated with a cytoskeletal component like the elongasome and divisome. The connection to a cytoskeletal component again suggests a convergent and re-occurring model of shape-determining cell wall modification by complex formation. Csd5 binds peptidoglycan by its SH3 domains and interacts with synthesis related proteins MurG, MurA and MurF, as well as hydrolases Csd1 and 2 (37,86,87).

Unrelated cell envelope proteins must affect peptidoglycan-membrane linkage

Whilst peptidoglycan is generally regarded as being composed of intramolecular crosslinks (24), in some organisms, *L,D*-transpeptidase enzymes (LDTPs) can catalyse the formation of intermolecular cross links used to attach it to other macromolecular structures (88). It was shown in the 1970s that at least one third of Braun's lipoprotein (Lpp), an outer membrane protein, is bound to peptidoglycan (Figure 1.1.2)(89). In Figure 1.1.3/4, an abundance of outer membrane interactors and regulators, are shown to be part of elongation, and division complexes. It is has been known for some time that in species with and without Lpp, the outer membrane protein OmpA interacts non-covalently with peptidoglycan (90,91), but more recently, the literature has shown multiple OMPs across species to be connected covalently to the peptidoglycan by LDTPs (89). Specifically for example in *C. burnetii* the *L,D*-transpeptidase Ldt2 is required for covalent attachment of OMPs BbpA and BbpB to peptidoglycan (92). This creates a static covalent link between the outer membrane, peptidoglycan and inner membrane proteins such as Tol machinery across species (73).

Further still, bacterial periplasmic complexes such as pili, transport systems and flagella penetrate through the peptidoglycan layers (Figure 1.1.2) and are often able to transport proteins from the inner to outer membrane despite this linkage. This activity requires dynamic pores in the peptidoglycan layer (38,93). Observed movement of large proteins laterally across these fixed peptidoglycan layers, linked to outer and inner membranes may involve cleaving of the peptidoglycan using lytic transglycosylases to facilitate this movement, in addition to L,D carboxypeptidases, for Lpp release (94,95). These protein-peptidoglycan interactions must all be regulated to avoid cell lysis, as shown by multiple periplasmic proteins that ensure the cell envelope remains structurally stable during the pili, flagella and other transport complex movements. These local changes in cell wall structure due to large complex movements present the cell envelope and peptidoglycan as a more multi-ordered structure than a simple mesh, which must require uncrosslinking and regulatory mechanisms for growth facilitation, and cell envelope stabilisation, by regulatory mechanisms still unknown.

Chapter 1.1 Cell wall networks review summary

The peptidoglycan polymer's complex and essential role to Gram-negative bacterial cells requires an intricate set of proteins within the periplasm; to maintain its role in response to growth, during division and to ensure a stabilising permeable barrier is maintained in tandem with the inner and outer membrane. The literature has shown this takes place through a series of protein interactions forming SEDS complexes, and this is reaffirmed in predictive genetic and experimental interactions presented (Figure 1.1.3, 1.1.4). However, the full picture of experiments, when investigating the roles of each protein have shown that these complex interactions are not static in composition but are instead part of a web of interactions that allow many variant complexes to be in dynamic equilibrium depending on cell growth stage and need. This model is not yet complete. Models have been postulated of wandering and cytoskeletal-associated complexes such as the elongasomes and divisomes that create and modify peptidoglycan dependent on growth needs (Figure 1.1.6, 1.1.7). Alternative complexes have also been shown to exist for the antibiotic insusceptible *L, D*-transpeptidase enzymes which can allow crosslinking of peptidoglycan in the absence of the antibiotic susceptible PBPs (Figure 1.1.8). These must all occur in the context of structures that cross the periplasm and connect the inner membrane and outer membrane in partnership with other processes (95,96).

This model of large protein complexes evolved to allow for peptidoglycan modification dynamically across a growth cycle. This model has also shown to repeat convergently in other species, even among Archaea. Peptidoglycan modification systems, such as the shape determining complex oriented by the cytoskeletal protein CcmA in *Caulobacter sp.* (Figure 1.1.9*)* exist as convergent versions of the *E.coli* MreB and FtsZ based models presented in this review. The cytoskeletal component of some of these dynamic complexes across species, (FtsZ and MreB) treadmill along the circumference of the cell and have been shown to exchange protein partners during their interactions, and cytoskeletal or regulator absence/inhibition leads to growth defects. This evidence among others, shows an exchange of proteins which facilitate a change of complex composition over time by the associated machineries. Sometimes these complex changes are driven by specific cytoplasmic events and cascades, such as those that control the divisome.

However not all modification relies upon these cytoskeletons, as shown by PBP1b wondering motion across the cell (77). Indeed, a single protein could be required for multiple functions and complexes that exist at once (Figure 1.1.3), therefore these multiple protein localisations are in part controlled by affinity to the cytoskeletal proteins or outer membrane proteins anchored such as NlpI. This allows for fine control of complex composition in addition to regulation by protein affinity to local substrate (10).

The peptidoglycan research and AMR field has come to place importance on specific protein structure, and singular relationships with inhibitory/activator proteins in future antibiotic design. Our meta-analysis has shown the full picture so far likely extends beyond the crystallised complexes and static complexes, and revealing a great deal of flexibility, but also indicating the importance of specific nodal proteins in peptidoglycan synthesis. Research into macro-regulation of the complicated cell envelope complexes showcased will be an important step in the creation of new drugs that can overcome known mechanism of antibiotic bypass by protein exchange, but also postulate new methods for peptidoglycan and cell envelope disruption. Viewing these proteins in a systems context will be an important step in combatting resistance to antibiotics *in vivo*.

Author Contributions for above Chapter:

CG was the concept originator and main writer throughout, with written additions in their specialisations and editorial support from the other authors, with significance reflected in order of names. Supervisors DR, MB and AE provided advice, reworded sections, and reviewed the piece to ensure quality throughout.

Chapter 1.2 The Outer membrane synthesis machinery

Introduction

As mentioned, Gram-negative bacteria are formed of an inner membrane, constituting a bilayer of polar fats, a peptidoglycan sugar-peptide structure which surrounds and connects to this, and finally an outer membrane of outer leaflet lipopolysaccharide sugar lipid (LPS) and an inner leaflet of other further polar lipids. The negative charge of this LPS repels many antimicrobial challenges (9,97), the cytoplasmic inner membrane compartmentalises metabolic processes, and the peptidoglycan gives the shape to the bacteria. Between these layers, proteins are required to form pores between the layers, and ensure that nutrients are available to feed the cell and to allow for replication and growth(97). Finally a set of proteins must facilitate this growth in the presence of environmental pressures through membrane growth, peptidoglycan growth and overall cellular growth and division (2).

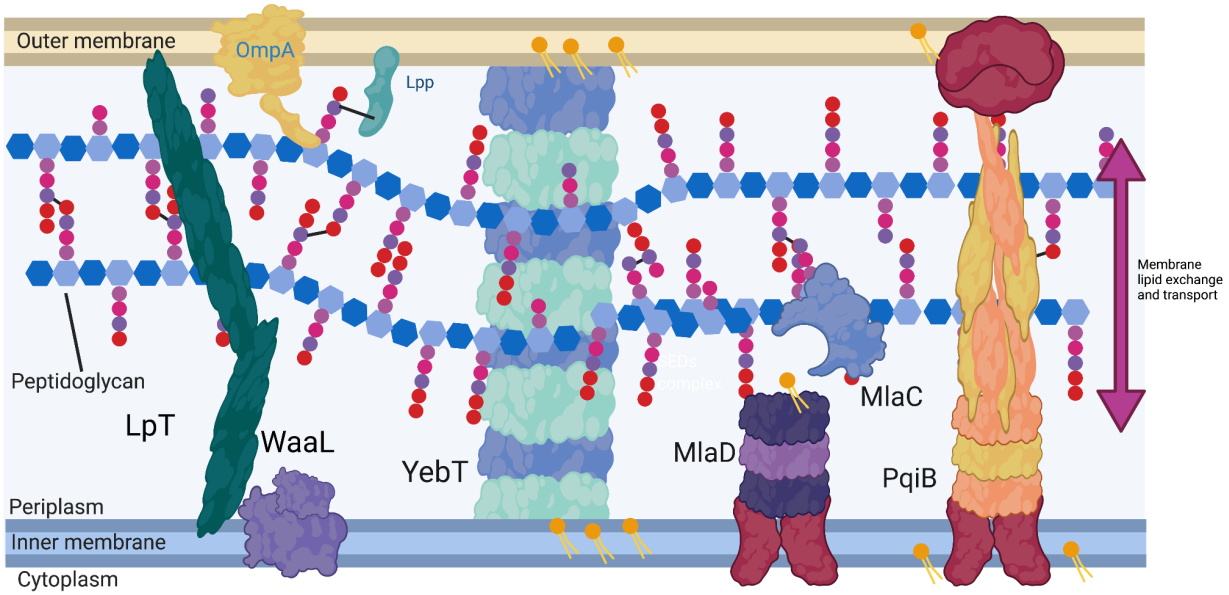


Figure 1.2.0 The cell envelope of Gram-negatives, and the MCE domain containing proteins that may facilitate the balance of lipid layers.

Cartoon representation produced by BioRender

The cytoplasmic processes, which make up the majority of enzymatic reactions and pre-cursors to the formation of bacterial cells, are adjacent to the inner membrane and facilitate this lipid layers growth (Figure 1.2.0) However the growth of the periplasmic layer peptidoglycan, and outer membrane lipid layers which are both separate from this cytosolic compartment (where the cellular control processes lie) must have proteins that move the cytosolic components out into the periplasmic space. This part of the chapter will focus on the proteins which are likely involved in lipid balance between the inner and outer membrane, as well as Lipopolysaccharide (LPS) insertion into the outer membrane, and outer membrane protein insertion.

Cell envelope overview and processes

The periplasm is the space between the inner membrane of Gram-negative bacteria, and the outer membrane. The periplasmic space has porins allowing entry of molecules less than 8Å across by passive diffusion, therefore this region is subject to conditions much like the space outside the

bacteria itself, with volatility in pH and osmolarity (98,99). This makes the proteins and processes within this space vulnerable to these changes and often evolved in tandem to these changes. The periplasm in Gram-negative bacteria, due to the presence of porins which allow for the movement of small molecules into the space therefore could be considered more difficult to regulate, with the majority of interactions so far noted as direct protein-protein interactions(2). The periplasm functions to protect the inner membrane from antibiotic challenge in Gram-negatives, as well act as a buffer zone between outside the cell and the vulnerable cytosolic interior. There is no DNA in the periplasm, and thus no translation or creation of proteins. This also means that metabolic processes, with the exception of peptidoglycan wall formation(59), are performed in the cytosol before export rather than within the periplasm.

Inner and Outer membrane formation and constitution

The inner membrane in Gram-negative bacteria is variable in terms of lipid composition, however in *Escherichia coli* contains PE/PG and CL ratio of 71.39:23.36:5.25 mol/mol%(100), these lipids are able to be simulated and are available on the CHARRM GUI (101) server for membrane building which has recently allowed for easy interrogation of the dynamics in this system along with other methods. These ratios and distributions are thought to not be homogenous however in fact compartmentalised into lipid rafts dependent on the flotilin class of proteins (102). Therefore, the locality of lipids is important to the construction and function of proteins.

The outer membrane in Gram-negative bacteria is variable in terms of lipid composition, also with a great deal of OMP protein, and a major difference in composition of the inner and outer leaflets (103), with LPS on the outer leaflet and a mixture of phospholipids on the innermost leaflet. (Figure 1.2.1) The most major of these outer lipids, and reactive component in humans is the LPS components, but LPS also provides rigidity and restricts lipid flow, as it interacts with itself (104). The inner leaflet of these two membranes, is composed of a variety of lipids, similar to the inner membrane, but dominated by phosphoethanolamine and cardiolipin. It is not yet known the difference in speed of the inner most leaflet lipids to that of the outer leaflet in the outer membrane, or the distribution of these lipids.

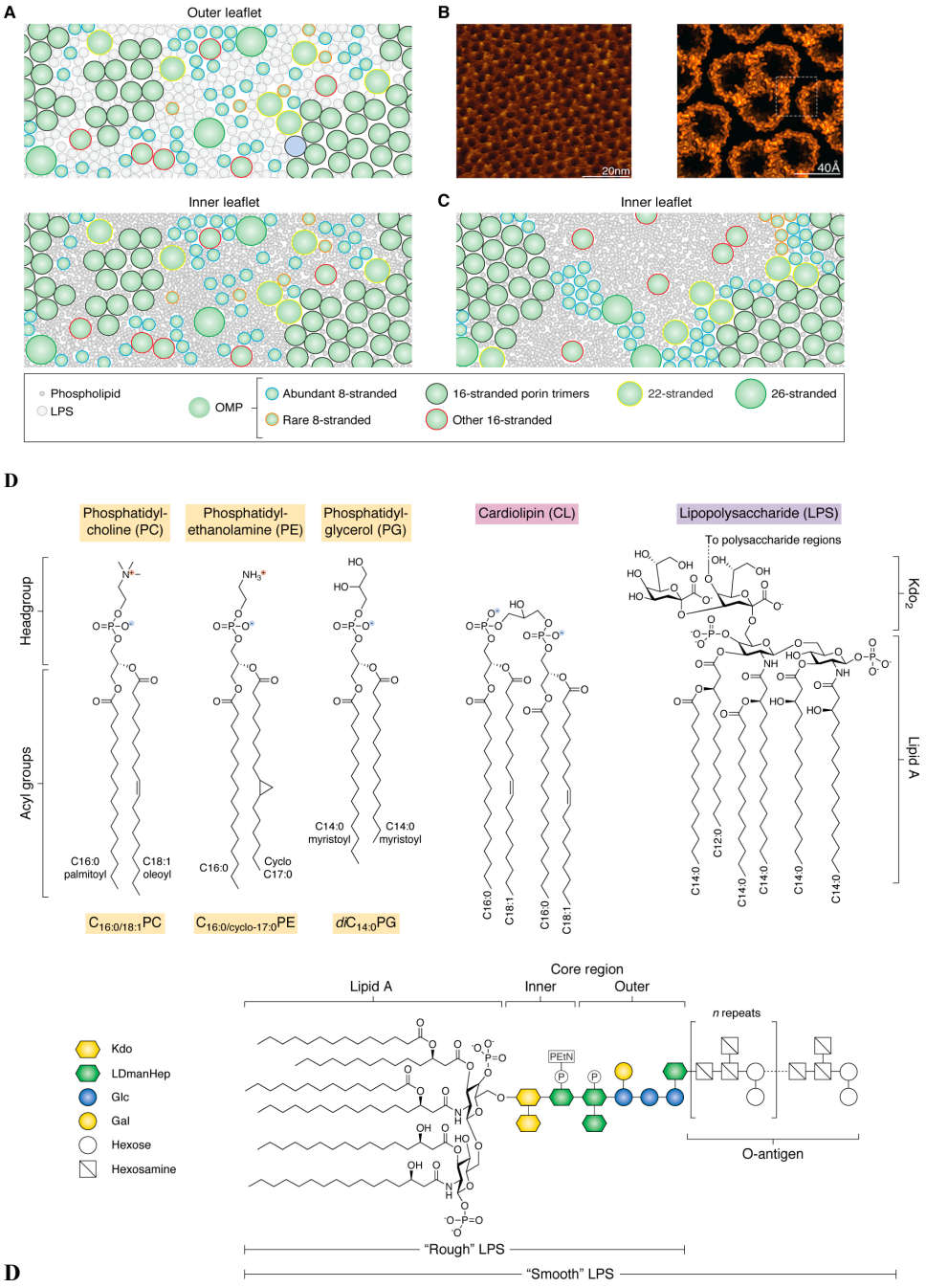


Figure 1.2.1 Lipid Structures and Outer membrane composition

A Inner and outer membrane composition *E.coli* **B** Atomic Force Microscopy of Outer membrane surface **C** Extreme clustering **D** Lipid structures PC, PE, PG, CL and general LPS. Adapted from "Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria' (105)

The MCE domain containing proteins and LPS transporters

The ratio of lipids therefore in the inner and outer membranes is important to cellular integrity as there are two distinctly stable layers, the peptidoglycan and the outer membrane opposed to a dynamic inner membrane. Thus the difference and constituents of cell envelope is catered for by a range of lipid transport systems, notably those that carry phospholipids and lipopolysaccharides (LPS) (Figure 1.2.2). MCE domain-containing proteins (10) and AsmA containing proteins such as TamB transport phospholipids (106), and the LptABDEFG is known to facilitate LPS transport (107).

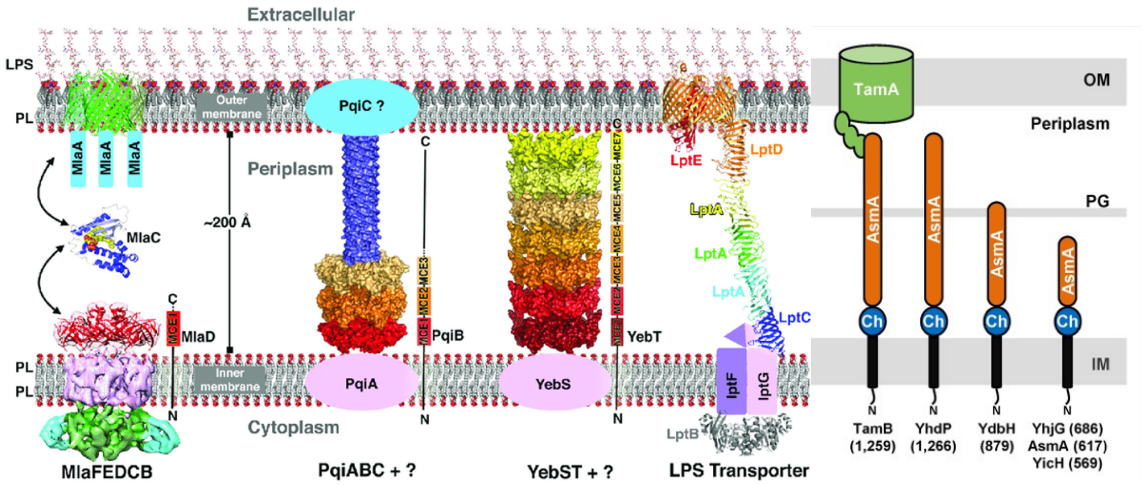


Figure 1.2.3 Outer membrane Homeostasis proteins.

MCE and AsmA domain-containing proteins stretch across the periplasm, creating shuttles or bridges by which phospholipids can travel the 200A membrane. (adapted from Ekiert et al 2017 and Douglas et al 2022)(108)

Phospholipid transport from the inner membrane is theorised for the most part to be made through interactions with MCE domain containing proteins (Figure 1.2.3) (10) and recently has also been postulated to be made by AsmA domain containing proteins. The *E. coli* MCE protein MlaD, forms a ring complex associated with an ABC transporter complex in the inner membrane, and has been crystallised with cardiolipin, and associated with other lipids such as phosphoethanolamine. MlaD is associated with MlaC, a lipid ferry which transports lipids to an OMP, with speculative connection to OmpF, facilitating transport through the MlaC interactions with phospholipid.

Another MCE protein 'PqiB' forms a hexamer ring associated with PqiA at the inner membrane, the C terminus forms a smaller non-MCE domain containing tube for lipid movement in a flexible manner, with which lipids transport through. PqiB has been associated with PqiC of the outer membrane, and acts as a lipid transport tube that crosses the entirety of the periplasm. The MCE protein YebS similarly forms a hexameric ring complex associated with YebT at the outer membrane. It acts as a lipid transport tube that crosses the entirety of the periplasm, with MCE domains making the majority of its structure. The AsmA domain was recently discovered as an integral component in TamB, present in a selection of other proteins which if deleted result in outer membrane instability (108)(Figure 1.2.3). The importance of these proteins to the formation of the inner leaflet of the outer membrane, suggests they are a possible target for localisation studies to determine the relevance of these proteins to divisional processes. The proteins are discussed in detail in Chapter 4 of this thesis, where their localisations and function are interrogated.

Porins and BAM apparatus

Porins, a protein framed around a Beta barrel domain, exist throughout the Outer membrane, and their insertion is facilitated by autotransport/self-transport in rare cases, and in the majority by the BAM machinery or Beta barrel insertion machinery(109,110), which accepts pre-folded Beta barrels and inserts them through a membrane distorting pore into the outer membrane. This family of Omp85-like proteins, that BamA is a part of, use carriers of OMPs such as SurA to carry their cargo, from the protein transport Sec/TAT secretion machinery partially pre-folded, to a series of POTRA or highly interactive protein domains, which accept the OMP protein and through the Omp-85 like protein insert the protein into the LPS and phospholipid containing outer membrane (Figure 1.2.3). Recent papers have shown BAM protein insertion to be division related, hinting at a connection between division and outer membrane biogenesis.(111)

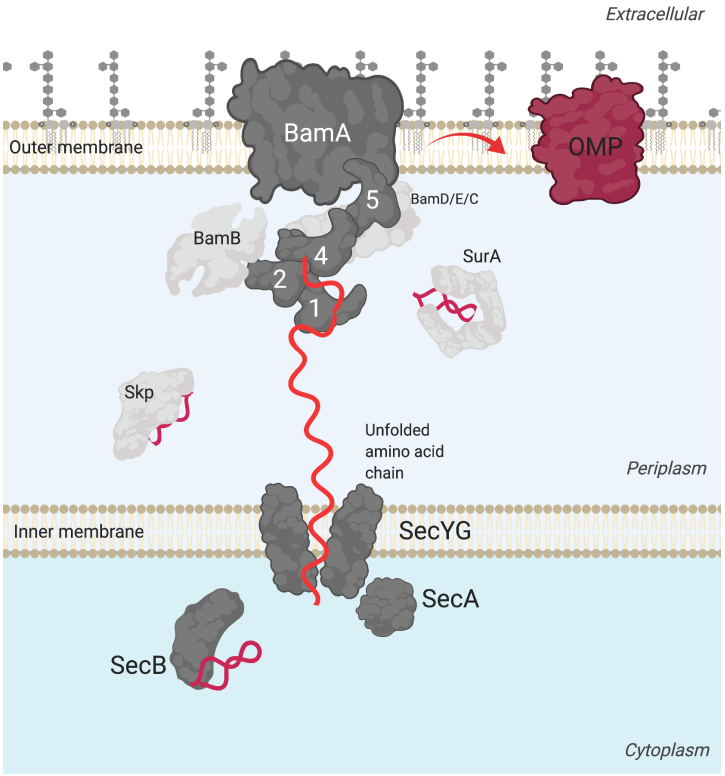


Figure 1.2.4 Outer membrane proteins and insertion by BamA

Insertion mechanism by BAM adapted from Graham et al 2021 "Autodock Vina In silica binding of loose chain amino acids to BamA POTRAs suggests their involvement is not selective" (112)

The collection of outer membrane proteins, which crowd the outer membrane(Figure 1.2.4), as well as an asymmetry in lipids in this layer, create an almost static collection of lipid, LPS and OMPs. This outer membrane layer must then also be coordinated with peptidoglycan to prevent vesicle formation, which is a symptom of over overexpression or lack of some of these proteins(76).

In summary both the peptidoglycan (Chapter 1.1) and cell outer membrane biogenesis machinery (Chapter 1.2) have a role in the integrity and function of the cell envelope, and therefore the research of both, from their foundation and molecular catalysis to the localisation and the connections between envelope layers is an important area of study.

Chapter 2: The mechanism of RodA glycosyltransferase and SEDS protein complexes.

As published in (2023) Cryo-EM structure of the Peptidoglycan Elongasome complex, and Lipid II biosynthesis mechanism, R Nygaard, CLB Graham et al, (Under 2nd Review Round- *Nature Communications*)

Chapter 2: Abstract

The peptidoglycan cell wall polymer of bacteria, is produced using two separate enzymatic reaction mechanisms. Firstly a glycosyltransferase reaction on the lipid II substrate creates a disaccharide sugar backbone chain from which pentapeptide side chains emanate to form the MurNAc sugar. Secondly a transpeptidation reaction occurs between pentapeptide side chains on adjacent sugar polymers to create a peptide crosslinked peptidoglycan layer. The penicillin binding proteins are responsible for the majority of the peptidoglycan peptide crosslinking found in cell division and cell elongation and are the target of beta-lactam antibiotics which have been a major force in antimicrobial chemotherapy for decades (4). Formerly, it was thought that the lipid II glycosyltransferase reaction was mainly catalysed by the corresponding activity of the bifunctional class A PBPs found in bacterial species. However, since 2016 it has become apparent

that the SEDS proteins are also able to catalyse this reaction (5,6) and are likely to be the major Lipid II glycosyltransferase activity responsible for cell division and cell elongation.

In this chapter I investigated the glycosyltransferase reaction used for the elongation and division of bacteria in SEDS proteins. In a collaborative manner, I have focused on RodA and its SEDs protein homologues, through a new *E.coli* cryo-EM structure of RodA and PBP2 in complex. This cryo-EM structure reveals flexibility of a channel within RodA through which two lipid II substrates bind before polymerization. In this study I have identified movement of RodA around the proposed lipid II cavities, have shown binding of a lipid in one of these cavities within our structure and simulated through informed (Double electron-electron resonance) DEER flexibility interrogation, the RodA-PBP2 *E.coli* complex active site based around the absolutely conserved D262A residue, in coordination with a series of conserved phosphate binding arginine residues. I have identified the catalytic residues responsible for the glycosyltransferase reaction conserved across species, through use of mutagenesis *in vitro* of this active complex in *E. coli*, and finally predicted how lipid II is transglycosylated by the RodA-PBP2 complex.

Chapter 2: Output

1. RodA-PBP2 Purification: I reproduced a NYCOMPS23-RodA-PBP2-HISX10 linked protein, allowing for purification of the complex in Warwick.
2. Co-evolution shows interaction of PBP2 and RodA:
3. RodA-PBP2 Initial Structure: I resolved an initial structure of the 4.7Å RodA-PBP2 complex using New York collaborative data, and my own predicted structure, prior to the existence of AlphaFold.
4. RodA-PBP2 activity testing: I managed to test RodA-PBP2 for glycosyltransferase activity using a gel electrophoresis based Schägger gel system
5. RodA-PBP2 mutant library: Using the initial structure, I designed and successfully created a series of mutants to be tested for glycosyltransferase activity.
6. RodA-PBP2-lipid II active site in silica:
7. TLC of RodA-PBP2 nanodiscs:

8. Solving the polymerisation of lipid II by RodA: I wrote a paper with co-author Rie Nygaard on a proposed mechanism using the above data, attaching my own self styled figures for a deeper analysis.
9. Revealing GT-C protein connection: Part of *Nature* paper on Waal Ligase UPP binding, I show that RodA is part of a family of similar proteins.

Chapter 2: Background

The Elongasome core: RodA and PBP2

One of the core peptidoglycan biosynthesis and modifying complexes is the elongasome; a biosynthetic complex of peptidoglycan manufacturing protein "machinery" spanning the inner cytoplasmic membrane into the periplasm with a distinctive role in cell growth during elongation of cells by peptidoglycan/cell wall polymerisation (RodA) and transpeptidation/crosslinking (PBP2). (5,6,29,40,47,113).

The core elongasome model centers on the membrane-bound central monofunctional class B transpeptidase PBP2 which is known to activate another membrane-embedded protein RodA, a processive lipid II transglycosylase. This creates a machine capable of both transpeptidation and glycosyltransferase activity, with additional debated accessory proteins mentioned previously (Chapter 1.1). PBP2 interacts at an anchor domain and at its membrane helix with RodA, although this wasn't known before the *T. thermophilus (6) structure*. This core machine is also transiently in interaction with MreC and therefore also indirectly the cytoskeletal protein MreB (Not in Figure 2.0 for simplicity) which continuously makes circumnavigations of the cell, this therefore orients the peptidoglycan synthesis by RodA-PBP2 to a cytoskeletally determined path. A similar protein partner set FtsW and PBP3 works in the same way, but with alternative function as a division-based apparatus and FtsZ determined localisation. A PBP2 knockout is lethal, and RodA knockout causes cells to round/ become short (4). Therefore, this elongation apparatus is essential.

As well as a PBP2 transpeptidase and RodA transglycocylase , in addition to some of the proteins and ligases shown by the model proposed in Figure 2.0, the elongasome also likely contains a host of other enzymes that interact with RodA and PBP2, already suggested in the Chapter 1.1 review on peptidoglycan synthesis by the SEDS complexes. In short, in addition to a lipid II polymerase (RodA) and transpeptidase (PBP2) which increase the chain length of sugars and crosslink these chains by peptides; carboxypeptidases alter peptide lengths, lytic transglycosylases cut these chains apart and structural proteins such as LpoA and RlpA may control the process, but also most likely there are other proteins involved in outer membrane maintenance.

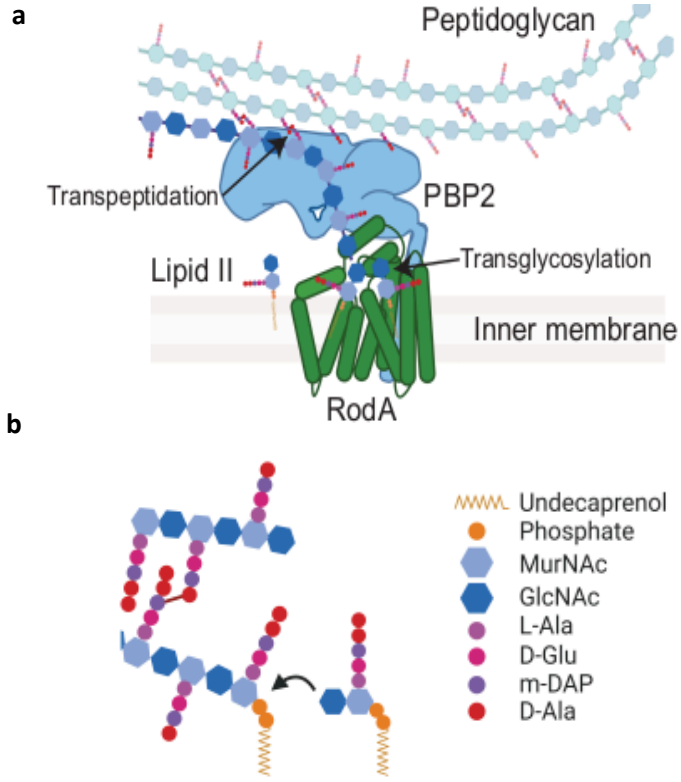


Figure 2.0. Peptidoglycan elongation by the RodA-PBP2 elongasome.

a RodA and PBP2 in complex, transpeptidation of de novo PG into existing PG by PBP2 and transglycosylation of Lipid II by RodA. **b** Transglycosylation mechanism by RodA (arrow) , through attachment of MurNac to GlucNac of adjacent polymer.

The catalytic mechanism of lipid II polymerisation by a SEDS glycosyltransferase was not known before the start of this PhD. Nor was any aspect of the mechanism other than it does not require

ATP to occur. As the majority of antibiotics are cell wall synthesis inhibiting in their action, the interrogation of this synthesis by the RodA-PBP2 elongasome will be vital to the production of new antibiotics, and will move forward to develop an understanding of not only RodA-PBP2, but its high homology SEDs complex paralogues such as SpoVDE and FtsW-PBP3 which are also vital for bacterial survival in other species or other time points in the cell cycle(49).

In this chapter for the first time, through a combination of co-evolution analysis, mapping of cryo-EM densities and molecular dynamic software, the cryo-electron atomic structure of the complex was resolved first to 4.7Å, which allowed us to make strategic functional mutations, interrogate the movement of the proteins and for the first time visualise *E.coli* RodA-PBP2 as a complex, then this structure was later refined to 3.6Å, which has allowed us to write a paper. In addition, two binding sites for lipid II were found *in silico* by myself *a*nd an active site facilitating processive glycosyltransferase activity proposed using my assays. The mechanism of lipid II glycosyltransferase reaction by RodA-PBP2 and by extension its homologues occurs at RodA along a conserved cleft and across the membrane spanning region of the protein. This is made clear by my *in vitro* results which guided and facilitated a consortium.

This chapter was made possible by a collaboration with Professor Filippo Mancia and his research assistant Dr Rie Nygaard of Columbia University in New York regarding the Cryo-EM structure of the *E.coli* RodA-PBP2 complex mentioned.

Chapter 2: Methods

RodA-PBP2 Expression construct.

The RodA-PBP2 construct used as the basis for these experiments was obtained from Professor Filippo Mancia at Columbia University and consists of the *E.coli* W3310 genes for RodA, linked by an SGSGSG flexible linker to PBP2 cloned in a modified pET23 based vector called pNYCOMPS23. The construct produces a fusion protein of RodA and PBP2 with a N-terminal FLAG and 10x His tag and TEV protease cleavage site (Figure 2.1)

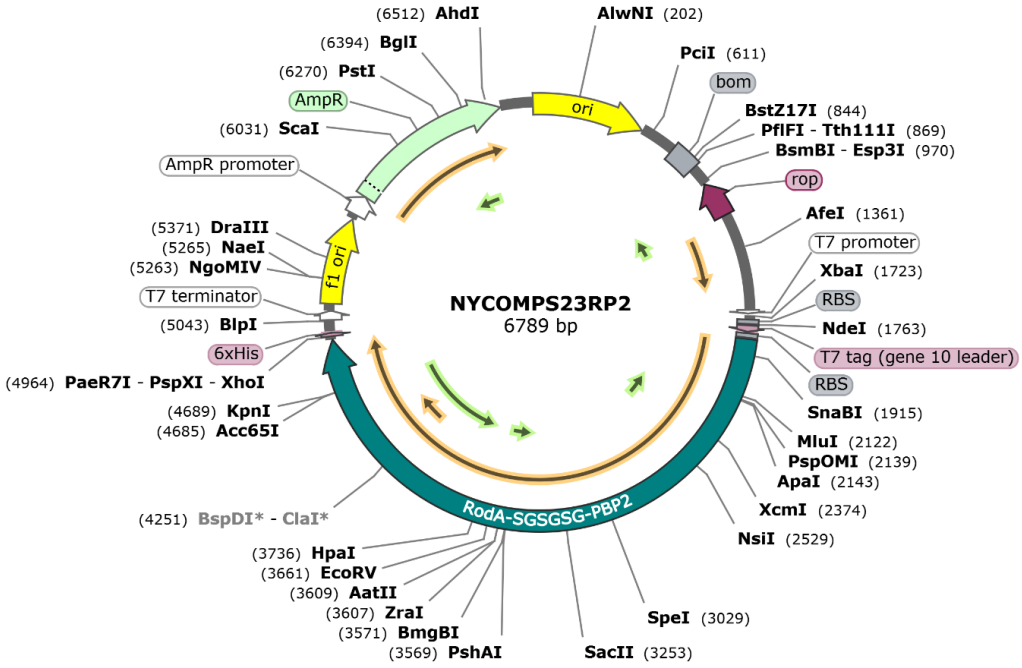


Figure 2.1 Plasmid Map of NYCMPS23 RodA-PBP2

Created using snapgene, based on sequenced data. RodA and PBP2 are connected by an SGSGSG peptide linker between RodA and PBP2 and the fusion protein is purified using IMAC utilizing the 10xHIS tag on RodA. The C terminal 6xHis is not expressed with the protein, due to a stop codon after the PBP2.

RodA-PBP2 expression and purification

RodA-PBP2 in pNYCOMPS-N23 was transformed into BL21 pLysS *E. coli* strain, a 10 mL starter culture (LB 35µg/ml Chloramphenicol, 100µg/ml Ampicillin) made using a freshly transformed colony was grown overnight at 37˚C. Mature starter cultures were used to inoculate 800ml 2xYT AutoInduction media cultures which were grown at 37˚C to reach an OD600 between 0.8 and 1.0, before the temperature was reduced to 22˚C for overnight incubation at 215-220 RPM. Cells were pelleted at 4000g for 10 minutes at 4˚C. Cell pellets were resuspended using Cell Lysis buffer [20 mM HEPES, pH 7, 200 mM NaCl, 10mg/mL solid DNase, 4mL/100mL RNase, 1:1000 Complete cocktail of protease inhibitors, 1mM TCEP, 1mM PMSF, 20% Glycerol] and lysed by cell disruption at multiple passes to ensure 100% cell lysis. Lysate was centrifuged using an ultracentrifuge 80,000G for 35 minutes at 4˚C and the membrane pellets kept. Membrane pellets were washed by resuspension in high salt buffer [20 mM HEPES, pH 7, 500 mM NaCl,10mg/mL solid DNase, 4mL/100mL RNase, 1 mM TCEP, 1:1000 Complete cocktail of protease inhibitors, 1mM PMSF, 20% Glycerol] and span again at 80,000G for 35 minutes and the pellet kept. [Pause step possible by freezing of membrane]. Salt washed pellets were resuspended using high salt buffer and homogenisation, then allowed to solubilise in 1% DDM. Solubilised membranes were then spun at 80,000G for 35 minutes and supernatant collected. Supernatant was then added to equilibrated Ni-NTA resin for binding for 2hrs, after being adjusted to 40mM Imidazole final concentration. After binding, the solute flow through was collected by gravity column leaving protein enriched resin. The protein enriched resin was washed using 10CV of Wash buffer [20 mM HEPES, pH 7, 500 mM NaCl, 60 mM Imidazole, pH 7, 20% glycerol, 0.1% DDM] and then eluted into 0.5CV using elution buffer [20 mM HEPES, pH 7, 500 mM NaCl, 300 mM Imidazole, pH 7, 20% glycerol, 0.1% DDM]. The resultant elution was then concentrated using a 100kDa 1ml total volume concentrator for 10minutes and resolved by SDS-page electrophoresis.

Lipid II Polymerisation by detergent soluble RodA-PBP2 and RodA

Glycosyltransferase activity by RodA-PBP2 was shown using visualisation of fluorescently labelled Dansyl Lipid II (amidated Lysine) molecules through a time point orientated gel-based method. 1.5µl of 2.7µM Detergent solubilized RodA-PBP2 or RodA protein [300mM Imidazole, 250mM NaCl, 20mM HEPES, 0.05% DDM, 20% Glycerol] final concentration 0.27µM, was added to 13.5µl [10mM MgCl2, 100mM NaCl, 50mM HEPES, 20% DMSO, 0.03% LDAO,10µM Dansyl amidated Lysine Lipid II]. The reaction mixture was incubated at 37°C for 1hr 30 minutes to allow for polymerisation.

The resultant polymer was denatured at 95°C with addition of 5x loading dye [50mM TrisHCl pH 8.8, 4% SDS, 40% Glycerol, an amount of bromophenol blue, DTT 200mM] and ran on a Criterion 16.5% gel at 110V for 80 minutes with anode gel running buffer [0.1M Tris-HCl pH 8.8] and cathode gel running buffer [0.1M Tris-HCl pH 8.25, 0.1M Tricine, 0.1% SDS] and visualised by 10 second exposure to UV on gel viewing apparatus.

Bocillin staining Protocol

Bocillin stains were performed by incubating 1µl DMSO dissolved Bocillin 1mM with 15ul 1µM of folded PBP domain containing proteins for 15 minutes. The resulting solutions were then unfolded using 5x loading dye [50mM TrisHCl pH 8.8, 4% SDS, 40% Glycerol, an amount of bromophenol blue, DTT 200mM] at 65°C for 10 minutes and on an SDS page gel, washed and then visualised by 4 seconds exposure to fluorescein staining on a Biorad gel visualiser. The images were visualised at maxima to include the bocillin marker of the ladder and no other marker, therefore only visualising Bocillin stained proteins.

Cryo-EM Structure Resolution and RodA-PBP2 structure building

Initial *E.coli* RodA-PBP2 structural models were created using the soluble *E. coli* PBP2 6G9F crystal structure stitched in PyMol to a SwissModel prediction of the *E. coli* RodA- and transmembrane helix PBP2 template of a manually SGSGSG linked 6PL5. This initial homology model was used as an input to the Cryo-EM structural refinement tool Flex-EM of CCPEM (114), in conjunction with guided manual protein refinement into our Cryo-EM maps over 30 successive equilibration cycles, with an assumption of 4.7Å resolution.

Maleimide labeling of proteins and other thiolated biomolecules

RodA-PBP2 in storage buffer was moved to a degassed buffer replacement by dialysis. Then a 100 molar TCEP excess of the RodA-PBP2 was added, and the mixture allowed to react for 20 minutes. Maleimide (MTSL) or FRET dye mixture, dissolved in DMSO at 10µM (20X molar excess) was added to the reaction and mixed overnight at 4°C, before passaging through a 100kD concentration into storage buffer to remove excess dye and maintain RodA-PBP2. The resultant mixture was used for DEER experiments, and polymerisation stability experiments.

Mutagenesis

RodA-PBP2 Mutagenesis was performed using the QuickChange II method (115). Complementary 75°C primers with desired sequences were primed. Phusion HF DNA polymerase reaction mixtures of 25µl [H20 to 25µl, 5xHF Buffer 5ul, 10mM dNTPs 0.4µl, 10µM Forward primer 1.25µl, 10µM Reverse primer 1.25µl. Template 1µl] 18 cycles [95°C denature 30s, 1 minute 55°C anneal, 3minutes 72°C elongation] were performed, and the above digested in DPNI, before transformation, miniprep and confirmation by sequencing.

Conservation analysis

Weblogo(116) was used, along with an *E.coli* K12 W3110 sequence to create a weblogo of the aligned sequences of RodA. Species specificity and range of these alignments was assigned by BLAST

GREMLIN Co-Evolution Analysis

Escherichia coli K12 W3110 sequences of RodA (mrdb) and PBP2 (mrdA) and their homologue proteins across gram negative species with E-10 cutoffs and C terminal amino acid trimming for increased sequence similarity 75% necessary were used as an input to measure species dependent co-evolution of amino acids on GREMLIN the co-evolution calculation server (117). Displayed interactions are those in the top 10 most interactive amino acid pairs.

Ligand Docking

Analysis of lipid II binding was conducted using Audodock Vina ligand docking software (118). Our structures were used as receptors, and lipid II was created through Chemdraw and then realised in PyMol in 3D. These 3D structures were flexibly docked into the conserved regions of RodA-PBP2 at the RodA site. The highest energy binding sites with correct theoretical ligand orientation in light of membrane orientation restrictions were chosen for later analysis.

RodA-PBP2 Transpeptidation assay

RodA-PBP2 transpeptidation assays were performed as previously described with a substitution of RodA-PBP2 as the Lipid II polymerisation and transpeptidation catalyst. The resulting reaction mixture was concentrated by freeze drying, and resuspended in 50 ml of water to which mutanolysin 1ml (0.5mg/ml|) was added for incubation overnight. A polymerization reaction using 0.5 mM Dansyl mDAP Lipid II as the substrate was also treated with lyzosyme for comparison with the mutanolysin digested sample product on a 16% Tris/Tricine gel

electrophoresis in order to enable to isolate the mutanolysin digested crosslink products at an equivalent position. The resulting gel cross-section lipids were then extracted by gel dehydration by ethanol, followed by a wash step and acetonitrile 50% addition (v/v) prior to MS analysis.

Preparation of RodA-PBP2 associated lipids for mass spectrometry and TLC analysis

RodA-PBP2 reconstituted in nanodisc were analyzed for lipid content, the lipid was purified using the Folch method (119) as follows. Firstly, 200 µL ice-cold MeOH was added to 200 µL sample followed by 400 µL chloroform and lastly 150µL H2O. Each sample was then vortexed and sonicated in an ice bath for 15 min to facilitate rigorous mixing and extraction. The samples were centrifuged at 6,000× g for 15 min at 4 °C to achieve phase separation. The lipid containing organic layer was collected. The resulting lipid extract was dried under nitrogen gas and stored at −20°C until analysis.

Dried lipid samples were re-suspended in acetonitrile 50% v/v and analyzed by negative mode mass spectrometry on an OrbiTRAP mass spectrometer, with m/s collected between 200-2000.

Between 32-35V were used to break samples components. Analysis of lipid content was conducted using LipidMAPs mass prediction and CM-ID (120,121) .Plant derived undecaprenyl di-phosphate control samples were gifted by Swiezewska laboratory Warsaw, Poland

Chapter 2: Strains, Plasmids and Oligonucleotide Primers

Mutant name/primer name	Primer 1 - 5' to 3'	Primer 2 - 5' to 3'
L61R	CGC TCT AAT GAA AAC CGC ATT AAG CGG GTG CCT ATC GCG CCC AGC CGC GGC	GCC GCG GCT GGG CGC GAT AGG CAC CCG CTT AAT GCG GTT TTC ATT AGA GCG
L278R	GCT GGG ATT AGT GGG CAT TCG GAT TCT GCT CGC TCT CTA CAT TCT GC	GCA GAA TGT AGA GAG CGA GCA GAA TCC GAA TGC CCA CTA ATC CCA GC
L277A	GCG GAA GAG CTG GGA TTA GTG GGC GCT CTG ATT CTG CTC GCT CTC TAC ATT C	GAA TGT AGA GAG CGA GCA GAA TCA GAG CGC CCA CTA ATC CCA GCT CTT CCG C

l274R	CTG GCG GAA GAG CTG GGA GCA GTG GGC ATT CTG ATT CTG CTC GCT C	GAG CGA GCA GAA TCA GAA TGC CCA CTC GTC CCA GCT CTT CCG CCA G
I322A	CGT TTA TGT CTT CGT AAA TGC TGG TAT GGT AAG CGG TAT TCT GCC GG	CCG GCA GAA TAC CGC TTA CCA TAC CAG CAT TTA CGA AGA CAT AAA CG
L22A	GGG GAT TTT GGC GCT GAC CGG CGT GCT TAT CGC CAA CCT G	CAG GTT GGC GAT AAG CAC GCC GGT CAG CGC CAA AAT CCC C
L30A	CAA TTT GCA GAT TAT ACG CGT TGG CGA TAA GCA CGC CGG TCA GCA	TGC TGA CCG GCG TGC TTA TCG CCA ACG CGT ATA ATC TGC AAA TTG
R210A	CTG ATG CAT GAT TAC CAG CGC CAG GCC GTA ATG ATG CTC CTG GAC CCG G	CCG GGT CCA GGA GCA TCA TTA CGG CCT GGC GCT GGT AAT CAT GCA TCA G
E255A	CTC AGT CAC AGC TTG AAT TTC TCC CCG CAC GCC ATA CTG ACT TTA TCT TCG CGG TAC TG	CAG TAC CGC GAA GAT AAA GTC AGT ATG GCG TGC GGG GAG AAA TTC AAG CTG TGA CTG AG
E270A	GCC ATA CTG ACT TTA TCT TCG CGG TAC TGG CGG CAG AGC TGG GAT TAG TGG GCA TTC	GAA TGC CCA CTA ATC CCA GCT CTG CCG CCA GTA CCG CGA AGA TAA AGT CAG TAT GGC
K117R	CGT TTT CAG CCG TCG GAA ATT GCC CGA ATA GCC GTA CCA CTG ATG GTT GCG	CGC AAC CAT CAG TGG TAC GGC TAT TCG GGC AAT TTC CGA CGG CTG AAA ACG
K117I	TAT TGT TCG TTT TCA GCC GTC GGA AAT TGC CAT AAT AGC CGT ACC ACT GAT GGT TGC G	CGC AAC CAT CAG TGG TAC GGC TAT TAT GGC AAT TTC CGA CGG CTG AAA ACG AAC AAT AC
E114A	CTG GAC CTC GGT ATT GTT CGT TTT CAG CCG TCG GCA ATT GCC AAA ATA GCC GTA CCA CTG	CAG TGG TAC GGC TAT TTT GGC AAT TGC CGA CGG CTG AAA ACG AAC AAT ACC GAG GTC CAG
D159V	GCC CAC GCT GCT GGT GGC TGC ACA GCC TGU CCT GGG AAC ATC AAT CCT CGT TGC G	CGC AAC GAG GAT TGA TGT TCC CAG GAC AGG CTG TGC AGC CAC CAG CAG CGT GGG C
W102A	CAT CTC TAA AGG TGC TCA ACG CGC GCT GGA CCT CGG TAT TGT TCG	CGA ACA ATA CCG AGG TCC AGC GCG CGT TGA GCA CCT TTA GAG ATG
K97A	CGG TAG ATG CTT TCG GTG CCA TCT CTG CAG GTG CTC AAC GCT GGC TGG ACC TCG	CGA GGT CCA GCC AGC GTT GAG CAC CTG CAG AGA TGG CAC CGA AAG CAT CTA CCG
R101A	GTG CCA TCT CTA AAG GTG CTC AAG CCT GGC TGG ACC TCG GTA TTG TTC G	CGA ACA ATA CCG AGG TCC AGC CAG GCT TGA GCA CCT TTA GAG ATG GCA C

K117N	CAG CCG TCG GAA ATT GCC AAC ATA GCC GTA CCA CTG	CAG TGG TAC GGC TAT GTT GGC AAT TTC CGA CGG CTG
R109A	CTG GCT GGA CCT CGG TAT TGT TGC TTT TCA GCC GTC GGA AAT TG	CAA TTT CCG ACG GCT GAA AAG CAA CAA TAC CGA GGT CCA GCC AG
K49A	GCG GTC AGG ATA TTG GCA TGA TGG AGC GTG CAA TCG GCC AAA TCG CGA TG	CAT CGC GAT TTG GCC GAT TGC ACG CTC CAT CAT GCC AAT ATC CTG ACC GC
R101F	CAA TAC CGA GGT CCA GCC AGA ATT GAG CAC CTT TAG AGA TGG CAC	GTG CCA TCT CTA AAG GTG CTC AAT TCT GGC TGG ACC TCG GTA TTG
T261S	GAA TTT CTC CCC GAA CGC CAT GCT GAC TTT ATC TTC GCG GTA CTG	CAG TAC CGC GAA GAT AAA GTC AGC ATG GCG TTC GGG GAG AAA TTC
W102F	GTG CCA TCT CTA AAG GTG CTC AAC GCT TTC TGG ACC TCG GTA TTG TTC G	CGA ACA ATA CCG AGG TCC AGA AAG CGT TGA GCA CCT TTA GAG ATG GCA C
H260A	TCA CAG CTT GAA TTT CTC CCC GAA CGC GCT ACT GAC TTT ATC TTC GCG GTA CTG GC	GCC AGT ACC GCG AAG ATA AAG TCA GTA GCG CGT TCG GGG AGA AAT TCA AGC TGT GA
C82G	GCT GGG CCC CCT ATC TCT ATA TCA TCG GTA TTA TTT TGC TGG TGG CGG TAG ATG	CAT CTA CCG CCA CCA GCA AAA TAA TAC CGA TGA TAT AGA GAT AGG GGG CCC AGC
C133A	CGC GCT TTA TCA ACC GCG ACG TTG CCC CGC CAT CGT TGA AGA ACA CTG	CAG TGT TCT TCA ACG ATG GCG GGG CAA CGT CGC GGT TGA TAA AGC GCG
P257G	A CRO was used to obtain this mutant (was made by a Azenta biosciences-Genewiz)) P257G: Replace CCC to GGC at this amino acid, the following region as modified is identified in the adjacent box	GTC ACA GCT TGA ATT TCT CGG CGA ACG CCA TAC TGA C
D262A	CCC GAA CGC CAT ACT GCC TTT ATC TTC GCG GTA	TAC CGC GAA GAT AAA GGC AGT ATG GCG TTC GGG
Q111A	A CRO was used to obtain this mutant (was made by a Azenta biosciences-Genewiz)) Q111A: Replace CAG to GCG at this amino acid, the following region as modified is identified in the adjacent box	TGG CTG GAC CTC GGT ATT GTT CGT TTT GCG CCG TCG GAA ATT G
S344A	A CRO was used to obtain this mutant	GGT CAG TTA TGG AGG AGC GGC GCT

	(was made by a Azenta biosciences-Genewiz)) S344A: Replace TCG to GCG at this amino acid, the following region as modified is identified in the adjacent box	AAT TGT GCT GAT GGC
R48A	A CRO was used to obtain this mutant (was made by a Azenta biosciences-Genewiz)) R48A: Replace CGT to GCT a at this amino acid, the following region as modified is identified in the adjacent box	GGC ATG ATG GAG GCT AAA ATC GGC CAA ATC GCG ATG GG
F401STOP	A CRO was used to obtain this mutant (was made by a Azenta biosciences-Genewiz))	
G452STOP	A CRO was used to obtain this mutant (was made by a Azenta biosciences-Genewiz))	
RodA-PBP2 sequencing Forward	ATC ACC ATC ACC ATC ACC ACC ATC ACG	
RodA-PBP2 middle sequence Forward	ATC GTT GAA GAA CAC TGG CAT CGC G	
RodA-PBP2 reverse	TTA ATG GTC CTC CGC TGC GGC	
Plasmids	Sequence	
pNYCOMPS 23-RodA-PBP2- AmpR	TTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTT TTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGT TAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAA GGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCG GACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAA CGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCT CGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTT GCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCC TTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGA AGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATAT GGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTA CGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGT CTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTT CACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATT CACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCT TCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGG GGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACT GATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGG	

	GACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAG GGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTT TCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGC AGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCG CCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCT GCCCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCT AGGATCATTTGTTTAACTTTAAGAAGGAGATATACCATGGATTATAAAGATGATGATGATAAACAT CATCACCATCACCATCACCACCATCACGAAAACCTGTATTTTCAATCCTACGTAATGACGGATAATC CGAATAAAAAAACATTCTGGGATAAAGTCCATCTCGATCCCACAATGCTGCTGATCTTACTGGCATT GCTGGTTTACAGCGCCCTGGTTATCTGGAGCGCCAGCGGTCAGGATATTGGCATGATGGAGCGTAA AATCGGCCAAATCGCGATGGGTCTGGTCATCATGGTGGTGATGGCGCAAATTCCTCCACGCGTTTAT GAAGGCTGGGCCCCCTATCTCTATATCATCTGTATTATTTTGCTGGTGGCGGTAGATGCTTTCGGTG CCATCTCTAAAGGTGCTCAACGCTGGCTGGACCTCGGTATTGTTCGTTTTCAGCCGTCGGAAATTGC CAAAATAGCCGTACCACTGATGGTTGCGCGCTTTATCAACCGCGACGTTTGCCCGCCATCGTTGAAG AACACTGGCATCGCGCTGGTGCTGATATTTATGCCCACGCTGCTGGTGGCTGCACAGCCTGACCTGG GAACATCAATCCTCGTTGCGCTTTCCGGTCTGTTTGTACTGTTCCTCTCTGGCCTTAGCTGGCGTCTG ATTGGCGTCGCAGTAGTGCTGGTAGCGGCGTTCATTCCGATTCTGTGGTTCTTCCTGATGCATGATT ACCAGCGCCAGCGCGTAATGATGCTCCTGGACCCGGAATCAGACCCACTCGGCGCGGGCTATCACA TTATTCAGTCTAAAATTGCTATTGGCTCCGGCGGATTACGCGGCAAAGGCTGGCTGCACGGCACTCA GTCACAGCTTGAATTTCTCCCCGAACGCCATACTGACTTTATCTTCGCGGTACTGGCGGAAGAGCTG GGATTAGTGGGCATTCTGATTCTGCTCGCTCTCTACATTCTGCTGATCATGCGCGGGCTGTGGATAG CCGCCAGAGCGCAAACCACCTTTGGTCGCGTCATGGCTGGCGGCTTAATGCTGATATTATTCGTTTA TGTCTTCGTAAATATTGGTATGGTAAGCGGTATTCTGCCGGTTGTAGGGGTTCCGCTCCCACTGGTC AGTTATGGAGGATCGGCGCTAATTGTGCTGATGGCTGGGTTCGGGATTGTAATGTCAATCCACACCC ACAGGAAAATGTTGTCGAAAAGCGTGACTAGTGGCTCTGGCTCTGGCTCGAAACTACAGAACTCTT TTCGCGACTATACGGCAGAGTCCGCGCTGTTTGTGCGCCGGGCGCTGGTCGCCTTTTTGGGGATTTT GCTGCTGACCGGCGTGCTTATCGCCAACCTGTATAATCTGCAAATTGTTCGCTTTACCGACTACCAG ACCCGCTCTAATGAAAACCGCATTAAGCTGGTGCCTATCGCGCCCAGCCGCGGCATTATCTACGAT CGTAACGGTATCCCTCTGGCCCTCAACCGCACTATCTACCAGATAGAAATGATGCCGGAGAAAGTC GATAACGTGCAGCAAACGCTGGACGCTTTGCGCAGCGTGGTAGATCTGACCGATGACGATATTGCT GCATTCCGAAAAGAGCGCGCACGTTCACACCGTTTCACCTCTATTCCGGTGAAAACTAACCTGACC GAAGTACAAGTAGCTCGCTTTGCCGTCAATCAGTACCGTTTTCCGGGTGTCGAAGTTAAAGGCTATA AACGTCGTTACTATCCTTACGGTTCGGCGTTGACCCACGTCATCGGCTATGTGTCGAAAATCAACGA TAAAGACGTCGAACGCCTGAATAATGACGGCAAACTGGCCAACTATGCGGCAACGCATGATATCGG TAAGCTGGGCATTGAGCGTTACTATGAAGATGTGCTGCACGGTCAGACCGGTTATGAAGAGGTTGA AGTTAACAACCGTGGGCGTGTTATTCGCCAGTTAAAAGAAGTACCACCGCAAGCCGGACACGATAT TTACCTGACGCTGGATCTCAAACTCCAGCAATATATTGAAACGCTGCTGGCGGGTAGCCGCGCAGC TGTGGTAGTCACCGATCCGCGTACAGGTGGGGTGCTGGCGCTGGTTTCCACGCCTAGTTATGACCCA AACTTGTTTGTTGACGGTATCTCCAGCAAAGATTATTCCGCCTTGTTGAACGATCCGAATACACCGC TGGTGAACCGCGCCACACAGGGGGTTTATCCTCCCGCGTCTACAGTTAAACCCTATGTGGCGGTTTC GGCATTGAGCGCCGGGGTGATCACGCGCAATACGACGCTGTTTGACCCAGGCTGGTGGCAACTGCC AGGTTCGGAAAAACGTTATCGTGACTGGAAAAAATGGGGCCACGGGCGTCTGAATGTCACAAGATC GCTGGAAGAATCTGCGGATACCTTCTTCTATCAGGTGGCCTACGATATGGGGATCGATCGCCTCTCC GAATGGATGGGTAAATTCGGTTATGGTCATTACACCGGTATCGACCTGGCGGAAGAACGTTCCGGC AACATGCCTACCCGCGAATGGAAACAGAAACGCTTTAAAAAACCGTGGTATCAGGGTGACACCATT CCGGTTGGTATCGGTCAGGGTTACTGGACAGCGACCCCAATCCAGATGAGTAAGGCACTGATGATC CTGATTAATGACGGTATCGTGAAGGTTCCTCATTTGCTGATGAGCACCGCCGAAGACGGCAAACAG GTGCCATGGGTACAGCCGCATGAACCGCCCGTCGGCGATATTCATTCCGGTTACTGGGAGCTGGCG AAAGACGGTATGTACGGTGTTGCTAACCGCCCTAACGGTACGGCGCATAAATACTTTGCTAGCGCA CCGTACAAAATTGCGGCGAAATCCGGTACCGCTCAGGTCTTCGGTCTGAAAGCGAACGAAACCTAT AATGCGCACAAAATTGCCGAGCGTTTACGTGACCACAAACTGATGACCGCCTTTGCGCCATACAAC AATCCGCAAGTGGCTGTCGCCATGATTCTGGAGAACGGTGGTGCGGGTCCGGCGGTTGGTACACTG ATGCGCCAGATCCTCGACCACATTATGCTGGGTGATAACAACACCGATCTGCCTGCGGAAAATCCA GCGGTTGCCGCAGCGGAGGACCATTAATAATATTGAGGGCTCGAGCACCACCACCACCACCACTGA GATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTA GCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCC GGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCA GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCC ACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTT ACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATA

	GACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGA ACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATT TCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAA ATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTA TGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTC ACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATC GAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATG AGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCG GTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTAC GGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAA CTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCA TGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACAC CACGATGCCTGCAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGC TTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGC CCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATT GCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCA ACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTG TCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA GGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCG TCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGC
Strains	Notes
BL21 DE3 pLysS competent cells	Allows tunable expression of plasmid, used for RodA-PBP2 protein purifications, Chloramphenicol resistant(Agilent) F–, *omp*T, *hsd*S_B (r_B–, m_B–), *dcm*, *gal*, λ(DE3), pLysS, Cm^r.

Chapter 2: Results and Discussion

2.1 Purification of RodA-PBP2 in detergent

The functional complex was designed prior to the start of the PhD and successfully purified in New York by our collaborators Dr Rie Nygaard and Professor Filippo Mancia. The fusion protein was tested for its lipid II glycosyltransferase activity in the membrane scaffold protein (MSP) encapsulated form used for Cryo-EM studies. Briefly the protein was purified from membranes after solubiliation in DDM as described above before addition of the MSP protein and detergent removal. The resulting stabilized form of the protein was subjected to size exclusion chromatography prior to further structural studies.

RodA-PBP2 had never been successfully purified in the UK, and the complex's activity was thought to be dependent on its stability within the lipid filled nano disc. After multiple attempts at purifying the complex unsuccessfully, it was eventually purified using pLysS BL21 (DE2) cells as the expression line, and upon changing our membrane purification protocol to increase lysis efficiency, after having previously used one step of cell lysis, and a clear lysate step at 25,000g, this step was skipped in favour of further lysis, in addition auto induction media was used. The SDS-PAGE electrophoresis gel of a RodA-PBP2 purification shortly after this initial success is below (Figure 2.1.0) Bocillin staining and use of a fluorescein ladder allowed for this to be visualised as a complex with a stable PBP2 protein. (Figure 2.1.0B). This purification, constituing 0.3mg/ml by BCA assay, allowed the first assays of the protein complex activity.

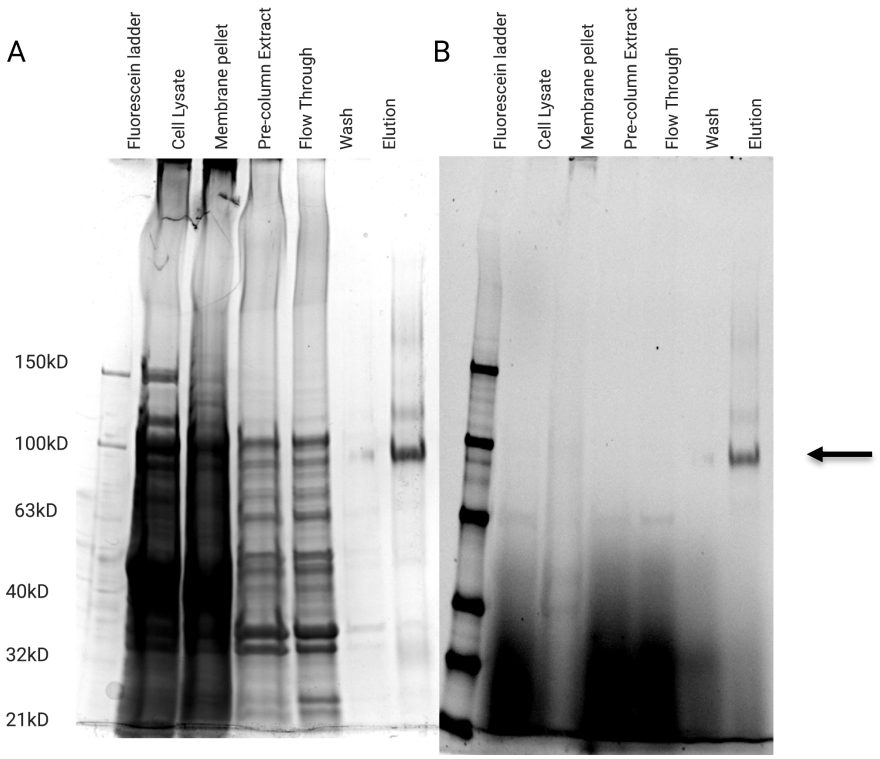
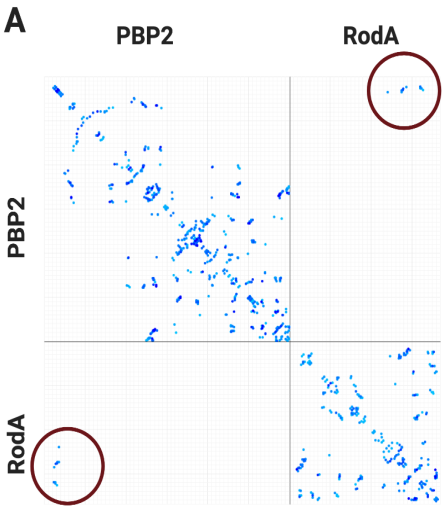


Figure 2.1.0 Coomassie stained gel of first Successful RodA-PBP2 purification.

A: Coomassie stain of 4-15% SDS page gel separated by 1D electrophoresis. Protein molecular weight indicated on ladder lane, arrow points to presence of RodA-PBP2 in sample. Image taken in BioRad imager. B Bocillin stained version, showing RodA-PBP2.

2.2 Genetic interaction of PBP2 and RodA

As well as a biochemical stream of work, the project also went along a bioinformatic path. The structure of RodA and PBP2 of *T. thermophilus* in complex was solved in March 2020, 6 months after the start of the PhD (6). Before this time, we hypothesised their interaction based on the literature. Co-evolution of residues across species indicate that in RodA and PBP2 there are a set of these interactions and a surface of contact between the proteins, which provides a fitness benefit necessary for bacterial survival. Here we used GREMLIN co-evolution discovery to annotate these co-evolved residues between the proteins. (Figure 2.2.0)



B

PBP2	RodA	Scaled Score
36_I	278_L	2.09
39_L	277_I	1.94
31_L	319_F	1.65
40_Y	274_L	1.5
35_L	322_I	1.33
44_I	240_L	1.32
32_T	285_I	1.28
44_I	274_L	1.28
35_L	319_F	1.15
39_L	326_S	1.01
35_L	328_I	1

Figure 2.2.0 Co-Evolution and interaction of RodA and PBP2.

A Contact map of co-evolved residues between RodA and PBP2 as determined by multisequence analysis engine GREMLIN **B** Top 11 co-evolved residues between RodA and PBP2 and their scaled interaction scores.

These residue contacts between 30-44aa PBP2 and 240- 328aa on RodA indicate the two proteins interact at the membrane helix of PBP2- predicted from the model structures, and therefore a construct of RodA and PBP2 we had created with a linker between these contacting helices theoretically would create a functional complex.

2.3 Initial Cryo-EM structure of *Escherichia coli* RodA-PBP2 complex

Immediately prior to the COVID19 lockdown period the structure of our RodA-PBP2 in *E. coli* was resolved to a nominal resolution of 4.7Å, shortly after publication of the *T. thermophilus* PBP2 and RodA crystal structure was published using their model to create our own similar model(6) (PDB: 6G9F) I was integral in the alignment of predicted structures within a cryo-EM density. This represented a change in the project, where my model of *E. coli* RodA-PBP2 allowed us as a collaborative consortium to interrogate the structure more insightfully. The reliability of this structure

was then tested through further molecular dynamics, for interrogation of stability by Dr Owen Vickery (Figure 2.3.0)Lack of density in the penicillin binding extremity of PBP2 in our own cryo-EM structure meant it likely there was a high degree of flexibility in this PBP2 extremity. (Red/Orange region Figure 2.3.0) in contrast to the density of RodA and PBP2 interacting region which is more stable.

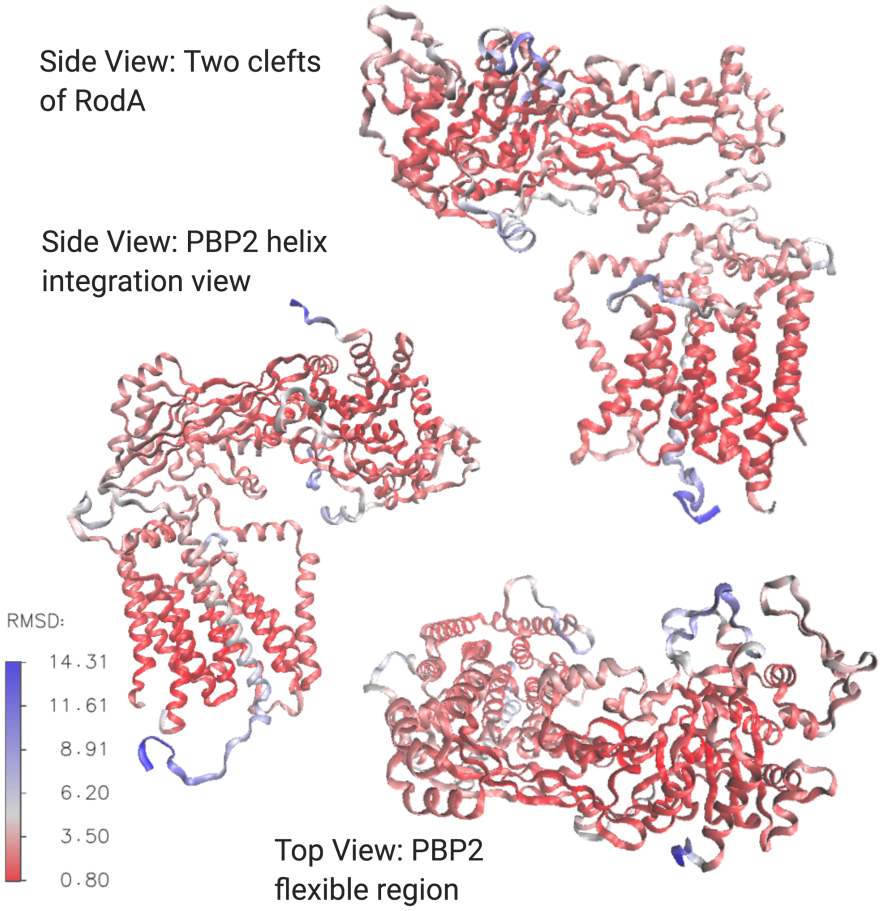


Figure 2.3.0 RMSD of E.coli RodA-PBP2 stability simulation mapped to structure.

Simulation performed by Dr Owen Vickery, and then analysed by myself using VMD in cartoon representation. Blue regions of high RMSD value, indicate instability, red regions indicate high stability, as shown by RMSD scale. Protein represented in Cartoon View.

Simulations of my predicted structure within a membrane by Dr Owen Vickery (Figure 2.3.0) and raw data analysed by myself, reveal the structurally unstable regions of the *E. coli* structure to be those already expected, PBP2 at its unresolved loops, along with the linker, and the N terminus of RodA (bright blue with 14Å RMSD), however overall the RMSD of the structure is low- between

3Å and 0.8Å RMSD after 1μs, indicating that the predicted structure is stable, and indicative of a real structure. The cryo-EM data was then re-analysed by AI and CryoSPARK by Dr Rie Nygaard at Columbia to produce a higher quality 3.2Å structure in Figure 2.3.1.

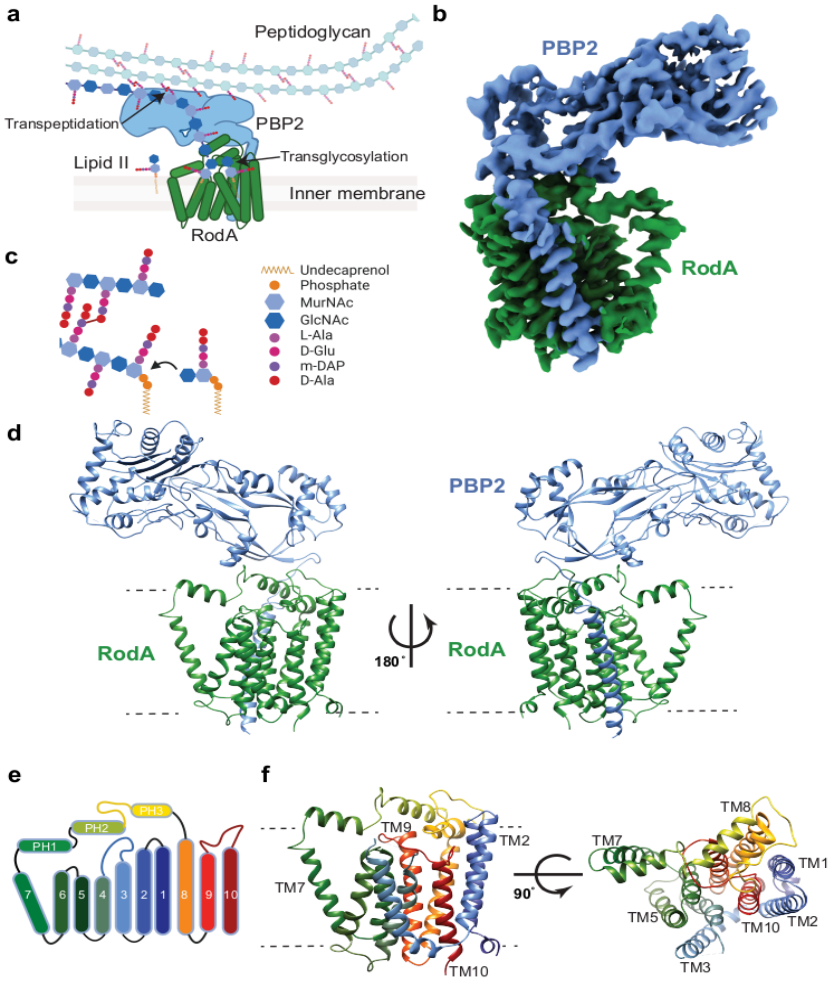


Figure 2.3.1. The Cryo-EM resolved structure of RodA-PBP2 complex

a. Synthesis of PG by RodA-PBP2 **b**. 3.2Å cryoelectron microscopy structure of RodA-PBP2, finalised in Columbia University. **c** Polymerisation of lipid II to existing peptidoglycan **d** Visualisation of PBP2 and RodA as predicted by Cryo-EM density data. **e** Cartoon flattened Helix number RodA **f** Corresponding RodA structure. Thanks to Rie Nygaard for the improved Cryo-EM model.

The most notable information from the 3.2A structure alone, was the high flexibility of the previously mentioned 101A loop between helices 3 and 4, which were difficult to resolve. Ours

was the first cryo-EM structure of publishable quality in *E.coli*. This structure, and the flexibility it indicated was later essential to future work.

2.4 Activity of RodA-PBP2 can polymerise Lipid II without MSP nanodiscs

After the short COVID related break, I focused on lab work again following the successful RodA-PBP2 purification, we then attempted purification with an in-house preparation of MSP nanodiscs, capable of lipid encapsulation in the presence of lipid, to replicate the purification protocol attempted by New York. However, this reduced efficiency of the purification 10 fold. As an alternative to this purification procedure, I tested detergent solubilised RodA-PBP2 construct immediately following IMAC purification and found that the complex was stable and was able to efficiently polymerise Lipid II. I compared to the glycosyltransferase activity of PBP1b as a positive control (Figure 2.4.0)

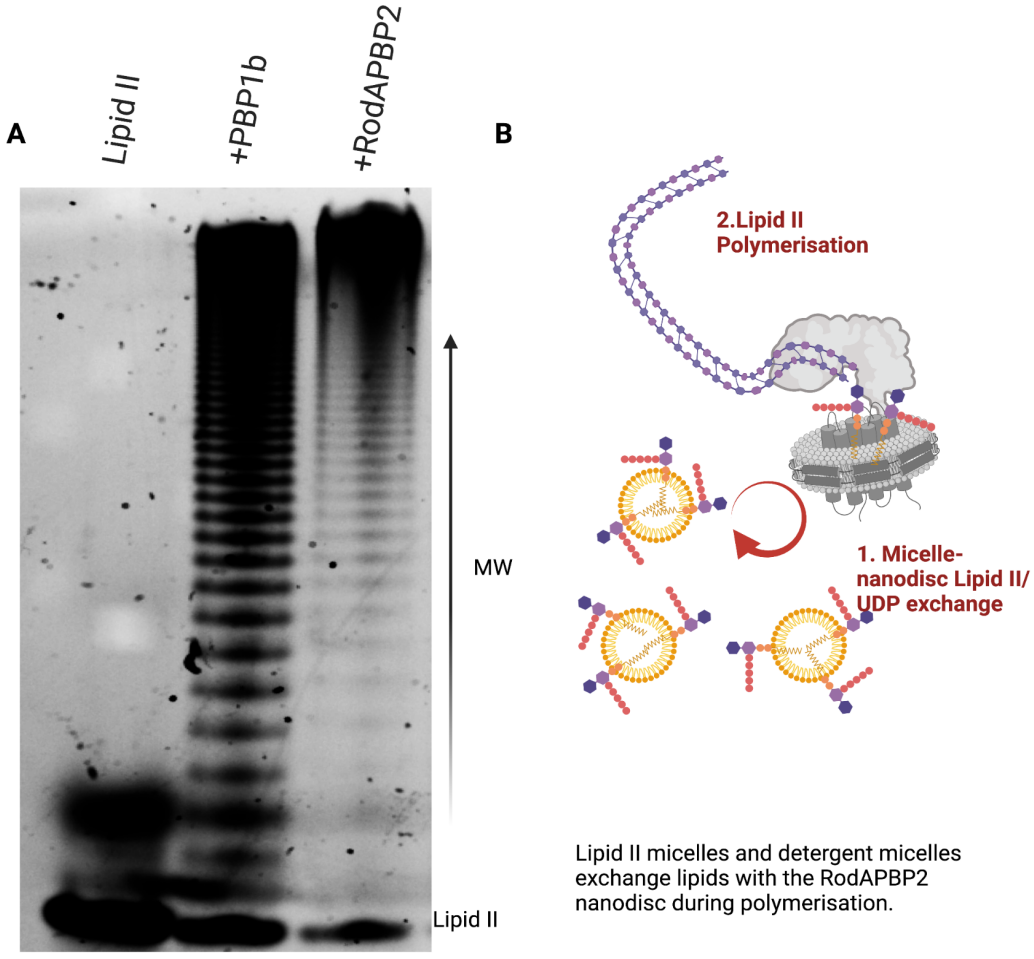


Figure 2.4.0 visualisation detergent solubilised RodA-PBP2 polymerisation of Lipid II by Schägger gel

A Polymerisation and increased product size of polymerised lipid II visualised on the vertical axis of each lane Lipid II corresponds to Dansyl amidated Lysine lipid II. Incubations were completed at 37 degrees celsius at 100µM lipid II, and resolved by gel electrophoresis as described in methods. 0.5µM RodA-PBP2 and PBP1b used in this case. PBP1b acts as a control for polymerisation. **B** Schematic of lipid II polymerisation in vitro

Verification that the fusion protein was active after IMAC purification was an important factor in our subsequent experiments as it afforded an easier route to purify and examine any site directed mutants required for further investigation. Therefore, I tested the concentration

dependence of the RodA-PBP2 I purified, to be certain this method was viable, and determine the reaction conditions which enabled this polymerisation.

2.4.5 Protein concentration dependence of RodA-PBP2 lipid II polymerisation.

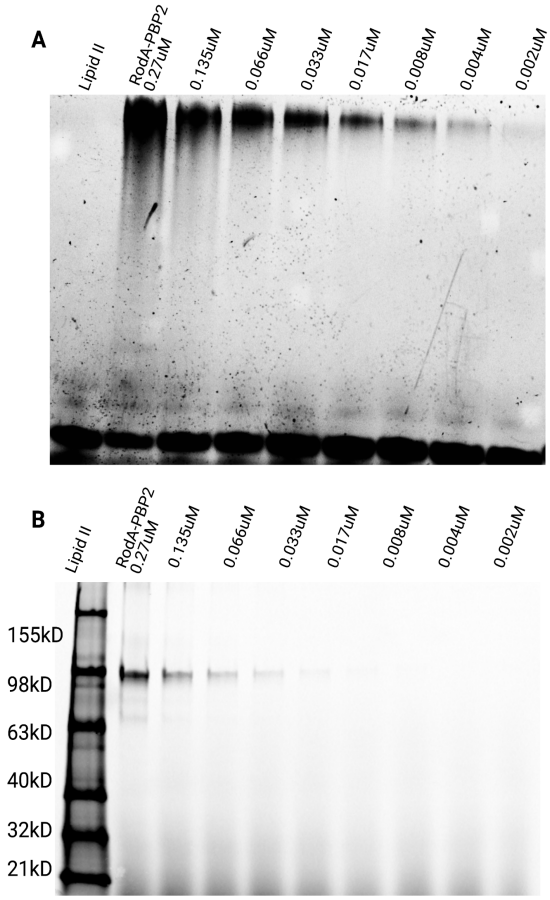


Figure 2.4.5 Polymerisation of Dansyl Lipid II by RodA-PBP2 at alternative concentrations

A Schägger gel visualised by UV excitation at fluorescein bandwidth, activity shown by presence of high molecular weight polymer. **B** A Bocillin gel of corresponding final RodA-PBP2 concentration, corresponding to initial enzyme added, 10x the final concentration.

Our RodA-PBP2 construct shows polymerisation ability at (Calculated from 0.3mg/ml addition at 110,000mw) 0.27μM to 0.033μM over a 2 hour period on 100μM lipid II, however below this concentration the polymerisation ability is less visible. This suggests pipetting and concentration

measurement errors of the magnitude of 10 fold differences should not cause false negatives of any mutants assays. (Figure 2.4.5)

2.5 Strategic RodA-PBP2 interaction and structural interrogation

Using the co-evolutionary and conservational knowledge between RodA and PBP2, in addition to my preliminary structure, and with no theory of lipid II polymerisation by RodA-PBP2 in the literature we started the creation of a mutant library, and purification of those mutants, so that we may test the RodA-PBP2 construct which we had shown to be active, and determine if amino acid changes would change this, thus revealing the most integral residues.

As the interactions of RodA to PBP2 were what formalised the production of the initial construct, these were re-interrogated based on the cryo-EM structure. In Figure 2.5.0 B, D, we see that the co-evolution and conservation can be mapped onto our structure to reveal essential interactions, these interactions can then be confirmed in Figure 2.5.0.A where it is revealed by mutation, the synthesis of lipid II is reduced upon their alteration *in vitro*.

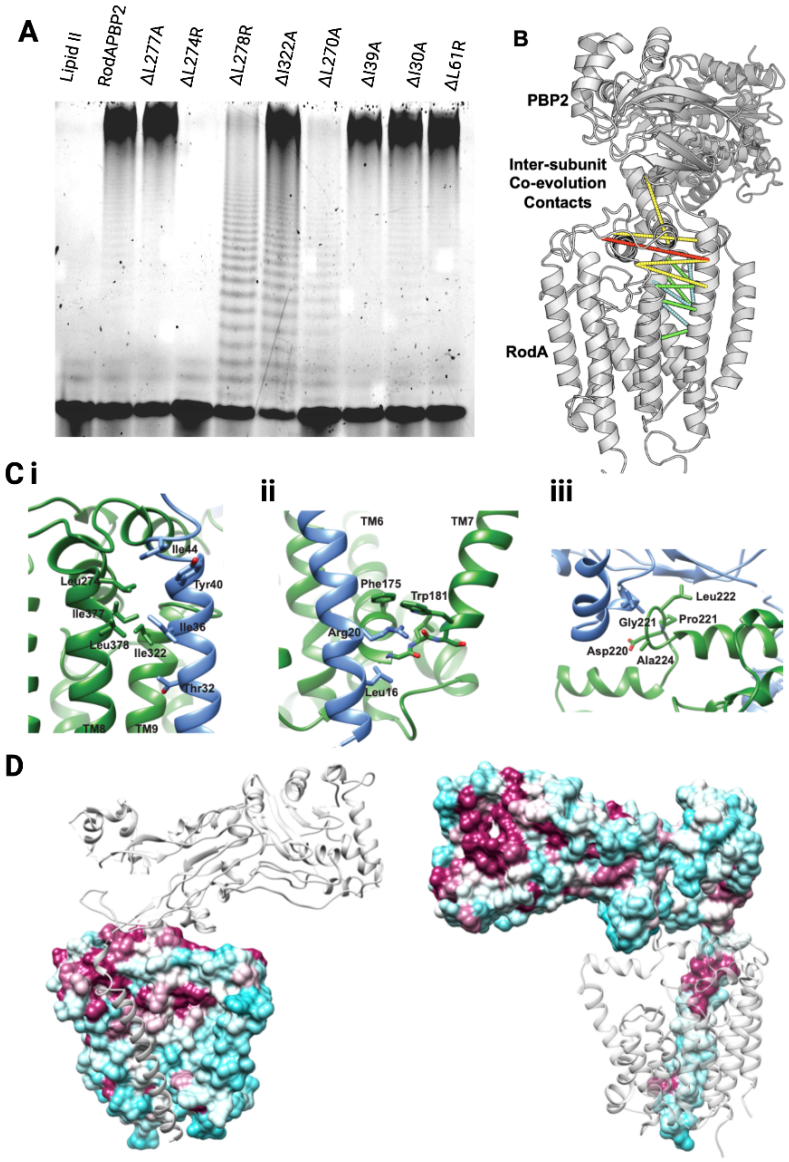


Figure 2.5.0a Movement of Transmembrane helix in RodA- upon activation by PBP2 reveals a highly conserved substrate binding cleft.

A *In vitro* activity of RodA-PBP2 mutants made at co-evolutionary pair sites, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. **B** Top 10 Co-evolutionary pairings between RodA and PBP2 reveal anchor domain and helix interactions. **C**i Co-evolutionary conserved residues visible in the Cryo-EM structure. Ii Conserved R20 of PBP2 interacts with RodA residues in structure. **D** Conservation as visualised by consurf across molecules.

In Figure 2.2.0 we revealed L274 and L278 to be interacting with PBP2's helix by a presence of co-evolution. In order to investigate this, we then interrogated these amino acids the substitution of these interactive residues *in vitro* on our construct. This revealed changes to L274, L278 and in RodA stopped and reduced activity respectively (Figure 2.5.0a) despite low conservation values, lead to loss or drop in activity in its lipid II polymerisation. This is not the case for residues not predicted to interact in our structure such as L277.

This data, in the context of a folded and covalently linked protein complex with specific amino acid conservation requirements, suggests that for RodA-PBP2 to function in complex, the interaction must be finely tuned to allow RodA's glycosyltransferase to be active. PBP2 activates RodA not only by proximity to RodA and helical attachment sterically, but also in a smaller specific PBP2-RodA amino acid sized effect across multiple binding points (Figure 2.2.0). In total, these interaction interrogating mutations lean towards a complex where both proteins are required in a specific conformational context to function.

In order to complement this, I then created a version of the RodA-PBP2 complex with a stop codon at the RodA-PBP2 linker G401STOP, as well as another construct which included the transmembrane helix of PBP2 F452STOP (Figure 2.5.0.b). Here I show, that despite the same molar concentration, there is little activity in a RodA only mutant, however when the transmembrane is connected, there is activity. This suggests that the membrane component is an important part in regulation of RodA activity.

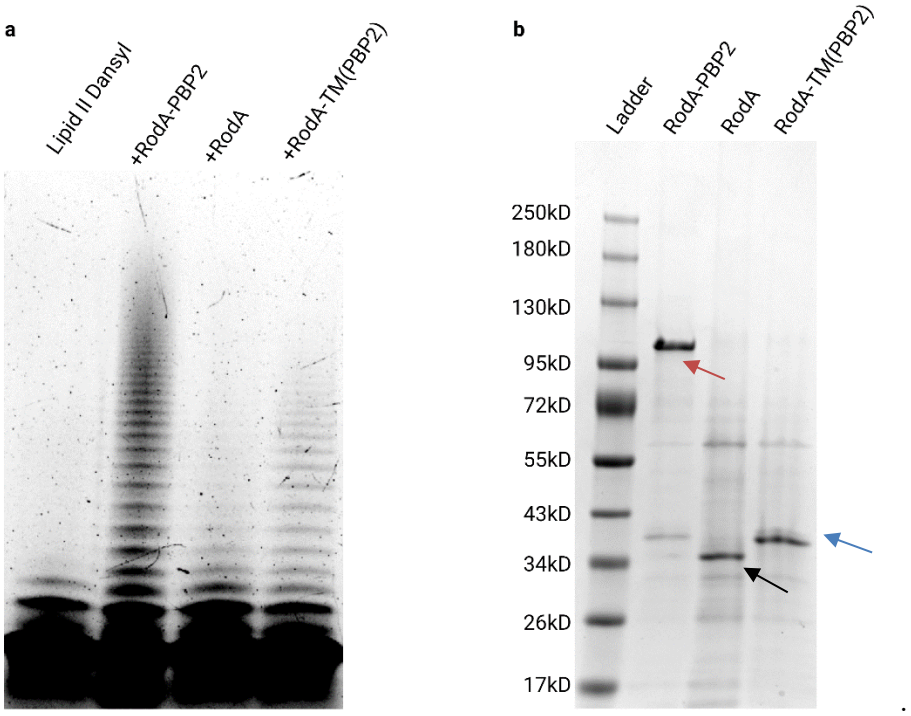


Figure 2.5.0b Helix of PBP2 required for RodA transglycosylase activity.

a 2.7µM RodA-PBP2, 2.7µM RodAG401STOP and 2.7µM RodA-PBP2 (Helix-only) F452STOP) activities on Lipid II after 1hr 30mins incubation**.** Polymerisation performed in presence of 20xmolar excess of moenomycin **b**. Coomassie stained gels of STOP mutant constructs. Arrows (Full length-orange, RodAG401STOP- black, RodA-PBP2 (Helix-only) F452STOP – blue)

2.5.1 Conservation analysis mapped onto *E. coli* RodA-PBP2 structure reveals a conserved central cleft

Our new structure also allowed for further visualisations such as the location of highly conserved regions, and comparison with other structures without PBP2.(Figure 1.8)

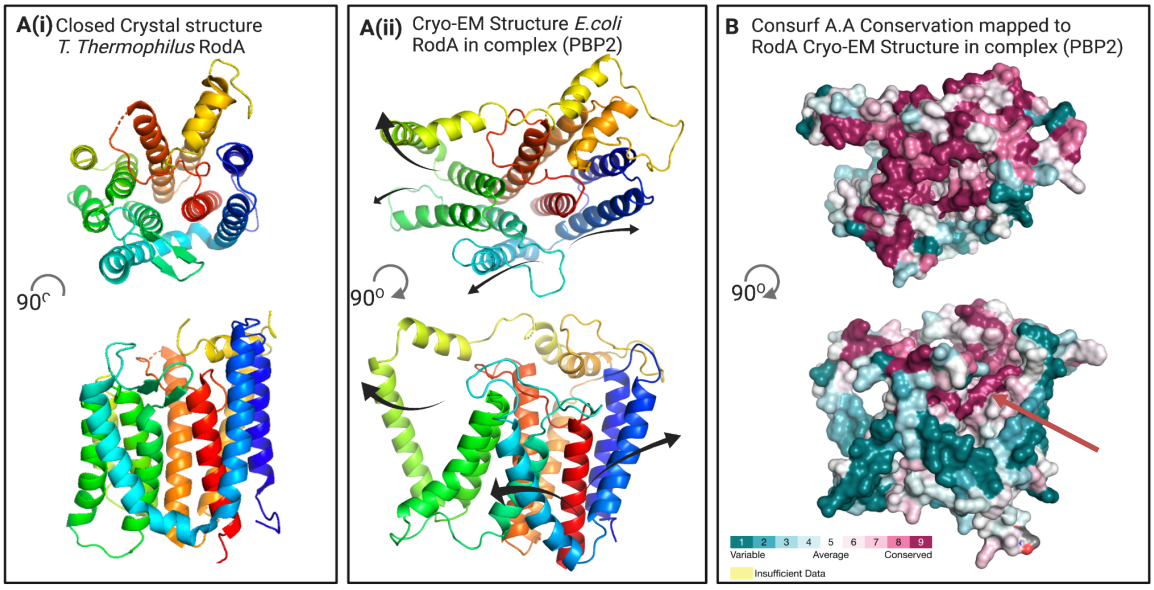


Figure 2.5.1.0. Movement of transmembrane helix in RodA- upon activation by PBP2 reveals a highly conserved substrate binding cleft.

A(i) X-ray crystal structure of closed form *T. thermophilus* RodA PDB id:6BAR (6) reveals a tightly packed structure, which upon membrane and PBP2 stabilisation, and visualisation by Cryo-EM in *E. coli* **A**(ii) opens up through swing of a membrane lateral and embedded helix to reveal two conserved clefts that form a channel across the protein **B** A highly conserved cleft and channel for lipid II is revealed in the centre of RodA by Consurf conservation analysis (Purple= Highly conserved, Turquoise = Highly Undivergent) red arrow- cleft

As shown in Figure 2.5.1.0 and as the key point of a recent paper on the *T. thermophilus* structure published part way through the PhD (6), the attachment of PBP2 to the RodA, reveals a swinging helix in our RodA-PBP2 structure, I realised that this conserved helix opens a highly conserved channel through the centre of the protein by Consurf analysis (as indicated by purple, as well as red arrow) (Figure 2.5.1.0). This highly conserved channel indicates a likely binding point for Lipid II.

Upon finalisation of our structure, our collaboration with Columbia was again fruitful, as we were also focussing on WaaL, a protein with similar activity to RodA, except which performs each transglycosylation reaction once not to form polymers and on LPS, not peptidoglycan, however also used a UPP carrier lipid.. It then became obvious from my conservation analysis and interest in this similar structure that there was a highly conserved arginine (R210) on the RodA helix that is moved out of the structure by PBP2 integration, and is retained in Waal. In Waal this was significant, as we pointed this out, among other arginine mutations contributed to O-antigen ligation by Waal (9), it also

bound to UPP. This integral residue, which shows functional reduction and conservation in Waal, and binding to UPP (Undecaprenyl Phosphate) a product of lipid II synthesis, clearly indicated to us that RodA also likely uses this as a lipid II binding site to bind the phosphate of lipid II's lipid tail. The two enzymes were linked, and interacted with UPP at the membrane.

2.5.2 RodA-PBP2 belongs to a conserved class of GT-C class proteins which bind Phosphate groups.

In Figure 2.5.2.0, and as written in the Waal-O antigen ligase paper, the conservation of an arginine adjacent to a UPP binding cleft became a characteristic similarity between the GT-C class of proteins which also includes ArnT an aminorabinose "lipid-to-lipid glycosyltransferase 4-amino-4-deoxy-L-arabinose transferase" important in polymyxin resistance, which also binds UPP (122). RodA-PBP2 of *E.coli* also contained this cleft and associated UPP binding Arginine. R210. I was the author to realise this link between proteins, and suggest the secondary binding site of the Lipid A molecule in LPS synthesis in Waal.

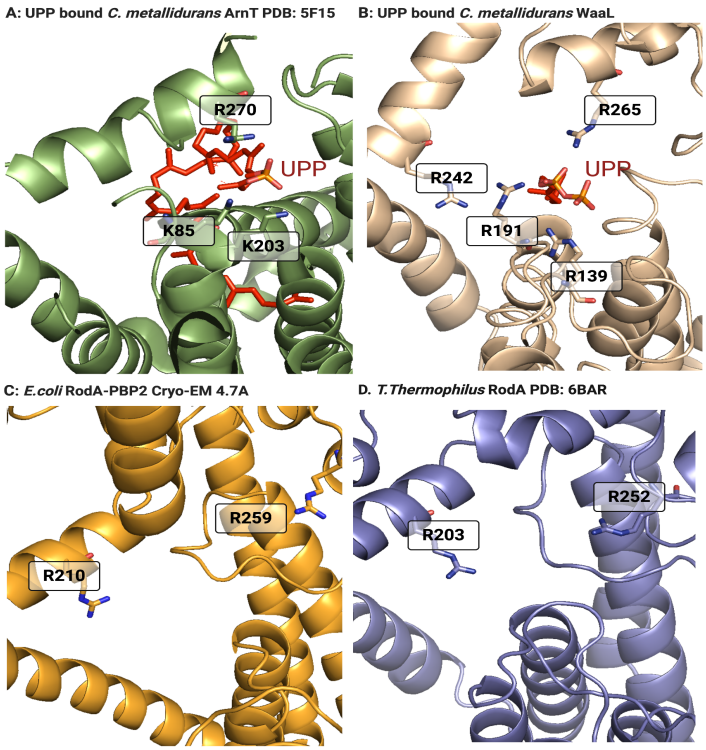


Figure 2.5.2.0 A group of GT-C proteins including ArnT, WaaL and RodA share a binding pocket, and a conserved phosphate coordinating Arginine residue

A UPP bound ArnT, with arginines shown **B** UPP bound Waal from collaborator structure with arginines shown **C**. RodA cryo-EM structure with predicted arginine orientation, **D** *T.Thermophilus* RodA. All structures visualised by Cartoon format.

It then became obvious, that indeed a group of GT-C type proteins bind phosphate groups using a coordinating arginine and helix, in coordination with a helical bundle. This again was highly suggestive of a lipid II binding site in the revealed groove. This indicated the first potential binding site of RodA and a good candidate for mutagenesis.

2.5.3 Ligand binding of RodA-PBP2 in Autodock Vina

Using the ligand docking tool Autodock, and models of lipid II constructed with ChemDraw, I then performed flexible ligand binding analysis on RodA, and found two regions within the conserved cleft as visualised in Figure 2.5.3.0 that might bind Lipid II and allow catalysis, Site A and Site B. Site A rests at a conserved region of phosphate binding by GT-C proteins, and contains conserved amino acids 262, 114, 159 and 117, and Site B was also hypothesised earlier by the flexibility of the protein as determined by Cryo-EM and its conservation.

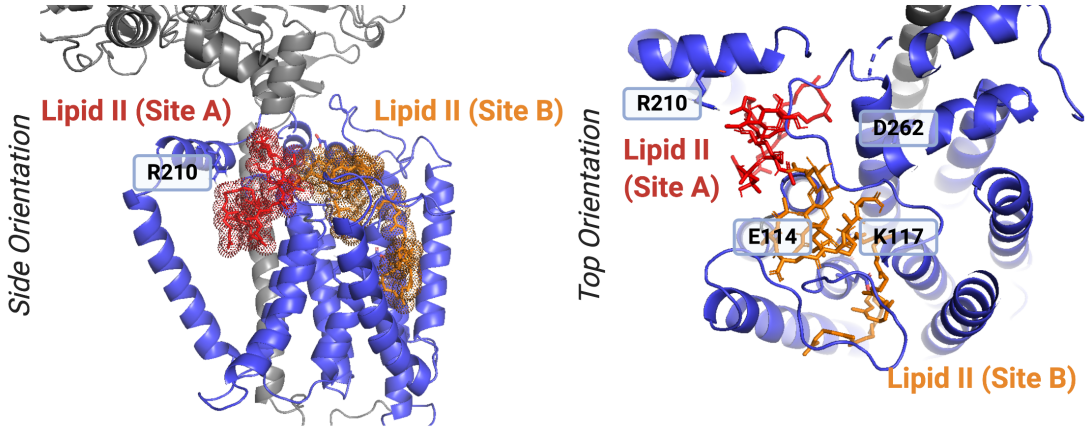


Figure 2.5.3.0. Ligand Docking and Molecular Simulation show Lipid II binding site in RodA-PBP2, and reveal lipid tail accessibility through the cleft of the protein.

RodA portion of 4.7Å predicted structure of RodA in Escherichia coli represented in cartoon format, with Autodock Vina predicted binding sites of Lipid II shown, (Site A and Site B). Important and conserved residues indicated by text boxes denoting amino acid.

My experiments indicated Lipid II is predicted to dock at two sites of the cleft A and B. After an Autodock Vina mediated automatic docking at both sides of the cleft, they reveal a potential binding site where two lipid II molecules may interact simultaneously and bind to initiate Lipid IV synthesis. This binding point is at the place of highly conserved and essential residues R101 and, D262, and adjacent R210. This also explains the high length of the C55 tail of lipid II which by our docking, seems to remain embedded in the membrane and also reach across the conserved RodA cleft simultaneously, without this length, it is likely the lipid II molecule simply could not reach the active site as the potential energy required would be too great. Two highly conserved residues within the membrane E114 and K117 are also present at this site, however we later assay these, to find they have no catalytic function *in vitro*. (Figure 2.5.5)

All this made it likely, that *E.coli* RodA (as in WaaL) recruits lipid II using its essential TM7 R210 residue and allows polymerisation to occur through two binding sites across the highly conserved cleft of the protein. This polymerisation would require PBP2 present to allow Site A to be accessible.

2.5.4 Parallel purification of RodA-PBP2 mutants and confirmation of RodA-PBP2 presence

Using the ligand binding information of Figure 2.5.3.0, and the conservation data of RodA-PBP2 in conjunction with co-evolution data of Figure 2.2.0 , we determined important mutations which would indicate the catalytic mechanism of the complex. These mutations were then made and the RodA-PBP2 complex of each purified in detergent. In Figure 2.5.4.0 the Bocillin stain of this purified protein (Bocillin binds PBP2 as a fluorescent penicillin-like inhibitor) binds at 110kD, concurrent with our protein, indicating these purifications were successful. The glycosyltransferase was assayed and controlled in the presence of MTSL dye (2.5.4.0. Hii,iii), for movement experiments, in the presence of Dansyl mDAP and Dansyl amidated Lysine lipid II as well as increased and reduced amounts of lipid II and protein, which confirmed the protein is active in a variety of conditions and thus the assays are reliable and independent of amino acid attachments at the murNac residue. The glycosyltransferase activity of the complex was assayed using a micelle lipid II exchange system outlined in Figure 2.5.4.0J, and developed prior. (Figure 2.4.5)

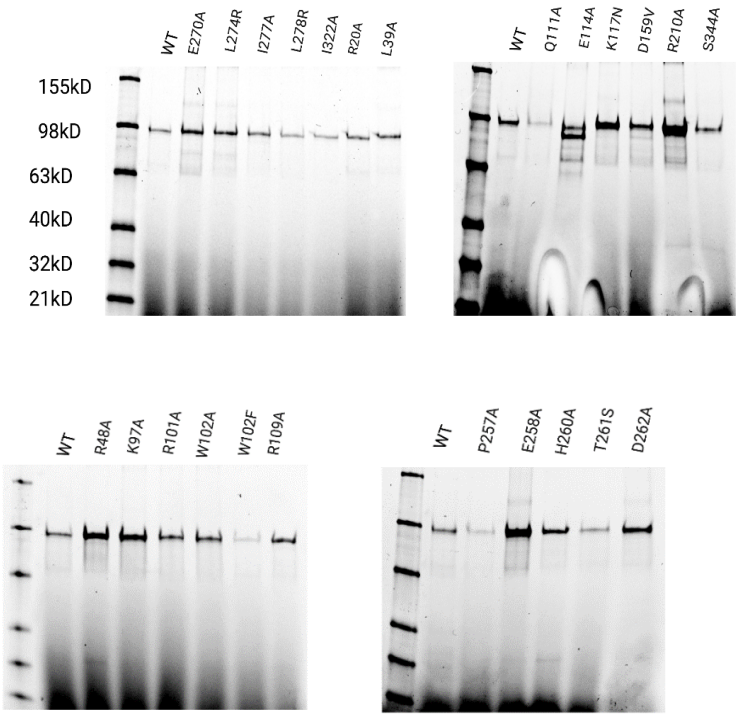


Figure 2.5.4.0 Purification of RodA-PBP2 constructs and confirmation by Bocillin Stain

Purified RodA-PBP2 mutants 15ul at 1uM visualised by Bocillin stained SDS page electrophoresis gel.

The cryo-EM data was then used to augment our existing conservation analysis (Figure 2.5.2) and identify two binding pockets based on Waal's mechanism of action (Figure 2.5.4a). The ligand binding experiment in 2.5.3, indicated amino acids adjacent to the predicted binding site, which we then mapped on the structure and labelled according to conservation (Figure 2.5.4b). I asked Phil Stansfields group, Jonny Colburn with coordination from David Roper, if we could further interrogate the binding of lipid II on the RodA molecule in its open state obtained via cryo-EM (with the flexible loop 97-110 modelled in for stability), to see the likelihood of the regions importance. This in Figure 2.5.5c clearly showed two regions of high binding between TM3 and TM2 which represented Cavity A, as well as TM7 and 5 which represented Cavity B. Rie Nygaard then found a density of lipid adjacent to our Cryo-EM structure (Figure 2.5.5d). I interrogated the charged and conserved residues to find the amino acids most likely involved in binding. (Figure 2.5.5e)

Using the cavity search and flexibility data gained using our *E.coli* Cryo-EM structure (Figure 2.3.1), and the conservation data of RodA-PBP2 in conjunction with co-evolution data Figure 2.5.5/2.5.3.0/2.5.2, we determined important mutations which would indicate the catalytic mechanism of the complex based on conservation across the molecule (Figure 2.5.5). Using the same technique as previously, these mutations were then made and the RodA-PBP2 complex purified in detergent. In Figure 2.5.4 the Bocillin stain of the detergent solubilised purified protein constructs and mutants indicate these purifications were successful, and of similar concentration, within bounds of our control concentration dependence experiments. This when assayed then indicated D159V, R101, R48A, and R210 to be important to ligand binding (Figure 2.5.5), and concurring with two large binding cavities.

In Figure 2.5.5e, which was also published, there were conserved residues in the binding pocket (Figure 2.5.5b) which affected polymerisation efficiency, by a small reduction in polymerisation. However what was more telling, was that there were residues that were very well conserved such as E114 and K117 which even when replaced still allowed *in vitro* protein to polymerise peptidoglycan, this then suggested these were likely not involved in the reaction itself. Other residues reduced polymerisation markedly despite being far away from any predicted catalytic site, on both sides of the RodA molecule responsible for lipid II binding. This confirmed that the region was important in general to catalysis.

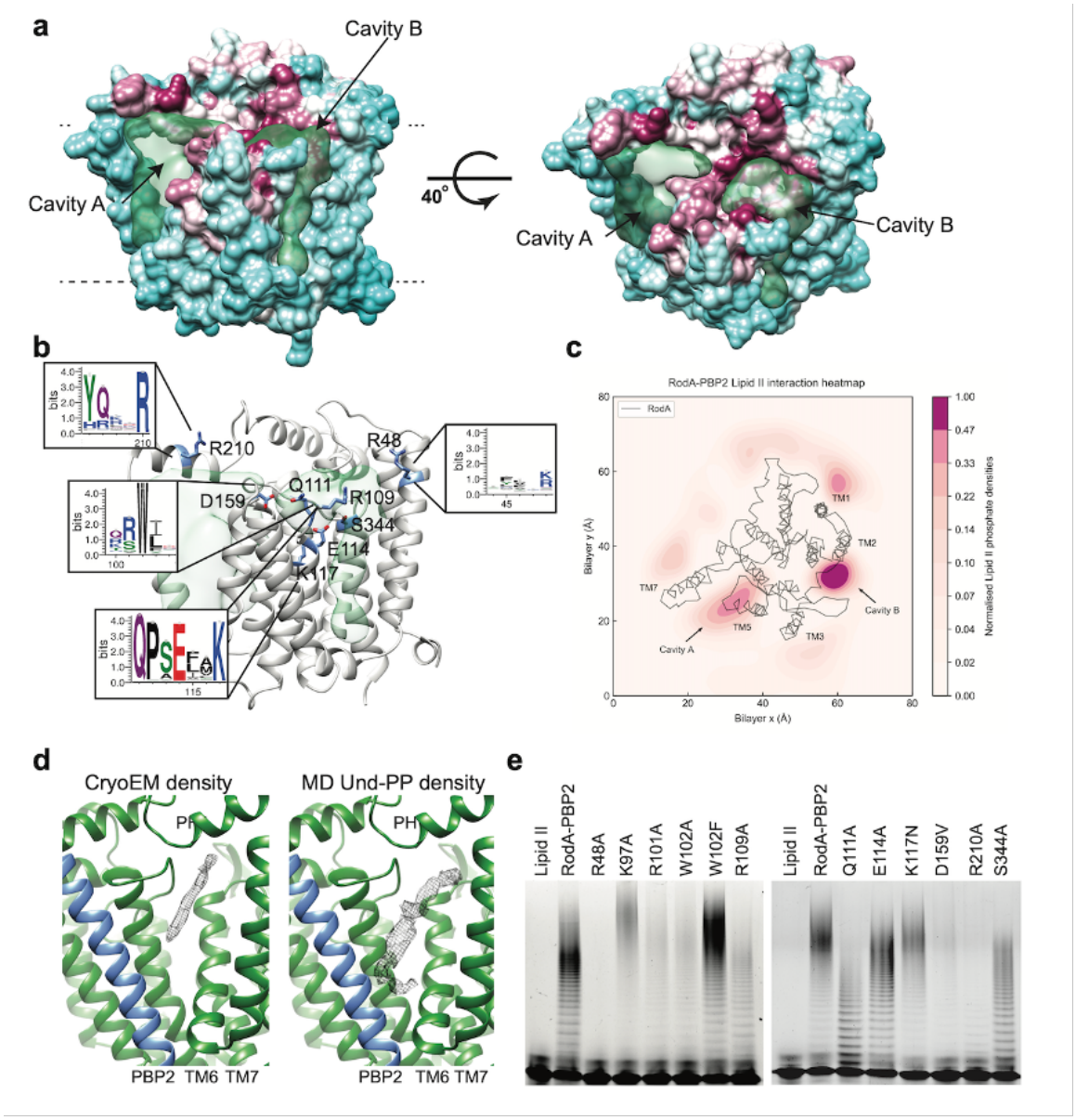


Figure 2.5.5. Ligand binding pocket of RodA-PBP2.

a Cavity search of RodA-PBP2 shows two binding sites across a conserved cleft. **b** Conserved Lysine and Arginine residues in region when modified lead to reduced or lost activity **c** Lipid II continues binding at both sites after MARTINI GROMACs simulation. **d** Cryo-EM Density reveals binding of UPP like molecule at GT-C active site **e** Conserved residues found at sites A and B, shown to have affect on lipid II binding *in vitro. Cryo-EM density of Und-PP by Rie, and ligand binding density of lipid II by Jonny Colburn.*

The presence in this binding site of arginines R48, R109 and R210 at critical potential ligand-binding points and existence of RodA-PBP2 complex within the GT-C family of UPP binding proteins (2.5.2.0) and for R210 its conservation (Figure 2.5.5b) suggests these residues, are also likely essential for high catalytic capability *in vivo* for lipid II recruitment within the membrane, and support the two lipid II binding sites hypothesis proposed at Cavities A and B. Figure 2.5.5e , which indicates the *in vitro* activity of these mutants suggested that K48 and R210, are important to the mechanism, similar to Waal.

Less well conserved residues such as K97 when mutated to alanine, as well as R109 do not have any effect on the activity of the complex and its ability to polymerise lipid II, however Q111 changes the efficiency of polymerisation and too R210 and R48. Despite previously expecting K117 and E114 mutations to have a negative effect on polymerisations due to their higher conservation and charged nature at the centre of the molecule (Figure 2.5.5b), both led to wildtype activities, suggesting catalysis and binding is not associated with these residues, perhaps these served another purpose the assay could not uncover, E114 mutants did seem to produce less stable protein, according to the bocillin assays.

This in all suggested that in cavity A the R210 residue was important to ligand binding, much like D159, which both reduced activities, however these residues were not involved in catalysis, as activity was still present. The *in silica* binding to cavity b shown in Figure 2.5.5e as well as changes in polymerisation by Q111, R101 and W102 mutations (rescued by W102F) was affirmative of two binding sites.

Finally, in Figure 2.5.5d we found a density in the cryo-EM at cavity A adjacent to D159 and R210 indicative of a bound lipid. Simulations by Phil Stansfeld, found this to correspond to a density produced by Und-PP, which mirrors existing densities found in ArnT(122) and Waal GT-C motif family glycosyltransferases to which RodA belongs(9).

After glycosyltransferase was assayed, the collaboration and data collection began, however to ensure the complex was entirely active including PBP2, we assayed the protein for transpeptidation activity.

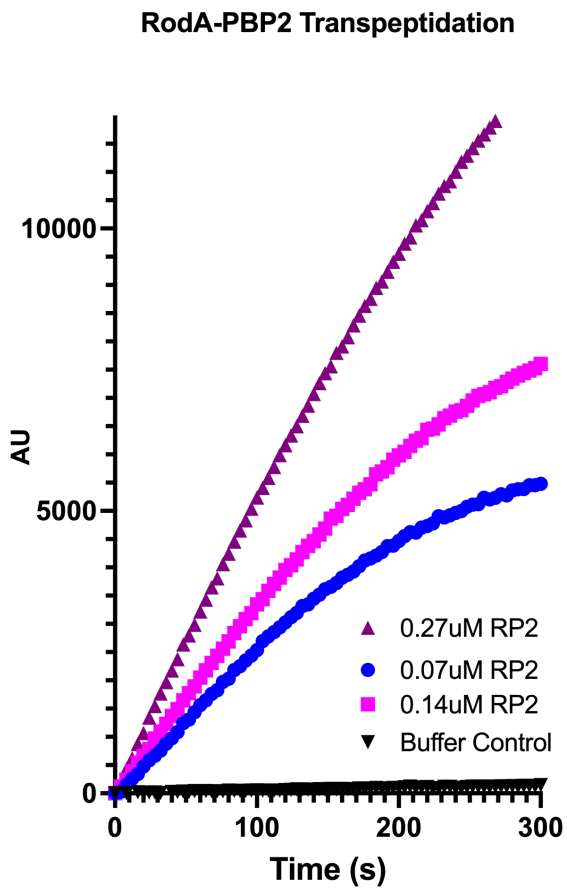


Figure 2.5.6.1 Transpeptidation activity of RodA-PBP2 assayed by D-alanine release assay

AU- arbitrary units of fluorescence indicated by horse radish peroxidase fluorescent product Resofulin from amplex red, facilitated by D-alanine release and hydrogen peroxide production by D-amino acid oxidase. D alanine release directly measured by change in fluorescence at 525nm. Concentration dependence indicated.

Change in absorbance of the solution after substrate addition allowed for the production of D-alanine to be measured. This showed that upon increasing RodA-PBP2, activity changes in line with an equivalent activity increase dependent on protein concentration, consistent with RodA-PBP2 itself is causing release of the D-alanine from substrate upon addition. However this experiment cannot be taken as direct evidence of PBP2 directed transpeptidation activity since it is also possible that D-alanine release is detected as a result of carboxypeptidase reaction (i.e. futile release of D-alanine from the formation of a PBP2-tetrapeptide intermediaet with water acting as an incoming nucleophile). I attempted to resolve this with mass spectrometry

experiments of the peptidoglycan products of such incubations, but we were unable to produce sufficient qualities of starting material and product in the time available. Future experiments and experimental verification are required to resolve this issue.

2.6 Amino acid substitutions and conservation reveal active site of RodA

After having confirmed the complex was active by transglycosylation, and PBP2 had enzymatic activity, (2.5.6.1 and 2.5.6.2) The protein could be interrogated for mutations which no longer allow activity. These mutations then highlighting the essential catalytic and ligand binding regions of the protein. Alongside this the wildtype detergent purified construct was assayed against methicillin and moenomycin to ensure it is active independent of any potential contaminants.

The revelation of RodA being part of a set GT-C type proteins that bind lipid II (Figure 2.5.20), indicated that R210 was essential despite its structural isolation on a helix, and our new hypothesised binding site at this cleft left us to interrogate the highly conserved residues along this cleft between Cavity A and B, as well as nearby conserved residues in coordination. I then through enzymatic analysis in Figure 2.6.1 showed that indeed these residues alter activity, which indicated further that our hypothesis is correct, and that it does not involve proposed residues suggested by *T. thermophilus* E258, which shows wildtype activity, and indeed modification in this membrane embedded cleft infact leads to activity loss, at the charged residue D262 as well as the helix positioning residue L274 and P257, which suggests D262 alone to be essential for Lipid II polymerisation. In the summary of mutations made (Figure 2.6.1), the essential and catalytically altering residues are in this cleft, with arginine R210, R101 and R48 showing altered and reduced catalysis, along with prolines along it which manage its shape.

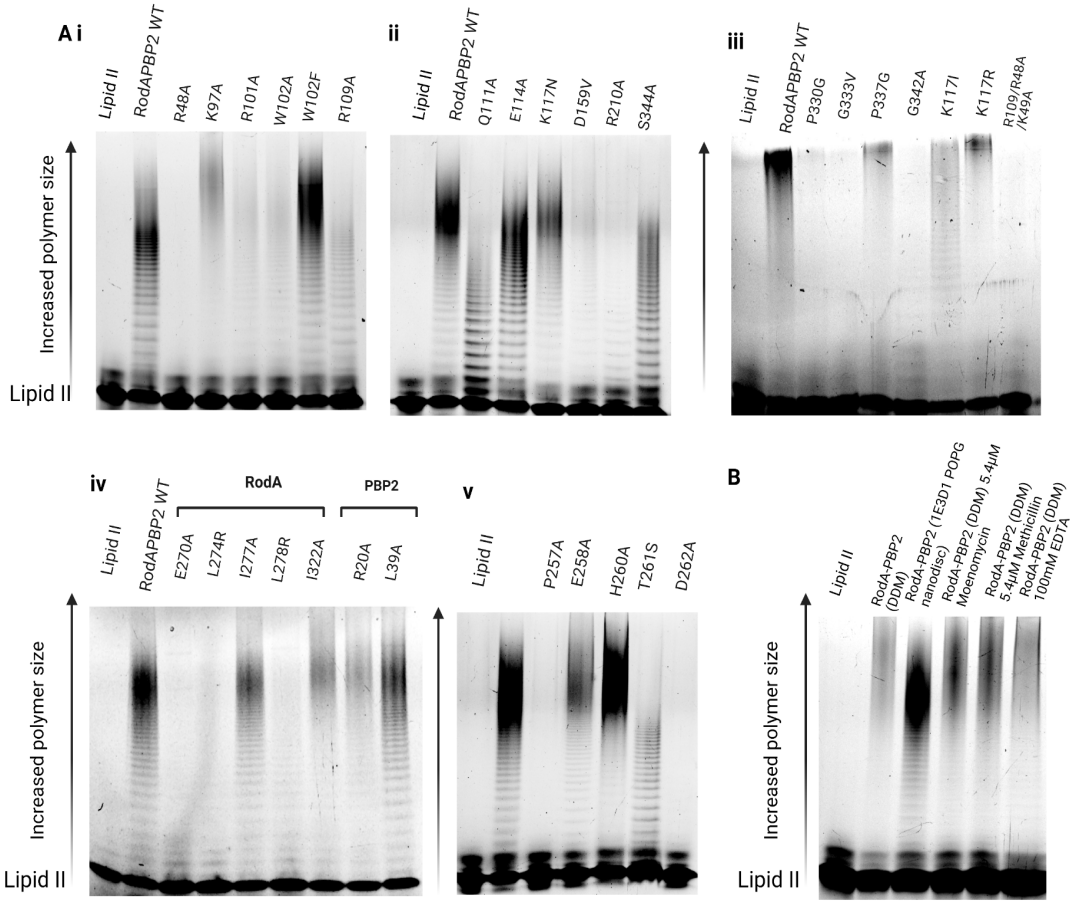


Figure 2.6.1 RodA-PBP2 active mutants reveal two important catalytic clefts and an absolutely necessary D262

A (i-v)16% Tris/Tricine gel electrophoresis analysis of RodA-PBP2 and mutants. B methicillin and moenomycin controls, as well as a repeated EDTA control, confirm that no type 1b enzymes are active in the wildtype RodA-PBP2 mutant.

In addition to the mutations, I also assayed the RodA-PBP2 complex to ensure that it was consistent between DDM preparations and nanodisc encapsulated protein, as well as to investigate if the complex was inhibited by moenomycin, a pbp1b inhibitor, or methicillin a pbp2 inhibitor (Figure 2.6.1 B) Although the complex produced alternative products, there was no appreciable quantitative difference between the products when the complex assayed under any of these conditions, therefore the activity seen of RodA-PBP2 was real, as indicated by

presence of moenomycin, an inhibitor of any PBP1b which may have been contaminating samples if it were shown to be affected. In addition methicillin, and therefore any transpeptidation, is confirmed to not be a part of the products we see, as shown by no difference in the products formed.

Table 2.6 Summary of activity loss

	GEL 1 (PUBLISHED)	GEL 2	GEL 3
E270A	NA	NA	NA
L274R	NA	NA	NA
I277A	WT	WT	WT
l278R	NA	NA	NA
I322A	WT	WT	WT
R20A	WT	WT	WT
l39A	WT	WT	WT
R48A	LA	LA	NA
K97A	WT	WT	WT
R101A	LA	LA	WT
W102A	LA	LA	WT
W102F	WT	WT	LA
R109A	WT	WT	WT
Q111A	Altered activity	Altered activity	Altered activity
E114A	LA	WT	WT
K117N	WT	WT	WT

D159V	LA	WT	WT
R210A	LA	LA	LA
S344A	WT	WT	WT
P257A	NA	NA	NA
E258A	WT	WT	WT
T261S	WT	WT	WT
D262A	NA	NA	NA
WT	WT	WT	WT

The catalytic summary of RodA-PBP2 mutations *in vitro* in Table 2.6 indicated that there are many highly conserved residues, which upon knockout still allow the protein to retain activity. These residues therefore were unlikely involved in the mechanism, however were involved in polymerisation, as indicated by conservation and reduction in efficiency of polymerisation.

In order to reduce the complexity of the possible mechanistic residues, I then created a summary of literature so far, conservation, FtsW studies- a similar protein, and finally our own *in vitro* data, to reveal a map of residue importance according to the literature, in equal part with highly reductive scores upon activity (Figure 2.6) This revealed residues 262A, W102 and R210 are those residues most likely to be essential for polymerisation. However, we had yet to form an argument for a processive mechanism required for chain polymerisation.

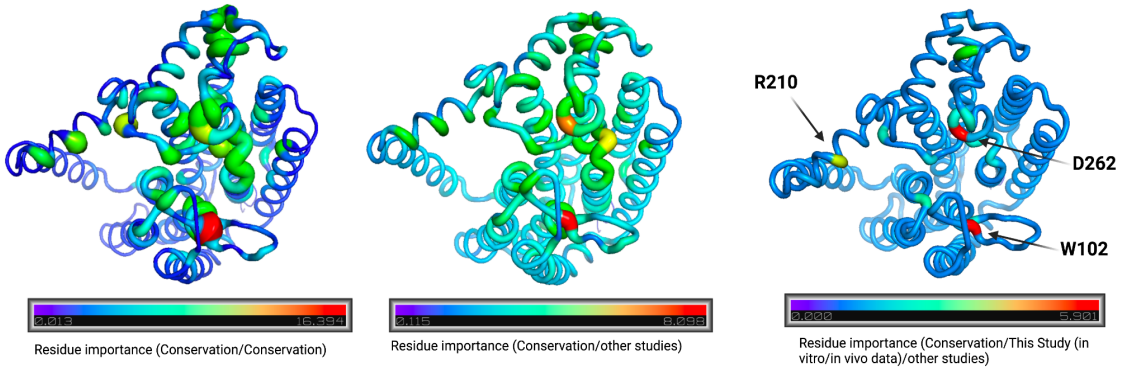


Figure 2.6 Activity loss in literature and our own studies reveals three integral residues. Residue importance mapped to intensity and size on our Cryo-EM.

Panel **A**, Conservation of RodA, Panel **B**, conservation of RodA multiplied by importance shown in existing papers, Panel **C**, Conservation of RodA, multiplied by importance in other papers and my data.

2.7 Flexibility of the RodA-PBP2 complex interrogated by DEER.

Mutations of the RodA-PBP2 complex revealed integral amino acids in RodA, however did not reveal the mechanism by which this likely occurred. I had isolated two likely binding sites through autodock ligand docking (Figure 2.5.3.0), however the movements of lipid II which would then allow for polymerisation were certainly not understood. Therefore, by collaborating with Dr Meagan Dulphrisne within the research group of Dr Linda Columbus at Virginia University, we devised and created a set of mutants for analysing the movements within the molecule, especially across the conserved cleft and TM7 conserved across GT-C proteins. These mutants were created to label cysteines with a maleimide label to interrogate self interaction. I made the first “cysteine-less” mutant myself, however Meagan mutated the other constructs onto this before I purified them in the UK.

After creating these mutants I purified each, and confirmed their stability by labelling with Bocillin, which binds PBP proteins at their transpeptidation site when folded. This indicated the PBP2 of the RodA-PBP2 was correctly folded, thus the complex was usable.

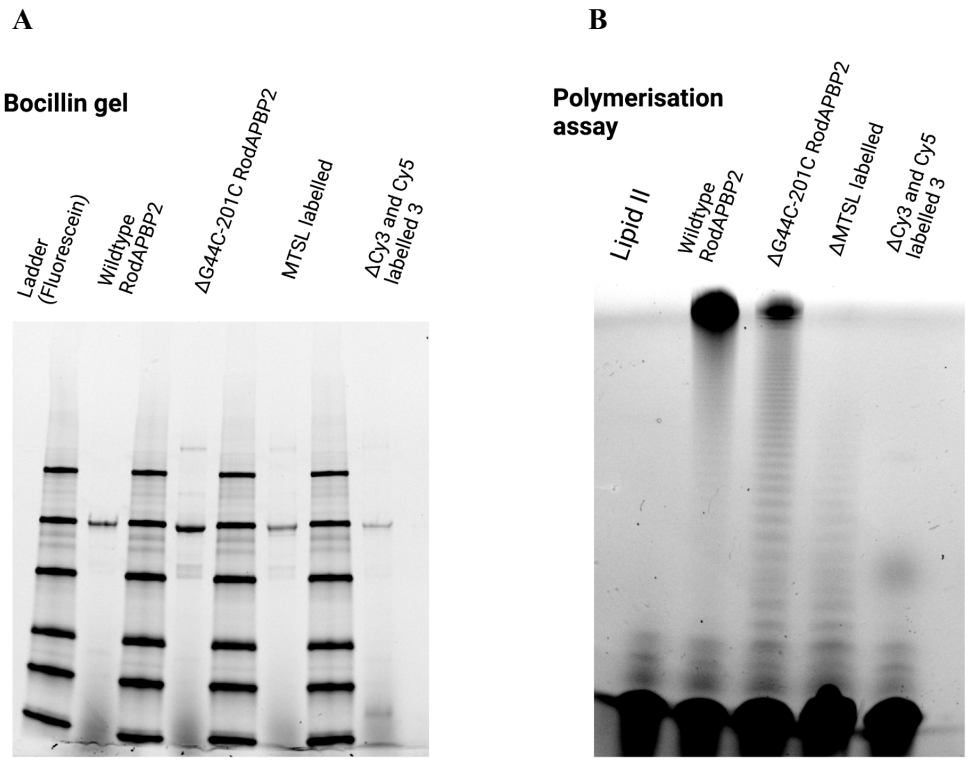


Figure 2.7.0 Use of MTSL DEER dye and Cy3/5 dye on catalysis proximal residues reduces activity.

A Maleimide and FRET labelled RodA-PBP2 Bocillin stained SDS page electrophoresis gel, purified protein constructs. **B** Confirmation of maintained activity of Maleimide labelled RodA-PBP2 used in DEER movement experiments

I then assayed my purified proteins for polymerisation, and found labelling of the RodA-PBP2 proteins by DEER and Cy3/5 Dyes then had an effect on polymerisation, this was likely due to the sterically hindering groups attached at times to regions near my predicted active site. Further

to this confirmation that the construct was still active and therefore it was useful to test via DEER (although less active), our collaborator Meagan at Virginia University took my mutants, which were for proximity produced in by Colombia University collaborators.

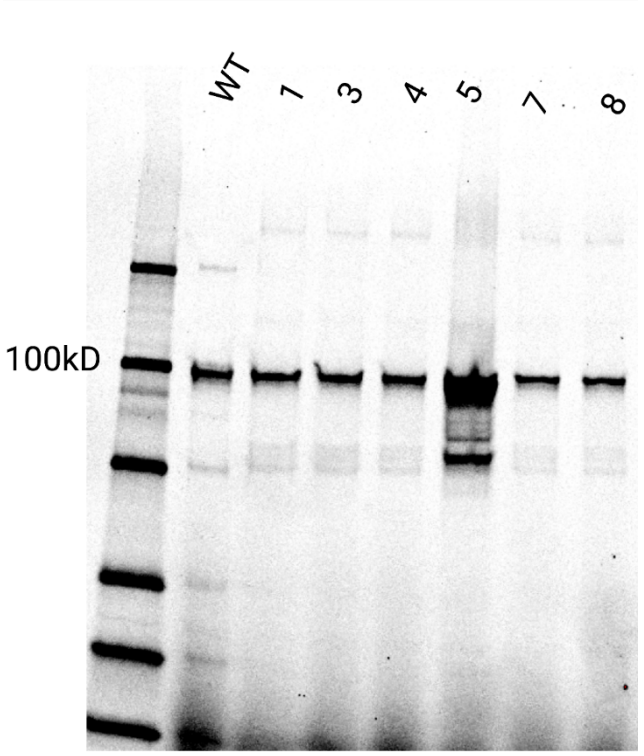


Figure 2.7.1 Quadruple Cysteine mutant bocillin binding and activity shows protein stability post cysteine insertion.

Bocillin stain gel of RodA-PBP2 mutants confirming stability. 1. G44 - D90, 3. G44 - F201 4. G44 - T48 PBP2, 5. G44 - Q99 PBP2, 7. T48 PBP2 - D90, 8. T48 PBP2 - S163

I then purified a set of eight protein pairs across the RodA-PBP2 protein, corresponding to unique points on the RodA-PBP2 molecule (Figure 2.7.1). After purification I ensured these proteins were still intact and did not degrade through use of a bocillin stain, which indicated the PBP2 component was folded and the construct expressed. This then allowed my collaborators in the US, to make the molecule in parallel and test these for flexibility under the Double Electron electron resonance system. The results of this experiment are shown in Table 2.7.2 and Figure 2.7.2

Table 2.7.2 Predicted movement of RodA and PBP2 summary – (Thanks to Meagan Dulphrisne for data collection)

Residues (fusion numbering)	distance (ca- ca)	Description	DEER result summary
1. G44 - D90	38 Å	Within RodA	Main distance distribution around 35 Å, however multiple peaks (Figure 2.7.3)
3. G44 - F201	56 Å	Within RodA	Main distance distribution centered at ~53 Å with a possible secondary distribution centered at ~43 Å
4. G44 - T48 PBP2	40 Å	Between G44 for RodA to PBP	Too noisy to reliably interpret.
5. G44 - Q99 PBP2	41 Å	Between G44 for RodA to PBP	Broad Gaussian distance distribution centered around 45-50 Å
7. T48 PBP2 - D90	50 Å	Between T425 of PBP and RodA	Main distance distribution centered at ~50 Å with a broad distribution of smaller distances (with lower confidence)
8. T48 PBP2 - S163	37 Å	Between T425 of PBP and RodA	Main distance distribution centered at ~47 Å with a broad distribution of smaller distances (with lower confidence)

A simulation of some of the most interesting interactions was then made, in addition to the data to reveal peaks of distance. The most interesting to me, was that of the TM7 related data (G44-F201), in addition to the cleft peaks of G44-D90 which reveals the conserved cleft of the protein moves apart.

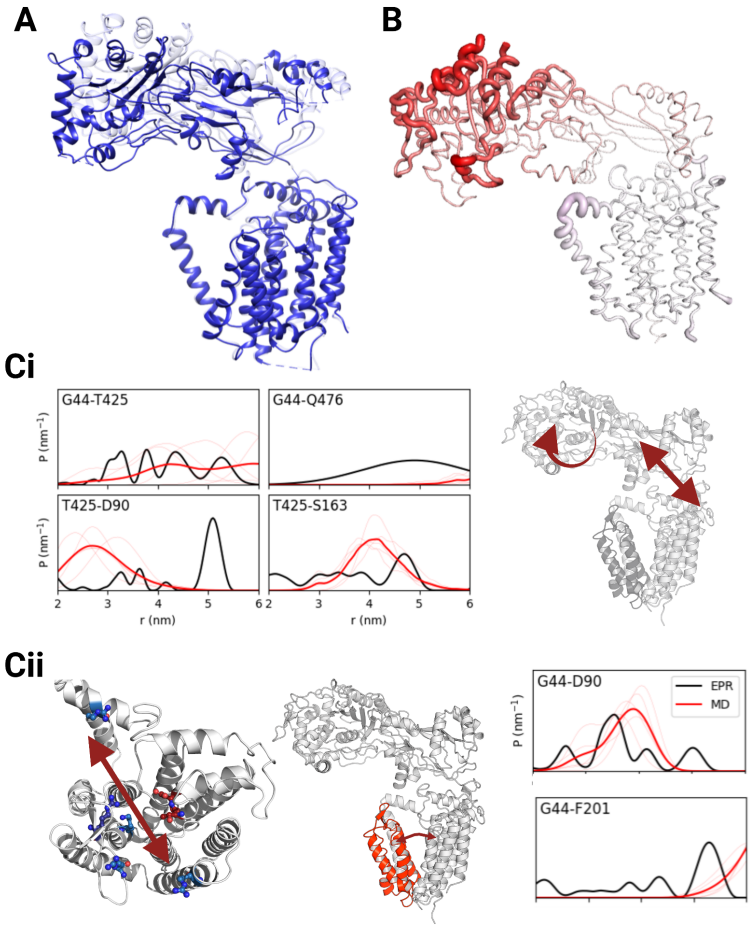


Figure 2.7.3 Flexible domains of RodA gate a lipid II synthesis as determined by DEER interactions.

A Extreme conformations of PBP2 as visualised in Cryo-EM data subsets , acknowledgement and thanks to Rie Nygaard. **B** Reversed Certainty values of Cryo-EM data by residue, indicating potential residue flexibility. **C**i DEER experiments between anchor domain of PBP2 and RodA, indicating multiple peaks and movement. **Cii** DEER experiments between TM7 and TM2 of RodA showing a range of peaks, and TM4 and TM2 (G44-D90) revealing cleft flexibility within molecule.

Cysteine labelling mutations of the RodA-PBP2 complex, with the cysteines removed, allowed for understanding in the conformational isomers of the molecule (Figure 2.7.3) purified and assayed in Figure 2.6.1 to confirm activity and folding. The dynamics revealed by these interactions is shown in Figure 2.7.3. The most interesting mechanistically was the G44-D90 mutation, this revealed the ability for transmembrane helices 3 and 1 to occur in more than one distance confirmation, suggestive of a flexible cavity at our proposed secondary site of lipid II binding. The G44-210 distance measurement also reveals the RodA-TM7's flexibility and multiple states of the protein when RodA is bound to PBP2 at the primary lipid II binding site shared with GT-C proteins as a phosphate binding site.
For illustrative purposes, I also attach Cryo-Em subclass data of PBP2. PBP2 reveals multiple conformations in our Cryo-EM structures, indicating a rotation and increased flexibility at the transpeptidase site, but also rotation of the PBP2 molecule. This rotation may shed light on the final transpeptidation action of the complex. (Figure 2.7.3)

The ability for the cavity predicted to be important in ligand binding to be flexible and move apart (Figure 2.7.3c) was indicative of a mechanism where perhaps the cavities could connect. This was formative in understanding the RodA molecule as a processive enzyme able to accept lipid II substrate molecules through one channel and release product through another.

2.8 UPP binding of GT-C protein RodA

Finally it was important to consider what lipids may be in the nanodisc and were seen in the density earlier in Figure 2.5.5d, therefore I ran a Folch lipid extraction of our RodA-PBP2 nanodisc samples on Thin layer Chromatography sheets. (Figure 2.8.0) Our nanodiscs were prepared in POPG, therefore we expected to see this, but also an Und-PP molecule.

Figure 2.8.0 TLC of purified RodA-PBP2 complex in nanodiscs.

TLC performed on XL Silica 60. Lipids labelled extracted by Solvent A before addition to TLC plate along X axis. TLC visualised through iodine staining method.

The RodA-PBP2 POPG filled nanodisc from Colombia – RodA-PBP2 nanodiscs had a similar TLC read-out to UPP (Figure 2.8.0) as perhaps expected, however this was not prepared from native lipid and was momentarily isolated in DDM, so shouldn't have been as marked as visualised. The TLC of RodAPBP2 extracted by DDM however showed a POPG like peak, inconsistent with the expected readout if DDM was being retained through purification. The nanodiscs, by lack of POPG presenec, indicated a failure of POPG incorporation or high presence of DDM or UPP in the sample. I therefore wanted to extract a selection of lipids proximate to the complex to increase bound lipid concentrations visible in any subsequent TLC. This led to isolation of the expressed RodA-PBP2 complex using SMALP, a non-detergent method for extracting native lipids(123) for an improved look on the presence of UPP in the nanodisc. I purified RodA-PBP2 in SMALP (Figure 2.8.1), however I was unsuccessful in capturing a good TLC picture with this sample, despite the effort. This may be used in the future

for the investigation of the complex, however could not yield a high amount of protein in my hands.

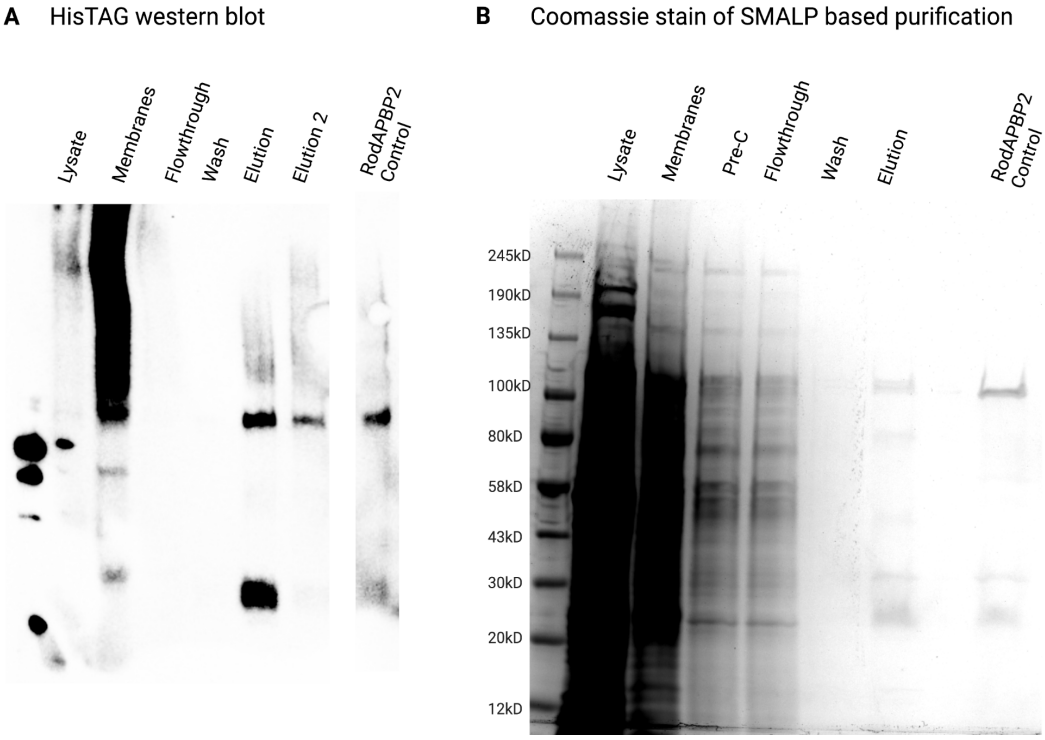


Figure 2.8.1 SMALP encapsulated RodaPBP2

Purification of SMALP encapsulated RodA visualised by western blot and histag visualisation A, and coomassie stain, RodA-PBP2 control DDM lane juxtaposed visually B, Coomassie stain of proteins purified, indicating RodA-PBP2 purification adjacent to correct control band, revealing potential use of SMALP to encapsulate RodA in future studies if necessary or in other homologous proteins.

Therefore in order to look for UPP in the nanodisc encapsulated sample , we moved to mass spectrometry of the samples, after lipid isolation.

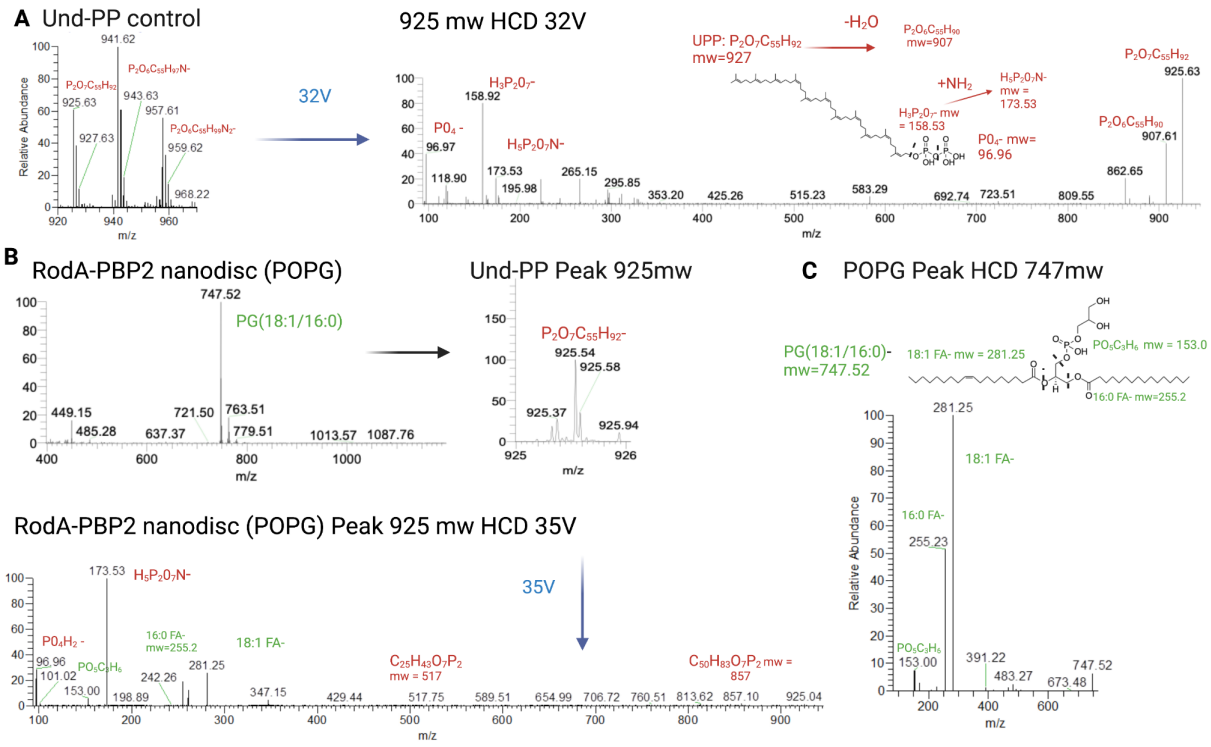


Figure 2.8.2. Undecaprenyl pyrophosphate (Und-PP) identification within RodA-PBP2 nanodisc sample through OrbiTrap Mass spec. (negative mode)

A Und-PP as visualised as ammonia salts, B Spectra of nanodisc sample between 400mw-2000mw, indicating small 925mw Und-PP peak, with masses indicative of Und-PP. C Main POPG peak of sample at 747mw.

In Figure 2.8.2, we analysed an undecaprenyl pyrophosphate (Und-PP) control by negative mode mass spectrometry and high collision energy products (120,121,124), to find that Und-PP does spray and create peaks, however even at high concentrations this was poor. However we do see expected peaks at 158mw, and 96.66mw, as well as higher molecule weights as predicted. This control when compared with smaller peaks in the nanodisc sample of RodA-PBP2 we had purified lipid from by chloroform extraction, indicates there is likely some Und-PP in the RodA-PBP2 nanodisc, which as identified in Figure 2.5.5, could be visible in the Cryo-EM density and hint at the mechanism. However this was low quantity and is not conclusive of a high amount of Und-PP trapped by the complex. Und-PP peaks were critically also identified in an empty nanodisc control, suggesting the POPG itself may have trace Und-PP present. However despite this, the conservation of the GT-C fold (9), and our simulations, and this mass spec data as well

as that of the TLC pointed towards Und-PP as a RodA-PBP2 binding partner, and the data serves to show the density could be Und-PP, independent of the source.

2.9 Mechanism of RodA-PBP2.

In addition to looking at the structure, to find a lipid binding region across two cavities of RodA in Autodock vina (Figure 2.5.3.0), and finding residues of importance identified by conservation and *in vitro* mutation (Figure 2.6) and Table 2.6., in order to find a mechanism, the physical characteristics of the interaction between lipid II substrates and RodA also needed to be characterised.

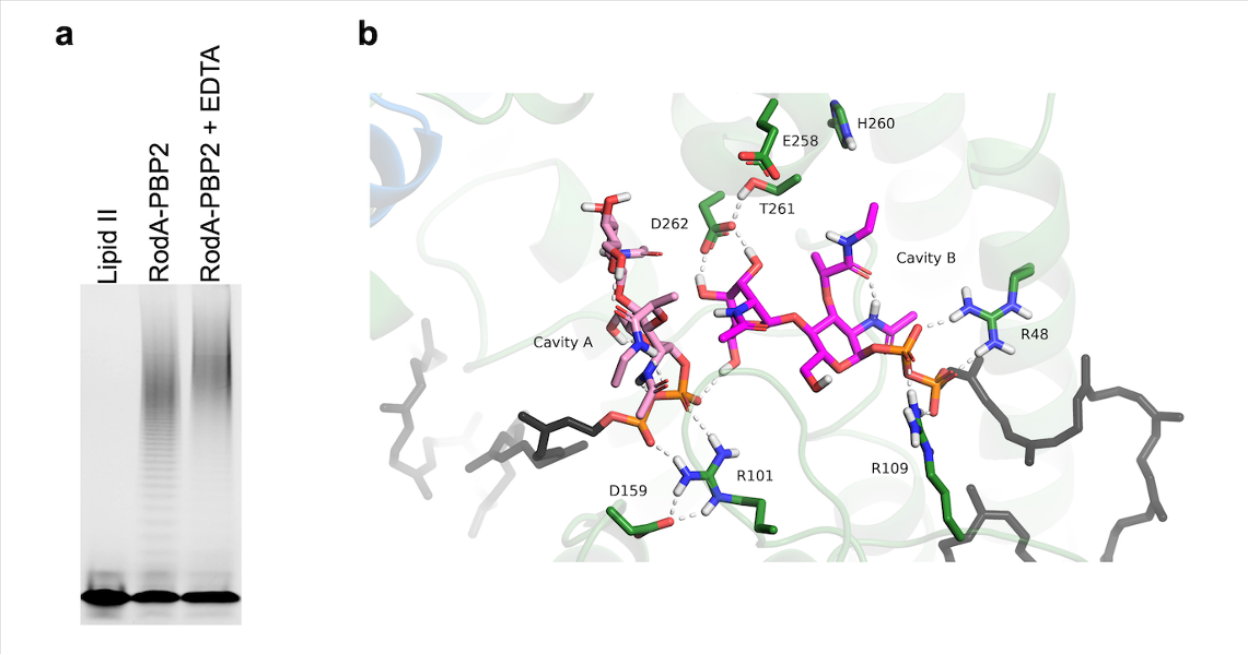


Figure 2.9.0 Chemical parameters allowing for a transglysocylation reaction in RodA

a- RodA-PBP2 polymerisation in the presence of EDTA, a metal chelating agent (35mM) , **b**- A simulated lipid II, Lipid II translgysosylation at the predicted active site across Cavity A and B, courtesy of Jonny Colburn. (Using Autodock vina poses of 2.2.0)

In Figure 2.9.0a we assayed the RodA-PBP2 apparatus to show that it is metal independent, through use of a metal chelator EDTA. This indicates there are no requirements for metal in the reaction carried out by RodA, unlike that of the transglycosylase of PBP1b (64,125,126). In

Figure 2.9.0b, performed by Dr Jon Colburn lipid II molecules were simulated at Autodock vina derived locations identified earlier, and allowed a molecule in each proposed cavity to relax in a simulation before a reaction mechanism forced to occur, with the resulting energies being favorable in the hypothesised pocked (Figure 2.9.0). The residues identified as central to the reaction were adjacent to D262, a residue of high importance identified earlier, and D159. This suggested a mechanism was possible involving the D262, and when the simulation was asked to perform a reaction of the glycosyltransferase using D262 as an electron donator it found this to be energetically favorable. This combined with prior data on the flexibility of the transmembrane domains involved in cavities A and B (Figure 2.7.3) and my mutation data, (Figure 2.9.1b) suggested this site was the catalytic centre. R101, R109 and R48 all interacted with the phosphate residues of the lipid II, similar to other GT-C motif containing proteins which interact with phosphates such as the O antigen ligase Waal (9)

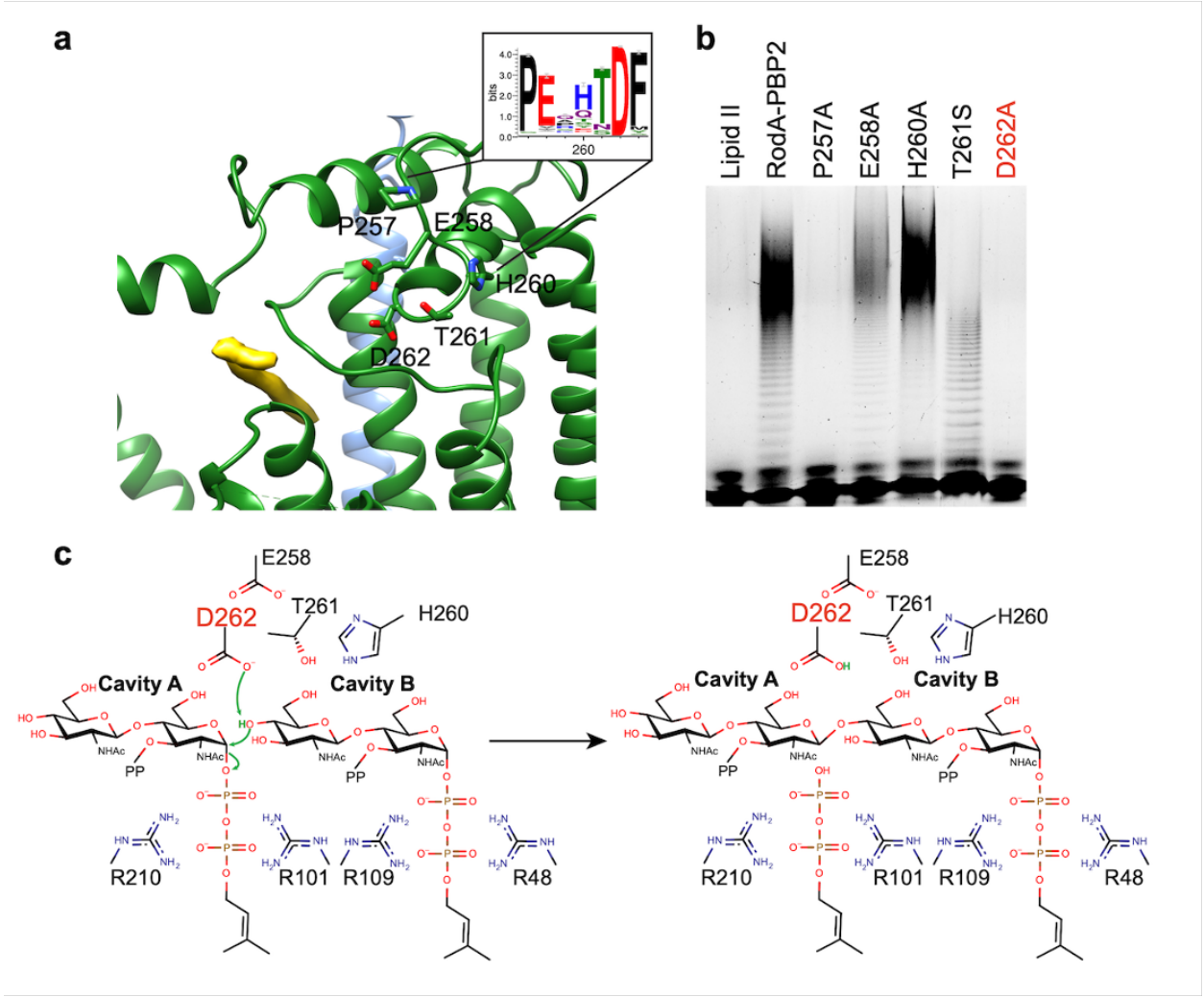


Figure 2.9.1. RodA active site of glycosyltransferase

a Proposed binding site of lipid II **b** Activity in vitro gel of RodA-PBP2 revealing on D262 mutant to be completely dead, and others of reduced activity. **c** Proposed catalytic mechanism confirmed in silica to be energetically favourable between D262, glycosyltransferase and R101 as represented by Chemdraw

Systematic mutagenesis of the *E.coli* RodA conserved cavities *in vitro* using the RodA-PBP2 construct (Table 2.6) revealed changes in polymerisation, specifically caused by the arginines of the flexible loop K97 to R109, the conserved W102 and other residues important to the binding as in Figure 2.9.3 such as K117, however full catalytic loss activity in correctly folded protein was only through mutagenesis of D262 and P257, with neighbouring amino acids within the cleft having little effect *in vitro*. This is suggestive, as has been suggested in RodA's paralogue FtsW(127) of a central D262 responsible for catalysis, which by our data and new evidence is supported by phosphate binding residues common to other GT-C proteins which may also act as potential electron acceptors. This new polymerisation mechanism positioned between two lipid II binding cavities, and supported by previous studies on its paralogue *in vivo* in FtsW of the importance of D262 however with no mechanism suggested, or identification of cavities is indicative of a single responsible D262 across species for electron donation and the addition of new lipid IIs at cavity B to the polymer held in cavity A.

Interrogation of the adjacent prolines to the D262 location as in our *E.coli* structure (Figure 2.9.1b), and concurrent with data in Figure 2.7.3 on the flexibility of the RodA molecule at its second and first cavities, is suggestive of a conserved lipid II channel through the centre of RodA which facilitates its processivity, with a chemical description of the reaction given in Figure 2.9.1c.

2.9.5 Full mechanism of RodA-PBP2 chain elongation in concert with PBP2

After *in vitro* data had shown flexibility of the inner cleft of RodA (Figure 2.7.3), in addition to two sites of Lipid II binding and an energetically favourable catalysis by D262 (Figure 2.9.0 and

Figure 2.9.1), a mechanism of movement was proposed by myself and others (Figure 2.9.2), which included PBP2, and acknowledges that the reaction elongates an existing lipid bound polymer onto a new lipid carrier donated by a new lipid II each reaction as determined by a previous study (128). It is shown to be energetically favourable through Phil Stansfeld's simulations that lipid II undecaprenyl migrates through the molecule's conserved cleft from Site B to Site A as polymerisation occurs at the D262 electron donator (Figure 2.9.2c). This then releases a Und-PP molecule and creates a Lipid IV, capable of the same glycosyltransferase reaction from its Und-PP base to a new glycan of introduced lipid II at cavity B. This facilitates a fixed chain of sugars that extends through a groove of PBP2 (Figure 2.9.2a), which extends to PBP2's transpeptidation site Ser330(129) to attach the nascent chain into existing peptidoglycan.

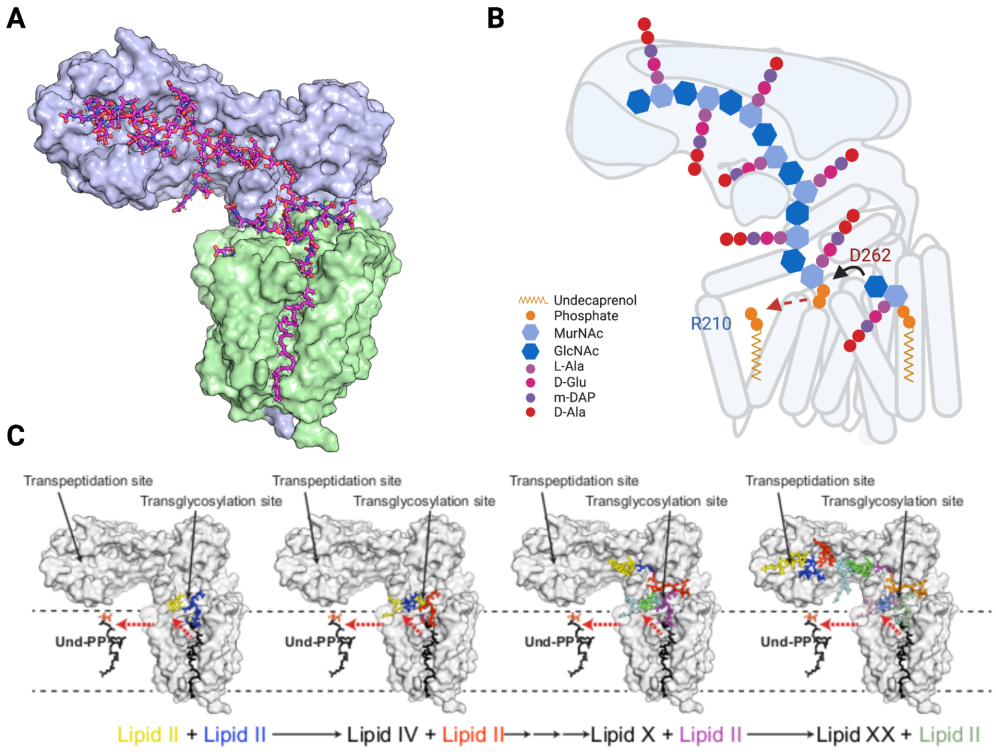


Figure 2.9.2. Mechanism of poIymerisation within the RodA-PBP2 complex

Collaboration with Philip Stansfield **A** the PBP2 molecule has a groove through which newly formed lipid II can be polymerised. **B** Cartoon representation of lipid II substrate, and PG product formation. **C** Extension of the glycan sugar chain over time.

Chapter 2: Conclusion

In collaboration, and through a new the Cryo-electron density map of the *E.coli* RodA-PBP2 complex provided by Rie Nygaard, I initially determined the structure of *E. coli* RodA-PBP2 by protein modelling, molecular dynamics and flexible map fitting to a cryo-EM map, giving us initially a 4.7Å structure of the entire complex. This new structure was verified by another collaborator Owen Vickery, and showed to be stable in the regions important for catalysis, with only extreme regions being flexible. As a consortium we then improved this structure to 3.2Å after use of CryoSpark.

Structural comparison with other proteins such as Waal O antigen ligase and the GT-C motif containing proteins which I published on, share an evolutionary past with RodA. This evolutionary past offered insights into the RodA mechanism of action. This revealed a conserved and *in vitro* essential residue, R210 that I hypothesised to coordinate the phosphate of lipid II, as too have conservation and co-evolution investigations of the complex. This data and link, suggested a conserved cleft across the periplasmic facing part of RodA and that this membrane embedded face of RodA is responsible for lipid II polymerisation, which extends a chain that then reaches the transpeptidase site of PBP2.

I managed to purify RodA-PBP2 in the UK in detergent and show it is active, in varying conditions including metal chelating conditions, suggesting metals are not required for synthesis and enabling further enzymatic interrogation. I also found that for the protein to function, it must have at least a membrane helix of PBP2 interacting with it, and many other residues in PBP2, even when the complex is attached by a loop, to ensure interaction, are shown to be essential.

Using *in silico* docking and site directed mutagenesis of the protein, I have predicted two binding sites for RodA and identified the most essential residues. This laid a firm theory for Lipid II polymerisation by RodA and PBP2 and the residues which carry out this catalysis. The reaction takes place using an electron donating D262 residues connecting lipid II molecules and later substrates, through a chain extension at the phosphate group. The phosphates involved of lipid II are stabilised by a series of

arginines, and we worked in collaboration with the University of Virginia and Meagan Dulphrisne to indicate the flexibility of the complex which allowed this to happen. This work is also applicable to the many members of the SEDS family of proteins which share an extremely close connection to RodA, indicated by the conserved catalytic residue and similarity in molecular shape. In addition this work indicates how peptidoglycan production and modification can be a multiprotein process, with RodA requiring PBP2 for function. In the future, it is likely antibiotics will be designed with this mechanism in mind, on a series of proteins which use the same catalytic process.

Chapter 3: Creation of a new technique to measure Lytic transglycosylase activity and purification of Lipid II

To be Published in "*E.coli Lytic transglycosylase recognition and substrate binding, assayed by fluorescent substrates*" , 2023 Gillet and Graham et al (Submitting to Access Microbiology)

Chapter 3: Summary

New assays are required to uncover the activity of many enzymes involved in peptidoglycan and cell envelope synthesis. These assays can then be built upon for new knowledge of the system and proteins that control cell growth. In part inspired by the polymerisation of fluorescent lipid II by the glycolsyltransferase activity of RodA-PBP2 as visualised by gel electrophoresis in Chapter 2, I realised that one could use the visualization of lipid II polymerisation to study the lytic transglycosylases (LTs) including what products they create (in terms of units of glycan sugar with dansylated mDAP or Lysine produced). This was successful in an initial attempt with the

hypothetical E.*coli* Lytic transglycosylase RlpA, after which I purified five other lytic transglycosylases and confirmed their activity in MltA, B, C, E and Slt70. This created the first specific assay for Lytic transglycosylase activity. Throughout these experiments I used Hen (*Gallus gallus domesticus)* egg white Lysozyme as a control to visualise LT activity, which indicates that the assay could be used to study any enzyme capable of breaking the glycosidic bonds which forms the backbone of the peptidoglycan polymer in both the lysozyme or lytic transglycosylase domain families. This assay has allowed for rudimentary inhibition testing of essential peptidoglycan modifying enzymes with the known lytic transglycosylase inhibitor "Bulgecin A" for the first time, as well as exolytic or endolytic activity investigation. The work has become the focus of a PhD for another student, but will also be disseminated into a paper based on the assay and data presented here. This chapter also includes data on a collaborator's own integral and highly expressed LT from another species rSlt70, which was part of confirming their activities *in vitro* for the first time. I also tested how the activity of these LTs changed under the influence of RodA-PBP2 presence to look for interactions. This work has been written into a grant and iCASE studentship of my fellow PhD student and previous masters student Francessca Gillet and will contribute to a future paper. In order to facilitate testing of these LTs with native *E.coli* lipid II, I also purified mDAP dansylated lipid II which is detailed as part of this chapter.

Chapter 3: Outputs

1. Synthesis of mDAP Dansylated Lipid II
2. Creation of new LT constructs
3. Purification of Lytic transglycosylases
4. Creation of Reverse Schägger gel assay
5. Inhibition testing of LTs by Bulgecin A
6. Co-incubation of RodA-PBP2 with LTs
7. Writing of paper on new reverse Schägger gel assay.

Chapter 3: Abstract

The growth and division of bacterial cells relies on the action of lytic transglycosylases to break apart strands of polymerised lipid II substrate peptidoglycan. A class of antibiotic adjuvant Bulgecin has been shown to be effective in stalling division, and has also been shown binding

efficiency to *E.coli* LTs. Catalytic investigation of lytic transglycosylases has been limited to mass spectrometry and *in vivo* knockout studies, however no enzymatic studies have been performed due to lack of available "clean substrate" with clearly defined product. Here we altered an assay previously used for lipid II polymerisation studies in transglycosylases, and used it's high molecular weight fluorescent product as a substrate to be visualised by images obtained through gel electrophoresis. I found that the Mlts have different binding capabilities and specificities, postulated LT RlpA has lytic activity, and some LTs show peptide peptidoglycan stem preference, and show that Bulgecin A does not inhibit all LTs, but infact inhibits only MltC *in vitro.* This method was then finally compared to an existing soluble fluorescence technique on peptidoglycan labelled with a lytic transglycosylase releasable dye.

Chapter 3: Background

In Chapter 1 we established the network of interactions and processes required to make peptidoglycan, then in Chapter 2, we discovered how peptidoglycan is polymerised from lipid II during elongation by the SEDS protein complex RodA-PBP2. However during synthesis of peptidoglycan, a lytic transglycosylase, an enzyme which breaks apart the glycosyl bonds between sugars in peptidoglycan is likely required to allow room for new peptidoglycan synthesis (Chapter 1). Therefore Slt35 (MltB), and RlpA which both share high levels of co-evolution with PBP2 and RodA, were initially cloned and expressed for investigation and then their binding assayed using simple his-tag binding to the complex (Gillet, MBio Thesis Warwick 2019) to interrogate this, which resulted in little publishable information, but hinted towards an MltB-(soluble) PBP2 interaction. However, after initial experiments on this coupling, it was realised a more sensitive assay for lytic transglycosylase activity was needed to assay how PBP2 and RodA may in turn effect their activity and *vice versa*. In this chapter, we will focus on the activity of the Lytic transglycosylases in *E.coli*, and create an assay for their activity using *in vitro* substrate. This chapter first takes the reader through the lytic transglysocylase functions determined by bioinformatics, then goes into the biosynthesis of the fluorescent lipid II polymer used in the assay, and then conducts a series of experiments to determine the information this new assay can give us.

Lytic Transglycosylase action

The action of lytic transglycosylases is that of breaking the glycosydic bond between the MurNac and GlucNac residues. This action is part of many other hydrolase actions (Figure 3.1.0), therefore occurs in the presence of different substrates, either as part of a peptidoglycan shell, or on glycan strand substrates of altered peptide stem and acetylation states. These altered peptide stems of peptidoglycan in bacteria, create a host of potential substrates and create a clear evolutionary reason for a range of active site types (130–137)and enzymes to cut peptidoglycan between the GlucNAc-MurNac sugars during growth, which have been studied in *E.coli* enzymes before, however not in specific assay such as the one presented in this chapter. The potential for an assay which visualises the size of product and substrate allows for the experimenter to differentiate product size, and point directly to lytic glycosyltransferase as the mechanism of action. Never before has Lytic transglycosylase specific activity been tested *in vitro* directly on a purified uniform nascent PG substrate. Their substrate preference, and product formation are yet to be uncovered fully, and no assay exists for easy lytic translglycosylase discovery and drug testing.

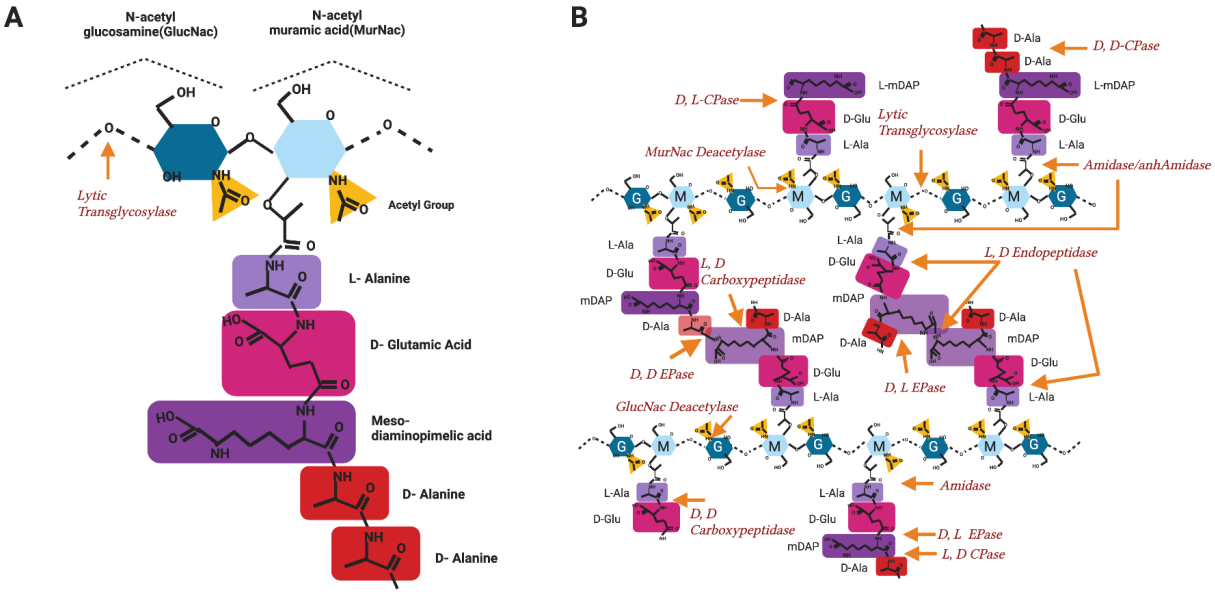


Figure 3.0.0 Lytic Transglycosylase action

A. Action of lytic transglycosylase on Gram-negative nascent peptidoglycan, **B**. Hydrolase activity on Gram-negative peptidoglycan

As mentioned the lytic transglycosylases and lysozyme enzymes are responsible for the breaking of the glycosyl bond between N-acetyl glucosamine(GlucNac) as well as N-acetyl muramic acid and rarely their de-aminated derivatives (Figure 3.0.0), with altered activity dependent on the peptide crosslinking and size of these chains.

There are eight known lytic transglycosylases in *E.coli* ; Slt, MltA, MltB, MltC, MltD, MltE (EmtA), MltF and MltG (YceG) in addition to one hypothetical LT RlpA (Figure 3.0.0/3.0.1). The cataloguing, expression and testing of these enzymes would enable the design of a new protocol for assays, and a screen for LT inhibitors across the families of protein, as well as helping determine their role in the cell.

The current literature and genetics of these proteins and genes which encode them, suggest an interactive network of function and shared interaction as shown in Figure 3.0.1. The catalytic action of these proteins differs between proteins, and family.

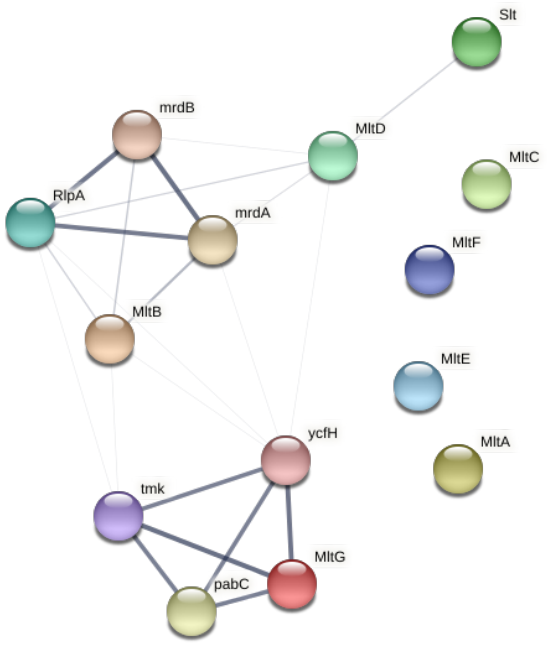


Figure 3.0.1 STRING interaction network of the lytic transglycosylases

Interaction score set to above 0.1 included, with gene co-occurance and textmining removed to remove conservation noise associated with all LTs and reviews. https://version-11-5.string-db.org/cgi/network?networkId=bdf2NgCrpj7i

Roles and interactions of the Lytic Transglycosylases

As mentioned, the lytic transglycosylases were originally hypothesised as a series proteins with lytic activity required to form the peptidoglycan cell wall and antagonise cell wall synthesis. but their catalytic activity was later confirmed by enzymes purified from the cell envelope on an array of DEAE beads charged with radioactive peptidoglycan substrates prepared from *E.coli,* effectively forming di and tri-peptide attached disaccharides (MurNac-GlucNac) in their action. This was different to lysozyme activity, due to its production of a 'no reducing end group and that the muramic acid residue possesses an internal 1 leads to 6 anhydro linkage.'(138). When the enzymes were assayed and crystallised, further members were later discovered.

There is a diversity of roles and structures among the Lytic transglycosylases to investigate. The domains, and active site of each of these proteins is summarised below (Figure 3.0.2). This also sorted by protein type, with soluble proteins such as Slt70, inner membrane bound (MltF) proteins and outer membrane palymitoylated proteins such as MltA.

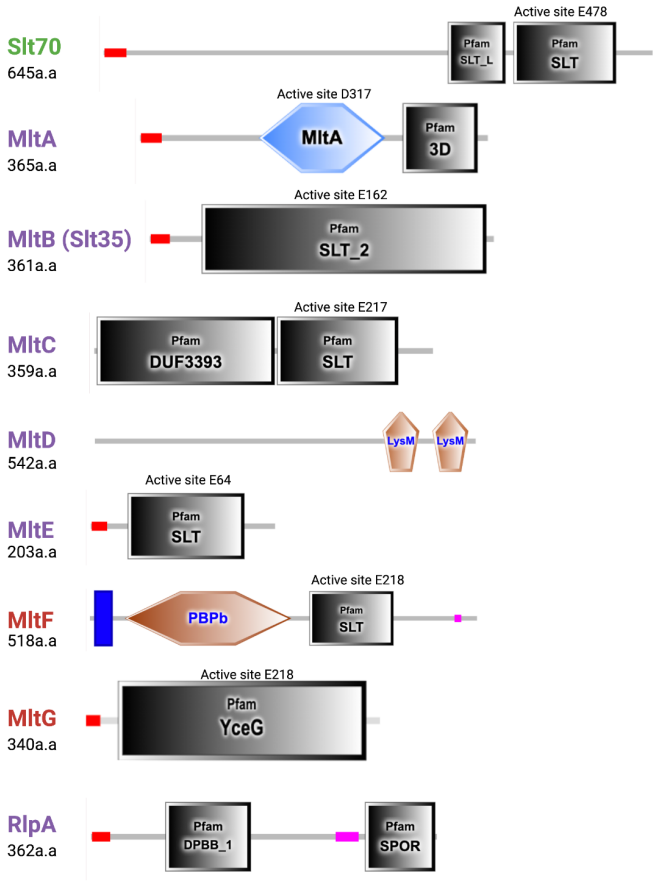


Figure 3.0.2 Lytic transglycosylases of *E.coli* and distribution across other organisms

Domain architecture of the known Lytic transglycosylases (LTs) in *E.coli*, Green= soluble protein, Purple = palymitoylated proteins, Red = Inner membrane bound proteins,

In *E.coli* there are many hypothesised roles of the Lytic transglycosylases. Slt70, MltA, MltB, MltE, MltG, MltF and RlpA all contain a signal peptide sequence identified by Interpro/PFAM(139), confirming their periplasmic localisation. The signal peptides and palymitolation points of each protein have been shown shown (Fig 3.0.2). Through domain searching similarity between proteins active sites similarity can be observed, and therefore drug target cross compatibility can be hypothesised between the Slt domain containing proteins Slt70, MltC, MltE and MltF, therefore when screening drugs later in the thesis this should be taken into consideration. In addition to a series of roles and differences which go beyond redundancy evidence by altered domain architecture, suggesting these lytic roles are paralogues of one another.

Soluble Lytic Transglycosylases

Lytic transglycosylases, when under no control should be dangerous to the cell, as they can break through the murein cell wall and leave vulnerabilities and cell envelope irregularities that may lead to rupture or loss of a cells morphological and stressor adaptions to the environment and therefore cell death. The same is true for a cell without lytic transglycosylases, as their removal means there is no way to insert new peptidoglycan. The solubility of these proteins therefore, and potential freedom to act away from directly cytoplasmically controlled partners could be a disadvantage. In *E.coli* there is only one 'true' soluble lytic transglycosylase in the periplasm, Slt70.

Slt - Slt is a 70kD protein, first of the identified lytic transglycosylases with endolytic and exolytic activity. It has been purified bound to substrate and its exact role is still unknown. However it shares genetic co-evolved connections to MltD. It also is one of the most conserved LTs across Gram- negative bacteria, therefore likely very signficant in survival Its halo like shape suggests it may interact with another protein in inactive form when not in use, similar to hypothesised EnvC, NlpD interactions proposed of the amidases. Its active site is a Glutamate at residue 478 in *E.coli.* (140) During this chapter, we assay rSlt an enzyme which when knocked out in a strain of potential bioweapon potential, reduces or stops pathogenicity.

Lipoprotein Lytic Transglycosylases (LTs)

As discussed in Chapter 1, many proteins that interact with peptidoglycan and/or bind proteins which alter peptidoglycan are attached to the outer membrane. The lytic transglycosylases are no exception, with the majority bound by a lipid, to the inner leaflet of the outer membrane, facilitating some movement when bound to mobile partners. The lipoprotein (outer membrane bound) LTs are MltA, B, C, D, E and RlpA

mltA -MltA has no predicted interactions on string, or closely conserved genetic neighbourhood partners. Therefore its exact role is not yet known. It has been implicated in recycling of murein peptides during cell elongation and division. Although not noted in the STRING profile in Figure 3.0.1 due to its interaction cut-off of 0.7. proteomic studies [Chapter 1 Figure 1.4] have

shown interactions with MipA(68,130) and then indirectly MrcB. This would implicate MltA by proximity in a number of roles associated with *mrcB*/PBP1b such as elongation and division. It has been noted that MltA performs exolytic cleavage. Its optimum pH is between 4.0 and 4.5, with reduced activity above 30 degrees. Its active site is formed of an Aspartate residue 317. (68,130)

mltB - MltB has been shown to interact with PBP2/MrdA in our STRING interaction network (Figure 3.0.1) (132,137,141) A calcium binding interaction has also shown to be implicated in *Pseudomonas sp.* PBP2 interactions , as well as our own experiments (MBio thesis Fran Gillet). In *E.coli* its active site is formed of a Glutamate at residue 162 (Figure 3.0.2). As well as being bound to the inner leaflet of outer membrane, it also has a soluble form, which is formed as a cleavage product of its full length derivative, known as Slt35.

mltC - MltC has no known partner but in proteomic studies may interact with NlpI, (131). In *E.coli* its active site centred at Glutamate 217 (Figure 3.0.2).

mltD - MltD has been shown to be genetically linked to the lytic transglycosylases Slt and RlpA but also the transglycosylase and transpeptidase of elongation RodA as well as PBP2 in Figure 3.0.1, this may implicate MltD as part of the elongation complex. (142)

mltE/emtA- MltE has no known interaction partners, however its activity has been assayed. It cleaves preferentially in a partially endolytic manner, cleaving at least four glycan sugars or two disaccharide MurNac GlucNac units from the end of the glycan chain, preferring crosslinked PG. Its active site is at Glutamate 64 in *E.coli* (Figure 3.0.2). (133)

rlpA - RlpA localises to the septal ring, with SPOR domains known to associate with denuded peptidoglycan, a peptide strand cleaved form of the nascent peptidoglycan found most often at cell division sites. This protein has a strong preference for naked glycan strands that lack stem peptide (143) It is often genetically fused to RodA and PBP2, as a neighbour. Its role is uncertain still.

Inner membrane bound LT’s

Some LTs likewise are associated with inner membrane, therefore may be directly positioned by cytoskeletal machineries. MltF and MltG are part of this class of LTs.

mltF - MltF is associated with flagella lytic action, this could be to create space during the assembly of toxin transport machinery and flagella. There are no known interaction partners, it contains a PBPb like domain. (144) Its preferred substrate action, and endolytic or exolytic nature are unknown, but it is likely able to digest high complexity murein due to its role as a flagella associated LT. In *E.coli* its active site is E218 (Figure 3.0.2).

mltG- MltG is hypothesised to be as a peptidoglycan terminase that cleaves nascent peptidoglycan strands endolytically to terminate their elongation, therefore if investigated in the future could have high activity on our de novo peptidoglycan substrate. It has many interaction partners (Figure 3.0.1), as noted recently with an active site of Glutamate 218 in *E.coli* (Figure 3.0.2). (135,136,145)

Bulgecin, and antibiotic adjuvant and combination therapy

These lytic enzymes have been implicated as necessary to cell envelope formation since their discovery. However in the 1980s, the drug Bulgecin A (146)(Figure 3.0.3B), was shown to cause bulges in Gram-negative bacteria (Figure 3.0.3A) upon division. Later co-crystals of bulgecin were formed in complex with the LTs(17), and thus it is now known Bulgecin likely affects multiple of these enzymes in a cell lethal manner, (Figure 3.0.3 C/D) binding not only one protein, but multiple. However this is only one drug, and others can be developped. Thus to test new drugs or fragments which have been implicated to inhibit or designed to work with these enzymes a new assay would be

an asset for drug discovery.

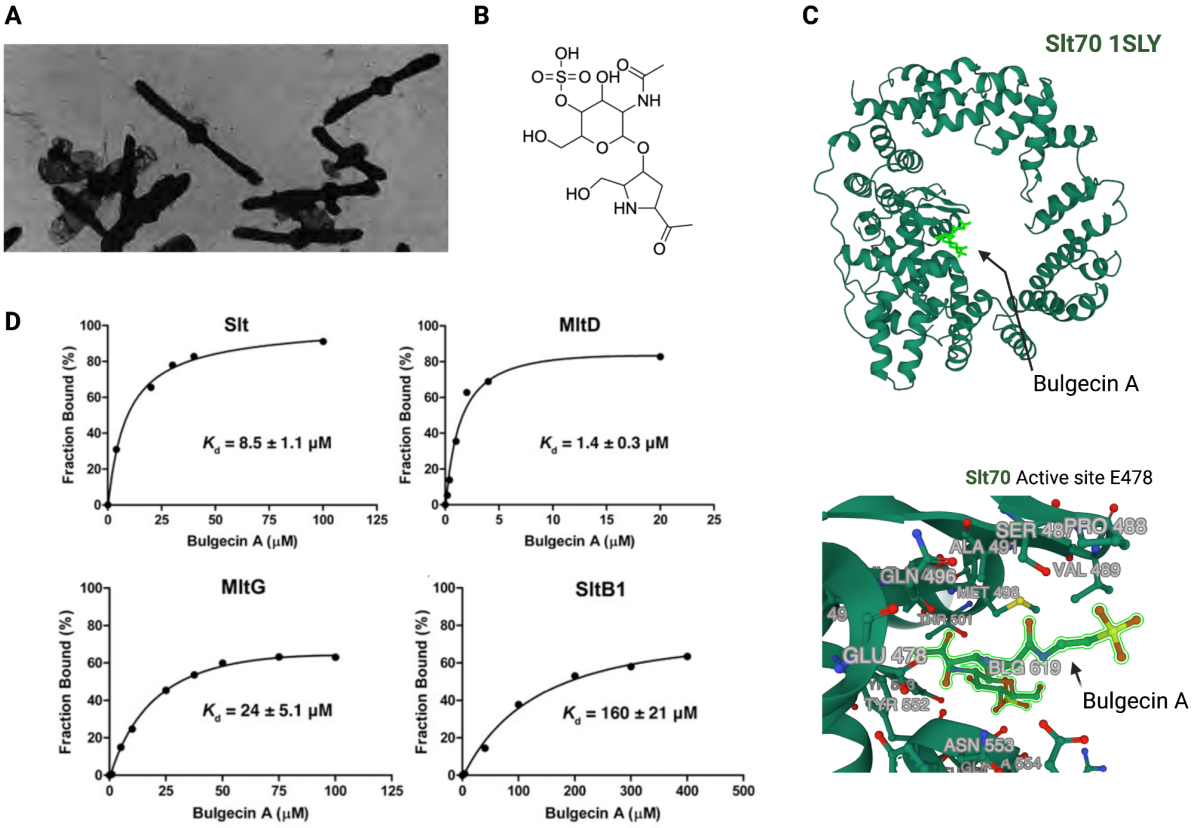


Figure 3.0.3 Bulgecin A as a Lytic transglycosylase inhibitor

A. First images of *E.coli* cells with high concentration of Bulgecin A, giving its namesake. (146) **B**. Structure of Bulgecin A **C**. Slt70 bound to Bulgecin A in its active site **D**. Bulgecin A binds *P. aeruginosa* Slt70, MltD, MltG (145)

As these enzymes have also been implicated in binding or associating with other essential crosslinking proteins such as the PBPs, which also play a part in peptidoglycan synthesis, a treatment of Bulgecin or another LT affecting drug, simultaneously with a transpeptidation or glycosyltransferase affecting drug could mean growth would be impossible, even in an uncontrolled manner as no compensatory resistance mechanisms could take place. This lends to the theory that the LTs as a target to be used in combination therapy. Currently fast methods which can resolve lytic transglycosylase activity, rely on change in fluorescence on a purified peptidoglycan substrate, made of a complex matrix of alternative side chain sizes and crosslinking, with changes in fluorescence possible from a range of enzymes from carboxypeptidases, to amidases, and lytic transglycosylases, this has meant there is room for a new assay in lytic transglycosylase activity resolution. Discovery

and confirmation of lytic transglycosylases and lysozyme enzymes could also be resolved this way, with prefered substrates determinable through altered substrate presentation.

In the future, if the assay produced in this chapter is successful, we can also use alternative enzymes which have only glycosyltransferase ability or alter the peptidoglycan with carboxypeptidases, thus allowing a pure substrate of desired peptide side chain or crosslinking to be assayed in Lytic transglycosylase activity, in an environment which accepts most inhibitors, without need for *in vivo* trials, and that must affect specific LT activity. In addition to this, inhibitors can be assayed in the same assay for action against transglycosylases.

However in order to create this assay, we have to form a pure lipid II substrate, for use by our enzymes in peptidoglycan synthesis. Our laboratory produced *in vitro* substrate.

Lipid II and peptidoglycan synthesis

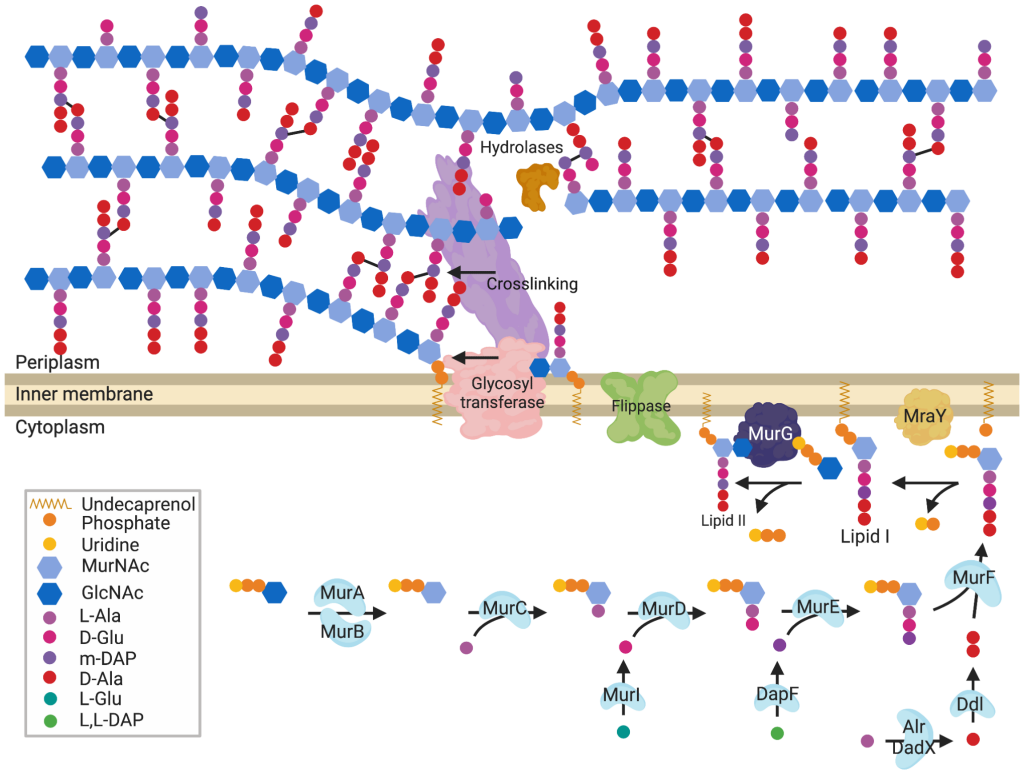


Figure 3.0.4 Synthesis of Lipid II, and Peptidoglycan

In order to create the peptidoglycan which is then suitable for polymerisation in our assays and then lytic glycosyltransferase, the substrate lipid II is required. Lipid II is first synthesised in the cytoplasm by a series of ligases, which convert a common glycolysis product Fructose-6-phosphate by four successive enzyme activities to uridine 5'-diphosphate-N-acetylglucosamine (UDP-GlcNAc). This is catalysed by GlmS, GlmM and GlmU (bifunctional enzyme) (140). UDP-MurNAc (5'-diphosphate-N-acetylmuramic acid) is formed from UDP-GlcNAc using Mur ligases MurA and MurB. This results in a sugar moiety ready for pentapeptide addition. A pentapeptide stem is then appended to the D-lactoyl carboxyl group of UDP-MurNAc by sequential addition of peptides by MurC-F: L-Ala (MurC), D-glutamic acid (MurD), meso-diaminopimelic acid (m-DAP) (MurE), dipeptide D-Ala-D-Ala (MurF), with D-Glu and m-DAP being synthesised from their L- or L,L- stereoisomers by MurI and DapF respectively, and D-ala-D-Ala being produced from L-Ala by alanine racemases and D-Ala-D-Ala ligase (140, 20). (Figure 3.0.4)

The UDP-MurNAc 5P produced by these reactions is then transferred to an undecaprenol, a membrane spanning lipid, yielding undecaprenyl diphospho MurNAc 5P (Lipid I), in a reaction catalysed by MraY at the inner membrane. Thereafter, MurG transfers a GlcNAc sugar moiety from UDP-GlcNAc to Lipid I, producing undecaprenyl diphospho MurNAc GlcNAc 5P (Lipid II) which is polymerised by the Type A PBPs and SEDS proteins for nascent chain peptidoglycan formation. It is this product when mature, after potentially other modifications such as 3-3 and 4-3 crosslinking which is acted on by hydrolytic enzymes such as the lytic transglycosylases, it is this action we will assay in this chapter. (Figure 3.0.4)

Chapter 3: Methods

Expression and purification of soluble proteins

Expressions of proteins were performed using pET plasmids (pET-47b) (147) transformed into BL21 star cells in 800ml autoinduction 2xYT Broth, inoculated by 10ml overnight cultures grown at 37°C 160rpm in their respective antibiotic and of the respective cell line. Induction of

the auto-induction broth took place over 6hrs at 37°C 180rpm up to OD 0.8, then cooled to 22°C and maintaining RPM to improve protein folding and incubated for a further 16hrs. Purification of proteins was completed by centrifugation of the above cultures into pellets at 6000g for 15mins, often using a floor standing centrifuge with 1L buckets. The pellets were resuspended in Soluble protein Lysis buffer [10% Glycerol, 10mM Imidazole, 50mM HEPES, 0.4M NaCl +protease inhibitor tablet +Dnase I pH 8.0] and sonicated at 45% capacity for 30 seconds for 6 cycles. After sonication, lysates were cleared of debris by centrifugation at 50,000g for 30mins and added to gravity flow columns containing 1ml of lysis buffer normalised nickel primed agarose beads. After a passage, the beads were washed using 20x Column volumes Soluble protein wash buffer.1 [10% Glycerol, 10mM Imidazole, 50mM HEPES, 0.4M NaCl pH 8.0] and 10x Column volumes Soluble protein wash buffer.2 [10% Glycerol, 50mM Imidazole, 50mM HEPES, 0.4M NaCl pH 8.0] before elution using 1x Column volume Soluble protein elution buffer. [10% Glycerol, 400mM Imidazole, 50mM HEPES, 0.4M NaCl pH 8.0]. The resultant proteins were size eluted by gel sepharose beads at various quantities using an AKTA if necessary, or otherwise dialysed into soluble protein storage buffer [10% Glycerol, 50mM HEPES, 0.4M NaCl +protease inhibitor tablet pH 8.0]

Expression and purification of hydrophobic proteins

Expression of hydrophobic, but not membrane bound proteins such as Slt70 was performed using a pET48a plasmid carrier plasmid transformed into Tuner cells expressed in 6x 800ml 2xYT Broth. The broth was inoculated by 60ml overnight cultures grown at 37C 160RPM in kanamycin 50mM antibiotic. Induction took place after the culture reached first reached OD 0.8 then was cooled by ice bath and re-incubated at 22°C. The culture was induced using 1ml 1M IPTG per 800ml broth and incubated for a further 16hrs. Purification of Slt70 protein was completed by centrifugation of the above cultures into pellets at 6000g for 15mins, using a floor standing centrifuge. The pellets were resuspended in 10% Glycerol, 10mM Imidazole, 50mM HEPES, 0.4M NaCl , 0.1% Triton X100 +protease inhibitor tablet +Dnase I pH 8.0 and sonicated at 45% capacity for 30 seconds for 6 cycles. After sonication, lysates were cleared of debris by centrifugation at 50,000g for 30 minutes and added to gravity flow columns containing 1ml of lysis buffer normalised Nickel primed Thermo Fischer agarose beads. After binding, the beads were washed using 20x Column volumes of 10% Glycerol, 10mM Imidazole, 50mM HEPES, 0.4M NaCl , 0.1% Triton X100 pH 8.0 and 10x Column volumes 10% Glycerol, 50mM Imidazole, 50mM HEPES, 0.4M NaCl , 0.1% Triton X100 pH 8.0 before elution using 1x Column volume 10% Glycerol, 400mM Imidazole, 50mM HEPES, 0.4M NaCl , 0.1% Triton

X100 pH 8.0. The resultant protein was not further purified by gel sepharose beads using an AKTA as it stuck to the sepharose and instead was dialysed into storage buffer (10% Glycerol, 50mM HEPES, 0.4M NaCl , 0.1% Triton X100 +protease inhibitor tablet pH 8.0) overnight before storage at -80°C.

SaMGT purification

SaMGT was transformed into BL21 PLysS *E. coli* strain, a 10 mL starter culture (LB 35ug/ml Chloramphenicol, 100ug/ml Kanamycin) made using a freshly transformed colony was grown overnight at 37°C. Mature starter cultures were used to innoculate 800ml 2XYT AutoInduction media and grown at 37˚C to reach an OD600 between 0.8 and 1.0 at 160-180RPM, the temperature was reduced to 22˚C and incubated at 22˚C overnight 215-220 RPM. Cells were pelleted at 4000g for 10 min at 4˚C. Cell pellets were resuspended using Cell Lysis buffer [20 mM HEPES, pH 7, 200 mM NaCl, 10mg/mL solid DNase, 4mL/100mL RNase, 1:1000 Complete cocktail of protease inhibitors, 1mM TCEP, 1mM PMSF, 20% Glycerol] and Lysed by cell disruption at multiple passes to ensure 100% cell lysis. Lysate was centrifuged using an ultracentrifuge 80,000G for 35 minutes at 4˚C and the membrane pellets kept. Membrane pellets were washed by resuspension in high salt buffer [20 mM HEPES, pH 7, 500 mM NaCl,10mg/mL solid DNase, 4mL/100mL RNase, 1 mM TCEP, 1:1000 Complete cocktail of protease inhibitors, 1mM PMSF, 20% Glycerol] and span again at 80,000G for 35mins and the pellet kept. [Pause step possible by freezing of membrane]. Salt washed pellets were resuspended using high salt buffer and homogenisation, then allowed to solubilise in 1% DDM by addition of 20% DDM proportionally for 2hrs. Solubilised membranes were then spun at 80,000G for 35 minutes and supernatant collected. Supernatant was then added to equilibrated Ni-NTA resin for binding for 2hrs, after being adjusted to 40mM Imidazole final concentration. After binding, the solute flow through was collected by gravity column leaving protein enriched resin. The protein enriched resin was washed using 10CV of Wash buffer [20 mM HEPES, pH 7, 500 mM NaCl, 60 mM Imidazole, pH 7, 20% glycerol, 0.1% DDM] and then eluted into 0.5CV using elution buffer [20 mM HEPES, pH 7, 500 mM NaCl, 300 mM Imidazole, pH 7, 20% glycerol, 0.1% DDM]. The resultant elution was resolved by SDS-page electrophoresis.

Mur ligase synthesis

Plasmid constructs containing Mur ligases of *P. aeruginosa*, required to form the pentapeptide subunit on N-acetyl muramic acid were used to express his tagged proteins MurA, MurB, MurC, MurE, MurF and MurG. His-tagged proteins were expressed in the same manner as all other soluble protein constructs in this thesis, via Auto-induction media for lactose concentration increase and protein expression. The high protein concentration cells were then harvested, span at 5,000g to pellet, and lysed by sonication in Lysis Buffer, Lysate was cleared using a 50,000g spin for 30 mins and flowed through a pre-equilibrated Nickel histrap column, then washed using Soluble protein Wash buffer A, Soluble protein Wash Buffer B and eluted with Soluble protein elution buffer. Mur ligases were verified for efficacy using their respective ligase assays.

mDAP pentapeptide synthesis

A “one pot” reaction mixture of 8mls [50mM PEP, 50mM HEPES, 10mM MgCl2 1mM DTT, 50mM KCl, MurA 0.21mg/ml, MurB 1.24mg/ml, MurC 0.24mg/.ml, MurD 0.17mg/ml, MurE(DAP) 0.36mg/ml, MurF (DAP), Pyruvate Kinase 5.53u/mg, IDH 1.48u/ml DAP 80mM, UDP-GlcNac 8.22mM, NADP 0.2mM, DL-Isocitrate 26mM, ATP 6mM, L-Ala 35mM, D-Glu 35mM, D-Ala-D-Ala 35mM pH 7.6] was incubated for 24hrs at 37°C.

mDAP pentapeptide purification

The one pot mDAP pentapeptide was filtered using a Vivaspin 10kD concentrator to remove protein by centrifugation and collection of flowthrough. The pentapeptide was purified using FPLC using a Source 30Q column and exchange of 10mM Ammonium acetate pH 7.6 and 1M Ammonium Acetate and monitored at 280nm and 254nm. 90 minute gradient 10mM-1M at 10ml/min. A peak displaying a 254:280nm absorbance of 1:2 was collected and freeze dried until all ammonium acetate was removed by successive dilution/freeze drying cycles.

mDAP pentapeptide purity assay

Concentration of pentapeptide found using 260nm and quartz cuvettes and a coefficient mM of 0.1, purity confirmed by DacB release assay.

Dansylation of mDAP pentapeptide

Freeze dried and purified mDAP pentapeptide was re-purified using a Q sepharose column and Sodium Carbonate buffer in preparation for dansylation (Buffer A) 10mM NaHCO3 and (Buffer

B) 500mM NaHCO3 at pH9 on a Q30 Sepharose column, and residual ammonia checked by Nesslers reagent assay after. When no ammonia was present, the fractions were assayed for concentration at 260nm and pooled. Pooled fractions were reacted with acetone dissolved Dansyl Chloride at a 42:1 Dansyl Chloride to mDAP pentapeptide ratio in a glass container overnight in the absence of light and neautralised with 2M Tris-HCl. The resulting mixture was rotorvapped, resuspended in 2ml H20 and separated by size fractionation in a superdex peptide 200 column in 0.1 M ammonium bicarbonate buffer. The first 340nm rich peak collected. This fraction was then purified as per mDAP pentapeptide purification.

Dansyl mDAP Lipid II synthesis of C35 and C55 variant

UDP-MurNac-DAP-Dansyl was reacted in a one pot Lipid II reaction mixture. monophosphate C55 and c35 were dried down by nitrogen gas, and resuspended in 5x Buffer (triton). The one pot lipid II reaction consisted of 2mM UDP-MurNac-DAP-Dansyl, 6mM UDP-GlcNac, 3.5mg/ml *M. Flavus* membranes (MraY rich), MurG 0.57mg/ml, 4mg Monophosphate C55 and c35 (3ml). This one pot mix is stopped by 3.5ml Pyridinium Acetate addition and 7ml n-Butanol addition and 10ml H20 addition. The lipid fraction appears as a yellow phase, for later purification

Dansyl mDAP Lipid II purification of C35 and C55 variants

Lipids are separated by affinity to DEAE sephacel beads. A gradient of Solvent A[35% Chloroform, 55% methanol, 10% H2O] and 1M Ammonium Bicarbonate. Lipid separation and fractions analysed by TLC and dansyl visualisation 340nm.

Dansyl mDAP Lipid II visualisation of C35 and C55 variants by TLC.

Lipid fractions of purification by DEAE affinity were analysed by TLC, using 88ml Chloroform, 48ml methanol, 7ml H20 3ml 30% Ammonia by Silica 60 TLC and Iodine exposure.

Gel based In vitro Lytic transglycosylase activity assay (Reverse Schägger Gel)

The reverse Schägger gel uses a similar gel based system for small protein visualization as Schägger et al 1987 (148) but for the visualisation of lipid II polmers. Glycosyltransferase activity by RodA-PBP2 was shown using visualisation of fluorescently labelled Dansyl Lipid II molecules through a time point orientated gel based method. 1.5µl 2.7µM detergent solubilized SaMGT/RodAPBP2 or PBP1B protein [, 300mM Imidazole, 250mM NaCl, 20mM HEPES,

0.05% DDM, 20% Glycerol] was added to 13.5ul [10mM MgCl2, 100mM NaCl, 50mM HEPES, 20% DMSO, 0.03% LDAO,10μM Dansyl amidated Lysine or mDAP Lipid II]. The reaction mixture was incubated at 37° C for 1hr 30 minutes to allow for polymerisation. After this, the reaction was heated to 95° C to deactivate the polymerase, and cooled to 37°C, before the Lytic transglycosylase was added.

The resultant polymer was denatured at 95°C with addition of 5x loading dye [50mM TrisHCl pH 8.8, 4% SDS, 40% Glycerol, an amount of bromophenol blue, DTT 200mM] and ran on a Criterion 16.5% gel at 110V for 80 minutes with anode gel running buffer [0.1M Tris-HCl pH 8.8] and cathode gel running buffer [0.1M Tris-HCl pH 8.25, 0.1M Tricine, 0.1% SDS] and visualised by 10 second exposure to UV on gel viewing apparatus.

Fluorescein isothiocyanate (FITC)-Peptidoglycan degradation assay

As performed in by H Maeda et al 1980(149), *E.coli* peptidoglycan was isolated as previously described (150) and attached covalently to Fluorescein isothiocyanate. This peptidoglycan substrate, was then assayed for two hours, at temperatures dependent on assay (37-25°C) , in 100mM NaCl, 50mM HEPES 10% Glycerol, at alternative pH dependent on assay (7.5pH if not mentioned). The insoluble fluorescent substrates were then separated from lytically removed products, by filtration in a 96 well plate format, before comparison with a Lysozyme and no enzyme control.

Chapter 3: Strains, Primers and plasmids

Table 3.0 Strains, Plasmids and Primers used for Lytic transglycosylase reverse assay development

Name	Sequence Information		Notes
Primers	Primer 1 - 5’ to 3’	Primer 2- 5’ to 3’	
Slt70 linearisation	CCC GGG TAC CAG GAT CCG AAT TCT ATG TAT TTA CAC TTA GAG GAT GCG	CCG CAA GCT TGT CGA CGG ACG TCG TCA GTA ACG ACG TCC CCA TTC	

pET47b linearisation	AGA ATT CGG ATC CTG GTA CCC GGG	CGA CGT CCG TCG ACA AGC TTG CGG	
Plasmids	Notes		Resistance
pET47b	Novagen		Kanamycin
E.coli MurA-pET28	Roper Lab plasmid		Kanamycin
P.*aeruginosa* MurB pMON 3006	Gift from Roger C Levesque lab		Kanamycin
P.*aeruginosa* MurC pMON 3004	Gift from Roger C Levesque lab		Kanamycin
P.*aeruginosa* MurD pMON 3013	Gift from Roger C Levesque lab		Ampicillin
P.*aeruginosa* MurE pMON 3014	Gift from Roger C Levesque lab		Ampicillin
P.*aeruginosa* MurF pMON 3009	Gift from Roger C Levesque lab		Kanamycin
E.coli MurG	Gift from Roger C Levesque lab		Ampicillin
E.coli sMltA	Gift from Mobashery Lab		Kanamycin
E.coli sMltB	Produced during PhD (with Fran Gillett)		Kanamycin
E.coli sMltC	Gift from Mobashery Lab		Kanamycin
Staphylococcus *aereus* *s*MTG	Roper lab plasmid		Kanamycin
E.coli sMltE	Gift from Mobashery Lab		Kanamycin
E.coli sRlpA	Produced during PhD (with Fran Gillett)		Kanamycin
E.coli Slt70	Produced during PhD		Kanamycin
E.coli sMreC – pet28b	Gift from Nic Briggs		Kanamycin
Strains/Genomes	Notes		

Escherichia *.coli* genome W3110	Used as a template for the multi-purpose gibson primers and cloning of LT genes. F^- lambda$^-$ IN(rrnD-rrnE)1 rph-1
E.coli BL21 DE3- Tuner cells	Novagen, F^- *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm lacY1*(DE3)
E. coli DH5a cells	Invitrogen *fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
E.coli BL21(DE3)Star	Thermofischer F$^-$*omp*T *hsd*S$_B$ (r_B^-, m_B^-) *galdcmrne*131 (DE3)
Micrococcus Flavus	Used for production of MraY/MurG containing membranes in lipid II synthesis

Chapter 3: Results

Before testing a set of lytic transglycosylases with our idea of a reversed Schägger gel, we required more lipid II mDAP dansyl substrate. Therefore, the initial part of this chapter covers its synthesis and confirmation. This process required the use of an *in vitro* system of Gram negative species derived Mur ligase enzymes, and substrate to create milligram quantities of dansylated fluorescent lipid suitable for visualisation and size separation post glycosyltransferase and lytic action on a polyacrylamide gel by excitation and emission fluorescence. This substrate has been created before at David Ropers laboratory, however this synthesis was novel, as in addition to the C55 undecaprenyl, we also created a C35 length lipid used for Chapter 2, in attempts to have lipid II encapsulated at the active site of RodA. I created both dansylation (not fluorescent) unlabelled and (fluorescent) labelled mDAP lipid II. In section 3.3 of this Chapter we return to the bioinformatics and synthesis of the lytic transglycosylases assayed in our study, as well as the reverse Schägger assay itself.

3.1 Dansylated Lipid II was synthesised

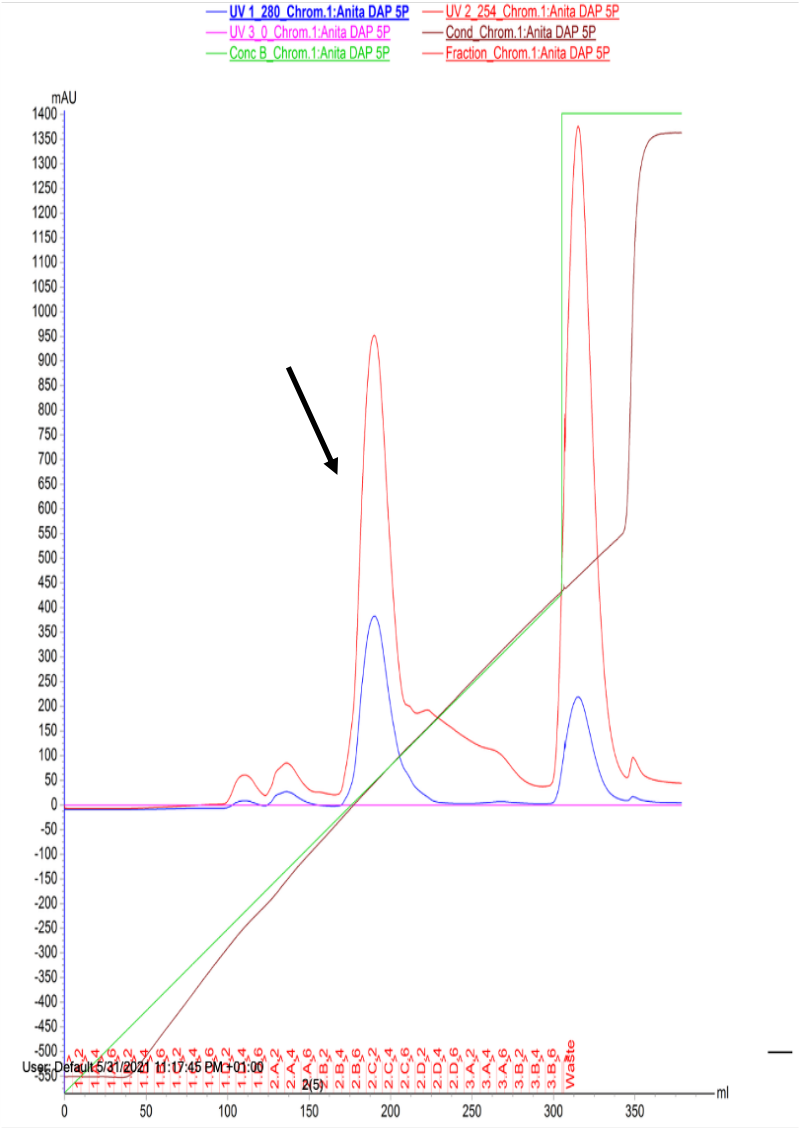


Figure 3.1.0 Purification of 5DAP purity Source Q30 Column

Y axis denoting absorbance intensity in AU, X axis denoting ml, Line colours correspond to those shown in key. Pentapeptide peak occurs between 150-225mls volume- Black arrow- collected fraction

As per the methods mentioned above, Mur ligases up to MurF were purified, and assembled in a “one pot” reaction of Mur ligases in addition to substrate. The resultant one pot reaction of products as shown (Figure 3.1.0) had a peak of expected value for pentapeptide, representing the enzymatic activities of MurA through to MurF (Figure 3.0.0), with a ⅓ ratio of 280nm to 254nm light absorption shown on the x-axis to cross roughly 200mls volume. This solution represented by the 280nm/254nm peak was collected at the heights of the peak between 150 and 225mls for further freeze drying and removal of ammonium acetate by which the purification was conducted. 40mls of soluble mDAP pentapeptide was isolated, at an unknown concentration. This was measured using its 260nm absorbance and extinction co-efficient to represent roughly 5mg.

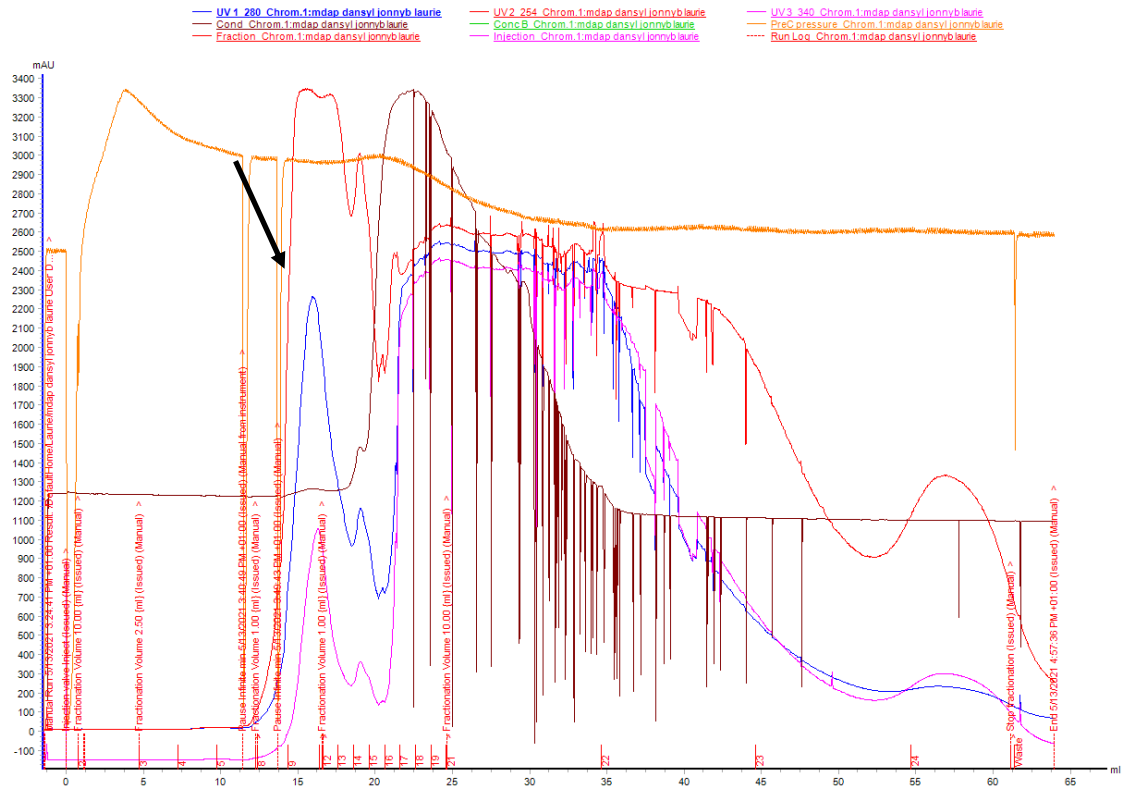


Figure 3.1.1 Purification of Dansyl 5DAP purity Superdex 300 peptide

Y axis denoting absorbance intensity in AU, X axis denoting ml, Line colours correspond to those shown in key. The initial, blue 280nm peak, with a 340nm peak within it, was isolated, between 12-17mls collected. A concatenate of products were derived after this which were not collected. Black arrow-collected fraction

After the initial pentapeptide was created, a dansyl group was added, with high amounts of excess dansyl-chloride. This reaction creates a range of dansyl-tris adducts, and impurities, as well as leaving behind dansyl-chloride, which show high 360nm absorbance peaks. The expected band for the mDAP pentapeptide was the first peak between 13 and 18mls as shown in Figure 3.1.1. This was collected, and freeze dried to remove excess ammonium bicarbonate, used during fractionation. The other peaks represented the range of adducts formed by dansylation were not useful to our work.

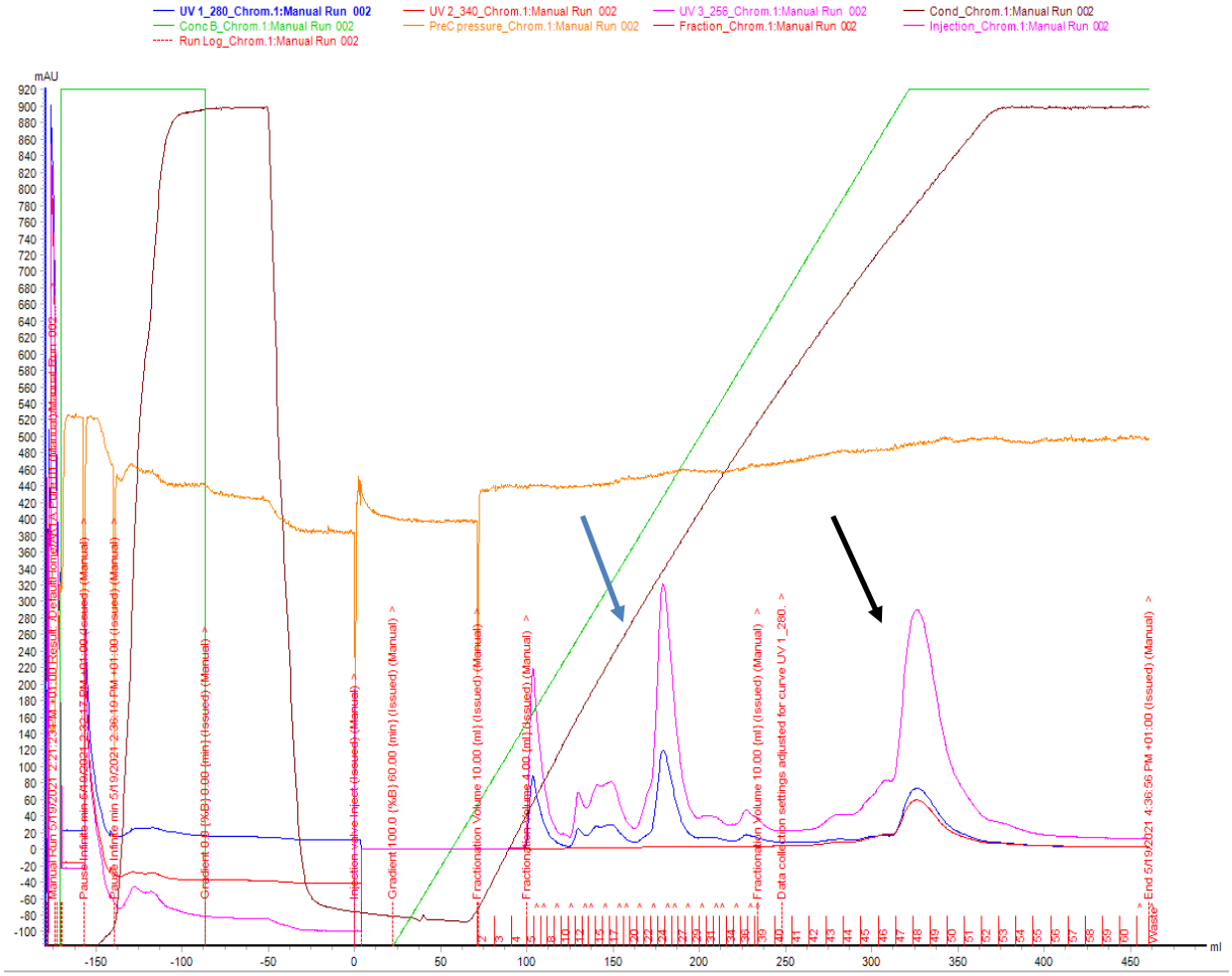


Figure 3.1.2 Confirmation of Dansyl 5DAP using Q30 Column.

Y axis denoting absorbance intensity in AU, X axis denoting ml, Line colours correspond to those shown in key. Peak 300-370mls was collected. Black arrow collected fraction. Blue arrow, undansylated 5DAP

The initial peak (Figure 3.1.1) was unresolvable in detail during the first instance as the column separated by size exclusion over a short gradient, therefore I analysed and repurified this fraction by Source MiliQ 30 fractionation to reveal a dansylated pentapeptide peak at 300-360mls, which showed dansyl fluorescence at 360nm in addition to expected ratios at 260nm and 180nm emission. I also collected unreacted lipid II with no peak at 360nm as well as other biproducts for future use and recycling.

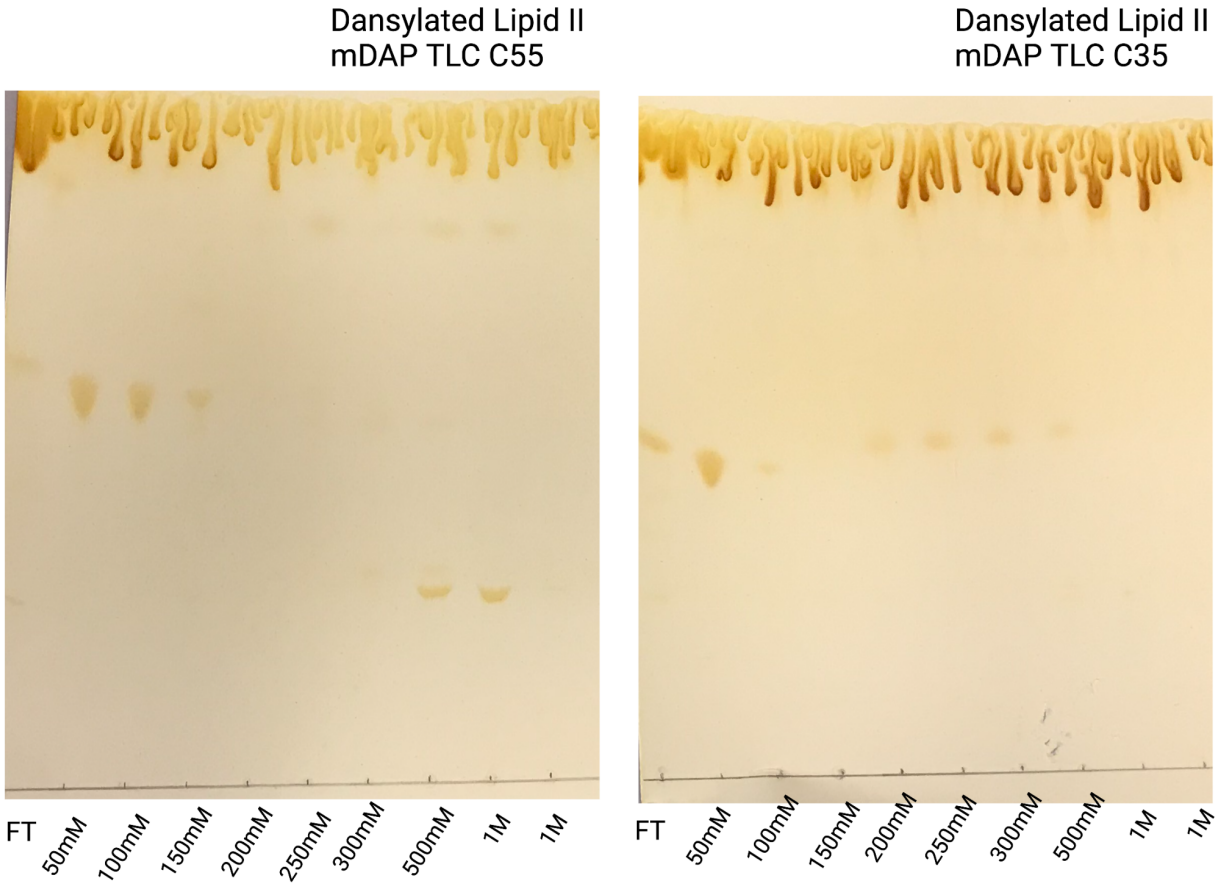


Figure 3.1.3 Thin layer chromatography of Lipid II lengths 55 and 35.

Y axis denoting position migrated in liquid phase, X axis denoting fraction in mM ammonium acetate used. FT denoting flow through of DEAE beads before Amonium acetate use.

The resultant dansyl-mDAP pentapeptide peak was then purified and freeze-dried over six cycles of water addition, freezing and sublimation until it was free of ammonium acetate, before being subject to a one pot membrane reaction for lipid II synthesis, with either C55 or c35 phosphate lipid tails (Undecaprenyl phosphate) to produce lipid II as described in the methods. This lipid was finally purified as shown in Figure 3.1.0 by DEAE affinity at the 1M ammonium

acetate fraction and visualised under UV light to confirm Dansyl fluorescence on the lipid group isolated as well as by TLC (Figure 3.1.3 and 3.1.4)

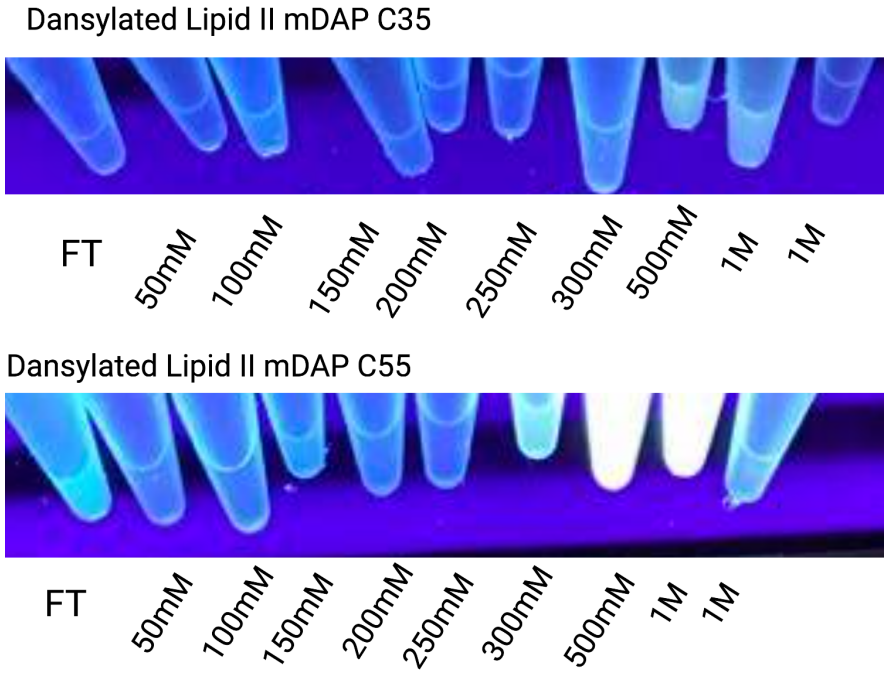


Figure 3.1.4 Dansylation confirmation by UV excitation lipid II wash

FT denoting flow through of DEAE beads before Amonium acetate use and concentration mM ammonium acetate used for elution. Fluorescence and therefore brightness of tube an indicator of dansylation present therefore dansyl mDAP lipid II.

The resultant lipid was resuspended in 1x Schägger buffer and LDAO 6mM at 100µM in preparation for any future Schägger gels and stored at -80C. 90ug of C35 was made and 45ug given to David Roper. Chapter 2 describes confirmation of the unlabelled lipid and its products by mass spec, which showed a 1919mw peak. I should take care to add a note that if stored in "Storage solution" useful for mass spectrometry analysis or long term storage, one should ensure ammonia is added in sufficient quantity to make the solution neutral and prevent degradation of the Undecaprenyl phosphate-MurNac bond. The product appeared to be labile.

630µg of 55 Carbon chain length Lipid II (C55) was also created and 315µg was gifted to Jonny Burnett, in exchange for undansylated lipid II for later experiments, also made from my initial preparations. This substrate was used for my native transpeptidation reactions of RodA-PBP2 of Chapter 2.

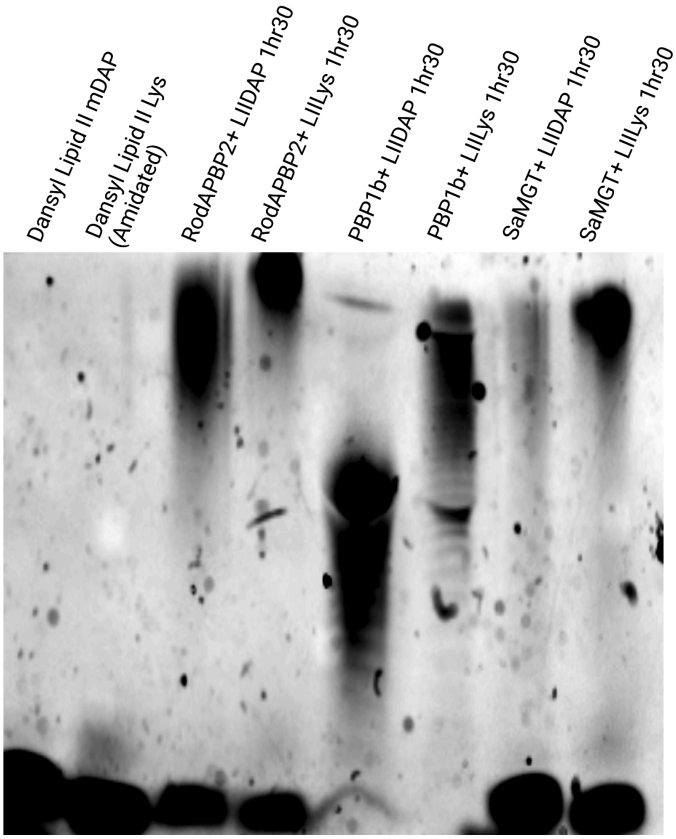


Figure 3.1.5 RodA-PBP2, saMGT and PBP1b Dansyl mDAP lipid II C55 and Lys Amidated polymerisation efficiency visualised by Schägger gel

In vitro activity of lipid II polymerisation enzymes, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel.

An initial Schägger gel was created, using the various polymerization apparatus available at the time of synthesis RodA-PBP2, saMGT and PBP1b(+LpoB), confirming the lipid II was suitable in our experiments as a substrate for polymerisation, and thus lytic glycosyltransferase by other proteins. (Figure 3.1.5). A high molecular weight band was visualised by mDAP dansylated liid II polymerised by each enzyme assayed.

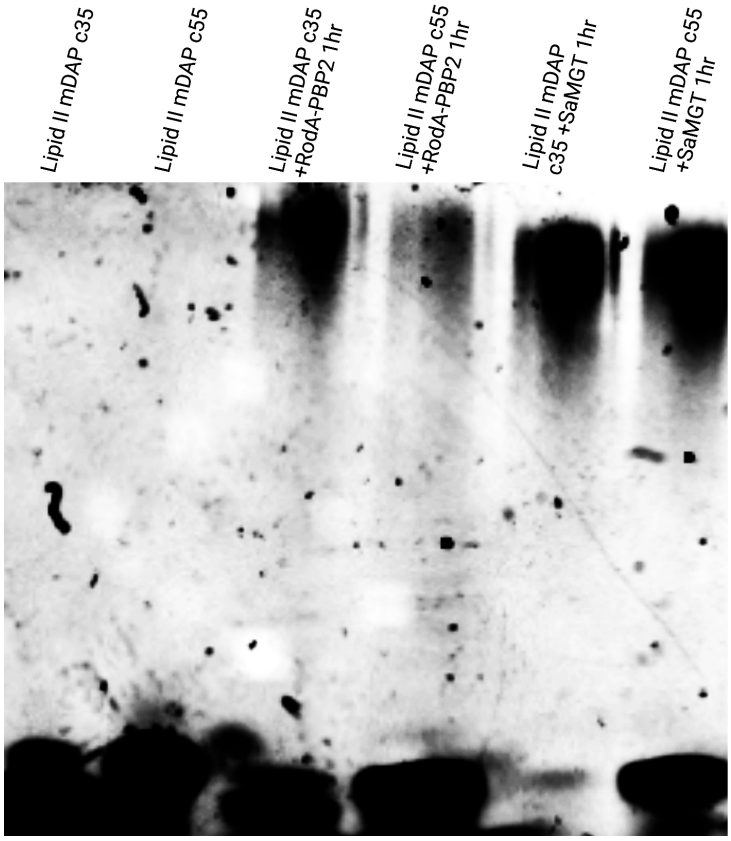


Figure 3.1.6 RodA-PBP2 Dansyl mDAP lipid II c35 and C55 polymerisation efficiency visualised by Shagger gel

In vitro activity of lipid II polymerisation enzymes, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. High molecular weight and Polymerised glycan strains visible at top of electrophoresis gel confirming activity.

In addition to confirmation of the C55 mDAP lipid II stock, which was to be used in the reverse Schägger gel assays, the c35 lipid II was also assayed. The difference between C55 and C35 polymerisation by RodA-PBP2 by activity of polymerisation protein was not visible between samples. This meant these available mDAP substrates could likely be interchanged and tail length was not a factor important to high molecular weight substrate creation, where lipid chains are lost. I sent half of my c35 sample to the US in chloroform for ligand binding experiments relevant to RodA-PBP2 mentioned in Chapter 2.

3.2 Optimisation of a polymerised lipid II substrate

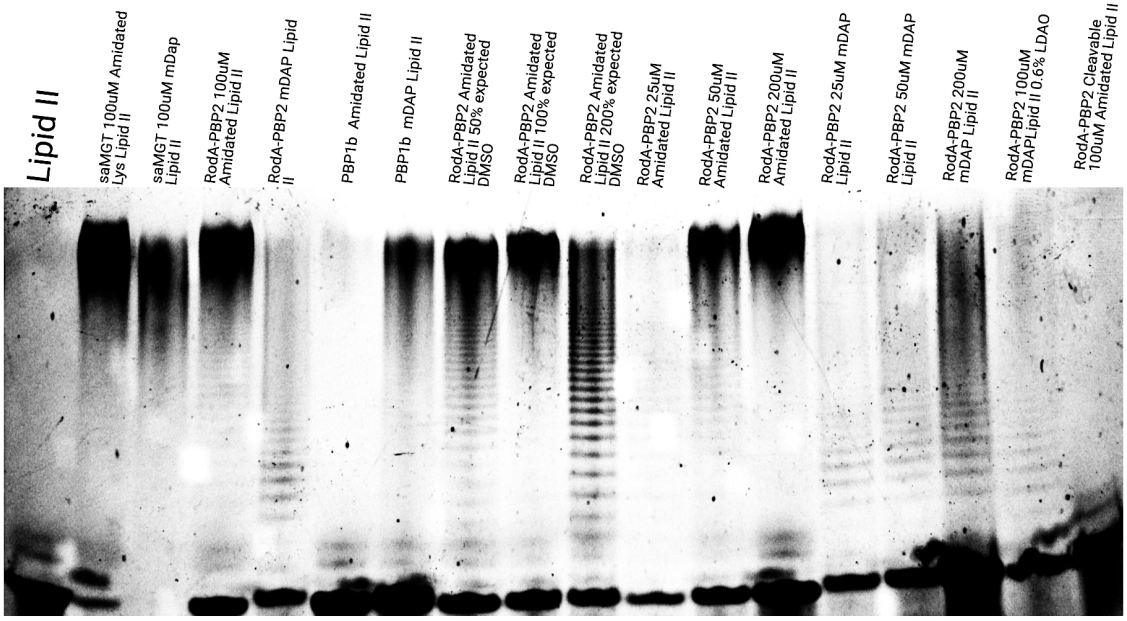


Figure 3.2.5 Characterisation of lipid II variants and polymerases shows saMGT, with mDAP substrate is optimal for lysis assays

In vitro activity of lipid II polymerisation enzymes, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. High molecular weight and polymerised glycan strains visible at top of electrophoresis gel confirming activity.

The amidated Lysine (donated by John Deering and available from the work of past lab members in a communal stock) and mDAP dansylated lipids created by myself were then interrogated under variations in detergent and concentrations, revealing preference or not in a variation of scenarios (Figure 3.2.5). Under the conditions initially used SaMGT and amidated Lysine lipid, made the best substrates for visualisation, however mDAP lipid polymerised by RodAPBP2 was also suitable for experiments dependent on peptide chain specific experiments such as the crosslink sensitive LTs. The polymer made by RodA-PBP2 using mDAP lipid II had a weaker banding, perhaps due to the concentration of polymerisable lipid, and PBP1b crosslinked lipid II of both species was of a reduced efficiency,

especially using PBP1b. This led us to two conclusions, firstly that Lipid II sidechains of DAP or Lysine are usable as a substrates for transglycosylases, and also that it could be useful to purify more SaMGT, , as the polymerising agent in the reverse Schägger gel assays if possible since it showed a greater degree of processivity in priducing high molecualr weight lipid II polymers (and had the advantage of ruling out and transpeptidation products since no transpeptidase is present)

3.3.0 Bioinformatics of the Lytic transglycosylases in *E.coli*

Whilst the compounds useful for the assay of lytic transglycosylase activity were created, the enzymes were also being purified. The central aspect of this PhD focuses on the elongasome (Chapter 1, 2) therefore the initial proteins to be investigated in terms of activity were MltB and RlpA, potential interactors with the RodA-PBP2 complex (Chapter 1). RlpA shares genomic context with RodA and PBP2 across many species (69)(Figure 3.3.0B). MltB has been implicated in binding PBP2 in *Pseudomonas* sp.(141)

Predicted structure and Position of Lytic transglycosylases in *E.coli*

In order to visualise the proteins more easily, and understand their active site, but also hypothesise each protein's role and design the protein expression constructs, we looked through the AlphaFold and PDB database(151) to uncover each LT 3D structure (Figure 3.3.2) . The proteins were then truncated based on these models for soluble protein purification, without removal of structured regions independent of membrane. No membrane bound proteins were used during assay development and only soluble proteins expressed. The differences to enzymatic function, dependent on the lipoprotein and inner membrane bound properties of the proteins was not required for basic inhibitor binding and activity testing for assay verification. Instead the focus of activity in this early work was to find enzymatic roles dependent on pH or salt concentration. Although these longer membrane embedded constructs would be preferred when in individual studies of interaction and function in the future.

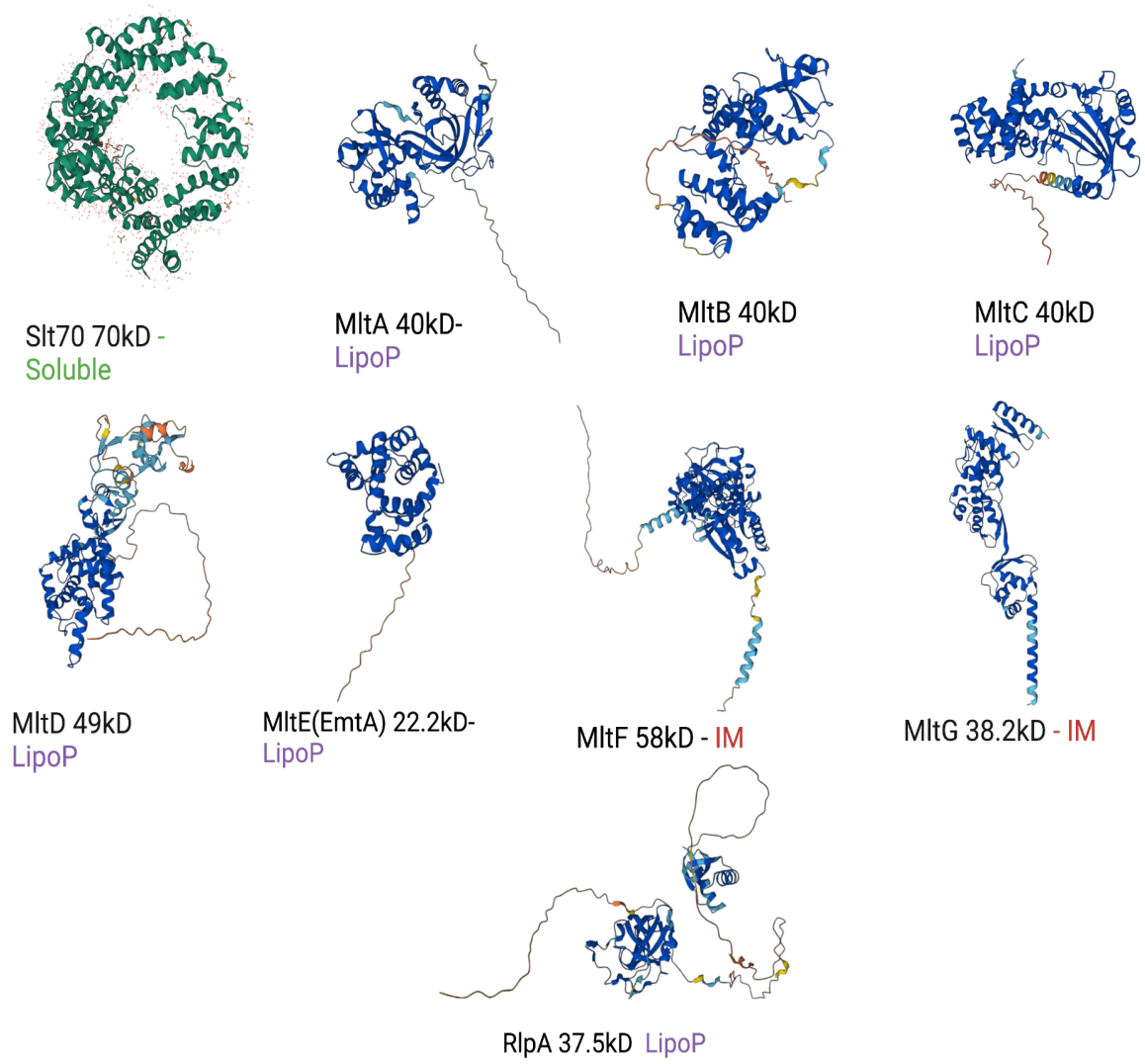


Figure 3.3.2. Predicted and existing structures of the E.coli lytic transglycosylases.

Size of protein annotated. Cartoon representation, structures derived from existing PDB models and AlphaFold predicted sequences. Molecular weight of uncleaved structures shown. LipoP – Outer membrane lipoprotein, Soluble- a soluble protein, IM – inner membrane embedded protein.

3.35 Initial Design and Purification Lytic transglycosylase expression constructs.

In order to move on from the set of assays used by our lab already, in polymerisation, to a lytic transglycosylase assay, testing a set of enzymes able to break apart the glycan strands of peptidoglycan, the protocol had to be adapted. I set out to purify further *E.coli* LTs which had been tested for activity previously, MltA, B, C , D and E as well as Slt70 in addition to a hypothetical LT RlpA. The method of their purification can be found in methods. A gel of the purification of all of these proteins, including Slt70 which required a great deal of troubleshooting and use of detergent due to its membrane interactive nature can be found below. I made the initial purifications, with Francessca Gillet, later purifying her own protein for use in assays in her own PhD. We found that RlpA and MltB were unstable after a few days at room temperature therefore for our experiments were made "fresh" each time.

MltA, C and E plasmids were requested from the Mobashery lab and the resulting proteins purified in Warwick, which were then tested in our system to confirm lipid II polymer degradation. (Figure 3.5.0) I designed the MltB and RlpA constructs in pET47b, which were then made for the first time by Francesca Gillet under my guidance. The Slt70 construct was entirely my design and effort using a pET47b vector. I restricted myself to these proteins, solely for assay development, as MltF, D and G were known to be more complex and inner membrane bound. These will be covered in future work, by Francesca.

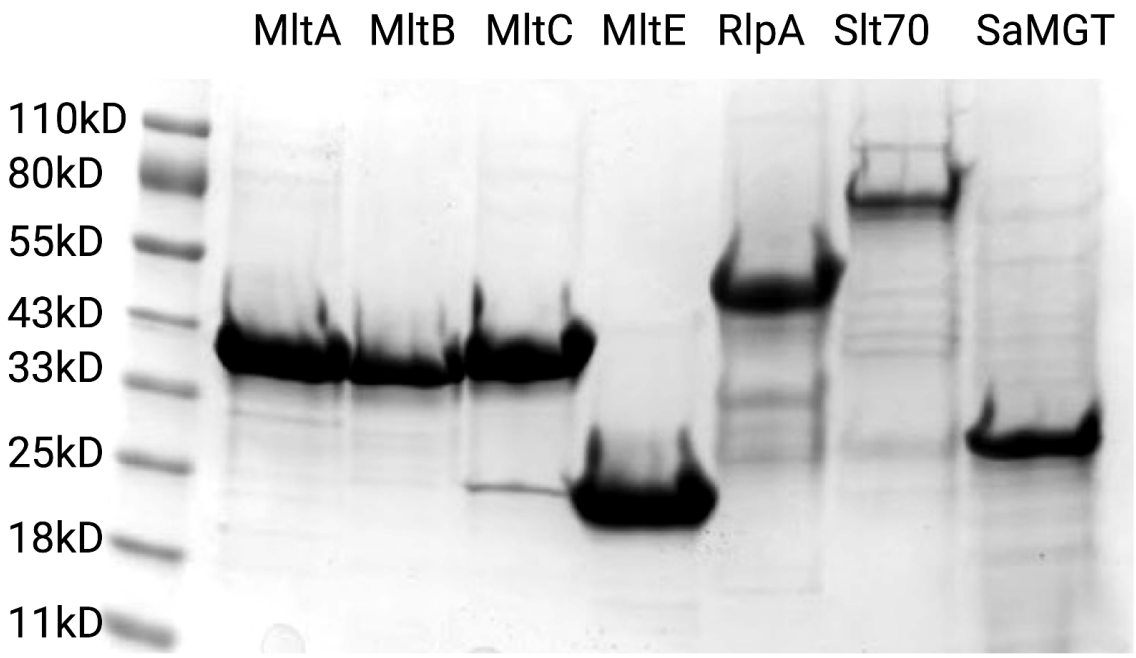


Figure 3.3.1 Purification of Lytic transglycosylase set

Coomassie stain of Lytic transglycosylases of E.*coli*. 10% SDS page gel separated by 1D electrophoresis. Protein molecular weight indicated on ladder lane. Image taken in BioRad imager. Proteins were separated by Nickel based IMAC in a one step purification. Proteins are visualised at 0.5mg/ml.

3.4 A new assay for Lytic transglycosylase activity was devised.

As mentioned previously, in order to further our understanding of the complexes that create the cell wall, it was realised the activity of the lytic transglycosylases needed to be assayed. In Chapter 2, we used a polymerase to polymerise *in vitro* de novo lipid II. In order to assay the activity of a hydrolytic lytic transglycosylases we could add an additional step where a high molecular weight polymer of de novo lipid II polymer could be first polymerised by a transglycosylase, then the proteins denatured and this product assayed on a lytic

transglycosylases. The size of these products could then be visualised through a gel based assay to visualise activity. This is effectively looking for activity of lytic enzymes in the manner reverse to a glycosyltransferase enzyme as we did of RodA-PBP2 in Chapter 2.

As summarised below (Figure 3.4.0), this involved the polymerisation of *in vitro* substrate dansyl lipid II, of which's synthesis is described in 3.2, polymerised by a transglycosylase such as PBP1b or RodA-PBP2. This reaction mixture after multiple time points, could then be heat inactivated at 95C to destroy proteins, then cooled and new enzymes added. After a set timepoint had been reached, SDS was added and the mixture frozen at -20C for later visualisation by 1D electrophoresis in 16.5 or 16% Tris-Tricine gels. These would give resolution of size of fluorescent preferred substrate and product formation by the enzymes dependent on step of reaction pipette into each lane.

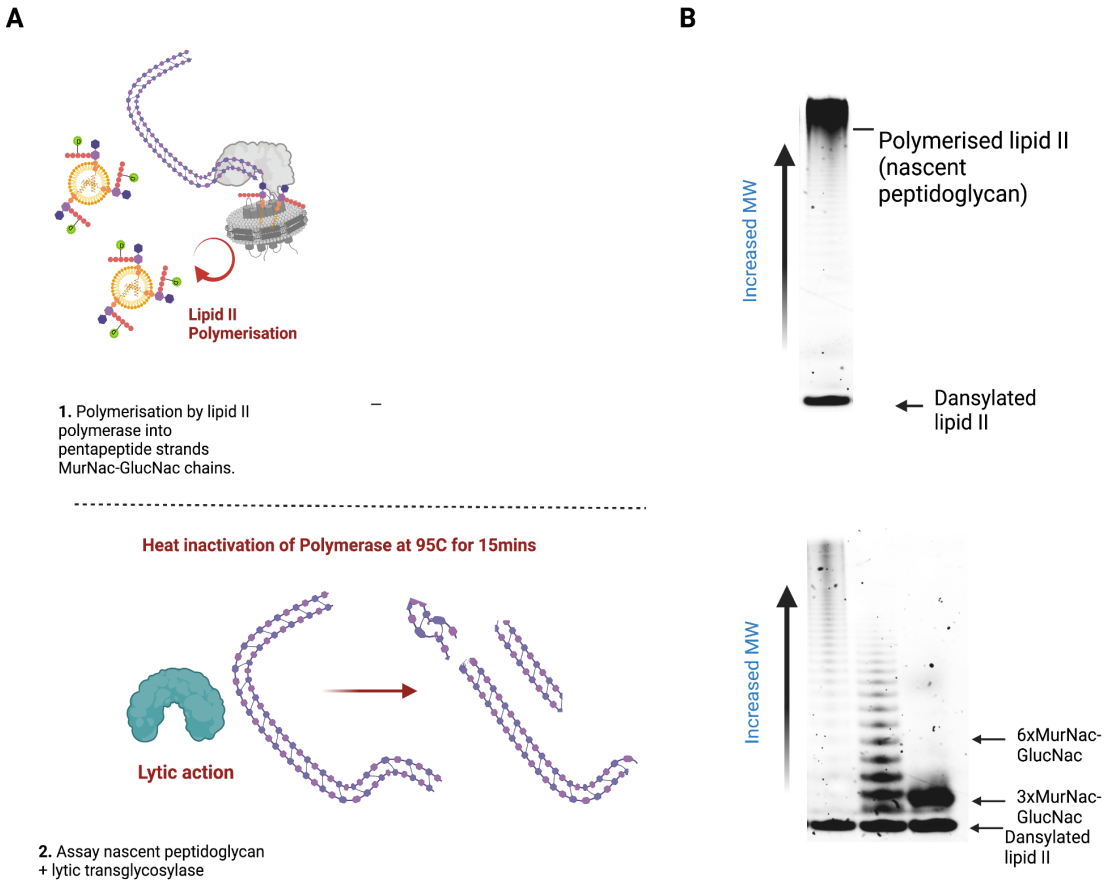


Figure 3.4.0 Method for Qualitative Lytic glycosyltransferase assay

A. Polymerisation of lipid II labelled at the Lysine or mDAP residue creates a nascent peptidoglycan product, the reaction is heat inactivated, and new lytic transglycosylases assayed for activity. B. Fluorescent Dansylated nascent peptidoglycan chains are mixed with Lytic transglycosylases, and assayed over time to reveal endolytic, exolytic nature as well as peptide preference. Sugar counts are visible e.g. 3xMurNac-GlucNac revealing preference for this product.

Using this method, I took RlpA, as the first available hypothetical LT produced by masters student Francesca Gillet, and performed a time course assay in high concentrations revealing that the recombinant sRlpA protein we have purified, is an active enzyme capable of lytic activity over long time periods. (Figure 3.4.1). After a similar length of time required to polymerise the synthetic product by E.coli PBP1b (1hr 30mins), the activity of RlpA became visible. After 2hrs, a reduction of higher molecular weight bands, as well as increase in intermediate sized bands was visible (5-10 disaccharide pairs), , with disaccharide pairs and tetrasaccharide peptide stems increasing in abundance. After incubation with the RlpA protein, the high molecular weight product initially provided, of a lipid II polymerised by the PBP1b was entirely lost, with decreases in band strength shown especially prominent at 8hrs. However, use of PBP1b (Gifted by Julie Tod) as a polymerase, was not ideal, as it provided a mixed substrate of variant chain length, and therefore was not very replicable.

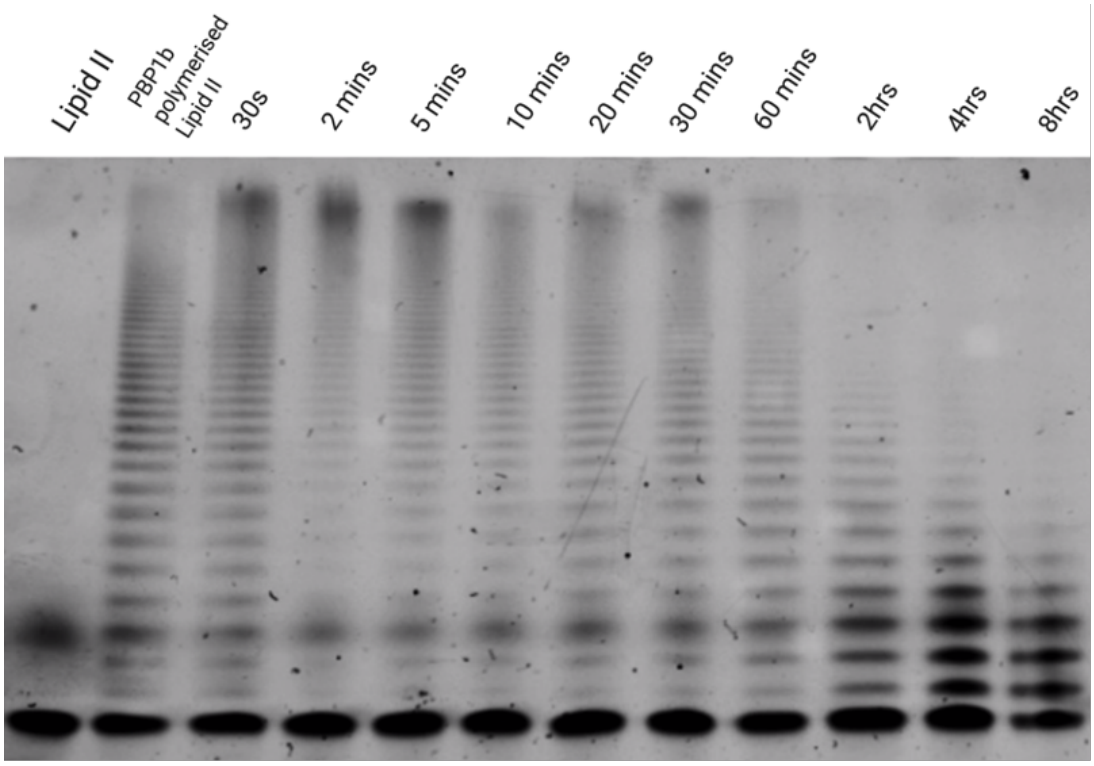


Figure 3.4.1 A new "Reverse" Schäggergel allows lytic transglycocylase assay of RlpA

in vitro activity of RlpA 0.5mg/ml on product of High molecular weight glycan polymer, visualised by fluorescent mDAP lipid II polymerisation in Schägge rgel. Sugar product size visualised. Gel performed with Francesca Gillett.

3.5 Assays of Lytic transglycosylase activity in *E. coli*

After initial success of RlpA activity in a time course assay with PBP1b as a transglycosylase, the remaining lytic transglycosylases were then purified by myself and assayed against the amidated lysine peptide side chain version of the Dansylated lipid II, each for 2 hours. Soluble MreC was added as a control protein incubation, which should not break down product. During typical protein purificaiton, chicken egg lysozyme is used as a cell lysis tool. In our experiments, no lysozyme was used during the methods for LT protein isolation (Figure 3.5.0) this was to reduce the potential of lysozyme, an enzyme which can break down PG, affecting our assays.

The first assay of the activity of all the lytic transglycosylases simultaneously, allowed for some measure of lytic transglycosylase activity categorisation (exolytic action versus endolytic action). Endolytic is cutting the peptidoglycan with no limit to length of peptidoglycan on either side, in the centre of the glycan chain molecule. Exolytic, is action that cuts peptidoglycan from the outside in. Throughout the assays, it was also noted that the peptidoglycan created was not crosslinked, and modified at the third peptide residue therefore represented a uniquely unbranched glycan chain, which may affect the activities of the enzymes, or binding, although a great distance from the active site that breaks down the sugar.

In this first assay, Lipid II could have been polymerised by any of our polymerases (Figure 3.2.5), however as we had a great deal of RodA-PBP2 from Chapter 2 and PBP1b was not ideal due to its product range, RodA-PBP2 was used for the initial polymerisation. After polymerisation for 1hr 30 as was typical for activity testing on our RodA-PBP2 construct, the samples were boiled at 95°C for 15 minutes to heat inactive the RodA-PBP2 polymerisation activity. Figure 3.5.0 reveals the first known activity of each enzyme after two hours in our assay system. MreC has little to no effect as a control, and likewise the polymerised lipid II sample left at 37°C was also unaffected. MltA, B, C, E, RlpA and Slt70 showed activity, with Lysozyme showing a significant amount of activity and creation of a three-glycan long polymer product. This specific product increase can also be seen in Slt70, however with a two-glycan product also visible, as with the other LTs, that is not visible in either the polymerised lipid II or Lysozyme, this indicates a difference in the mechanism of catalysis. RlpA perhaps due to degradation or non-ideal reaction conditions was the most poor performing, but like the earlier experiment, did reduce the fluorescence of the high molecular weight product presented.

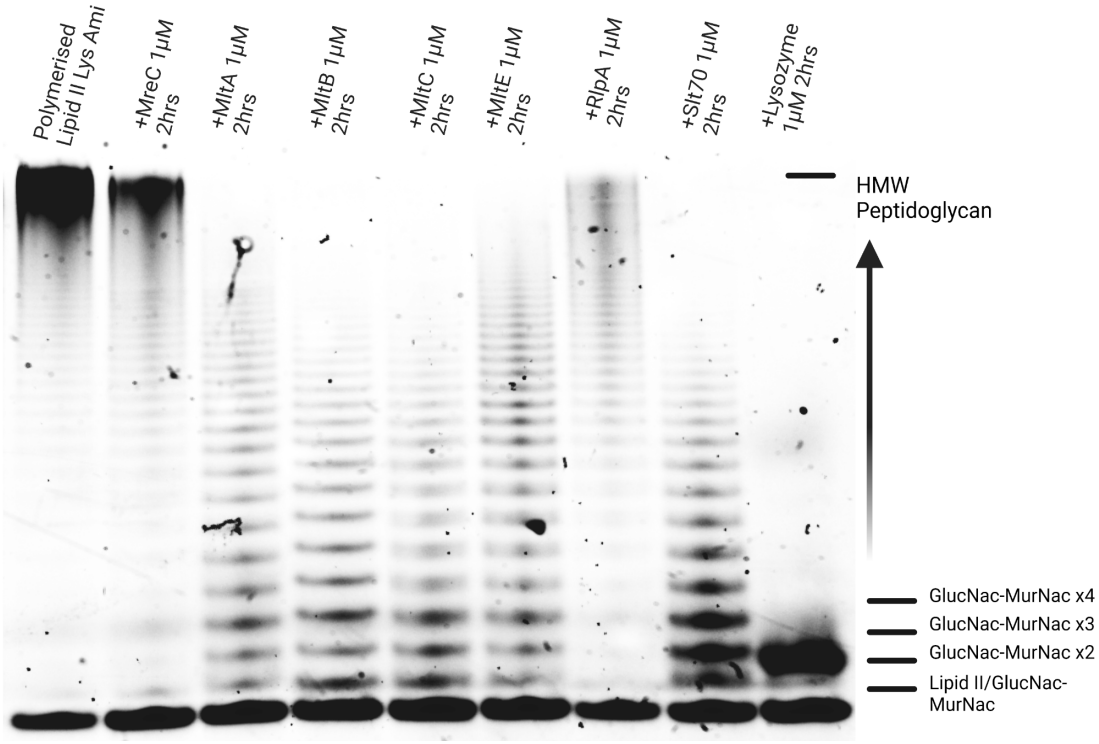


Figure 3.5.0 New Schägger Lytic Transglycosylase assay on *E.coli* LTs with Lipid II Lys substrate

In vitro activity of Lytic transglycosylases on product of High molecular weight glycan polymer, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. Sugar product size visualised by fluorescence.

The assay, presents a range of product sizes, however as it is not a timepoint assay, can give no indication of preferred glycan chain length. It can be indicated from the assay, that it is suitable for use by all the above lytic transglycosylases. The pH used as 7.5, with a background of imidazole present. However as noted, a pH dependence on activity is known and could be investigated by the system, if the pH was changed. This will be addressed in F.Gillets PhD and the methods paper with my suggested values of pH investigated indicating gut pH such as 3.5, against 5.5 and 7.5pH controls, as well as alternative salt concentrations and metal ions.

3.6 Qualitative activity change on RodA-PBP2 addition by Lytic transglycosylases.

Lysis and polymerisation can occur simultaneously, and allow an alternatively processive RodA-PBP2, and finer vision of lysis type by LTs.In order to test the potential of protein-protein dynamics within a cell at the elongasome between lytic transglycosylases and the SEDs proteins, and test the limits of the assay, we polymerised lipid II simultaneous to lytic action by the LT set. (Figure 3.6.0). This in addition revealed the exolytic or endolytic nature of each individual LT alone more clearly. When the lytic transglycosylases were competing against RodA-PBP2 in polymerisation, at similar molar concentrations this may have led to interactions at the apparatus responsible for glycosyltransferase activity, or caused interactions of the LT and their polymerised substrate, as well as the RodA-PBP2, where an exolytic activity would interact directly with polymerisation, and endolytic activity would require larger substrates.

The result would also change the visualisation seen in the gels, with endolytic enzymes likely to destroy high molecular weight products or prevent their appearance, and exolytic enzymes simply reducing the upper limit on high molecular weight chains. For example, an endolytic LT would produce a higher concentration of medium molecular weight products, as it requires a larger substrate to function, therefore is intrinsically required to hydrolyse the PG after the SEDs activity. In contrast an exolytic LT would immediately produce low molecular weight products regardless of the average chain size. Here in Figure 3.5.0 Slt70 is seen as an endolytic LT, with RlpA an exolytic LT, as seen by a gradual change in product length. MltA, B and C also appear to be endolytic. These activities may also change dependent on pH and temperature.

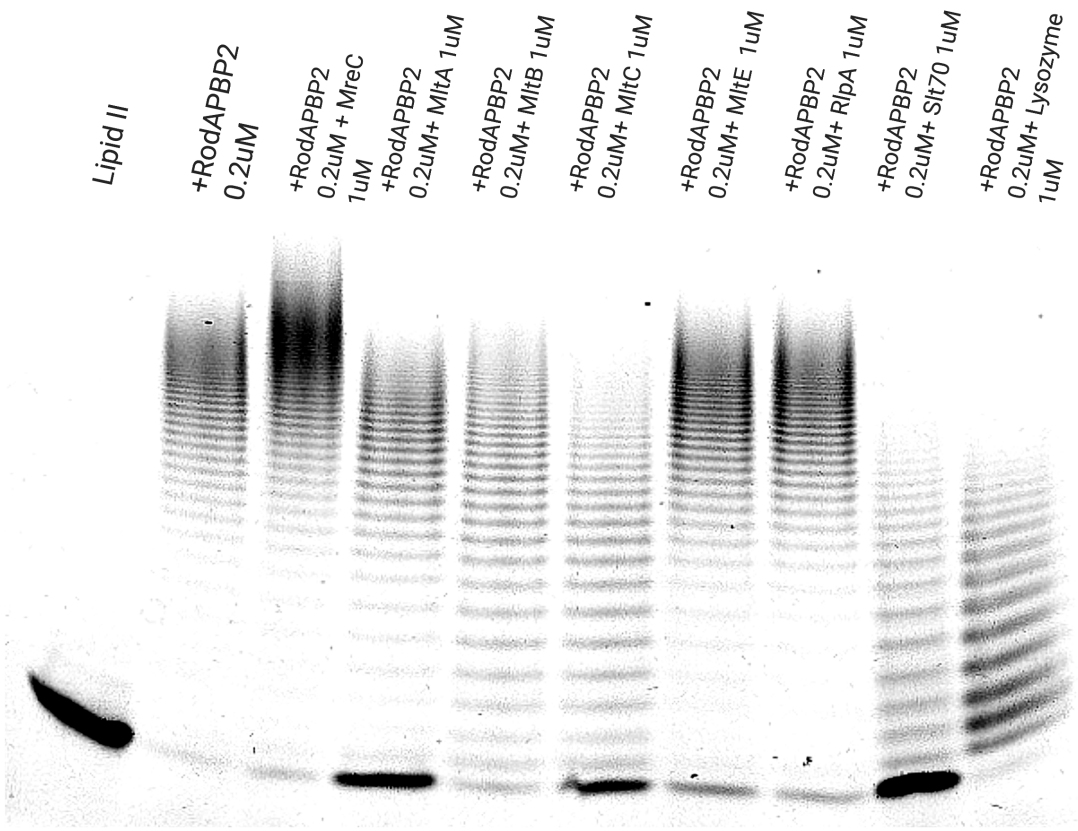


Figure 3.6.0 Lipid II polymerisation by RodA-PBP2 in the presence of MreC and potential accessory lytic transglycosylases

In vitro activity of lytic transglycosylases on product of high molecular weight glycan polymer in the presence of active RodA-PBP2, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. Sugar product size visualised.

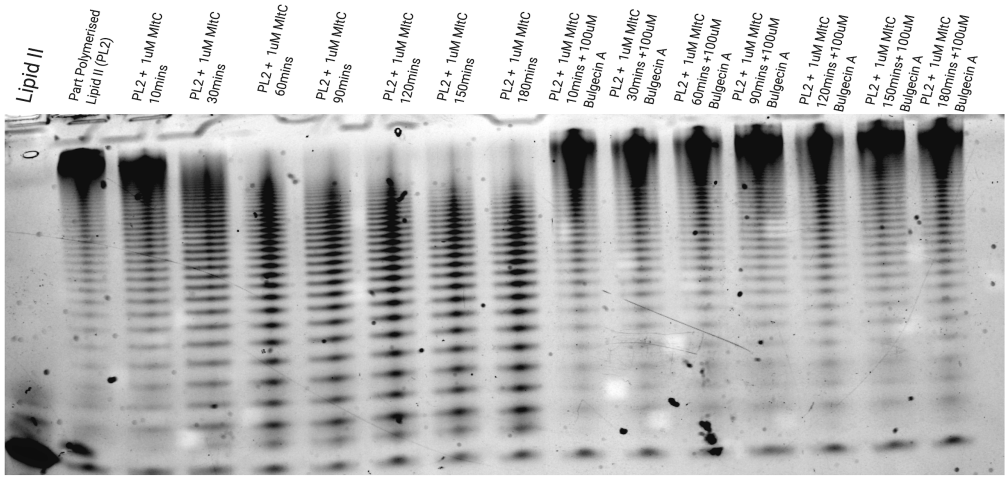
The alternative product sizes predicted in the introduction to the gel, and the experiment on co-incubation can be visualised in MltA, MltE and RlpA which show a clear increase compared to the control of the two-sugar product similar in size to lipid II, alongside high molecular weight PG as exolytic enzymes, creating a single product, and MltB and Lysozyme as endolytic enzymes creating 3-10 sugar products of similar molecular weight but a reduction in high molecular weight bands. MltC action is unclear on either exolytic or endolytic activity in this assay, with increases in both 2xSugars and mid-sized bands compared to RodA-PBP2 alone. The sMreC control, with a co-incubation of the two proteins seems to reveal an increase in

activity by RodA-PBP2 when in the presence of RodA, as shown by the darker and higher molecular weight band. Overall, co-incubation in follow up experiments, could be a way to visualise differences in lytic activity more easily. However due to the difficulty in convoluting the reasoning behind the differences in activity and the difficulty in removing factors such as RodA-PBP2 to Lytic transglycosylase interactions it was not pursued.

3.8 New Lytic transglycosylase assay allows for inhibitor screening and furthers understanding of antibiotic and adjuvant action

Our new lytic transglycosylase assay and ability to show activity of the LTs tested *in vitro* for the first time, allowed us to also test the lytic transglycosylase inhibitor and potential adjuvant, used in conjunction with beta lactamase inhibitors Bulgecin A, which has been shown to be dependent on lytic transglycosylases and bind to LTs in crystals, but has never been directly linked to their activity and instead their degree of binding. Therefore In Figure 3.8.1 a timecourse assay of the LTs was set up in the presence of Bulgecin for MltB (control) and MltC, MltB was again not inhibited by Bulgecin A (Figure 3.8.1) in the conditions we used. A focus on MltC activity affected by Bulgecin (Figure 3.8.1) reveals Bulgecin does inhibit MltC *in vitro*, but not MltB, similar to the literature.

A



B

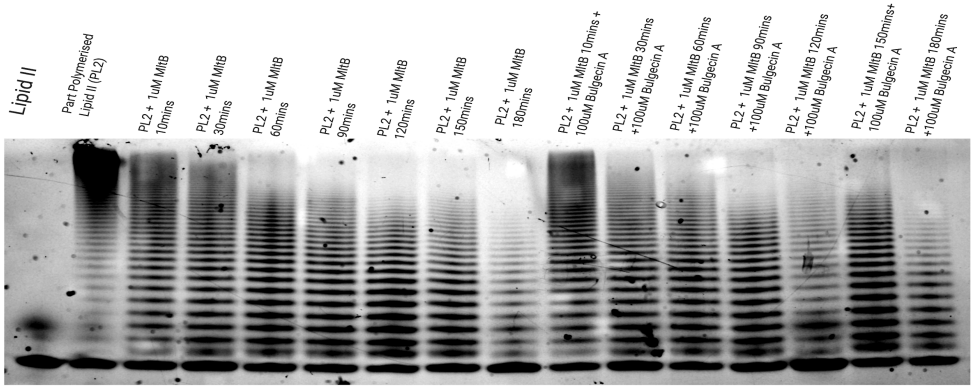


Figure 3.8.1 New Lytic transglycosylase assay indicates No inhibition of sMltB by Bulgecin A , but Inhibition of sMltC *In vitro* activity of MltB and C in the presence of Bulgecin on product of High molecular weight Part Polymerised Lipid II glycan polymer PPL2- Part polymerised lipid II – 10µM polymerised by RodA-PBP2 0.5µM 1hr 30, visualised by reverse of fluorescent lysine lipid II polymerisation in Schäggergel. Sugar product size visualised over time as indicated.

Lack of inhibition of LT activity on MltB at pH 7.5, shows that Bulgecin may not affect all LT enzyme activity. Conversely Figure 3.8.2 shows MltC is inhibited completely by Bulgecin A presence. A study of various bulgecin activities, in various conditions would be a future study point in future work.

Novel rSLT activity determined by reverse Schägger gel.

After this initial assay series, we collaborated with the US army research department, to assay a new LT, in the presence of mDAP and amidated lysine. This would be the first test on an unknown protein “rSLT” to confirm its role as an Lytic transglycosylase.

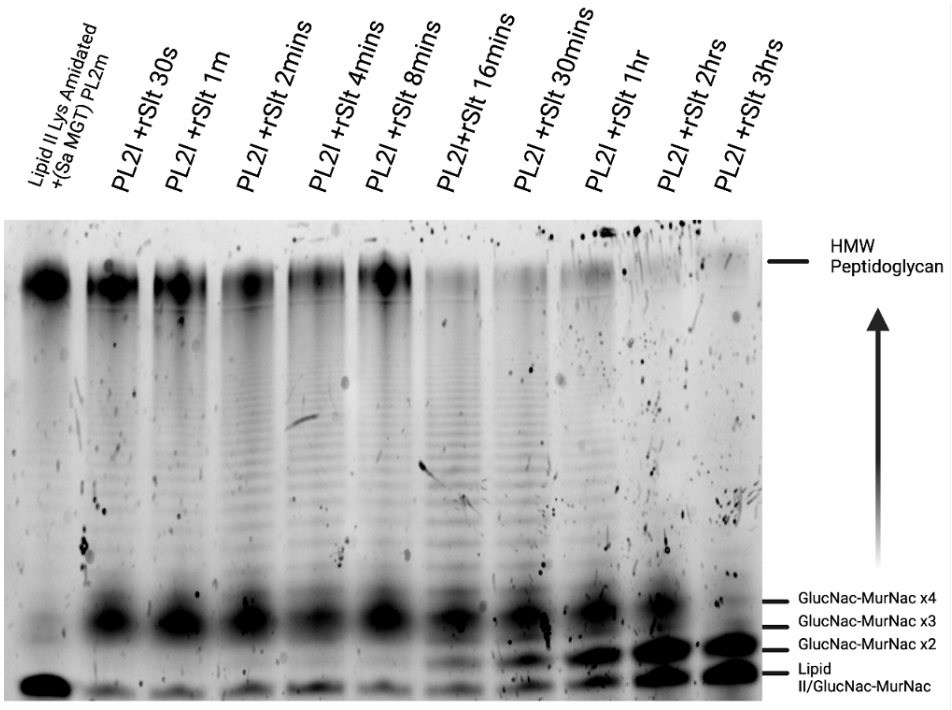


Figure 3.8.2 Novel LT activity testing rSLT from USAMRID

In vitro activity of rSlt in the presence of fluorescent High molecular weight glycan polymer, visualised by fluorescent lysine lipid II polymerisation in Schägger gel. Sugar product size visualised.

When we subject 10µM polymerised lipid II to 0.5µM rSlt in this timecourse performed at 37C [20% DMSO 50mM HEPES, 50mM NaCl, pH 7.5, 10mM MgCl2] immediate breakdown of peptidoglycan occurs. The enzyme initially produces a GlucNac-MurNacx4 product, exolytically as all other strand lengths remain similar concentrations. As the reaction continues this increases in abundance, reducing the polymerised lipid II concentration until the substrate changes, to the initial product GlucNac-MurNacx4, a new product is then created GlucNac-MurNacX2, and similarly and finally GlucNac-MurNac, showing an order of preference and catalytic preference. Similar uses of the assay on other enzymes and conditions may reveal the role and activity of other cryptic species enzymes or proposed Lytic transglycosylase pseudogenes.

3.9 A FITC label confirmation of Lytic transglycosylase activity

In order to validate the results of the schagger gel assay as detecting lytic transglycosylase activity, and have our enzyme activity independently shown through other methods, we decided to use another technique for measuring general hydrolase activity on peptidoglycan. We therefore confirmed the activity of MltA, B,C and E as well as RlpA and Slt70 using an existing assay, the FITC- peptidoglycan release assay. In this assay purified proteins were added to a FITC labelled *E.coli* peptidoglycan mix (gifted by Manuel Banzhaf) and the fluorescence of the solution, constituted from released FITC substrate was measured as an approximate measure of hydrolytic activity on intact peptidoglycan in Figure 3.9.0.The protocol for this is described in methods. Controls on the assay include a no hydrolase control, as well as the use of multiple tests for each protein. The major difference in these assays is in our reverse Schägger gels, we create de novo pentapeptide dansyl labelled substrate, however in the FITC label assays, it is a mix of substrates, and therefore a range of hydrolytic activities as well as substrates can be assayed. This includes a marked reduction in the concentration of de novo pentapeptide substrate.

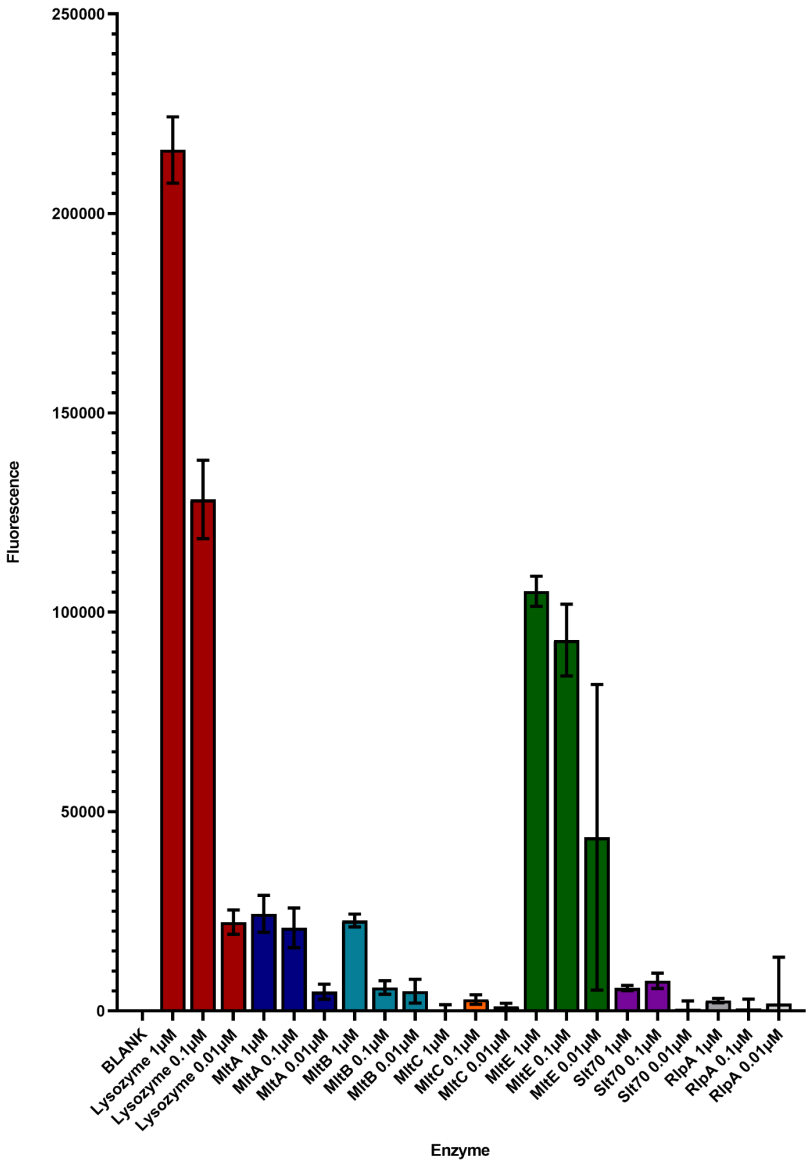


Figure 3.9.0 A FITC- fluorescence based assay validates our in vitro gel system.

Fluorescence measured at an excitation of 490nm and emission at 519nm. FITC-PG 1mg/ml Anomylous values removed. Assay performed together with Francesca Gillett, using Banzhaf lab substrates.

The assay was able to confirm the activity of our enzymes, however the FITC assay involving MltC was less affected and could not be confirmed in this assay, unlike our own reverse Schägger assay, possibly due to the difference in roles between the proteins and the substrate type, which will be discussed later. The primary drawback of the FITC assay results on our proteins is that a tenfold difference in enzymatic concentration was not conducive to a 10 fold

difference in activity, therefore suggesting the FITC fluorescence release assay is not a quantitative assay, but rather a test of hydrolase activity in general.

Another drawback to the assay is the substrate is not uniform but instead a combination of many potential peptidoglycan states; de novo, division and highly denuded peptidoglycan through to highly crosslinked murein. The fluorescent readout of released fluorescence from the substrate therefore is not specific, also including endopeptidase activity, as the FITC label is linked to the peptide stem. This fluorescent readout is related to a filtration system, with the newly solublised peptidoglycan formed by the hydrolase, not filtered by a 0.1μm filter allowed through to fluoresce and shown as a reading of the assay. This does mean that activity could be difficult to detect if the enzymatic activity is not conducive to peptidoglycan being cleaved from the main body, and for example being connected through remaining peptide crosslinking.

However, on an individual basis per protein, it could be considered to be useful in individual changes in activity such as with the introduction of binding partners and inhibitors on a set protein concentration which may increase or decrease the readout, irrespective of how this extra fluorescence is achieved. The assay also involves the use of an undigested complete peptidoglycan polymer, with crosslinks, therefore differences in activity in the assay and our own gel-based system would be informative on the preferred target type, and as a validation technique. Our gel-based system is a single substrate of de novo peptidoglycan, with altered stem peptide, however some lytic transglycosylases, as referred to earlier, seem to be sensitive to the substrate peptide type provided. MltC for example acts upon *de novo* peptidoglycan in our assay, but causes minimal change to the FITC-*E.coli* PG mixed substrate.

The use of the FITC assay, alongside our *in vitro* gel system, could allow the potential of a highthroughput inhibitor screen, and activity dissemination not yet available, as well as drug testing specific to LT activity.

Chapter 3 Discussion: Future research in Lytic transglycosylases

In the first half of this chapter, the purification of the Mur ligases, followed by synthesis of fluorescent lipid II was documented, which enables *in vitro* monitoring of polymerisation and compounds which may inhibit it. As well as this *in vitro* creation of lipid II through a series of enzymatic steps, and polymerisation and crosslinking of peptidoglycan, it then made sense to continue the journey of cell wall creation and the modification in PG to lytic glycosyltransferase by the LTs in *E.coli.* This enabled us to create an assay for the discovery and creation of new antibiotics suited to LTs.

The investigation of the lytic transglycosylases as enzymes will be essential in tackling antibiotic resistance. As already discussed (Chapter 1) the division and elongation of bacteria relies upon a cascade of interactions and mechanisms. In the context of lytic transglycoslases, these are typically roles most likely involved in shortening chain length and creating space for new peptidoglycan as well as large structures which cross it such as flagellar or Pili. However in addition to this, lytic transglycosylases and its related lysozyme molecules have the capability to break down large amounts of peptidoglycan (as shown in our assay) and therefore have antimicrobial activity. The creation of assays able to determine the effectiveness of these enzymes, along with potential predatory secretion systems not yet discovered may allow for the development of topical medicine used in conjunction with other antibiotics as a general, or that specific target to certain strains of bacteria.

Our assays have allowed for the confirmation of two new enzymes as lytic transglycosylases (RlpA and rSlt (Figure 3.8.0), however have also shown that some LTs have a specific preference for certain amino acid peptide chains attached to their glycan chain target even with other chemical groups potentially blocking the interaction. Specifically, the Gram-negative lytic transglycosylases have some preference for Gram-negative dansylated peptidoglycan over Gram-positive. (Figure 3.8.0, Figure 3.8.2). Revealing the degree of binding of substrate to the LTs is more specific than the transglycocylases investigated in Chapter 2.In addition, the use of the lytic transglycosases in tandem with polymerase enzymes such as RodA, allows for

diagnosis of increased activity or altered activity of these partners, which may be useful given more enzymes in creating a full model of PG synthesis *in vitro* in the future.

We investigated the use of a known antibiotic adjuvant, able to supplement existing beta lactam antibiotics “Bulgecin”. We show in our data, that Bulgecin acts on MltC, as one example to reduces or halts LT activity for the first time. This will be investigated in further studies, however, here in this thesis, indicates the use of the gel based assay for LT specific activity inhibition testing. This will be expanded to pH and temperature differences in the future to reveal the roles of each LT protein in the cell.

Chapter 3: Conclusion

In conclusion, our assay of LT activity has for the first time shown the activity of LTs directly, indicating the processivity and type of enzyme activity (exolytic vs endolytic) but also allowed for a new screening method of antibiotics such as the bulgecins which inhibit LTs, to discern their *in vivo* targets. This has been verified by a FITC labelled peptidoglycan assay, which also indicates our LTs are active hydrolytically. Two advantages exist in our assay, substrate/product visualisation, and specificity of action type.

This assay was made possible by our ability to synthesise and polymerize lipid II in the different chemical forms used predominantly in Gram-positive (amidated, lysine) and Gram negative bacteria (DAP). The project will form the basis for a new assay/method paper for lytic transglycosylases, and continued as a PhD by my colleague and friend Francesca Gillet. The future of this assay may also find utility in characterization of specific LT inhibitors such as Bulgecin, and for new lytic transglycosylase discovery, as exampled by my collaboration on rSlt.

Chapter 4: Fluorescent labelling of OM balance and PBPs in *Pseudomonas aeruginosa* and *Escherichia coli*

To be submitted for publication after PhD as two papers

'*Pseudomonas* MCE domaining protein localisation and phospholipid insertion', CLB Graham et al 2023 (Submitting to Microbiology- Attached)

And

'Phosphatidyl choline mimic propargyl-choline for membrane staining and phospholipid tracking in *Pseudomonas aeruginosa*'. CLB Graham et al 2023 (Submitting to Access Microbiology - Attached)

Chapter 4: Summary

In Chapters 2 and 3 I have focused on the peptidoglycan related proteins of the cell envelope. However as discussed in Chapter 1, we hypothesise a link between these proteins and the proteins involved in the formation of the outer membrane. The localisation of PG related proteins in *E.coli* is well studied, and is division related with concentrations of protein increasing at the division site, or along the cell wall during crucial growth periods. (Chapter 1.1/1.2). However at the beginning of the PhD the localisation of outer membrane synthesis proteins was not yet known in Gram-negative bacteria, especially *E.coli* or in *Pseudomonas aeruginosa*, two model and pathogen relevant Gram-negative bacteria. This chapter looks at the localisation of the proteins and lipids involved.

The first step to determine the existence of a link between these two processes was to determine the outer membrane biogenesis localisation patterns throughout the cell cycle. Using the knowledge gained in bioinformatic and literature exercises of Chapter 1 I labelled the proteins MlaD, PqiB, YebT and other MCE containing proteins. This was done first genetically in *Pseudomonas aeruginosa* with a C-terminal mCherry fluorescent protein dependent on native expression, and then with a pBAD30 complementation system in *E.coli*. These experiments revealed that indeed the MCE domain containing lipid transporting proteins are specifically localised. In the case of *Pseudomonas aeruginosa*, the proteins follow a similar polar pattern. This polar link is significant in that the substrate for the MCE domain containing proteins has been shown to be in part cardiolipin, a lipid found more often at the poles.

In addition to this, I developed a method for *in vivo* labelling *Pseudomonas aeruginosa* membranes, with a phosphatidylcholine mimic ‘propargyl-choline’ which can be labelled with click chemistry to track lipid movements over time. This method has now been characterised, so others may track and visualise the lipids in bacteria. This is different from dyes which have an affinity to lipids such as FM464 in that is specifically labels only a certain lipid and not all membrane layers, in addition to tracking individual lipid movements.

The results and theory of this research chapter were the subject of two successful grant applications by myself to Mitacs- NERC and to ANTRUK. I also presented my work at the GRC Cell Envelope conference Vermont, US aswell as a conference I organised at the University of Warwick

Chapter 4 Abstract

If the linkage between peptidoglycan and membrane biogenesis is explored it may allow us to postulate the most efficient combinations of antibiotics that affect both systems to address Gram-negative bacterial pathogens in a novel way. Although the existence of separate systems for outer membrane lipid biogenesis and peptidoglycan biosynthesis has been known for some time, there is still a great deal of information missing regarding the regulation and coordination of these processes. Here for the first time, we interrogated localisation of phosphatidyl choline (PC) in *Pseudomonas aeruginosa* using a chemically pliable phosphatidylcholine headgroup mimic 'propargyl-choline' (PCho), but also tracked the localisation of Waal an O-antigen ligase and MCE domain containing lipid transport proteins MlaD and PqiB. In this study, we couldn't definitively show the insertion of new outer membrane lipids, and instead create a new phospholipid labelling technique. However my experiments have shown that among the MCE domain-containing proteins pliable for mCherry tagging, the transport machinery PqiB and MlaD are specifically localised to poles and longitudinally in *P. Aeruginosa*, this localisation also does not align with HADA (3-[[(7-Hydroxy-2-oxo-2H-1-benzopyran-3-yl) carbonyl]amino]-D-alanine hydrocholoride) insertion, a fluorescent peptidoglycan D-Alanine mimic, therefore these transport machineries are not likely associated directly with peptidoglycan insertion. I also indicate that there are likely many roles for the MCE domain-containing proteins; PqiB and Mla associated more strongly with cell integrity elsewhere.

Chapter 4 Output

1. New localised Phosphatidylcholine insertion mimic and tracking technique
2. Created *Pseudomonas aeruginosa* Genomic mCherry labelled strains
3. Localisation of Waal, PqiB, MlaD *in vivo*
4. Construction of *Escherichia Col*i pBAD30-mCherry constructs across 14 proteins, for future localisation and labelling study.
5. Two successful grant applications.

Chapter 4: Background

Chapter 1.2, a review of the outer membrane biogenesis and lipid exchange proteins can give a full review of the outer membrane. However the most concise background context for this chapter is to focus on a set of MCE domain (lipid interacting domain) containing proteins, which have been shown to interact with lipids and form parts of outer membrane and inner membrane homeostatic pathways. They have little known of them in terms of interaction partners or connections to other cell envelope maintenance pathways such as polymerisation of PG by RodA, FtsW, PBP1b or a (refer to often in Chapter 1.1 and Chapter 2).

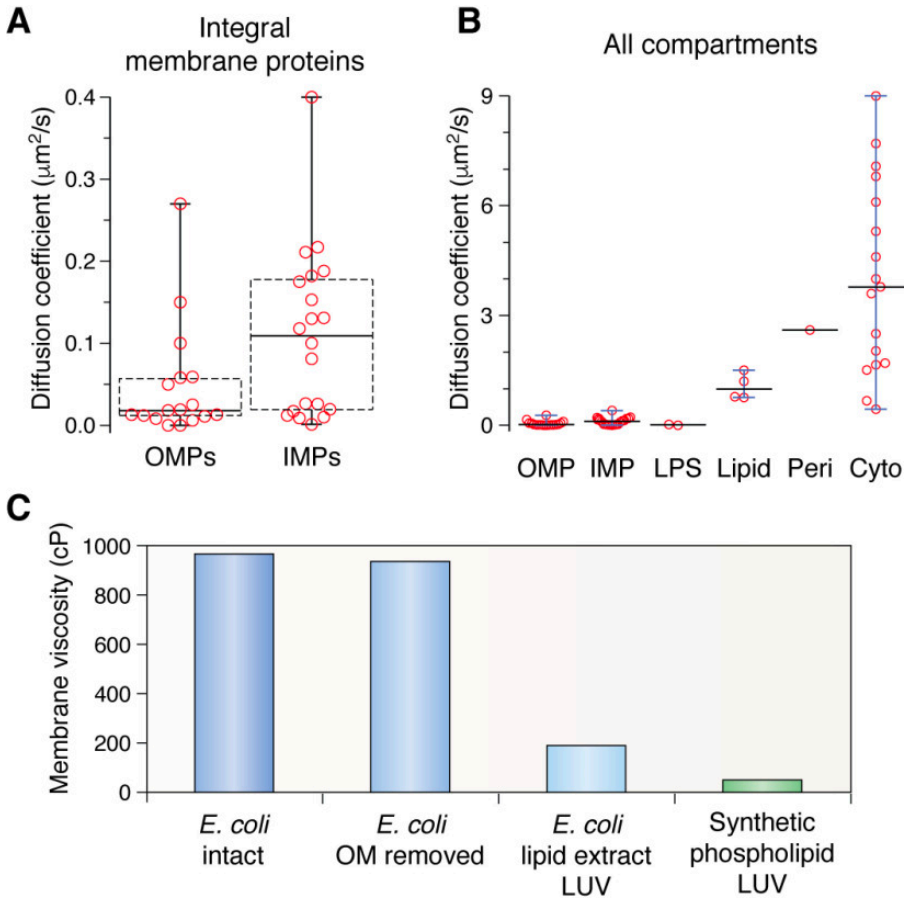


Figure 4.0.5 Lipid viscosity

A Box plots showing the range of diffusion coefficients reported for OMPs and IMPs (see "A crowded environment"). Boxes show interquartile range calculated by the Tukey method with the median indicated as a boldface horizontal line. Whiskers show the minimum and maximum values. **B** comparison of the diffusion coefficients of membrane proteins with other components of bacteria. Whiskers are only shown for components that have three or more values reported in the literature. All values are reported from in vivo studies. LPS, diffusion of LPS molecules in *S. typhimurium*. Lipid, diffusion rate of a fluorescent lipid reporter probe in *E. coli* membranes. 'Peri', diffusion of soluble protein in the *E. coli* periplasm. Cyto, diffusion of soluble protein in the *E. coli* cytoplasm. **C**, viscosities of different membrane environments as measured by the use of fluorescent BODIPY C10 lipid reporter probes.. Adapted from 'Role of the lipid bilayer in Outer membrane protein folding of Gram negative Bacteria' J Horne et al 2020 (105)

The outer membrane and peptidoglycan must be made in tandem if the lipids are fluid enough as suggested in J Horne et al 2020 and Figure 4.0.5, or there will be cause of blebbing and bubbling

in the membranes in response to differences between inner membrane, outer membrane and peptidoglycan surface areas. The coordination of division between the outer membrane and peptidoglycan has been shown before with BAM (beta barrel insertion) and LpT LPS insertion, however not yet for the phospholipids (Chapter 1.2).

Maintenance of balance between the cell envelope layers, specifically of the inner leaflet lipids within the outer membrane would require coordination if lipid rafts are moving slowly through this layer as the literature has suggested. An inner leaflet expansion would be caused by too many lipids inserted in a confined space compared to the LPS would promote curvature, which in turn would affect anchoring partners which hold the lipid membranes in place and the rigid peptidoglycan.

The MCE domain containing proteins YebT, MlaD and PqiB were all hypothesised to be involved in the movement and exchange of lipids from the inner membrane to the outer membrane, thus were prime candidates for the creation and growth of new outer membrane cell envelope(10). Recently another paper has also shed light on AsmA domain containing proteins which have been suggested to be also involved in outer membrane lipid balance and insertion.(108,152)

The insertion and maintenance of lipids in the inner membrane, and inner leaflet of the outer membrane in Gram-negative bacteria and maintenance of the balance of lipids, is not yet fully understood. Whilst labelling of cells has revealed lipid raft localisation, and cardiolipin localisation, as well as phospholipid abundances over time, no method has been able to track this *in vivo* or track individual lipid movements or metabolism, but are instead normally monitored after insertion has already taken place using probes. Propargyl-choline has been suggested as a useful label, able to mimic the lipid ‘Phosphatidylcholine’ after long incubation times in mammals, also recently *E.coli* cells have been modified to include phosphatidyl choline, and these cells were suggested to have phosphatidyl choline in both inner and outer membranes, however this was not a wildtype system, and the dying technique was not shared however instead used for a method of outer membrane interrogation. Here for the first time, we study the distribution of propargyl-choline, a phosphatidyl choline mimic in wildtype bacterial cells, and indicate this method may have wider uses in microscopy.

Relevance of Pseudomonas Aeruginosa in lipid labelling experiments

Pseudomonas aeruginosa, is an especially pliable species to the interrogation of phopholipids, and their localisation and dynamics. It is a Gram-negative bacteria, with a connection between the outer membrane lipid leaflet formation and the peptidoglycan, worth investigating across bacteria, and many homologues to model organisms such as *Escherichia coli*. Its peptidoglycan formation process are also known, and is tractable with fluorescent *D*-amino acid mimics, so new labels would allow for study of both these processes simultaneously. Most essential to this study is that it is also one a few pathogenic bacteria that are regularly genetically manipulated, which also contain the lipid Phosphatidylcholine (PC). This lipid presence is relevant as mimics for this lipid have been developed for mammalian studies, and although also pliable for teichoic acid labelling in Gram-positive bacteria propargyl-choline has not yet been used in wildtype gram-negative bacteria for PC pulse-chase labelling.

Therefore, in the first half of the chapter, we formalise the use of propargyl-choline, a PC headgroup mimic, to determine the cytoplasmic membrane insertion localisation as well as further our understanding and inner leaflet formation and incorporation of PC into the outer membrane. The insertion of the lipid PC was uniform through growth, with no obvious increases in fluorescence locally. PCho shows to be of similar technical use to existing membrane labels, however with potential for tracking individual lipids rather than detecting or binding to lipids, and without need for membrane perturbation when using full phospholipid additionor use of mutant cells with alternative enzyme pathways to wildtype.

MCE domain containing proteins PqiB and MlaD

As mentioned in earlier chapters, part of the hypothesese investigated by this thesis was to suggest that the outer membrane and peptidoglycan process are linked. Therefore using fluorescent gene-fusion constructs of Waal an O-antigen ligase and MCE domain containing lipid transport proteins MlaD and PqiB mentioned in Chapter 1.2, required to form the outer membrane, their location during division and throughout the cell cycle was tracked, with a hypothesis that this would coordinate with division at the cell centre. These experiments conducted in *Pseudomonas aeruginosa*, as a collaboration with Dr Lori Burrows. Our experiments have since shown that among the MCE domain-containing proteins liable for mCherry tagging, the transport machinery PqiB and MlaD are instead of targeting the division

site, specifically localised to poles and longitudinally in *P. Aeruginosa*, MlaD also localised to the poles during division.

In this chapter, we also used propargyl-choline a new lipid dye to determine the inner leaflet formation and incorporation of PC into the OM during division, however we were unable to deconvolute the insertion of propargyl choline into the outer membrane specifically. This meant that use of the new dye was unable to provide any further insight.

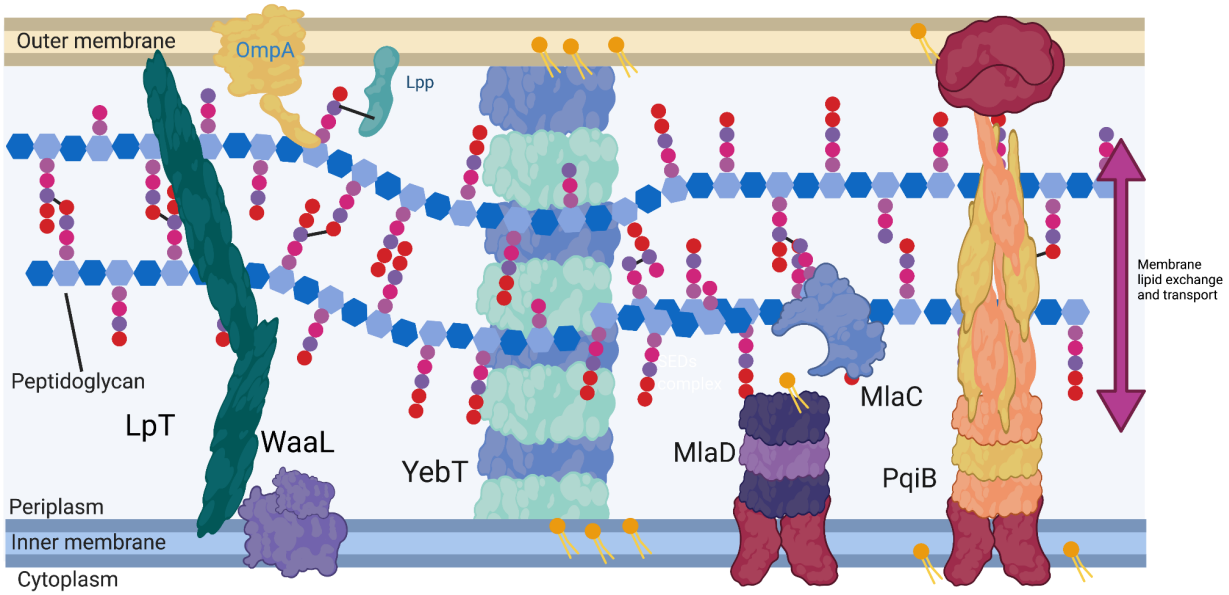


Figure 4.1.0 Outer membrane transport systems PqiB, MlaD, Lpt and YebT.

The peptidoglycan related proteins, have been tracked before in *E.coli* and *Pseudomonas* sp. to reveal a range of localisation specific roles which (Chapter 1.1) depending on their local environment changes their activity. Therefore the aim of this chapter, was to add the MCE domaining proteins in Gram negatives, (in addition to TULIP domain (lipid interacting domains which are soluble such as that within the LPT family)) to this categorization so that links may be interrogated and their roles understood. (Figure 4.1.1).

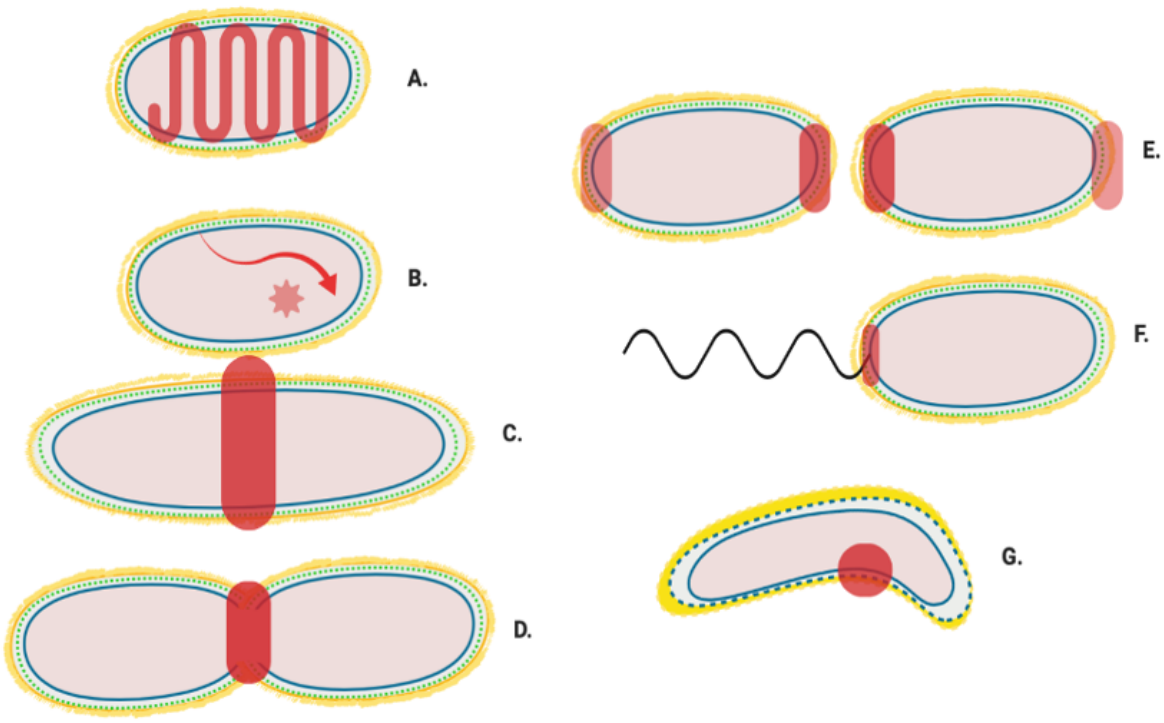


Figure 4.1.1 Division and elongation related localisations of Peptidoglycan

Localisation regions of known and potential peptidoglycan modifying enzymes in Gram-negative bacteria. Localisation sites are highlighted in red. **A** Helical and MreB associated Elongasome **B.** Free diffusion (unlocalised) **C**. Pre-septal machinery, **D**. Division machinery, **E**. Post septal polar machinery and polar growth, **F**. Flagella peptidoglycan modification machinery, and **G**. Shape determining pinpoint/seam

In this chapter, in conclusion it was found that despite the hypothesis that division coordinates with lipid insertion, we could provide no evidence to support this hypothesis, with both propargyl-choline concentrations and MCE domain protein localisations not correlating with division. However a new lipid dye was created, that will enable more research into the dynamics of membranes in bacteria.

Chapter 4: Methods

Fluorescent labelling of OM proteins *in vivo* - *P. Aeruginosa*

pEX18Gm plasmids were created through synthesis of geneblocks of the FtsW, MlaD, PqiB and YebT proteins, found through gene searching and use of online webtool “xBase” including 600bp flanking regions. The geneblocks were then cloned using gibson assembly, once amplified by PCR into an existing pEX18Gm plasmid. Once gibson cloning was successful for each vector, they were interrogated by sequencing and then once confirmed, the suicide vectors were transformed into *Pseudomonas aeruginosa*, with only meridiploid bacteria, thus genomically incorporating bacteria surviving an antibiotic challenge by gentramycin post transformation. The merodiploid conjugates were then purified by sucrose selection, and removal of the gentramycin challenge over successive generations to produce strains with the full genomic conversion of wildtype, to fluorescently tagged genes.

Imaging of fluorescence in *Pseudomonas aeruginosa*

Strains were streaked from glycerol stocks onto LB Agar plates, and incubated at 37°C overnight. One PA colony was grown overnight at 37°C, at 180 rpm in 2 ml LB. The following morning the a 1/10 dilution of the samples were then grown at 37°C, 180 rpm for a period of time until the samples had all reached an OD_{600} of 0.6, typically 2 hours. 1% agarose in PBS was heated in a microwave until piping hot. Microscope slides were topped with a solution of 1% agarose in PBS which was flattened with a coverslip and left to cool. The coverslip was removed and 10 µl of pre-prepared sample was added to the slide, the coverslip was placed on top to spread the sample across the slide. Samples were then analysed using confocal microscopy specifically to identify the fluorescence, with corresponding filters dependent on expected fluorescence. mCherry fluorescence was detected at 2 seconds laser exposure, with 0.1s exposure for B/W channels. The resulting images were taken in clusters of 15 across the sample at random to reduce bias, and allow for quantitive cell measurements.

HADA labelling

Cells were labelled with HADA (3-[[(7-Hydroxy-2-oxo-2H-1-benzopyran-3-yl)carbonyl]amino]-D-alanine hydrocholoride) a fluorescent D- amino acid as previously(153), for 2 minutes at 100μM at 0.1 OD, the cells were then washed with PBS solution twice by centrifugation.

Analysis of fluorescence localisation

In Figure 4.1.5 A, images are shown of cells, imported at LIFs or TIF libraries, which were then, analysed by MicrobeJ software to determine cell width and boundaries, (visually no filamentation). The points of increased fluorescence to background, were then tracked across the cell and mapped per individual point across thousands of cells, dependent on cell stage. The points were then mapped in Figure 4.1.5 C for each strain, giving a quantitave analysis of cells compared to controls. Automatic, cellular counting and size determination by MicrobeJ and BactMAP allowed for quantitative analysis. Fluorescent points were tracked using custom tolerance and intensity filters maintained throughout study. Scripts and conditions for image analysis attached in the Supplementary Data file. Correlation between fluorescent channels was found using the Colocalisation threshold feature of ImageJ.

R Script for BACTMAP fluorescent peak assignment

After installing BACTMAP on R, simply run the following two commands. Then play with outputs and groups.

```
     “
#First create a pixel conversion factor - found using the pixel conversion located in .lif imageJ meta data.

library(bactMAP)
Pseudo<- addPixels2um("Pseudo",0.065)
     meshmladstat<-
extr_MicrobeJ("/Users/u1760231/Desktop/PhD_Documents/Outer_membrane_project/mladstatcontour.csv","/Users/u1760231/Desktop/PhD_Documents/Outer_membrane_project/mladstatmaxima.csv")
```

```
#This calls my files from MicrobeJ contour (meshdata) and
MicrobeJ maxima (spotdata), which I have previously selected
using MicrobeJ in imageJ and exported into csv.

> plots_mladstat <-
createPlotList(spotdata=meshmladstat$spots_relative,
meshdata=meshmladstat$mesh, groups = 4, AllPlot = F, mag =
“Pseudo”)
```

#This gives an output of cells described in https://github.com/veeninglab/BactMAP/wiki/Subcellular-localization#createPlotlist “

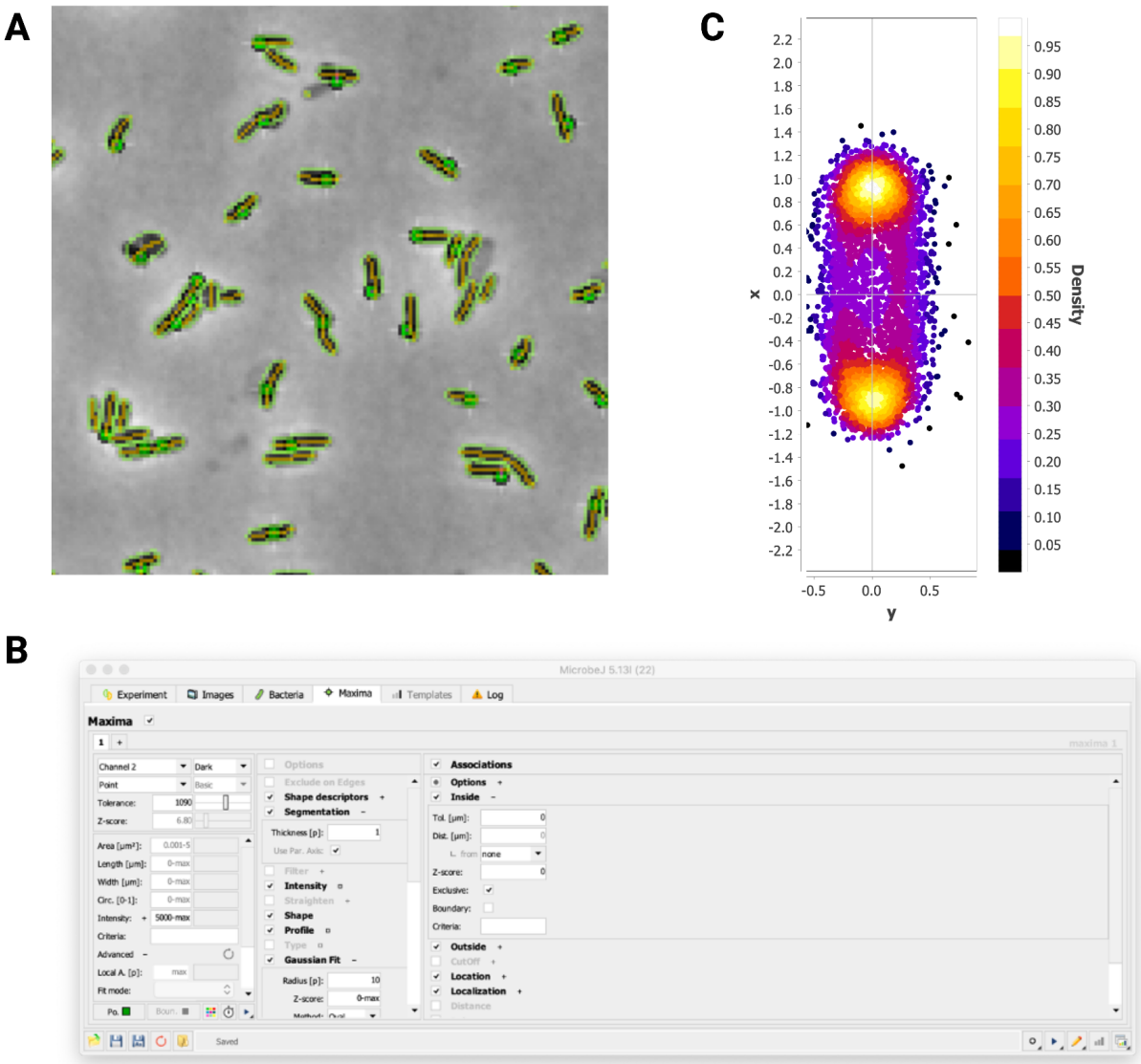


Figure 4.1.5 Fluorescent localisation analysis

A Automatic, cellular counting and size determination by MicrobeJ **B** Fluorescent points were tracked using custom Tolerance and intensity filters maintained throughout study, **C** fluorescent point frequency mapping to cell averages revealed hotspots of intensity. x- cell length, y- cell width.

Propargyl-choline incorporation

0.1 OD cells were incubated with 1mM propargyl-choline (Dissolved in DMSO) for 30seconds, 5 minutes and 40mins respectively. These cells were then concentrated by centrifugation for

10mins at 5000xg and fixed by 4% paraformaldehyde PBS for 30minutes . After fixation, the cells were washed by centrifugation at 5000xg by pelleting, and resuspended in 100mM Tris-HCL pH 8.8, 1mM CuS04 50µl and ascorbic acid 0.1M . The cells were then reacted with 100µM 598nm fluorescent azide, for a click chemistry reaction, then after 30 minutes at room temperature washed by centrifugation four times to remove the fluorescent azide remaining , before imaging.

TrackMatev6.02

Trackmate, implemented through a FIJI package was used to identify foci and track their movement overtime. Octanol at 1mM was compared to H20 only to determine if lipid speed was being tracked. Threshold 30,000, increased to 60,000, diameter 0.1µm. Linking distance 0.3µm, gap closing 0.3µm, Simple LAP tracking, LoG detector, subpixel localisation.

Transformation and Storage Solution transformation

0.3-0.6 OD Cells were mixed with an equal volume 2x TSS (2xLB, 20% Polyethylene glycol, 10% dimethyl sulphoxide pH 6.5 100mM MgCl2)(154) and 1µl of between 1ng and 200ng/µl plasmid added. After 20minutes at 37°C cells were plated onto appropriate selective media.

Chemical genomics assays

In evaluating phenotypic differences between the *Pseudomonas aeruginosa* mutants (ΔmlaD-mCherry, ΔpqiB-mCherry and Δwaal-mCherry) and the wildtype, a chemical genomics assay was conducted. In the chemical genomics assay, phenotypic observations under different stress conditions were linked to the genotypic differences between the mutants and the wildtype(162). The mutants and the wildtype were inoculated on LB agar plate with each mutant replicated 96 times . In addition, the mutants and the wildtype were inoculated on LB agar plates with each plate containing sub-minimum inhibitory concentration of different membrane stresses and antibiotics. Each condition plate has 4 replicas. The plates were incubated at 37^0C for 12 hours and images were taken afterwards. The images of the plates were subjected to an image analysing software 'Iris'(163). Iris identified each colony on the plate and measures the fitness in terms of size and opacity. The fitness of the mutants under different stress conditions were then statistically analysed and compared to that of wildtype with a software called ChemGAPP(164). ChemGAPP normalised the colonies around the edges of the plates and fitness scores were generated for each mutant.

Colony size data for each plate was checked for plate effects via a Wilcoxon rank sum test between the distributions of colony sizes for the outer two edges of the plate and the centre colonies. Where $p < 0.05$ the difference in the distributions was considered as statistically significant and the outer edge colonies were scaled to the plate middle mean (PMM). The plate middle mean is defined as the mean colony size of colonies within the centre of the plate, within the 40^{th} and 60^{th} percentile of size. Each plate was then scaled such that the PMM was equal to the median colony size of all colonies within the screen. Colony sizes greater than 10,000 were removed as outliers. For the bar plots and heatmaps, the average colony size was taken for each set of mutants within each plate and divided by the average colony size of the wildtype of the same plate to produce a fitness ratio. The average fitness ratios for each mutant in each condition was calculated between replicate plates for display. The 95% confidence intervals between replicates was calculated and displayed as error bars. For the swarm plots, each individual mutant colony within each plate was divided by the mean colony size of the wildtype

Chapter 4: Strains, plasmids and Primers

Table 4.0 Primers, Plasmids and Strains used for Fluorescence invstigation of Outer membrane proteins

Name	Sequence		Notes
Primers	Primer 1	Primer 2	Description of primer function
pEX18seq	GAT TAA GTT GGG TAA CGC CAG	GTG GAA TTG TGA GCG GAT AAC	Suicide Vector used for propagation in E.coli and genetic insertion in *Pseudomonas* sp.
Pseudogeneamplifier	AAG CTT GCA TGC CTG CAG G	GAA TTC GAG CTC GGT ACC CGG	A primer for the amplification of mCherry constructs
linpEX	CGG GTA CCG AGC TCG AAT TCG TAA TCA TGG	GTC GAC CTG CAG GCA TGC AAG CTT GG	Primers for amplification of the pEX18 plasmid
PqiBlin	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG	CCA AGC TTG CAT GCC TGC AGG TCG ACG GCC TTG CCG	Primers to linearise and amplify PqiB-

	CGC TTC GGC GTG CTG TTC CAG CAG	AAG GCG CTG ACC AC	mCherry for gibson cloning
linRodA	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG GTA CCG CCC AGG TGG TGG CGA TCA AGC AG	CCA AGC TTG CAT GCC TGC AGG TCG ACC GGG TAG TCG AAG CCG AGG GTG GAC AG	Primers to linearise and amplify RodA-mCerulean for gibson cloning
linFtsW	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CGT GAC GCC GGG CTG ATT GCC CAG	CCA AGC TTG CAT GCC TGC AGG TCG ACC CTC TGG CAA CAG CTT GTT CAG CGG TTC CG	Primers to linearise and amplify FtsW-mCerulean for gibson cloning
linWaal	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CCA ACT CGA CCG CAA CGA CCC GGA G	CCA AGC TTG CAT GCC TGC AGG TCG ACA TCG AGA TCA TCG ACT CCA ATG GCC GGG TCA TG	Primers to linearise and amplify WaaI-mCherry for gibson cloning
linYebT	CCA AGC TTG CAT GCC TGC AGG TCG ACA TCG AGA TCA TCG ACT CCA ATG GCC GGG TCA TG	CCA AGC TTG CAT GCC TGC AGG TCG ACG CTT GCT GCT GGT CGC CTA TGC CCT G	Primers to linearise and amplify YebT-mCherry for gibson cloning
linMlaD	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CAT TGG CAT GGT GCT GGC CCT GCA G	CCA AGC TTG CAT GCC TGC AGG TCG ACG GTC GTT GCG GTT CTT CTG CAT GGT GTC GG	Primers to linearise and amplify MlaD-mCherry for gibson cloning
Plasmid	Sequence		
pEX18	Gifted by Prof Lori Burrows		
FtsW-mCerulean_pUC57	GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGT CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGT TCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTA TTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAAC AACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCC GTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG ATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATT TAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCC TTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT		

	CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCG AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGA TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTT TGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG CACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCA CACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGGCTGCTCGC CGGCGGAGACGGCAAGGGCGCCGATTTCCATGACCTGCGCGAGCCGGTCGCGCGCTTCTGCCGGGCGG TGGTACTGCTTGGCCGTGACGCCGGGCTGATTGCCCAGGCACTGGGCAACGCGGTACCGCTGGTGCGC GTCGCAACGCTGGACGAAGCAGTCCGGCAGGCCGCCGAGCTGGCCCGCGAAGGCGATGCGGTGCTGTT GTCGCCGGCCTGCGCGAGCCTGGACATGTTCAAGAACTTCGAAGAACGCGGACGCCTGTTCGCCAAAG CCGTAGAGGAGCTAGCGTGATGCTGTCGGTGTTGCGCCCCTTCCCGTCGCCGCTGTTGAGCCGGCACGG CATCGACCTGGACTTCCCGCTGCTGGCGGGTTGCCTGGCGCTGCTCGGCCTCGGCCTGGTGATGGTGAC TTCCGCCTCTTCCGAGGTGGCGGCGGCGCAGTCGGGCAACCCGCTGTACTTCTCCGTCCGTCACCTGAT CTACCTGGTGATCGGCCTGATTTCCTGCGGCCTGACCATGATGGTCCCGATGGCCACCTGGCAGCGCTG GGGCTGGAAGCTGTTGCTGGTCGCCTTCGGCCTGCTGGTGCTGGTGATCACCCCCGGTATCGGCCGCGA GGTGAACGGCTCGATGCGCTGGATCGGCTTCGGCCTGTTCAACATCCAGCCTTCGGAGATCGCCAAGGT CTGCGTGGTGATCTTCATGGCCGGCTACCTGATCCGCCGCCAGCAGGAGGTCCGGGAAAGCTGGATGG GCTTCTTCAAGCCCTTCGTGGTGTTGCTGCCGATGGCCGGCCTGCTGCTGCGCGAGCCGGACTTCGGCG CCACCGTGGTAATGATGGGCGCTGCCGCGGCCATGCTGTTCCTCGGCGGGGTGGGGCTGTTCCGCTTCG GCCTGATGGTGCTGCTGGCGGTCGGCGCGGTGGTCCTGCTGATCCAGACCCAGCCCTACCGGATGGCAC GCCTGACCAACTTCACCGACCCCTGGGCCGACCAGTTCGGCGCCGGCTATCAGTTGAGCCAGGCGCTG ATCGCCTTCGGCCGCGGCGGCTGGTTGGGCATGGGACTGGGCAACAGCATCCAGAAGCAGTTCTACCT GCCGGAGGCGCACACCGACTTCGTCTTCGCCGTGCTTGCAGAGGAGCTGGGCATCGTCGGCGCCCTGG CGACCGTCGCCCTGTTCGTGTTCGTCAGCCTGCGCGCGCTGTACATCGGCATCTGGGCCGAGCAGGCCA AGCAGTTCTTCTCCGCCTACGTGGCCTACGGCCTGGCCTTCCTCTGGATCGGCCAGTTCCTGATCAACAT CGGGGTGAACGTCGGCCTGCTGCCGACCAAGGGCCTGACCCTGCCGTTCCTCAGCTACGGCGGCAGTTC GCTGGTGATCTGCTGTGCCTGTCTGGGGATGTTGCTGCGCATCGAATGGGAACGGCGAACCCACCTGGG TAGCGAGGAATATGAGTTCAACGAAGAGGACTTTGCCGATGAAAGGTCGGGCTCCGGAAGCGGGATG GTGTCGAAAGGCGAAGAACTGTTCACCGGCGTCGTGCCGATCCTGGTCGAGCTGGATGGCGATGTCAA TGGGCACAAGTTCTCGGTGTCGGGGGAGGGCGAGGGGGATGCGACGTACGGCAAGCTCACCCTGAAAT TCATCTGCACCACCGGCAAACTGCCCGTGCCCTGGCCCACCCTCGTGACGACCCTCACGTGGGGGGTGC AATGCTTCGCGCGCTATCCCGATCACATGAAACAGCATGACTTCTTCAAGAGCGCGATGCCCGAAGGG TATGTCCAGGAACGGACGATCTTCTTCAAAGACGATGGGAATTATAAAACGCGCGCCGAGGTCAAGTT CGAAGGGGACACCCTCGTGAATCGGATCGAACTCAAGGGGATCGATTTCAAGGAGGACGGGAACATCC TCGGCCACAAACTCGAGTACAACGCGATCTCCGATAATGTGTATATCACGGCGGACAAGCAAAAGAAT GGGATCAAAGCCAATTTCAAGATCCGCCATAACATCGAAGATGGGTCCGTCCAACTGGCGGATCATTA CCAGCAGAACACGCCGATCGGCGATGGGCCGGTCCTCCTCCCCGATAACCACTACCTGAGCACCCAGA GCAAACTCTCCAAAGATCCCAACGAGAAGCGGGATCATATGGTCCTCCTGGAGTTCGTCACGGCGGCC GGCATCACCCTCGGCATGGATGAACTGTATAAATCCGGCTCCGGAAGCGGCTAATGTCCTGATCATGGC GGGTGGCACCGGCGGACACGTGTTCCCGGCGCTCGCCTGTGCGCGGGAGTTCCAGGCGCGCGGCTATG CCGTGCACTGGCTGGGGACGCCGCGCGGCATCGAGAATGACCTGGTACCCAAGGCCGGCCTGCCGTTG CACCTGATCCAGGTCAGCGGGCTGCGCGGCAAGGGCCTGAAGTCGCTGGTCAAGGCGCCGCTGGAACT GCTCAAGTCGCTGTTCCAGGCGCTGCGGGTGATCCGCCAGCTGAGGCCGGTCTGCGTGCTCGGCCTGGG TGGCTATGTCACCGGCCCGGGCGGCCTGGCTGCGCGCCTCAACGGCGTGCCGCTGGTGATCCACGAGC AGAACGCCGTGGCCGGCACCGCCAACCGCAGCCTGGCGCCGATCGCCAGGCGCGTCTGCGAGGCATTC CCGGATACCTTCCCGGCCAGTGACAAGCGTCTGACCACCGGCAATCCGGTGCGCGGCGAGCTGTTCCTC GACGCGCACGCCCGCGCGCCGCTGACCGGCCGTCGGGTCAATCTGCTGGTGCTCGGCGGCAGCCTTGG CGCGGAACCGCTGAACAAGCTGTTGCCAGAGGCCCTGGCAGGATCCCCGGGTACCGAGCTCGAATTCA CTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCA CATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGC AGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGC ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACA CCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCC

	GGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
RodA-mCerulean_pUC57	GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGT CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGT TCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTA TTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAAC AACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCC GTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG ATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATT TAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCC TTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCG AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGA TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTT TGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG CACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCA CACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGTCGTGCTGC GCGACCCGGCCAACTGGGACCGCGTCGACTACGGCATGCAGCAGGTGGTGCATGGCGCCCGCGGCACG GCGCGCAAGGTCGGCGCCACCTCGGCCTACCTGATCGCCGGCAAGAGCGGTACCGCCCAGGTGGTGGC GATCAAGCAGAACGAGCGCTACGACCGCTCCAAGCTGCTCGAGCGGCACCGCGACCACGCCCTGTTCG TCGGCTTCGCGCCGGCCAACAACCCGCAGATCGCGGTGGCGGTGATGGTGGAGAACGGCGAGTCCGGC TCCGGCGTCGCCGCCCCGGTGGTCAAGCAGGTGATGGACGCCTGGCTGCTCGACGAGCACGGCAAGCT CAAGGCCGAATACGCGGAACCCGTCGCGCCGCTCGCGGCGGCGGTGGCCAAGCCCGAACCCACGGCA GCCGAGCCGGAGGCGCCCGCCCTTGAACAGTAACTTCGACCGCACCCTGTCCAACGAAGACGTGATGC GCCGGCGCGCCAGCCTGTTGCAGCGCCTGCACATCGACGGGATTCTCCTGCTGCTGTTGCTGATCCTCG GCGCCACCGGCCTGTTCGTCCTCTATTCGGCGAGCGGCAAGAGCTGGGACATGCTGATGAAGCAGGCC ACCTCGTTCGGCCTGGGACTGGGCATGATGGTGGTCATCGCCCAGATCGAGCCGCGCTTCATGGCCCGC TGGGTGCCGCTGGGCTACCTGGTGGGGGTCGCGCTGCTGGTGGTGGTCGACGTGATCGGCCACGACGC CAAGGGGGCGACGCGCTGGATCAACATCCCCGGGGTGATCCGCTTCCAGCCCTCGGAGTTCATGAAGC TGCTGATGCCGATGACGGTGGCCTGGTACCTGTCCAAGCGCAACCTGCCGCCGGGGCTCAAGCACATG GTGATCAGCCTGGCGATCATCATCACCCCGTTCGTGCTGATCCTGAAACAACCCGACCTCGGCACCGCC ATGCTGATCCTCGCTTCCGGCGGCTTCGTGCTGTTCGTCGGCGGCCTGCGCTGGCGCTGGATCGTCGGC GCGGTGTCGGCGGCGGTGCCGATCGCCGTGGCGATGTGGTTCTTCATCATGCACGACTACCAGAAACA GCGGGTGCTGACCTTCCTCGACCCGGAAAGCGACCCGCTGGGCACCGGCTGGAACATCATCCAGTCTA AGGCGGCCATCGGTTCCGGCGGGGTATTCGGCAAGGGCTGGCTGCTAGGCACCCAGTCGCACCTGGAT TTTTGCCGGAAAGCCACACCGACTTTATCATTGCGGTCCTCGGCGAGGAGTTCGGCCTGGTCGGCGTC TGCCTGCTGCTGGTGCTCTACCTGCTGCTGATCAGTCGTGGCCTGGTGATCACCGCCCAGGCGCAGACG CTCTACGGCAAACTGCTGGCCGGTGGCATCACCATGACGTTTTTCGTTTATGTCTTCGTGAATATCGGTA TGGTGAGCGGGTTGTTGCCGGTGGTGGGGGTACCGCTGCCTTTCATTAGTTATGGCGGAACTTCGCTGG TTACCCTGTTGTCAGGGTTCGGGGTTCTGATGTCGATCCATACCCATCGGAAGTGGATCGCCCAGGTTT CGGGCTCCGGAAGCGGGATGGTGTCGAAAGGCGAAGAACTGTTCACCGGCGTCGTGCCGATCCTGGTC

	GAGCTGGATGGCGATGTCAATGGGCACAAGTTCTCGGTGTCGGGGGAGGGCGAGGGGGATGCGACGT ACGGCAAGCTCACCCTGAAATTCATCTGCACCACCGGCAAACTGCCCGTGCCCTGGCCCACCCTCGTGA CGACCCTCACGTGGGGGGTGCAATGCTTCGCGCGCTATCCCGATCACATGAAACAGCATGACTTCTTCA AGAGCGCGATGCCCGAAGGGTATGTCCAGGAACGGACGATCTTCTTCAAAGACGATGGGAATTATAAA ACGCGCGCCGAGGTCAAGTTCGAAGGGGACACCCTCGTGAATCGGATCGAACTCAAGGGGATCGATTT CAAGGAGGACGGGAACATCCTCGGCCACAAACTCGAGTACAACGCGATCTCCGATAATGTGTATATCA CGGCGGACAAGCAAAAGAATGGGATCAAAGCCAATTTCAAGATCCGCCATAACATCGAAGATGGGTCC GTCCAACTGGCGGATCATTACCAGCAGAACACGCCGATCGGCGATGGGCCGGTCCTCCTCCCCGATAA CCACTACCTGAGCACCCAGAGCAAACTCTCCAAAGATCCCAACGAGAAGCGGGATCATATGGTCCTCC TGGAGTTCGTCACGGCGGCCGGCATCACCCTCGGCATGGATGAACTGTATAAATCCGGCTCCGGAAGC GGCTGAGACCAAGAGTGAAGAACGCAATGCAAGTACTGCGTACATGGGCGGCCAGGGGCGTCCAATG GGTCGGCGTAGCCGGCGTCATTGGCCTGTCCGGGGCGGCCCAGGCGGGGGACTACGACGGCTCGCCGC AAGTGGCCGAGTTCGTCAGCGAAATGACCCGCGACTACGGCTTCGCCGGAGAGCAGCTGATGGGGCTG TTCCGCGACGTGAACCGCAAGCAGTCGATCCTCGATGCGATCTCGCGCCCGGCCGAGCGGGTCAAGCA GTGGAAGGAATACCGGCCGATCTTCATCAGCGACGCGCGCATCAGTCGTGGCGTCGACTTCTGGAACA AGCATGCCGAAGACCTGGCGCGGGCGGAGAAGGAATACGGCGTGCCGGCCGAGATCATCGTCTCGATC ATCGGCGTGGAAACCTTCTTCGGCCGCAACACCGGCAGTTACCGGGTGATGGACGCGCTGTCCACCCTC GGCTTCGACTACCCGCCGCGGGCCGACTTCTTCCGCAAGGAGTTGCAGGATCCCCGGGTACCGAGCTCG AATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTT GCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACA GTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTC ACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCC GCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGAC CGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
PqiB-mCherry_pUC57	GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGT CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGT TCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTA TTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAAC AACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCC GTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG ATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATT TAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCC TTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCG AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGA TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTT TGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG CACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCA CACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGTGCGGGTG TTCGGCGAAGACCTGCTGCAACTGCCAGAGGAGCGCCGCTCCCTGGTCGAACGGCGCTTCGGCGTGCT GTTCCAGCAGGGCGCGCTGTTTTCCTCGCTTACCGTGGTGGAGAACGTCGCCCTGCCGCTGATCGAAAA

	CGCCGGCCTGCCGCGCGCCGACGCCGAGCACCTGGCCCAGGTCAAGCTGGCCCTGGCCGGGCTACCGG CCAACGCCGGCGACAAGTATCCGGCCTCGCTTTCCGGCGGCATGATCAAGCGCGCCGCCCTGGCCCGC GCCCTGGCGCTGGACCCGGACATCCTGTTCCTCGACGAGCCCACCGCCGGCCTCGACCCGATCGGCGCG GCGGCCTTCGACAACCTGATCCGTACCCTGCGCGACGCCCTCGGCCTGACCGTGTTCCTGGTCACCCAT GACCTGGACACCCTGTACACCATCTGCGACCGGGTCGCCGTGCTGTCGCAGAAGAAGGTCCTGGTAGT CGATACCCTGGAACGGGTCGCCGCCACCGACGACGCCTGGGTCCGCGAATACTTCCACGGGCCGCGCG GGCGCGCCGCCCACCAAGCCGCCGCCGTGCCGGAGAATCACTGAGATGGAAACCCGAGCCCACCATGT GTTGATCGGCCTGTTCAGCGTGATCGTGATCGGCGCCGCCCTGCTCTTCGGCCTGTGGCTGGCCAAGTC CGGCTCGGAGGGCAAGTTCAACTACTACGACATCGTCTTCAACGAAGCGGTCAGTGGCCTTTCCCAGG GCAGCTCGGTGCAGTACAGCGGGATCAAGGTCGGCGACGTGGCCTTCCTCCGCCTCGATCCGAAGGAC CCGCGCAAGGTCTGGGCACGCATCCGCGTGGTGGCCAGCGCGCCGATCAAGCAGGACACCACCGCCAA GCTGGCGCTCACCGGGATCACCGGCACCTCGATCATCCAGCTCAGCAGCGGCACGCCCGCCAGCCCGA TGCTGGAGGGCAAGGACGGGAAGATCCCGGTGATCGTCGCCACGCCCTCGCCGCTGACCCAGTTACTG AGCAACGGCGAAGACCTGATGGGCAACATCAACCAGCTCATCGCGCGCTTCAGCAACCTGCTCTCCGA GGAAAACACAGCGCGCATCAGCCGCACCCTCGACCATCTCGACCAGGCCACCGGCGCGCTCTCCGCCG AGCGCGAGAACGTCAGCGCGGTGATGCAGCAGTTGGCCCAGGCGAGCAGGCAGGCCAACGCCGCCCT GGCCCAGGCCAGCGAGTTGATGCGCAGCGCCAACGGCCTGCTCAACGAGCAAGGCAAGGGGATGCTG GAAAACGCCAACAAGACCATGGCTTCGCTGGAGCGCACCAGCGCCACCCTCGACCAGTTGATCAGCGA GAATCGCCACTCGCTGGACGGCGGCATCCAGGGACTCGCCGAGCTGGGCCCGGCGGTCAGCGAACTGC GCGATACCCTGGCGGCACTGCGCGGAATCAGCCGGCGCCTGGAAGAGAACCCGGCGAACTACCTGCTG GGCCGGGAAAAAACCAAGGAGTTCACCCCATCGGGGTCCGGATCGGGGATGGTCTCGAAGGGGGAGG AAGACAACATGGCCATCATCAAAGAGTTCATGCGGTTCAAGGTCCATATGGAAGGCAGCGTGAATGGC CACGAATTCGAGATCGAAGGGGAGGGGGAGGGGCGGCCGTATGAAGGGACCCAAACCGCGAAACTGA AGGTGACCAAAGGGGGGCCGCTGCCGTTCGCGTGGGATATCCTGTCCCCGCAATTCATGTACGGCTCG AAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACTACCTCAAGCTGTCCTTCCCCGAGGGCTTCAA ATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGACGGTGACCCAGGATAGCAGCCTCCAGG ATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGGACCAACTTCCCGTCCGATGGGCCCGTCATGCAA AAGAAGACGATGGGCTGGGAGGCCTCGAGCGAGCGGATGTATCCCGAAGACGGGGCCCTCAAAGGGG AGATCAAGCAACGCCTGAAGCTGAAAGATGGGGGCCATTATGATGCGGAAGTGAAGACGACCTACAA AGCCAAGAAACCGGTGCAACTGCCGGGCGCGTACAATGTCAATATCAAACTGGATATCACGTCCCACA ATGAAGACTATACCATCGTGGAACAGTATGAACGCGCCGAAGGGCGGCACTCGACGGGGGGGATGGA TGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGGCTCGCCCTTCGCCCGCTGCGTCGCCTGTCCCTG GCCGCCGGCCTGGCCGCGCTCGCCACCCTCGGCGCCTGCTCGATCCTGCCGGAGGCGCAGGTCCTCCAG GTCTACCTGCTACCGGTGCACAACCCTCCGGCCAGCGCCGCCGCGCGGCCGGTCGACTGGTCGCTGCGG ATCGCCCGGCCGCGTACCAGCCTGGTGCTGGAGAGCCCGCGCATCGCGGTGCGCCCGCACGGCGACGA AATCAGCGTGTACCAGGGCGCGCGCTGGAGCGATCCGGCGCCGTCGCTGCTGCGCGATCGTCTGATGC AGGCGTTCCAGGCCGACGGCCGGGTCCGCGGCCTGAGCAGCGACGACAGCAACCTGCAGGCCGATTTC GAACTGGGCGGCGACCTTCGCGCCTTCCAGACCGAGTACCCGAACGGCCAGGCCAGCGCGCTGATCCG CTACGACGCGCGCCTGGTGCGTACCGACGACAAGCGCGTGGTCGCCAGCCGGCGTTTCGAGGTCAGCC AGCCGGTGGATGGCAAGAAGGTCGCGGCGGTGGTCAGCGCCTTCGGCAAGGCCGGCGATACGCTGTCG GCCCAGGTTCTCGACTGGAAGGATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTTTACAACGT CGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGG CGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCG CCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACA ATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGG GCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG TTTTCACCGTCATCACCGAAACGCGCGA
Waal-mCherry_pUC57	GCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCA GAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA CCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTC CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCAAGCTTGCAT GCCTGCAGGTCGACTCTAGGGATCCCACATATCCAGCGCGACGCGGTGCACTACGAGCCGCTACCGGG GCAGACCCTCCGCCAACTCGACCGCAACGACCCGGAGTGCGGCGACGCAGTCCGGACGCAACTGGGCG AGCTGATCGCCGCGCTGCACGCCAAGGGGGTGTACTTCCGCTCGCTGCACCTGGGCAACGTGGTGCGC ACGCCGGAAGGCCGGATGGGCCTGATCGACATCGCCGACCTGCGGGTCCAGCGCTCGGCGCTCAGCGC CGCCCAGCGCATTCGCAATTTCAAGCACCTGCTGCGCTACGAACAGGATCGAACCTGGTTCCTCGACGA CAGCGAGCACCTGGTGCTCAAGGCCTATCTCCGCAGCAGCCAGGTCCACTGGACACTCGAGCAGCTCC TCAAGCAACTGCAACTGGATTGAGCAAGGAATGTTCGCAGCTACCCGACTGTCCCGTCTTCGTCACGAT

	ACCAGCCGCATCCTCAGCCACTGGATCCTGCCGCTGGGCTGGCTGGCCCTGCTCACCGGAATGTTCTGG GTCGGCGACCGTTCCGACTATCACCGGCTTTTCTATATCTTGCTGGCTGCACCGACGCTGCTGTACGTGA TACTGCAACCGCGTCTGCTTCGCCCGCTAACGGGCTCGCCGCTGTTCATCGCCTTCCTCGCCTTCAGCAG CTACATGATGCTGAGCCTGTCCTGGTCGACCCCCGAGAACTCCACCGGCTCGCTGCTCAAGCGCCCCCT GTACATCGCCCTCCTGTTCTTCTGCGCCGCCATCCTGGCGCTCGAAGCTCCCCTCCGCCTCAAGACCGCG ACCTGGCTGGCAGCCCTTGGCGCGGTGATCTCCGCGGCAGCCACCTTGCTGCGATACTACTGGGACGCC AACCCGCTGCGCCTTACCGGTTATGGCGCTCTCTATAACCCATTACTGAGCGCCCATGTCTATGGCGCC TTCACGGCACTCTGGCTGGCCTATTGGATGCAATCCCGCCCCATCCTGGCCCCCTTGCCACTGATATCTC TGGCGCTGCTCGGTGGCCTTCTCATCGCGACCGGTTCACGTACTCCCCTGGTAGGGCTCACAGCGGCCC TTATGTGGCTGGTCCTGGCCGGAGACAGGAAAAAAGCCCTCATCGCCCTGGCACTCGCCCTGGCTGGA GCGCTACTGGGCTACATCCTGTACCCGGAAGTGATCACCCAGAGAGGGGCATCGTTCCGCCCGGAAAT CTGGGCCGACGCCCTACGCCAGATCAGCGAGCATCCATGGCTCGGACACGGCTACGATCATCCGATGC GAATCGTCCTGAGCAACGGCATGCTGCTCGCCGACCCGCACAACATCGAACTGGGTGTGCTTTTCGCTG GAGGGATCATCGGGCTCCTGCTGTGGGTGGCGATCTACGCACTGGCCTTCGGCTTCTCCTGGAAAAACC GGAAGTCTCCAGCCGTTCTGCTAGCCTCGACCTGGCTGGTGTTCGGCCTGGCAGCCGGACTCACGGAGG GCAACGCCTTCCTGCCCCGTCCCAAGGAACACTGGTTCCTGATCTGGATTCCCATGGCGCTGCTGTATG CGCTGTGGATCCAGCAAAGGTTCGCCGCCAGCAGGCGCGGTGAGGATATCGCCGCGCCTTCGGGGTCC GGATCGGGGATGGTCTCGAAGGGGGAGGAAGACAACATGGCCATCATCAAAGAGTTCATGCGGTTCAA GGTCCATATGGAAGGCAGCGTGAATGGCCACGAATTCGAGATCGAAGGGGAGGGGGAGGGGCGGCCG TATGAAGGGACCCAAACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCTGCCGTTCGCGTGGGATAT CCTGTCCCCGCAATTCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACTACCT CAAGCTGTCCTTCCCCGAGGGCTTCAAATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGA CGGTGACCCAGGATAGCAGCCTCCAGGATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGGACCAAC TTCCCGTCCGATGGGCCCGTCATGCAAAAGAAGACGATGGGCTGGGAGGCCTCGAGCGAGCGGATGTA TCCCGAAGACGGGGCCCTCAAAGGGGAGATCAAGCAACGCCTGAAGCTGAAAGATGGGGGCCATTAT GATGCGGAAGTGAAGACGACCTACAAAGCCAAGAAACCGGTGCAACTGCCGGGCGCGTACAATGTCA ATATCAAACTGGATATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGCCGAA GGGCGGCACTCGACGGGGGGGATGGATGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGTGCGGC GATATCAGCGATCAGGAGAAAAACAGCCGCGCAATGCCGCGGAAGGTCTTCTCGTTCCAGTACTTCAG CGGCAAGCCGGAGAGCAGCTCCCGGGCCAGTACCTTGTCCCGGTTCGAGCACTTGAGGAACATCGAAT TGCGGAAGCGCATGATCACCTGCTCGTAATCGGGGTGATCGCTGAACTGGCTGTACGTCTTGAATACGT TGTCCACCATGAAACGGGCGTTCTTGTACGTATTGGTGGGATGCTTGCGGTACTGGGCGAGGACCTCGC CAAGAATGTCGATGACGTAGCCAGCCTTGGTCACCGCCAGTTCGATATAGACATCTTCCAGGCGTATGT CCGGATCGAAACCGCCGACCTTCTCCAGGGCTTCCCGCCGGAACATCAAGGTGGGCGCCATCGGGCCC GGCTTGCGGTCCAGGAACAGGTCATCGAAGTCGAGACGGCGGAAAGGCAGGTCGCGTTTGCGCTGCTC CTTTCCCGGCATGACCCGGCCATTGGAGTCGATGATCTCGATATTGCCAGGATCCCCGGGTACCGAGCT CGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTG TTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAAT GAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC AGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTT TGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCG TTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTA ATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGC ATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATA TTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTC AATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAG AAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAAC GGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGG ATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTAC GGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAA ATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTC CATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCA TTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGC TTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATT CGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAG TGACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACG TTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTATAC CGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTACACGAAGGTT TTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCGATG CTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATATGCT

	GTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAAAAT CTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTT CTGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTC GTGCGGTGTTACGTTGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAA CCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCTCG GCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTCGA AGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTCAC TGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGAGG GGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTT TACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAG GGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAG GGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGC CCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTAA CCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCC GGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAA AACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGATC TTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCC TTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAA ATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGA CAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGA AACGTAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGG GATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACAT GACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGT AAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGA TGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGG GCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAG TTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAG TATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTG CCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAAC AGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGTATT TTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGGGTG GTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATG GATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGAT TTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGAT TATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGA TATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGC TCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTGAGGTATGTGCTCTTCT TATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTG TTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATG AAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGC ACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCG GATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGA TATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGCGTG ATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATA AAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTG TTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTC ATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAA AGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTT AATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAA CTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATG TCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCA GTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATT TTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGC CTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGA TGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTTCTA TTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAA GCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACC AGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTT GCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTC TCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCATCC

	CTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTG CCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGC TGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAAC GTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGG AAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTC TTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAAAG AGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAGCT GAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCA CGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAG GTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCAC CCTGATGCGACCACGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAG CATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTGG GACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGG TCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGT CCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATT AGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTA TCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGAC CACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAG ACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACCT GTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCA CAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGG CTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCA GTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCT GCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATCA TTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTA TTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATA CCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTCT GTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTAT ACGAAGTTATATTCGAT
YebT-mCherry_pUC57	GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGT CAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATAT GTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTA TTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAA ACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCT CAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGT TCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTA TTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCG GAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGG AACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAAC AACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGAT GGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAA ATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCC GTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAG ATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATT TAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCC TTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTT CTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTG CTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGA TAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCG AACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGA GAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGA TGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTT TGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTT TGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG

	GAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGA CAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGG CACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCA CACAGGAAACAGCTATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGGCACCTGG GCGTTGCTGATCACCGCGGCGATTCTCTACATACCCGCCAACCTGTTGCCGATCATGACGGTCAACCTC TTCGGCAGCGGAATGCCCGCCACCATCATGGAGGGCGTGGTCGAACTGGTGCATGCCGACATGTTCCC CATCGCCATGGTGGTGTTCGTCGCCAGTATCCTGGTTCCTACTTTCAAGCTGGTGGGCATCGCCCTGCTG CTCTATTCGGTACAGCGGCACCAGCCGATGTCGGCCCGCCAGCGTATCCTGATGTATCGCTTCATCGAA TGGATCGGGCGCTGGTCGATGCTCGATATCTTCGTCATCGCCATCCTCGTGGCGCTGGTGAATTTCGGC AACCTGGCGAGCATCGAGGCCAACCTCGGCGCCGCCGCCTTCGCCAGCGTGGTGGTGCTCACCATGCTC GCCGCCGTGACGTTCGACCCCCGGCTGATCTGGGACAACACCAATCTGGATGACGAGAATGAGTGATC TGCCAAGTCCGAAGAAACACAAGACCTCGAACTGGTCGGCGATCTGGGTGCTGCCGCTGGTGGCCCTG GCCATCGGTGCCTGGCTGGCGTGGCGTGCCTTCGACCAGGCTGGCGTGGACATCCAGGTTCGCTTCGAG AGCGGCGACGGGATCCAGGCCAACAAGACCGAGGTGCTGTACAAGGGCATCTCGGTGGGCAAGGTGA CCGACCTGCATGTCAGCAAGGACATCAAGGGCGTGGTCGCGACCATCGAGATCAAGAAGGAAGCCCA GGAGTACCTGAGCAAGGACACCCGCTTCTGGCTGGTCAAGCCGCGGGTTTCCCTGGCTGGCGTCACCG GCCTGGAGACCCTGGTCTCCGGCGTCTACATCGCGGTCGATCCGGTGAAGGGCGAGAAGGAAGAACGC TACTTCACCGCGTTGAAGGAGCCGCCACCCCTTTCCGACAAGCTGCCCGGCCTGCACCTGACCCTCAAG GCCGATCGCCTGGGCTCCCTGGAGCAGGGCAGCCCGGTGTTCTACCGGCAGATCCAGGTCGGCCAGGT GAAGAGCTTCCAGTTAGGCGATGACCAGAGGACCATCGAGATCAAGGTGCACATCGAGCCGGCCTACG CCGACCTGGTGCGCAAGCACACGCGATTCTGGAACGCCAGCGGCATCTCCATTTCCGGCGGCCTGTCCG GCTTCAAGGTGCACAGCGAGTCGCTGCTGACCCTGGTCGCCGGCGGTATCGCCTTCTCCACCCCGGAGA ACCGCACCGACAGCCCGCCGACCGATCCGAGCAAGCCGTTCCGCCTGTACGACGACTACGACGCGGCC CAGGCCGGTCTGCGCGTGAAGCTGAAGATGAACGACGTCAGTGGTATCGATCCGGGCCGCACTCCGGT GATGTTCAACGGCGTGCAGGTCGGCCTGGTCAAGTCGATCGACATGGGCAAGGACTACAGCAGCGCCA CTGCCGACCTGGCGATGGATCCGCGGGTCGAGGACATGCTGCTGGAAGGCACCGAATTCTGGACCGTC AAGCCAAGCATCTCCCTGGCCGGCATCACCGGCCTCGAGGCCCTGGTGAAAGGTAACTATATCGACGT GCGCTTCGCCAAGAGCGGCGCGCCGAGTCGCGAGTTCACCATCCGCCCGAAAGCGCCGCCGCTGAACA CCGACGCGCCCGGCCTGCACCTGGTGCTGACCAGCGACAAGCTCGGCTCGATCGATATCGGCGCGCCG ATCCTCTACCGCCAGGTCCGCGTCGGCAGCGTGCAGAGCTACCAGCTGTCGCGCGACCGCCAGCGAGT GGTGGTGGGTGTGCACATCGAGCCGGAGTACGCGCACCTGGTGAATACCTCGACGCGCTTCTGGAACT CCAGCGGGATCACCCTGACCGGCAACCTCAGCGGCGTGCAGGTGAAGAGCGAGTCGCTACAGACCCTG ATCACCGGCGGGATTTCCTTCGACACCCTCGATCCGAAGGCGCCGACCGTGACCAAGGTGCGTCGCTTC ACCCTCTTCGACAGCGAGGAGGCGGCGATGGCCCGTGGCGTGGAAATCCAGCTGAGCATCGACAATGC CGACGGCCTGCGCGAAGGCACGCCGATCCGCTTCAAGGGCCTGGACATCGGCAAGATCGAAAGCGTCG AGCTGAACCCGGACCTCTCCGGGGTACTGATGAAGGCCCGCCTGACTTCCGCCGGCGAGCGCGTCGCC CGCAGCGGAACCCGCTTCTGGGTGGTGCGTCCGGCGCTCGGCCTGCTGCGTACCGAGAACCTCGGAAC CCTGGTCAGCGGCCCCTATATCGAGGCGCTGCCGTCGAGCACGCCGGGCGAGCGCCAGGCACGCTTCC AGACTCTCGCCGAAGCACCCAACCTGCTGGGTCGGGAAAACGGCCTGCGGCTGACCCTCAGTGCGCCG CGCAAGGGCTCGATCAAGCCGGGCAACCTGGTGACCTACCGGCAGATCCCGGTGGGCAAGGTCGTCGA CCTGGCCTTGGGCGAACAGGCCGACCGGGTGCTGATCAGCATCCTCATCGAACCGCGCTACGTACCGCT GGTGCGTACCGGCAGCCGTTTCTGGAATGCCAGCGGCTTCGGCGTCGACGCCAGCCTGTTCAAGGGCCT GTCGCTACGCACCGAGTCGATGGAGGCGCTGATGGAGGGCGGCATTGCCTTCGCCACCCCGAACAATG CGCAGATGGGCGAGCCCGCCAAGCCGGGGCAGACCTTCGCCCTGTTCGATTCGGCCAATGACGAATGG TTGGAGTGGGCGCCGAAGATTCCTTTGAAGGAAACCGCCCGCCGCTCGGGGTCCGGATCGGGGATGGT CTCGAAGGGGGAGGAAGACAACATGGCCATCATCAAAGAGTTCATGCGGTTCAAGGTCCATATGGAAG GCAGCGTGAATGGCCACGAATTCGAGATCGAAGGGGAGGGGGAGGGGCGGCCGTATGAAGGGACCCA AACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCTGCCGTTCGCGTGGGATATCCTGTCCCCGCAAT TCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACTACCTCAAGCTGTCCTTCC CCGAGGGCTTCAAATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGACGGTGACCCAGGAT AGCAGCCTCCAGGATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGGACCAACTTCCCGTCCGATGG GCCCGTCATGCAAAAGAAGACGATGGGCTGGGAGGCCTCGAGCGAGCGGATGTATCCCGAAGACGGG GCCCTCAAAGGGGAGATCAAGCAACGCCTGAAGCTGAAAGATGGGGGCCATTATGATGCGGAAGTGA AGACGACCTACAAAGCCAAGAAACCGGTGCAACTGCCGGGCGCGTACAATGTCAATATCAAACTGGAT ATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGCCGAAGGGCGGCACTCGAC GGGGGGGATGGATGAACTGTATAAAAGCGGGTCCGGGTCCGGTGATTGTCGGACCGGCAAAAAGAAA CGCCGCTTACCTGGGTAAGCGGCGTTTTCGTTCATGCCTGCCTGCTGGGGAGTGGGTGTCAGTGAGTGC TGAGATGCGCCCGGGTGGTCAGCAGGTACACCGGCAAGCCGGAGATCAGGATCAGCAGCGCGGCGTA GGGTGCGGCGGCAGCGTACTCCAGGTTGGCGGTGTGCACCCAGACCGAGGTGGCCAGGGTCGACAGCC CGGTGGGACCTAGCACCAGGGTGGCGGTCAGTTCCTTCATGGTGTCGAGGAACACCAGCACGAAGCCG

	GCGCCGATGGCCGGCGAGATGATCGGCAGGGTGATCCGCAGGAAGGCCATCAAGGGCGTCTGGCCGA GAGTGCGGGCGGCTTCCTCCAACTGCGGCGAGGCCTTTTCCAGGGCCACGCGGATCGGTGCCTGGGTC AGCGGCATGAACAGCAGGGCATAGGCGACCAGCAGCAAGCCGGTGGTCTAGGATCCCCGGGTACCGA GCTCGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATC GCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCC AACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTA TTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGAC ACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTG TGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA
MlaD_pUC57	GCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCA GAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA CCGCATCAGGCGCCATTCGCCATTCAGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCG CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTC CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCAAGCTTGCAT GCCTGCAGGTCGACTCTAGGGCCTGTTCATTGGCATGGTGCTGGCCCTGCAGGGCTACAACATCCTGAT TTCCTATGGTTCCGAGCAGGCGGTCGGGCAGATGGTGGCCCTGACCCTGCTGCGCGAACTGGGGCCGG TGGTGACCGGCCTGCTGTTCGCCGGCCGTGCCGGCTCTGCGCTTACCGCGGAAATCGGCAACATGAAG GCCACCGAACAGCTTTCCAGCCTGGAAATGATCGGTGTCGACCCGCTCAAGTACATCGTCGCGCCGCGC CTGTGGGCCGGTTTCATCTCCATGCCGCTGCTCGCCGCGATCTTCAGCGTGGTCGGCATCTGGGGTGGG GCGATGGTCGCGGTCGACTGGCTGGGCGTATACGAGGGGTCGTTCTGGGCGAACATGCAGAACAGCGT GCAGTTCACCGAGGATGTGCTCAACGGCGTGATCAAAAGTATCGTATTCGCCTTCGTGGTGACCTGGAT CGCGGTCTACCAAGGCTACGACTGCGAGCCGACCTCGGAAGGAATCAGCCGGGCGACGACCCGGACCG TGGTCTATGCCTCCCTGGCGGTGCTGGGGCTCGACTTCATTCTGACTGCTTTGATGTTTGGAGATTTCTG AATGCAAACCCGCACCCTGGAAATCGGTGTCGGCCTGTTCCTCCTGGCCGGCCTGCTGGCCCTGTTGCT GCTGGCCCTGCGGGTCAGCGGCCTGAGCGTGGGCAACGCCGGCGATACCTACAAGGTCTACGCCTACT TCGACAACATCGCCGGTGTTACCGTGCGCGGCAAGGTCACCCTGGCCGGCGTGACGATCGGCAAGGTG ACGGCGGTCGACCTGGATCGCGACAGCTACACTGGTCGCGTGACCATGGAGATCAACCAGAACGTGAA CAACCTGCCGGTCGATTCCACGGCGTCGATCCTGACCGCCGGCCTGCTGGGCGAGAAATACATCGGCA TCAGCGTCGGCGGCGACGAGGACGTTCTGAAGGACGGCAGCACCATCCACGACACCCAGTCGGCGCTG GTGCTGGAAGACCTGATCGGCAAGTTCCTGCTGAACTCGGTTAACAAAGACGAAGCCAAAAAGTCGGG GTCCGGATCGGGGATGGTCTCGAAGGGGGAGGAAGACAACATGGCCATCATCAAAGAGTTCATGCGGT TCAAGGTCCATATGGAAGGCAGCGTGAATGGCCACGAATTCGAGATCGAAGGGGAGGGGGAGGGGCG GCCGTATGAAGGGACCCAAACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCTGCCGTTCGCGTGG GATATCCTGTCCCCGCAATTCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGAC TACCTCAAGCTGTCCTTCCCCGAGGGCTTCAAATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGT CGTGACGGTGACCCAGGATAGCAGCCTCCAGGATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGA CCAACTTCCCGTCCGATGGGCCCGTCATGCAAAAGAAGACGATGGGCTGGGAGGCCTCGAGCGAGCGG ATGTATCCCGAAGACGGGGCCCTCAAAGGGGAGATCAAGCAACGCCTGAAGCTGAAAGATGGGGGCC ATTATGATGCGGAAGTGAAGACGACCTACAAAGCCAAGAAACCGGTGCAACTGCCGGGCGCGTACAAT GTCAATATCAAACTGGATATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGC CGAAGGGCGGCACTCGACGGGGGGGGATGGATGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGGT TTTCTTCCATGCTGACTCTTCTGCGTCGCGGCCTGCTGGTGTTCCTGGCGGCTTTCCCGCTGCTTTCCATG GCGGCGCCGACCCCGCAACAGGTGGTGCAGGGCACGGTCGACGAACTGCTTTCCGACATCAAGGCCAA CAAGGCCGCCTACAAGGCCGATCCGCAAAAGCTCTACGCCACTCTCGACCGTATCCTTGGGCCGGTGGT CGATGCCGAAGGCATCGCCAAGAGCGTGATGACCGTCAAGTACTCGCGCCAGGCCTCACCCGAGCAGA TCAAGCGCTTCGAGGAAGTGTTCAAGAACAGCCTGATGCAGTTCTACGGCAACGCGCTGCTCGAATAC GACAACCAGGACATCCGCGTGCTGCCTAGTTCGGCCAAGCCGAGCGACGATCGCGCCAGCGTCAACAT GGAGATCCGTGACAGCAAGGGCACGGTCTATCCGGTCTCCTACACCATGACCAACCTGGCCGGTGGCT GGAAGGTCCGCAACGTGATCATCAACGGCATCAACATCGGCAAGCTGTTCCGCGACCAGTTCGCCGAC ACCATGCAGAAGAACCGCAACGACCTCGAGAAGACCATCGCCGGCTGGGAGGATCCCCGGGTACCGA GCTCGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT GCCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCAT GTTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAG GCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTT GTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGC GGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCC ATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATT

	CTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTG TAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAA AACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTC CGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTT TACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTG AAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTC TCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTT CATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGG GCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTT ATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGG AAGTGACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTG ACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTA TACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTACACGAAG GTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCG ATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGGAACTGAGTGGATAT GCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGAA AATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGT GTTCTGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCT TCGTGCGGTGTTACGTTGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAG AACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCT CGGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGCGCCAGAAACGCCGTC GAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCCCTC ACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGA GGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGAT TTTACGCGAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGA GGGGCGCGACTACTGACAGATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGA GGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCG CCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTTGTTTTTA ACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTC CCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAA AAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGA TCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGC CCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGAC AAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGT GACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAA GAAACGTAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCC GGGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAA CATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCC AGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGA GGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCCATCCAG AGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACG CAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGT CAGTATCGTAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAG CTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATC AACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGCCAGACGACTCATATCGT ATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTTGAGG GTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGC ATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTT GATTTCCTTCTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTT GATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGT GGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGCCAGTT CGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTGAGGTATGTGCTCT TCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTG TTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGA ATGAAACTCATGGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCAT TGCACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCA GCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCC GGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGC GTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCC ATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTAC

	GTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGAT CTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCA ATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTG AGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTG AAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTA ATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACT GCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACG TATTTTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGA TGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGTAAAGGTCAGATCC GGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTT CTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAAT GAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGAC ACCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTAC TTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGAC CTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGCACTGGATGATCTCAT CCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAAT TGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGA GCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGA ACGTGGTCAGCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGC GGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGC TCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAAGCCTTTACAGATAA AGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAG CTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTC ACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACA GGTCTCGTGTTCCAACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCA CCCTGATGCGACCACGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAA GCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCTGATAATCAGACTG GGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCG GTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGG TCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTAT TAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGT ATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGA CCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCA GACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACC TGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCC ACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGG GCTGCATCCGATGCAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTC AGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACC TGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAATC ATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCT ATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAAT ACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTC TGTTTTTGTCCGTGGAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTA TACGAAGTTATATTCGAT
Strain	Notes
Pseudomonas aeruginosa PAO1	Gifted by Professor Lori Burrows, McMaster University.
Escherichia *coli* MG165	Gifted by Manuel Banzhaf

5	

Chapter 4: Results and Discussion

4.2 Phosphatidyl choline incorporation - tracking through its mimic: propargyl-choline

As part of this chapter, as mentioned in the abstract, I created a new method for staining bacterial membranes and assayed this on the bacteria *Pseudomonas aeruginosa*, in an attempt to see where new lipid insertion occurs, not only at the inner membrane, but the outer membrane. The use of membrane specific dyes for fluorescent microscopy of organisms, *in vivo* has been done, however much of the time these labels are non-specific, and cannot track individual lipid movements, often instead acting as non-specific lipid affinity labels, or require insertion of whole lipids. I was not successful in visualising local membrane insertion with this label, therefore the hypothesis for outer membrane inner leaftlet insertion and peptidoglycan coordination was not answered, however did create a new membrane labelling technique.

Here I describe a short method for labelling using a phosphatidylcholine mimic, “propargyl-choline” on *Pseudomonas aeruginosa*. This click chemistry pliable mimic is visible under microscopy at bacterial membranes once reacted with a suitable click-label such as Cy3, has a fluorescence dependent on PC concentrations and is found in the membrane fraction of cells, therefore is suitable for cell imaging, and membrane labelling. Our analysis of this lipid headgroup mimic hints at movement of phosphatidylcholine, but also has potential in the monitoring of potential differences in phosphatidylcholine distribution over time in bacterial populations. The method of phospholipid labelling is especially relevant, as *Pseudomonas aeruginosa* pathogenicity has been shown to be partly dependent on the phosphatidylcholine pathway, and enzymes that degrade it, inducing inflamation, (166, 167), and there are a lack of tools for understanding lipid movement in these bacteria.

In order to determine propargyl-choline efficacy as a membrane localising phosphatidylcholine mimic the click chemistry mimic for the PC headgroup, “propargyl-choline” was incubated with cells during exponential growth for 5 minutes, representing roughly ¼ of a growth cycle.
In Figure 4.2.0, we show that in cells incubated with propargyl-choline for 5 minutes there is an increase in fluorescence at the membrane after click labelling, indicating that the lipid head group mimics are localised to the membrane, however in specific cases, an increase in fluorescence can be visualised at the division site, possibly due to increased membrane vulnerability and staining access. When cells were averaged using BACTmap software, no specific increase in fluorescence could be seen at any region except the membrane, and division plane of new membrane formation. A titration of propargyl-choline concentration in the cells Figure 4.2.5, revealed no fluorescence above background when no propargyl-choline was present, and visible fluorescence in cells from 1μM propargyl-choline abundance, through to 100μM.

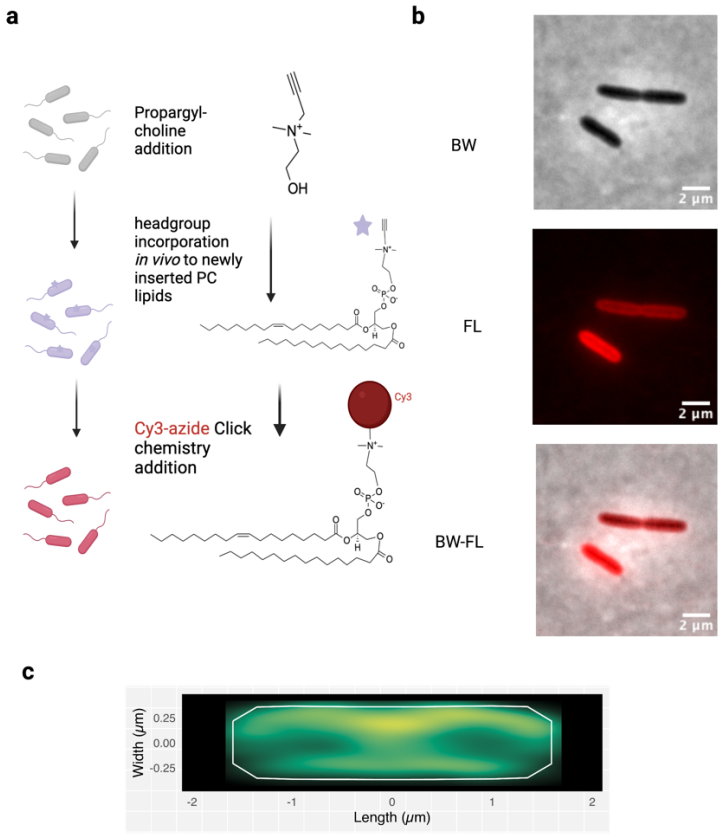


Figure 4.2.0. Phosphatidylcholine Lipid insertion can be visualised by Propargyl-Choline

a- propargyl-choline mimics phosphatidyl choline and can be labelled by click chemistry, b- fluorescence to the membrane facilitated by soluble propargyl-choline click label. c- BACTmap average of fluorescence.

In order to associate our fluorescence with any outer membrane inner leaflet fluorescence, we prepared cells incubated with propargyl-choline for an outer membrane isolation. A TLC of these lipids (Figure 4.2.1) indicated the presence of propargyl-choline head groups, which we identified through click chemistry and fluorescence.

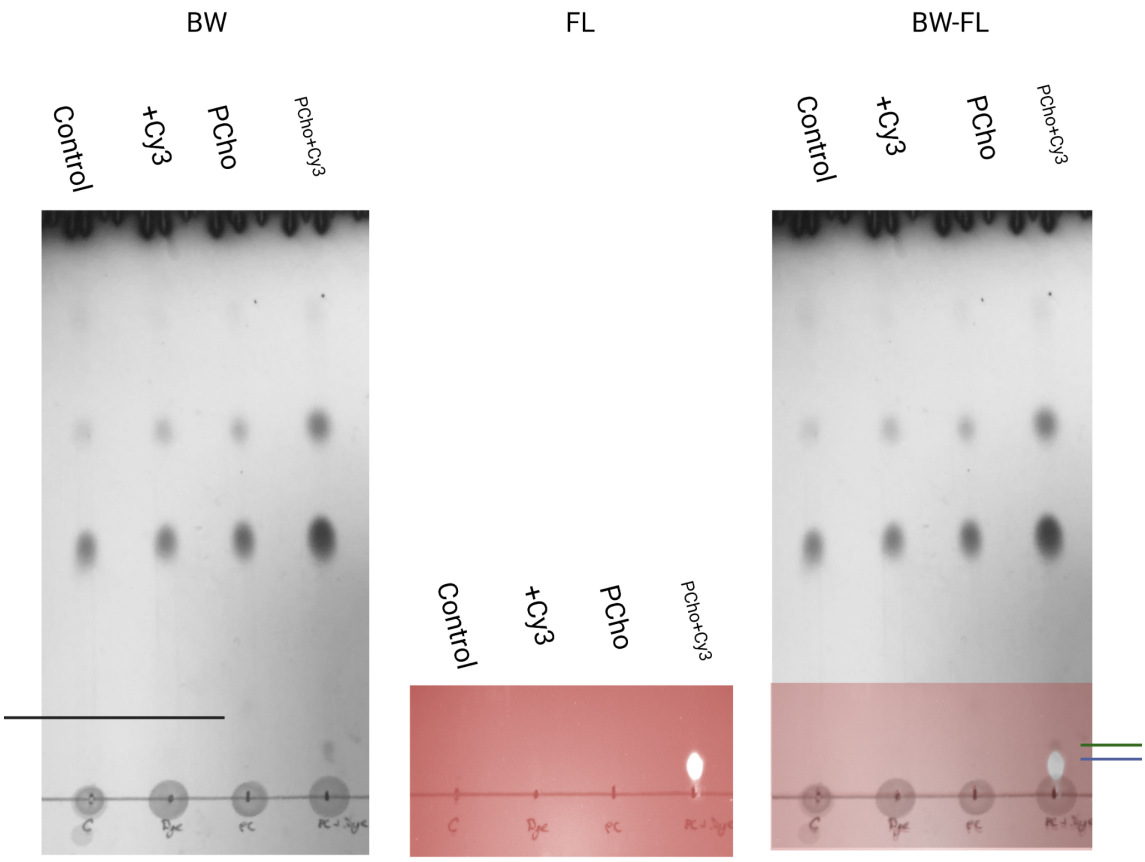


Figure 4.2.1 Insertion of propargyl-choline and Cy3 in P. aeruginosa membranes.

FL- Fluorescent channel of cells, indicating Cy3 lipid attachment, BW- TLC of whole cells indicates abundance of lipid in each sample. (Black line)- PCho (Green/Blue)- PCho-Cy3. The TLC plate fluorescence was recorded using Cy3 fluoremetric TXR filters on a 5x Zeiss Axio zoom microscope, and posed adjacently.

Figure 4.2.1, highlighting this TLC of the lipid fraction of whole cell *Pseudomonas aeruginosa* cells, indicates the fluorescent headgroup of propargyl-choline-cy3 was incorporated in only cells with both propargyl-choline and Cy3 exclusively, and propargyl-choline was incorporated in the lipid fraction of cells Figure 4.2.1. This TLC also indicates a peak for cells without Cy3 addition, where a heavier phosphatidylcholine mimic may migrate.

In order to find differences in fluorescent intensity over larger cell stretches, we were able to repeat our experiments in elongated Astreonam challenged cells, to indicate regions of high

intensity of PC insertion compared with no insertion over this same time period (Figure 4.2.3), this did show small differences in intensity, with banded regions of higher propargyl-choline concentration, however not to a significant degree when used in wildtype cells , or phenotypically interesting as a label.

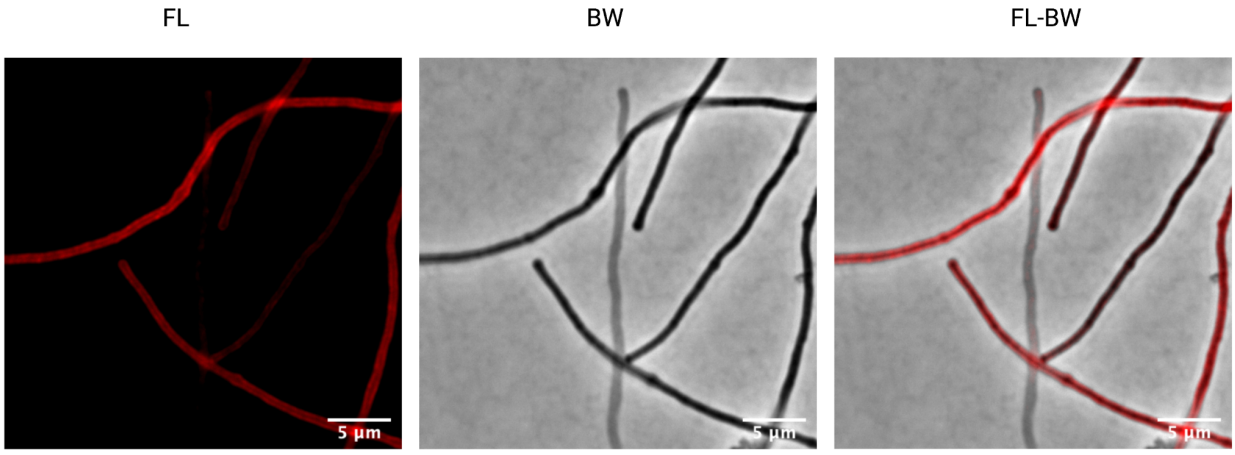


Figure 4.2.3 Astreonam elongated cells show localised fluorescent incorporation

BW- Confocal, FL- fluorescent wavelength, BW-FL - merged Confocal Fluorescence images. All images were brightness and contrast adjusted manually for the highest level detail observable in fluorescence channels. 5 minute incubation.

Propargyl-choline-Cy3 fluorescence has a similar localisation pattern to FM-464X

In order to associate Cy3-mediated fluorescence of labelled propargyl-choline with membrane fluorescence and compare this with a widely used membrane localisation method, the membrane dye FM-464X was used as a control (Figure 4.2.1). In our experiments, we found that the FM-464X membrane localisation was a visually clearer membrane label than our own propargyl-choline alternative, perhaps as not all the propargyl-choline had been incorporated into the membrane. However we also found that our lipid dye, clearly localised to the membrane across the cell cycle, indicating the propargyl-choline headgroups were being recruited to the membrane. However, this fluorescence was heterogenous among cells, suggesting differences in fluorescence of our label, or patterns of the dye between cells.

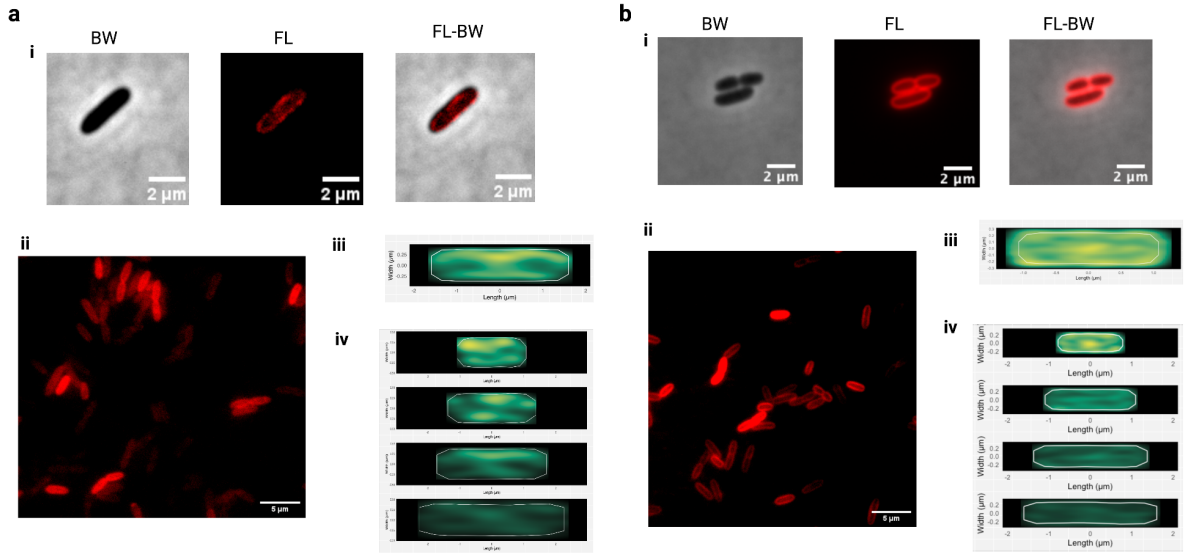


Figure 4.2.2 Lipid insertion to outer membrane during division is divisionally localised Propargyl- Choline

a- localisation of fluorescence in propargyl-choline timepulsed cells after 5 minutes incubation. **b**- FM4-64X comparison

The propargyl-choline click labellable lipids can track individual lipid fluorescence.

The Cy3 label that had been click-labelled to the propargyl-choline group, photobleached over time, however the localisation of the fluorescence in the cell changed over the course of miliseconds. This was suggestive of the 3% potential phospholipids labelled changing localisation. These lipid particles, or groups of lipid particles could be tracked in imageJ, using Trackmate, to give a population of potential lipid speeds over time. When these speeds, were tracked over time, they revealed movement around a periplasmic track between 0.25-0.75µm/s which is similar to the speeds seen in mammalian cells previously, suggesting these are lipid movements, these movements were also increased by addition of octanol, a membrane fluidiser. Without suitable controls, this lipid speed cannot be verified, however studies tracking individual lipids in wildtype environment membranes like this could be taken advantage of in future studies.

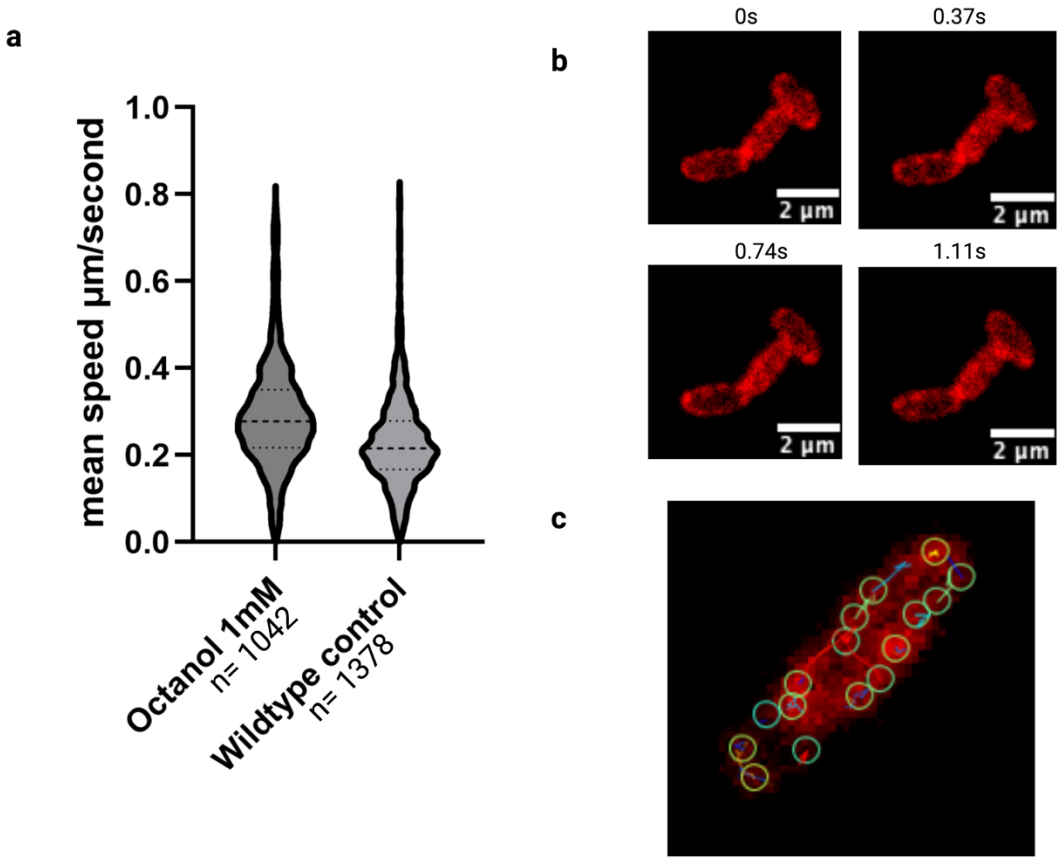


Figure 4.2.4 Use of propargyl-choline as a potential lipid domain visualisation tool.

a Trackmate mean track speed µm/s and fluorophore counts in a *Pseudomonas aeruginosa* PAO1 propargyl-choline and Cy3 labelled cells. Significant difference $p < 0.0001$ **b** Snapshot of fluorescence change over time in periplasm. Red indicating fluorescence, 100µs exposure Cy3 labelled PCho *Pseudomonas aeruginosa* cell **c** Trackmate fluorescence domains capture, and tracks shown

Finally, we assayed the concentration of propargyl-choline that could be added to cells, whilst allowing for unrestricted growth during growth phase, therefore useful for experiments. All the concentrations used in our experiments did not affect growth, however above 175µM, and with increasing DMSO concentration, growth was slowed (Figure 4.2.5). This would suggest any future study in *Pseudomonas aeruginosa* do not use excessive propargyl-choline, but anywhere between our assayed values, with visible fluorescent read out between 1µm and 100µm in concentration, if the propargyl-choline suspension is from DMSO. In addition, by changing the amount of propargyl-choline used for the pulse labelling prior to cell fixing staining and

fluorescence monitoring, this was shown to be associated with a change in observed fluorescence over background. (Figure 4.2.5).

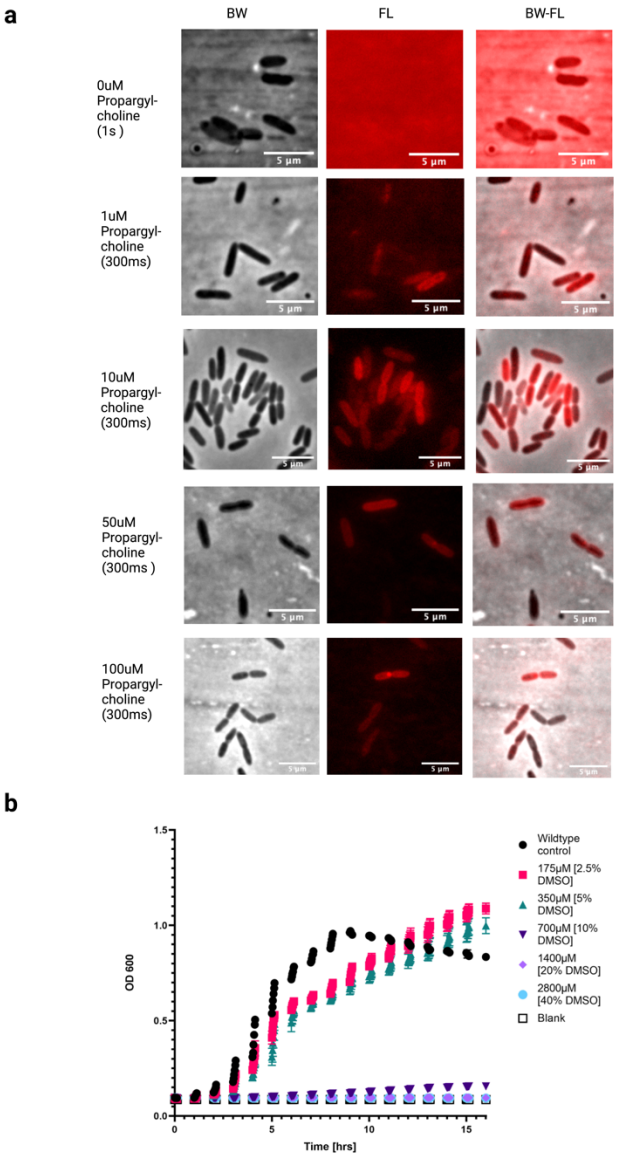


Figure 4.2.5 Cell Viability and propargyl-choline concentration

a propargyl choline-cy3 fluorescence across alternate PCho concentrations **b** The viability of cells in the presence of propargyl-choline remains high until DMSO concentrations for dissolution impact growth.

4.2.6 Use of propargyl-choline as a *Pseudomonas aeruginosa* membrane marker

In Figure 4.2.0 by pulse labelling *P. aeruginosa* cells with propargyl-choline, we were able to visualise new phosphatidylcholine (PC) insertion in the cell membrane, with no specific preference of insertion. This insertion, at the membrane (Figure 4.2) was present across the cell and did not represent the localised patterns hoped to be achieved by any increased fluorescence in the inner leaflet of the outer membrane, however could also be visualised in heterogeneous membrane localisations (Figure 4.2.4).

This is the first time propargyl-choline has been used as a membrane marker in bacteria, with a previous study reporting its use for teichoic acid incorporation in Gram-positive bacteria and mammalian cells, therefore characterisation of the fluorescence localisation by TLC (Figure 4.2.1) and other suitable controls were conducted to remove the possibility of background fluorescence, or mislabelling. There are no other structures such as teichoic acid in *Pseudomonas aeruginosa* capable of using the propargyl-choline click-label chemical groups, therefore only the membrane was labelled, however we did not have the time to ask at what level of insertion the PC had at the inner membrane and outer membrane, which would be the next experiment if the project was to continue from a biological questions in pseudomonas standpoint. With these controls, and on a PC utilising species we had focused on this lipid fluorescence. I assayed lipid speed, determined by the fluorescence and was intrigued to discover if there were two lipid domain movement velocities, when single lipids were tracked, as these lipids potentially could be in the inner membrane and also inner leaflet of the outer membrane, but could not find a discernible separation of velocities which may be suggestive of modes of lipid rafts with lower diffusion rates such as the outer membrane inner leaflet.

Although a previous study has used propargyl-choline and an alternative exogenous 1-azidoethyl-choline to label a specifically modified *E.coli* strain with upregulated pcs, in our studies we have shown clearly that propargyl-choline has insertion into the membrane of wildtype bacteria, through its replacement as a phosphatidylcholine mimic *in vivo.* I have shown this through lipid extraction of *Pseudomonas aeruginosa* cells, which on thin layer chromatography revealed a fluorescently labelled group, only when both Cy3 and propargyl-choline were present, as well as microscopy which reveals a membrane localisation.

Most notably, our labelling technique through use of a pulse-chase compatible biomarker could allow for monitoring of lipid movements in *Pseudomonas aeruginosa* cells, with concentrations of propargyl-choline below 100μM compatible with wild type growth, this could allow for studies in phospholipid exchange within bacterial populations and phospholipid fluidity changes without use of mass spectrometry or for single cell sorting by uptake of propargyl-choline headgroups. I have also shown that the labelling is compatible with changes in cell length and antibiotic challenge, as in cells treated with Astreonam (Figure 4.2.3)The importance of phosphatidylcholine in pathogenicity for *P.aeruginosa*, makes its use as a biomarker relevant, not only for membrane fluidity as discussed, but in future understanding of Gram-negative cell envelopes that contain phosphatidylcholine.

4.3 Identification of MCE domain containing proteins in *Pseudomonas aeruginosa*

Pseudomonas aeruginosa PAO1 MCE domain containing proteins have not been studied before, therefore the initial step of the experiment was to look for homologues of the MCE domain containing ORFs within *P. Aeruginosa*, and to assess likely functionality. A PFAM search, and Blast search revealed three proteins, corresponding to the bridge like mechanism of YebT and PqiB, as well as a boat like mechanism of the Mla system of transport (Figure 4.3.0-A) these were well conserved across the Pseudomonas family (Figure 4.3.0-C) among other taxa, but were not always present (Figure 4.3.0). The domain distribution of these genes (PA3213, PA4454, and PA4689) was suggestive of these roles, in addition to an MlaC like protein adjacent to PA4454, and PqiA protein adjacent to PA3213.

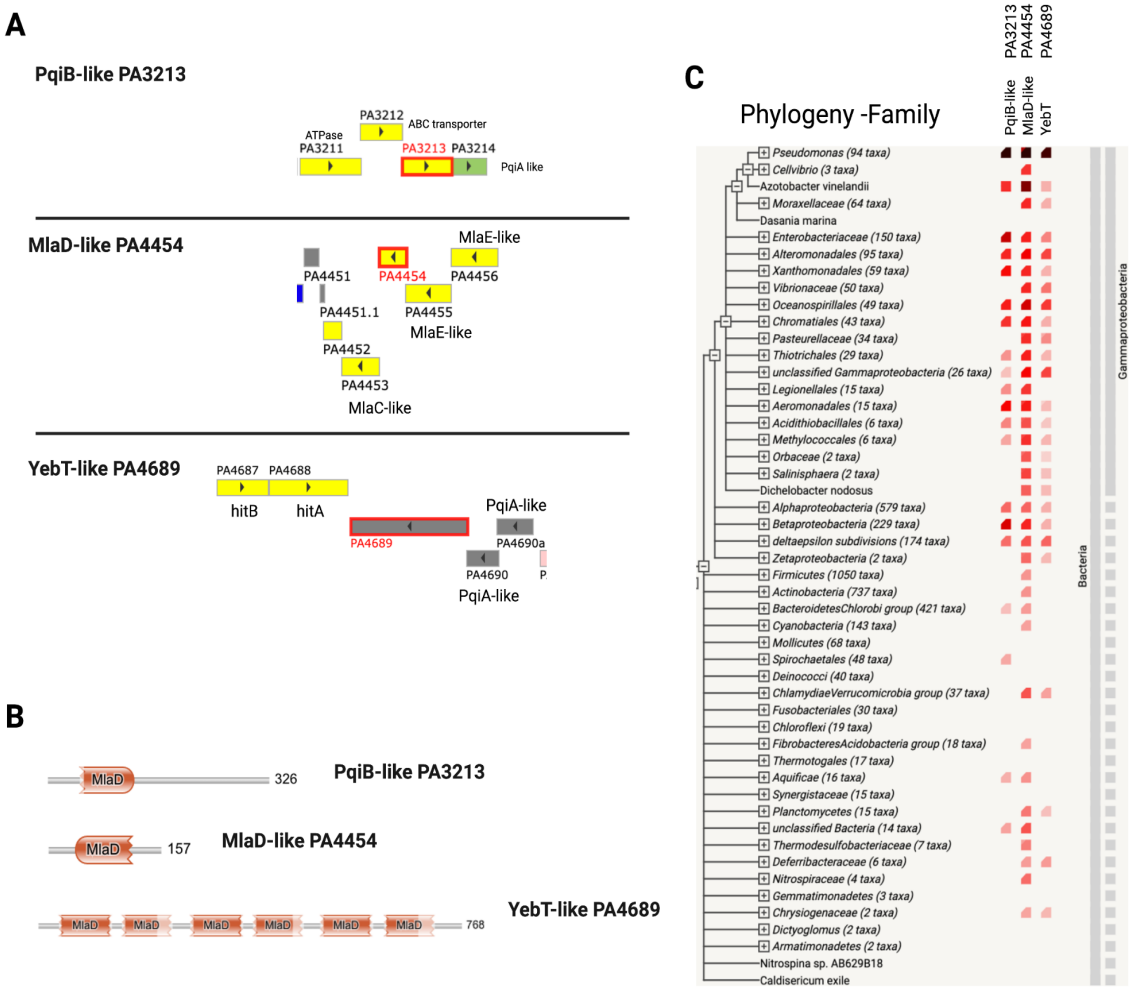


Figure 4.3.0 Identification of MCE domain containing proteins in *Pseudomonas aeruginosa PAO1*

In addition to the domain analysis, we predicted the 3D structure of these proteins using the AlphaFold-multimer package on a Co-lab server. The predictions generated (Figure 4.2.1) were also suggestive of these roles of bridge and boat. Most interestingly the PqiB-like protein PA3213 was annotated as MlaD due to the presence of only one MCE domain, however it's coiled coil region extends far out into the periplasm, as *E.coli* PqiB does, suggesting that although the protein is altered in the majority of the well conserved MCE domain containing region, the overall fold is more similar to PqiB. Similarly the gene annotated PqiB PA4454 is

more similar to MlaD, with a soluble lipid carrier protein MlaC, adjacent. *P. aeruginosa* YebT, was well annotated and its predicted structure very similar to *E.coli* YebT.

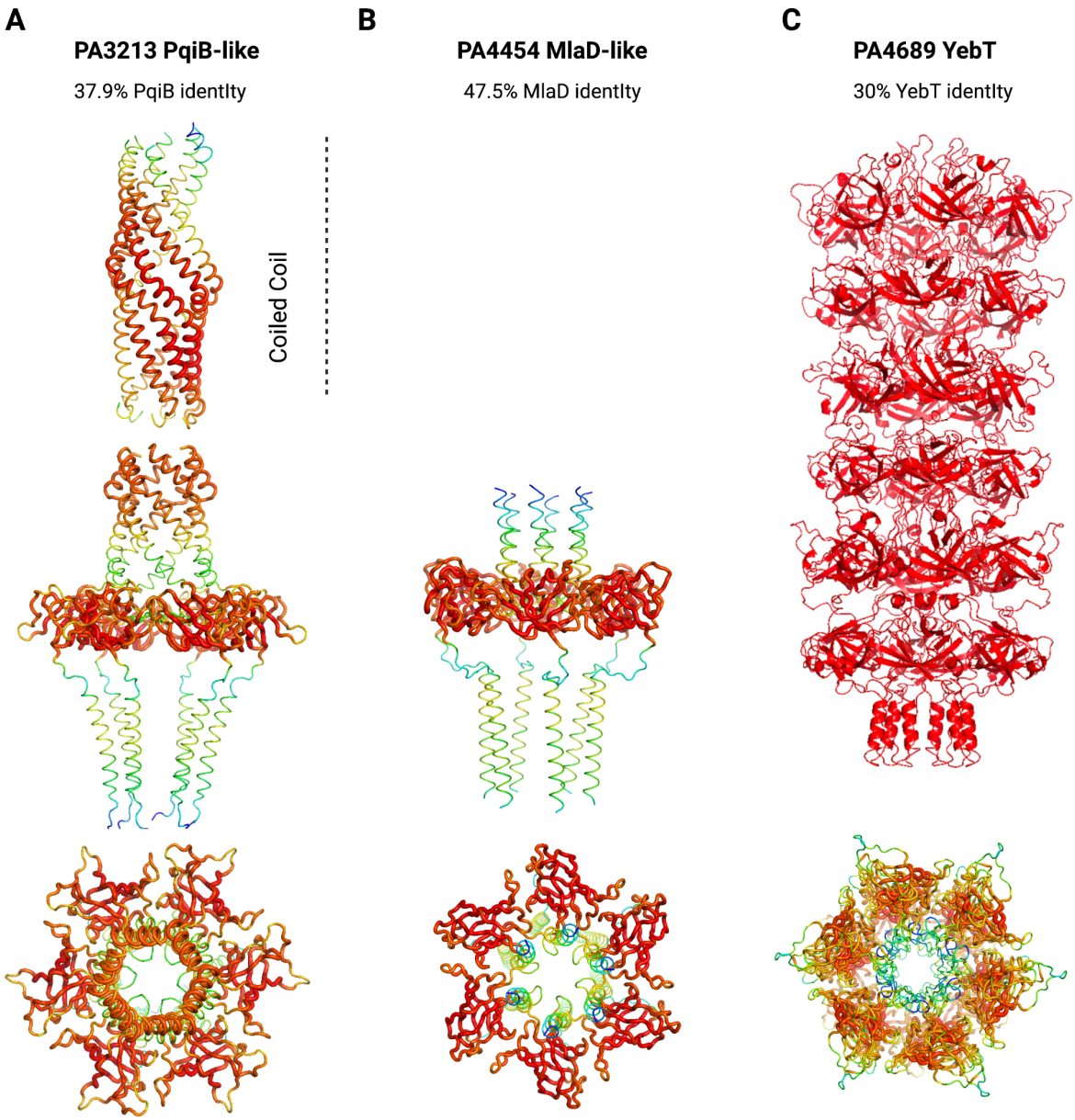


Figure 4.3.1 Structural prediction of MCE domain containing proteins in Pseudomonas aeruginosa PAO1

After finding the proteins, the proteins were labelled at the C terminal domain by a SGSGSG linker, in addition to FtsW and RodA contols, and their localisation investigated, to uncover more of their role in P. *aeruginosa*, as well as other similar systems.

4.4 Genetic knockins of *Pseudomonas aeruginosa*

After creating the fluorescent chromosomal fusions of each gene in pseudomonas, we had to ensure the cells were not sick, therefore behaving as wildtype. I did this using a colony size indicator across 288 replicates across conditions which effect cell viability as seen in (Figure 4.4.1) This indicated that P. *aeruginosa* have a range of growth rates under these conditions.

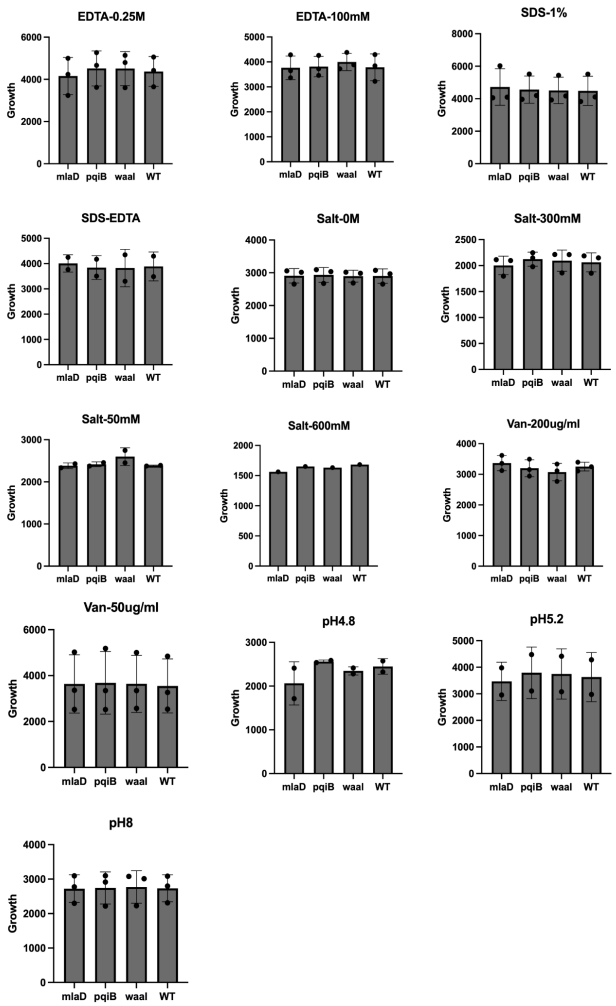


Figure 4.4.1 Phenotype of fluorescent mutants is consistent with wildtype

Although these on glance appeared consistent, we then confirmed using a two way ANOVA test that these phenotypying results in our chromosomal fluorescent knockins were consistent with wildtype. (Figure 4.4.1) and in Table 4.4.1

Table 4.4.1 Anova test of strain growth

Dunnett's multiple comparisons test	Predicted (LS) mean diff.	95.00% CI of diff.	Below threshold ?	Summary	Adjusted P Value
EDTA-025					
WT vs. mlaD	217.4	-1054 to 1489	No	ns	0.9562
WT vs. pqiB	-142.3	-1414 to 1129	No	ns	0.9868
WT vs. waal	-137.2	-1409 to 1135	No	ns	0.9882
EDTA-1					
WT vs. mlaD	20.55	-1251 to 1292	No	ns	>0.9999
WT vs. pqiB	-29.13	-1301 to 1243	No	ns	>0.9999
WT vs. waal	-209.7	-1481 to 1062	No	ns	0.9604
SDS-1					
WT vs. mlaD	-241.9	-1514 to 1030	No	ns	0.9414
WT vs. pqiB	-76.98	-1349 to 1195	No	ns	0.9978
WT vs. waal	-26.98	-1299 to 1245	No	ns	>0.9999
SDS-EDTA					

WT vs. mlaD	-115.3	-1673 to 1442	No	ns	0.9961
WT vs. pqiB	52.25	-1505 to 1610	No	ns	0.9996
WT vs. waal	66.16	-1491 to 1624	No	ns	0.9993
Salt-0					
WT vs. mlaD	-8.792	-1281 to 1263	No	ns	>0.9999
WT vs. pqiB	-35.47	-1307 to 1236	No	ns	0.9999
WT vs. waal	3.045	-1269 to 1275	No	ns	>0.9999
Salt-300					
WT vs. mlaD	59.93	-1212 to 1332	No	ns	0.999
WT vs. pqiB	-61.15	-1333 to 1211	No	ns	0.9989
WT vs. waal	-31.81	-1304 to 1240	No	ns	0.9999
Salt-50					
WT vs. mlaD	1.734	-1556 to 1559	No	ns	>0.9999
WT vs. pqiB	-33.96	-1592 to 1524	No	ns	>0.9999
WT vs. waal	-215	-1773 to 1343	No	ns	0.9759
Salt-600					
WT vs. mlaD	119.8	-2083 to 2323	No	ns	0.9984
WT vs. pqiB	31.05	-2172 to 2234	No	ns	>0.9999
WT vs. waal	48.51	-2154 to 2251	No	ns	>0.9999
Van-200					

WT vs. mlaD	-114.4	-1386 to 1157	No	ns	0.993
WT vs. pqiB	53.93	-1218 to 1326	No	ns	0.9993
WT vs. waal	177.9	-1094 to 1450	No	ns	0.975
Van-50					
WT vs. mlaD	-88.66	-1360 to 1183	No	ns	0.9967
WT vs. pqiB	-135.4	-1407 to 1136	No	ns	0.9886
WT vs. waal	-92.04	-1364 to 1180	No	ns	0.9963
pH4.8					
WT vs. mlaD	387	-1171 to 1945	No	ns	0.8822
WT vs. pqiB	-112.6	-1670 to 1445	No	ns	0.9963
WT vs. waal	102.1	-1455 to 1660	No	ns	0.9972
pH5.2					
WT vs. mlaD	161.7	-1396 to 1719	No	ns	0.9894
WT vs. pqiB	-162.3	-1720 to 1395	No	ns	0.9893
WT vs. waal	-117.1	-1675 to 1440	No	ns	0.9959
pH8					
WT vs. mlaD	9.441	-1262 to 1281	No	ns	>0.9999
WT vs. pqiB	-11.57	-1283 to 1260	No	ns	>0.9999
WT vs. waal	-40.86	-1313 to 1231	No	ns	0.9996

This confirmed our strains were behaving as wildtype, indicating strongly that we would see the correct fluorescent localisations of our proteins when investigated. There was no significant difference between any strain in any condition.

4.5 Localisation of MCE domaining proteins in *Pseudomonas aeruginosa*

Images taken during stationary phase indicated a localisation of the MCE domain containing protein and periplasm spanning protein PqiB and also the periplasmic and inner membrane associated MlaD at polar and divisional regions (Figure 4.4.1 B, therefore maxima localisation analysis of the cells was performed using MicrobeJ image analysis software (Figure 4.4.1 A) This indicated PqiB and MlaD had asynchronous division related and polar growth related locality. Waal had no such locality, with exception for increased divisional presence. This for the first time showed the machinery are localised. However stationary phase cells are not representative of growth, and MCE domain containing proteins show no connection to PG apparatus, except cardiolipin increases at stationary phase which may be accommodate by this change.

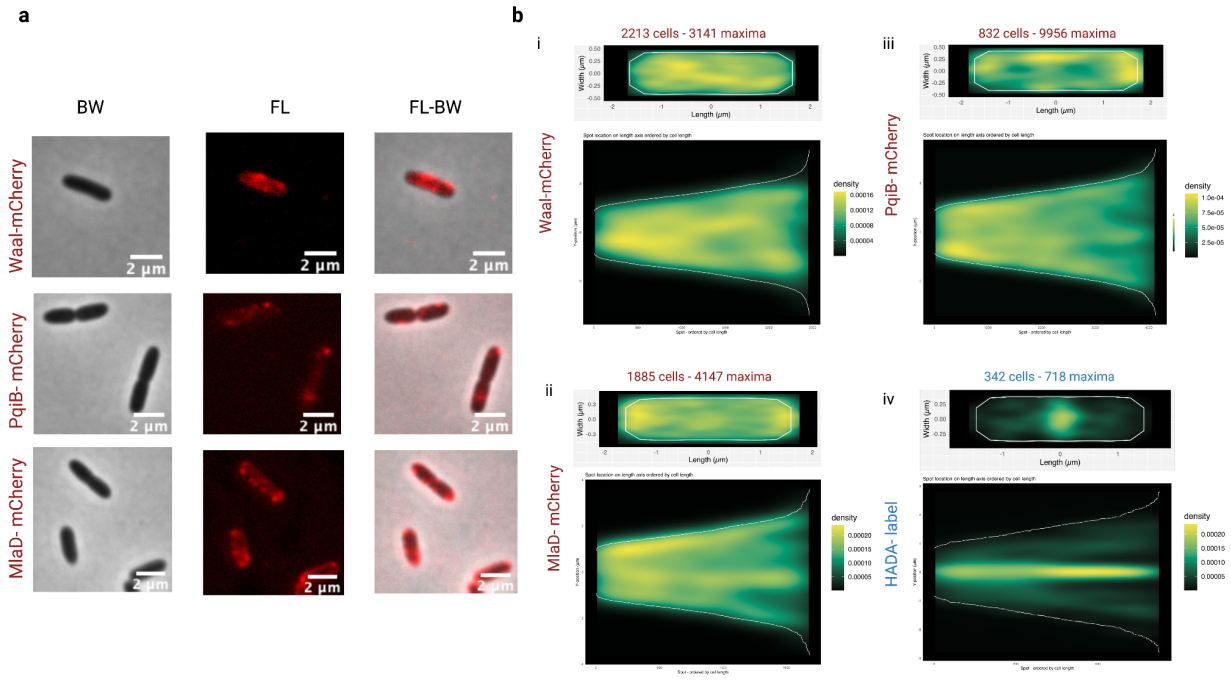


Figure 4.5.1 Fluorescent localisation at exponential phase of fluorescently labelled proteins.

a Fluorescent maxima plotting using microbeJ averaging, at stationary phase. **b** Phase contrast, Appropriate Fluorescent channel, and merged images indicating select fluorescence patterns observed on majority of cells.

The change in fluorescence dependent on stationary and non stationary cells was also interrogated (Figure 4.5.2) , with essentially a loss in localisation preference and localisation across the cell after inoculation and during exponential phase. This showed foci of protein spread through the cell.

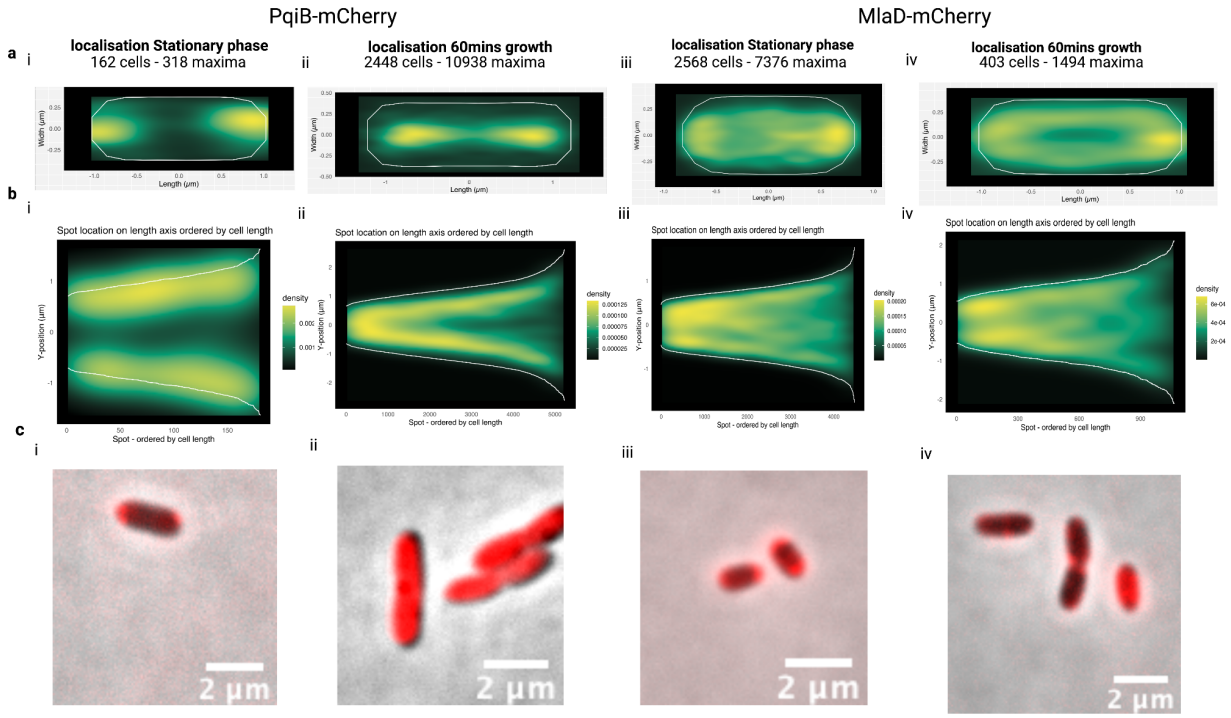


Figure 4.5.2 Timecourse of PqiB and MlaD localisation

a Histogram of Fluorescent maxima visualisation by average location (green-yellow density) as analysed by MicrobeJ and BactMAP b longitudinal distribution based on cell length c Fluorescent and BW channels merged, showing localisation as seen through microscope.

The polar localisation of the MCE domain containing proteins in pseudomonas MlaD and PqiB in experiments was not expected, and at stationary phase was shown to be accentuated. Therefore we analysed stationary phase cells, and recovered these cells to see changes in fluorescence. This revealed the polar localisation to be partially lost after incubation with new media, suggesting that this polar increase is associated with aging cells (Figure 4.5.2).

In order to determine if the localisation of OM phospholipid machinery MCE domain containing proteins was coordinating with division at all, considering the increase in fluorescence at the division site in the length sorted cells of Figure 4.4.2, the localisation of HADA and MlaD-mCherry was compared in vivo. (Figure 4.5.3) In dividing cells, MlaD did not localise to the division site, as shown by a HADA control label, and was instead abundant in foci throughout thr cell. Per-dividing cells were also imaged showing little fluorescent overlap (Figure 4.5.3aii)

and there were regions of fluorescent overlap, however when cell fluorescence was compared the correlation between mcherry and HADA was negative, suggestive of no relation to HADA insertion and MlaD localisation. MlaD therefore is not associated with division.

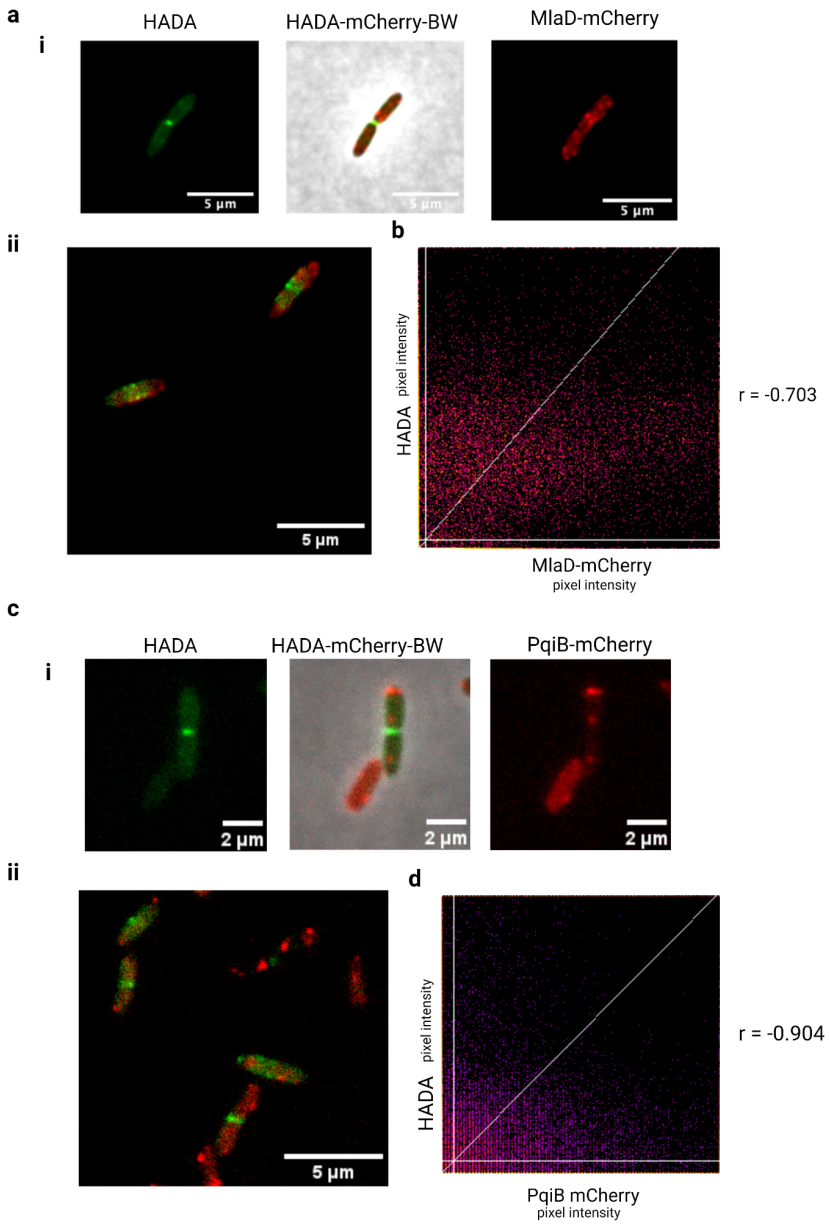


Figure 4.5.3 Localisation of MCE domain containing proteins is independent of cell division

a- MlaD Localisation of HADA(Green) and MlaD(Red) in example dividing cell i and pre-division **b** Scatterplot of localisation in thresholded regions within cells, indicate no correlation, with a pearsons correlation of -0.703 **c-**PqiB Localisation of HADA(Green) and PqiB(Red) in example dividing cell i and pre-division **d** Scatterplot of localisation in thresholded regions within cells, indicate no correlation, with a pearsons correlation of -0.904

Our HADA labelling correlation experiment indicated PqiB and MlaD have locality independent of division (a/b), but we have also shown in our data that during exponential growth they appear in nodes similar to the elongasome. This is suggestive of similar roles played by elongasome in the MCE domain containing protein. This distribution is suggestive of MlaD and PqiB's involvement in cell envelope repair during growth, throughout the cell envelope may be dependent on cell needs.

However the tendency of some of our strains to be polar in nature was similar to that of the elongasome associated protein PBP1a in another(159) PG related study in *Pseudomonas aeruginosa* which was also polar. Our own experiments have shown this polar localisation to have reduced HADA insertion, but not no insertion, as shown by some fluorescent overlap. This is suggestive of similar housekeeping roles of growth perhaps played by PBP1a and our MCE proteins in their respective envelopes and of potential interaction, if PG and lipid synthesis is linked.

According to transcriptome data during heat shock. The increased expression pattern of MlaD PA4454 has also shown to be associated with heat shock, similar to the Waal O antigen ligase and other LPS forming and outer membrane synthesis proteins, however PqiB is not increased in expresssion(160). This suggests that although both MlaD and PqiB proteins have similar localisations and domains capable of binding cardiolipin and phosphatidylethanolamine, they likely have different roles in outer membrane stabilisation if this is their purpose.

MCE Domain proteins localisation and function summary

We labelled the MCE domain containing proteins with a fluorescent protein tag, however this fluorescence whilst showing a small increase during division, does not significantly indicate any co-ordination with division labels such as HADA (Figure 4.5.1) and in fact, a small polar preference of localisation is seen of the MCE domain containing proteins targeted. I then determined this polar localisation is exaggerated in stationary phase cells, and reduced upon innoculation with new media and subsequent new growth. PqiB and MlaD have both been associated with cardiolipin, therefore it could be speculated that their polar localisation, or increase in fluorescence at the poles is related to a role in promoting and maintaining polar

curvature with cardiolipin and other lipids. Cardiolipin presence at the poles is also noted to increase during stationary phase, which would fit this narrative, however we have no evidence they are undertaking this role in this localised region.

It is likely that if division and lipid synthesis are linked, *which this data suggests is not the case*, this linked process would be performed by another protein or set of proteins, such as the AsmA domain containing proteins recently discovered. Studies into their localisation during division and throughout the cell cycle, considering this increase in localisation of phosphatidylcholine observed, would bring us closer into uncovering the mechanism of cell envelope formation.

4.6 Localisation of *Escherichia coli* lipid proteins

In order to then assay the conservation of this system in *E.coli*, I designed 12 additional constructs based on bioinformatics and recent advances in lipid transport proteins. This would include all phospholipid transporters including the AsmA proteins, in a complementation style system confirmed by cell wall integrity assays at Manuel Bhanzafs lab, as well as key peptidoglycan modifying proteins such as PBP1b and RodA, but also membrane-peptidoglycan interaction proteins such as YhdB and OmpA. I predicted their structure using Alphafold, to determine which point to label the proteins fluorescently (Figure 4.3.0)

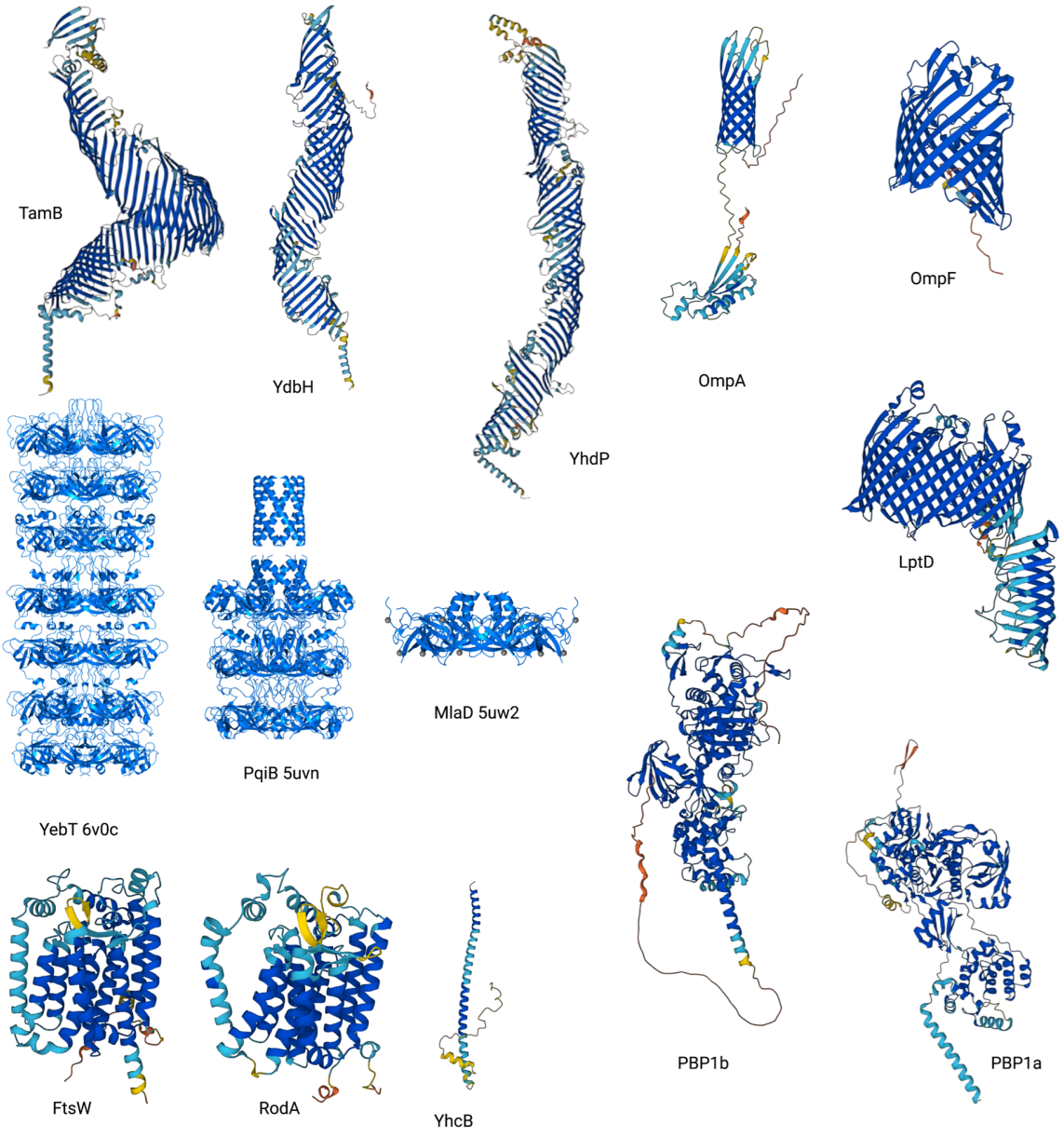


Figure 4.4.0 AlphaFold2 predicted structures of *E.coli* peptidoglycan and outer membrane phospholipid balance systems.

The AsmA proteins TamB, YdbH and YhdP are predicted by AlphaFold to have a single perisplasmic spanning region, which bridges phospholipid across the periplasmic gap, with a single TM helix, therefore were labelled N terminally as to be labelled at the inner membrane. Peptidoglycan modifying enzymes FtsW, RodA, PBP1b and PBP1A were similarly labelled,

however C-terminal labels were given to the MCE domain containing proteins, as their coiled coil regions predicted allow for space after linker attachment. In addition their N terminal is within the protein making this less efficient.

Table 4.3 *E.coli* Chloramphenicol pBAD30 RFP constructs

Strain
OmpF (Control)
YhdP-N (New OM assembly AsmA containing)
TamB-N (New OM assembly AsmA containing)
YdbH-N (New OM assembly AsmA containing)
YebT
PqiB
MlaD
LptD-N (OM inserted LPS tulip protein)
YhcB-N (Special OM-PG link protein)
RodA
PBP1a
FtsW
PBP1b
OmpA (Control)

The constructs were constructed with a Chloramphenicol marker so as to be compatible with Kanamycin constructs, as well as Histadine markers so pulldowns could be made to confirm protein presence. A leaky expression, pBAD30 vector was used as wildtype and knockout strains without DE3 T7 polymerase lysogenic addition could be used. They will be used after the PhD in further studies, however we did not have the time to investigate and analyse them.

Chapter 4: Conclusion

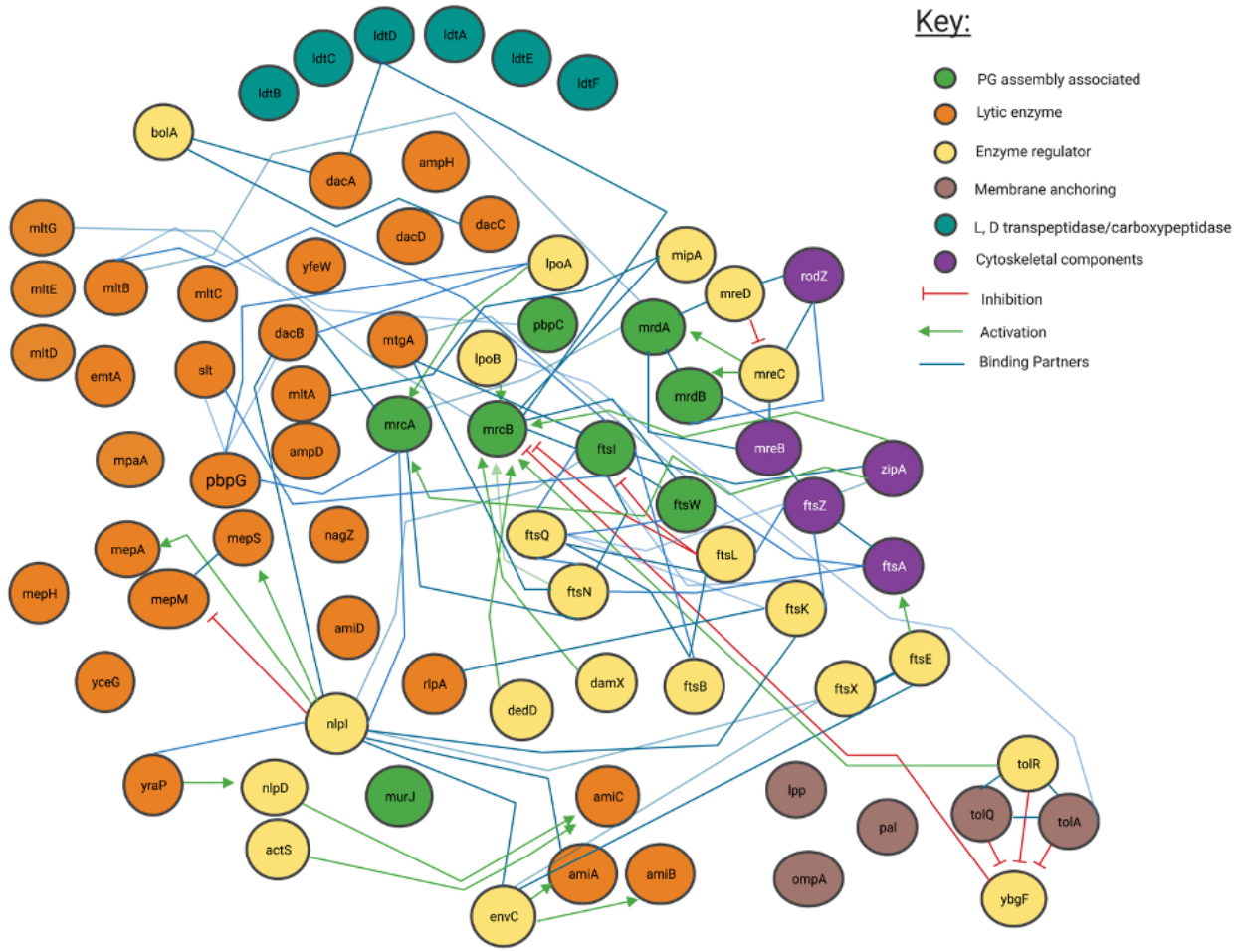
In this chapter, I created a new technique for investigating lipid movements in wildtype bacteria, specifically species containing the lipid 'phosphatidylcholine'. I then tested this on *Pseudomonas aeruginosa PAO1*. The goal of this technique was to uncover if a noticeable increase in lipid fluorescence was detectable when the cells were pulsed with propargyl-choline label at the mid-cell of the bacteria, therefore uncovering if insertion of new phosphatidylcholine lipids was a division related process. I confirmed this technique was localising Cy3 labelled propargyl-choline to the lipid fraction of cells by TLC, and cell envelope by microscopy imaging. However the localisation of lipids was likely that of both cytoplasmic and outer membrane lipids. I used TrackMATE to track the speed of the lipids and found them to correspond roughly to accepted lipid speeds in mammals. I characterised this method for use in other studies, however the hypothesis that lipids were inserted into the inner leaflet of the outer membrane during division at the division site was not supported by my work.

I then for the first time identify the MCE domain containing proteins and showed the cellular localisation of MlaD and PqiB, in *Pseudomonas aeruginosa*, which indicated they are specifically localised, supporting the need for localised insertion, and in a similar fashion to the known localisations in *Pseudomonas* sp. of existing PG enzymes, however with a preference to the poles especially during stationary phase. In my work we discuss how this is likely related to the cardiolipin and phosphatidylethanolamine, this is due to the affinity of these lipids to MCE domain containing proteins in crystal structures, pointing their role towards this balance, with other systems such as the AsmA domain containing proteins likely involved in other lipid insertion types .

The chapter hypothesis, to link the peptidoglycan synthesis with that of the outer membrane however, did not find any explicit links, but in the effort to find it created a new method for visualising for lipid movement in wildtype cells, and uncovered the likely function of the MCE domain-containing proteins as potential polar maintenance machines. This work also hinted at the need for localised lipid insertion by the polar nature of the MCE domain-containing proteins. I have set up a body of work to uncover the localisation of the AsmA domain-containing proteins, with a series of plasmids uncovered by my bioinformatics, which may one day help to find the other carriers involved in support of the cell envelopes lipid to peptidoglycan connection and homeostasis. In conclusion, during the chapter, no link between peptidoglycan synthesis and outer membrane lipid synthesis was established, however a new method, as well as a new role for MCE domains was hypothesised.

Thesis Conclusion

In Chapter 1, we used bioinformatics and literature trawling approaches to establish the current state of the art in interactions and growth of the cell envelope. Here it was found there is a network of dynamic interactions between proteins during growth and dependent on antibiotic and environmental challenges such as pH. This chapter also contributed to the literature in the form of a meta-analysis and review on peptidoglycan binding proteins, in addition to outer membrane synthesis. RodA and PBP2, also known as mrdB and mrdB were part of this analysis.

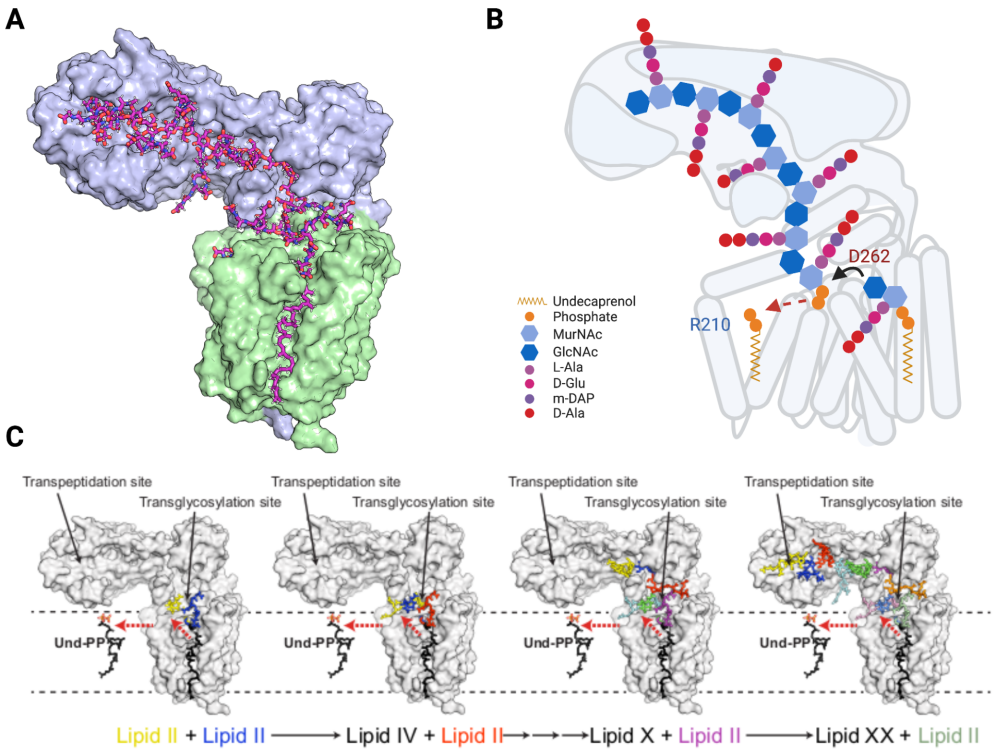


RECAP Figure 1.1.3. Interaction network of peptidoglycan modifying enzymes and their partners from Literature inspection.

Interaction map of peptidoglycan-associated proteins sorted by enzymatic action. Network structure determined by STRING, with manual addition of interactions through literature associated with each protein. Reference matrix available in this Chapters Appendix data.

In Chapter 2, with a focus on the central SEDs protein of bacterial elongation “RodA”and its partner “PBP2”, shown to be at the central of many of the interacting nodes of Chapter 1 we collaborated with Dr Rie Nygaard and a consortium of others, with a new detergent based assay and Schägger gel of fluorescent peptidoglycan pre-cursors to visualise RodA-PBP2 activity. I predicted the RodA-PBP2 structure of *E.coli*, analysed this structure alongside electron densities provided by New York and then provided initial structures of the complex. These initial structures were used by myself and with the consortium to predict carefully chosen conserved and/or co-evolutionarily valuable residues to mutate. I bound a lipid II substrate to RodA-PBP2 and thus predicted two sites of binding which were then confirmed by molecular dynamics simulation in the Phil Stansfield group.

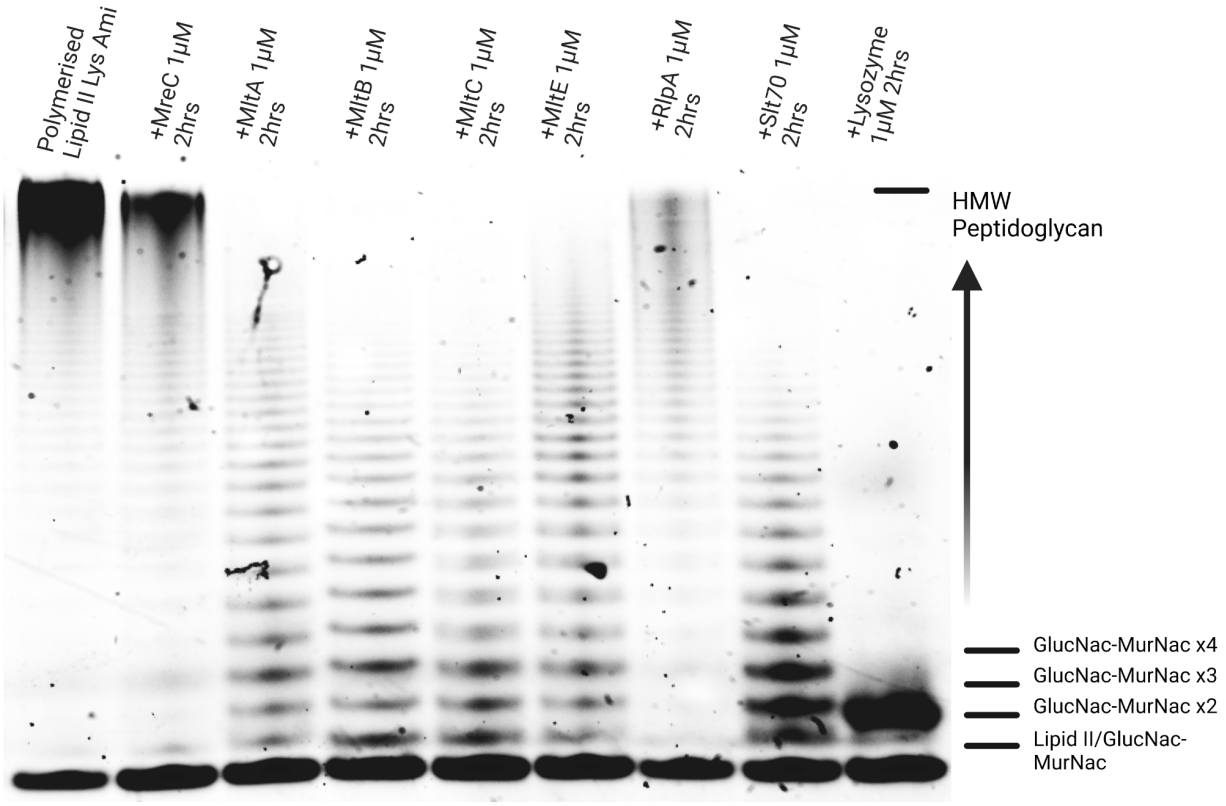
In addition to the cryo-electron structure which we then improved with better density filtering thanks to new AI software CryoSpark, we worked with Meagan Dulphrisne to visualise the flexibility of RodA and thus once and for all predict where Lipid II attaches and see the flexibility of RodA itself. Finally in Chapter 2 I detail a paper for which I am chiefly responsible for in mutagenesis, mechanistic analysis and initial structures and predictions in my own words, which predicts how lipid II is polymerised by RodA.



RECAP- Figure 2.9.2. Mechanism of polymerisation within the RodA-PBP2 complex

Collaboration with Philip Stansfield A the PBP2 molecule has a groove through which newly formed lipid II can be polymerised. B Cartoon representation of lipid II substrate, and PG product formation. C Extension of the glycan sugar chain over time.

In Chapter 3, and inspired by the central PG synthesis enzyme of RodA in addition to bioinformatic and literature searches of Chapter 1, I created a new assay for lytic transglycosylase activity, an activity essential to PG laydown and control in Gram negatives and gram positives. In this chapter I synthesised more lipid II and then visualise this being degraded by a set of eight different LTs that we purified and confirmed activity for the first time with and without an inhibitor.

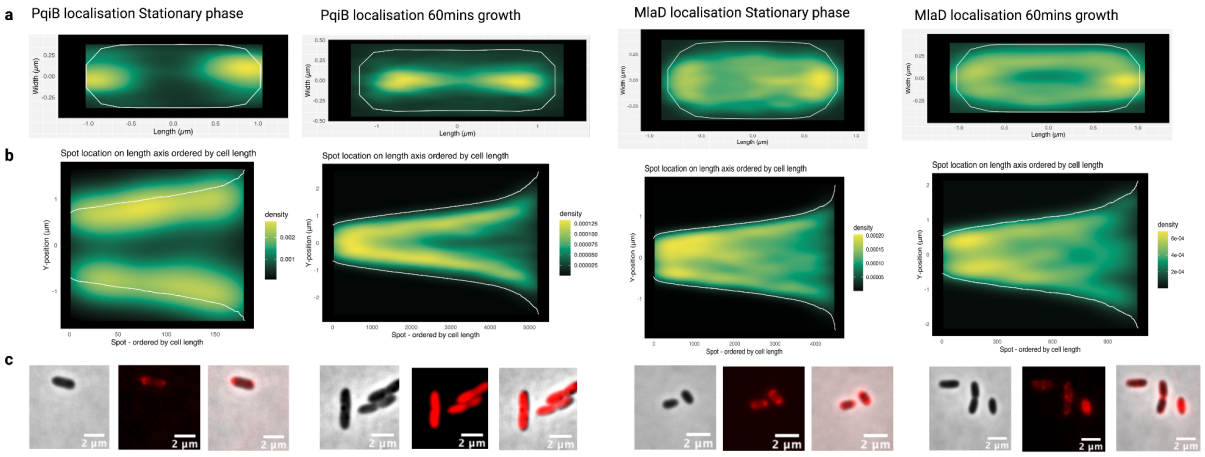


RECAP- Figure 3.5.0 New Schägger Lytic Transglycocylase assay on *E.coli* LTs with Lipid II Lys substrate

In vitro activity of Lytic transglycosylases on product of High molecular weight glycan polymer, visualised by fluorescent mDAP lipid II polymerisation in Schägger gel. Sugar product size visualised by fluorescence.

In Chapter 4, we returned to the central dogmatic argument that the cell envelope is a connected structure and investigated for the first time the localisation of the outer membrane creation systems of the MCE domains within *Pseudomonas aeruginosa*. Here I created six suicide vectors for genetic manipulation of the *Pseudomonas aeruginosa* system which were then transfected by Professor Lori Burrows in Canada and then analysed and imaged by myself. This revealed for the first time that the membrane laydown and exchange systems were indeed localised dependent on cell stage and shared common localisation patterns to that of PBP1a another integral and essential PG laydown enzyme in *Pseudomonas aeruginosa*, and linking the MCE domain containing proteins further with cardiolipin, a polar lipid. This was suggestive of a coordinated response at the outer membrane during different growth stages, strengthening the

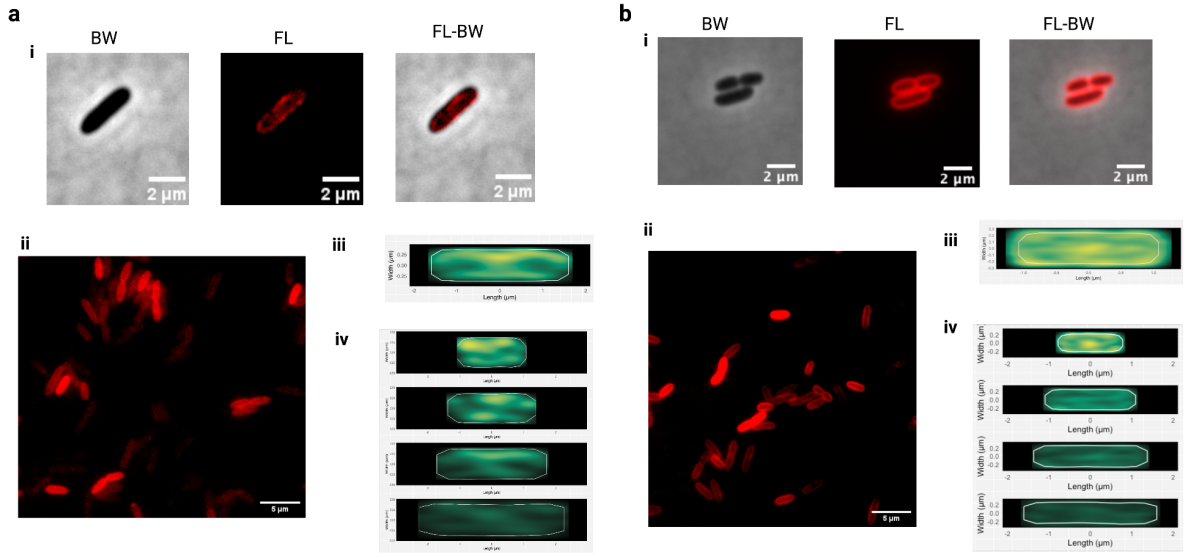
argument for an altered lipid speed at the inner leaflet of the outer membrane, requiring regulationt. This then developed into two grants and a paper which investigated this in *Pseudomonas* as well as in *E.coli* using a complementation system (Figure 4.5.2). However it was counter to the original hypothesis that outer membrane lipids were inserted in concert with peptidoglycan, therefore more work will need to be done in this area to understand either why, or what other proteins may be facilitating this process.



RECAP- Figure 4.5.2 Timecourse of PqiB and MlaD localisation

a Histogram of Fluorescent maxima visualisation by average location dependent on time as determined by MicrobeJ b Fluorescent maxima mapping by density determined through MicrobeJ

I also later in Chapter 4 refined and developed a method for *Pseudomonas aeruginosa* specific membrane labelling *in vivo*, capable of visualising the fluidity of the membranes. This technique also allowed us to visualise individual lipid movements. (RECAP- Figure 4.2.2)



RECAP Figure 4.2.2 Lipid insertion to outer membrane during division is divisionally localised Propargyl- Choline

a- localisation of fluorescence in propargyl-choline timepulsed cells after 5 minutes incubation. b- FM4-64X localisation shows no midcell preference.

A connection between Outer membrane biogenesis and peptidoglycan

This thesis began with the hypothesis that the cell envelope is constructed through a series of interconnected architectures between the outer membrane and peptidoglycan. I have continued to work on this hypothesis. Throughout the thesis I have worked on aspects of the cell envelope architecture and how it is formed, in doing so showing a great degree of redundancy, for example in the lytic transglycosylases in Chapter 3 specifically, but also in Chapter 1 bioinformatically I show that a great deal of proteins interact and have similar roles. In Chapter 4, again this was shown in very similar growth data of mCherry fusion mutants to wildtype.

The hypothesis among this redundancy was an interconnection between layers, specifically peptidoglycan and the outer membrane. I postulated that if you increase or produce one layer of the cell envelope more than any other, you will get an imbalance. As mentioned in Chapter 1.2 and Chapter 4, the outer most lipid layer is less fluid than the cytoplasmic lipid layer, and migrates at a slow speed, with intermolecular bonds joining the LPS and beta barrel proteins together, the peptidoglycan too is a covalent supramolecular structure providing cell rigidity, but

is also added to and modified by a series of enzymes. I hypothesised in 2019, and presented in a successful grant application that this rigidity of both structures meant that systems which produce the outer most layer must work in tandem with peptidoglycan machinery, and later again presented this at a conference in Vermont in 2022.

However contrary to my hypothesis, in Chapter 4 and the work I presented, I have shown that the specific outer membrane balance proteins in *Pseudomonas aeruginosa* I investigated; PqiB, MlaD, which shuttle phospholipids and cardiolipin between the inner leaflet of the outer membrane and inner membrane were in fact not associated with a coordinated division role between outer membrane and peptidoglycan. This must be taken in the context of redundancy, and plurality of proteins with similar roles however; the proteins that I chose to investigate were one of only many proteins that facilitate a balance between the outer membrane and peptidoglycan, other research groups around the world did show an OM-PG connection in its outermost layer (111,165).

It is now known that the BAM complex which inserts beta barrels into the outer membrane is division coordinated (111), so too is the LPS insertion machinery the Lpt proteins which insert LPS, co-localising with peptidoglycan formation at new poles in polar growth bacteria(165). Therefore whilst I have not yet found a link, the hypothesis that linked my thesis chapters was sound. Although in my thesis work specifically I showed the MCE domain containing proteins are not part of a division process, the AsmA domain containing proteins (108,152) may show to be. I have set up a collaboration to have a set of constructs imaged and analysed in my absence by another group.

Whilst I did not have time to investigate the outer membrane-peptidoglycan connection further using the new lipid tracking method I developed in Chapter 4. I did show that finding lipid speeds and labelling bacterial membranes through pulsing cells with a eukaryote inspired click-labelable and soluble choline headgroup mimic is possible. This new technique of addition of ‘propargyl-choline’, a phosphatidylcholine mimic, to bacteria and then visualising click labelled lipids is replicable and works, and therefore in the future using super resolution microscopy it may be possible, especially using TIRF single molecule tracking to determine populations of lipids at different speeds which may reflect behaviours different cell envelopes, inner and outer. This new method therefore may in time develop my central hypothesis further to understand how they behave.

Therefore from a lipid standpoint my thesis work has furthered the field in multiple aspects, from a new method of lipid labelling in bacteria, to a nuanced understanding of OM-PG relations and the roles of MCE domain containing proteins. These MCE domain proteins are just one of multiple groups which have been hypothesised to have a role in OM biogenesis, and may not necessarily be division coordinated. This makes up part of a new literature with further understanding, shown by biochemical data which support the central hypothesis that the OM is connected to PG, but with exceptions to the rule, and a need for more research.

Peptidoglycan and the transglycosylation machinery

Although the central hypothesis of the thesis was that of a cell envelope connection, it is a project related to this, specifically on peptidoglycan, that will create the most scientific impact in terms of citations and potential drug discovery in the shorter term. As mentioned in Chapter 2, I principally in biochemical understanding, through assays, *in silica* prediction, modelling and cryo-EM structure building, with collaborators uncovered the mechanism of action for the RodA-PBP2 complex, more specifically on RodA. This action is responsible for the production of new cell wall by transglycosylation. Transglycosylation in the context of peptidoglycan is the creation of a bond between two sugar molecules, and this in a processive mode of continued enzymatic activity is what facilitates the formation of long chain glycan polymers. These polymers, when linked to existing peptidoglycan become part of the macromolecular structure that make up the cell wall. I visualised the activity of RodA and changes in activity in mutants, through a fluorescently labelled glycan group attached to RodA's substrate Lipid II, as well as through cryo-EM maps, flexibility simulations, thin layer chromatography and molecular dynamic docking experiments.

The RodA protein is a member of the SEDs protein complex family, and has a high similarity to other related proteins which are hypothesised to polymerise Lipid II, therefore not only is the mechanism of peptidoglycan formation proposed of RodA by my Chapter 2 impactful to this protein complex, which uses data in two organisms and systems SpoVDE, *B. subtilis* and RodA-PBP2 *E. coli*, but it is also useful for other SEDs complexes such as the division machinery protein and RodA orthologue FtsW, and the sporulation peptidoglycan formation machine SpoVDE. RodA is the protein most associated with elongation, however similar proteins work on sporulation, elongation and division, and as shown by our data working with *B. subtilis* they

also have very similar active sites. Interestingly, I show that RodA requires PBP2 to function efficiently, again showing how connected the two enzymes are, which was initially realised by my own co-evolution *in silica* work. This may have some bearing on future interactions uncovered between FtsW and FtsN/I which as indicated in my Chapter 1 (2), have also shown to have an array of interactions with other proteins and genes. FtsW is similar to RodA in structure, therefore likely follows the same mechanism of polymerisation of peptidoglycan with conserved residues taking part in this interaction, but it's action is likely modulated differently by other partner proteins. In conclusion to this work, my biochemical and *in silica* data was then added to a cryo-EM structure, to form the basis of a paper in Nature Communications, which I contributed most to in terms of data, mechanistic insight and understanding.

Also during my thesis, I was part of uncovering the similarity of SEDs proteins such as RodA, with a new family of GT-C transmembrane proteins, which bind Undecaprenyl phosphate through a shared motif, these GT-C transmembrane proteins also having at least seven transmembrane helices and located at the cytoplasmic membrane. This discovery in context of the literature, was published in Nature and focussed on the story of *E.coli*'s Waal O antigen ligase, a protein that is part of the lipopolysaccharide synthesis pathway, to form the outer membrane, in a paper which principally focussed on Waal's mechanism of action. Waal has a similar tertiary structure to RodA, but most prominently includes a TM helix similar to RodA's TM7, which interacts with Undecaprenyl phosphate. In this paper (9), my contribution was principally in the annotation and discoved similarity between RodA and Waal, through a shared conservation of an arginine residue, R210 in RodA and R242 in Waal, which was noted in the paper on Waals mechanism, and has already been referred to in the literature. This led in time to my later experiments in Autodock vina, which then were replicated in Molecular Dynamics simulation by Stansfeld group, using a similar active site location to Waal but instead in RodA, indicating a similar mechanism of action as well as inspiring many experiments aforementioned, which ended up pushing our mechanistic understanding of the complex forward.

RodA, and SEDs proteins often also have a partner protein. RodA has a transpeptidase, PBP2 which binds at a transmembrane helix, but also at the periplasmic face of RodA, adjacent to the active site. I made mutations in PBP2 in a linked protein, and also removed PBP2 from the construct to show that alone the RodA protein is less active. This work is more understood in the context of papers which suggest a movement of PBP2 that may affect RodA's activity (27,47). It is likely, that without a partner to make transpeptidase crosslinks, the transglycosylated products

would not be incorporated into new peptidoglycan, therefore PBP2 which has been shown to be inhibited by methicillin, and is lethal when knocked out must have a role in ensuring this insertion. However if the activity of PBP2 is knocked out alone, other proteins may be able to crosslink nascent peptidoglycan from RodA into the cell wall, such as the LD transpeptidases and other transpeptidases mentioned in Chapter 1 (2).

In terms of AMR, this work on RodA's mechanism of action may have a future drug discovery role. In Appendix 3, I suggest how new antibiotics might be found for this active site identified in RodA, with *in silica* docking, as well as active site prediction re-affirming the mechanism of action and suggesting potential drugs as well as binding sites. I believe, that new molecule discovery on the active site could be fruitful now the active site is known. I would suggest in future work if possible to have a lipid II molecule bound in a RodA or similar protein structure, I synthesised lipid II to this purpose, however during my PhD my collaborators were unable to attempt to make grids for Cryo-EM mapping of this ligand-protein complex.

Lytic mechanisms of action

Related to this work, my new method of using 'Reverse' Schägger gel assays to determine the mechanism by which peptidoglycan is later processed and hydrolysed by lytic enzymes, has been optimised by myself and a fellow lab member, and has already been used to confirm more than the example enzyme I present 'rSlt', as well the active site of new proteins. I hope that this assay will have a great impact in the field, especially on the lytic transglycosylases, which are a target for antibiotics and adjuvants in the future, such as Bulgecin, which has previously been shown to improve the efficiency of other antibiotics when given in coordination (17,146). By understanding the processivity of these enzymes, as well as optimal conditions for polymerisation, we can understand more on the cell envelope as a whole and its redundancy, and how these proteins might be balanced in a complex environment, explaining their redundancy touched on in Chapter 1.

Overall in the diversity of research presented, my PhD has begun to link the PG and OM processes across multiple species in literature, using bioinformatics, microscopic localisation tools, biochemistry and mass spectrometry. This thesis has also provided tools for research of a set of these enzymes in the Lytic transglycosylases as well as a new lipid dye for lipid study. My peptidoglycan work is being continued by the Roper group therefore will likely well integrate

into the field, alongside work by other groups on the lipid element which have confirmed the central hypothesis of the PhD.

The cell envelope and the processes that create each layer of which it is composed are clearly linked, and although all these connections are not yet known, through my work we are a small step closer to understanding its formation of its constituent peptidoglycan, but also lipid layers.

Appendices

Appendix 1: Creating Communities in Open Science

In addition to a focus on RodA-PBP2, a novel lytic transgylycosylase assay and the cell envelope connectivity of lipid homeostasis, during my PhD I worked on the creation of new open science programmes and movements. These movements were powered by non academics and technicians whom were empowered by grants to do good. Particularly interesting was that of a COVID19 relief and research fund, which we conducted that uses applicants as reviewers to speed up reviews and allow for large scale calls for applicants without a reviewer bottleneck. At the back of the thesis, and relevant to AMR, I attach this study, its predecessor in vaccine hesitancy, as well as a study on antimicrobrial resistance education and curricula creation in Africa, as they are relevant to the creation of new antibiotics in open source environments independent of private companies or NGOs.

Masselot C, Greshake Tzovaras B, Graham C LB, Finnegan G, Jeyaram R, Vitali I, Landrain T, Santolini M Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study

J Particip Med 2022;14(1):e32125 DOI:10.2196/32125: https://jopm.jmir.org/2022/1/e32125 Altmetric: 5

Chris LB Graham, Harry Akligoh and Joy King'Ori et al. Education-based grant programmes for bottom-up distance learning and project catalysis: Antimicrobial Resistance in Sub-Saharan Africa. *Accesss Microbiology.* AltMetric: 11

Chris LB Graham, Thomas E. Landrain, Amber Vjestica, Camille Masselot, Elliot Lawton, Leo Blondel, Luca Haenel, Bastian Greshake Tzovoras, Marc Santolini Community review: a robust and scalable selection system for resource allocation within open science and innovation communities *bioRxiv* 2022.04.25.489391; doi: https://doi.org/10.1101/2022.04.25.489391 https://www.biorxiv.org/content/10.1101/2022.04.25.489391v2 AltMetric : 16

Appendix 2: Thesis Supplementary Data:

Supplementary Data for Chapter 1.

Supplementary Information 1. Mur Ligase pathway

Fructose-6-phosphate is converted by four successive enzyme activities to uridine 5'-diphosphate-N-acetylglucosamine (UDP-GlcNAc). This is catalysed by GlmS, GlmM and GlmU (bifunctional enzyme) . UDP-MurNAc (5'-diphosphate N-acetylmuramic acid) is formed from UDP-GlcNAc using Mur ligases MurA and MurB. This results in a sugar moiety ready for pentapeptide addition.

A pentapeptide stem is then appended to the D-lactoyl carboxyl group of UDP-MurNAc by sequential addition of peptides by MurC-F: L-Ala (MurC), D-glutamic acid (MurD), meso-diaminopimelic acid (m-DAP) (MurE), dipeptide D-Ala-D-Ala (MurF), with D-Glu and m-DAP being synthesised from their L- or L,L- stereoisomers by MurI and DapF respectively, and D-ala-D-Ala being produced from L-Ala by alanine racemases and D-Ala-D-Ala ligase .

The UDP-MurNAc 5P produced by these reactions is then transferred to an undecaprenol, a membrane spanning lipid, yielding undecaprenyl diphospho MurNAc 5P (Lipid I), in a reaction catalysed by MraY at the inner membrane. Thereafter, MurG transfers a GlcNAc sugar moiety from UDP-GlcNAc to Lipid I, producing undecaprenyl diphospho MurNAc GlcNAc 5P (Lipid II).

Supplementary Information 2. After initial synthesis, peptidoglycan is modified

The modification of peptidoglycan extends far beyond the model in Figure 3, and continues after *de novo* peptidoglycan insertion, by a series of proteins and protein complexes referred to as PBPs. During synthesis a lytic transglycosylase (SI Figure 2) separates the existing strands of peptidoglycan to make space for new synthesis by cleaving the links between sugars, the glycan strand does not continue indefinitely and are typically between 7 and 32 sugars in length. New peptidoglycan must also be attached to the outer membrane by L, D transpeptidase action through linking peptidoglycan to outer membrane proteins like Lpp (SFigure 1.21), roughly once every 100 Å as suggested in Figure 1.1.2, to maintain cell envelope stability.

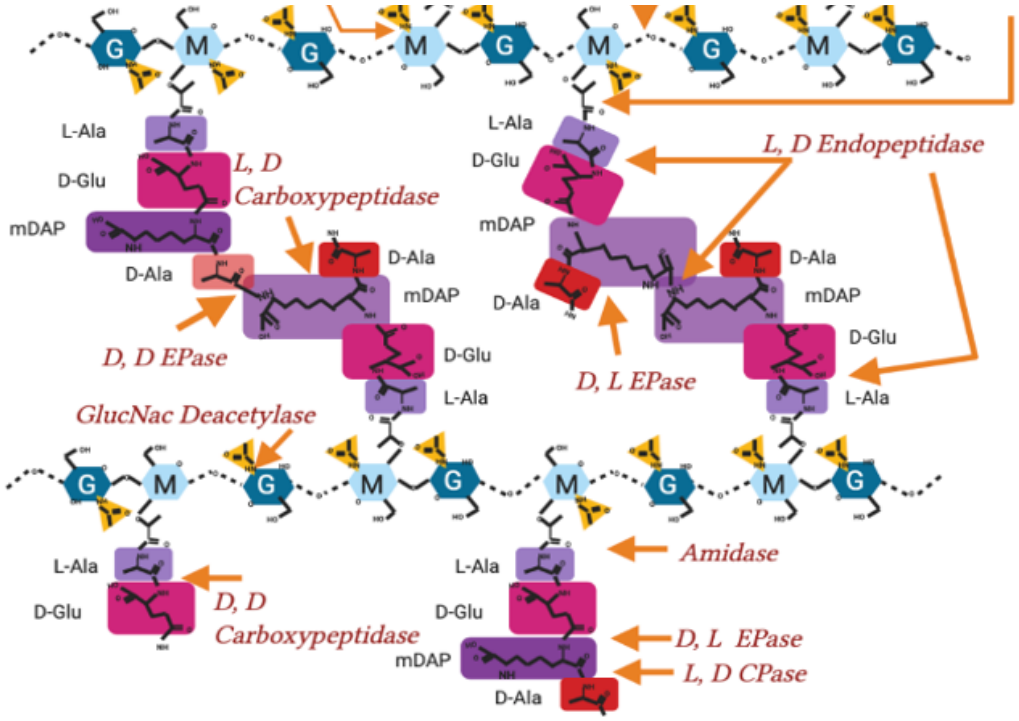
During division, the peptidoglycan crosslinks are continually broken and remade to relieve overall cell wall stress and facilitate growth (SI Figure 2). 3-4 crosslinks being predominant when the cell is not in stress and early in growth, whereas 3-3 crosslinking is found in cells increasingly during the stationary phase, or following antibiotic exposure and osmotic shock (SI Figure 2)

A shift from 3-4 to 3-3 crosslinking often also occurs during cell growth. Fluorescent D-amino acids (FDAA) have been used to label transpeptidase activity through the ability of PBPs to exchange amino acids, revealing that during stationary phase and growth, the entire peptidoglycan is “lit up” by incorporation of new FDAA indicating new modifications are being made throughout growth.

Bacterial cell elongation and division requires the peptidoglycan layer to be constantly modified and cleaved to allow for growth. The attachment of protein partners to the peptidoglycan layer and the peptidoglycan recycling process additionally requires peptidoglycan cleavage. The cleavage and modifications of peptidoglycan varies across species and can be broadly split into two classes of enzymatic action: hydrolase and transferase. (SI Figures 1 and 2)

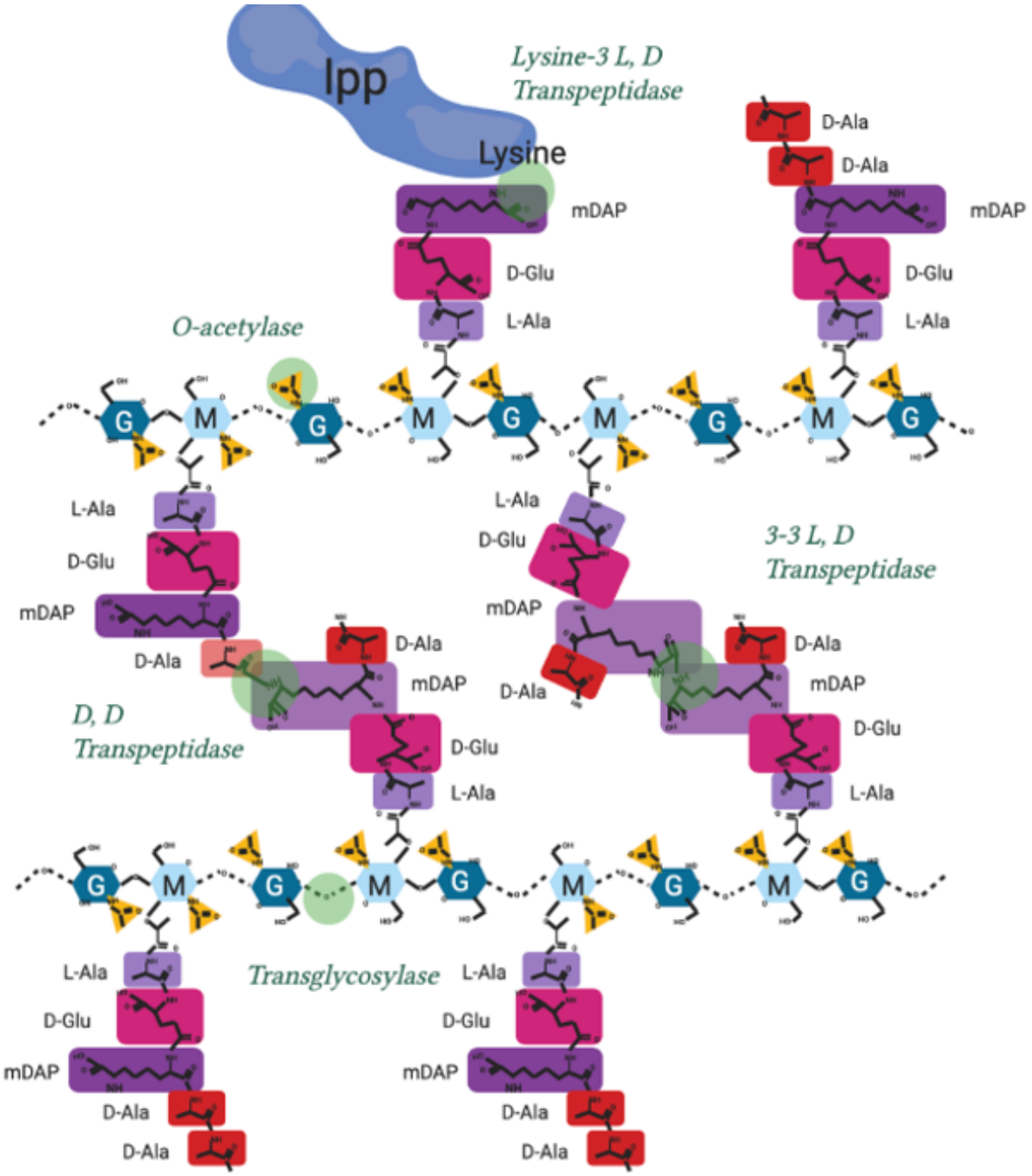
Hydrolases carry out a range of lytic modifications to peptidoglycan, including cleavage of the peptide stem at the glycosidic bond between glycan molecules, and the removal of acetyl groups (a lysozyme resistance factor in pathogenic strains) (SI Figure 1). The hydrolases so far characterized are dispersed throughout the cell periplasm at the lateral cell wall or the division plane

Hydrolases are controlled by regulatory proteins and each hydrolase has their own distinct role within the cell and the cell's complexes (Table 1).



SI Figure 1. Hydrolase activity on peptidoglycan in Gram-negative bacteria.

Peptidoglycan molecular structure, with sites of hydrolysis and enzyme nomenclature labelled. EPase- Endopeptidase, CPase- Carboxypeptidase

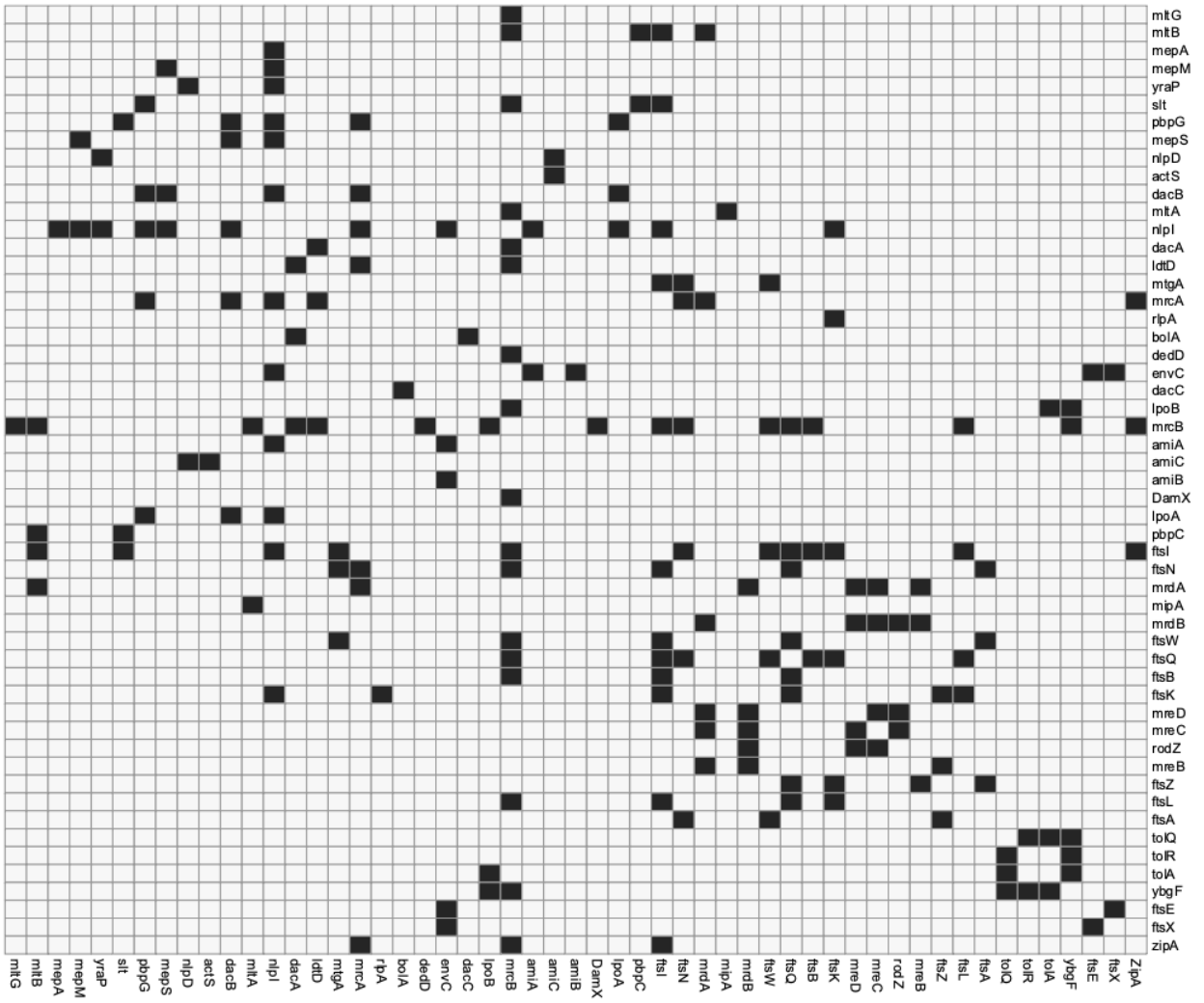


SI Figure 2 Peptidoglycan crosslinking sites and activity in Gram-negatives

Peptidoglycan molecular structure, with sites of crosslinking actions and enzyme nomenclature labelled.

Peptidoglycan is also polymerised and modified by a series of crosslinking enzymes such as the penicillin binding proteins (PBPs) or L, D transpeptidases at different points in the bacterial cell cycle as well as after stress events. The modification sites shown in Figure 1.21 are the points of peptidoglycan crosslinking and attachment known in Gram-negative bacterial species.

Their individual activities are elaborated in Table 2.



SI Figure 3 Interaction matrix of PG related proteins

genes or proteins known to interact and confirmed by the literature. Black = confirmed interaction, White = unknown/self/no interaction. Please find attached pg_network.csv file for DOI matrix of confirmed interactions for above.

SI Table 1. Characterised Hydrolases in Gram-negative bacteria

Peptidoglycan degradation/Hydrolases		
Function	Enzymatic Action	Known Genes/Protein
D, D Carboxypeptidases	D-Ala D- Ala Cleavage 4-5	dacA, yfeW, dacC, dacD, vanY, ampH,Csd3*
MurNac DeAcetylase	Deacetylation of N-acetyl Muramic acid	pgdA
GlutNac Deacetylase	Deacetylation of N-acetyl Glucosamine	
Amidase	Cleavage of peptide stem from Glycan strand	amiA, amiB, amiC, amiD,ampD, mpaA
Lytic Transglycocylase	Breaking Glycan strand at GlucNac-MurNac(endo)	Slt, MltA, MltB, MltC, MtD, MltE, PilT, traB, virB1, rlpA, MltG
	Breaking Glycan strand at GlucNac-MurNac(exo)	NagZ
L, D Carboxy/Endopeptidase	mDAP mDAP cleavage 3-3	mepA
	mDAP- Lpp Cleavage	YafK/LdtF

	mDAP D-Ala cleavage 3-4	pgp2*, csd6*
	mDAP-D Glucosamine cleave 3-2	csd4*
D, D Endopeptidase	Cleavage of D-Ala-mDAP crosslink 3-4	dacB, pbpG, MepS, MepM, PBP7, MepH

*Not present in Escherichia *coli* MG1655 genome, but present in other species.

SI Table 2. Characterised peptidoglycan synthases in Gram-negative bacteria

Function	Enzymatic action	Known Genes/Protein (E.coli)
DD Transpeptidase and Transglycocylase	Adds lipid II to nascent strand and crosslinks into existing PG	mrcA
		mrcB
DD Transpeptidase	Crosslinks nascent strand into existing peptidoglycan	mrdA FtsI
Transglycosylase	Adds lipid II to nascent strand	mtgA, rodA, ftsW
Flippase	Flips Lipid II to periplasm	murJ
LD Transpeptidases	Peptidoglycan Brauns lipoprotein crosslinkers	LdtA/ErfK,
		YbiS/LdtB

			Ycfs/LdtC
	Peptidoglycan crosslinkers	3mDAP-3mDAP	YnhG/LdtE
			YcbB/LdtD
O- acetylation	O-acetylates nam		oatA*
			adr
			pacA

*Not present in Escherichia *coli* MG1655 genome, but present in other species.

SI Table 3. Peptidoglycan associated and PBP regulatory proteins

Theoretical Function	Known genes/protein (E.coli)
Moderate class A PBP activity	LpoA
	LpoB
Alter interactor ability	CpoB
Bind OM with peptidoglycan	OmpA

Moderate OM linkage with peptidoglycan	lpp
Periskeletal elongasome component, moderates PBP2 activity	MreC
Periskeletal elongasome component, moderate PBP2 activity	MreD
Treadmilling Cytoskeletal elongasome component	MreB
Elongasome staple component	RodZ
SPOR domain containing proteins, protein interaction	Rlpa
	FtsN
	DamX
	SpoX
Hydrolase binding activity	NlpI
	ActS
	NlpD

	EnvC
Inner membrane peptidoglycan moderation	TolA,TolR, TolQ
	palA, palB
Division moderation and EnvC control	FtsX, FtsE
Treadmilling cytoskeletal component for divisome	FtsZ
Division moderation	FtsA
Helicase and PBP interactor	FtsK
FtsN interactor and division start	FtsB
	FtsL
	FtsQ
FtsZ interactor and PIPs mediator	ZipA

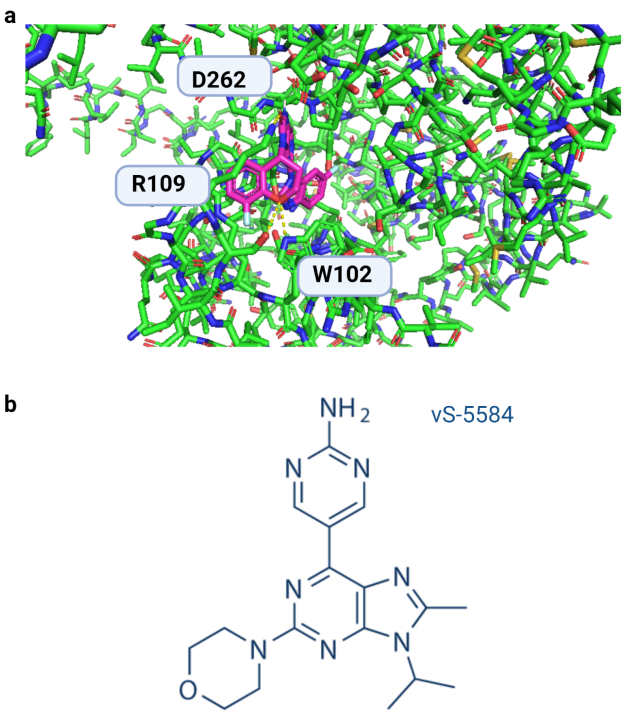
Supplementary Data for Chapter 4.

GIFs of propargylcholine click labelled lipid movements

https://drive.google.com/file/d/1c77Fqm3cbA5RYsM94975deZDZDoUE0mc/view?usp=sharing and https://drive.google.com/file/d/1ArmTBPJKB4bpIuFkKCnvepX-5EVovovY/view?usp=sharing

Appendix 3: New inhibitors to bind RodA-PBP2

After the PhD, I recommend the Roper laboratory and my collaborators continue research into RodA. After having uncovered the active site in Chapter 2, I then continued research on the RodA molecule i*n silico* using OpenMTI and AutodockVina using existing techniques(118). Briefly, this method, along with future techniques using tools such as DiffDock are a starting point for new inhibitor discovery. (SI Figure 4).



SI Figure 4. Binding of a potential inhibitor to the RodA , SEDs transglycosyltransferase active site.

a Binding of vs-5584 to *E.coli* RodA-PBP2 according to Autodock Vina binding energy -9.9kcal
b vS-5584 chemical formula

In SI Figure 4, I assayed 3100 potential inhibitors from existing antimicrobrial and drug libraries. The binding of these molecules were tested across the hypothesised Cavity A and B established in Chapter 2. An inhibitor which looks best in the top 15 of these inhibitors terms of interacting with multiple predicted active site residues is vs-5584 (SI Table 4), with a predicted binding energy of -9.9kcal/mol. This molecule fits between the integral residues identified in peptidoglycan catalysis, the active site Aspartate , R109 and W102, with a highly aromatic structure.

Compound	Model ID	Energy	nRot

R428_ZINC000051951669	1	-11.1	5
Golvatinib_ZINC000043195317	1	-10.6	8
Irinotecan_ZINC000001612996	1	-10.6	6
Mk3207_ZINC000103760981	1	-10.5	4
Mk3207_ZINC000103760984	1	-10.2	4
Telcagepant_ZINC000028827350	1	-10.2	5
Vindesine_ZINC000008214470	1	-10.1	10
Pamaqueside_ZINC000257972190	1	-10.1	13
Tg100-801_ZINC000029136020	1	-10	10
Proscillaridin_ZINC000056897526	1	-10	7
Azd3514_ZINC000101673084	1	-9.9	7
Vs-5584_ZINC000049090010	1	-9.9	2
Lifitegrast_ZINC000084668739	1	-9.9	7
Sarolaner_ZINC000103297729	1	-9.9	5
Merestinib_ZINC000095926668	1	-9.8	6

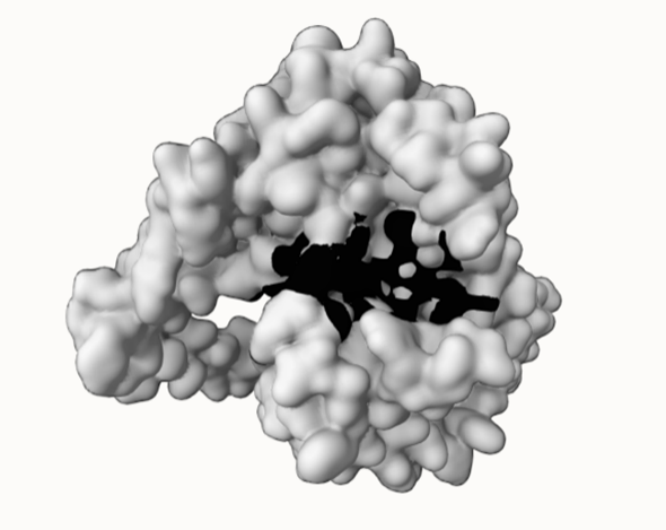
SI Table 4 Top 15 RodA inhibitors.

The inhibitor with the highest binding to RodA at the proposed binding site (Table of potential RodA inhibitors) selected from currently used antibiotics and drugs, was R428_ZINC000051951669 along with a range of other potential inhibitors, with Kcal values above -10. As these drugs are already used for other therapies, R428 especially (for chemotherapy) it would be wise to look through the top 100 hits of the 3100 drugs assayed for binding, and filter for those drugs with no toxic effect on humans, which chemotherapy drugs for example innately have. A screen of these drugs using our existing assay established in Chapter 2 would be a potential inhibitor finding opportunity. If these inhibitors needed to be reduced in number, use of another *in silico* docking tool such as Diffdock could act as a

verification for the Autodock Vina methodology. Diffdock is a molecular dynamics based ligand binding tool, and has potential for any project with an *in silico* aspect to ligand binding.

Whilst looking for alternative programmes for confirmation of the ligand binding in RodA, I also came across PrankWeb. PrankWeb uses a conservation based approach, which I touch on in my conservation and co-evolution ligand finding model, as a ligand binding site predictor, but also incorporates deep learning to rank these poses.

After a submission of RodA, with the K97- R109 loop added to this server, it found the predicted binding site of Chapter 2 to be the most likely binding area, as a single binding pocket. 'Pocket rank:1, Pocket score: 45.14, Probability score:0.970, AA count: 33, Conservation: 2.4, Residues: A_101, A_102, A_108, A_109, A_111, A_114, A_117, A_159, A_160, A_161, A_162, A_258, A_260, A_261, A_262, A_317, A_331, A_332, A_333, A_334, A_335, A_341, A_342, A_343, A_344, A_345, A_347, A_348, A_48, A_52, A_55, A_96, A_97' - PrankWeb predicted site of E.coli RodA- Lipid II binding. I have highlighted the residues identified by ourselves and confirmed to be important in this ligand binding. Therefore, after writing the paper. I would suggest that future work incorporate PrankWeb as a preliminary binding site finding tool which can then be used for box selection for Autodock ligand binding predictions. Alternatively AlphaFill would be useful for any conserved binding sites already known to science.



SI Figure 5. RodA predicted binding by PrankWeb

The PrankWeb server, surface representation of RodA, with the loop modelled in. The binding site agrees with the data and theory of Chapter 2, including D262, R101and R48 respectively.

Appendix 4: Sum neighbourhood co-evolution and amino acid conservation as an improved catalytic site determinant

Abstract

The elucidation of active sites in proteins is especially difficult without known protein structures. A traditional tool for discovery of active site integral residues involves site directed mutagenesis, conservation of proteins amino acids however already shows residues likely needed for function. Here we use the sum scores of neighbouring residues, in combination with neighbourhood conservation scores to multiply an amino acid's conservation score and determine amino acids and regions of particularly high interest in a set of peptidoglycan modifying proteins. I demonstrate in this paper the ability for this simple transformation to discover active sites without any experimental, or structural information.

Introduction

Co-evolution scores are determined by an amino acids likelihood of evolving at the same time as another amino acid, this indicates this amino acid pair is important- and likely has contact spatially or in mechanisms of action/interaction.

Conservation scores/Entropy of an amino acid within a protein is the measure of the likelihood or importance of that amino acid in the protein.

Highly conserved amino acids have extremely high entropy scores- as it takes a lot for them to remain the same amino acid, if these amino acids can change between others in that group the entropy remains high- therefore I choose entropy as the measurement to care about as opposed to raw conservation.

The problem is, co-evolution can happen just due to spatial importance in structure with little bearing on active site residues, so to find out if a residue is important we can create a co-evolution umbrella and add up all of its relative co-evolution scores (found using gremlin using a very simple algorithm but a big computer).

What we do here is we take the co-evolution score and conservation scores and give bins of co-evolution score to high entropy amino acids. I also donate bins of local conservation scores to the high co-evolution scores. This creates a combined local conservation co-evolution score which we then multiply by our entropy to create a region relevant. This could be a new way to pick integral residues for binding site searching, with essentially an super enriched conservation score and visualisation on 3D structures. Further to this, clustering of such high score residues could be an identification technique for new binding sites. This work after its first iteration, was abandoned in its first principles state, after it was later confirmed by P2Rank as a method for improvement in binding site identification.

Method

Amino acid “Evolutionary significance” scoring

Co-evolution scores were pulled from the pre-existing “ECOLI” database on the GREMLIN server by Bakerlab. Conservation scores are entropy scores determined by the Weblogo3’s algorithm performed on the ECOLI database pre-existing MSA.

1. Search for Gene- - https://gremlin2.bakerlab.org/preds.php?db=ECOLI

Search for protein of interest, there is a list of available proteins on GREMLIN, including pre-existing multiple sequence alignments.
Find domain classified Multiple Sequence alignment MSA- https://gremlin2.bakerlab.org/db/ECOLI/fasta/P0AD65.fas Pull MSA as available for each gene of interest from Gremlin

2. Entropy- http://weblogo.threeplusone.com/create.cgi

Copy conservation, or entropy, calculated from MSA’s pulled from publicly available datasets to Weblogo and print as plain text,

3. Contacts- https://gremlin2.bakerlab.org/preds_cst.php?db=ECOLI&id=P0AD65

On the Gremlin site, with the listed protein, on self-co-evolution extract amino acid and entropy- then vectorise with i and j, and add column entropy

i	j	i_id	j_id	rawscore	s_sco	prob	prob_cb	prob_min5

4. Co-evolution and conservation data matching After having collected co-evolution of the protein, as well as the conservation scores, the scores can be combined to create a score of genetic significance.

5. Transformation of raw data to an amino acid value Before moving to step 5, to allow for a scoring function that is applicable to all proteins, a new value must be made, *"Specific Entropy"* - referring to the entropy special of each amino acid in this protein, scores for each amino acid "n" were determined by the following where "n" refers to amino acid number.

(*Maximum Entropy score of protein* "E" ÷ *Entropy score of amino acid* "e_n")^2 = *Specific Entropy of amino acid "*ϵ_n*"*

The co-evolution value of an amino acid, instead refers to the specific value of an amino acid to other amino acids, in the top 1000 co-evolution interactions. *Sum Co-evolution score* "C_n"- referring to the sum score of all co-evolution interactions "c" for each residue "n" with all other protein residues.

6. Value of localised regions Very conserved or integral residues are often accompanied by highly valued residues. It is therefore valuable to find specific peaks of localised regions with value to a gene.

The value of each amino acid, before localised value is, the *Specific Entropy-C and "Sum Co-evolution score*" of each amino acid. This is then in the absence of a deep learning algorithm to find the optimum localised amino acids for such an algorithm given a shifting box of scores all within a shifting 10 amino acid sequence of one another referred to here as "Local Entropy/Co-evolution" or "L". This is calculated as follows where "n" refers to the amino acid number.

"Local Sum Co-evolution" ϵ*[n -5, n+5]* * *"Local Sum Co-evolution"* $C_{[n-5,\, n+5]} = L_n$

or "Local Entropy/Co-evolution"

The "Local Entropy/Co-evolution" L_n, is then multiplied by the specific entropy of amino acid n, "ϵ_n" to give the "Evolutionary significance" score G_n, which is a compound value of both local and amino acid specific entropy and co-evolution scores, thereby highlighting amino acids which have multiple evolutionary partners locally in the amino acid sequence, but themselves are also of high conservation, thereby reducing signal noise and highlighting the residues of highest "Evolutionary significance"

$$\left(\epsilon_{[n-5,\, n+5]} * C_{[n-5,\, n+5]}\right) * \epsilon_n = G_n \quad \textit{"Evolutionary significance Score"}$$

Normalisation of above performed using G_n/G_{max} = Normalised G_n

6. Visualisation of amino acid "Evolutionary significance" scoring on to structural information
Once a table of amino acid significance by evolution is established 'Normalised G' for each amino acid 'n', to create a usable GUI, the amino acids are coloured in ribbon format, to produce visible regions of high interest.

Code in Python- using newBfactors, and order as amino acid number, vector of Normalised Gn in list.

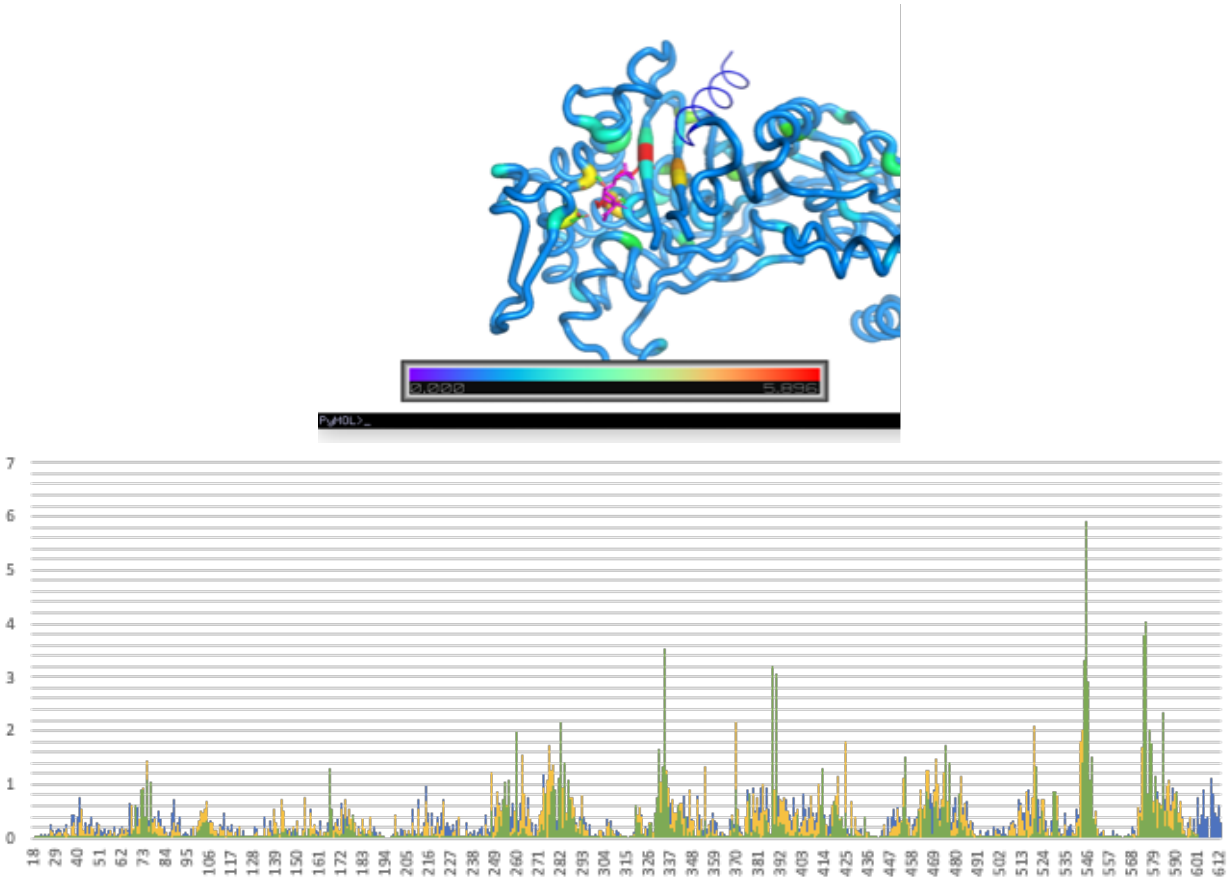
"

```
from pymol import cmd, stored, math

def loadBfacts (mol,startaa=1,source="newBfactors.txt", visual="Y"):
        """
        Replaces B-factors with a list of values contained in a plain txt file

        usage: loadBfacts mol, [startaa, [source, [visual]]]

        mol = any object selection (within one single object though)
        startaa = number of first amino acid in 'new B-factors' file (default=1)
        source = name of the file containing new B-factor values (default=newBfactors.txt)
        visual = redraws structure as cartoon_putty and displays bar with min/max values
(default=Y)

        example: loadBfacts 1LVM and chain A
        """
        obj=cmd.get_object_list(mol)[0]
        cmd.alter(mol,"b=-1.0")
        inFile = open(source, 'r')
        counter=int(startaa)
        bfacts=[]
        for line in inFile.readlines():
```

```
                bfact=float(line)
                bfacts.append(bfact)
                cmd.alter("%s and resi %s and n. CA"%(mol,counter), "b=%s"%bfact)
                counter=counter+1
        if visual=="Y":
                cmd.show_as("cartoon",mol)
                cmd.cartoon("putty", mol)
                cmd.set("cartoon_putty_scale_min", min(bfacts),obj)
                cmd.set("cartoon_putty_scale_max", max(bfacts),obj)
                cmd.set("cartoon_putty_transform", 0,obj)
                cmd.set("cartoon_putty_radius", 0.2,obj)
                cmd.spectrum("b","rainbow", "%s and n. CA " %mol)
                cmd.ramp_new("count", obj, [min(bfacts), max(bfacts)], "rainbow")
                cmd.recolor()

cmd.extend("loadBfacts", loadBfacts);
“
```

Results

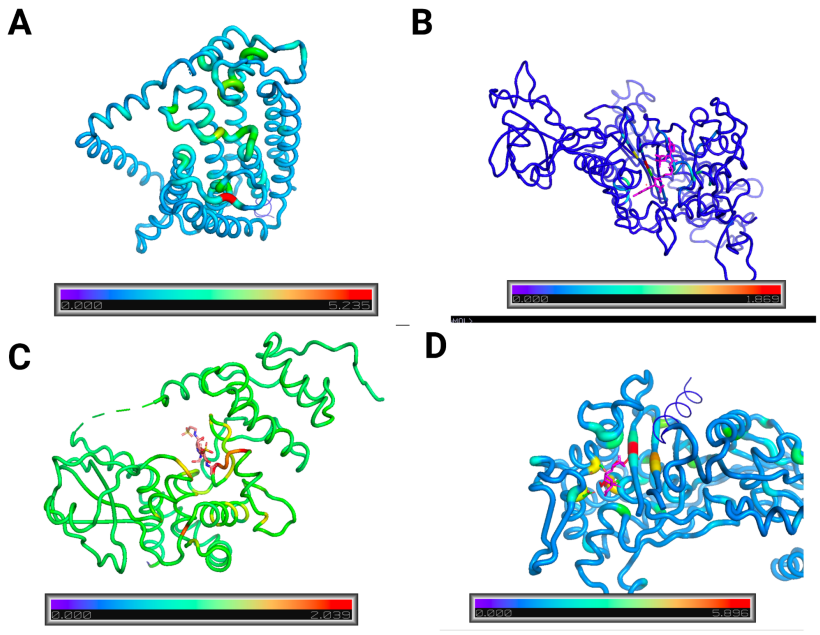
In SI Figure 1, the co-evolutionary likelihood sum score ‘C’ is compared to conservation score, and combined localised scores, when mapped to a protein. This reveals a highly pinpointed visible representation of a potential binding site. PBP2 is used in this case as an example.



SI Figure 6. Improved datasets created by co-evolution and conservation mixing, allow for sequence only active site discovery.

Panel A E.coli PBP2 representation of the penicillin binding site, correctly predicted by the conservation value determinant algorithm. Panel B Co-evolution scores- blue, Entropy- yellow , Combined evolutionary score- green.

In SI Figure 6, the peaks of increased evolutionary value are shown to be clearer than Entropy or co-evolution information alone, with regions of specifically high scores highlighting a potential binding pocket. The use of PBP2 as an initial test however could not show the re-usability of such a method in ligand binding determination. Therefore the method was then applied to other proteins in the PhD, RodA, PBP1a and MltB (SI Figure 7). This was shown to correctly predict binding sites.



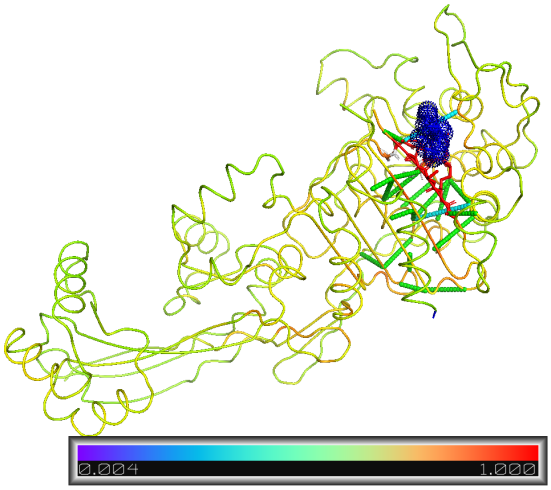
SI Figure 7. Repeat of formula across protein types and catalytic specificity

The success of the use of a combined evolutionary value score ‘G’ was an indicator that sequence information alone, could predict ligand binding siteswithout any information or outside manual assistance for ligand searching. However the automated prediction of amino acidsclose to, or involved in binding was not yet possible. However, if the distance between amino acids with high ‘G’ scores could be calculated and clustered, binding sites could be discovered without use of GUI information, or automatically annotated.

Co-evolution cluster analysis of identified regions allows for binding site prediction without docking.

Use of the PDB, or AlphaFold, and distance modelling of members of the high G scores in a protein, could allow for a new additional score, similar to sequence locality, but in the 3D space for localised ligand binding prediction. As an initial test for this method, I used a script by Phil Stansfeld capable of uncovering distance clusters after a distance value is assigned. This revealed all amino acids in the top 20 co-evolution pairs within 10A, as a new layer on the pdb images as determined by Gremlin , overlaying the G score calculation, this initial, low computational load of local clustering could be used to find regions with both high co-evolutionary clustering, suggesting amino acid partnerships as well as conservation. SI Figure 8 represents this result in PBP2. I have since created a new co-evolution labelling tool based on

these ideas, but for protein-protein interaction called CoEvFold, in a google collab notebook, and shared this among my lab group. This is available on request.



SI Figure 8 Top 20 co-evolution pairs chosen for visualisation, Global evolution score visualised using colour bar.

Whilst this was informative and set an underlying precedent for use of Conservation in ligand binding studies, I did not go any further. Alongside my own first-principles based work, the 'PrankWeb' server (161) later released a paper in July 2022 with a similar notion and improved throughput. P2Rank and later PrankWeb, is based on Deep learning algorithms of existing ligand binding in proteins, and performs ligand docking searching similar to Autodock (Appendix 5), however these are also supplemented by conservation of amino acids as a ligand site determinant as well as deep learning. This use of conservation scores, as a supplement to ligand binding studies increased the accuracies of ligand binding prediction compared to any previous model, and confirms the use of conservation as a part of ligand docking. This algorithm has been applied to the entire AlphaFold and PDB database to reveal the top 6 possible binding sites for each protein. Understanding some of the mechanisms behind this work, and likely shortcuts taken by the deeplearning, I hope will make my future work in ligand binding mode identification useful.

Appendix 5: Papers published during PhD

(2023). Education-based grant programmes for bottom-up distance learning and project catalysis: Antimicrobial Resistance in Sub-Saharan Africa, CLB. Graham* Harry Akligoh, Joy King'Ori, Gameli Adzaho, Linda Salekwa, Courage KS Saba, Thomas Landrain Marc Santolini, *Access Microbiology* https://doi.org/10.1099/acmi.0.000472.v1

(2022) Structural basis of Lipopolysaccharide Maturation by the WaaL O-Antigen Ligase, KU. Ashraf, R Nygaard, ON. Vickery, SK. Erramilli, CM. HT McConville, V Petrou, S Giacometti, M B. Dufrisne, K Nosol, AP. Zinkle, CLB. Graham, B Kloss, K Skorupinska-Tudek, E Swiezewska, David I. Roper,, Anne-Catrin Uhlemann, OB. Clarke, A Kossiakoff, M. Stephen Trent, P Stansfeld and F Mancia *Nature* 604, 371–376 (2022). https://doi.org/10.1038/s41586-022-04555-x

(2021) A dynamic network of proteins facilitate cell envelope biogenesis in Gram-negative bacteria. Review, CLB. Graham, H Newman, F Gillet, K Smart, N Briggs, M Bhanzaf, D I Roper *Int. J. Mol. Sci.* 2021, *22*(23), 12831; https://doi.org/10.3390/ijms222312831

(2021) Co-Immune: a case study on open innovation for vaccination hesitancy and access, C M. Masselot, BG Tzovaras, CLB. Graham, M Santolini Journal of Participatory medicine https://doi.org/10.1101/2021.03.29.20248781

(2021) Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs Beal J, Baldwin GS, Farny NG, Gershater M, Haddock-Angelli T, et al. Comparative analysis of three studies measuring fluorescence from engineered bacterial genetic constructs. PLOS ONE 16(6): e0252263. https://doi.org/10.1371/journal.pone.0252263

(2022). A tested model for an open and applicant driven grant allocation system CLB. Graham* and T Landrain, A Vjestica, C Masselot, B G Tzovoras, L Blondel, L Haenel, M Santollini* , F1000 research and BioRxiv (https://www.biorxiv.org/content/10.1101/2022.04.25.489391v1.full and https://f1000research.com/articles/11-1440/v1)

Publications in review

(2023) Cryo-EM structure of the Peptidoglycan Elongasome complex, and Lipid II biosynthesis mechanism, R Nygaard, CLB Graham et al, (Under Revision- *Nature Communications*) - Attached

(2023) Pseudomonas MCE domaining protein localisation and phospholipid insertion, CLB Graham et al (Attached- submitting to *Microbiology*)

(2023) Phosphatidyl choline mimic propargyl choline for membrane staining and phospholipid tracking in *Pseudomonas Aeruginosa*. CLB Graham et al (Attached- submitting to *Microbiology Access*)

Publications in Progress

(2023) *E.coli* Lytic transglycosylase recognition and substrate binding, assayed by fluorescent substrates , F Gillet and CLB Graham et al, (In progress)

The above papers are attached at the back of the thesis for discussion.

References

1. Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. Nat Commun. 2019 Dec 2;10(1):5477.
2. Graham CLB, Newman H, Gillett FN, Smart K, Briggs N, Banzhaf M, et al. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-Negative Bacteria. Int J Mol Sci. 2021 Jan;22(23):12831.
3. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet. 2022 Feb;399(10325):629–55.
4. Cho H, Wivagg CN, Kapoor M, Barry Z, Rohs PDA, Suh H, et al. Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-Autonomously. Nat Microbiol. 2016;
5. Contreras-Martel C, Martins A, Ecobichon C, Trindade DM, Matteï PJ, Hicham S, et al. Molecular architecture of the PBP2-MreC core bacterial cell wall synthesis complex. Nat Commun. 2017;
6. Sjodt M, Rohs PDA, Gilman MSA, Erlandson SC, Zheng S, Green AG, et al. Structural coordination of polymerization and crosslinking by a SEDS–bPBP peptidoglycan synthase complex. Nat Microbiol. 2020 Jun;5(6):813–20.
7. Guilmi AMD, Bonnet J, Peiβert S, Durmort C, Gallet B, Vernet T, et al. Specific and spatial labeling of choline-containing teichoic acids in Streptococcus pneumoniae by click chemistry. Chem Commun. 2017;53(76):10572–5.
8. Jao CY, Roth M, Welti R, Salic A. Metabolic labeling and direct imaging of choline phospholipids in vivo. Proc Natl Acad Sci. 2009 Sep 8;106(36):15332–7.
9. Ashraf KU, Nygaard R, Vickery ON, Erramilli SK, Herrera CM, McConville TH, et al. Structural basis of lipopolysaccharide maturation by the O-antigen ligase. Nature. 2022 Apr;604(7905):371–6.
10. Isom GL, Davies NJ, Chong ZS, Bryant JA, Jamshad M, Sharif M, et al. MCE domain proteins: Conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. Sci Rep. 2017;
11. Lambert C, Lerner TR, Bui NK, Somers H, Aizawa SI, Liddell S, et al. Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. Sci Rep. 2016;
12. Graham CL, Akligoh H, Ori JK, Adzaho G, Salekwa L, Campbell P, et al. Education-based grant programmes for bottom-up distance learning and project catalysis: Antimicrobial Resistance in Sub-Saharan Africa. Access Microbiol [Internet]. 2023 [cited 2023 Jan 24]; Available from: https://www.microbiologyresearch.org/content/journal/acmi/10.1099/acmi.0.000472.v1
13. Graham CL. Community review: a robust and scalable selection system for resource allocation within open science and innovation communities. bioRxiv. 2022;
14. Lambert C, Chang CY, Capeness MJ, Sockett RE. The first bite - Profiling the predatosome in the bacterial pathogen Bdellovibrio. PLoS ONE. 2010;

15. Brives C, Pourraz J. Phage therapy as a potential solution in the fight against AMR: obstacles and possible futures. Palgrave Commun. 2020 May 19;6(1):1–11.
16. Fleming A. Penicillin. Br Med J. 1941 Sep 13;2(4210):386.
17. Williams AH, Wheeler R, Thiriau C, Haouz A, Taha MK, Boneca IG. Bulgecin A: The Key to a Broad-Spectrum Inhibitor That Targets Lytic Transglycosylases. Antibiotics. 2017 Feb 22;6(1):8.
18. Dickman R, Mitchell SA, Figueiredo AM, Hansen DF, Tabor AB. Molecular Recognition of Lipid II by Lantibiotics: Synthesis and Conformational Studies of Analogues of Nisin and Mutacin Rings A and B. J Org Chem. 2019 Sep 20;84(18):11493–512.
19. Sousa MC. New antibiotics target the outer membrane of bacteria. Nature. 2019 Dec;576(7787):389–90.
20. Egan AJF, Cleverley RM, Peters K, Lewis RJ, Vollmer W. Regulation of bacterial cell wall growth. FEBS J. 2017;
21. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiol Rev. 2008;
22. Lutkenhaus J, Pichoff S, Du S. Bacterial cytokinesis: From Z ring to divisome. Cytoskeleton. 2012;
23. Barreteau H, Kovač A, Boniface A, Sova M, Gobec S, Blanot D. Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiology Reviews. 2008.
24. Vollmer W. Structural variation in the glycan strands of bacterial peptidoglycan. FEMS Microbiology Reviews. 2008.
25. Bernal-Cabas M, Ayala JA, Raivio TL. The Cpx envelope stress response modifies peptidoglycan cross-linking via the L,D-transpeptidase LdtD and the novel protein YgaU. J Bacteriol. 2015;
26. Pazos M, Peters K, Vollmer W. Robust peptidoglycan growth by dynamic and variable multi-protein complexes. Current Opinion in Microbiology. 2017.
27. Mueller EA, Egan AJ, Breukink E, Vollmer W, Levin PA. Plasticity of Escherichia coli cell wall metabolism promotes fitness and antibiotic resistance across environmental conditions. eLife [Internet]. 2019 Apr;8. Available from: https://elifesciences.org/articles/40754
28. Gray AN, Egan AJF, van't Veer IL, Verheul J, Colavin A, Koumoutsi A, et al. Coordination of peptidoglycan synthesis and outer membrane constriction during Escherichia coli cell division. eLife. 2015;
29. Wollrab E, Özbaykal G, Vigouroux A, Cordier B, Simon F, Chaze T, et al. Transpeptidase PBP2 governs initial localization and activity of major cell-wall synthesis machinery in Escherichia coli. bioRxiv. 2019;
30. Vischer NOE, Verheul J, Postma M, van den Berg van Saparoea B, Galli E, Natale P, et al. Cell age dependent concentration of Escherichia coli divisome proteins analyzed with ImageJ and ObjectJ. Front Microbiol. 2015;
31. van der Ploeg R, Goudelis ST, den Blaauwen T. Validation of FRET assay for the screening of growth inhibitors of escherichia coli reveals elongasome assembly dynamics. Int J Mol Sci. 2015;
32. Van der Ploeg R, Verheul J, Vischer NOE, Alexeeva S, Hoogendoorn E, Postma M, et al. Colocalization and interaction between elongasome and divisome during a preparative cell division phase in Escherichia coli. Mol Microbiol. 2013;
33. Daniel RA, Errington J. Control of cell morphogenesis in bacteria: Two distinct ways to make a rod-shaped cell. Cell. 2003;
34. Roure S, Bonis M, Chaput C, Ecobichon C, Mattox A, Barrière C, et al. Peptidoglycan maturation enzymes affect flagellar functionality in bacteria. Mol Microbiol. 2012;
35. Sycuro LK, Rule CS, Petersen TW, Wyckoff TJ, Sessler T, Nagarkar DB, et al. Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence

Helicobacter pylori morphology. Mol Microbiol. 2013;
36. Sycuro LK, Pincus Z, Gutierrez KD, Biboy J, Stern CA, Vollmer W, et al. Peptidoglycan Crosslinking Relaxation Promotes Helicobacter pylori's Helical Shape and Stomach Colonization. Cell. 2010 May;141(5):822–33.
37. Kris Blair. Exploring Mechanisms of Cell Shape Control in Helicobacter pylori. Thesis. 2018;
38. Turner RD, Vollmer W, Foster SJ. Different walls for rods and balls: The diversity of peptidoglycan. Molecular Microbiology. 2014.
39. Turner RD, Hobbs JK, Foster SJ. Atomic Force Microscopy Analysis of Bacterial Cell Wall Peptidoglycan Architecture BT - Bacterial Cell Wall Homeostasis: Methods and Protocols. In: Methods in Molecular Biology. 2016.
40. Liu X, Biboy J, Vollmer W, Blaauwen den T. MreC and MreD balance the interaction between the elongasome proteins PBP2 and RodA. bioRxiv. 2019;
41. den Blaauwen T, Luirink J. Checks and Balances in Bacterial Cell Division. mBio [Internet]. 2019 Feb;10(1). Available from: http://mbio.asm.org/lookup/doi/10.1128/mBio.00149-19
42. Typas A, Banzhaf M, van den Berg van Saparoea B, Verheul J, Biboy J, Nichols RJ, et al. Regulation of Peptidoglycan Synthesis by Outer-Membrane Proteins. Cell. 2010 Dec;143(7):1097–109.
43. Moynihan PJ, Clarke AJ. O-Acetylated peptidoglycan: Controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems. International Journal of Biochemistry and Cell Biology. 2011.
44. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;
45. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;
46. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;
47. Rohs PDA, Buss J, Sim SI, Squyres GR, Srisuknimit V, Smith M, et al. A central role for PBP2 in the activation of peptidoglycan polymerization by the bacterial cell elongation machinery. PLoS Genet. 2018;
48. Boes A, Olatunji S, Breukink E, Terrak M. Regulation of the Peptidoglycan Polymerase Activity of PBP1b by Antagonist Actions of the Core Divisome Proteins FtsBLQ and FtsN. den Blaauwen T, Salama NR, editors. mBio [Internet]. 2019 Jan;10(1). Available from: http://mbio.asm.org/lookup/doi/10.1128/mBio.01912-18
49. Meeske AJ, Riley EP, Robins WP, Uehara T, Mekalanos JJ, Kahne D, et al. SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature. 2016 Sep;537(7622):634–8.
50. Ranjit DK, Jorgenson MA, Young KD. PBP1B glycosyltransferase and transpeptidase activities play different essential roles during the de novo regeneration of rod morphology in Escherichia coli. J Bacteriol. 2017;
51. Kim SY, Gitai Z, Kinkhabwala A, Shapiro L, Moerner WE. Single molecules of the bacterial actin MreB undergo directed treadmilling motion in Caulobacter crescentus. Proc Natl Acad Sci. 2006 Jul;103(29):10929–34.
52. Adams DW, Errington J. Bacterial cell division: Assembly, maintenance and disassembly of the Z ring. Nature Reviews Microbiology. 2009.
53. Potluri LP, Kannan S, Young KD. ZipA is required for FtsZ-dependent preseptal peptidoglycan synthesis prior to invagination during cell division. J Bacteriol. 2012;
54. Wang J, Galgoci A, Kodali S, Herath KB, Jayasuriya H, Dorso K, et al. Discovery of a

Small Molecule that Inhibits Cell Division by Blocking FtsZ, a Novel Therapeutic Target of Antibiotics. J Biol Chem. 2003;
55. Yan Liao SI Jan Löwe, Iain G Duggin. Two FtsZ proteins orchestrate archaeal cell division through distinct functions in ring assembly and constriction. bioRxiv. 2021;
56. Van Teeffelen S, Gitai Z. Rotate into shape: MreB and bacterial morphogenesis. EMBO J. 2011;
57. Bean GJ, Flickinger ST, Westler WM, McCully ME, Sept D, Weibel DB, et al. A22 disrupts the bacterial actin cytoskeleton by directly binding and inducing a low-affinity state in MreB. Biochemistry. 2009;
58. McCausland JW, Yang X, Lyu Z, Söderström B, Xiao J, Liu J. Treadmilling FtsZ polymers drive the directional movement of sPG-synthesis enzymes via Brownian ratchet mechanism. bioRxiv. 2019 Jan;857813.
59. Banzhaf M, van den Berg van Saparoea B, Terrak M, Fraipont C, Egan A, Philippe J, et al. Cooperativity of peptidoglycan synthases active in bacterial cell elongation. Mol Microbiol. 2012 Jul;85(1):179–94.
60. Divakaruni AV, Baida C, White CL, Gober JW. The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes. Mol Microbiol. 2007;
61. Martins A, Contreras-Martel C, Janet-Maitre M, Miyachiro MM, Estrozi LF, Trindade DM, et al. Self-association of MreC as a regulatory signal in bacterial cell wall elongation. Nat Commun. 2021 Dec;12(1):2987.
62. Jean NL, Bougault CM, Lodge A, Derouaux A, Callens G, Egan AJF, et al. Elongated structure of the outer-membrane activator of peptidoglycan synthesis LpoA: Implications for PBP1A stimulation. Structure. 2014;
63. Paradis-Bleau C, Markovski M, Uehara T, Lupoli TJ, Walker S, Kahne DE, et al. Lipoprotein Cofactors Located in the Outer Membrane Activate Bacterial Cell Wall Polymerases. Cell. 2010 Dec;143(7):1110–20.
64. Catherwood AC, Lloyd AJ, Tod JA, Chauhan S, Slade SE, Walkowiak GP, et al. Substrate and Stereochemical Control of Peptidoglycan Cross-Linking by Transpeptidation by Escherichia coli PBP1B. J Am Chem Soc. 2020 Mar;142(11):5034–48.
65. Egan AJF, Errington J, Vollmer W. Regulation of peptidoglycan synthesis and remodelling. Nat Rev Microbiol. 2020 Aug;18(8):446–60.
66. Bendezú FO, Hale CA, Bernhardt TG, De Boer PAJ. RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in E. coli. EMBO J. 2009;
67. White CL, Kitich A, Gober JW. Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. Mol Microbiol. 2010;
68. Banzhaf M, Yau HCL, Verheul J, Lodge A, Kritikos G, Mateus A, et al. The outer membrane lipoprotein NlpI nucleates hydrolases within peptidoglycan multi-enzyme complexes in Escherichia coli. bioRxiv. 2019;
69. Berezuk AM, Glavota S, Roach EJ, Goodyear MC, Krieger JR, Khursigara CM. Outer membrane lipoprotein RlpA is a novel periplasmic interaction partner of the cell division protein FtsK in Escherichia coli. Sci Rep. 2018;
70. Caveney NA, Caballero G, Voedts H, Niciforovic A, Worrall LJ, Vuckovic M, et al. Structural insight into YcbB-mediated beta-lactam resistance in Escherichia coli. Nat Commun. 2019;
71. Laddomada F, Miyachiro MM, Jessop M, Patin D, Job V, Mengin-Lecreulx D, et al. The MurG glycosyltransferase provides an oligomeric scaffold for the cytoplasmic steps of peptidoglycan biosynthesis in the human pathogen Bordetella pertussis. Sci Rep. 2019;
72. Gerding MA, Liu B, Bendezú FO, Hale CA, Bernhardt TG, De Boer PAJ. Self-enhanced accumulation of FtsN at division sites and roles for other proteins with a SPOR domain (DamX, DedD, and RlpA) in Escherichia coli cell constriction. J Bacteriol. 2009;

73. Petiti M, Serrano B, Faure L, Lloubes R, Mignot T, Duché D. Tol Energy-Driven Localization of Pal and Anchoring to the Peptidoglycan Promote Outer-Membrane Constriction. J Mol Biol. 2019;
74. Bisson-Filho AW, Hsu YP, Squyres GR, Kuru E, Wu F, Jukes C, et al. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. Science. 2017;
75. Fenton AK, Gerdes K. Direct interaction of FtsZ and MreB is required for septum synthesis and cell division in Escherichia coli. EMBO J. 2013;
76. Schwechheimer C, Rodriguez DL, Kuehn MJ. NlpI-mediated modulation of outer membrane vesicle production through peptidoglycan dynamics in Escherichia coli. MicrobiologyOpen. 2015;
77. Pazos M, Peters K, Casanova M, Palacios P, VanNieuwenhze M, Breukink E, et al. Z-ring membrane anchors associate with cell wall synthases to initiate bacterial cell division. Nat Commun. 2018;
78. Bendezú FO, De Boer PAJ. Conditional lethality, division defects, membrane involution, and endocytosis in mre and mrd shape mutants of Escherichia coli. J Bacteriol. 2008;
79. Varma A, Young KD. In Escherichia coli, MreB and FtsZ direct the synthesis of lateral cell wall via independent pathways that require PBP 2. J Bacteriol. 2009;
80. Dion MF, Kapoor M, Sun Y, Wilson S, Ryan J, Vigouroux A, et al. Bacillus subtilis cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems. Nat Microbiol. 2019;
81. Shiomi D, Toyoda A, Aizu T, Ejima F, Fujiyama A, Shini T, et al. Mutations in cell elongation genes mreB, mrdA and mrdB suppress the shape defect of RodZ-deficient cells. Mol Microbiol. 2013;
82. Eberhardt C, Kuerschner L, Weiss DS. Probing the Catalytic Activity of a Cell Division-Specific Transpeptidase In Vivo with β-Lactams. J Bacteriol. 2003 Jul;185(13):3726–34.
83. Bernal-Cabas M. The Cpx pathway causes changes in the peptidoglycan structure, turnover, and recycling. MSc Thesis. 2014;
84. Mohammadi T, Karczmarek A, Crouvoisier M, Bouhss A, Mengin-Lecreulx D, Den Blaauwen T. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in Escherichia coli. Mol Microbiol. 2007;
85. Yang DC, Blair KM, Salama NR. Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. Microbiol Mol Biol Rev. 2016;
86. Frirdich E, Biboy J, Adams C, Lee J, Ellermeier J, Gielda LD, et al. Peptidoglycan-modifying enzyme Pgp1 is required for helical cell shape and pathogenicity traits in campylobacter jejuni. PLoS Pathog. 2012;
87. Frirdich E, Vermeulen J, Biboy J, Soares F, Taveirne ME, Johnson JG, et al. Peptidoglycan LD-carboxypeptidase Pgp2 influences Campylobacter jejuni helical cell shape and pathogenic properties and provides the substrate for the DL-carboxypeptidase Pgp1. J Biol Chem. 2014;
88. Hugonnet JE, Mengin-Lecreulx D, Monton A, den Blaauwen T, Carbonnelle E, Veckerlé C, et al. Factors essential for L,D-transpeptidase-mediated peptidoglycan cross-linking and β-lactam resistance in Escherichia coli. eLife. 2016;
89. Braun V, Sieglin U. The Covalent Murein-Lipoprotin Structure of the Escherichia coli Cell Wall: The Attachment Site of the Lipoprotein on the Murein. Eur J Biochem. 1970;
90. Samsudin F, Boags A, Piggot TJ, Khalid S. Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. Biophys J. 2017;
91. A.T. B, F. S, http://orcid.org/0000-0002-3694-5044 Khalid S. AO - Khalid SO. Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the E. coli Cell Wall. Structure. 2019;
92. Sandoz KM, Moore RA, Beare PA, Patel AV, Smith RE, Bern M, et al. β-Barrel proteins

tether the outer membrane in many Gram-negative bacteria. Nat Microbiol. 2021 Jan;6(1):19–26.
93. Sandoz KM, Moore RA, Beare PA, Patel VA, Smith RE, Bern M, et al. β-Barrel proteins tether the outer membrane in many Gram-negative bacteria. Nat Microbiol. 2021 Jan 2;6(1):19–26.
94. Yu YC, Lin CN, Wang SH, Ng SC, Hu WS, Syu WJ. A putative lytic transglycosylase tightly regulated and critical for the EHEC type three secretion. J Biomed Sci. 2010;
95. Bahadur R, Chodisetti PK, Reddy M. Cleavage of Braun's lipoprotein Lpp from the bacterial peptidoglycan by a paralog of l,d-transpeptidases, LdtF. Proc Natl Acad Sci. 2021 May;118(19):e2101989118.
96. 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;
97. Breijyeh Z, Jubeh B, Karaman R. Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It. Molecules. 2020 Jan;25(6):1340.
98. Wilks JC, Slonczewski JL. pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry. J Bacteriol. 2007 Aug;189(15):5601–7.
99. van den Berg B, Prathyusha Bhamidimarri S, Dahyabhai Prajapati J, Kleinekathöfer U, Winterhalter M. Outer-membrane translocation of bulky small molecules by passive diffusion. Proc Natl Acad Sci. 2015 Jun 9;112(23):E2991–9.
100. Rowlett VW, Mallampalli VKPS, Karlstaedt A, Dowhan W, Taegtmeyer H, Margolin W, et al. Impact of Membrane Phospholipid Alterations in Escherichia coli on Cellular Function and Bacterial Stress Adaptation. J Bacteriol. 2017 Jun 13;199(13):e00849-16.
101. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem. 2008;29(11):1859–65.
102. Zielińska A, Savietto A, de Sousa Borges A, Martinez D, Berbon M, Roelofsen JR, et al. Flotillin-mediated membrane fluidity controls peptidoglycan synthesis and MreB movement. Mignot T, Storz G, Hamoen L, editors. eLife. 2020 Jul 14;9:e57179.
103. Goose JE, Sansom MSP. Reduced Lateral Mobility of Lipids and Proteins in Crowded Membranes. PLOS Comput Biol. 2013 Apr 11;9(4):e1003033.
104. Vos AF de, Pater JM, Pangaart PS van den, Kruif MD de, Veer C van 't, Poll T van der. In Vivo Lipopolysaccharide Exposure of Human Blood Leukocytes Induces Cross-Tolerance to Multiple TLR Ligands. J Immunol. 2009 Jul 1;183(1):533–42.
105. Horne JE, Brockwell DJ, Radford SE. Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. J Biol Chem. 2020 Jul 24;295(30):10340–67.
106. Neuman SD, Levine TP, Bashirullah A. A novel superfamily of bridge-like lipid transfer proteins. Trends Cell Biol. 2022 Nov 1;32(11):962–74.
107. Sperandeo P, Martorana AM, Polissi A. The lipopolysaccharide transport (Lpt) machinery: A nonconventional transporter for lipopolysaccharide assembly at the outer membrane of Gram-negative bacteria. J Biol Chem. 2017 Nov 3;292(44):17981–90.
108. Douglass MV, McLean AB, Trent MS. Absence of YhdP, TamB, and YdbH leads to defects in glycerophospholipid transport and cell morphology in Gram-negative bacteria. PLOS Genet. 2022 Feb 28;18(2):e1010096.
109. Plummer AM, Fleming KG. BamA Alone Accelerates Outer Membrane Protein Folding in Vitro through a Catalytic Mechanism. Biochemistry. 2015;
110. Tomasek D, Rawson S, Lee J, Wzorek JS, Harrison SC, Li Z, et al. Structure of a nascent membrane protein as it folds on the Bam complex. Nature. 2020 Jul;583(7816):473–8.
111. Consoli E, Luirink J, den Blaauwen T. The Escherichia coli Outer Membrane β-Barrel Assembly Machinery (BAM) Crosstalks with the Divisome. Int J Mol Sci. 2021 Jan;22(22):12101.

112. In silica binding of loose a.a chains to BamA POTRAs suggests no selective involvement.pdf. 2021 Sep 21 [cited 2022 Nov 1]; Available from: https://osf.io/https://osf.io/rt823
113. Özbaykal G, Wollrab E, Simon F, Vigouroux A, Cordier B, Aristov A, et al. The transpeptidase PBP2 governs initial localization and activity of the major cell-wall synthesis machinery in E. coli. eLife. 2020;
114. Wood C, Burnley T, Patwardhan A, Scheres S, Topf M, Roseman A, et al. Collaborative Computational Project for Electron cryo-Microscopy. Acta Crystallogr D Biol Crystallogr. 2015 Jan 1;71(1):123–6.
115. Liu H, Naismith JH. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 2008 Dec 4;8(1):91.
116. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004 Jun;14(6):1188–90.
117. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. Roux B, editor. eLife. 2014 May 1;3:e02030.
118. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. 2021 Aug 23;61(8):3891–8.
119. Breil C, Abert Vian M, Zemb T, Kunz W, Chemat F. "Bligh and Dyer" and Folch Methods for Solid–Liquid–Liquid Extraction of Lipids from Microorganisms. Comprehension of Solvatation Mechanisms and towards Substitution with Alternative Solvents. Int J Mol Sci. 2017 Apr;18(4):708.
120. Fahy E, Alvarez-Jarreta J, Brasher CJ, Nguyen A, Hawksworth JI, Rodrigues P, et al. LipidFinder on LIPID MAPS: peak filtering, MS searching and statistical analysis for lipidomics. Bioinformatics. 2019 Feb 15;35(4):685–7.
121. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. Anal Chem. 2021 Aug 31;93(34):11692–700.
122. Petrou VI, Herrera CM, Schultz KM, Clarke OB, Vendome J, Tomasek D, et al. Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. Science. 2016 Feb 5;351(6273):608–12.
123. Pollock NL, Rai M, Simon KS, Hesketh SJ, Teo ACK, Parmar M, et al. SMA-PAGE: A new method to examine complexes of membrane proteins using SMALP nano-encapsulation and native gel electrophoresis. Biochim Biophys Acta Biomembr. 2019 Aug 1;1861(8):1437–45.
124. Zubarev RA, Makarov A. Orbitrap Mass Spectrometry. Anal Chem. 2013 Jun 4;85(11):5288–96.
125. Boes A, Olatunji S, Breukink E, Terrak M. Regulation of the peptidoglycan polymerase activity of PBP1b by antagonist actions of the core divisome proteins FtsBLQ and FtsN. mBio. 2019;
126. Schwartz B, Markwalder JA, Seitz SP, Wang Y, Stein RL. A Kinetic Characterization of the Glycosyltransferase Activity of Eschericia coli PBP1b and Development of a Continuous Fluorescence Assay. Biochemistry. 2002 Oct 1;41(41):12552–61.
127. Li Y, Boes A, Cui Y, Zhao S, Liao Q, Gong H, et al. Identification of the potential active site of the septal peptidoglycan polymerase FtsW. PLOS Genet. 2022 Jan 5;18(1):e1009993.
128. Welsh MA, Schaefer K, Taguchi A, Kahne D, Walker S. Direction of Chain Growth and Substrate Preferences of Shape, Elongation, Division, and Sporulation-Family Peptidoglycan Glycosyltransferases. J Am Chem Soc. 2019 Aug 21;141(33):12994–7.
129. Takasuga A, Adachi H, Ishino F, Matsuhashi M, Ohta T, Matsuzawa H. Identification of

the penicillin-binding active site of penicillin-binding protein 2 of Escherichia coli. J Biochem (Tokyo). 1988 Nov;104(5):822–6.
130. van Straaten KE, Barends TRM, Dijkstra BW, Thunnissen AMWH. Structure of Escherichia coli Lytic transglycosylase MltA with bound chitohexaose: implications for peptidoglycan binding and cleavage. J Biol Chem. 2007 Jul 20;282(29):21197–205.
131. Artola-Recolons C, Lee M, Bernardo-García N, Blázquez B, Hesek D, Bartual SG, et al. Structure and cell wall cleavage by modular lytic transglycosylase MltC of Escherichia coli. ACS Chem Biol. 2014 Sep 19;9(9):2058–66.
132. van Asselt EJ, Dijkstra AJ, Kalk KH, Takacs B, Keck W, Dijkstra BW. Crystal structure of Escherichia coli lytic transglycosylase Slt35 reveals a lysozyme-like catalytic domain with an EF-hand. Structure. 1999 Oct 15;7(10):1167–80.
133. Dik DA, Batuecas MT, Lee M, Mahasenan KV, Marous DR, Lastochkin E, et al. A Structural Dissection of the Active Site of the Lytic Transglycosylase MltE from Escherichia coli. Biochemistry. 2018 Oct 23;57(42):6090–8.
134. Scheurwater E, Reid CW, Clarke AJ. Lytic transglycosylases: Bacterial space-making autolysins. International Journal of Biochemistry and Cell Biology. 2008.
135. Bohrhunter JL, Rohs PDA, Torres G, Yunck R, Bernhardt TG. MltG activity antagonizes cell wall synthesis by both types of peptidoglycan polymerases in Escherichia coli. Mol Microbiol. 2021 Jun;115(6):1170–80.
136. Yunck R, Cho H, Bernhardt TG. Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria. Mol Microbiol. 2016 Feb;99(4):700–18.
137. Crépin S, Ottosen EN, Peters K, Smith SN, Himpsl SD, Vollmer W, et al. The lytic transglycosylase MltB connects membrane homeostasis and in vivo fitness of Acinetobacter baumannii. Mol Microbiol. 2018;
138. Höltje JV, Mirelman D, Sharon N, Schwarz U. Novel type of murein transglycosylase in Escherichia coli. J Bacteriol. 1975 Dec;124(3):1067–76.
139. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023 Jan 6;51(D1):D418–27.
140. van Asselt EJ, Thunnissen AM, Dijkstra BW. High resolution crystal structures of the Escherichia coli lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment. J Mol Biol. 1999 Aug 27;291(4):877–98.
141. Legaree BA, Clarke AJ. Interaction of Penicillin-Binding Protein 2 with Soluble Lytic Transglycosylase B1 in Pseudomonas aeruginosa. J Bacteriol. 2008 Oct;190(20):6922–6.
142. Bateman A, Bycroft M. The structure of a LysM domain from E. coli membrane-bound lytic murein transglycosylase D (MltD)11Edited by P. E. Wight. J Mol Biol. 2000 Jun 16;299(4):1113–9.
143. Alcorlo M, Dik DA, De Benedetti S, Mahasenan KV, Lee M, Domínguez-Gil T, et al. Structural basis of denuded glycan recognition by SPOR domains in bacterial cell division. Nat Commun. 2019 Dec 5;10(1):5567.
144. Domínguez-Gil T, Lee M, Acebrón-Avalos I, Mahasenan KV, Hesek D, Dik DA, et al. Activation by Allostery in Cell-Wall Remodeling by a Modular Membrane-Bound Lytic Transglycosylase from Pseudomonas aeruginosa. Struct Lond Engl 1993. 2016 Oct 4;24(10):1729–41.
145. Dik DA, Madukoma CS, Tomoshige S, Kim C, Lastochkin E, Boggess B, et al. Slt, MltD and MltG of Pseudomonas aeruginosa as Targets of Bulgecin A in Potentiation of β-Lactam Antibiotic. ACS Chem Biol. 2019 Feb 15;14(2):296–303.
146. Templin MF, Edwards DH, Höltje JV. A murein hydrolase is the specific target of bulgecin in Escherichia coli. J Biol Chem. 1992 Oct 5;267(28):20039–43.
147. Mierendorf RC, Morris BB, Hammer B, Novy RE. Expression and Purification of Recombinant Proteins Using the pET System. Methods Mol Med. 1998;13:257–92.
148. Schägger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel

electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem. 1987 Nov 1;166(2):368–79.
149. MAEDA H. A New Lysozyme Assay Based on Fluorescence Polarization or Fluorescence Intensity Utilizing a Fluorescent Peptidoglycan Substrate 1. J Biochem (Tokyo). 1980 Oct 1;88(4):1185–91.
150. Kühner D, Stahl M, Demircioglu DD, Bertsche U. From cells to muropeptide structures in 24 h: Peptidoglycan mapping by UPLC-MS. Sci Rep. 2014 Dec 16;4(1):7494.
151. Bank RPD. RCSB PDB: Homepage [Internet]. [cited 2023 Jan 9]. Available from: https://www.rcsb.org/
152. Wong MWK, Braidy N, Pickford R, Sachdev PS, Poljak A. Comparison of Single Phase and Biphasic Extraction Protocols for Lipidomic Studies Using Human Plasma. Front Neurol [Internet]. 2019 [cited 2023 Jan 3];10. Available from: https://www.frontiersin.org/articles/10.3389/fneur.2019.00879
153. Kuru E, Velocity Hughes H, Brown PJ, Hall E, Tekkam S, Cava F, et al. In situ Probing of Newly Synthesized Peptidoglycan in Live Bacteria with Fluorescent D-Amino Acids. Angew Chem Int Ed Engl. 2012 Dec 7;51(50):12519–23.
154. Chung CT, Miller RH. Preparation and storage of competent Escherichia coli cells. Methods Enzymol. 1993;218:621–7.
155. Nilsson I, Lee SY, Sawyer WS, Baxter Rath CM, Lapointe G, Six DA. Metabolic phospholipid labeling of intact bacteria enables a fluorescence assay that detects compromised outer membranes[S]. J Lipid Res. 2020 Jun 1;61(6):870–83.
156. Chen W, Zhang YM, Davies C. Penicillin-Binding Protein 3 Is Essential for Growth of Pseudomonas aeruginosa. Antimicrob Agents Chemother. 2016 Dec 27;61(1):e01651-16.
157. Chan KG, Priya K, Chang CY, Rahman AYA, Tee KK, Yin WF. Transcriptome analysis of Pseudomonas aeruginosa PAO1 grown at both body and elevated temperatures. PeerJ. 2016 Jul 19;4:e2223.158. Nilsson I, Lee SY, Sawyer WS, Baxter Rath CM, Lapointe G, Six DA. Metabolic phospholipid labeling of intact bacteria enables a fluorescence assay that detects compromised outer membranes[S]. J Lipid Res. 2020 Jun 1;61(6):870–83.
159. Chen W, Zhang YM, Davies C. Penicillin-Binding Protein 3 Is Essential for Growth of Pseudomonas aeruginosa. Antimicrob Agents Chemother. 2016 Dec 27;61(1):e01651-16.
160. Chan KG, Priya K, Chang CY, Rahman AYA, Tee KK, Yin WF. Transcriptome analysis of Pseudomonas aeruginosa PAO1 grown at both body and elevated temperatures. PeerJ. 2016 Jul 19;4:e2223.
161. Jakubec D, Skoda P, Krivak R, Novotny M, Hoksza D. PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. Nucleic Acids Res. 2022 Jul 5;50(W1):W593–7.
162. Cacace E, Kritikos G, Typas A. Chemical genetics in drug discovery. Curr Opin Syst Biol. 2017 Aug 1;4:35–42.
163. Kritikos G, Banzhaf M, Herrera-Dominguez L, Koumoutsi A, Wartel M, Zietek M, et al. A tool named Iris for versatile high-throughput phenotyping in microorganisms. Nat Microbiol. 2017 Feb 17;2:17014.
164. Doherty H, Kritikos G, Galardini M, Banzhaf M, Moradigaravand D. ChemGAPP; A Package for Chemical Genomics Analysis and Phenotypic Profilin [Internet]. 2023 [cited 2023 Feb 28]. Available from: https://europepmc.org/article/PPR/PPR593530
165. Servais, C., Vassen, V., Verhaeghe, A. et al. Lipopolysaccharide biosynthesis and traffic in the envelope of the pathogen Brucella abortus. Nat Commun 14, 911 (2023). https://doi.org/10.1038/s41467-023-36442-y
166. Aktas M, Wessel M, Hacker S, Klüsener S, Gleichenhagen J, Narberhaus F. Phosphatidylcholine biosynthesis and its significance in bacteria interacting with eukaryotic cells. Eur J Cell Biol. 2010 Dec;89(12):888-94. doi: 10.1016/j.ejcb.2010.06.013. Epub 2010 Jul 24. PMID: 20656373.

167. Wargo MJ, Gross MJ, Rajamani S, Allard JL, Lundblad LK, Allen GB, Vasil ML, Leclair LW, Hogan DA. Hemolytic phospholipase C inhibition protects lung function during Pseudomonas aeruginosa infection. Am J Respir Crit Care Med. 2011 Aug 1;184(3):345-54. doi: 10.1164/rccm.201103-0374OC. Epub 2011 May 11. PMID: 21562128; PMCID: PMC3175536.

Thus ended the thesis of Chris Graham, PhD student of Professor David Roper, so brilliant its scientific spirit wherein, imbued by the power of science, it is said to have caught fire and breathe flame into the hearts of its reader.

Structural basis of peptidoglycan synthesis by *E. coli* RodA-PBP2 complex

Rie Nygaard[1], Chris L.B. Graham[2], Meagan Belcher Dufrisne[6], Jonathan D. Colburn[2,3], Joseph Pepe[1], Molly A. Hydorn[5], Silvia Corradi[1,8], Chelsea M. Brown[2,3], Khuram U. Ashraf[1], Owen N. Vickery[2,3], Nicholas S. Briggs[2], John J. Deering[2], Brian Kloss[7], Bruno Botta[8], Oliver B. Clarke[1,4], Linda Columbus[6,*], Jonathan Dworkin[5,*], Phillip J. Stansfeld[2,3,*], David I. Roper[2,*] and Filippo Mancia[1,*]

[1]Department of Physiology and Cellular Biophysics, Columbia University Irving Medical Center, New York, NY 10032, USA.

[2]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK

[3]Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK

[4]Department of Anesthesiology, Columbia University Irving Medical Center, New York, NY, 10032, USA

[5]Department of Microbiology and Immunology, Columbia University Irving Medical Center, New York, NY, 10032, USA

[6]Department of Chemistry and Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA, 22904, USA

[7]New York Consortium on Membrane Protein Structure, New York Structural Biology Center, 89 Convent Avenue, New York, NY, 10027, USA

[8]Faculty of Pharmacy and Medicine, Sapienza University of Rome, Rome, Italy

[*]Correspondence to be addressed to: fm123@columbia.edu (F.M.), David.roper@warwick.ac.uk (D.I.R.), phillip.stansfeld@warwick.ac.uk (P.J.S.), jonathan.dworkin@columbia.edu (J.D.), columbus@virginia.edu (L.C.)

KEYWORDS

Cryo-EM, MD simulations, site-directed spin labeling, DEER EPR, genetics, antimicrobial resistance, beta-lactam, shape, elongation, division and sporulation (SEDS), penicillin binding protein (PBP)

SUMMARY

Peptidoglycan (PG) is an essential structural component of the bacterial cell wall that is synthetized during cell division and elongation. PG forms an extracellular polymer crucial for cellular viability, the synthesis of which is the target of many antibiotics. PG assembly requires a glycosyltransferase (GT) to generate a glycan polymer using a Lipid II substrate, which is then crosslinked to the existing PG via a transpeptidase (TP) reaction. A Shape, Elongation, Division and Sporulation (SEDS) GT enzyme and a Class B Penicillin Binding Protein (PBP) form the core of the multi-protein complex required for PG assembly. Here we used single particle cryo-electron microscopy to determine the structure of a cell elongation-specific *E. coli* RodA-PBP2 complex. We combine this information with biochemical, genetic, spectroscopic, and computational analyses to identify the Lipid II binding sites and propose a mechanism for Lipid II polymerization. Our data suggest a hypothesis for the movement of the glycan strand from the Lipid II polymerization site of RodA towards the TP site of PBP2, functionally linking these two central enzymatic activities required for cell wall peptidoglycan biosynthesis.

INTRODUCTION

Cell shape in bacteria is determined and maintained by the extracellular polymer peptidoglycan (PG), a mesh-like sacculus surrounding the cytoplasmic membrane composed of polymerized glycan chains cross-linked by short peptides[1]. PG synthesis is rate limiting for bacterial growth and its disruption results in cell lysis or cessation of growth as exploited by many natural product and semi-synthetic antibiotics[2-5]. These include β-lactams, the most clinically successful antibiotics to date[6, 7]. The cytoplasmic proteins that synthesize the PG precursor, Lipid II – an undecaprenyl (C_{55}) pyrophosphate (Und-PP)-linked disaccharide of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc)-pentapeptide – and the extracellular proteins responsible for the subsequent polymerization of PG, have been individually characterized, biochemically and structurally[8].

In the periplasm, PG biosynthesis begins with a Lipid II-specific glycosyltransferase (GT) which forms a glycan strand polymer by linking the disaccharides of two Lipid II molecules, one termed the donor and the other the acceptor, and thereby releasing Und-PP from the donor site (Fig. 1a). After the two initial Lipid II molecules have been linked together, the resulting tetra-disaccharide attached to Und-PP, termed Lipid IV, becomes the donor for another Lipid II acceptor, in turn linking its tetra-saccharide to the Lipid II di-saccharide to yield Lipid VI. This cycle repeats in a processive manner creating progressively longer polysaccharide chains attached to Und-PP (the roman numeral denotes the number of monosaccharide groups in the polysaccharide chain). Once the growing glycan polymer reaches a sufficient length, it is attached to the existing PG sacculus via peptide crosslinks between the pentapeptide of the glycan strand and a peptide stem on the existing PG sacculus by a transpeptidase (TP) to yield crosslinked PG (Fig. 1a). In *E. coli*, the GT RodA from the Shape, Elongation, Division, and Sporulation (SEDS) family, and PBP2, the

monofunctional TP class B Penicillin Binding Protein, mediate these respective enzymatic tasks[9]. RodA is an integral membrane protein consisting of ten transmembrane (TM) helices[10], while PBP2 has a single TM helix and an extracellular domain with a classical class B PBP fold containing the TP active site[11, 12]. Together, they comprise the core of the elongasome[13], the complex responsible for the determination of bacterial rod shape. Despite recent advances in our understanding of this molecular machine[14], not least as derived from the crystal structure of a *Thermus thermophilus* RodA-PBP2 complex[15], fundamental mechanistic questions remain unresolved. These include characterization of molecular determinants and conformational states required for i) Lipid II binding, ii) GT polymerization of glycan strands, and iii) subsequent translocation of the glycan polymer to the TP active site.

Here, we used single-particle cryo-electron microscopy (cryo-EM) to determine the structure of the *E. coli* RodA-PBP2 complex, expressed as a functional fusion and reconstituted in lipid-filled nanodiscs, to 3.0 Å resolution. We used an integrated approach – combining structural information with biochemical and genetic assays, molecular dynamics (MD) simulations and site-directed spin labeling (SDSL) double electron-electron resonance (DEER) experiments – to investigate the molecular determinants of substrate binding and catalysis, and conformational changes required for this processive machinery to function. Our studies suggest a mechanism that would facilitate migration of a growing glycan polymer towards the TP site of PBP2, enabling TP-dependent crosslinking to form the PG layer.

RESULTS

Structure determination of RodA-PBP2

Although most bacteria contain separate open reading frames (ORFs) encoding the GT (SEDS) and TP (PBP) activities, examples exist of a single ORF encoding a SEDS-PBP fusion[16]. We previously demonstrated that a synthetic construct consisting of a SEDS protein fused to its cognate PBP functionally complements gene deletions of both enzymes[12]. We reasoned that a SEDS-PBP fusion would be biochemically more tractable than the individual components, and that its structure in a lipid environment would greatly facilitate a mechanistic understanding of the overall complex. We screened 189 SEDS orthologs for expression and stability in detergents to identify those amenable for structural studies[17]. From this initial screen, SEDS proteins from five different species with the highest expression levels and stability were selected to design a set of fusions with one or more species-matched PBPs, resulting in eight unique constructs (Supplementary Table 1). These were again evaluated for expression and stability in detergent after metal-affinity chromatography (using a genetically encoded poly-histidine tag), and mono-dispersity as assessed by size-exclusion chromatography (SEC). Based on these criteria, a fusion construct between *E. coli* RodA and PBP2 (Supplementary Fig. 1a) had the most promising profile for structure determination.

Next, we confirmed that this *E. coli* RodA-PBP2 fusion has GT activity – which is not inhibited by moenomycin, in contrast to class A bifunctional PBPs – as observed by polymerization of Lipid II modified with a fluorescent dansyl group (dansyl lysine Lipid II; Supplementary Fig. 1b, c, e). The TP domain binds (and is covalently modified by) a fluorescent penicillin mimic bocillin[18] Supplementary Fig. 1d, f), suggesting that the TP domain is intact. We purified the RodA-PBP2 fusion protein, and as controls RodA alone (terminating at residue 373 of RodA) and a fusion of RodA with only the single TM helix of PBP2 (terminating at residue 47 of PBP2) in detergent (Supplementary Fig. 1g). These proteins were assayed for their Lipid II polymerization activity

(Supplementary Fig. 1h). RodA in isolation has residual GT activity but is significantly stimulated by the presence of the transmembrane helix of PBP2 in both the full length fusion and the truncated version. These results are consistent with what was previously shown for the *T. thermophilus* proteins[15].

The RodA-PBP2 fusion was purified to homogeneity in detergent by metal-affinity chromatography followed by SEC, and then reconstituted into lipid-filled nanodiscs (Supplementary Fig. 1i-l) for subsequent vitrification and cryo-EM analysis. This resulted in two maps locally refined around the TM and periplasmic regions, both at 3.0 Å resolution (Fig. 1b, Supplementary Table 2 and Supplementary Fig. 2). In the TM map, we could reliably build the RodA structure, assigning the sequence from residues 9 to 93 and 109 to 364 as well as the adjacent PBP2 single TM helix from residue 10 to 40 (Supplementary Fig. 3). The structure shows ten RodA TM helices, with intracellular N- and C-termini. TM helices 1-6 and TM helices 8-10 form a tight helical bundle, with TM helix 7 extending away from it, stabilized by three periplasmic juxtamembrane helices (PH1, PH2 and PH3) (Fig. 1d, e). Additionally, the PBP2 single TM helix, is ordered and packs against RodA TM helices 8 and 9 (Fig. 1c).

The PBP2 soluble domain is less well resolved, but using Namdinator[19] and its previously published crystal structure as the input model[20], we were able to build residues 10-343, 401-430, 457-540 and 569-612. We were unable to observe interpretable density for residues 344-400 and 431-456 at the tip of PBP2, proximal to the TP active site (Supplementary Fig. 4a). The binding site for MreC[21] – a scaffolding protein that binds to PBP2 and is thought to impact its activity[21, 22] – also known as the head domain, has the lowest local resolution in the map (Supplementary Fig. 4b). Nevertheless, the resolution was sufficient to fit the helices in the head domain to the density

map, despite the fact that we observe substantial structural changes compared to the previously published X-ray crystal structure of PBP2 from *E. coli*[20] (Supplementary Fig. 4a).

Comparing the *T. thermophilus* model[15] with our model, we observe that the overall structure of the TM region is very similar between the two, despite having a sequence identity of only 39% (Supplementary Fig. 4c). This includes the positioning of TM helix 7, which in both extends similarly away from the helical TM core. In contrast, in the periplasmic domain of PBP, we observe quite substantial structural differences. There is both a relative turn and a tilt between the two, and a closing of the head and anchor domain in our structure as opposed to an opening in the *T. thermophilus* model (Supplementary Fig. 4c).

Putative substrate binding cavities

Synthesis of the glycan polymer requires an acceptor and a donor binding site in RodA, both initially accommodating Lipid II. The two initial Lipid II substrates are linked via a GT reaction, coupling the MurNAc sugar of the Lipid II in the donor site to the GlcNAc of the Lipid II in the acceptor site, with Und-PP as the product in the donor site (Fig. 1a). Since RodA is a processive enzyme, Lipid IV (polymerized Lipid II after the first reaction) now becomes the donor so that the MurNAc directly bound to Und-PP in Lipid IV can be transferred to an incoming Lipid II acceptor. For this to occur, Lipid IV needs to transition from the acceptor to the donor sites. This movement of the growing glycan chain from the acceptor to the donor site is processive, thus allowing the cycle to repeat producing progressively longer chains (i.e., Lipid VI, Lipid VIII, etc.).

Analysis of the RodA portion of the structure for putative substrate-binding sites reveals two major cavities (termed cavity A and B) (Fig. 2a). Cavity A is located between TM helices 6, 7 and 9, and

is framed on one side by PH1 and on the other by TM helices 5 and 6 and the periplasmic loop (PL3) connecting the two. The rim of cavity A is lined with conserved residues (Fig. 2a) and overall is polar in nature (Supplementary Fig. 4d). It extends into the membrane, where it becomes exposed to the lipid bilayer, and is lined with hydrophobic residues (Fig. 2a and Supplementary Fig. 4d). Cavity B is located on the opposite side of RodA relative to cavity A, between TM helices 2, 3, 4 and 10 and is also exposed to the membrane (Fig. 2a). Residues lining cavity B are less conserved, but as for cavity A we observe several positively charged residues on its periplasmic side (Fig. 2a and Supplementary Fig. 4d). The region connecting the two cavities has some of the highest degree of conservation in the entire structure (Fig. 2a, b). Given their orientation in respect to the TP site of PBP2, we suggest that cavities A and B are the donor and acceptor sites for the two substrates, respectively[22].

To probe this hypothesis, we performed 50 repeats of 10 μs unbiased coarse-grained (CG) MD simulations to identify the location and interactions the two Lipid II substrates have with the surface of RodA. Analysis of these simulations identified two preferred binding sites for Lipid II that directly map to cavities A and B (Fig. 2c). Of the two binding sites, cavity B appears to have higher particle density for Lipid II than cavity A over the total simulation time. This is consistent with cavity B being the acceptor site, continuously recruiting Lipid II from the membrane. The density analysis also revealed further sites of interaction, but when evaluated with PyLipID[23], cavity B, followed by cavity A, had the highest occupancies by Lipid II during the simulations (Supplementary Fig. 5a-d). Specifically, over the CG simulations, Arg48 and Arg109 (cavity B) and Arg210 (cavity A) displayed the highest occupancy of interactions with Lipid II (Supplementary Fig. 5b). In all cases, the predominant protein-Lipid II interactions were with the peptidoglycan-pyrophosphate headgroup, while the lipid tail engaged with both protein and

surrounding lipid membrane. We combined these data with ligand docking to refine our model for how Lipid II binds to both cavities A and B (Supplementary Fig. 5e-f).

In agreement with our simulations, we observed an elongated density in the cryo-EM map within cavity A (Fig. 2d). The shape of the density is consistent with Und-PP, an obligate product of each GT reaction (Fig. 1a). To further elaborate on this observation, we ran three repeats of 1 μs atomistic MD simulations of RodA-PBP2 inserted into a phospholipid membrane with Und-PP docked into cavity A (Supplementary Fig. 5g). During these experiments, we observed that Und-PP stayed tightly associated within this cavity and that its average occupancy overlapped with the density present in the cryo-EM data (Fig. 2d).

Biochemical characterization of residues in and around cavity A and B

Both cavities have positively charged residues on their periplasmic side (Arg210 in cavity A; Arg48 and Arg109 in cavity B) (Fig. 2b and Supplementary Fig. 6). Our simulations suggest that these residues interact with the pyrophosphate group of Lipid II. An equivalent of Arg210 on PH1 appears to be part of a common structural element for GT-C glycosyltransferases that use Und-PP as a carrier, as for example the O-antigen ligase WaaL[24, 25]. To test whether Arg210 is essential for function, we mutated it to alanine and found that GT activity *in vitro* was significantly reduced (Fig. 2e and Supplementary Fig. 7a). We also examined the phenotype of a mutation in the corresponding arginine in *B. subtilis* using an *in vivo* sporulation assay[12]. Here, spore heat resistance is dependent on the function of the RodA homolog SpoVE (Supplementary Fig. 8) that synthesizes spore PG. We observed that mutating SpoVE Arg212 (corresponding to *E. coli* RodA Arg210) to alanine had a severe effect on sporulation (Supplementary Table 3). This result is consistent with the requirement of this arginine for GT activity (Fig. 2e). Arg48 and Arg109 both

point towards cavity B (Fig. 2b). GT activity is severely reduced *in vitro* and *in vivo* when Arg48 is mutated to alanine (Fig. 2e, Supplementary Fig. 7a and Table 3) whereas it is only moderately affected *in vitro* for an Arg109Ala mutant. These results are consistent with an important role of Arg48 and to a lesser extent Arg109, in coordinating the pyrophosphate of Lipid II within or entering cavity B.

The twenty amino acid periplasmic loop connecting TM helices 3 and 4 (PL2) is positioned adjacent to cavity B and could potentially reach from one cavity to the other and interact with the substrates in both cavities, and/or play a role in their transition from the acceptor to the donor sites. This loop is intrinsically flexible, as observed in our MD simulations (Supplementary Fig. 5b), and we could only partially assign the sequence to the density in this region of the cryo-EM map. PL2 has several highly conserved residues, including Trp102 and Gln111, that are invariant in all species analyzed (Supplementary Fig. 6). PL2 also contains two positively charged residues, Lys97 and Arg101, which while only moderately conserved, are appropriately positioned to engage with Lipid II. To probe the function of PL2, we mutated these two charged residues to alanine. Only Arg101, which is the more conserved of the two, showed significant reduction in activity compared to wild type (Fig. 2e and Supplementary Fig. 7a). Next, we mutated Trp102 or Gln111 to alanine and showed that both mutants had reduced GT activity (Fig. 2e and Supplementary Fig. 7a). However, the GT activity of Trp102 appeared to be maintained with a mutation to phenylalanine. Mutation of the equivalent Trp104 in *B. subtilis* SpoVE to alanine or phenylalanine severely affected the sporulation phenotype (Supplementary Table 3), consistent with the functional importance of this residue.

Finally, we probed the potential roles of other highly conserved residues located in or between the two cavities (Glu114, Lys117, Asp159 and Ser344) (Fig. 2b). GT activity in our *in vitro* GT assay

was not affected when Ser344 was mutated to alanine whereas we observed a reduction in activity when Asp159 was mutated to valine, consistent with the severe effect of a mutation of *B. subtilis* SpoVE Asp163 (the equivalent of *E. coli* Asp159) to valine in our *in vivo* sporulation assay (Supplementary Table 3). Finally, while mutations of Glu114 to alanine and Lys117 to asparagine had only a modest impact on GT activity *in vitro* (Fig. 2e and Supplementary Fig. 7a), identical mutations of the corresponding residues in *B. subtilis* SpoVE (Glu116Ala, Lys119Asn) severely affected sporulation efficiency (Supplementary Table 3). The differences observed between the *in vivo* and *in vitro* assays suggest that even modestly reduced *in vitro* activity may not be sufficient for maintaining *in vivo* function, and highlight the importance of characterizing mutant phenotypes *in vivo*.

Conformational flexibility within RodA

In order for the glycan strand to form and extend towards the TP active site of PBP2, the growing Und-PP-linked glycans (Lipid II, IV, VI, etc.) must transition from the acceptor site (cavity B) to the donor site (cavity A) site at each round of catalysis[26]. For this to occur, a conformational change in RodA must take place to open a passageway between TM helices 1-2 and 8-10 on one side, and the helical bundle of TM helices 3-7 on the other (Fig. 3a). The cryo-EM density map reveals that the bundle composed of TM helices 3-7 has an overall lower resolution than the rest of the protein, suggesting some degree of flexibility within this region (Fig. 3b).

We used SDSL DEER spectroscopy to further probe this hypothesis. The resulting DEER-derived probability distribution of distances between two nitroxide spin labels (MTSL; R1) introduced at specific sites provides information about the number and population of conformational states, as well as a distance restraint for each one. We replaced the two native cysteines in RodA by

mutagenesis to glycine and alanine (Cys82Gly and Cys133Ala) and confirmed that the cysteine-free fusion RodA-PBP2 complex was expressed and functional (Supplementary Fig. 9a, b). We then introduced a spin label pair – Gly44R1 on the periplasmic end of TM helix 2 and Asp90R1 on the periplasmic end of TM helix 3 – in the RodA fusion with the cysteine-free background (Fig. 3c and Supplementary Fig. 9a-c). From the resulting DEER-derived distance distribution, three populations were observed with dominant distances at 26 Å, 35 Å, and 45 Å (Fig. 3d and Supplementary Fig. 9d, e). Predicted distances calculated by adding sterically allowed MTSL rotamers to the cryo-EM structure *in silico* yielded a distance distribution between 34 and 48 Å (Fig. 3d), aligning best with the longest experimental distance population. Side chain rotamers, which can contribute up to +/- 8 Å (and accounted for in the *in silico* modeling) may be able to partially explain the middle-distance population (centered at 35 Å). However, a backbone conformational change from the cryo-EM structure is needed to sample the majority of the distances between 20 and 35 Å, implying that multiple conformations of the helical bundle exist. These populations have shorter distances between Gly44R1 and Asp90R1 than observed in the cryo-EM map so further investigation is needed to determine the structure and physiological relevance of these states.

The RodA active site and mechanism of catalysis

Within the highly conserved region between the two cavities and centrally located in periplasmic loop 4 (PL4) is Asp262, previously identified as a catalytic residue[14, 15, 27] (Fig. 4a). To further define the active site, we performed systematic mutagenesis of residues surrounding Asp262, and tested their effect on function via our biochemical assay. Consistent with a previous report[9], mutation of Asp262 to alanine renders RodA-PBP2 enzymatically inactive (Fig. 4b and

Supplementary Fig. 7a). Mutations Glu258Ala, His260Ala or Thr261Ser did not affect activity despite the high degree of conservation and spatial-proximity of these residues to Asp262, whilst mutation of Pro257 – also proximal to Asp262 – to alanine resulted in complete loss of GT activity (Fig. 4b, Supplementary Fig. 6 and 7a). A similar severe phenotype is observed when the equivalent of Asp262 is mutated in our *in vivo B. subtilis* assay (Asp263Ala) together with a more modest reduction in activity for the equivalent of Glu258 (Glu259Ala) (Supplementary Table 3).

Based on our modelled and simulated coordinates of Lipid II bound to both cavities (Supplementary Fig. 5a), we propose a mechanism, in which Asp262 plays a central role in enabling the GT reaction to occur (Fig. 4c), analogous to His338 in WaaL[24]. In our scheme, Asp262 abstracts a proton from the 4' hydroxyl of Lipid II in cavity B. This abstraction of the proton would then allow the 4' oxygen to perform a nucleophilic attack on the 1' carbon of MurNAc of Lipid II in cavity A. This breaks the bond between the carbon and the oxygen of the phosphate resulting in a Und-PP product in cavity A and Lipid IV (or Lipid VI, Lipid VII etc.) in cavity B. Our MD simulations suggest that Arg210 in cavity A and Arg48 in cavity B, as well as the arginine residues in PL2 (97-111) (Arg101 and Arg109) engage with the pyrophosphate head groups of Lipid II to coordinate of the substrate within the active site. We tested whether the reaction is metal dependent, but observed no change in activity by adding the divalent-ion chelator EDTA during the assay, indicating that the GT reaction catalyzed by the *E. coli* RodA-PBP2 fusion is metal-independent (Supplementary Fig. 7b). This contrasts with the GT51 family of enzymes, which includes class A PBPs[28-30].

We probed the feasibility of the proposed mechanism by performing semi-empirical Density Functional Tight Binding (DFTB)[31] calculations on a cluster model of the putative RodA active site (Supplementary Fig. 5a). From these calculations, we observed that Asp262-activated

formation of Lipid IV from two Lipid II substrates bound in cavities A and B is plausible given the geometric constraints of the RodA structure as sampled by our atomistic simulations. By using partially-converged nudged elastic band calculations, we present the reaction visually within Supplementary Movie 1. While this does not validate the energetic feasibility of the proposed mechanism, it does illustrate that the coordination of the bound substrates within the active site is geometrically suitable to enable the reaction.

Formation of a glycan chain and movement of polymer to TP site

The processive mechanism of RodA involves a the shuttling of substrates from the acceptor to the donor site to progressively extend the growing glycan strand, until the leading pentapeptide stem reaches the TP site of PBP2[32, 33]. Our structure contains a groove that extends from the extracellular surface of the RodA GT site to the PBP2 TP active site (Fig. 6). Ligand docking and modelling different lengths of the glycan strand reveal that Lipid XX (i.e., ten disaccharides) bound to cavity A is of sufficient length to reach the TP active site and for the peptide stem to connect to the already existing PG. We have modelled RodA-PBP2 bound sequentially to Lipid II, Lipid IV, up to Lipid XXII and MD simulations show that the polysaccharides are stably coordinated within this groove (Supplementary Fig. 10a). The pentapeptide stem appears more mobile when not bound to the TP site, and the positions of the polyprenyl tails in and around both cavities A and B anchors the nascent PG to the membrane through interactions both within the protein and with the phospholipid membrane (Supplementary Fig. 10a). The periplasmic portion of the complex has a much higher RMSF than the TM domain, however, the secondary structure of RodA-PBP2 is stable during the simulations (Supplementary Fig. 5b and 10b).

Dynamics between RodA and PBP2 in the complex

While analyzing the cryo-EM data, we observed considerable structural heterogeneity between RodA and PBP2 (Fig. 5a and Supplementary Fig. 9n). To further investigate this, we performed a 3D variability analysis with a mask around PBP2, which has the lowest local resolution and seems to be the most mobile of the two proteins (Fig. 5a and Supplementary Fig. 9n). After separating the particles into 6 clusters, we observe a vertical tilt of ~10° of PBP2 with respect to the bilayer, as well as a rotation around PBP2 (Fig. 5a and Supplementary Fig. 9n). To complement this observation, we ran MD simulations of RodA-PBP2. Consistent with the cryo-EM data, the tip of PBP2, harboring the TP active site, appears to be highly dynamic in these simulations (Fig. 5c). For the simulations with glycan chains included, especially true for those with longer ones, we observe a large conformational change in the extracellular part of PBP2 that swings upwards with the tip of PBP2 pointing away from the membrane (Supplementary Fig. 10c). We do not observe an extended conformation like this in our cryo-EM data, but as discussed above, we observe heterogeneity in the sample and discarded numerous particles that may adopt more transient conformations of PBP2.

To further study this, we investigated the dynamics between PBP2 and RodA with DEER using an engineered cysteine pair between residue Gly44 at the periplasmic side of TM helix 2 in RodA and residue Gln99 in the head domain of PBP2 (residue 476 in the fusion) (Fig. 5b). PBP2 Gln99 is located in the proposed MreC binding site[14, 21, 22], the region where we observe the most variability in the cryo-EM data (Supplementary Fig. 9n). This pair (Gly44Cys, Gln99Cys) was labelled with cysteine-reactive nitroxide spin-label (R1; MTSL), confirmed to be functional (Supplementary Fig. 9n), and DEER data were collected. The resulting DEER data for Gly44R1 and Gln99R1 show a broad distribution of distances ranging from 30 to 65 Å (Fig. 5d and Supplementary Fig. 9f-h). Residue Gly44 in RodA is in a relatively static region of the structure

(Supplementary Fig. 5b), so most motions contributing to the DEER distances likely arise from the dynamics of the MreC-binding head domain of PBP2 and/or potentially from the dynamic nature of the entire soluble domain of PBP2 with respect to RodA. Sterically allowed MTSL rotamers attached to residues of the cryo-EM structure *in silico* (Fig. 5b) result in a calculated distance distribution ranging from 18 to 40 Å but do not sample the longer distances of the broad DEER distribution (Fig. 5d). Furthermore, when MTSL rotamers were computationally attached to a structure from MD (50 ns frame from Supplementary Fig. 10c) that has an extended conformation of the head domain of PBP2, the resulting distance distribution was between 46 and 70 Å (Fig. 5d). In comparison, a DEER pair within the PBP2 domain (Lys185R1 and Ser330R1) yields a single more narrow DEER distance distribution (Supplemental Figure 9i-m). These data support the notion that the flexibility observed between the RodA and PBP2 domains is likely due to movement of the head domain with respect to RodA. Collectively, these data suggest that the DEER distance distribution cannot be explained solely by side-chain conformers or small backbone differences from the cryo-EM structure and must represent substantial backbone or domain conformational changes as observed in the MD simulations.

DISCUSSION

The GT RodA and the TP PBP2 constitute the core of the *E. coli* elongasome, the multiprotein complex that catalyzes formation of the essential PG layer. In combination, our results provide data-supported hypotheses for the architecture, catalysis, and structural rearrangements required for peptidoglycan synthesis. The cryo-EM structure identified two major cavities (A and B) on the periplasmic side that coarse grained MD analysis supports as two Lipid II binding sites, with cavity A and B as the donor and acceptor site, respectively. In our cryo-EM map, within the putative

donor site there is a lipid-like density, which we tentatively assigned to Und-PP, a product of the GT reaction. MD simulations show that Und-PP can stably occupy both cavities and that its average occupancy overlaps well with the observed experimental density.

By combining *in vitro* and *in vivo* functional assays that measure GT activity on purified protein and sporulation efficiency in *B. subtilis*, respectively, we have begun to characterize the role of some of the most conserved and/or charged residues in cavity A and B. The structural homology between RodA and other GT-C type glycosyltransferases that utilize Und-PP-coupled ligands[24], suggests that the positively-charged residues could participate in coordinating the pyrophosphate to either facilitate its recruitment into the binding sites or position the substrate for catalysis to occur. Some, such as Arg210, seem to be essential for function and others less so, reflecting perhaps the fact that substrate affinity and specificity are determined by multiple sites of interaction with the protein. Of interest is the conservation and functional significance of a tryptophan at position 102 in PL2 in RodA, which may play a similar role to Trp383 in the bacterial cellulose synthase BscA[34]. In BscA, this tryptophan residue was proposed to act as part of a "finger helix" interacting with the cellulose substrate sugars, enabling a path for the growing glycan strand polymer out of the active site after its formation, which is characteristic of a processive GT enzyme[35, 36]. Trp102 in *E. coli* RodA could perform an analogous role in the processive Lipid II polymerization mechanism. In between the two cavities and bridging them, we also find the active site residue Asp262 and several conserved residues delineating the active site. We probed the functional relevance of these residues both *in vitro* and *in vivo*, and also showed that the GT reaction is metal-independent. By combining these results with semi-empirical DFT calculations on the putative RodA active site, we propose a mechanism for how RodA catalyzes the reaction between the two Und-PP-linked glycans.

To enable processivity, the growing Und-PP-linked glycan chain must transition at each catalytic cycle from the acceptor to the donor cavity. Based on the relative flexibility of the structure as observed in the cryo-EM maps, we hypothesize a movement to create a passageway between TM helices 3-7 on one side and 1-2, 8-10 on the other to form a conduit for the lipid tail of Und-PP. We investigated this hypothesis by DEER, inserting probes on cysteine mutants, designed based on the structure and introduced on a cysteine-less background. These experiments yielded results consistent with flexibility of the TM 3-7 sub-domain.

Our structure suggests how the growing glycan polymer can extend from the RodA GT site to the TP active site within PBP2. By combining modelling of Lipid II, Lipid IV and so forth up to Lipid XX with MD simulations, we determined that that the growing glycan is stable within the groove leading to the TP site, that the polyprenyl tail can be positioned in either cavity as this occurs, and that once Lipid XX has been synthesized, the TP reaction can take place. Cryo-EM, MD, and DEER determine that PBP2 is dynamic and flexible, which could allow for accommodation of the growing glycan strand and facilitate its attachment to the pre-existing PG.

RodA has structural similarities with a number of distinct GTs[24], including a ligand binding cavity created by a TM helix protruding away from the main helical bundle and a short amphipathic helix which lies parallel to the membrane, containing a highly conserved arginine residue, which provides coordination to the Und-PP product of the reaction[24]. Proteins which retain this functional motif for Und-PP binding include WaaL O-antigen ligase[24], and other members of the GT-C family of Und-PP dependent transferase and polymerases for which there until recently was poor structural definition[25]. Sequence conservation strongly indicates that this motif is retained across species. We propose that the mechanism described here for RodA is conserved across other SEDS proteins, including *B. subtilis* SpoVE that serves as *in vivo* system for the present study.

In summary, using an integrated approach centered around the cryo-EM structure of the *E. coli* RodA-PBP2 complex in the close to native environment of a nanodisc, combined with biochemical assays, genetics analysis, MD simulations and DEER experiments, we have generated a model for how this processive machinery functions to synthesize PG. This work will facilitate design structure-based inhibitors for the essential GT activity of RodA and other SEDS Lipid II polymerases for which none are known.

ACKNOWLEDGMENTS

This work was supported by NIGMS grant R35GM132120 (To F.M.), R35GM141953 (J.D.), R35GM131829 (L.C), and F32GM136076 (M.B.D). DIR was funded by a Schaefer Research Scholars Program Award to Columbia University, Biological Sciences Research Council (BBSRC) Grant BB/N003241/1 and Medical Research Council (MRC) Grant MR/N002679/1. Research in PJS's lab is funded by Wellcome (208361/Z/17/Z), the MRC (MR/S009213/1) and BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1). JJD's studentship was sponsored by the MRC (MR/N014294/1) and studentships for NSB and CLBG were sponsored by the BBSRC (BB/M01116X/1). CLBG also received funded from NERC grant NE/T014717/1 and the Antibiotic Research UK Charity. CMB's studentship is sponsored by the MRC (MR/N014294/1). This project made use of time on ARCHER and JADE granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (http://hecbiosim.ac.uk), supported by EPSRC (grant no. EP/R029407/1). PJS acknowledges Sulis at HPC Midlands+, which was funded by the EPSRC on grant EP/T022108/1, and the University of Warwick Scientific Computing Research Technology Platform for computational access.

AUTHOR CONTRIBUTIONS

SEDS-PBP fusion design and cloning was undertaken by M.B.D with help from J.D. and B.K. R.N optimized expression of the *E. coli* RodA-PBP2 construct with help from K.A, S.C. and M.B.D. R.N purified protein for cryo-EM sample. R.N. collected and analysed the cryo-EM data. R.N. built the model with help from C.L.B.G and O.B.C. Gene editing for *E. coli* functional assays was performed by C.L.B.G. and J.P., C.L.B.G. and N.B performed protein expression and purification for *E. coli* functional assays. Functional *in vitro* Lipid II polymerization assays in *E.*

coli were carried out by C.L.B.G. Gene editing for *B. subtilis* was performed by M.A.H and J.D. *In vivo B. subtilis* sporulation assays were conducted by M.A.H with supervision from J.D. All MD simulations were performed by P.J.S and J.D.C, with some scripts provided by O.V, parameters for Lipid II dynamics were designed by C.M.B. Autodock Lipid II binding experiments were performed by C.M.B. and C.L.B.G. The Lipid II substrate was synthesised by C.L.B.G and J.J.D.. Co-evolution analysis was conducted by P.J.S and C.L.B.G. Gene editing for DEER experiments was done by M.B.D and C.L.B.G., and DEER data were collected and analysed by M.B.D. under supervision of L.C. R.N, C.L.B.G, M.B.D, J.D.C, P.J.S, L.C, J.D., F.M and D.I.R designed experiments and wrote the paper. Oversight for the entire project was provided by F.M.

DECLARATION OF INTEREST

The authors declare no competing interest.

DATA AND MATERIALS AVAILABILITY

All raw movie frames, micrographs, the particle stack and relevant metadata files will be deposited into EMPIAR. The electron density map will be deposited into EMDB. The model will be deposited into PDB. All data are available in the manuscript or the supplementary materials.

FIGURES

Figure 1

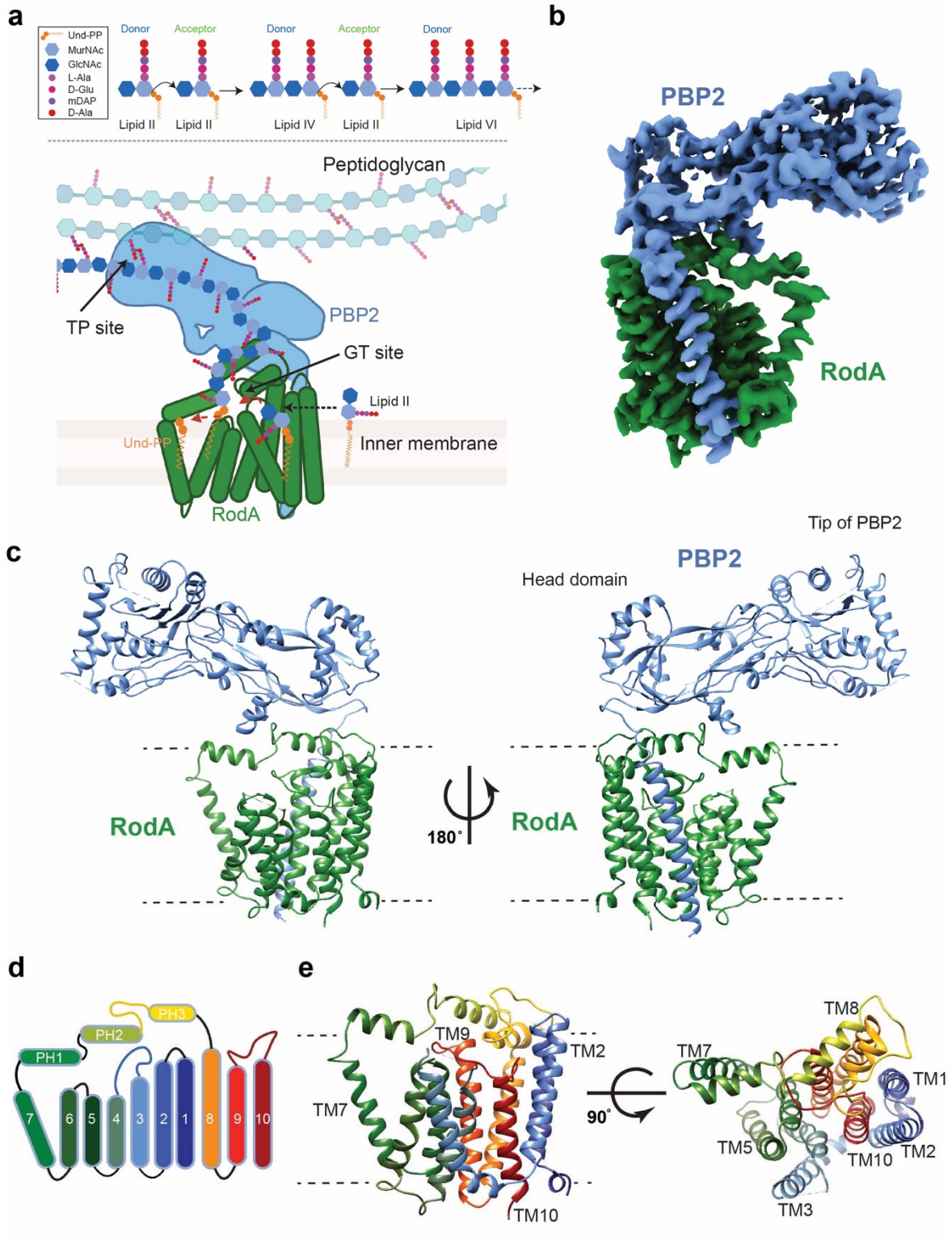


Figure 1 | Mechanism and overall structure of RodA-PBP2 complex. a) Above, schematic representation of the reaction catalyzed by RodA (green) and PBP2 (blue). Below, chemical representation of peptidoglycan synthesis with Lipid II building blocks and chemical components of Lipid II are shown in a box on the top left. **b)** Cryo-EM density map of the RodA-PBP2 complex. Density corresponding to RodA and PBP2 is shown in green and blue respectively. **c**) Structure of the RodA-PBP2 complex shown as ribbon with RodA in green and PBP2 in blue. Approximate membrane boundaries are represented as dotted lines. **d**) Schematic diagram showing the topology of RodA colored in rainbow from C (blue) to N terminus (red), consisting of ten TM helices and a well-ordered periplasmic region between TM helix 7 and 8. **e)** Structure of RodA shown rotated 90°, with the ten TM helices colored as in **e**. Approximate membrane boundaries are again shown as dotted lines.

Figure 2

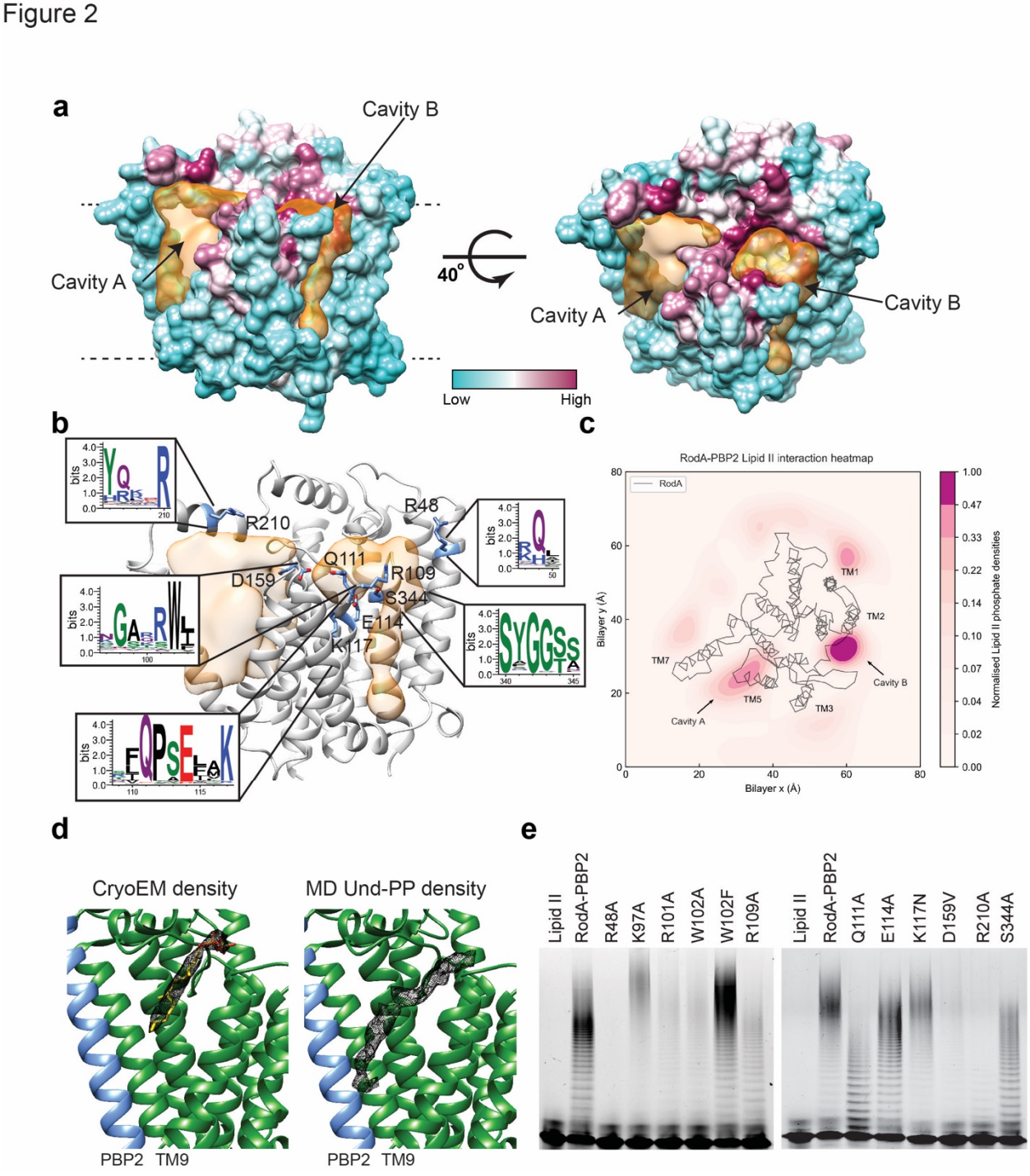
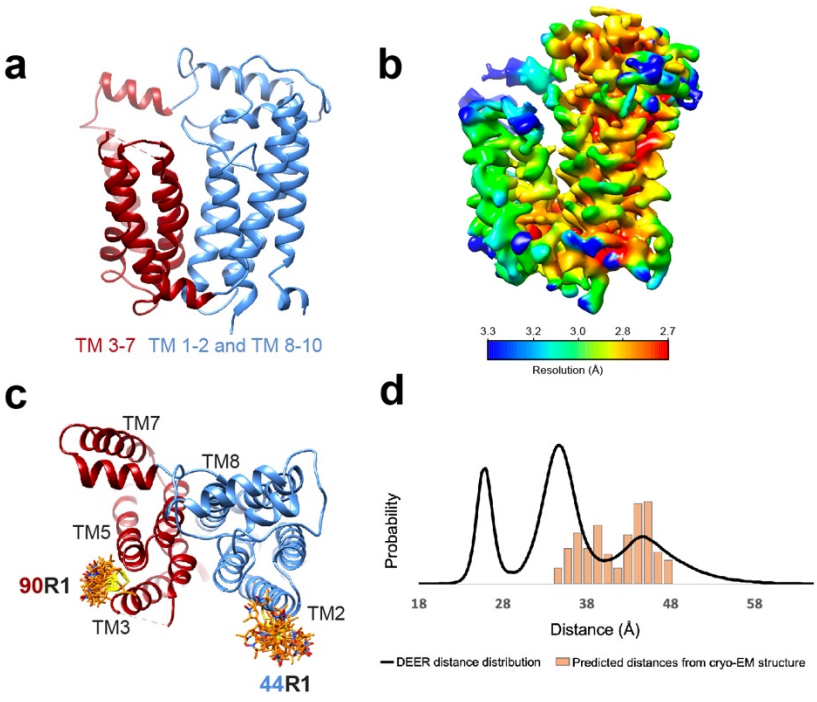


Figure 2 | Lipid II binding cavities. a) Cavity analysis of the RodA-PBP2 structure with RodA-PBP2 shown as surface colored by conservation on a green (no conservation) to purple (absolute conservation) scale. The cavities are shown as a semi-transparent surface in orange. Volumes were

calculated using the Voss Volume Voxelator (3V) server[37] using probes with 10 and 2 Å radii, corresponding to the outer and inner probe, respectively. **b)** Structure of the transmembrane region of RodA-PBP2 shown as a ribbon with residues of interest shown as sticks, and the cavities shown as in **a**. Cut-outs from Weblogo plots[38], as in Supplementary Fig. 6, shown for regions of interest. **c)** Density plot of Lipid II from 50 repeats of 10 μs in unbiased CG MD simulations of the RodA-PBP2 complex. The plot shows two preferred binding sites of Lipid II overlapping with cavity A and cavity B. **d)** Comparison of (left) lipid-like density observed in cavity A in the cryo-EM map (only the region of the map corresponding to the lipid like density is shown) with (right) average density of Und-PP from unbiased atomistic MD simulation of RodA-PBP2 with Und-PP initially docked into cavity A. **e)** Functional analysis of RodA-PBP2 GT activity using *in vitro* Lipid II polymerization assay where the effect of different residues was studied by introducing point mutations and testing for activity.

Figure 3

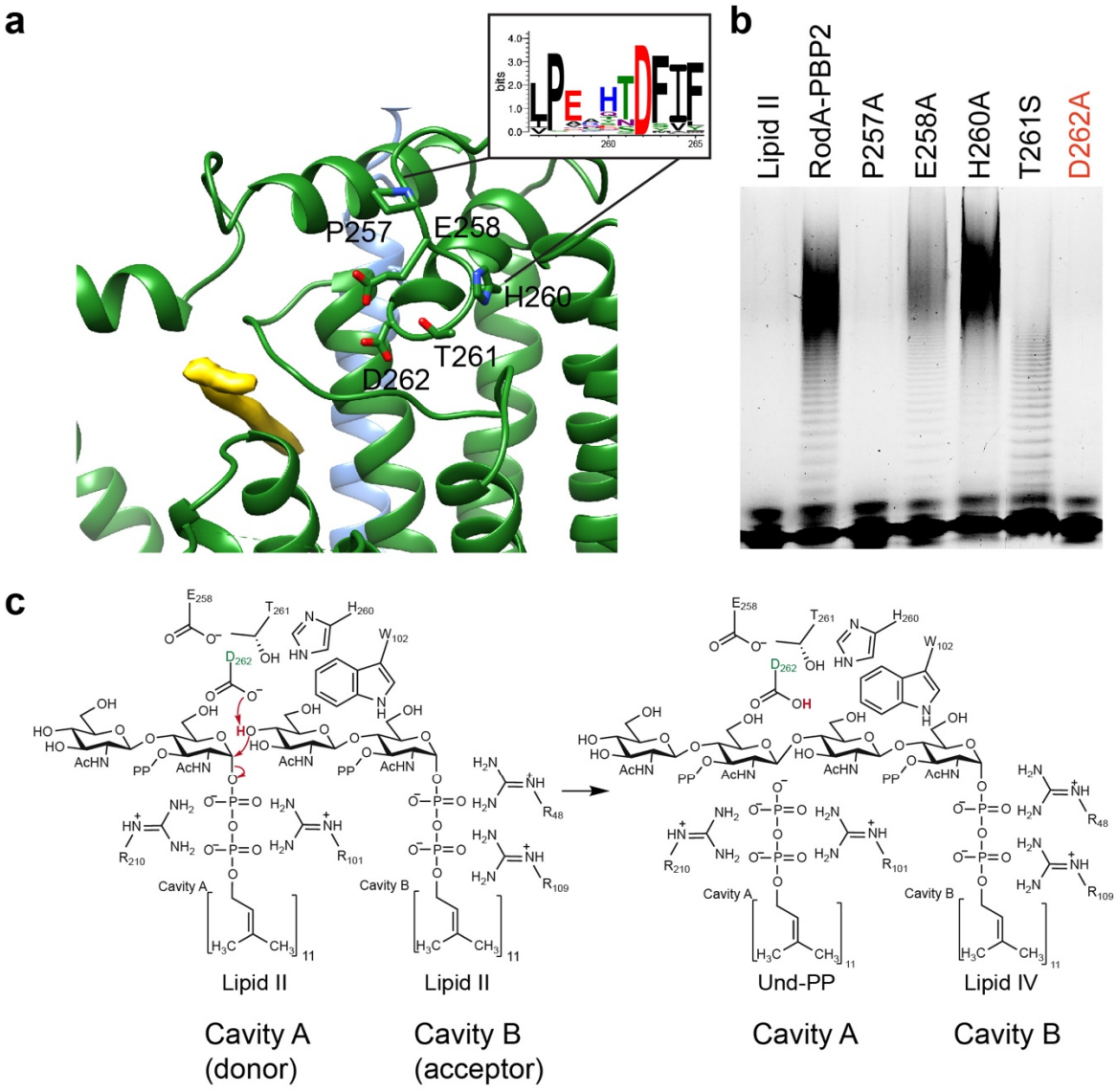


498

499 **Figure 3 | Dynamics within the RodA. a)** RodA shown as ribbon colored by domain with TM
500 helices 3-7 red and TM helices 1-2 and TM helices 8-10 blue. **b**) Local resolution of the
501 transmembrane region of the RodA-PBP2 complex from the final local resolution refinement with
502 a mask around the transmembrane part of the complex. It is apparent that the domain consisting of
503 TM helices 3-7 overall has lower local resolution. **c**) Sterically allowed rotamers of nitroxide spin
504 labels (R1) are illustrated for Gly44R1/Asp90R1 on the cryo-EM structure (orange sidechains). **d**)
505 Resulting DEER distance distributions (solid black line) are compared to predicted distance
506 distributions (bars) based on the sterically allowed rotamers shown in the structures (colored as
507 sidechains).

508

Figure 4



509

510 **Figure 4 | Mechanism of Lipid II transglycosylation. a)** RodA active site with RodA shown as

511 green ribbon and PBP2 shown as blue ribbon, residues of interest represented as sticks. Cut-out

from Weblogo plot, as in Supplementary Fig. 6, shown for residues of interest, residues are represented as sticks. Cryo-EM density in cavity A proposed to correspond to Und-PP displayed as a yellow surface. **b)** Functional analysis of RodA-PBP2 GT activity using *in vitro* Lipid II polymerization assay, monitoring the effect of different mutations on selected residues. **c)** Schematic representation of the mechanistic model for the GT reaction between two Lipid II molecules by RodA viewed from the periplasmic side of the membrane. The active site Asp262 is highlighted in green.

Figure 5

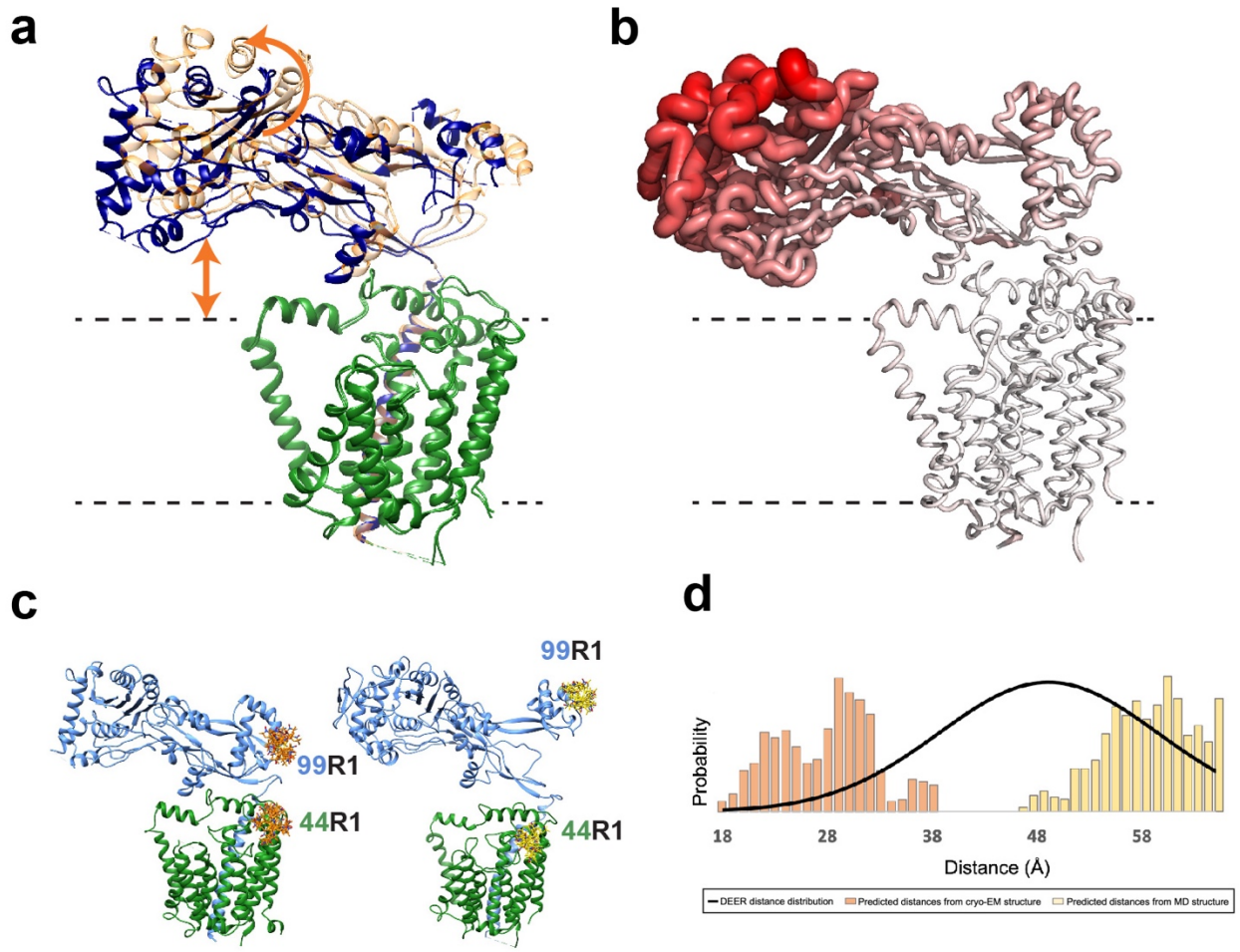


Figure 5 | Dynamics within the RodA-PBP2 complex. a) The two extreme conformations of PBP2 from the cryo-EM 3D variability analysis. **b**) RodA-PBP2 shown as ribbon colored by RMSF over 3 MD simulations of 1 μs aligned to the RodA backbone, with the ribbon increasing in thickness with greater RMSF. **c**) Sterically allowed rotamers of nitroxide spin labels (R1) are illustrated for Gly44R1/Gln99R1 on the cryo-EM structure (left; orange sidechains) and a structure from an MD simulation (right; yellow sidechains). **d**) Resulting DEER distance distributions (solid black line) are compared to predicted distance distributions (bars) based on the sterically allowed rotamers shown in the structures (colored as sidechains).

Figure 6

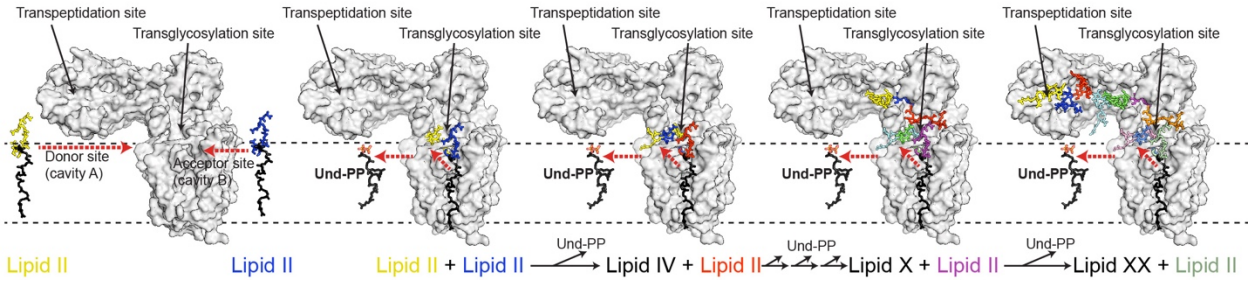
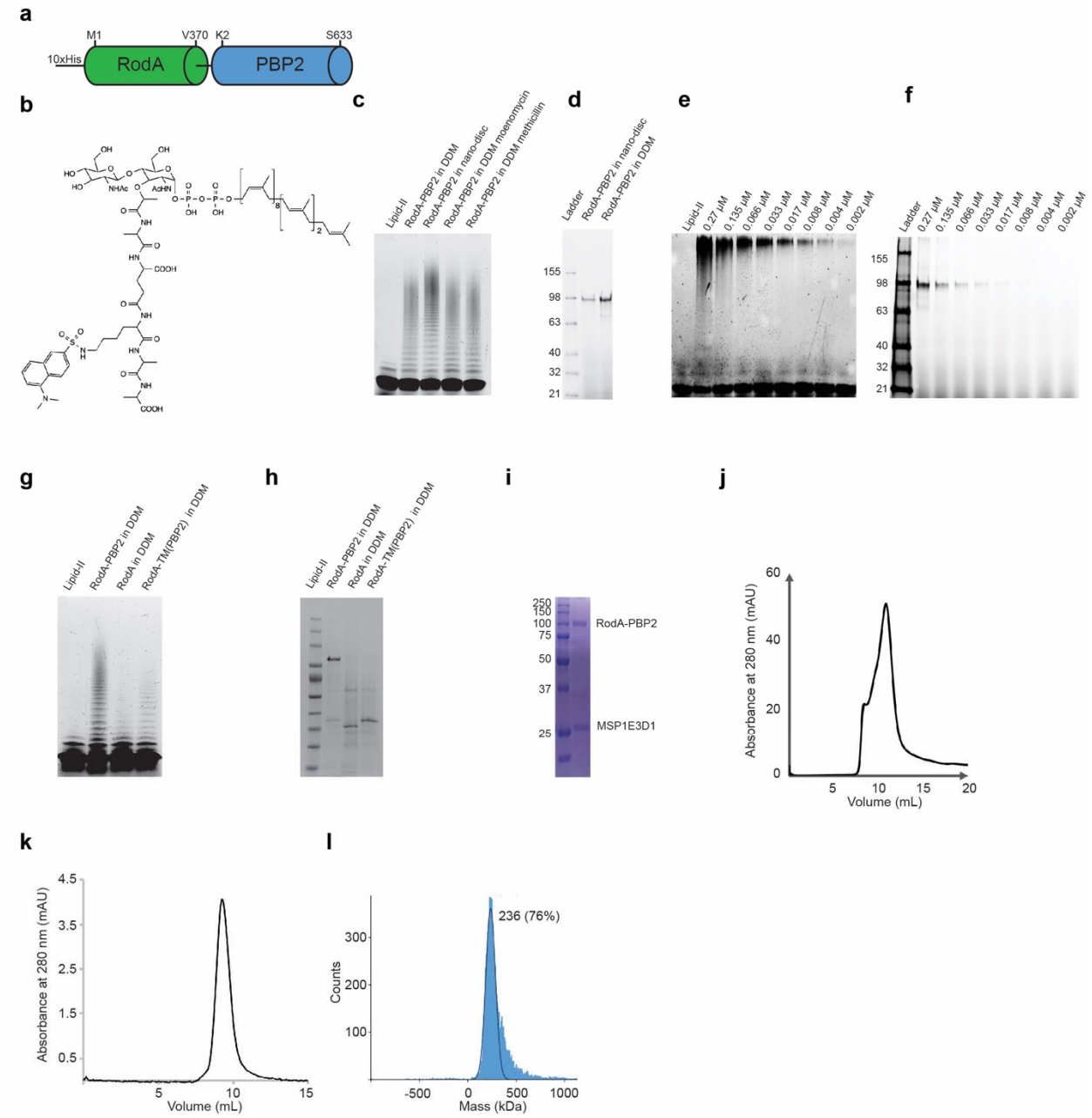


Figure 6 | Elongation of the glycan strand. Modelling of the different steps that occur during elongation of the glycan strand. Starting from the, left RodA-PBP2 is shown in complex with two Lipid II molecules. Once these two molecules are connected through the GT activity of RodA, Und-PP is released from cavity A either between TM helix 7 and the core of RodA, or under the JM helices into the membrane. We have modelled RodA-PBP2 bound sequentially to Lipid II, Lipid IV, up to Lipid XXII, and here we show different complexes during the elongation until the glycan chain reaches the active site of PBP2. RodA-PBP2 is shown as surface representation in grey. The previously suggested opening and closing of the head and anchor domain of PBP2 upon binding of the regulator protein MreC is not modelled here[14, 21, 22, 39]. The Und-PP lipid is represented in black sticks and the disaccharides with the pentapeptide stem attached as sticks in different colors.

Supplementary Figure 1

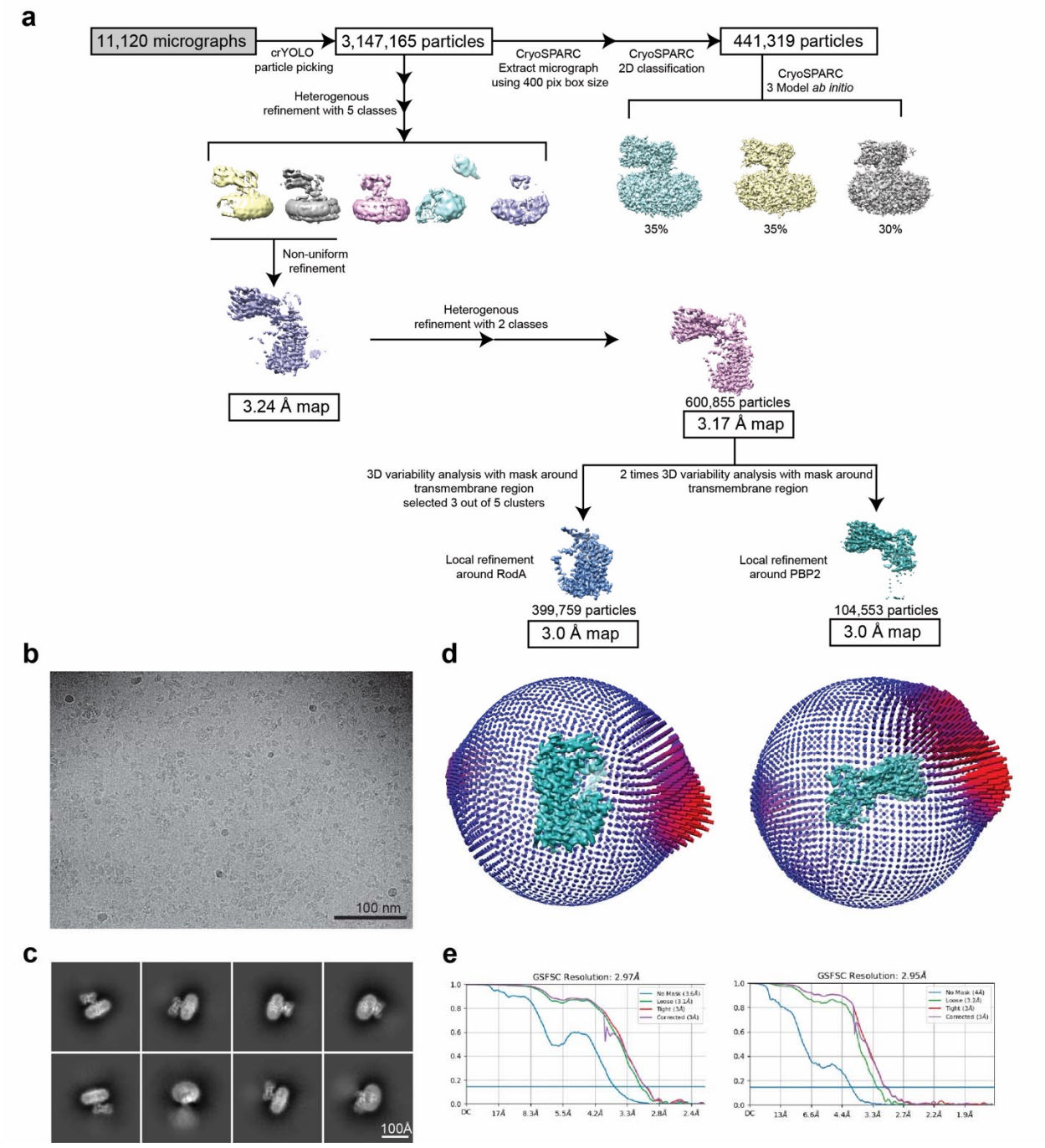


Supplementary Figure 1 | RodA-PBP2 purification and functional characterization.

a) Construct design for the RodA-PBP2 fusion with a His tag genetically fused to the N-terminus of RodA. **b**) Chemical structure of dansyl lysine Lipid-II. The fluorescent dansyl-label is attached at position 3 of the peptide stem. **c**) Electrophoresis analysis of Lipid II polymerization products of RodA-PBP2 in detergent and nanodisc. In the last two lanes shown, we added either

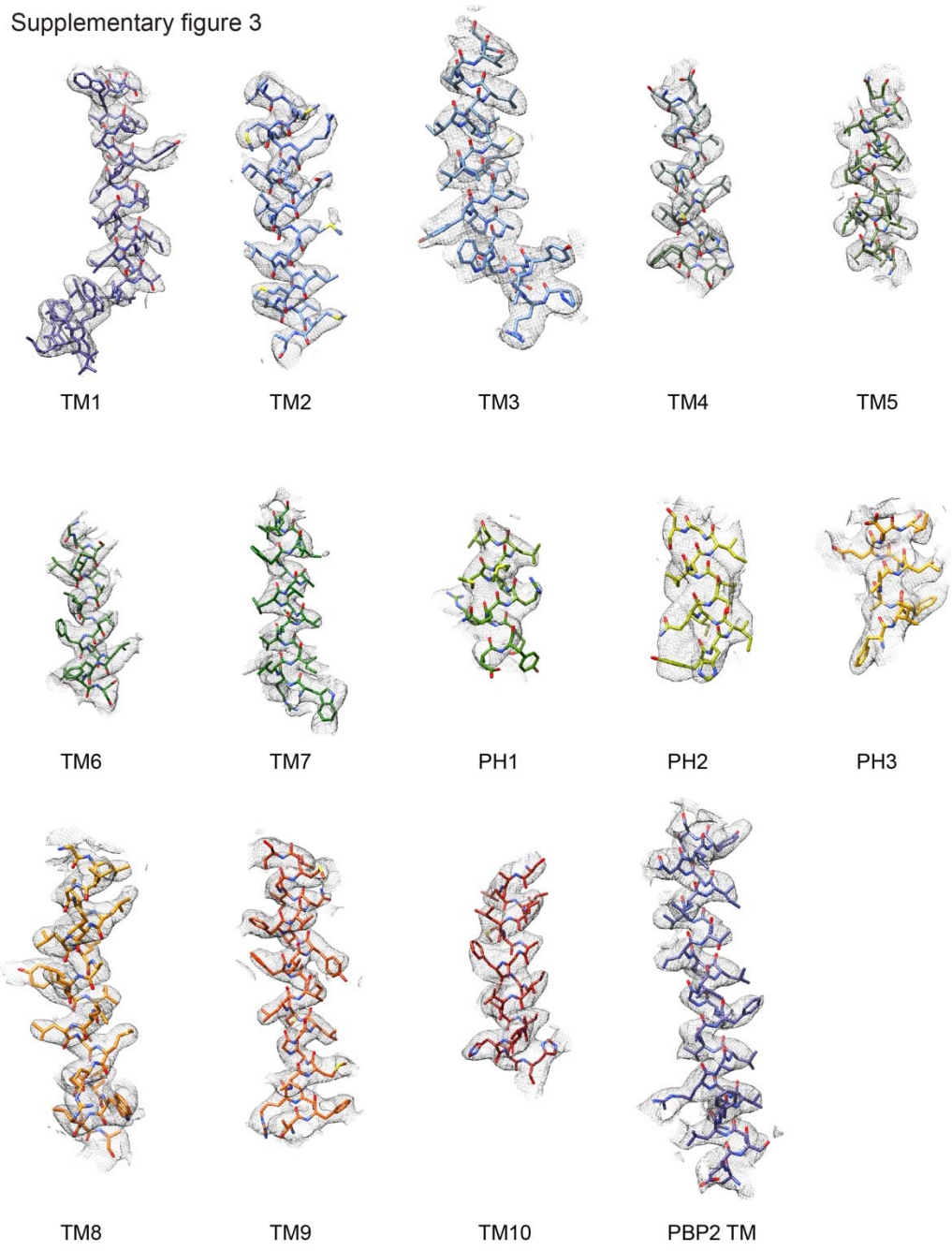
549 moenomycin or methicillin at a concentration of 5.4 μM. **d**) Visualization of RodA-PBP2 protein
550 used for polymerization assays. Bocillin staining confirms integrity of PBP2 transpeptidase active
551 site and reflects protein folding. **e**) Electrophoresis analysis of Lipid II polymerization products of
552 RodA-PBP2 with varying amounts of RodA-PBP added. **f**) Visualization of RodA-PBP2 protein
553 used for polymerization assays using bocillin staining of SDS-PAGE gel. **h**) Electrophoresis
554 analysis of Lipid II polymerization for full length RodA-PBP2 fusion, RodA alone (terminating at
555 residue 373 of RodA) and RodA-TM(PBP2) (terminating at residue 47 of PBP2) in detergent. **g**)
556 SDS-PAGE gel of purified RodA-PBP2 and truncated constructs. **i**) SDS-PAGE gel of the RodA-
557 PBP2 complex after final SEC. **j**) SEC elution profile of RodA-PBP2 in nanodisc at the final
558 purification step. **k**) SEC elution profile of fraction 11 from the final SEC purification step, used
559 for freezing on grids, re-run on analytical SEC. **l**) Mass photometry profile of the sample used for
560 freezing on grids showing that 76% of the particles fit under the peak at 236 KDa corresponding
561 to one RodA-PBP2 fusion complex in nanodisc.
562
563

Supplementary Figure 2



Supplementary Figure 2 | cryo-EM analysis of the RodA-PBP2 complex in nanodisc. a) Flow chart of cryo-EM image acquisition and data processing of the RodA-PBP2 complex. **b**) Representative micrograph. **c**) Representative 2D class averages from cryoSPARC 2D classification. **d**) Euler angle distribution of all particles used in the final map reconstruction. Final

map shown in green. Each orientation is represented by a cylinder, with each cylinder's height and color (from blue to red) proportional to the number of particles for that specific direction.

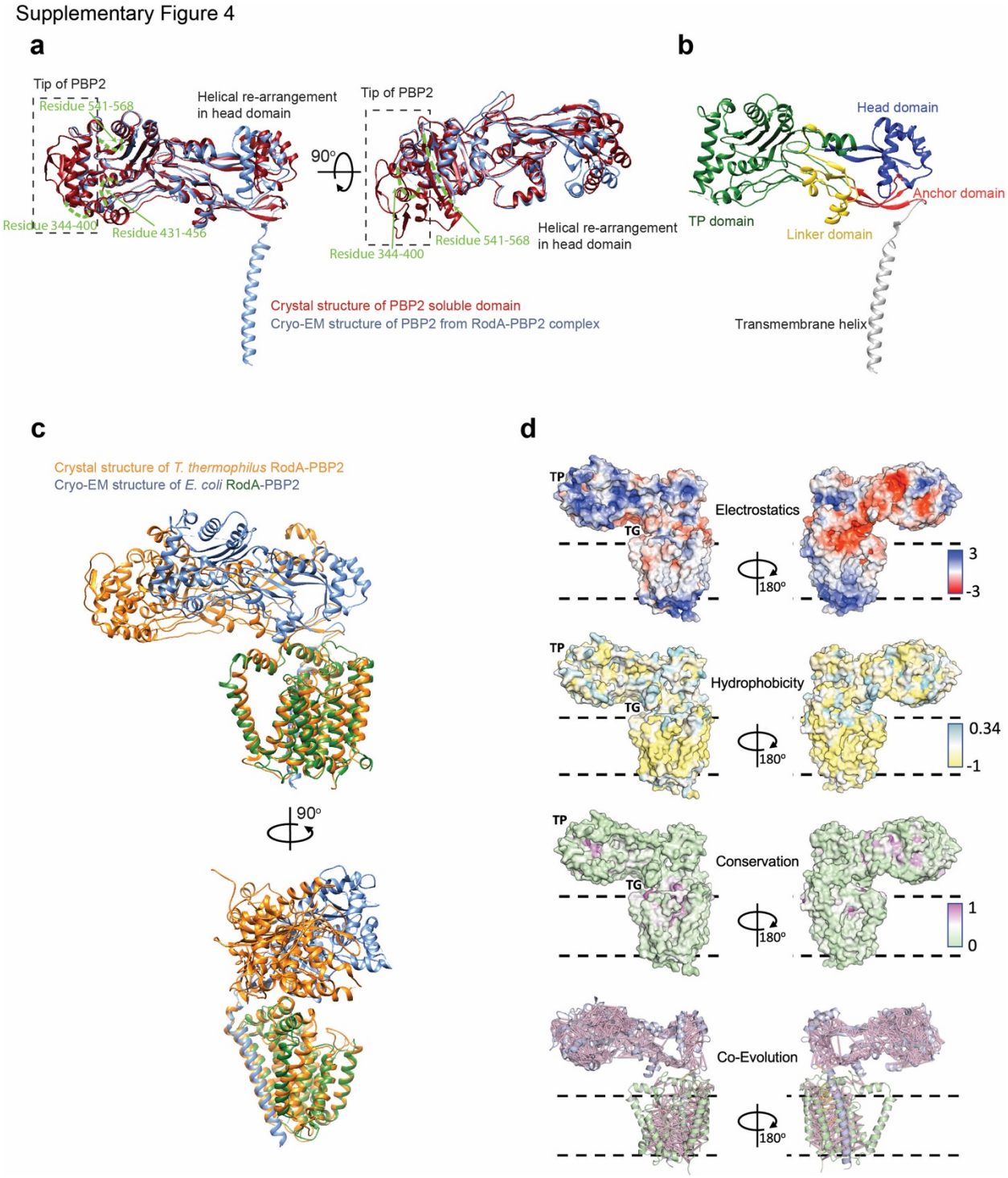


572

573 **Supplementary Figure 3 | Fit of cryo-EM density with model.**

574 Cryo-EM densities (mesh) are superimposed on secondary structure elements from RodA. The

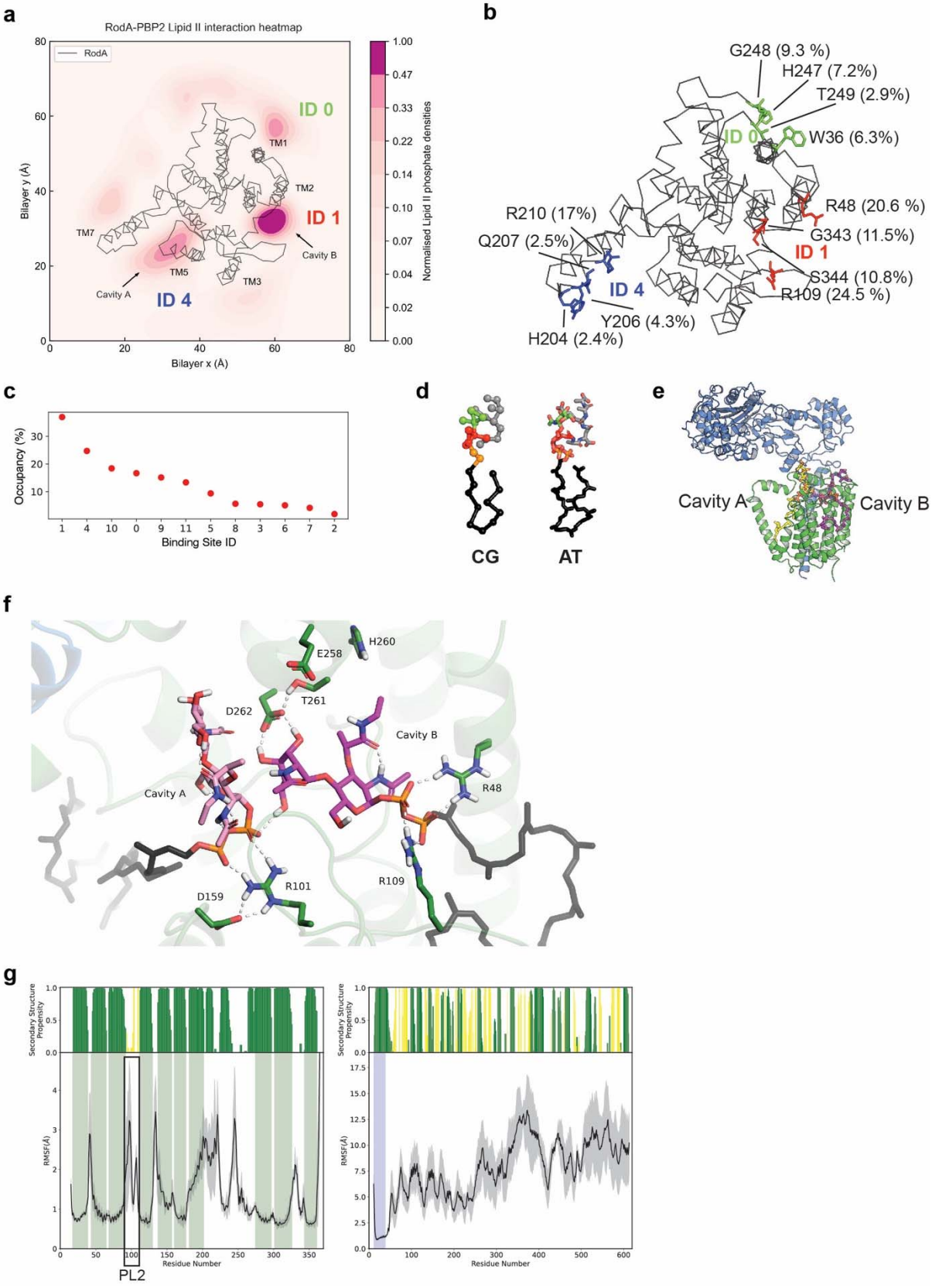
575 model is rendered in sticks using the color scheme as in Fig. 1d and e.



Supplementary Figure 4 | Structural features of the RodA-PBP2 complex. a) Comparison of the previously published X-ray crystallography structure of PBP2 without the transmembrane helix (PDB ID 6G9S) in red and PBP2 from our current cryo-EM structure of the RodA-PBP2 complex in blue. The tip of PBP2 looks to be partially unfolded in our structure and could not be resolved.

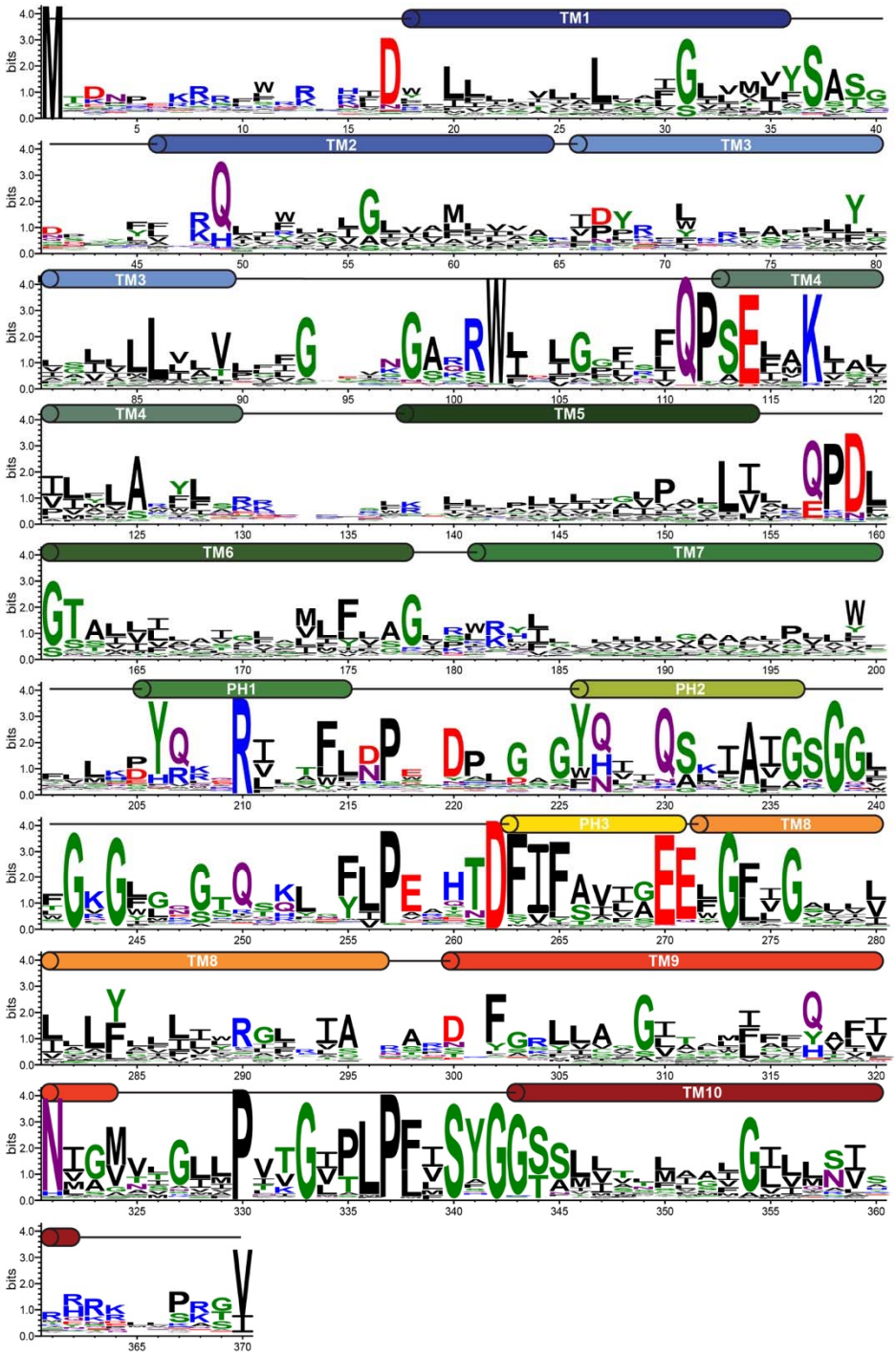
b) PBP2 from the RodA-PBP2 complex colored by sub-domain as in Levy et al[20]. **c**) Comparison of our structure of RodA-PBP2 with the one from *T. thermophilus*. The *T. thermophilus* structure is shown in orange and the *E. coli* one in blue (PBP2) and green (RodA). **d**) RodA-PBP2 rendered in surface representation coloured by electrostatic potential on a range of ±5 kBT/e, by Wimley-White hydrophobicity, on a cyan (very hydrophilic) to gold (very hydrophobic) scale, by conservation on a green (no conservation) to purple (absolute conservation) scale and by co-evolutionary analysis calculated using MapPred[40] and mapped onto the cryo-EM structure of RodA-PBP2, using a threshold of 0.272. Intramolecular predicted contacts between Cα are shown as purple dashes and intermolecular predicted contacts are shown as yellow dashes.

Supplementary Figure 5



Supplementary Figure 5 | Docking and MD simulation analysis. a) Density plots from CG MD simulations of Lipid II binding, with binding site IDs noted from PyLipID analysis. **b)** PyLipID occupancy analysis of the three main binding sites to RodA, with the top 4 residues with the greatest occupancy shown for each site. **c)** Ranking of the binding site IDs based on occupancy. Binding site ID 10 is not shown in **a** or **b**, as it is predominantly bound to the PBP2 head, not RodA. **d)** CG and atomistic (AT) representations of Lipid II. **e)** Representations of the bound Lipid II molecules to both cavity A and B. **f**) Close-up of the RodA active site from a minimized DFTB cluster model. Lipid II molecules bound in cavites A and B as well as key residues are shown as sticks and labelled. Not shown are the backbone atoms of the peptide fragments constituting the cluster. Hydrogen-bonds are indicated with white dashes. **g**) Root Mean Square Fluctuation (RMSF) analysis of the Cα trace of RodA-PBP2. The RMSF measurements were averaged across 3 repeats of 1 μs simulation. The grey shading refers to the standard deviation across the repeats. The bar graph on top of the RMSF plot shows the α-helix propensity over the course of the simulation as with the green bars. Black box highlights the residues in PL2.

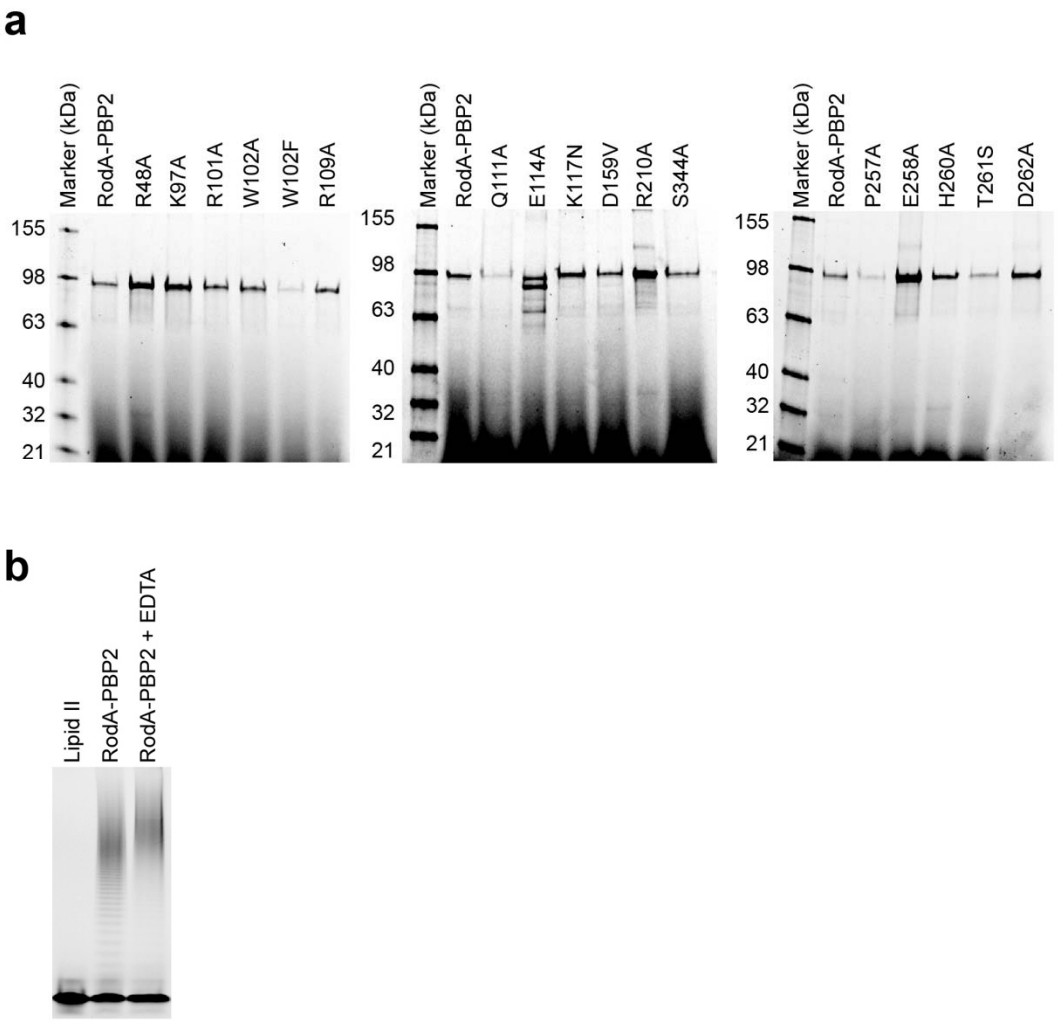
Supplementary figure 6



607

Supplementary Figure 6 | Conserved residues of RodA. Weblogo analysis of RodA created by Weblogov3[38] using the *E. coli* RodA sequence with MMseqs2[41] to search for homologous sequences within both UniRef100 and the environmental sequence databases. This resulted in ~14,000 sequences. Positions of RodA α-helical regions shown above weblogo coloured as in Fig. 1d. Key to residues colors: Black, hydrophobic; purple, hydrophilic; red, acidic; blue, basic; green hydroxyl & glycine.

Supplementary figure 7

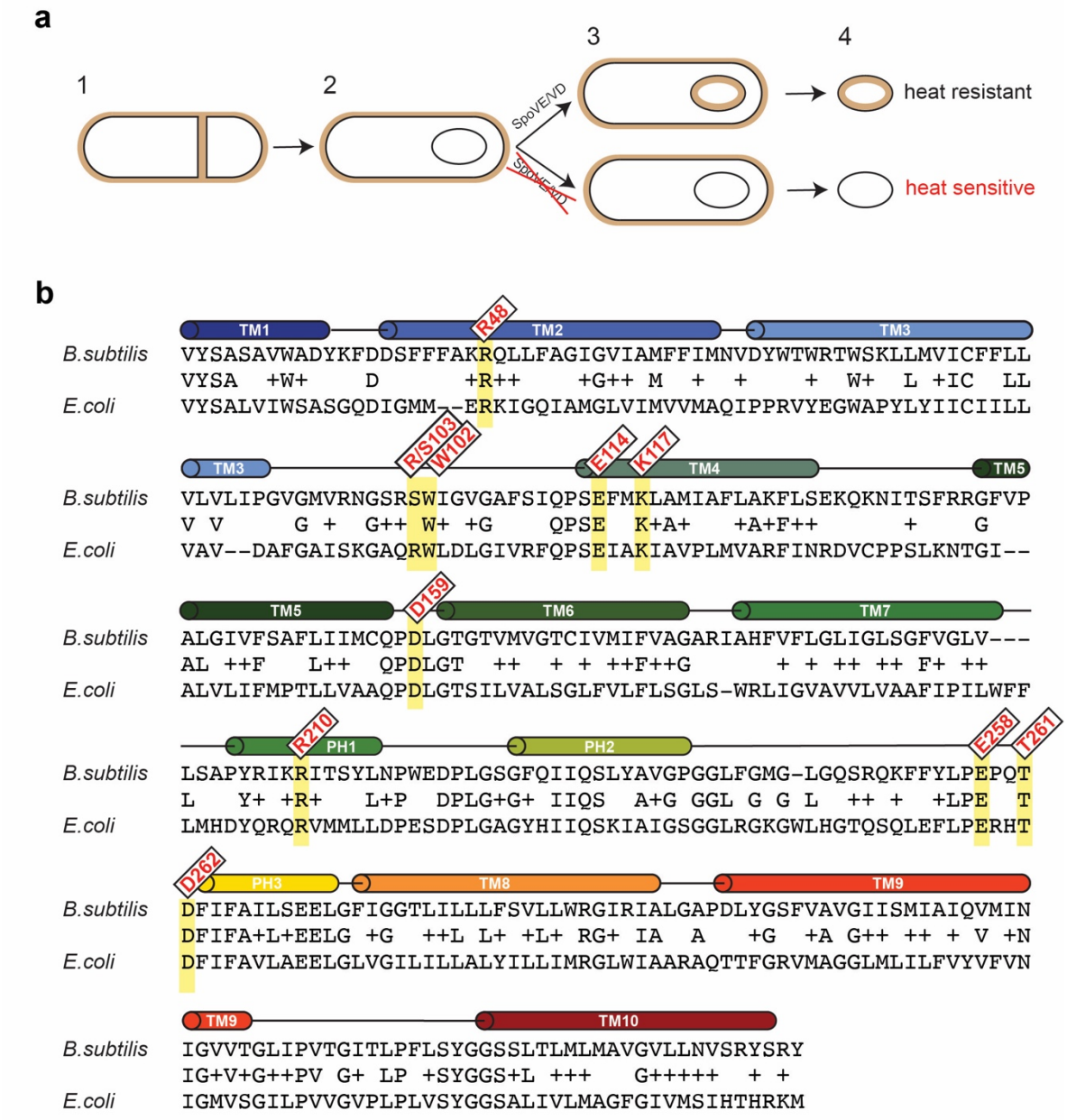


615

616 **Supplementary Figure 7 | Expression of mutant protein and metal-ion independence. a**)
617 Visualization of RodA-PBP2 protein mutants used for polymerization assays in Fig. 2e and 4c.
618 Bocillin staining confirms integrity of PBP2 transpeptidase active site and approximates protein
619 folding. **b**) Electrophoresis analysis of Lipid II polymerization products of RodA-PBP2 with and
620 without EDTA (35 mM).

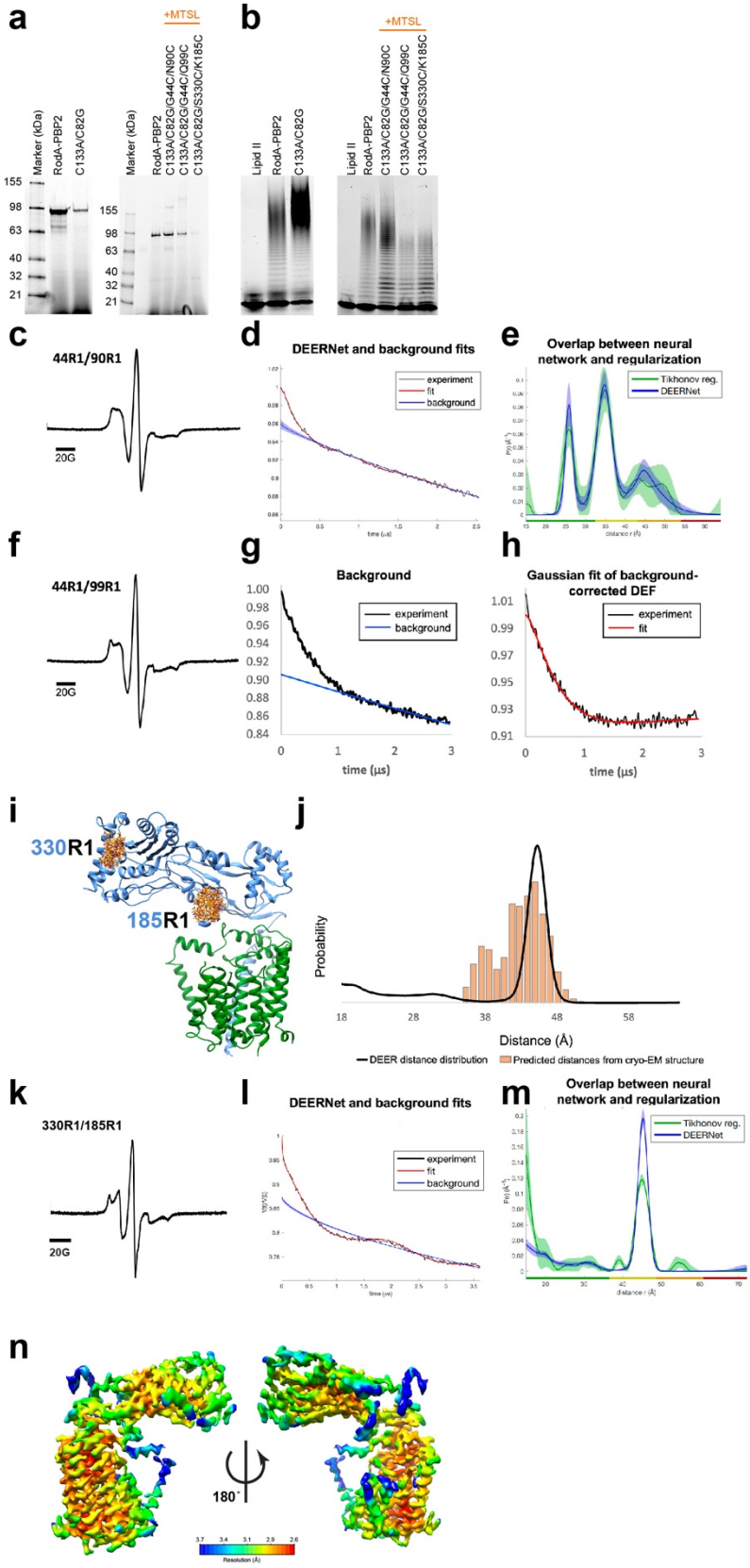
621

Supplementary Figure 8



Supplementary Figure 8 | Genetic analysis of SEDS-PBP function. a) Assay of sporulation-specific SpoVE-SpoVD function. (1) asymmetric division during sporulation; (2) formation of a membrane bound forespore compartment; (3) synthesis of spore PG is dependent on SpoVE/VD; (4) heat sensitivity of spores lacking SpoVE/VD. **b)** BLAST alignment of *E. coli* RodA-PBP2 fusion and *B. subtilis* SpoVE-SpoVD fusion.

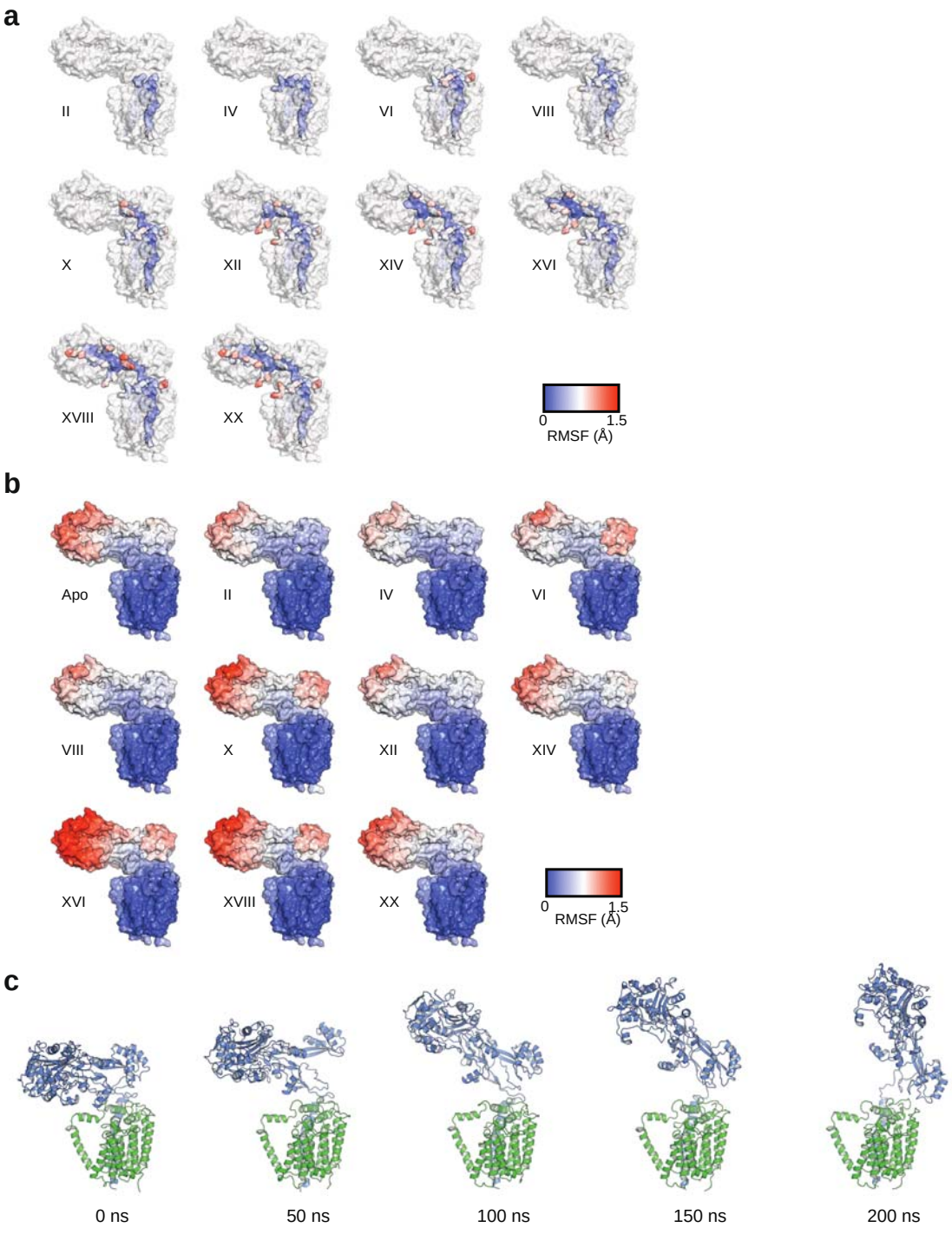
Supplementary Figure 9



Supplementary Figure 9 | Dynamics of RodA-PBP2. a) Visualization of RodA-PBP2 protein mutants used for used for the DEER experiments presented in Fig. 3 and 5 by Bocillin staining, confirming integrity of PBP2 transpeptidase active site and approximates protein folding. For the S300A mutant, there is no evidence of bocillin binding because S300 is its binding site in PBP2[20]. **b**) Electrophoresis analysis of Lipid II polymerization products of RodA-PBP2 for the cysteine-free background and double-cysteine constructs used for DEER experiments presented in Fig. 3 and 5 with spin label (MTSL). **c**) Continuous wave (CW) EPR spectra of spin labeled Gly44R1/Asp90R1 RodA-PBP2 mutant (44R1/90R1). Background estimation and fits of the resulting experimental DEER dipolar evolution function (DEF) (**d**) and resulting DEER distance distributions (comparison of Tikhonov regularization and DEERNet neural network) for 44R1/90R1 RodA-PBP2 calculated with the DEERNet extension of DEERAnalysis[42] (**e**). The DEERNet distance distribution was used for main figures and analysis. **f**) CW EPR spectra of spin labeled Gly44Cys/Gln99Cys RodA-PBP2 mutant (44R1/99R1). **g**) Background estimation of the resulting DEER data and **h**) Gaussian fit of the background-corrected DEF for 44R1/99R1 RodA-PBP2 done with DEERAnalysis[43]. **i**) Sterically allowed rotamers of nitroxide spin labels (R1) are illustrated for Lys185R1/Ser330R1 on the cryo-EM structure (orange sidechains). **j**) Resulting DEER distance distributions (Lys185R1/Ser330R1; solid black line) are compared to predicted distance distributions (orange bars) based on the sterically allowed rotamers shown on the structure. **k**) Continuous wave (CW) EPR spectra of spin labeled Lys185R1/Ser330R1 RodA-PBP2 mutant (185R1/330R1). **l**) Background estimation and fits of the resulting experimental DEER dipolar evolution function (DEF) and **m**) resulting DEER distance distributions (comparison of Tikhonov regularization and DEERNet neural network) for Lys185R1/Ser330R1

RodA-PBP2 calculated with the DEERNet extension of DEERAnalysis[42]. **n**) Local resolution display of orthogonal view of RodA-PBP2 from non-uniform refinement without a mask.

Supplementary figure 10



Supplementary Figure 10 | Analysis of MD simulations with varying length of the glycan strand. a) RodA-PBP2 is shown as a light grey surface representation with Lipid II-Lipid XX shown as surface representation colored by RMSF from blue to red. The polysaccharides are stably coordinated within the binding groove whereas the pentapeptide stem is more mobile. **b**) RodA-PBP2 shown as a surface representation colored by RMSF from blue to red. **c)** Snapshots from a single MD simulation of RodA-PBP2 with Lipid II and Lipid XVIII bound. The protein is shown in cartoon representation at 0, 50,100,150 and 200 ns timepoints.

663 **Supplementary Table 1. SEDS-PBP fusion proteins**

Bacterial species	SEDS UniProt ID	PBP UniProt ID	MW (kDa)
Enterococcus faecalis	Q820T8	Q836V7	126.8
Enterococcus faecalis	Q820T8	Q830D1	122.8
Klebsiella pneumoniae	A0A0H3GLD1	A0A0W8ASI8	112.0
Escherichia coli	**P0ABG7**	**P0AD65**	**111.8**
Streptococcus pneumoniae	A0A062WN05	P14677	128.2
Streptococcus pneumoniae	A0A062WN05	A0A0Y0HFB8	119.7
Streptococcus pneumoniae	A0A062WN05	A0A0H2URT5	126.5
Escherichia fergusonii	B7LLH8	B7LLH7	111.7

664

Supplementary Table 2. Summary of Cryo-EM Data Collection and Model Refinement

	RodA-PBP2 (EMD-XXX) (PDB XXX)
Data collection and processing	
Magnification	105,000
Voltage (kV)	300
Electron exposure (e–/Å^2)	58.5
Exposure time (s)	2.5
Dose rate (e–/pixel/s)	16.1
Nominal defocus range (μm)	-2.5 to -1
Defocus range (μm)	-2.4 to -1.5
Pixel size (Å)	0.83
Symmetry imposed	C1
Number of micrographs	11,120
Initial particle images (no.)	3,462,335
Final particle images (no.)	399,759 (Transmembrane region local refinement) and 104,553 (PBP2 periplasmic domain local refinement)
Map resolution (Å)	2.97 (Transmembrane region local refinement) and 2.95 (PBP2 periplasmic domain local refinement)
FSC threshold	0.143
Refinement	
Map sharpening *B* factor (Å^2)	-82.4 (Transmembrane region local refinement) and -84.3 (PBP2 periplasmic domain local refinement)
Residue range	RodA (9 to 93 and 109 to 365) and PBP2 (10-343, 401-430, 457-540 and 569-612)
Model composition	
Non-hydrogen atoms	6,475
Protein residues	833
Ligands	0
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.725
Validation	
MolProbity score	2.03
Clashscore	9.52
Poor rotamers (%)	1.16
Ramachandran plot	
Favored (%)	92.33
Allowed (%)	7.55
Disallowed (%)	0.12

Supplementary Table 3. Table summarizing genetic and biochemical activity assays for SpoVE-SpoVD and RodA-PBP2 fusions

Mutant	*In vivo* Sporulation	*In vitro* RodA GT activity
R48A	strong (~5%) (R48A)	NDA (Fig. 2)
K97A	ND	WTA (Fig. 2)
R101A	no defect (S103A)	LA (Fig. 2)
W102A	severe (W104A)	LA (Fig. 2)
W102F	severe (W104F)	WTA (Fig. 2)
R109A	ND	WTA (Fig. 2)
E114A	severe (E116A)	WTA (Fig. 2)
K117N	severe (K119N)	WTA (Fig. 2)
D159V	severe (D163V)	LA (Fig. 2)
R210A	severe (R212A)	LA (Fig. 2)
P257A	severe (P258A)	NDA (Fig. 4)
E258A	moderate (~10%)(E259A)	WTA (Fig. 4)
H260A	ND	WTA (Fig. 4)
T261S	moderate (~10%)	LA (Fig. 4)
D262A	severe (D263A)	NDA (Fig. 4)
S344A	ND	WTA (Fig. 2)

In vivo Sporulation: Sporulation efficiency was assayed by heat resistance relative to wild type. "severe" is when sporulation efficiency of is less than 10^{-7}; “strong” is <10%; “moderate” is 10%-50%; no defect is >50%. Mutant is in reference to the *E. coli* RodA sequence; corresponding residues in *B. subtilis* SpoVE are noted in brackets.
ND: not determined.

In vitro RodA GT activity: WTA is the GT activity corresponding to wild type (without mutations) RodA-PBP2 fusion (green), NDA is no detectable GT activity (red), LA is Low GT activity judged as <50% activity of wild type RodA-PBP2 fusion (amber).

Supplementary Table 4.

Mutant	Primer Sequences
G44C	Forward: AGC GGT CAG GAT ATT TGC ATG ATG GAG CGT AAA Reverse: TTT ACG CTC CAT CAT GCA AAT ATC CTG ACC GCT
R48A	Synthesized by Azenta Biosciences - Genewiz. Nucleotide sequence changed from CGT to GCT.
C82G	Forward: GCT GGG CCC CCT ATC TCT ATA TCA TCG GTA TTA TTT TGC TGG TGG CGG TAG ATG Reverse: CAT CTA CCG CCA CCA GCA AAA TAA TAC CGA TGA TAT AGA GAT AGG GGG CCC AGC
D90C	Forward: TTG CTG GTG GCG GTA TGT GCT TTC GGT GCC ATC Reverse: GAT GGC ACC GAA AGC ACA TAC CGC CAC CAG CAA
K97A	Forward: CGG TAG ATG CTT TCG GTG CCA TCT CTG CAG GTG CTC AAC GCT GGC TGG ACC TCG Reverse: CGA GGT CCA GCC AGC GTT GAG CAC CTG CAG AGA TGG CAC CGA AAG CAT CTA CCG
R101F	Forward: CAA TAC CGA GGT CCA GCC AGA ATT GAG CAC CTT TAG AGA TGG CAC Reverse: GTG CCA TCT CTA AAG GTG CTC AAT TCT GGC TGG ACC TCG GTA TTG
R101A	Forward: GTG CCA TCT CTA AAG GTG CTC AAG CCT GGC TGG ACC TCG GTA TTG TTC G Reverse: CGA ACA ATA CCG AGG TCC AGC CAG GCT TGA GCA CCT TTA GAG ATG GCA C
W102A	Forward: CAT CTC TAA AGG TGC TCA ACG CGC GCT GGA CCT CGG TAT TGT TCG Reverse: CGA ACA ATA CCG AGG TCC AGC GCG CGT TGA GCA CCT TTA GAG ATG
W102F	Forward: GTG CCA TCT CTA AAG GTG CTC AAC GCT TTC TGG ACC TCG GTA TTG TTC G Reverse: CGA ACA ATA CCG AGG TCC AGA AAG CGT TGA GCA CCT TTA GAG ATG GCA C
R109A	Forward: CTG GCT GGA CCT CGG TAT TGT TGC TTT TCA GCC GTC GGA AAT TG Reverse: CAA TTT CCG ACG GCT GAA AAG CAA CAA TAC CGA GGT CCA GCC AG
Q111A	Synthesized by Azenta Biosciences - Genewiz. Nucleotide sequence changed from CAG to GCG.
E114A	Forward: CTG GAC CTC GGT ATT GTT CGT TTT CAG CCG TCG GCA ATT GCC AAA ATA GCC GTA CCA CTG Reverse: CAG TGG TAC GGC TAT TTT GGC AAT TGC CGA CGG CTG AAA ACG AAC AAT ACC GAG GTC CAG
K117I	Forward: TAT TGT TCG TTT TCA GCC GTC GGA AAT TGC CAT AAT AGC CGT ACC ACT GAT GGT TGC G Reverse: CGC AAC CAT CAG TGG TAC GGC TAT TAT GGC AAT TTC CGA CGG CTG AAA ACG AAC AAT AC
K117R	Forward: CGT TTT CAG CCG TCG GAA ATT GCC CGA ATA GCC GTA CCA CTG ATG GTT GCG Reverse: CGC AAC CAT CAG TGG TAC GGC TAT TCG GGC AAT TTC CGA CGG CTG AAA ACG
K117N	Forward: CAG CCG TCG GAA ATT GCC AAC ATA GCC GTA CCA CTG Reverse: CAG TGG TAC GGC TAT GTT GGC AAT TTC CGA CGG CTG
C133A	Forward: CGC GCT TTA TCA ACC GCG ACG TTG CCC CGC CAT CGT TGA AGA ACA CTG Reverse: CAG TGT TCT TCA ACG ATG GCG GGG CAA CGT CGC GGT TGA TAA AGC GCG
D159V	Forward: GCC CAC GCT GCT GGT GGC TGC ACA GCC TGT CCT GGG AAC ATC AAT CCT CGT TGC G Reverse: CGC AAC GAG GAT TGA TGT TCC CAG GAC AGG CTG TGC AGC CAC CAG CAG CGT GGG C
R210A	Forward: CTG ATG CAT GAT TAC CAG CGC CAG GCC GTA ATG ATG CTC CTG GAC CCG G Reverse: CCG GGT CCA GGA GCA TCA TTA CGG CCT GGC GCT GGT AAT CAT GCA TCA G
E255A	Forward: CTC AGT CAC AGC TTG AAT TTC TCC CCG CAC GCC ATA CTG ACT TTA TCT TCG CGG TAC TG Reverse: CAG TAC CGC GAA GAT AAA GTC AGT ATG GCG TGC GGG GAG AAA TTC AAG CTG TGA CTG AG
P257G	Synthesized by Azenta Biosciences - Genewiz. Nucleotide sequence changed from CCC to GGC.
H260A	Forward: TCA CAG CTT GAA TTT CTC CCC GAA CGC GCT ACT GAC TTT ATC TTC GCG GTA CTG GC Reverse: GCC AGT ACC GCG AAG ATA AAG TCA GTA GCG CGT TCG GGG AGA AAT TCA AGC TGT GA

T261S	Forward: GAA TTT CTC CCC GAA CGC CAT GCT GAC TTT ATC TTC GCG GTA CTG Reverse: CAG TAC CGC GAA GAT AAA GTC AGC ATG GCG TTC GGG GAG AAA TTC
D262A	Forward: CCC GAA CGC CAT ACT GCC TTT ATC TTC GCG GTA Reverse: TAC CGC GAA GAT AAA GGC AGT ATG GCG TTC GGG
E270A	Forward: GCC ATA CTG ACT TTA TCT TCG CGG TAC TGG CGG CAG AGC TGG GAT TAG TGG GCA TTC Reverse: GAA TGC CCA CTA ATC CCA GCT CTG CCG CCA GTA CCG CGA AGA TAA AGT CAG TAT GGC
S344A	Synthesized by Azenta Biosciences - Genewiz. Nucleotide sequence changed from TCG to GCG.
Q476C*	Forward: AAA GTC GAT AAC GTG TGC CAA ACG CTG GAC GCT Reverse: AGC GTC CAG CGT TTG GCA CAC GTT ATC GAC TTT
F401STOP - RodA Only	Synthesized by Azenta Biosciences - Genewiz.
G452STOP RodA and PBP2 helix only	Synthesized by Azenta Biosciences - Genewiz.

Amino acid sequence of the RodA-PBP2 fusion construct:

MTDNPNKKTFWDKVHLDPTMLLILLALLVYSALVIWSASGQDIGMMERKIGQIAMGLV
IMVVMAQIPPRVYEGWAPYLYIICIILLVAVDAFGAISKGAQRWLDLGIVRFQPSEIAKIA
VPLMVARFINRDVCPPSLKNTGIALVLIFMPTLLVAAQPDLGTSILVALSGLFVLFLSGLS
WRLIGVAVVLVAAFIPILWFFLMHDYQRQRVMMLLDPESDPLGAGYHIIQSKIAIGSGGL
RGKGWLHGTQSQLEFLPERHTDFIFAVLAEELGLVGILILLALYILLIMRGLWIAARAQTT
FGRVMAGGLMLILFVYVFVNIGMVSGILPVVGVPLPLVSYGGSALIVLMAGFGIVMSIHT
HRKMLSKSVTSGSGSGSKLQNSFRDYTAESALFVRRALVAFLGILLLTGVLIANLYNLQI
VRFTDYQTRSNENRIKLVPIAPSRGIIYDRNGIPLALNRTIYQIEMMPEKVDNVQQTLDAL
RSVVDLTDDDIAAFRKERARSHRFTSIPVKTNLTEVQVARFAVNQYRFPGVEVKGYKRR
YYPYGSALTHVIGYVSKINDKDVERLNNDGKLANYAATHDIGKLGIERYYEDVLHGQT
GYEEVEVNNRGRVIRQLKEVPPQAGHDIYLTLDLKLQQYIETLLAGSRAAVVVTDPRTG
GVLALVSTPSYDPNLFVDGISSKDYSALLNDPNTPLVNRATQGVYPPASTVKPYVAVSA
LSAGVITRNTTLFDPGWWQLPGSEKRYRDWKKWGHGRLNVTRSLEESADTFFYQVAY
DMGIDRLSEWMGKFGYGHYTGIDLAEERSGNMPTREWKQKRFKKPWYQGDTIPVGIG
QGYWTATPIQMSKALMILINDGIVKVPHLLMSTAEDGKQVPWVQPHEPPVGDIHSGYW
ELAKDGMYGVANRPNGTAHKYFASAPYKIAAKSGTAQVFGLKANETYNAHKIAERLR
DHKLMTAFAPYNNPQVAVAMILENGGAGPAVGTLMRQILDHIMLGDNNTDLPAENPA
VAAAEDH

METHODS

SEDS-PBP fusion design

Using the *E. coli* FtsW sequence as a seed, 189 different bacterial SEDS protein orthologs were identified by homology. These orthologs were cloned in a high throughput pipeline with a decahistidine tag at the N or C terminus separated from the target gene by a TEV cleavage site by NYCOMPS/COMPPÅ scientists (New York Consortium on Membrane Protein Structure/Center on Membrane Protein Production and Analysis) housed at the New York Structural Biology Center (NYSBC). These targets were screened for expression and monodispersity by size exclusion chromatography in a variety of detergents as previously described[17, 44, 45]. Five resulting SEDS targets that were most favorable for structural characterization were used to design SEDS-PBP fusion proteins, based on the *Bacillus subtilis* SpoVE-SpoVD and *E. coli* FtsW-PBP3 fusions genetically and functionally characterized previously [12]. The PBP partners of the SEDS proteins were identified, and PCR amplified from genomic DNA (provided by NYCOMPS/COMPPÅ) from the following bacterial strains: *Enterococcus faecalis* V583, *Klebsiella pneumoniae subsp. pneumoniae* MGH 78578, *Escherichia coli* str. K-12 substr. MG1655, *Streptococcus pneumoniae* TIGR4, and *Escherichia fergusonii* ATCC 35469. PBPs were inserted by Gibson assembly[46, 47] at the C-terminal end of SEDS genes with a TSGSGSGS linker between the SEDS gene and the PBP gene (see Supplementary Table 4). Construct design is illustrated in Supplementary Fig. 1a. All resulting clones were sequence verified by Sanger Sequencing (Macrogen). UniProt IDs for SEDS proteins and PBP proteins of the eight unique fusion proteins are given below in Supplementary Table 1. The *Escherichia coli* fusion (RodA UniProt ID 0ABG7; PBP2 UniProt ID P0AD65) in the pNYCOMPS-N23 vector resulted in the best expression levels and was carried forward for structural studies by cryo-EM.

Protein expression, purification in detergent, and reconstitution in nanodisc

The RodA-PBP2 fusion from *E. coli* in the pNYCOMPS-N23 vector was used to transform 50 µL of BL21(DE3) pLysS *E. coli* competent cells, which were grown overnight in 20mL of 2xYT media supplemented with 100 µg/mL ampicillin and 35 µg/mL chloramphenicol at 37 °C with shaking at 220 revolutions per minute (RPM). The next day 800 mL of 2xYT media was inoculated with 10 mL of starter culture and left to grow at 37 °C with shaking at 220 RPM until OD_{600} was 0.8-1.0. Temperature was reduced to 22 °C and protein expression was inducted with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and the culture was incubated for 4 hours with shaking at 220 RPM. Cells were harvested by centrifugation at 3220 *g for 15 minutes at 4 °C, resuspended in 1x Phosphate Buffered Saline (PBS) and centrifuged again, then stored at -80 °C. Cell pellets were thawed and resuspended in buffer containing 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.0, 200 mM NaCl, 20 mM $MgCl_2$, 4 µL/100mL RNase, 10 µg/mL solid Dnase, 1:1000 cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche), 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 1 mM phenylmethylsulfonyl fluoride (PMSF). A glass homogenizer was used to resuspend the cell pellets using 6 mL of buffer for each 1 gram of cell pellet mass. The resuspended cells were lysed by passing through an Emulsiflex C3 homogenizer (Avestin) at 15000 psi for 3 passages. Lysate was ultracentrifuged at 134000 g for 30 minutes at 4 °C. Supernatant was discarded and membrane pellets were resuspended using a glass homogenizer in 30 mL High Salt wash buffer (20 mM HEPES, 500mM NaCl, 10 µg/mL solid Dnase, 4 µL/100 mL RNAse, 1 mM TCEP, 1:1000 cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche), 1 mM PMSF) per 800 mL of cell culture. Resuspended membranes were ultracentrifuged at 134000 g for 30 minutes at 4 °C. Supernatant was discarded, and membrane pellets were resuspended and solubilized using a glass homogenizer and 30 mL buffer

(20 mM HEPES pH 7, 500 mM NaCl, 20% glycerol, 10 µg/mL solid Dnase, 4 µL/100mL Rnase, 1:1000 cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche), 1mM TCEP, and 1mM PMSF and 1% n-Dodecyl β-D-maltoside (DDM; Anatrace) per 800 mL of cell culture. Solubilizing membranes were kept rotating for 2 hours at 4 °C. Subsequently, the sample was ultracentrifuged at 134000*g for 30 minutes at 4 °C. Supernatant is collected and combined with 500 µL of Ni-NTA agarose beads (Qiagen) for every 800 mL of cell culture that was harvested. The mixture rotates for 2 hours at 4 °C. The beads were then loaded onto a column and washed with 10 column volumes of buffer containing 20 mM HEPES pH 7, 500 mM NaCl, 60 mM imidazole, pH 7.5, 20% glycerol, and 0.1% DDM. Protein was eluted with 3 column volumes of buffer containing 20 mM HEPES pH 7, 500 mM NaCl, 300 mM imidazole pH 7.5, 20% glycerol, and 0.1% DDM. Imidazole was removed from the eluted protein by exchanging buffer to 20 mM HEPES pH 7, 500 mM NaCl, 20% glycerol, 0.1% DDM and 1 mM TCEP using a PD-10 desalting column (Cytiva). The concentration of RodA-PBP2 fusion protein was measured using a nanodrop spectrophotometer (Thermo Fisher). RodA-PBP2 fusion protein was combined with 1-Palmitoyl-2-oleoyl-sn-glycero-3-(57hosphor-rac-(1-glycerol)) (POPG) and Membrane Scaffold Protein 1E3D1 in the molar ratios of 1:300:5 respectively and rotated for 2 hours at 4 °C. Bio-beads (Bio-Rad) were added to the mixture, which was left to rotate at 4 °C overnight. Bio-beads were removed, and the mixture was combined with Ni-NTA agarose beads (same volume as used previously). The mixture rotated for 2 hours at 4 °C. The Ni-NTA beads were then loaded onto a column and washed with 10 column volumes of buffer containing 20 mM HEPES pH 7, 500 mM NaCl, and 60 mM imidazole. Reconstituted nanodiscs were eluted by adding 3 column volumes of buffer containing 20 mM HEPES pH 7, 500 mM NaCl, 300 mM imidazole. Eluted sample is concentrated to 500 µL using a 100kDa MWCO centrifugal filter (Amicon). The nanodisc complex

was further purified by loading onto a Superdex 200 increase 10/300 size-exclusion column (Cytiva) with a filtered and degassed buffer containing 20 mM HEPES pH 7, 150 mM NaCl and 1 mM TCEP.

Single-particle cryo-EM vitrification and data acquisition.

Purified RodA-PBP2 complex was concentrated to 0.66 mg/ml (5.9 μM) using a 100-kDa concentrator (Amicon). The sample was frozen using a Vitrobot (Thermo Fisher) by adding 3 μL of the purified protein complex to previously plasma cleaned (Gatan Solarus) 0.6/1-μm holey gold grid (Quantifoil UltrAuFoil) and blotted using 595 filter paper (Ted Pella, Inc) for 7.5 s with a blot force of 3 and a wait time of 30 sec at 4°C with >95% humidity. Images were recorded using a Titan Krios electron microscope (FEI), at the Columbia University Cryo-Electron Microscopy Center, equipped with an energy filter and a K3 direct electron detection filter camera (Gatan K3-BioQuantum) using a 0.83Å pixel size. An energy filter slit width of 20 eV was used during the collection and was aligned automatically every hour using Leginon[48]. Data collection was performed using a dose of ~58.5 e-/Å^2 across 50 frames (50 ms per frame) at a dose rate of approximate 16.1 e–/pix/s, using a set defocus range of -1 μm to -2.5 μm. A 100 μm objective aperture was used. 11,120 micrographs were recorded over a two-day collection.

Cryo-EM Data processing

Movie frames were aligned using Patch Motion Correction implemented in cryoSPARC v.2.12[49] using a B-factor during alignment of 500. Contrast transfer function (CTF) estimation was performed using Patch CTF as implemented in cryoSPARC v.2.12. CrYOLO[50] was used to pick particles. Particle picking resulted in 3,147,165 particles which were then extracted in cryoSPARC with a 400 pixel box size binned 4 times. The particles were classified using 2D classification in cryoSPARC v.3.2 using a batch size per class of 400 and ‘‘Force Max over poses/shifts’’ turned

off with 40 online-EM iterations and one full iteration. 2D classes with well-defined high-resolution features were selected resulting in a particle stack of 441,319 particles. The particles were then re-extracted using a 400 pixel box size without binning. One round of *ab initio* reconstruction was performed in cryoSPARC v.3.2 using three classes, with a maximum resolution set at 4 Å and an initial resolution at 9 Å. This resulted in two classes that were mirror images and one class with a shortened PBP2. We then went back to the binned particle stack of 3,147,165 particles and ran heterogenous refinement in cryoSPARC v3.2 using the three classes from the *ab initio* reconstruction and two decoy classes and a "Batch size per class" of 30,000. This heterogenous refinement was run three times using the particles from the three *ab initio* classes from the previous heterogenous refinement as input for the next heterogenous refinement. From this final heterogenous refinement the top two classes were selected (927,369 particles) and the particles were re-extracted without binning. This was followed by a non-uniform refinement with a final resolution of 3.24 Å. The particles were further sorted by two class heterogenous refinement in cryoSPARC v3.2 using the map from the 3.24 Å non-uniform refinement and a 10 Å lowpass filtered map as input. We used a "Batch size per class" of 30,000 and an initial resolution of 5 Å and a "Resolution of convergence criteria" of 100. This was followed by a non-uniform refinement with a final resolution of 3.17Å (600,855 particles). To achieve higher resolution for the TM region, the particles were further separated by a 3D variability analysis with a mask around the TM region of the protein excluding the nanodisc and a filter resolution of 4.5Å. The particles were clustered into five clusters and three of these clusters (399,759 particles) were used as input for a non-uniform refinement with a final resolution of 3.10Å followed by a local refinement with a mask around the TM region of the molecule to a final resolution of 2.97Å for the TM region. To achieve higher resolution for the periplasmic part of PBP2 we used the 600,855 particles from the

previous non-uniform refinement and ran 3D variability analysis with a mask around the periplasmic part of PBP2. The particles were clustered into ten clusters and the six best clusters were used as input for non-uniform refinement (236,435 particles) to a final resolution of 3.23Å. This was followed by a beam tilt refinement by image shift groups, followed by another non-uniform refinement to a final resolution of 3.08 Å. We further sorted the particles with another round of 3D classification in cryoSPARC using four classes, target resolution of 4 Å and 10,000 particles per epochs online expectation maximization (O-EM). This identified a stack of particles which we further refined using non-uniform refinement to a final resolution of 3.14 Å. Finally a local refinement was performed using a mask around the periplasmic region to a final resolution of 2.95 Å.

3D variability analysis for PBP2 movement

The 3D variability analysis shown in Fig. 5a is based on a subset of particles where a local refinement with a mask around the transmembrane region was imposed. From this map and with a mask around the periplasmic part of PBP2 a 3D variability was performed, the particles were divided into 6 clusters.

Structural model building and refinement

An initial model of RodA and the TM helix of PBP2 was built as a homology model to the published RodA-PBP2 complex X-ray crystal structure[15]. The model was fitted to the map as a rigid body in Chimera[51], the model was subsequently adjusted to the current density using Namdinator[19] and further refined using Coot[52-54] and PHENIX[55, 56] iteratively. For the soluble part of PBP2 the previously published X-ray structure of the soluble part of PBP2[20] was used as input in Namdinator[19] and further refined using Coot[52-54] and PHENIX[55, 56] iteratively.

Model analysis

A cavity search using the Solvent Extractor from Voss Volume Voxelator server[37] was performed using an outer-probe radius of 10 Å and inner-probe radius of 2 Å. Chimera[51], PyMOL and ChimeraX[57] were used to visualize the structures in the figures.

Lipid II preparation.

Dansylated lysine version for gel visualization studies of GT activity (dansyl lysine Lipid II), was produced by *in vitro* recapitulation of the synthetic pathway as detailed previously[58].

Polymerization of Lipid II by RodA-PBP2 and RodA alone

The glycosyltransferase activity by RodA-PBP2 was demonstrated using visualization of fluorescently labelled dansyl Lipid II molecules using a Tris-Tricine acrylamide gel based electrophoresis method[59]. Detergent solubilized RodA-PBP2 protein (1.5 µL) at a concentration of 2.7 µM in 300 mM imidazole, 250 mM NaCl, 20 mM HEPES, 0.05% n-Dodecyl-B-D-Maltoside (DDM), 20% Glycerol was added to 13.5 µL reaction buffer (10 mM $MgCl_2$, 100 mM NaCl, 50 mM HEPES, 20% DMSO, 0.03% Lauryldimethylamine oxide (LDAO), 10 µM dansyl lysine Lipid II) to a final protein concentration of 0.27 µM. The reaction mixture was incubated at 37 °C for 1 hr 30 min to allow for RodA dependent polymerization of Lipid II.

The resultant Lipid II polymer was denatured at 95 °C to stop the reaction and mixed with 5x loading dye (50 mM Tris-HCl pH 8.8, 4 % SDS, 40 % glycerol, 0.01 % bromophenol blue, DTT 200 mM) prior to electrophoresis on a Biorad Criterion 16.5 % gel run at 110 V for 80 mins with anode gel running buffer (0.1 M Tris-HCl pH 8.8) and cathode gel running buffer (0.1 M Tris-HCl pH 8.25, 0.1 M Tricine, 0.1% SDS). The gel was visualized by 10 second exposure to UV on a BioRad GelDoc imaging system using Biorad imaging capture software. Concentration dependent

assays were performed in the same manner, with altered concentrations of RodA-PBP2 or Lipid II added. MTSL labelled sample was prepared as described in EPR sample preparation below, before assay.

Bocillin labeling of PBP2

Fluorescent Bocillin™ labeling of PBP2 was performed by incubating purified 25 µL RodA-PBP2 fractions with 1 µL bocillin (1 mg/mL) for 15 min. The resulting mixtures were run on an 4-15% SDS page gel, then visualized for 4 s by the using a BioRad gel documentation system[60]. The images were visualized at maxima to include the fluorescein ladder, and no other marker, therefore only visualizing bocillin stained proteins on the gel.

Mutagenesis of pNYCOMPS-N23-RodAPBP2 *E.coli* for functional analysis

Mutagenesis of pNYCOMPS-N23 RodA-PBP2 was performed using the QuickChange II Site-Directed Mutagenesis Kit (Agilent) with custom primers (see Supplementary Table 4).

Conservation and Co-evolution analysis

A multiple sequence alignment for RodA was generated using MMseqs2[41] by searching both UniRef100 and environmental sequence sets for homologous proteins. Weblogo3[38] was used to analyze and represent the conservation of the aligned sequences of RodA as a Weblogo, with the *E. coli* K12 W3110 sequence used as reference. Consurf was used to represent the amino acid conservation on the surface of RodA[61].

Co-evolution analysis was performed using GREMLIN[62]. The *E.coli* K12 W3110 sequences of RodA (*mrdb*) and PBP2 (*mrdA*) were used as input sequences to identify homologues with a cut-off of E^{-10}. Based on this paired sequence alignment, GREMLIN was used, with default parameters

to find the co-evolutionary contacts within either RodA or PBP2 and also between RodA and PBP2.

Mutagenesis, spin labeling and sample preparation for EPR

Two native cysteines in *E. coli* RodA-PBP2 fusion (pNYCOMPS-N23 vector) were mutated to glycine and alanine (Cys82Gly and Cys133Ala) and cysteines were then introduced to the resulting cysteine-free background for site-directed spin labelling (SDSL) at residues hypothesized to create informative distance distributions by DEER EPR spectroscopy based on the available structural information. All mutagenesis was performed using the QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent) or QuickChange II Site-Directed Mutagenesis Kit (Agilent) with custom primers (see Supplementary Table 4). All mutants were tested for enzymatic function and bocillin binding as described above. In preparation for spin labelling, RodA-PBP2 mutants were purified in detergent as described above but were exchanged to a non-reducing buffer containing 20 mM HEPES pH 7, 500 mM NaCl, 20% glycerol, 0.1% DDM using a PD-10 desalting column (Cytiva) after elution. S-(2, 2, 5, 5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothiolate (MTSL; Santa Cruz Biotechnology) was added at a 15:1 MTSL to protein molar ratio. The reaction was incubated at 4°C overnight with agitation while protected from light. Excess spin label was removed with a PD-10 desalting column (Cytiva) using the same non-reducing buffer. The resulting protein was concentrated to 100-150 mM. For continuous-wave (CW) EPR experiments, 7 µL of concentrated protein was loaded into pyrex capillaries (0.6 mm id x 0.84 mm od; Vitrocom) and measured at room temperature. For pulsed (DEER) EPR experiments, deuterated glycerol was added to the concentrated protein sample to a final concentration of 20% (v/v) and 15-20 µL of sample was frozen in quartz capillary tubes (1.6 mm

od x 1.1 mm id; Vitrocom) using a bath of dry ice and isopropanol. Frozen samples were stored at -80 °C until pulsed EPR data were collected.

Continuous-Wave and pulsed EPR data acquisition and analysis

CW EPR measurements were taken using an X-band Bruker EMX continuous wave spectrometer with an ER4123D dielectric resonator (Bruker Biospin) at room temperature. CW Spectra were baseline corrected and normalized using Lab-VIEW software (provided by C. Altenbach, University of California at Los Angeles). Pulsed (DEER) EPR measurements were taken at 50K with a Q-band Bruker E580 EPR Spectrometer (Bruker Biospin) equipped with a 300 Watt traveling wave tube amplifier (Applied Systems Engineering) and an EN5107D2 resonator. A standard four-pulse DEER sequence was used for all measurements with p/2 and p pulse lengths varying with sample. A pump frequency is set at the maximum of the nitroxide spectrum and the observed frequency is set to 75 MHz lower. Increasing inter-pulse delays at 16 ns increments were utilized with a 16-step phase cycle during data collection. Accumulation times were typically between 12 and 36 hours, with a dipolar evolution time between 3 and 3.5 ms. Dipolar evolution data were processed using DEERAnalysis[43] with Gaussian model fitting or the DEERNet neural network plugin[42]. MTSL rotamers were attached to structures obtained from cryo-EM data or from MD simulations *in silico* with Multiscale Modeling of Macromolecules (MMM)[63] using the default rotamer library. Nitroxide-to-nitroxide distances were calculated from resulting structures and binned to the closest angstrom for comparison with experimental data.

In vivo* studies in *Bacillus subtilis

Assays were carried out as described in Fay et al[12]. Briefly, mutations were introduced into a SpoVE-SpoVD construct in a strain lacking *spoVD* and *spoVE* and sporulation (heat sensitivity) was assessed.

Lipid II Ligand Docking

Autodock vina-carb[64] was used to dock Lipid II, Lipid II C55 mDAP, and peptidoglycan fragments to the identified cavities and crevices of the RodA-PBP2 complex. Docked poses were converted to CHARMM36m and energy minimized to optimize the binding orientations and to ligate docked fragments of peptidoglycan to allow the formation of the longer polymerized lipids, e.g., Lipid XX, with parameters developed based on those previously published[65].

Molecular dynamics simulations

Coarse grained simulations

All coarse-grained (CG) MD simulations used the Martini 3 forcefield[66, 67]. Martini 3 topologies with elastic networks were generated for RodA and PBP2 protein chains using Martinize2[66]. The DSSP program was used for secondary structure assignment[68], and intra-chain elastic network force constants were set to 500 kJ mol^{-1} nm^{-2} with upper and lower elastic bond cut-offs of 1.0 and 0.5 nm, respectively.

Martini 3 bead types and mapping to the Lipid II molecule were performed manually, converting from previously published Martini 2 Lipid II parameters[69]. Changes were made according to the amino acid beading in Martini 3[67] and the suggested bead types from the protocol for creating small molecules in Martini 3[70]. To refine the parameters for Lipid II, atomistic simulations (3 x 100 ns) were performed using previously published parameters[65] obtained from the authors in a PE:PG (3:1) membrane with a single copy of Lipid II. CG simulations were also performed with the same composition (5 x 1 μs). Representations of CG and atomistic Lipid II molecules are shown in Supplementary Fig. 5d. Distributions of distances and angles were measured using gmx tools distance and gangle. For the all-atom simulations, the atoms were grouped according to their

bead types and the center of geometry measured. The solvent surface accessible area was measured using the gmx tool sasa. Plots were created using Matplotlib[71].

All model-systems were prepared in 3D periodic boxes by using the Memembed[72] and insane[73] methods to generate protein-solvating symmetric PE/PG bilayers (4:1 ratio of PE:PG). Where applicable, two molecules of Lipid II were positioned at random positions in the upper (periplasmic) leaflet. Initial box dimensions were set to 5.0 nm plus the maximum protein diameter, approximately 11x11x11 nm for RodA and 13x14x15 nm for RodA-PBP2 systems. Using insane, remaining voids were filled with water beads, with sodium and chloride ions at placed random positions representing a neutralizing salt concentration of 0.15 M. Systems were then subjected to steepest descent minimization with a tolerance of F_{max} = 100 kJ mol^{-1} nm^{-2}. In total, the RodA and RodA-PBP2 systems consisted of approximately 11,000 and 24,000 beads, respectively.

All production CG simulations were conducted using GROMACS 2021[74] using the built-in leap-frog integrator with a timestep of 0.02 ps unless otherwise stated. All production simulations sampled isothermic-isobaric ensembles at 310 K using the V-rescale thermostat (τ_t = 1.0)[75] and the C-rescale barostat for semi-isotropic pressure coupling at 1.0 bar (τ_p = 12.0)[76]. Pre-production equilibration runs used the Berendsen barostat and a timestep of 0.01 ps. Separate coupling groups were used for protein, lipid and solvent molecules (i.e., waters and ions). For electrostatics, the reaction-field method was used with a Coulomb cut-off of 1.1 nm (ε_r = 15 and $\varepsilon_r = \infty$ for r > 1.1 nm), and van der Waals (VdW) interactions also used a cut-off of 1.1 nm (both with the Verlet cut-off scheme). The P-LINCS algorithm expanded up to 4th order was used for the treatment of

holonomic constraints[77]. Each system was equilibrated for 10 ns, after which 10 μs production runs were prepared from the coordinates and velocities of the final frames of the equilibration trajectories. For RodA and RodA-PBP2, 50 repeats of the production simulations were conducted.

Density maps for the MD simulation in Fig. 3d were prepared by concatenating all repeat trajectories centered on RodA, then least-squares fitting to RodA backbone beads (using MDanalysis[78]) and using the open-source, community-developed PLUMED library[79] to log the x and y coordinates of the geometric centers of Lipid II phosphate beads. Because of the standard orientation of the protein in the bilayer (from Memembed), the z-axis of the simulation box is always perpendicular to the membrane plane and the resulting histogram of the x and y coordinates is functionally a bilayer-facing 2D projection of the particle density. The seaborn API was used to perform the kernel density estimate presented in the final figures. Lipid II interaction analysis was performed using PyLipID[23] to identify binding sites, interacting residues and occupancy times for the bound Lipid II molecules from the CG simulations.

Atomistic simulations

Atomistic simulations for apo RodA and RodA-PBP2 were set up using the same pipeline as CG simulations, except using the Martini 2.2 forcefield[66] and position restraints for protein backbone beads with force constants set to 1000 kJ mol^{-1} nm^{-2}. After 50 ns of CG equilibration the CG systems were converted into atomistic systems using the CG2AT program with the built-in "align" method to direct protein geometry towards the prepared EM structure (pre-CG-relaxation) during the conversion[80]. Alignment suggested a typical RMSD of only 0.14 nm over the course of CG equilibration. Overall, this protocol enabled preparation of all-atom membrane protein systems in well-equilibrated bilayers.

Where relevant, Lipid II and its polymerized forms (i.e., Lipid IV, Lipid XX etc) were added to the atomistic systems post-conversion (i.e., before atomistic equilibration) by aligning and energy minimizing the docked conformations in the two binding cavities observed (A and B).

Production simulations consisted of three repeats of 500 ns, each continued from a distinct equilibration trajectory. The system representing the cryo-EM structure of RodA-PBP2 with bound Und-PP was simulated for three repeats of 1 μs each. All atomistic simulations used the GROMACS 2021 leap-frog integrator with a timestep of 0.002 ps, the CHARMM36m forcefield[81] and TIP3P water model[82, 83]. Simulations were performed in the isothermal-isobaric ensemble at 310 K with the V-rescale thermostat (τ_t = 0.1), and the Parrinello-Rahman barostat for semi-isotropic pressure coupling at 1.0 bar (τ_p = 1.0)[84, 85]. Particle-mesh Ewald (PME) was used for treatment of electrostatics[86], and VDW interactions used a (Verlet) cutoff of 1.2 nm. All bonds involving hydrogens were converted to holonomic constraints which were treated with the 4th order P-LINCS algorithm.

For the PBP2 dynamics heatmap in Figure 4, each apo RodA-PBP2 production simulation was root-mean-squared fit to RodA in the first trajectory frame using MDAnalysis before calculation of simulated B-factors using the GROMACS utility gmx rmsf. The values presented in the final figure are from averaging over the five repeat simulations and have been mapped to the minimized RodA-PBP2 structure before equilibration.

Lipid II polymerization video

A linear interpolation of Lipid II polymerization was created using GROMACS tool gmx morph. This movie illustrates the elongation of PG, possible conformational changes of RodA associated with polymerization and the potential transition of Lipid-linked products from Cavity B to Cavity A to allow the next Lipid II to bind.

DFTB Cluster Model

From the atomistic MD simulations with Lipids-II-VI bound, trajectory frames were filtered based on simple distance cut-offs of 6.5 Å between D262, the site B lipid C4 hydroxyl (donor) moiety and the site A lipid C1 (acceptor) atom to highlight suitable initial geometries of the catalytically relevant atoms. From these, selected snapshots were used to construct cluster models for QM minimizations. QM atoms included a slab-shaped region of RodA surrounding D262 and the bound Lipid head-groups at each site. RodA atoms consisted of 6 separate whole peptide fragments spanning residues R48-K49, K97-W102, R109-Q111, D159-L160, E258-F263, and S340-G343. For each substrate molecule, two complete glycan units (GlcNAc-MurNAc) were included (i.e., Lipid II) but undecaprenyl lipid tails were truncated after the first isoprenyl fragment and similarly pentapeptide chains were truncated at the L-Ala Cα atom (inclusive of the sidechain). When defining the boundaries of the peptide fragments, only (nonpolar) bonds between main-chain carbon atoms were cut, so cuts were always made between Cα and amide (carbonyl) carbons at both ends (such that N-terminal ends of the fragments always contain the carbonyl moiety from the preceding residue). All cut bonds were capped with added hydrogen atoms, and all boundary atoms were fixed in-place during optimizations (i.e., with frozen Cartesian coordinates). Overall, the cluster model consisted of 590 atoms in total, with an overall charge of -1 and singlet multiplicity. All QM calculations reported made used the ORCA program (version 5.0.3) and the GFN2-xTB semiempirical density-functional with ALPB implicit solvation (water). To obtain

product structures from reactant structures, minimizations were steered with curated harmonic restraints extending the pyrophosphate leaving-group bond and pulling along the coordinate forming the glycosidic bond, after which the cluster geometry was re-relaxed without the restraints, converging on the nearest product minimum. The supplementary video is interpolated from a partially converged NEB calculation consisting of 8 Images connecting product and reactant geometries minimized from a representative snapshot.

References

1. Typas, A., Banzhaf, M., Gross, C.A. & Vollmer, W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol* **10**, 123-136 (2011).
2. Skalweit, M.J. & Li, M. Bulgecin A as a β-lactam enhancer for carbapenem-resistant Pseudomonas aeruginosa and carbapenem-resistant Acinetobacter baumannii clinical isolates containing various resistance mechanisms. *Drug Des Devel Ther* **10**, 3013-3020 (2016).
3. Page, J.E. & Walker, S. Natural products that target the cell envelope. *Current Opinion in Microbiology* **61**, 16-24 (2021).
4. Novak, R., Charpentier, E., Braun, J.S. & Tuomanen, E. Signal Transduction by a Death Signal Peptide: Uncovering the Mechanism of Bacterial Killing by Penicillin. *Molecular Cell* **5**, 49-57 (2000).
5. Zgurskaya, H.I., López, C.A. & Gnanakaran, S. Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. *ACS Infectious Diseases* **1**, 512-522 (2015).
6. Silver, L.L. Viable screening targets related to the bacterial cell wall. *Ann N Y Acad Sci* **1277**, 29-53 (2013).
7. Antibiotic Discovery and Development, Vols 1 and 2. *Antibiotic Discovery and Development, Vols 1 and 2*, 79-117 (2012).
8. Egan, A.J.F., Errington, J. & Vollmer, W. Regulation of peptidoglycan synthesis and remodelling. *Nat Rev Microbiol* **18**, 446-460 (2020).
9. Meeske, A.J. *et al.* SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537**, 634-638 (2016).
10. Sjodt, M. *et al.* Structure of the peptidoglycan polymerase RodA resolved by evolutionary coupling analysis. *Nature* **556**, 118-121 (2018).
11. Sauvage, E., Kerff, F., Terrak, M., Ayala, J.A. & Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiol Rev* **32**, 234-258 (2008).
12. Fay, A., Meyer, P. & Dworkin, J. Interactions between late-acting proteins required for peptidoglycan synthesis during sporulation. *J Mol Biol* **399**, 547-561 (2010).
13. Egan, A.J., Cleverley, R.M., Peters, K., Lewis, R.J. & Vollmer, W. Regulation of bacterial cell wall growth. *FEBS J* **284**, 851-867 (2017).
14. Rohs, P.D.A. *et al.* A central role for PBP2 in the activation of peptidoglycan polymerization by the bacterial cell elongation machinery. *PLoS Genet* **14**, e1007726 (2018).
15. Sjodt, M. *et al.* Structural coordination of polymerization and crosslinking by a SEDS-bPBP peptidoglycan synthase complex. *Nat Microbiol* **5**, 813-820 (2020).
16. McLeod, M.P. *et al.* The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. *Proc Natl Acad Sci U S A* **103**, 15582-15587 (2006).
17. Mancia, F. & Love, J. High-throughput expression and purification of membrane proteins. *J Struct Biol* **172**, 85-93 (2010).
18. Kocaoglu, O. & Carlson, E.E. Penicillin-binding protein imaging probes. *Curr Protoc Chem Biol* **5**, 239-250 (2013).

19. Kidmose, R.T. *et al.* Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526-531 (2019).
20. Levy, N. *et al.* Structural Basis for E. coli Penicillin Binding Protein (PBP) 2 Inhibition, a Platform for Drug Design. *Journal of Medicinal Chemistry* **62**, 4742-4754 (2019).
21. Contreras-Martel, C. *et al.* Molecular architecture of the PBP2–MreC core bacterial cell wall synthesis complex. *Nature Communications* **8**, 776 (2017).
22. Liu, X., Biboy, J., Consoli, E., Vollmer, W. & den Blaauwen, T. MreC and MreD balance the interaction between the elongasome proteins PBP2 and RodA. *PLoS Genet* **16**, e1009276 (2020).
23. Song, W. *et al.* PyLipID: A Python Package for Analysis of Protein–Lipid Interactions from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **18**, 1188-1201 (2022).
24. Ashraf, K.U. *et al.* Structural basis of lipopolysaccharide maturation by the O-antigen ligase. *Nature* (2022).
25. Alexander, J.A.N. & Locher, K.P. Emerging structural insights into C-type glycosyltransferases. *Curr Opin Struct Biol* **79**, 102547 (2023).
26. Galley, N.F., O'Reilly, A.M. & Roper, D.I. Prospects for novel inhibitors of peptidoglycan transglycosylases. *Bioorg Chem* **55**, 16-26 (2014).
27. Li, Y. *et al.* Identification of the potential active site of the septal peptidoglycan polymerase FtsW. *PLoS Genet* **18**, e1009993 (2022).
28. Schwartz, B., Markwalder, J.A., Seitz, S.P., Wang, Y. & Stein, R.L. A Kinetic Characterization of the Glycosyltransferase Activity of Eschericia coli PBP1b and Development of a Continuous Fluorescence Assay. *Biochemistry* **41**, 12552-12561 (2002).
29. Zawadzka-Skomiał, J. *et al.* Characterization of the Bifunctional Glycosyltransferase/Acyltransferase Penicillin-Binding Protein 4 of Listeria monocytogenes. *Journal of Bacteriology* **188**, 1875-1881 (2006).
30. Terrak, M. & Nguyen-Distèche, M. Kinetic Characterization of the Monofunctional Glycosyltransferase from Staphylococcus aureus. *Journal of Bacteriology* **188**, 2528-2532 (2006).
31. Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB-An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J Chem Theory Comput* **15**, 1652-1671 (2019).
32. Yakovlieva, L. & Walvoort, M.T.C. Processivity in Bacterial Glycosyltransferases. *ACS Chem Biol* **15**, 3-16 (2020).
33. Welsh, M.A., Schaefer, K., Taguchi, A., Kahne, D. & Walker, S. Direction of Chain Growth and Substrate Preferences of Shape, Elongation, Division, and Sporulation-Family Peptidoglycan Glycosyltransferases. *J Am Chem Soc* **141**, 12994-12997 (2019).
34. Knott, B.C., Crowley, M.F., Himmel, M.E., Zimmer, J. & Beckham, G.T. Simulations of cellulose translocation in the bacterial cellulose synthase suggest a regulatory mechanism for the dimeric structure of cellulose. *Chem Sci* **7**, 3108-3116 (2016).
35. Morgan, J.L., McNamara, J.T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nat Struct Mol Biol* **21**, 489-496 (2014).

36. Hudson, K.L. *et al.* Carbohydrate-Aromatic Interactions in Proteins. *J Am Chem Soc* **137**, 15152-15160 (2015).
37. Voss, N.R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res* **38**, W555-562 (2010).
38. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).
39. Martins, A. *et al.* Self-association of MreC as a regulatory signal in bacterial cell wall elongation. *Nature Communications* **12**, 2987 (2021).
40. Wu, Q. *et al.* Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics* **36**, 41-48 (2020).
41. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology* **35**, 1026-1028 (2017).
42. Worswick, S.G., Spencer, J.A., Jeschke, G. & Kuprov, I. Deep neural network processing of DEER data. *Sci Adv* **4**, eaat5218 (2018).
43. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Applied Magnetic Resonance* **30**, 473-498 (2006).
44. Mancia, F. & Love, J. High throughput platforms for structural genomics of integral membrane proteins. *Curr Opin Struct Biol* **21**, 517-522 (2011).
45. Bruni, R. & Kloss, B. High-throughput cloning and expression of integral membrane proteins in Escherichia coli. *Curr Protoc Protein Sci* **74**, 29 26 21-29 26 34 (2013).
46. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
47. Gibson, D.G. Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol* **498**, 349-361 (2011).
48. Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J Struct Biol* **151**, 41-60 (2005).
49. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
50. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology* **2**, 218 (2019).
51. Pettersen, E.F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
53. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
54. Casanal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* **29**, 1069-1078 (2020).
55. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367 (2012).
56. Afonine, P.V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).
57. Goddard, T.D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25 (2018).

58. Catherwood, A.C. *et al.* Substrate and Stereochemical Control of Peptidoglycan Cross-Linking by Transpeptidation by Escherichia coli PBP1B. *Journal of the American Chemical Society* **142**, 5034-5048 (2020).
59. Schägger, H. Tricine–SDS-PAGE. *Nature Protocols* **1**, 16-22 (2006).
60. Zhao, G., Meier, T.I., Kahl, S.D., Gee, K.R. & Blaszczak, L.C. BOCILLIN FL, a sensitive and commercially available reagent for detection of penicillin-binding proteins. *Antimicrob Agents Chemother* **43**, 1124-1128 (1999).
61. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350 (2016).
62. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030 (2014).
63. Jeschke, G. MMM: A toolbox for integrative structure modeling. *Protein Sci* **27**, 76-85 (2018).
64. Nivedha, A.K., Thieker, D.F., Hu, H. & Woods, R.J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *Journal of Chemical Theory and Computation* **12**, 892-901 (2016).
65. Kim, S., Pires, M.M. & Im, W. Insight into Elongation Stages of Peptidoglycan Processing in Bacterial Cytoplasmic Membranes. *Sci Rep* **8**, 17704 (2018).
66. de Jong, D.H. *et al.* Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* **9**, 687-697 (2013).
67. Souza, P.C.T. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods* **18**, 382-388 (2021).
68. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637 (1983).
69. Oluwole, A.O. *et al.* Peptidoglycan biosynthesis is driven by lipid transfer along enzyme-substrate affinity gradients. *Nature Communications* **13**, 2278 (2022).
70. Alessandri, R. *et al.* Martini 3 Coarse-Grained Force Field: Small Molecules. *Advanced Theory and Simulations* **5**, 2100391 (2022).
71. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **9**, 90-95 (2007).
72. Nugent, T. & Jones, D.T. Membrane protein orientation and refinement using a knowledge-based statistical potential. *BMC Bioinformatics* **14**, 276 (2013).
73. Wassenaar, T.A., Ingolfsson, H.I., Bockmann, R.A., Tieleman, D.P. & Marrink, S.J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J Chem Theory Comput* **11**, 2144-2155 (2015).
74. Abraham, M.J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25 (2015).
75. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101 (2007).
76. Bernetti, M. & Bussi, G. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* **153**, 114107 (2020).
77. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J.G.E.M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472 (1997).

78. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32**, 2319-2327 (2011).
79. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **185**, 604-613 (2014).
80. Vickery, O.N. & Stansfeld, P.J. CG2AT2: an Enhanced Fragment-Based Approach for Serial Multi-scale Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation* **17**, 6472-6482 (2021).
81. Huang, J. & MacKerell, A.D., Jr. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34**, 2135-2145 (2013).
82. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935 (1983).
83. MacKerell, A.D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616 (1998).
84. Nosé, S. & Klein, M.L. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055-1076 (1983).
85. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52**, 7182-7190 (1981).
86. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577-8593 (1995).

International Journal of
Molecular Sciences

MDPI

Review

A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-Negative Bacteria

Chris L. B. Graham [1], Hector Newman [1], Francesca N. Gillett [1,2], Katie Smart [1], Nicholas Briggs [1], Manuel Banzhaf [2] and David I. Roper [1,*]

[1] School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK; Chris.graham@warwick.ac.uk (C.L.B.G.); H.Newman@warwick.ac.uk (H.N.); F.Gillett@warwick.ac.uk (F.N.G.); katiesmart93@gmail.com (K.S.); N.Briggs@warwick.ac.uk (N.B.)
[2] School of Biosciences, University of Birmingham, Birmingham B15 2TT, UK; m.banzhaf@bham.ac.uk
* Correspondence: David.Roper@warwick.ac.uk

Abstract: Bacteria must maintain the ability to modify and repair the peptidoglycan layer without jeopardising its essential functions in cell shape, cellular integrity and intermolecular interactions. A range of new experimental techniques is bringing an advanced understanding of how bacteria regulate and achieve peptidoglycan synthesis, particularly in respect of the central role played by complexes of Sporulation, Elongation or Division (SEDs) and class B penicillin-binding proteins required for cell division, growth and shape. In this review we highlight relationships implicated by a bioinformatic approach between the outer membrane, cytoskeletal components, periplasmic control proteins, and cell elongation/division proteins to provide further perspective on the interactions of these cell division, growth and shape complexes. We detail the network of protein interactions that assist in the formation of peptidoglycan and highlight the increasingly dynamic and connected set of protein machinery and macrostructures that assist in creating the cell envelope layers in Gram-negative bacteria.

Keywords: peptidoglycan; interactions; *Escherichia coli*; outer membrane; envelope; network; protein-protein; seds; complexes; dynamic; gram-negative; cell division; ytoskeleton

Citation: Graham, C.L.B.; Newman, H.; Gillett, F.N.; Smart, K.; Briggs, N.; Banzhaf, M.; Roper, D.I. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-Negative Bacteria. *Int. J. Mol. Sci.* **2021**, *22*, 12831. https://doi.org/10.3390/ijms222312831

Academic Editor: Paola Brun

Received: 14 October 2021
Accepted: 3 November 2021
Published: 27 November 2021

1. Peptidoglycan in Gram-Negatives

Peptidoglycan plays a vital role in the maintenance of cell envelope integrity in bacteria generally, and in Gram-Negative bacteria it acts as a stabilising structure that is attached to both the inner and outer membrane lipid bilayers [1]. The peptidoglycan layer is formed of a repeating beta-1–4-linked *N*-acetylmuramic acid N-acetylglucosamine disaccharide glycan polymer (MurNAc-GlcNAc) with crosslinked peptide side chains. The peptide side chains of each of these polymers can extend from the MurNac sugar and crosslink to create a macroscopic mesh-like structure (Figure 1) [2]. The cell constantly modifies this mesh-like macromolecule with a set of hydrolases to break the bonds involved and transferases to form new polymers allowing for cell expansion, shape changes, and septation. A recent review covers these modifications and the proteins involved in detail [3].

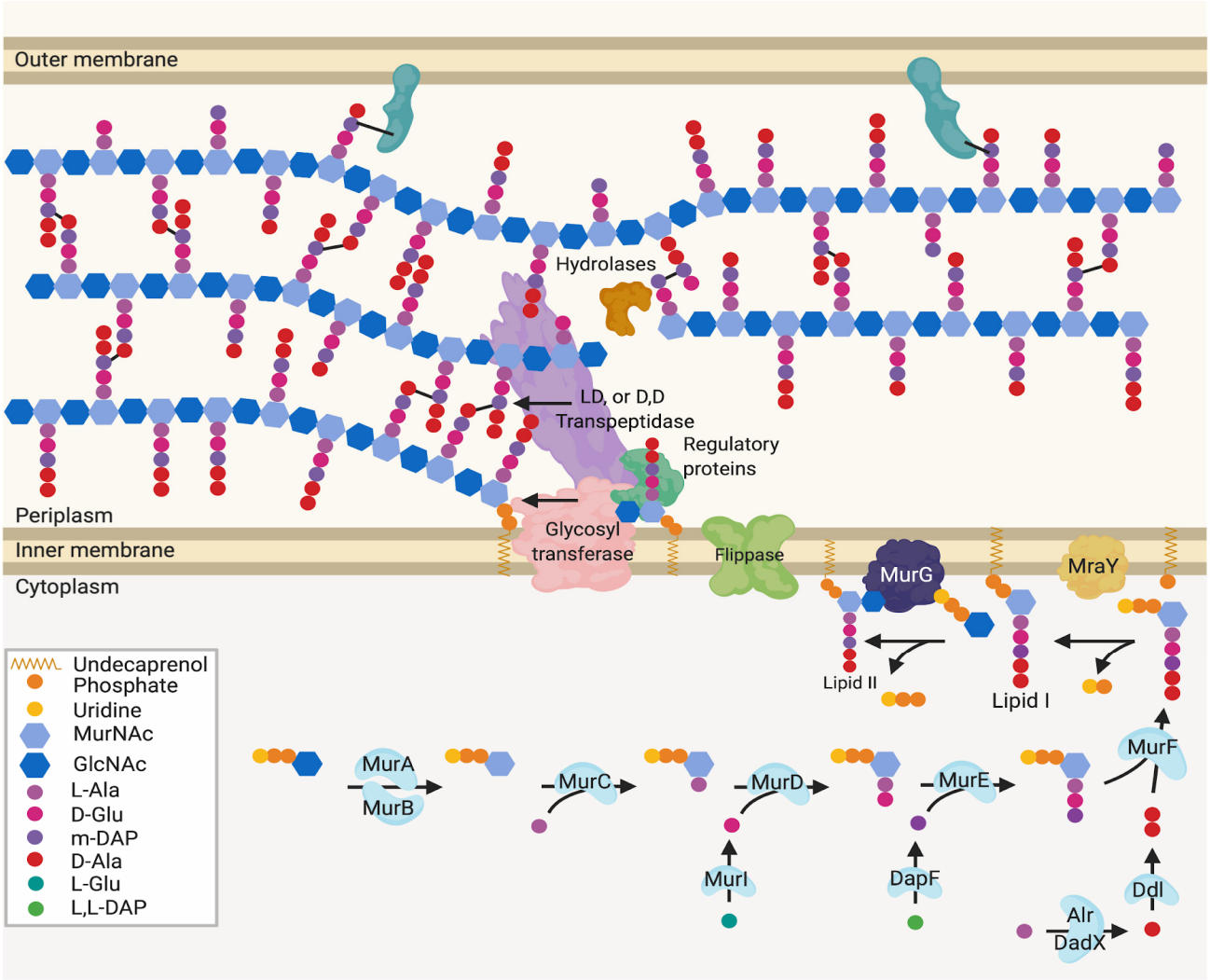


Figure 1. Generalised peptidoglycan synthesis and insertion pathway. Lipid II is the peptidoglycan building block precursor. This precursor is synthesised in the cytoplasm by sequential enzymatic steps then attached to undecaprenyl phosphate in the inner membrane [2,4]. The newly formed Lipid II is then flipped across the inner membrane and polymerised into glycan chains by the glycosyltransferase (GT) action of class A bifunctional penicillin-binding proteins (PBPs), Sporulation, Elongation or Division proteins (SEDS) in complex with class B monofunctional PBPs or monofunctional glycosyltransferases [5].

2. Cell Wall Modifying Enzymes and Complexes Have Altered Localisation during Growth Which Is Essential for Specialised Peptidoglycan Biosynthesis

The location of the enzymes required for the synthesis of peptidoglycan and its later modification (Figure 1) can vary, dependent upon cellular events and conditions [6,7]. The proteins and complexes involved are also dynamic, as many of their locations have been shown to change during the cell cycle. Studies using fluorescent gene fusions within the chromosome and peptidoglycan protein tracking approaches [8–10] now provide indications of coordinated peptidoglycan protein complex movement during the cell cycle [11,12] (Figure 2A–C). Localisation of these complexes presumably ensures that peptidoglycan is synthesized at particular regions for either overall growth or highly specialised growth situations such as cell division (Figure 2D); cell curvature; (Figure 2G) polar growth and maintenance (Figure 2E); as well as flagella associated regions (Figure 2F).

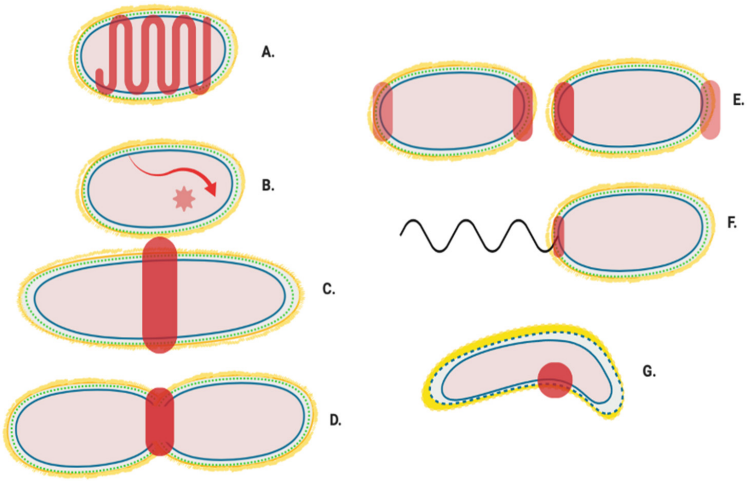


Figure 2. Generalised localisation of peptidoglycan modifying proteins. Localisation regions of known and potential peptidoglycan modifying enzymes in Gram-negative bacteria. Localisation sites are highlighted in red. (**A**) Helical and MreB associated Elongasome; (**B**) Free diffusion (unlocalised) [11]; (**C**) Pre-septal machinery; (**D**) Division machinery; (**E**) Post septal polar machinery and polar growth [13–15]; (**F**) Flagella peptidoglycan modification machinery [16], and (**G**) Shape determining pinpoint/seam [17,18].

3. Regulation of Peptidoglycan Modifying Enzymes by Their Interacting Partners

To achieve such diversity in the form and location of peptidoglycan, its synthesis and subsequent modification must be highly coordinated. Peptidoglycan is a complex three-dimensional molecule with architecture and chemistry which is dependent on host species and environmental localisation [19,20]. Cells also respond (via modified synthetic pathways) to antibiotic challenge and changing osmotic conditions by changes in their cell wall architecture and peptidoglycan biochemistry [21,22] (Figure 3).

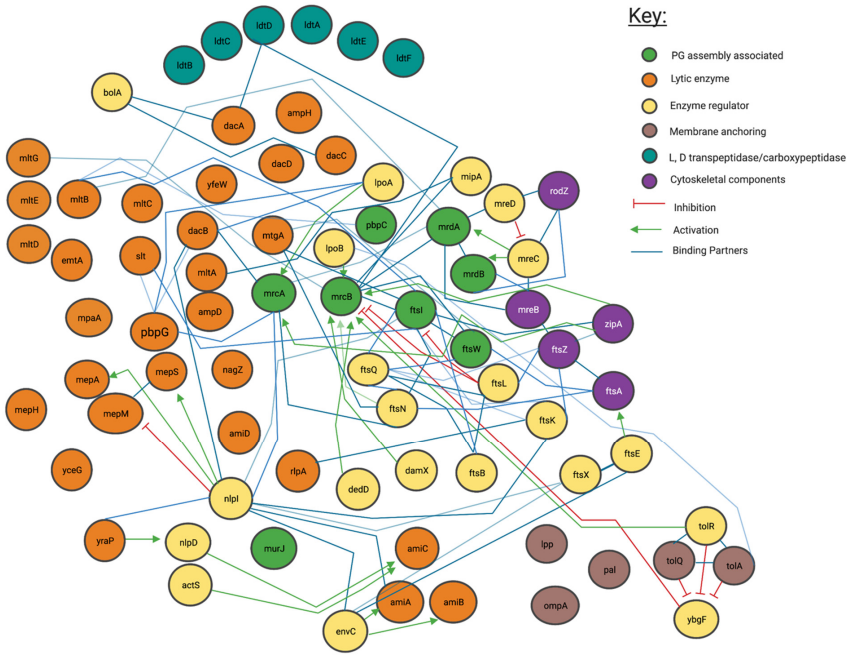


Figure 3. Interaction network of peptidoglycan modifying enzymes and their partners from Literature inspection. Interaction map of peptidoglycan-associated proteins sorted by enzymatic action. Network structure determined by STRING, with manual addition of interactions through literature associated with each protein. Reference matrix available in Appendix A.

The specialisation of peptidoglycan has been postulated to be driven by pathways that are regulated by local enzyme concentrations and protein: protein interactions. Integral peptidoglycan synthesis complexes such as RodA-PBP2 and FtsW-PBP3 have been shown to have non-enzymatic regulatory partners such MreC/MreD [23] and FtsN/L/Q respectively [24]. In addition enzymatic regulatory pairs exist e.g., PBP1A-PBP2 and PBP1b-PBP3. These networks of peptidoglycan synthesizing enzymes and regulatory proteins is still not understood either structurally or functionally.

4. Method Used to Visualise PG Synthesis Networks for This Meta-Review

To visualise the interactions of the genes and proteins relevant to peptidoglycan synthesis and allow a full informative meta-analysis, we have performed a network analysis of relevant genes using contemporary bioinformatic approaches [25].

4.1. Genetic and Protein Interactions Confirmed by the Literature

Peptidoglycan modifying and related genes, as listed in cell division, peptidoglycan biosynthesis, and peptidoglycan related papers centred on *E. coli* were collated to create a peptidoglycan relevant gene list [6,24,26]. We submitted this list of genes as a joint submission to the gene data trawling engine "STRING" (STRING) to create an interaction map centred around our listed proteins' data. Genes that interacted with this initial list, or were not in our initial list and given a combined "STRING" score of $\geq$0.7 (determined by co-occurrence data among species, gene neighbourhood scores and/or experimental data) [26,27] were then added and this new list was inputted as a meta-submission. Those genes which after meta-submission were found to have compound interaction scores with other listed genes > 0.9 were used in our final literature analysis (Figures 3 and 4). This has created a comprehensive picture of the current literature representing peptidoglycan synthesis and modification in Gram-Negative bacteria (Figure 3).

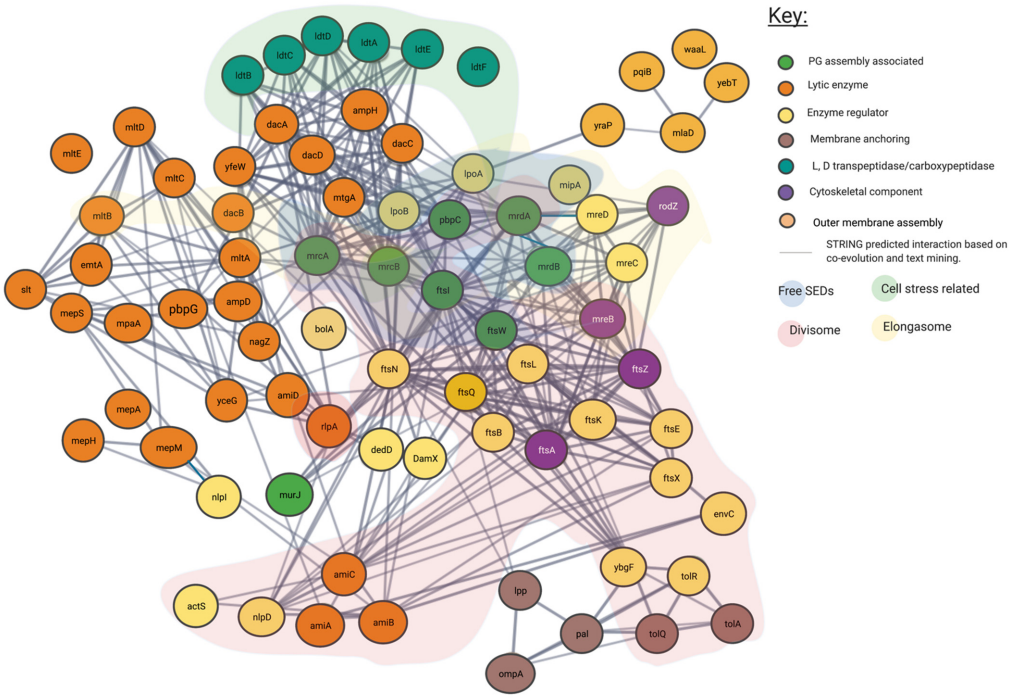


Figure 4. Predicted interaction network of peptidoglycan associated proteins using STRING. Interaction map of peptidoglycan associated proteins, sorted by enzymatic action. Network structure determined by STRING. SEDs complex interaction type is indicated by area colouring and arrows. Interactive network link: https://version-11-5.string-db.org/cgi/network?taskId=bIzLkBRoqjLb&sessionId=bBi0rwtoih3p (Last accessed on 4 November 2021).

4.2. Proposed Genetic Interactions

In addition to the confirmed interactions uncovered by the literature as shown in Figure 3, we show the predicted network used to create it. In Figure 4 unconfirmed STRING determined interactions of 0.7 or higher are shown. This more expansive and connected network represented in Figure 4 was constructed using a database of known and predicted gene-gene/protein-protein interactions derived from direct (physical) and indirect (functional) genetic associations, along with interactions aggregated from other (primary) databases [26,28]. We have grouped proteins currently linked to specific peptidoglycan assembly machinery and their cellular locations into their respective groups through the layering of background colour. This creates a gene-gene interaction pattern network, contextualised by the large protein assemblies they associate, such as the "divisome" and the "elongasome", as well as other potential environment dependent complexes.

5. Most Peptidoglycan Synthases and Modifiers Are Members of Multiple Local Complexes as Predicted by Genetic Interactions and Confirmed by Literature

Our network analysis (Figures 3, 4 and A3) suggests that the overlapping protein complex localisations of Figure 2 such as those related to cell stress, the "divisome" and "elongasome" involve some of the same proteins, which are shared among complexes. In particular our analysis of Figure 3, which focuses on known literature-verified interactions, indicates that some proteins interact with multiple complexes that facilitate coordination at the cytoplasmic membrane and outer membrane. These interactions can involve lytic, regulator, anchoring, cytoskeletal, and anabolic enzymes, often acting as partners to the same proteins.

The central message revealed by the networks suggests that regulation of proteins occurs in complexes, but also through protein exchange and sharing occurs to enable a range of possible additional ensembles. The roles of each protein, and an interaction table for these proteins in the network above are discussed in Appendix A, Tables A1–A3 as well as Figures A1–A3. A datafile of the literature in context of this interaction matrix is available (Supplementary Materials).

The networks of literature and genetic interaction in Figure 3 created through the use of an interaction matrix from the literature confirms already widely held theories, that the differences in central peptidoglycan formation units beyond the elongasome and divisome, in addition to other peptidoglycan formation nodes are often based upon changes in accessory proteins and exchange of core proteins and there is rarely a fixed static complex. Figure 4 is derived from a variety of different data sources and shows a complex web of possible interactions. However, all these interactions are unlikely to happen simultaneously in the cell and represent a large spectrum of possible connections. Some of these interaction have been observed experimentally as shown in Figure 3, with proteins such as PBP1b making interactions with a large array of other protein partners, actions highly unlikely to all occur at the same time. This concept exemplifies the observation of several proteins theoretically occurring in multiple protein complexes as indicated by Figure 4. These figures show the overwhelming complexity of the peptidoglycan protein network, as a collection of many complexes.

The pattern of protein re-use and interdependence however is not constant, for example, the monofunctional glycosyltransferase activity of FtsW and RodA have already been shown to be reliant on their partner class B PBP interactions [24,29,30]. These partner dependent glycosyltransferase proteins form a complex with specific class B PBPs such as *mrdA*/PBP2 for RodA and *ftsI*/PBP3 for FtsW (Figure 2) and in doing so produce codependent glycosyltransferase and transpeptidase peptidoglycan machines [10,31,32]; which are regulated by additional cytoskeletal and regulatory proteins. In *E. coli*, simplistically this includes the RodA-PBP2 (*mrdB-mrdA*) complex for elongation and the FtsW-PBP3 (*ftsW-ftsI*) complex for cell division [10]. These complexes act as nodes, (displayed in Figures 3 and 4) creating the basis of the dependent complexes of the elongasomes and divisomes, which can also interact with each another which we will discuss later in this article.

Existing alongside these functionally relevant codependent multimers, are bifunctional transpeptidases and glycosylase class A PBPs such as PBP1B and PBP1a with the codependent multimers can form interaction partnerships. These have also been shown to form their own complexes and act independently of these peptidoglycan machine nodes to modify and synthesis peptidoglycan, introducing further complexity [33].

All the complexes shown, including the Class A PBP centric ones, contain a host of additional regulatory, structural and enzymatically essential proteins that are shared among them, with interactions that span across the network, each interaction determining their specific overall activity dependent on interacting protein concentration, local substrates and their overall lipid environment [10,29–31] (Figure 4). Complexes of proteins such as the web of potential interactions shown in Figures 3 and 4 drive cell envelope synthesis. This review attempts to explain these interactions and their importance through stories presented by the current literature and investigates the specific nodes to which they centre.

6. Cytoskeletal Proteins Create Nodes of Complex Formation

As shown in Figure 4, some of the peptidoglycan modifying enzymes can be directly linked to cytoskeletal proteins. These cytoskeletal proteins also act as molecular treadmills [32,34,35] (Figure 4). Typically, two treadmills run around the circumference of the Gram-negative cell, one for division and one for elongation made of FtsZ and MreB filaments respectively [34,36]. FtsZ is a cytoskeletal component, that has been shown to localise and move along the cell circumference during cell division in individual strands that make up a macromolecular "Z ring" [37,38]. Soon after the discovery of FtsZ as a division-essential component, it was shown to interact with the peptidoglycan synthesis machinery of FtsW and FtsI/PBP3 and the regulatory proteins FtsQ, B and L [13,14]. Since then it has been associated with many other penicillin-binding proteins including PBP1b [11,12] (Figure 3). It is thought that the polymerization of FtsZ is responsible for the directional movement of the SEDs complexes during division, constriction and septation [36,39]. As a result, this is an important cell division protein. An antagonist of FtsZ polymerisation, viriditoxin causes cell division defects [40]. In Archaea, its analogous FtsZ and often multiple of its paralogues are integral as division orchestrating proteins associated with pseudomurein laydown. These genes cause cell division defects if not genetically present [41]. Similarly, MreB is another cytoskeletal component implicated in cell shape [42] shown to co-localise with the elongasome associated proteins during cell growth. It has been shown to bind the mur ligases which produce the lipid II precursor. The MreB polymerization antagonist A22 too causes cell morphology defects, highlighting the importance of both MreB and FtsZ as shape determining proteins [43].

Though FtsZ and MreB are important for correct cell growth and cell division, the literature has shown some of the peptidoglycan modifying machinery may only transiently attach to FtsZ and MreB treadmills, although it is not always necessary for their function [12,42]. Recent models of transient FtsI-FtsZ interactions by "Brownian-ratcheting" would suggest the peptidoglycan production and modification complexes move in and out of interaction with cytoskeletal components by a transient system of attachment to a cytoskeletal component, followed by protein wandering, allowing the peptidoglycan altering complex speeds to differ dependent on the degree of cytoskeletal attachment [44]. This "Brownian-ratcheting" model hypothesises a zone of active peptidoglycan producing, slower-moving complexes near the faster moving FtsZ rings or MreB filaments that transport inactive complexes in a dynamic equilibrium of interaction with the cytoskeletal nodes [10,36,37].

7. The "Elongasome" Is a Collection of Multiple Complexes

Figure 4 indicates all the known interactions of the complex machinery creating peptidoglycan. To understand the general mechanism for peptidoglycan synthesis and modification across species in the context of all steps involved in peptidoglycan creation, a specific example can be called upon. One of the core peptidoglycan biosynthesis and modifying

complexes is the "elongasome"; a biosynthetic complex of peptidoglycan manufacturing machinery spanning the periplasm and inner membrane, used during cell growth.

The elongasome complex (Figure 5) contains the components of generalised peptidoglycan synthesis (Figure 1), but we postulate as have others, based on our bioinformatic analysis and the literature (Figure 3), that the complex also contains class C PBP D, D-carboxypeptidases and lytic transglycosylases to modify peptidoglycan structure and prime it for attachment with new peptidoglycan [1,45]. This model contains the core monofunctional class B transpeptidase PBP2 which inserts its single transmembrane helix into the seven transmembrane helices of RodA, activating it as a glycosyltransferase [46]. This central peptidoglycan biosynthetic capability of the core of RodA-PBP2 is then augmented by the bifunctional Class A PBP, PBP1a also, which associates with the core complex [47].

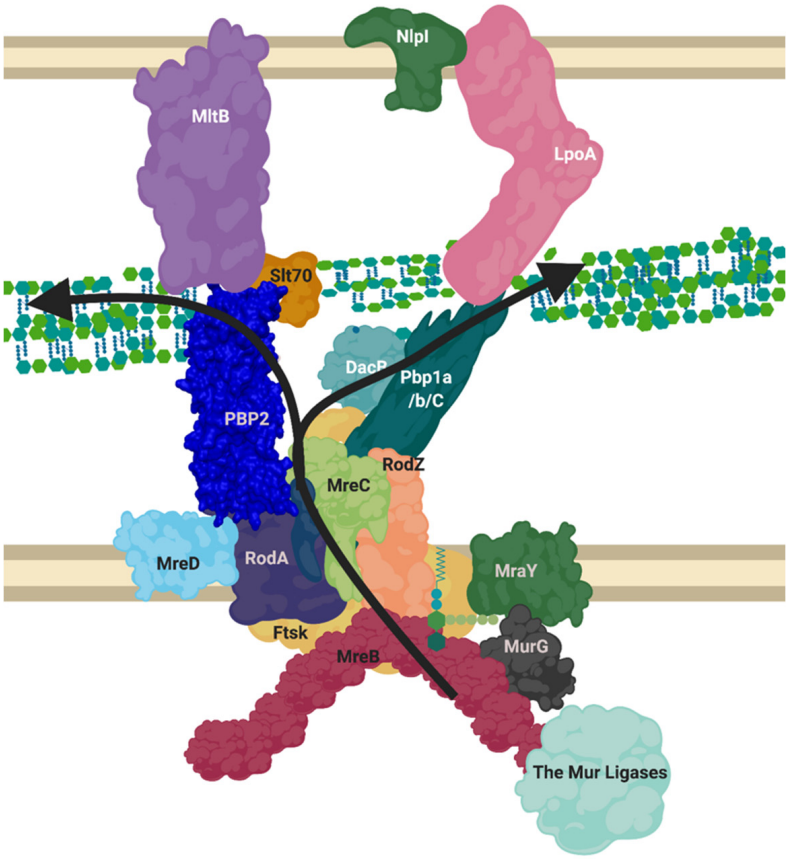


Figure 5. Proposed layout of the RodA orientated elongasome complex. A sketch of peptidoglycan insertion by a proposed formation of the elongasome complex. MreB and C sequester enzymes to the elongasome complex, including RodZ. MurG transforms Lipid I to Lipid II, MurJ/FtsW/RodA flip this into the periplasm. The PBP1a/b/c and/or PBP2-RodA complexes transglycosylate the lipid II into the nascent strand. DacB may remove the terminal D-Alanine from pentapeptide and a transpeptidation reaction occurs through catalysis from transpeptidases PBP2 or PBP1a/b/C. Lytic transglycosylases MltB or Slt70 open new nascent strands for modification as machinery moves forward. LpoA/B along with other regulatory mechanisms listed in Table A3 control PBP activities. This diagram disregards the dynamic nature of the SEDS complexes and MreB ratcheting. Figure created in BioRender.

The scaffolding and regulatory proteins of this elongasome, RodZ and MreC respectively, both interact with cytoplasmic MreB, as well as binding the transpeptidase PBP2 [48]. It has also been shown that MreC, anchored in the membrane with a single transmembrane helix, regulates the crosslinking transpeptidase activity of PBP2, and transglycosylation activity of RodA via interaction of its own periplasmic globular domain with the pedestal domain of PBP2. MreC may also have a role in binding to PBP1a among other components,

this is especially interesting as a recent paper shows that *P. aeruginosa* MreC forms large self-storage filaments within the periplasm to likely regulate MreC concentration in the membrane [48,49]. The integral membrane protein MreD has been shown to act as a co-ordinating partner to MreC in its interaction with RodA and PBP2, with an antagonising effect, however the interaction interface and the regulatory mechanism they perform itself is not yet known. Ribosomal studies suggest MreD levels are half that of MreC, indicating MreC's storage filaments are likely integral to proper regulation of this system among other possibilities [23].

The elongasome is transiently linked to the cytoplasmic MreB cytoskeleton of Gram-negative bacterial cells [14] (Figure 5). The "Brownian-ratcheting" mechanism of FtsZ (see above) could also apply to MreB interaction, considering the similarity of FtsZ and MreB as cytoskeletal protein homologues. This would suggest that the elongasome complex may instead move in and out of interaction with MreB, rather than remaining always associated [44]. This model would agree with the RodA-PBP2 complex moving along the cell circumferentially alongside MreB bi-modally active and inactive, at different speeds and with alternative partners [11].

Beyond the cytoskeletal interactors, regulation of this elongation apparatus has been shown to require the outer membrane regulatory lipoproteins LpoA and LpoB [25] (Figure 3). LpoA and LpoB span the periplasm to make contact with PBP1a [50] and PBP1b respectively and form synthetically lethal pairings upon genetic deletion, underlying their essential regulatory role [25]. LpoA stimulates the transpeptidase activity of PBP1a specifically, this turn upregulates PBP1a's glycosyltransferase activity and peptidoglycan production [51] and by contrast, LpoB has been shown to stimulate both PBP1b transglycosylase and transpeptidase activity [52]. Recent analysis of the kinetics of the related PBP1b-LpoB pairing required for cell division, shows that LpoB is an effective on/off kinetic switch for peptidoglycan transpeptidation by PBP1b [25,52]. Therefore, the elongasome contains multiple overlapping and seeming duplicate activities and interactions, but this almost certainly belies the complex network of interactions required for optimal peptidoglycan biosynthesis. One interpretation of this complexity is that the central RodA-PBP2 complex is required for the production of a peptidoglycan scaffold for elongation which the PBP1a-LpoA pairing (connecting inner membrane based synthesis with outer membrane control) then overlays with additional glycan stands and crosslinks, required to produce a complete layered structure [53].

8. NlpI Acts as a Facilitator of PBP Nucleation and Complex Interaction

The literature has shown peptidoglycan associated enzymes interact with a great deal of enzymatically inactive structural proteins (Figure 3). A recent paper, has shown there to be an outer membrane protein with a large number of protein-protein interactions, dominating our interaction networks called NlpI [54]. It is postulated to act as a scaffold for peptidoglycan associated proteins and is required for their formation and control. Microscale thermophoresis, pull-down and bacterial two-hybrid studies have shown that NlpI can form trimeric complexes with PBPs, for example, MepS-NlpI-DacA, MepS-NlpI-PBP7 and LpoA/PBP1a/NlpI among many others [54]. NlpI regulates a set of peptidoglycan hydrolases, as well as being able to form a trimer with PBP1A and LpoA. Its absence leads to increased vesicle creation [55] suggesting its importance to the cell envelope. Banzhaf et al. concluded NlpI may facilitate the interaction and/or change the composition of the peptidoglycan editing complexes, controlling the potentially harmful hydrolases and facilitating regulation of other proteins [54].

NlpI's dispersal around the cell indicates it is likely involved in many of the complexes responsible for creating peptidoglycan, including the divisome and elongasome, and possible intermediary complexes that likely exist between those in turn (Figure 6) [56]. As a result, its abundance across the entirety of the cell and regulatory ability suggests it may be part of the system of dynamic protein complex formation this review focuses upon (Figures 3, 4, 6 and 7) and is thus worth noting, however, its role beyond this is not well known [56].

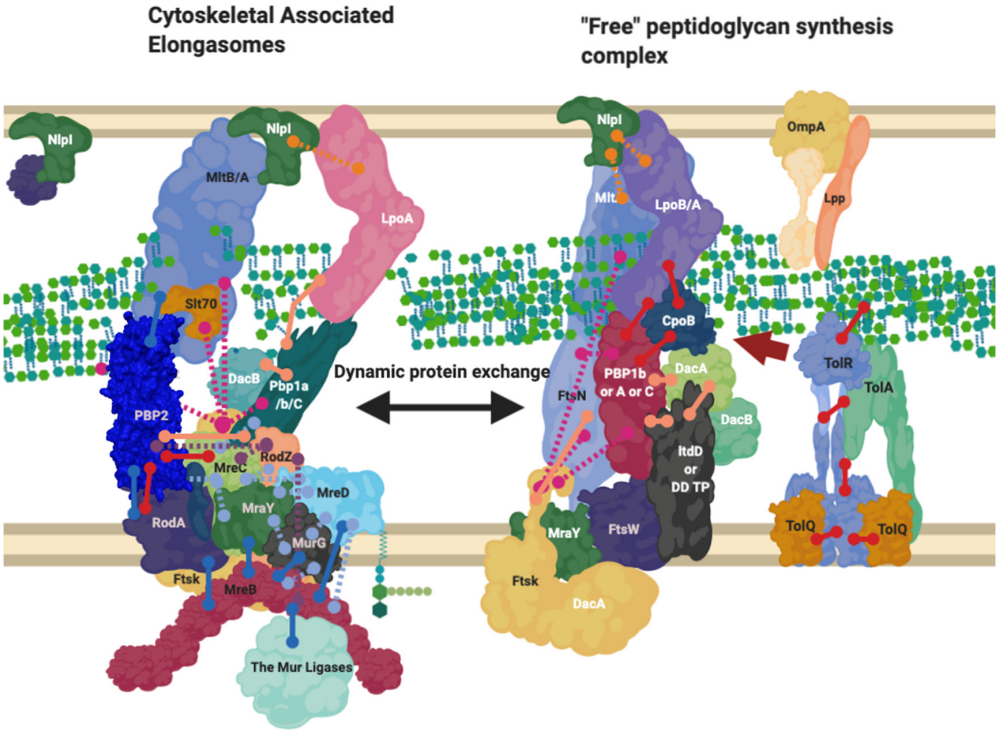


Figure 6. The Elongasome complex interactions. Interactions of the elongasome proteins. The elongasome is a series of complexes that are either attached to or not attached to the MreB filament. Localisation-dependent studies have shown the enzymes involved co-localise at the MreB filament [9,10], and each interaction is cited. This diagram is one of many possible configurations based on the current data, with the incorporation of the flippase not mentioned. Proteins are shared between complexes. Diagram rendered using BioRender. Connections: Purple [57], Red [9,23,31], Cyan [53], Orange [54], Pink [58], Peach [9,29,47,50,59,60], Blue [10,32,61–63].

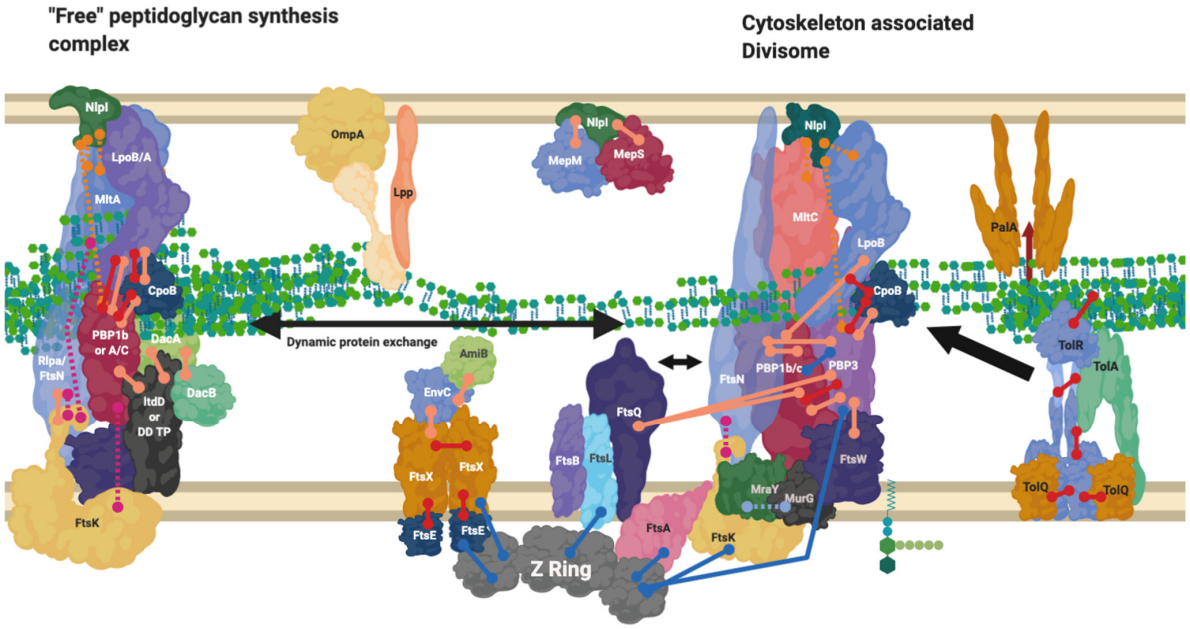


Figure 7. Divisome interaction network at the mature divisome. A diagram representing the interactions of the divisome proteins. Certain proteins within the divisome interact directly with the FtsZ filament. Localisation dependent studies have shown the enzymes co-localise at the FtsZ filament or in the pre-septal region (PIPs). Models suggest interactions are only brief [44]. This diagram is one of many possible configurations based on the current data. Diagram rendered using BioRender. Connections: Purple [57], Red [9,23,31] Cyan [53] Orange [54], Pink [58], Peach [9,29,47,50,59,60], Blue [10,32,61–63].

9. The Divisome Is a Series of Complexes Controlled by Cytoplasmic Events

The reasoning behind the complex series of interactions in Figures 3 and 4 can be more fully understood in the context of the cell cycle, as not all interactions must occur simultaneously, but rather on a cell stage basis. The divisome is responsible for the division of cells, it is a peptidoglycan modifying complex that has been studied for some time and exists as a series of complex protein-protein interactions (Figure 3), but these have been shown to occur at intermediate stages and be dynamic [3]. The divisome's function is similar to that of the elongasome's with analogous flippase, transglycosylase and transpeptidase partners, which are dependent upon a cascade of interactions [9,13,14,55,64]. The divisome proteins that modify peptidoglycan such as PBP3 and PBP1b are not always present with the divisomes central transglycosylase FtsW, as they change their cellular localisation dependent on the cell cycle and by their interactors (Figure 7) [9]. There is a higher concentration of peptidoglycan synthesis and hydrolysis enzymes at the septa during cell division, in a series of stages and cascades, suggesting a dynamic system much like the elongasome system, where protein composition changes over time as need and function changes [63].

Networks of interaction presented in Figures 3, 4 and 7 make clear the abundance of protein interactions possible. Division must account for osmotic conditions, cytoplasmic events, antibiotic challenge, and periplasmic protein complexes, whilst also maintaining the stability of cell envelope layers to prevent cell lysis, and finally allow septation. Recent reviews on the cascade of proteins and steps involved make the changes in the division complex over time clear [12,29,39].

10. Proteins Interchange between Complexes, and Complexes Interact

Throughout this review, it is mentioned that proteins can exist in more than one complex (Figure 4). Despite the notion that PBP2 and PBP1a are normally associated with the elongasomes as discussed above, they have also been found at the division site. A hypothesis involving a balance between the elongasome and the free-floating or unbound elongasome was investigated in the Gram-Positive, *Bacillus subtilis*, which found that PBP1a's homologue can move independently of MreC homologue or RodZ in the cell [64,65]. In the same study, the quantity of MreC and PBP1a also determined the cell width, suggesting this balance of two systems: one elongasome free of cytoskeletal proteins and RodZ, which can diffuse across the circumference of the cell to allow radial expansion, and one which interacts transiently with the cytoskeleton dominates and dominates the elongation process, in addition to morphology determination in *B. subtilis*, which could also be indicative of Gram-Negative systems (Figures 4 and 6) [64,65].

PBP1b/*mrcB* is a bifunctional glycosyltransferase and transpeptidase enzyme that interacts and plays part in the regulation of the divisome. It dominates Figures 3 and 4 as a node with high levels of interaction, beyond which is reasonable to exist at any one time simultaneously [9]. In contrast to PBP1a/*mrcA*, PBP1b/*mrcB* has been postulated to have division complex roles as well as a wandering role [10]. PBP1b and its partner activator LpoB are essential to peptidoglycan rebuilding in peptidoglycan-deficient spheroplasts of *E. coli* [33]. Their essentiality outside of division processes to create new peptidoglycan in spheroplasts suggests that PBP1b must play a major role in the creation of new peptidoglycan, which in wildtype cells (*E. coli*) is carried out 70–80% by the bifunctional PBPs such as PBP1b and PBP1a which have roles in the elongasome and divisome [10]. This dependence suggests either; the cytoskeleton-bound or free "elongasome" for cell growth including PBP2 and RodA involves PBP1b more than just transiently, or that more than the static model of the elongation machinery RodA-PBP2 exists and PBP1b has a separate role (Figure 6).

A "free" diffused PBP1b, and PBP2 have been observed independent of MreB/FtsZ systems by fluorescent localisation [10,11]. The interactions shown in Figures 3 and 4 suggest many possible complexes, that vary in their composition, position, and association with the cytoskeleton by PBP1b, PBP1a and PBP2.

The complexes that contain these interactive proteins may also interact. The elongation and division machinery share common protein components and interactors (Figures 3 and 4) with the elongation machinery associated "PBP2" even transiently localising to the Z-ring during cell division [63]. PBP2, a protein known to be integral to the elongation machine, has also been shown to interact with PBP3s division related role (*ftsI*), with PBP2 knockout studies revealing division defects. In addition to division and elongation related localisation, the peptidoglycan synthase proteins PBP2, PBP1a, PBP1b and FtsW have also been shown to localise diffusely around the cell, moving independently to the cytoskeletal-associated elongation and division complexes [9,10,12].

During the midstage of division, MreB and FtsZ appear to co-localise at the Z-ring whilst treadmilling [63]. It has been postulated that enzyme exchange of these proteins between the divisome and elongasome may occur through an interaction with the cytoskeletal components MreB and FtsZ [13,63]. This would support the "Brownian ratchet" model theory of cytoskeletal protein control, citing a transient interaction rather than permanent interaction of the PG machinery with FtsZ and MreB, allowing for the exchange of proteins between cytoskeletons more easily [44].

PBP1b and PBP2 localise to the septum adjacent to the Z-ring during division, and become delocalised from the septum in an *mreB* knockout strain [13]. Mutation of the FtsZ-interacting residue of MreB similarly delocalises PBP1b and PBP2 from the FtsZ rings, despite successful MreB and Z-ring formation. The unused Z-rings remain as "locked" stripes of unsuccessful division sites, and cells containing these Z-rings stripes become filamentous cells. These "locked" Z-rings fail to incorporate fluorescent single D-amino acid probes such as HADA denoting new peptidoglycan biosynthesis, and thus do not actively synthesise peptidoglycan, while the elongation enzymes along the rest of the cell remain functional and successfully incorporate HADA throughout the rest of the cell [13]. This may be due to the absence of the PBP1b-FtsN interaction which would normally interact transiently with MreB's PBP2, and transition to divisome interactions to relieve the FtsQLB inhibition of PBP3 [24,29,66], without this MreB-FtsZ transient reaction this inhibition remains in place. The cytoskeletal component amino acid knockouts described above, in conjunction with a known lethal PBP2 knockout phenotype, show the necessity of elongation enzymes such as PBP2 to also be required for division and highlights dynamic interchange between complexes [67].

11. Alternate Protein Complexes Exist, Containing 3-3 Crosslinking L, D Transpeptidases as an Alternative to 3-4 Crosslinking PBPs Important for Antibiotic Resistance

Figures 3 and 4 and the literature they represent indicates enzymes not yet confirmed to be integral machineries to be involved in interactions in multiple complexes. There is good evidence that during extended growth, osmotic cell stress and some instances of β-lactam challenge, 3-3 crosslinking increases in the peptidoglycan of many Gram-negative bacteria [68]. This form of crosslinking is catalysed by L,D-transpeptidases (Figure A2). The literature, and genetic predictions indicate these proteins interact with many other PBP related proteins (Figures 3 and 4).

The L, D-transpeptidase LdtD, has recently been shown to interact with peptidoglycan endopeptidase DacA and bifunctional synthase PBP1b by microscale thermophoresis [59]. Following PBP1b inhibition by β-lactams, LdtD compensates for the loss of 3-4 crosslinking by 3-3 crosslinking, enabling cell survival in the presence of β-lactams [69]. In this situation LdtD could compensate for part of PBP1b's normal bifunctional role as a transpeptidase (when in complex with the transglycosylase FtsW and PBP3) by replacing the PBP catalysed 3-4 transpeptidase activity with 3-3 crosslinking, using the PBP1b and FtsW transglycosylase glycan chain products as substrates. In this scenario the D, D-carboxypeptidase of DacA, shown to be essential for β-lactam resistance mediated by LdtD [69], modifies the pentapeptide by removal of the terminal amino acid to provide a suitable tetra peptide substrate to LdtD. This is necessary because LdtD requires a tetrapeptide as a substrate [69]. Depending upon the availability of suitable substrates and some environmental conditions,

PBP1b will also generate tetrapeptide products which could become substrates for the Ldt's [52,69]. It is likely this isolated complex is only one of many complexes incorporating the non-canonical peptidoglycan crosslinkers L, D-transpeptidases (Figure 8).

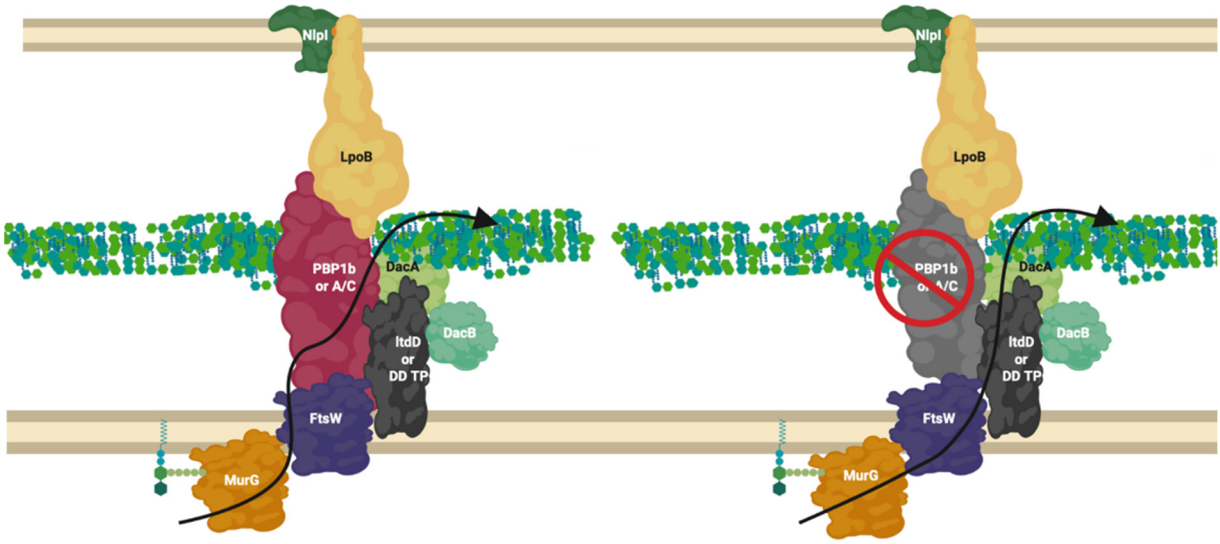


Figure 8. LdtD in complex with PBP1b and DacA. (**a**) PBP1b active complex performs 3-4 crosslinking; (**b**) On β-lactam challenge PBP1b activity is reduced, allowing LtdD, in complex with DacA and DacB, to take over crosslinking and allow cell viability [59,69]. The complex represented above is one of many possible complexes. The flippase is represented by FtsW due to its potential bifunctional role for simplification but likely involves MurJ, this is one of many possible configurations. MurG ligase has been shown to interact with FtsW [60,66].

12. A "Shapeosome" Complex Synthesises Peptidoglycan in Curved Gram-Negatives

Beyond the simple models presented as elongation or division mechanisms, many species of Gram-negative bacteria are more morphologically complex than just rods or spheres [70] and the complex of proteins presented for *E. coli* (Figure 4). These species require the peptidoglycan machinery to be altered compared to these classical exemplar species. *Campylobacter jejuni* and *Helicobacter pylori* have been shown to contain hydrolytic L, D carboxypeptidase proteins essential to cell curvature [71] in addition to a new cytoskeletal component analogue to MreB, CcmA. The most well studied of these systems is *Helicobacter pylori's* "Shapeosome" and the conserved shape determinant (Csd) protein family. Knockouts of Csd6, CcmA and Csd5 all lead to curvature loss, along with peptidoglycan peptidases Csd1, 2, 3, 4 and 6 [18,72] (Figure 9).

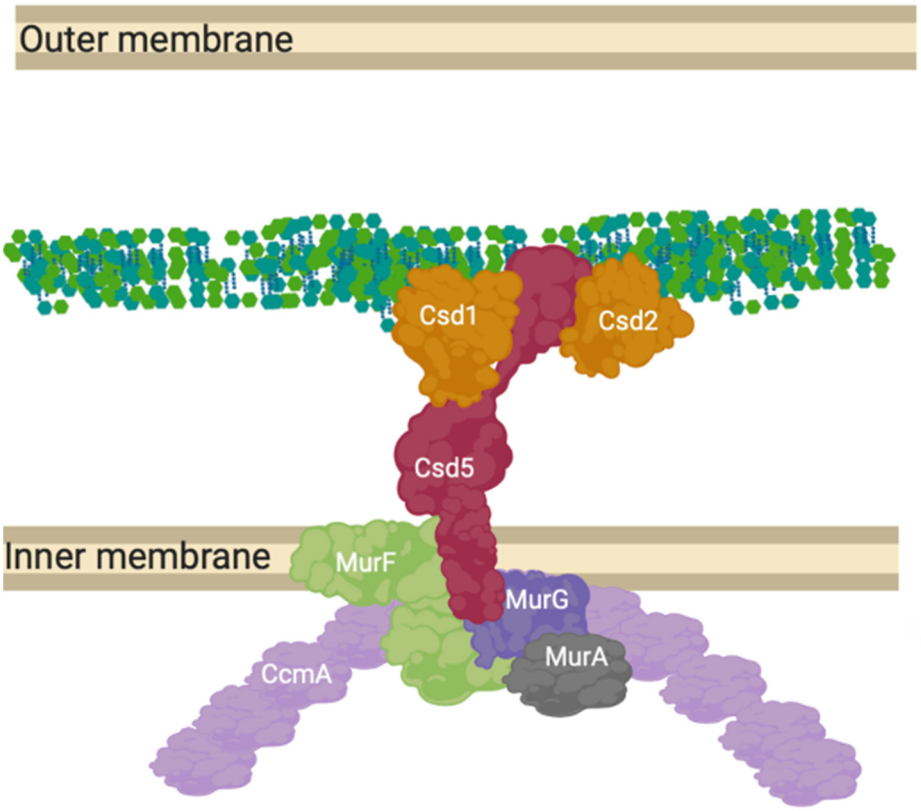


Figure 9. CcmA curvature promoting complex in *Helicobacter pylori*. The prospective anchored "Shapeosome" of *Helicobacter pylori*. The Shapeosome is a non-canonical peptidoglycan synthesis complex that facilitates cell shape of curved bacteria, associated with a cytoskeletal component like the elongasome and divisome. The connection to a cytoskeletal component again suggests a convergent and re-occurring model of shape-determining cell wall modification by complex formation. Csd5 binds peptidoglycan by its SH3 domains and interacts with synthesis related proteins MurG, MurA and MurF, as well as hydrolases Csd1 and 2 [18,72].

13. Unrelated Cell Envelope Proteins must Affect Peptidoglycan-Membrane Linkage

Whilst peptidoglycan is generally regarded as being composed of intramolecular crosslinks [2], in some organisms, L, D transpeptidase enzymes (LDTPs) can catalyse the formation of intermolecular cross links used to attach it to other macromolecular structures [73]. It was shown in the 1970s that at least one-third of Braun's lipoprotein (Lpp), an outer membrane protein, is bound to peptidoglycan [74] (Figure 1). In Figures 3 and 4, an abundance of outer membrane interactors and regulators, are shown to be part of elongation, and division complexes. It is has been known for some time that in species with and without Lpp, the outer membrane protein OmpA interacts non-covalently with peptidoglycan, but more recently, the literature has shown multiple OMPs across species to be connected covalently to the peptidoglycan by LDTPs [75]. Specifically for example in *C. burnetii* the L, D-transpeptidase ldt2 is required for covalent attachment of OMPs BbpA and BbpB to peptidoglycan [76]. This creates a static covalent link between the outer membrane, peptidoglycan and inner membrane proteins such as Tol machinery across species [62].

Further still, bacterial periplasmic complexes such as pili, transport systems and flagella penetrate through the peptidoglycan layers (Figure 1) and are often able to transport proteins from the inner to outer membrane despite this linkage. This activity requires dynamic pores in the peptidoglycan layer [20,77]. Observed movement of large proteins laterally across these fixed peptidoglycan layers, linked to outer and inner membranes may involve cleaving of the peptidoglycan using lytic transglycosylases to facilitate this movement [16,78,79], in addition to L, D carboxypeptidases, for Lpp release [80]. These

protein-peptidoglycan interactions must all be regulated to avoid cell lysis, as shown by multiple periplasmic proteins that ensure the cell envelope remains structurally stable during the cross-periplasmic complex movement. These local changes in cell wall structure due to large complex movements present the cell envelope and peptidoglycan as a more multi-ordered structure than a simple mesh. It must require uncrosslinking and regulatory mechanisms for growth facilitation, and cell envelope stabilisation.

14. Review Summary

The peptidoglycan polymer's complex and essential role to Gram-Negative bacterial cells requires an intricate set of proteins within the periplasm; to maintain its role in response to growth, during division and to ensure a stabilising permeable barrier is maintained in tandem with the inner and outer membrane. The literature has shown this takes place through a series of protein complexes, and this is reaffirmed in predictive genetic and experimental interactions presented (Figures 3 and 4). However, the full picture of experiments, when investigating the roles of each protein have shown that these complex interactions are not static in composition, but are instead part of a web of interactions that allow many variant complexes to be in dynamic equilibrium depending on cell growth stage and need. This model is not yet complete.

Models have been postulated of wandering and cytoskeletal-associated complexes such as the elongasomes and divisomes that create and modify peptidoglycan dependent on growth needs (Figures 6 and 7). Alternative complexes have also been shown to exist for the antibiotic insusceptible L, D-transpeptidase enzymes which can allow crosslinking of peptidoglycan in the absence of the antibiotic susceptible PBPs (Figure 8). These must all occur in the context of structures that cross the periplasm and connect the inner membrane and outer membrane in partnership with other processes [80,81].

This model of large protein complexes evolved to allow for peptidoglycan modification dynamically across a growth cycle and repeats convergently in other species, even among Archaea. Peptidoglycan modification systems, such as the shape determining complex oriented by the cytoskeletal protein CcmA in *Caulobacter* sp. (Figure 9) exist as convergent versions of the *E. coli* MreB and FtsZ based models presented in this review. The cytoskeletal component of some of these dynamic complexes across species, (FtsZ and MreB) treadmill along the circumference of the cell and have been shown to exchange protein partners during their interactions, and cytoskeletal or regulator absence/inhibition leads to growth defects. This evidence among others, shows an exchange of proteins which facilitate a change of complex composition over time by the associated machineries. Sometimes these complex changes are driven by specific cytoplasmic events and cascades, such as those that control the divisome.

However not all modification relies upon these cytoskeletons, as shown by PBP1b wondering motion across the cell [56]. Indeed, a single protein could be required for multiple functions and complexes that exist at once (Figure 3), therefore these multiple protein localisations are in part controlled by affinity to the cytoskeletal proteins or outer membrane proteins anchored such as NlpI. This allows for fine control of complex composition in addition to regulation by protein affinity to local substrate [30].

The peptidoglycan research and anti-microbrial resistance field has come to place importance on specific protein structure, and singular relationships with inhibitory/activator proteins in future antibiotic design. Our meta-analysis has shown the full picture so far likely extends beyond the crystallised complexes and static complexes, revealing a great deal of flexibility, but also indicating the importance of specific nodal proteins in peptidoglycan synthesis. Research into macro-regulation of the complicated cell envelope complexes showcased will be an important step in the creation of new drugs that can overcome known mechanism of antibiotic bypass by protein exchange, but also postulate new methods for peptidoglycan and cell envelope disruption. Viewing these proteins in a systems context will be an important step in combatting resistance to antibiotics in vivo.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms222312831/s1.

Author Contributions: C.L.B.G. was the concept originator and main writer throughout. Supervisor D.I.R. and M.B. provided advice, reworded sections, and reviewed the piece to ensure quality throughout. H.N., F.N.G., K.S. and N.B. confirmed and made additions in their specialisations in addition to editorial support, with significance reflected in order of names. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the BBSRC through studentships of F.N.G., C.L.B.G. and N.B. [BB/M01116X/1], ANTRUK[ANTRUK_SRG_05-2021] funding supported C.L.B.G, the MRC supported a studentship to [MR/J003964/1] to K.S. and a Collaborative Postgraduate award between the University of Warwick and Diamond Light Source [STU0212] for funding H.N, This work was supported by a UKRI Future Leaders Fellowship [MR/V027204/1] for the support of M.B.

Data Availability Statement: The interaction network of proteins, with DOI citations at each interaction pair can be found in the paper's supplementary data file. The predicted gene network used for Figure 4 is available at STRING's website using the permalink: https://version-11-5.string-db.org/cgi/network?taskId=bIzLkBRoqjLb&sessionId=bBi0rwtoih3p (Last accessed on 4 November 2021).

Acknowledgments: We thank Andrew Lovering, Tyler Baverstock and Alexander Egan for editorial suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1. Mur Ligase Pathway

Fructose-6-phosphate is converted by four successive enzyme activities to uridine 5′-diphosphate-N-acetylglucosamine (UDP-GlcNAc). This is catalysed by GlmS, GlmM and GlmU (bifunctional enzyme) UDP-MurNAc (5′-diphosphate *N*-acetylmuramic acid) is formed from UDP-GlcNAc using Mur ligases MurA and MurB. This results in a sugar moiety ready for pentapeptide addition [78].

A pentapeptide stem is then appended to the D-lactoyl carboxyl group of UDP-MurNAc by sequential addition of peptides by MurC-F: L-Ala (MurC), D-glutamic acid (MurD), meso-diaminopimelic acid (m-DAP) (MurE), dipeptide D-Ala-D-Ala (MurF), with D-Glu and m-DAP being synthesised from their L- or L,L-stereoisomers by MurI and DapF respectively, and D-ala-D-Ala being produced from L-Ala by alanine racemases and D-Ala-D-Ala ligase [78].

The UDP-MurNAc 5P produced by these reactions is then transferred to an undecaprenol, a membrane-spanning lipid, yielding undecaprenyl diphospho MurNAc 5P (Lipid I), in a reaction catalysed by MraY at the inner membrane. Thereafter, MurG transfers a GlcNAc sugar moiety from UDP-GlcNAc to Lipid I, producing undecaprenyl diphospho MurNAc GlcNAc 5P (Lipid II) [66,78].

Appendix A.2. After Initial Synthesis, Peptidoglycan Is Modified

The modification of peptidoglycan extends far beyond the model in Figure 1, and continues after de novo peptidoglycan insertion, by a series of proteins and protein complexes referred to as PBPs. During synthesis a lytic transglycosylase (Figure A1) separates the existing strands of peptidoglycan to make space for new synthesis by cleaving the links between sugars [79,82], the glycan strand does not continue indefinitely and are typically between 7 and 32 sugars in length [83]. New peptidoglycan must also be attached to the outer membrane by L, D transpeptidase action through linking peptidoglycan to outer membrane proteins like Lpp (Figure A2), roughly once every 100 Å to maintain cell envelope stability [69,80].

During division, the peptidoglycan crosslinks are continually broken and remade to relieve overall cell wall stress and facilitate growth (Figure A1). 3-4 crosslinks being predominant when the cell is not in stress and early in growth, whereas 3-3 crosslinking is

found in cells increasingly during the stationary phase, or following antibiotic exposure and osmotic shock [84] (Figure A2).

A shift from 3-4 to 3-3 crosslinking often also occurs during cell growth. Fluorescent D-amino acids (FDAA) have been used to label transpeptidase activity through the ability of PBPs to exchange amino acids, revealing that during stationary phase and growth, the entire peptidoglycan is "lit up" by incorporation of new FDAA indicating new modifications are being made throughout growth [85].

Bacterial cell elongation and division requires the peptidoglycan layer to be constantly modified and cleaved to allow for growth. The attachment of protein partners to the peptidoglycan layer and the peptidoglycan recycling process additionally requires peptidoglycan cleavage. The cleavage and modifications of peptidoglycan varies across species and can be broadly split into two classes of enzymatic action: hydrolase and transferase [12,82] (Figures A1 and A2).

Hydrolases carry out a range of lytic modifications to peptidoglycan, including cleavage of the peptide stem at the glycosidic bond between glycan molecules, and the removal of acetyl groups (a lysozyme resistance factor in pathogenic strains) [85] (Figure A1). The hydrolases so far characterised are dispersed throughout the cell periplasm at the lateral cell wall or the division plane [12].

Hydrolases are controlled by regulatory proteins and each hydrolase has their own distinct role within the cell and the cell's complexes [86,87] (Table A1).

Peptidoglycan is also polymerised and modified by a series of crosslinking enzymes such as the penicillin-binding proteins (PBPs) or L, D transpeptidases at different points in the bacterial cell cycle as well as after stress events. The modification sites shown in Figure A2 are the points of peptidoglycan crosslinking and attachment known in Gram-negative bacterial species [9,47,88].

Their individual activities are elaborated in Table A2.

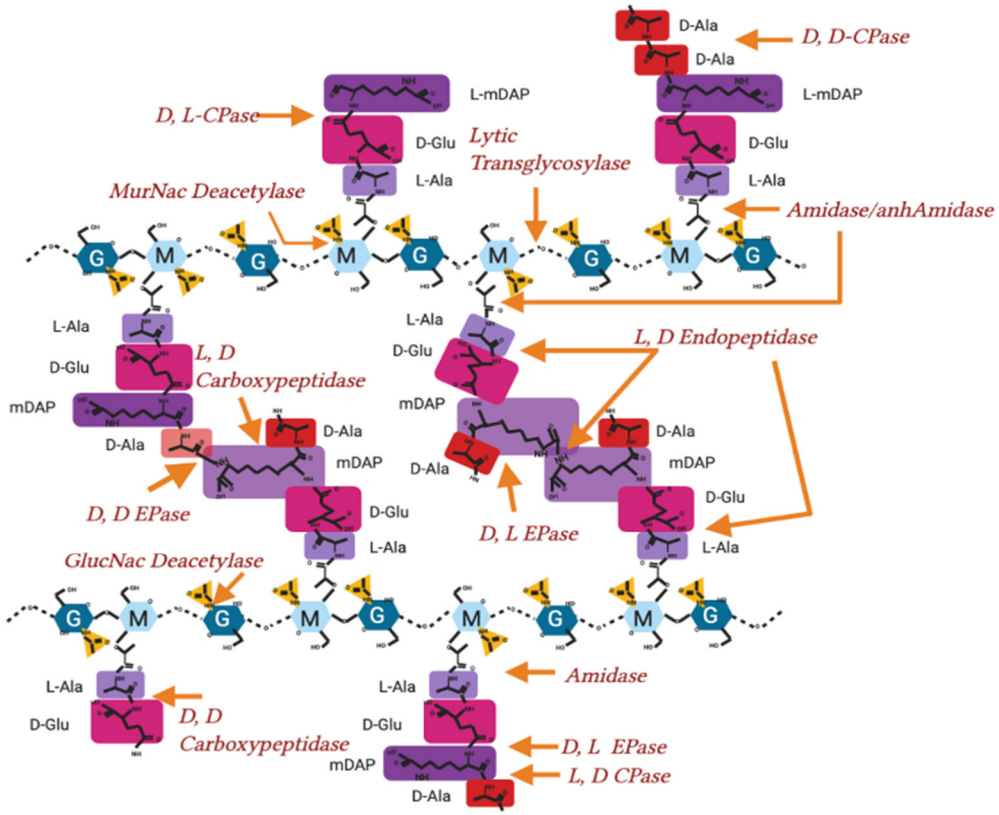


Figure A1. Hydrolase activity on peptidoglycan in Gram-negative bacteria. Peptidoglycan molecular structure, with sites of hydrolysis and enzyme nomenclature labelled. EPase-Endopeptidase, CPase-Carboxypeptidase.

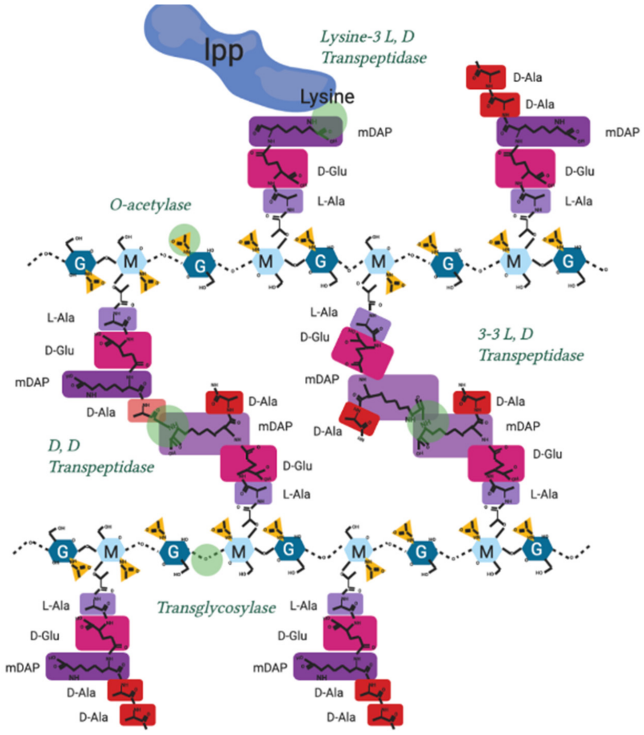


Figure A2. Peptidoglycan crosslinking sites and activity in Gram-negatives. Peptidoglycan molecular structure, with sites of crosslinking actions and enzyme nomenclature labelled.

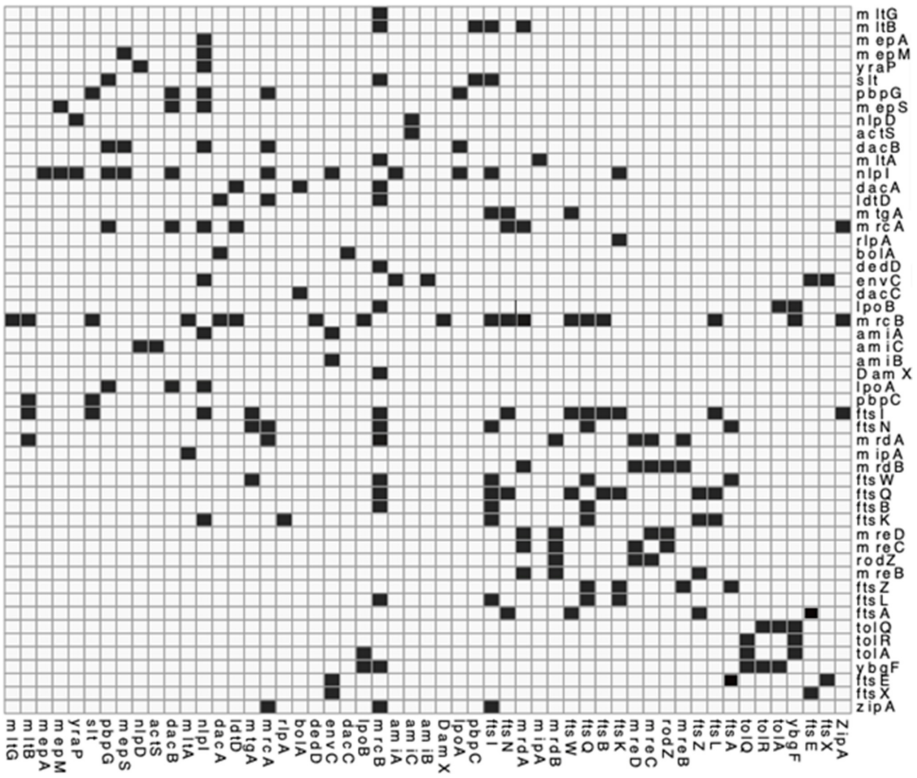


Figure A3. Interaction matrix of PG related proteins. genes or proteins known to interact and confirmed by the literature. Black = confirmed interaction, White = unknown/self/no interaction. Please find attached pg_network.csv file for DOI matrix of confirmed interactions for above.

Table A1. Characterised Hydrolases in Gram-negative bacteria.

Peptidoglycan Degradation/Hydrolases			
Function	**Enzymatic Action**	**Known Genes/Protein**	**References**
D,D Carboxypeptidases	D-Ala D-Ala Cleavage 4-5	dacA, yfeW, dacC, dacD, vanY, ampH, Csd3 *	[69,89–91]
MurNac de-Acetylase	Deacetylation of N-acetyl Muramic acid	pgdA	[92]
GlutNac de-acetylase	Deacetylation of N-acetyl Glucosamine		
Amidase	Cleavage of peptide stem from Glycan strand	amiA, amiB, amiC, amiD, ampD, mpaA	[88,90,93,94]
Lytic Transglycocylase	Breaking Glycan strand at GlucNac-MurNac (endo)	Slt, MltA, MltB, MltC, MtD, MltE, PilT, traB, virB1, rlpA, MltG	[16,58,79,82,83,86,95,96]
	Breaking Glycan strand at GlucNac-MurNac(exo)	NagZ	[97]
L, D Carboxy/Endopeptidase	mDAP mDAP cleavage 3-3	mepA	[98]
	mDAP-Lpp Cleavage	YafK/LdtF	[80]
	mDAP D-Ala cleavage 3-4	pgp2 *, csd6 *	[71,91]
	mDAP-D Glucosamine cleave 3-2	csd4 *	[72]
D, D Endopeptidase	Cleavage of D-Ala-mDAP crosslink 3-4	dacB, pbpG, MepS, MepM, PBP7, MepH	[31,99]

* Not present in *Escherichia coli* MG1655 genome, but present in other species.

Table A2. Characterised peptidoglycan synthases in Gram-negative bacteria.

Function	**Enzymatic Action**	**Known Genes/Protein (*E. coli*)**	**References**
D, D Transpeptidase and Transglycocylase	Adds lipid II to nascent strand and crosslinks into existing PG	mrcA/PBP1a	[29,100]
		mrcB/PBP1b	[78]
D, D Transpeptidase	Crosslinks nascent strand into existing peptidoglycan	mrdA/PBP2, FtsI/PBP3	[11,101]
Transglycosylase	Adds lipid II to nascent strand	mtgA, rodA, ftsW	[23,102,103]
Flippase	Flips Lipid II to periplasm	murJ	[104]
L, D Transpeptidases	Peptidoglycan Brauns lipoprotein crosslinkers	LdtA/ErfK,	[105]
		YbiS/LdtB	
		Ycfs/LdtC	
	Peptidoglycan 3mDAP-3mDAP crosslinkers	YnhG/LdtE	
		YcbB/LdtD	
O-acetylation	O-acetylates nam	oatA *	[2,92]
		adr	
		pacA	

* Not present in *Escherichia coli* MG1655 genome, but present in other species.

Table A3. Peptidoglycan associated and PBP regulatory proteins.

Function	Known Genes/Protein (*E. coli*)	Reference
Moderate class A PBP activity	LpoA, LpoB	[9]
Alter interactor ability	CpoB	[9]
Bind OM with peptidoglycan	OmpA	[106]
Moderate OM linkage with peptidoglycan	lpp	[73]
Periskeletal elongasome component, moderates PBP2 activity	MreC	[23]
Periskeletal elongasome component, moderate PBP2 activity	MreD	
Treadmilling Cytoskeletal elongasome component	MreB	[15]
Elongasome staple component	RodZ	[57]
SPOR domain containing proteins, protein interaction	Rlpa, FtsN, DamX, SpoX	[61,85,107]
Hydrolase binding activity	NlpI	[55]
	ActS	[108]
	NlpD	[109]
	EnvC	[93]
Inner membrane peptidoglycan moderation	TolA, TolR, TolQ, palA, palB	[9,109]
Division moderation and EnvC control	FtsX, FtsE	[110]
Treadmilling cytoskeletal component for divisome	FtsZ	[44,56]
Division moderation	FtsA	[56]
Helicase and PBP interactor	FtsK	[58]
FtsN interactor and division start	FtsB, FtsL, FtsQ	[25]
FtsZ interactor and PIPs mediator	ZipA	[39]

References

1. Egan, A.J.; Cleverley, R.M.; Peters, K.; Lewis, R.J.; Vollmer, W. Regulation of bacterial cell wall growth. *FEBS J.* **2017**, *284*, 851–867. [CrossRef]
2. Vollmer, W. Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol. Rev.* **2008**, *32*, 287–306. [CrossRef] [PubMed]
3. Egan, A.J.; Errington, J.; Vollmer, W. Regulation of peptidoglycan synthesis and remodelling. *Nat. Rev. Microbiol.* **2020**, *18*, 446–460. [CrossRef] [PubMed]
4. Barreteau, H.; Kovač, A.; Boniface, A.; Sova, M.; Gobec, S.; Blanot, D. Cytoplasmic steps of peptidoglycan biosynthesis. *FEMS Microbiol. Rev.* **2008**, *32*, 168–207. [CrossRef] [PubMed]
5. Bernal-Cabas, M.; Ayala, J.A.; Raivio, T.L. The Cpx envelope stress response modifies peptidoglycan cross-linking via the L,D-transpeptidase LdtD and the novel protein YgaU. *J. Bacteriol.* **2015**, *197*, 603–614. [CrossRef] [PubMed]
6. Pazos, M.; Peters, K.; Vollmer, W. Robust peptidoglycan growth by dynamic and variable multi-protein complexes. *Curr. Opin. Microbiol.* **2017**, *36*, 55–61. [CrossRef]
7. Mueller, E.A.; Egan, A.J.; Breukink, E.; Vollmer, W.; Levin, P.A. Plasticity of *Escherichia coli* cell wall metabolism promotes fitness and antibiotic resistance across environmental conditions. *eLife* **2019**, *8*, e40754. [CrossRef] [PubMed]
8. Lambert, C.; Chang, C.Y.; Capeness, M.J.; Sockett, R.E. The first bite—Profiling the predatosome in the bacterial pathogen Bdellovibrio. *PLoS ONE* **2010**, *5*, e8599. [CrossRef] [PubMed]
9. Gray, A.N.; Egan, A.J.F.; van't Veer, I.L.; Verheul, J.; Colavin, A.; Koumoutsi, A.; Biboy, J.; Altelaar, M.A.; Damen, M.J.; Huang, K.C.; et al. Coordination of peptidoglycan synthesis and outer membrane constriction during *Escherichia coli* cell division. *eLife* **2015**, *4*, e07118. [CrossRef] [PubMed]
10. Cho, H.; Wivagg, C.N.; Kapoor, M.; Barry, Z.; Rohs, P.D.; Suh, H.; Marto, J.A.; Garner, E.C.; Bernhardt, T.G. Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-Autonomously. *Nat. Microbiol.* **2016**, *1*, 16172. [CrossRef] [PubMed]
11. Wollrab, E.; Özbaykal, G.; Vigouroux, A.; Cordier, B.; Simon, F.; Chaze, T.; Matondo, M.; van Teeffelen, S. Transpeptidase PBP2 governs initial localization and activity of major cell-wall synthesis machinery in *Escherichia coli*. *bioRxiv* **2019**, 716407. [CrossRef]

12. Vischer, N.O.; Verheul, J.; Postma, M.; van den Berg van Saparoea, B.; Galli, E.; Natale, P.; Gerdes, K.; Luirink, J.; Vollmer, W.; Vicente, M.; et al. Cell age dependent concentration of *Escherichia coli* divisome proteins analyzed with ImageJ and ObjectJ. *Front. Microbiol.* **2015**, *6*, 586. [CrossRef] [PubMed]
13. Van der Ploeg, R.; Verheul, J.; Vischer, N.O.; Alexeeva, S.; Hoogendoorn, E.; Postma, M.; Banzhaf, M.; Vollmer, W.; Den Blaauwen, T. Colocalization and interaction between elongasome and divisome during a preparative cell division phase in *Escherichia coli*. *Mol. Microbiol.* **2013**, *87*, 1074–1087. [CrossRef]
14. van der Ploeg, R.; Goudelis, S.T.; den Blaauwen, T. Validation of FRET assay for the screening of growth inhibitors of *Escherichia coli* reveals elongasome assembly dynamics. *Int. J. Mol. Sci.* **2015**, *16*, 17637–17654. [CrossRef] [PubMed]
15. Daniel, R.A.; Errington, J. Control of cell morphogenesis in bacteria: Two distinct ways to make a rod-shaped cell. *Cell* **2003**, *113*, 767–776. [CrossRef]
16. Roure, S.; Bonis, M.; Chaput, C.; Ecobichon, C.; Mattox, A.; Barrière, C.; Geldmacher, N.; Guadagnini, S.; Schmitt, C.; Prévost, M.C.; et al. Peptidoglycan maturation enzymes affect flagellar functionality in bacteria. *Mol. Microbiol.* **2012**, *86*, 845–856. [CrossRef]
17. Sycuro, L.K.; Pincus, Z.; Gutierrez, K.D.; Biboy, J.; Stern, C.A.; Vollmer, W.; Salama, N.R. Peptidoglycan Crosslinking Relaxation Promotes Helicobacter pylori's Helical Shape and Stomach Colonization. *Cell* **2010**, *141*, 822–833. [CrossRef] [PubMed]
18. Blair, K. Exploring Mechanisms of Cell Shape Control in Helicobacter pylori. Ph.D Thesis, University of Washington, Seattle, WA, USA, 2018.
19. Turner, R.D.; Vollmer, W.; Foster, S.J. Different walls for rods and balls: The diversity of peptidoglycan. *Mol. Microbiol.* **2014**, *91*, 862–874. [CrossRef]
20. Turner, R.D.; Mesnage, S.; Hobbs, J.K.; Foster, S.J. Molecular imaging of glycan chains couples cell-wall polysaccharide architecture to bacterial cell morphology. *Nat. Commun.* **2018**, *9*, 1263. [CrossRef] [PubMed]
21. Yao, X.; Jericho, M.; Pink, D.; Beveridge, T. Thickness and elasticity of gram-negative murein sacculi measured by atomic force microscopy. *J. Bacteriol.* **1999**, *181*, 6865–6875. [CrossRef] [PubMed]
22. Vollmer, W.; Blanot, D.; De Pedro, M.A. Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **2008**, *32*, 149–167. [CrossRef] [PubMed]
23. Liu, X.; Biboy, J.; Vollmer, W.; den Blaauwen, T. MreC and MreD balance the interaction between the elongasome proteins PBP2 and RodA. *PLoS Genet.* **2020**, *16*, e1009276. [CrossRef] [PubMed]
24. den Blaauwen, T.; Luirink, J. Checks and Balances in Bacterial Cell Division. *MBio* **2019**, *10*, e00149-19. [CrossRef] [PubMed]
25. Typas, A.; Banzhaf, M.; van den Berg van Saparoea, B.; Verheul, J.; Biboy, J.; Nichols, R.J.; Zietek, M.; Beilharz, K.; Kannenberg, K.; von Rechenberg, M.; et al. Regulation of Peptidoglycan Synthesis by Outer-Membrane Proteins. *Cell* **2010**, *143*, 1097–1109. [CrossRef]
26. Moynihan, P.J.; Clarke, A.J. O-Acetylated peptidoglycan: Controlling the activity of bacterial autolysins and lytic enzymes of innate immune systems. *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1655–1659. [CrossRef] [PubMed]
27. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45*, D362–D368. [CrossRef]
28. Zhang, B.; Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **2005**, *4*, 17. [CrossRef]
29. Boes, A.; Olatunji, S.; Breukink, E.; Terrak, M. Regulation of the Peptidoglycan Polymerase Activity of PBP1b by Antagonist Actions of the Core Divisome Proteins FtsBLQ and FtsN. *MBio* **2019**, *10*, e01912-18. [CrossRef] [PubMed]
30. Rohs, P.D.; Buss, J.; Sim, S.I.; Squyres, G.R.; Srisuknimit, V.; Smith, M.; Cho, H.; Sjodt, M.; Kruse, A.C.; Garner, E.C.; et al. A central role for PBP2 in the activation of peptidoglycan polymerization by the bacterial cell elongation machinery. *PLoS Genet.* **2018**, *14*, e1007726. [CrossRef] [PubMed]
31. Sjodt, M.; Rohs, P.D.; Gilman, M.S.; Erlandson, S.C.; Zheng, S.; Green, A.G.; Brock, K.P.; Taguchi, A.; Kahne, D.; Walker, S.; et al. Structural coordination of polymerization and crosslinking by a SEDS–bPBP peptidoglycan synthase complex. *Nat. Microbiol.* **2020**, *5*, 813–820. [CrossRef]
32. Meeske, A.J.; Riley, E.P.; Robins, W.P.; Uehara, T.; Mekalanos, J.J.; Kahne, D.; Walker, S.; Kruse, A.C.; Bernhardt, T.G.; Rudner, D.Z. SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **2016**, *537*, 634–638. [CrossRef] [PubMed]
33. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef] [PubMed]
34. Ranjit, D.K.; Jorgenson, M.A.; Young, K.D. PBP1B glycosyltransferase and transpeptidase activities play different essential roles during the de novo regeneration of rod morphology in *Escherichia coli*. *J. Bacteriol.* **2017**, *199*, e00612-16. [CrossRef]
35. Müller, P.; Ewers, C.; Bertsche, U.; Anstett, M.; Kallis, T.; Breukink, E.; Fraipont, C.; Terrak, M.; Nguyen-Distèche, M.; Vollmer, W. The essential cell division protein FtsN interacts with the murein (peptidoglycan) synthase PBP1B in *Escherichia coli*. *J. Biol. Chem.* **2007**, *282*, 36394–36402. [CrossRef] [PubMed]
36. Yang, X.; McQuillen, R.; Lyu, Z.; Phillips-Mason, P.; De La Cruz, A.; McCausland, J.W.; Liang, H.; DeMeester, K.E.; Santiago, C.C.; Grimes, C.L.; et al. A two-track model for the spatiotemporal coordination of bacterial septal cell wall synthesis revealed by single-molecule imaging of FtsW. *Nat. Microbiol.* **2021**, *6*, 584–593. [CrossRef]

37. Kim, S.Y.; Gitai, Z.; Kinkhabwala, A.; Shapiro, L.; Moerner, W.E. Single molecules of the bacterial actin MreB undergo directed treadmilling motion in Caulobacter crescentus. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10929–10934. [CrossRef] [PubMed]
38. Adams, D.W.; Errington, J. Bacterial cell division: Assembly, maintenance and disassembly of the Z ring. *Nat. Rev. Microbiol.* **2009**, *7*, 642–653. [CrossRef]
39. Lutkenhaus, J.; Pichoff, S.; Du, S. Bacterial cytokinesis: From Z ring to divisome. *Cytoskeleton* **2012**, *69*, 778–790. [CrossRef] [PubMed]
40. Potluri, L.P.; Kannan, S.; Young, K.D. ZipA is required for FtsZ-dependent preseptal peptidoglycan synthesis prior to invagination during cell division. *J. Bacteriol.* **2012**, *194*, 5334–5342. [CrossRef]
41. Wang, J.; Galgoci, A.; Kodali, S.; Herath, K.B.; Jayasuriya, H.; Dorso, K.; Vicente, F.; González, A.; Cully, D.; Bramhill, D.; et al. Discovery of a Small Molecule that Inhibits Cell Division by Blocking FtsZ, a Novel Therapeutic Target of Antibiotics. *J. Biol. Chem.* **2003**, *278*, 44424–44428. [CrossRef] [PubMed]
42. Liao, Y.; Ithurbide, S.; Löwe, J.; Duggin, I.G. Two FtsZ proteins orchestrate archaeal cell division through distinct functions in ring assembly and constriction. *bioRxiv* **2021**. [CrossRef]
43. Van Teeffelen, S.; Gitai, Z. Rotate into shape: MreB and bacterial morphogenesis. *EMBO J.* **2011**, *30*, 4856–4857. [CrossRef] [PubMed]
44. Bean, G.J.; Flickinger, S.T.; Westler, W.M.; McCully, M.E.; Sept, D.; Weibel, D.B.; Amann, K.J. A22 disrupts the bacterial actin cytoskeleton by directly binding and inducing a low-affinity state in MreB. *Biochemistry* **2009**, *48*, 4852–4857. [CrossRef]
45. McCausland, J.W.; Yang, X.; Lyu, Z.; Söderström, B.; Xiao, J.; Liu, J. Treadmilling FtsZ polymers drive the directional movement of sPG-synthesis enzymes via Brownian ratchet mechanism. *bioRxiv* **2019**. [CrossRef]
46. Contreras-Martel, C.; Martins, A.; Ecobichon, C.; Trindade, D.M.; Matteï, P.J.; Hicham, S.; Hardouin, P.; El Ghachi, M.; Boneca, I.G.; Dessen, A. Molecular architecture of the PBP2-MreC core bacterial cell wall synthesis complex. *Nat. Commun.* **2017**, *8*, 776. [CrossRef] [PubMed]
47. Banzhaf, M.; van den Berg van Saparoea, B.; Terrak, M.; Fraipont, C.; Egan, A.; Philippe, J.; Zapun, A.; Breukink, E.; Nguyen-Distèche, M.; den Blaauwen, T.; et al. Cooperativity of peptidoglycan synthases active in bacterial cell elongation. *Mol. Microbiol.* **2012**, *85*, 179–194. [CrossRef] [PubMed]
48. Divakaruni, A.V.; Baida, C.; White, C.L.; Gober, J.W. The cell shape proteins MreB and MreC control cell morphogenesis by positioning cell wall synthetic complexes. *Mol. Microbiol.* **2007**, *66*, 174–188. [CrossRef]
49. Martins, A.; Contreras-Martel, C.; Janet-Maitre, M.; Miyachiro, M.M.; Estrozi, L.F.; Trindade, D.M.; Malospirito, C.C.; Rodrigues-Costa, F.; Imbert, L.; Job, V.; et al. Self-association of MreC as a regulatory signal in bacterial cell wall elongation. *Nat. Commun.* **2021**, *12*, 2987. [CrossRef] [PubMed]
50. Jean, N.L.; Bougault, C.M.; Lodge, A.; Derouaux, A.; Callens, G.; Egan, A.J.; Ayala, I.; Lewis, R.J.; Vollmer, W.; Simorre, J.P. Elongated structure of the outer-membrane activator of peptidoglycan synthesis LpoA: Implications for PBP1A stimulation. *Structure* **2014**, *22*, 1047–1054. [CrossRef]
51. Paradis-Bleau, C.; Markovski, M.; Uehara, T.; Lupoli, T.J.; Walker, S.; Kahne, D.E.; Bernhardt, T.G. Lipoprotein Cofactors Located in the Outer Membrane Activate Bacterial Cell Wall Polymerases. *Cell* **2010**, *143*, 1110–1120. [CrossRef] [PubMed]
52. Catherwood, A.C.; Lloyd, A.J.; Tod, J.A.; Chauhan, S.; Slade, S.E.; Walkowiak, G.P.; Galley, N.F.; Punekar, A.S.; Smart, K.; Rea, D.; et al. Substrate and Stereochemical Control of Peptidoglycan Cross-Linking by Transpeptidation by *Escherichia coli* PBP1B. *J. Am. Chem. Soc.* **2020**, *142*, 5034–5048. [CrossRef]
53. White, C.L.; Kitich, A.; Gober, J.W. Positioning cell wall synthetic complexes by the bacterial morphogenetic proteins MreB and MreD. *Mol. Microbiol.* **2010**, *76*, 616–633. [CrossRef]
54. Banzhaf, M.; Yau, H.C.; Verheul, J.; Lodge, A.; Kritikos, G.; Mateus, A.; Hov, A.K.; Stein, F.; Wartel, M.; Pazos, M.; et al. The outer membrane lipoprotein NlpI nucleates hydrolases within peptidoglycan multi-enzyme complexes in *Escherichia coli*. *bioRxiv* **2019**, 609503. [CrossRef]
55. Fenton, A.K.; Gerdes, K. Direct interaction of FtsZ and MreB is required for septum synthesis and cell division in *Escherichia coli*. *EMBO J.* **2013**, *32*, 1953–1965. [CrossRef] [PubMed]
56. Schwechheimer, C.; Rodriguez, D.L.; Kuehn, M.J. NlpI-mediated modulation of outer membrane vesicle production through peptidoglycan dynamics in *Escherichia coli*. *Microbiologyopen* **2015**, *4*, 375–389. [CrossRef] [PubMed]
57. Bendezú, F.O.; Hale, C.A.; Bernhardt, T.G.; De Boer, P.A. RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in *E. coli*. *EMBO J.* **2009**, *28*, 193–204. [CrossRef] [PubMed]
58. Berezuk, A.M.; Glavota, S.; Roach, E.J.; Goodyear, M.C.; Krieger, J.R.; Khursigara, C.M. Outer membrane lipoprotein RlpA is a novel periplasmic interaction partner of the cell division protein FtsK in *Escherichia coli*. *Sci. Rep.* **2018**, *8*, 12933. [CrossRef]
59. Caveney, N.A.; Caballero, G.; Voedts, H.; Niciforovic, A.; Worrall, L.J.; Vuckovic, M.; Fonvielle, M.; Hugonnet, J.E.; Arthur, M.; Strynadka, N.C. Structural insight into YcbB-mediated beta-lactam resistance in *Escherichia coli*. *Nat. Commun.* **2019**, *10*, 1849. [CrossRef] [PubMed]
60. Laddomada, F.; Miyachiro, M.M.; Jessop, M.; Patin, D.; Job, V.; Mengin-Lecreulx, D.; Le Roy, A.; Ebel, C.; Breyton, C.; Gutsche, I.; et al. The MurG glycosyltransferase provides an oligomeric scaffold for the cytoplasmic steps of peptidoglycan biosynthesis in the human pathogen Bordetella pertussis. *Sci. Rep.* **2019**, *9*, 4656. [CrossRef]
61. Gerding, M.A.; Liu, B.; Bendezú, F.O.; Hale, C.A.; Bernhardt, T.G.; De Boer, P.A. Self-enhanced accumulation of FtsN at division sites and roles for other proteins with a SPOR domain (DamX, DedD, and RlpA) in *Escherichia coli* cell constriction. *J. Bacteriol.* **2009**, *191*, 7383–7401. [CrossRef] [PubMed]

62. Petiti, M.; Serrano, B.; Faure, L.; Lloubes, R.; Mignot, T.; Duché, D. Tol Energy-Driven Localization of Pal and Anchoring to the Peptidoglycan Promote Outer-Membrane Constriction. *J. Mol. Biol.* **2019**, *431*, 3275–3288. [CrossRef]
63. Bisson-Filho, A.W.; Hsu, Y.P.; Squyres, G.R.; Kuru, E.; Wu, F.; Jukes, C.; Sun, Y.; Dekker, C.; Holden, S.; VanNieuwenhze, M.S.; et al. Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science* **2017**, *355*, 739–743. [CrossRef]
64. Bendezú, F.O.; De Boer, P.A. Conditional lethality, division defects, membrane involution, and endocytosis in mre and mrd shape mutants of *Escherichia coli*. *J. Bacteriol.* **2008**, *190*, 1792–1811. [CrossRef]
65. Pazos, M.; Peters, K.; Casanova, M.; Palacios, P.; VanNieuwenhze, M.; Breukink, E.; Vicente, M.; Vollmer, W. Z-ring membrane anchors associate with cell wall synthases to initiate bacterial cell division. *Nat. Commun.* **2018**, *9*, 5090. [CrossRef] [PubMed]
66. Eberhardt, C.; Kuerschner, L.; Weiss, D.S. Probing the Catalytic Activity of a Cell Division-Specific Transpeptidase In Vivo with β-Lactams. *J. Bacteriol.* **2003**, *185*, 3726–3734. [CrossRef]
67. Varma, A.; Young, K.D. In *Escherichia coli*, MreB and FtsZ direct the synthesis of lateral cell wall via independent pathways that require PBP 2. *J. Bacteriol.* **2009**, *191*, 3526–3533. [CrossRef] [PubMed]
68. Dion, M.F.; Kapoor, M.; Sun, Y.; Wilson, S.; Ryan, J.; Vigouroux, A.; van Teeffelen, S.; Oldenbourg, R.; Garner, E.C. Bacillus subtilis cell diameter is determined by the opposing actions of two distinct cell wall synthetic systems. *Nat. Microbiol.* **2019**, *4*, 1294–1305. [CrossRef]
69. Shiomi, D.; Toyoda, A.; Aizu, T.; Ejima, F.; Fujiyama, A.; Shini, T.; Kohara, Y.; Niki, H. Mutations in cell elongation genes mreB, mrdA and mrdB suppress the shape defect of RodZ-deficient cells. *Mol. Microbiol.* **2013**, *87*, 1029–1044. [CrossRef]
70. Bernal-Cabas, M. The Cpx Pathway Causes Changes in the Peptidoglycan Structure, Turnover, and Recycling. Master's Thesis, University of Alberta, Edmonton, AB, Canada, 2014.
71. Hugonnet, J.E.; Mengin-Lecreulx, D.; Monton, A.; den Blaauwen, T.; Carbonnelle, E.; Veckerlé, C.; Yves, V.B.; van Nieuwenhze, M.; Bouchier, C.; Tu, K.; et al. Factors essential for L,D-transpeptidase-mediated peptidoglycan cross-linking and β-lactam resistance in *Escherichia coli*. *eLife* **2016**, *5*, e19469. [CrossRef]
72. Mohammadi, T.; Karczmarek, A.; Crouvoisier, M.; Bouhss, A.; Mengin-Lecreulx, D.; Den Blaauwen, T. The essential peptidoglycan glycosyltransferase MurG forms a complex with proteins involved in lateral envelope growth as well as with proteins involved in cell division in *Escherichia coli*. *Mol. Microbiol.* **2007**, *65*, 1106–1121. [CrossRef] [PubMed]
73. Yang, D.C.; Blair, K.M.; Salama, N.R. Staying in Shape: The Impact of Cell Shape on Bacterial Survival in Diverse Environments. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 187–203. [CrossRef] [PubMed]
74. Frirdich, E.; Vermeulen, J.; Biboy, J.; Soares, F.; Taveirne, M.E.; Johnson, J.G.; DiRita, V.J.; Girardin, S.E.; Vollmer, W.; Gaynor, E.C. Peptidoglycan LD-carboxypeptidase Pgp2 influences Campylobacter jejuni helical cell shape and pathogenic properties and provides the substrate for the DL-carboxypeptidase Pgp1. *J. Biol. Chem.* **2014**, *289*, 8007–8018. [CrossRef]
75. Blair, K.M.; Mears, K.S.; Taylor, J.A.; Fero, J.; Jones, L.A.; Gafken, P.R.; Whitney, J.C.; Salama, N.R. The Helicobacter pylori cell shape promoting protein Csd5 interacts with the cell wall, MurF, and the bacterial cytoskeleton. *Mol. Microbiol.* **2018**, *110*, 114–127. [CrossRef]
76. Samsudin, F.; Boags, A.; Piggot, T.J.; Khalid, S. Braun's Lipoprotein Facilitates OmpA Interaction with the *Escherichia coli* Cell Wall. *Biophys. J.* **2017**, *113*, 1496–1504. [CrossRef] [PubMed]
77. Braun, V.; Sieglin, U. The Covalent Murein-Lipoprotin Structure of the *Escherichia coli* Cell Wall: The Attachment Site of the Lipoprotein on the Murein. *Eur. J. Biochem.* **1970**, *13*, 336–346. [CrossRef] [PubMed]
78. Zahrl, D.; Wagner, M.; Bischof, K.; Bayer, M.; Zavecz, B.; Beranek, A.; Ruckenstuhl, C.; Zarfel, G.E.; Koraimann, G. Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. *Microbiology* **2005**, *151*, 3455–3467. [CrossRef] [PubMed]
79. Yu, Y.C.; Lin, C.N.; Wang, S.H.; Ng, S.C.; Hu, W.S.; Syu, W.J. A putative lytic transglycosylase tightly regulated and critical for the EHEC type three secretion. *J. Biomed. Sci.* **2010**, *17*, 52. [CrossRef]
80. Sandoz, K.M.; Moore, R.A.; Beare, P.A.; Patel, A.V.; Smith, R.E.; Bern, M.; Hwang, H.; Cooper, C.J.; Priola, S.A.; Parks, J.M.; et al. β-Barrel proteins tether the outer membrane in many Gram-negative bacteria. *Nat. Microbiol.* **2021**, *6*, 19–26. [CrossRef]
81. Turner, R.D.; Hurd, A.F.; Cadby, A.; Hobbs, J.K.; Foster, S.J. Cell wall elongation mode in Gram-negative bacteria is determined by peptidoglycan architecture. *Nat. Commun.* **2013**, *4*, 1496. [CrossRef] [PubMed]
82. Bahadur, R.; Chodisetti, P.K.; Reddy, M. Cleavage of Braun's lipoprotein Lpp from the bacterial peptidoglycan by a paralog of L,D-transpeptidases, LdtF. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2101989118. [CrossRef]
83. 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. USA* **2018**, *115*, E7389–E7397. [CrossRef] [PubMed]
84. Isom, G.L.; Davies, N.J.; Chong, Z.S.; Bryant, J.A.; Jamshad, M.; Sharif, M.; Cunningham, A.F.; Knowles, T.J.; Chng, S.S.; Cole, J.A.; et al. MCE domain proteins: Conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. *Sci. Rep.* **2017**, *7*, 8608. [CrossRef] [PubMed]
85. Yahashiri, A.; Jorgenson, M.A.; Weiss, D.S. Bacterial SPOR domains are recruited to septal peptidoglycan by binding to glycan strands that lack stem peptides. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11347–11352. [CrossRef]
86. Born, P.; Breukink, E.; Vollmer, W. In vitro synthesis of cross-linked murein and its attachment to sacculi by PBP1A from *Escherichia coli*. *J. Biol. Chem.* **2006**, *281*, 26985–26993. [CrossRef] [PubMed]
87. Heidrich, C.; Ursinus, A.; Berger, J.; Schwarz, H.; Höltje, J.V. Effects of multiple deletions of murein hydrolases on viability, septum cleavage, and sensitivity to large toxic molecules in *Escherichia coli*. *J. Bacteriol.* **2002**, *184*, 6093–6099. [CrossRef]

88. Silva, A.M.; Otten, C.; Biboy, J.; Breukink, E.; Van Nieuwenhze, M.; Vollmer, W.; den Blaauwen, T. The fluorescent D-Amino Acid NADA as a tool to study the conditional activity of transpeptidases in *Escherichia coli*. *Front. Microbiol.* **2018**, *9*, 2101. [CrossRef]
89. Scheurwater, E.; Reid, C.W.; Clarke, A.J. Lytic transglycosylases: Bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **2008**, *40*, 586–591. [CrossRef] [PubMed]
90. Vollmer, W. Structure and Biosynthesis of the Murein (Peptidoglycan) Sacculus. In *The Periplasm*; ASM Press: Washington, DC, USA, 2014; Chapter 11. [CrossRef]
91. Morè, N.; Martorana, A.M.; Biboy, J.; Otten, C.; Winkle, M.; Serrano, C.K.; Montón Silva, A.; Atkinson, L.; Yau, H.; Breukink, E.; et al. Peptidoglycan remodeling enables *Escherichia coli* to survive severe outer membrane assembly defect. *MBio* **2019**, *10*, e02729-18. [CrossRef]
92. Lambert, C.; Lerner, T.R.; Bui, N.K.; Somers, H.; Aizawa, S.I.; Liddell, S.; Clark, A.; Vollmer, W.; Lovering, A.L.; Sockett, R.E. Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. *Sci. Rep.* **2016**, *6*, 26010. [CrossRef]
93. Meisel, U.; Höltje, J.V.; Vollmer, W. Overproduction of inactive variants of the murein synthase PBP1B causes lysis in *Escherichia coli*. *J. Bacteriol.* **2003**, *185*, 5342–5348. [CrossRef] [PubMed]
94. Vermassen, A.; Leroy, S.; Talon, R.; Provot, C.; Popowska, M.; Desvaux, M. Cell Wall Hydrolases in Bacteria: Insight on the Diversity of Cell Wall Amidases, Glycosidases and Peptidases Toward Peptidoglycan. *Front. Microbiol.* **2019**, *10*, 331. [CrossRef] [PubMed]
95. Ahangar, M.S.; Furze, C.M.; Guy, C.S.; Cooper, C.; Maskew, K.S.; Graham, B.; Cameron, A.D.; Fullam, X.E. Structural and functional determination of homologs of the Mycobacterium tuberculosis N-acetylglucosamine-6-phosphate deacetylase (NagA). *J. Biol. Chem.* **2018**, *293*, 9770–9783. [CrossRef]
96. van Heijenoort, J. Peptidoglycan Hydrolases of *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 636–663. [CrossRef] [PubMed]
97. Sycuro, L.K.; Rule, C.S.; Petersen, T.W.; Wyckoff, T.J.; Sessler, T.; Nagarkar, D.B.; Khalid, F.; Pincus, Z.; Biboy, J.; Vollmer, W.; et al. Flow cytometry-based enrichment for cell shape mutants identifies multiple genes that influence Helicobacter pylori morphology. *Mol. Microbiol.* **2013**, *90*, 869–883. [CrossRef] [PubMed]
98. Rae, C.S.; Geissler, A.; Adamson, P.C.; Portnoy, D.A. Mutations of the Listeria monocytogenes peptidoglycan N-Deacetylase and O-acetylase result in enhanced lysozyme sensitivity, bacteriolysis, and hyperinduction of innate immune pathways. *Infect. Immun.* **2011**, *79*, 3596–3606. [CrossRef] [PubMed]
99. Rocaboy, M.; Herman, R.; Sauvage, E.; Remaut, H.; Moonens, K.; Terrak, M.; Charlier, P.; Kerff, F. The crystal structure of the cell division amidase amic reveals the fold of the AMIN domain, a new peptidoglycan binding domain. *Mol. Microbiol.* **2013**, *90*, 267–277. [CrossRef]
100. Bernhardt, T.G.; De Boer, P.A. The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Mol. Microbiol.* **2003**, *48*, 1171–1182. [CrossRef]
101. Yang, D.C.; Tan, K.; Joachimiak, A.; Bernhardt, T.G. A conformational switch controls cell wall-remodelling enzymes required for bacterial cell division. *Mol. Microbiol.* **2012**, *85*, 768–781. [CrossRef] [PubMed]
102. Derouaux, A.; Wolf, B.; Fraipont, C.; Breukink, E.; Nguyen-Distèche, M.; Terrak, M. The monofunctional glycosyltransferase of *Escherichia coli* localizes to the cell division site and interacts with penicillin-binding protein 3, FtsW, and FtsN. *J. Bacteriol.* **2008**, *190*, 1831–1834. [CrossRef]
103. Mohammadi, T.; Van Dam, V.; Sijbrandi, R.; Vernet, T.; Zapun, A.; Bouhss, A.; Diepeveen-De Bruin, M.; Nguyen-Distéche, M.; De Kruijff, B.; Breukink, E. Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO J.* **2011**, *30*, 1425–1432. [CrossRef] [PubMed]
104. Crépin, S.; Ottosen, E.N.; Peters, K.; Smith, S.N.; Himpsl, S.D.; Vollmer, W.; Mobley, H.L. The lytic transglycosylase MltB connects membrane homeostasis and in vivo fitness of Acinetobacter baumannii. *Mol. Microbiol.* **2018**, *109*, 745–762. [CrossRef] [PubMed]
105. Moynihan, P.J.; Cadby, I.T.; Veerapen, N.; Jankute, M.; Crosatti, M.; Mukamolova, G.V.; Lovering, A.L.; Besra, G.S. The hydrolase LpqI primes mycobacterial peptidoglycan recycling. *Nat. Commun.* **2019**, *10*, 2647. [CrossRef] [PubMed]
106. Keck, W.; Van Leeuwen, A.M.; Huber, M.; Goodell, E.W. Cloning and characterization of mepA, the structural gene of the penicillin-insensitive murein endopeptidase from *Escherichia coli*. *Mol. Microbiol.* **1990**, *4*, 209–219. [CrossRef]
107. Liu, B.; Hale, C.A.; Persons, L.; Phillips-Mason, P.J.; de Boer, P.A. Roles of the DedD Protein in *Escherichia coli* Cell Constriction. *J. Bacteriol.* **2019**, *201*, e00698-18. [CrossRef] [PubMed]
108. Alcorlo, M.; Martínez-Caballero, S.; Molina, R.; Hermoso, J.A. Carbohydrate recognition and lysis by bacterial peptidoglycan hydrolases. *Curr. Opin. Struct. Biol.* **2017**, *44*, 87–100. [CrossRef] [PubMed]
109. Kocaoglu, O.; Carlson, E.E. Profiling of β-lactam selectivity for penicillin-binding proteins in *Escherichia coli* strain DC2. *Antimicrob. Agents Chemother.* **2015**, *59*, 2785–2790. [CrossRef]
110. Egan, A.J.; Biboy, J.; van't Veer, I.; Breukink, E.; Vollmer, W. Activities and regulation of peptidoglycan synthases. *Philos. Trans. R. Soc. B Biol. Sci.* **2015**, *370*, 20150031. [CrossRef] [PubMed]

Localisation of MCE domain containing proteins PqiB and MlaD does not co-ordinate with the division site.

Chris LB Graham, Hanjeong Harvey, Katie Graham, Michael Alao, Hannah Doherty, Jack Bryant, Manuel Banzhaf, Lori Burrows, David Roper

1. University of Warwick, UK
2. University of Birmingham, UK
3. McMaster University, Canada

Abstract

If the linkage between peptidoglycan and membrane biogenesis is explored it may allow us to postulate the most efficient combinations of antibiotics to address Gram-negative bacterial pathogens in a novel way. Here we interrogated localisation of the MCE domain containing lipid transport proteins MlaD and PqiB to determine if during division, like other studies the proteins are divisionally localised. Our experiments have shown that among the MCE domain-containing proteins liable for mCherry chromosomal fusion, the transport machinery PqiB and MlaD are more localised to poles and longitudinally in *P. aeruginosa,* not at the division site. This polar localisation increased in stationary phase cells especially. In addition to these experiments, we could not show a coordinated insertion of peptidoglycan and MlaD localisation. We also repeated the experiments in *E.coli* with results which hinted that this is likely a shared localisation trait in both species. This suggests the role of these proteins is necessary perhaps for maintaining lipid curvature at the poles, irrespective of division, and their is a localised role for MlaD and PqiB dependent on cell need.

Introduction

The insertion and maintenance of lipids in the inner leaflet of the outer membrane in Gram-negative bacteria, is not yet fully understood(1–3). Whilst labelling of cells has revealed lipid raft localisation after insertion has already taken place, no publications have yet fixed this insertion or increase in insertion as a timed or localised event, much like other parts of cell division such as peptidoglycan growth, which has been shown to be facilitated by a localisation dependent set of peptidoglycan synthesis and maturation machinery(4). Recently it has been shown that mycobacterial lipids, as well as outer membrane protein insertion by the BAM complex(5), and the Lpt system of LPS insertion are inserted in concert with peptidoglycan(6).

The outer membrane and peptidoglycan must be made simultaneously if the lipids are not fluid enough as is the case with the outer membrane(7), or there will be cause of blebbing and bubbling in the membranes

without the flexibility or fluidity, and weakening which can lead to loss of pathogenicity and change in phenotype(8). Specifically to our paper, we investigate the insertion of inner leaflet lipids into the outer membrane. We hypothesise, based on previous data indicating a high OMP density and low membrane fluidity(7), that an over-expansion of the inner leaflet of the outer membrane would be caused by too many lipids in a confined space compared to the LPS, perturbing balance and promoting membrane curvature, which in turn would affect anchoring partners that hold the lipid membranes in place as well as the rigid peptidoglycan mesh. YebT, MlaD and PqiB were all hypothesised to be involved in the movement and exchange of lipids from the inner membrane to the outer membrane, thus were prime candidates for the creation and growth of new outer membrane cell envelope at the inner leaflet(1,9). Recently another paper has also uncovered AsmA domain containing proteins to be also involved in outer membrane lipid balance and insertion(3) which although not talked about in this paper, these can be assumed to be taking a role.

In this article, we were able to track the MCE domain containing proteins MlaD and PqiB to determine if their fluorescence was coordinated to this division process. We found instead these proteins to be located along the periplasm during growth and more pole related in localisation, suggesting they may be more involved in pole maintenance and elongation, particularly of pole associated lipids such as cardiolipin and phosphatidyl ethanolamine, as previously suggested(10).

Methods

Fluorescent labelling of OM proteins *in vivo- P. aeruginosa*

pEX18Gm plasmids were created through synthesis of geneblocks of FtsW, MlaD, PqiB and YebT genes, found through gene searching and use of online webtool “xBase” including 600bp flanking regions. The geneblocks were then cloned using Gibson assembly, once amplified by PCR into an existing pEX18Gm plasmid. Once gibson cloning was successful for each vector, they were interrogated by sequencing and then once confirmed, the suicide vectors were transformed into *Pseudomonas aeruginosa*, with only meridiploid bacteria (or genomically incorporating bacteria surviving an antibiotic challenge by gentramycin post transformation.)
The merodiploid conjugates were then purified by sucrose selection, and removal of the gentramycin challenge over successive generations to produce strains with the full genomic conversion of wildtype, to fluorescently tagged genes.

Imaging of fluorescence in *Pseudomonas aeruginosa*

Strains were streaked from glycerol stocks onto LB Agar plates, and incubated at 37 degrees C overnight. One PA colony was grown overnight at 37 °C, at 180 rpm in 2 ml LB. The following morning the a 1/10

dilution of the samples were then grown at 37°C, 180 rpm for a period of time until the samples had all reached an OD_{600} of 0.3, typically 2 hours. 1% agarose in PBS was heated in a microwave until piping hot. Microscope slides were topped with a solution of 1% agarose in PBS which was flattened with a coverslip and left to cool. The coverslip was removed and 10 µl of pre-prepared sample was added to the slide, the coverslip was placed on top to spread the sample across the slide. Samples were then analysed using a Leica DMi8 confocal microscope specifically to identify the fluorescence, with corresponding filters dependent on expected fluorescence. HADA labelling was detecting using a DAPI filter set for 200ms, mCherry fluorescence was detected at 2 seconds laser exposure on a TXR filter set, with 100ms exposure for Bright-Field/BF detection.

HADA labelling

Cells were labelled as previously(11) with HADA (3-[[(7-Hydroxy-2-oxo-2H-1-benzopyran-3-yl)carbonyl]amino]-D-alanine hydrocholoride) a fluorescent D- amino acid, for 2 minutes at 100uM at 0.1 OD, the cells were then washed with PBS solution twice by centrifugation.

Analysis of fluorescence localisation

Images were imported at LIFs or TIF libraries, which were then, analysed by MicrobeJ software to determine cell width, determine contouring and maxima points. The points of increased fluorescence to background, were then tracked across the cell and mapped per individual point across thousands of cells, dependent on cell length. The points were then mapped for each strain. Automatic, cellular counting and size determination by MicrobeJ(12) and BactMAP(13) allowed for quantitative analysis. Fluorescent points were tracked using custom tolerance and intensity filters maintained throughout study. Scripts and conditions for image analysis attached in the Supplementary Data file. Correlation between fluorescent channels was found using the Colocalisation threshold feature of ImageJ.

Chemical genomics assays

In evaluating phenotypic differences between the *Pseudomonas aeruginosa* mutants (ΔmlaD-mCherry, ΔpqiB-mCherry and Δwaal-mCherry) and the wildtype, a chemical genomics assay was conducted. In the chemical genomics assay, phenotypic observations under different stress conditions were linked to the genotypic differences between the mutants and the wildtype(14). The mutants and the wildtype were inoculated on LB agar plate with each mutant replicated 96 times . In addition, the mutants and the wildtype were inoculated on LB agar plates with each plate containing sub-minimum inhibitory concentration of different membrane stresses and antibiotics. Each condition plate has 4 replicas. The plates were incubated at 37^0C for 12 hours and images were taken afterwards. The images of the plates were subjected to an image analysing software 'Iris'(15). Iris identified each colony on the plate and measures the fitness in terms of size and opacity. The fitness of the mutants under different stress conditions were then statistically analysed and compared to that of wildtype with a software called ChemGAPP(16). ChemGAPP normalised the colonies around the edges of the plates and fitness scores were generated for each mutant.

Colony size data for each plate was checked for plate effects via a Wilcoxon rank sum test between the distributions of colony sizes for the outer two edges of the plate and the centre colonies. Where $p < 0.05$ the difference in the distributions was considered as statistically significant and the outer edge colonies were scaled to the plate middle mean (PMM). The plate middle mean is defined as the mean colony size of colonies within the centre of the plate, within the 40^{th} and 60^{th} percentile of size. Each plate was then scaled such that the PMM was equal to the median colony size of all colonies within the screen. Colony sizes greater than 10,000 were removed as outliers. For the bar plots and heatmaps, the average colony size was taken for each set of mutants within each plate and divided by the average colony size of the wildtype of the same plate to produce a fitness ratio. The average fitness ratios for each mutant in each condition was calculated between replicate plates for display. The 95% confidence intervals between replicates was calculated and displayed as error bars. For the swarm plots, each individual mutant colony within each plate was divided by the mean colony size of the wildtype

Escherichia coli fluorescent tagging

The pBAD30 vector modified to have kanamycin resistance was expressed with a genetic copy of the *pqiB* and *mlaD* from the *Escherichia coli* MG1655 genome. A tagRFP label was attached after an 'SGSGSG' amino acid linker. The strains were visualised at 0.1OD using the TXR filter of a Leica DMi8 confocal microscope for 1 second exposure before analysis on ImageJ software.

Results

MCE domain containing proteins PqiB and MlaD are associated with cell poles.

In order to associate this fluorescence and insertion of inner leaflet lipids with a protein, the MCE domain containing proteins PqiB and MlaD of *P. aeruginosa PAO1* were identified as PA3213 and PA4454 respectively and genetically replaced with mCherry C-terminal fusions using a suicide vector and sucrose selection. These new *Pseudomonas aeruginosa* strains have been assayed for growth and found to behave as wildtype in a variety of membrane stress conditions, indicating they are likely successfully mimicking wildtype behaviour, with no protein aggregation. In Figure 1, we tracked MlaD and PqiB fluorescence as well as Waal O antigen ligase in these strains, as a relevant inner membrane-linked control.

Our fluorescent imaging (Figure 1) has shown that the MCE domain containing proteins MlaD and PqiB, rather than having a division-based localisation pattern, are dispersed throughout the periplasm, with a slight polar preference, both when observed qualitatively (Fig1a), and in quantitative imaging through MicrobeJ(12)(Fig1b).

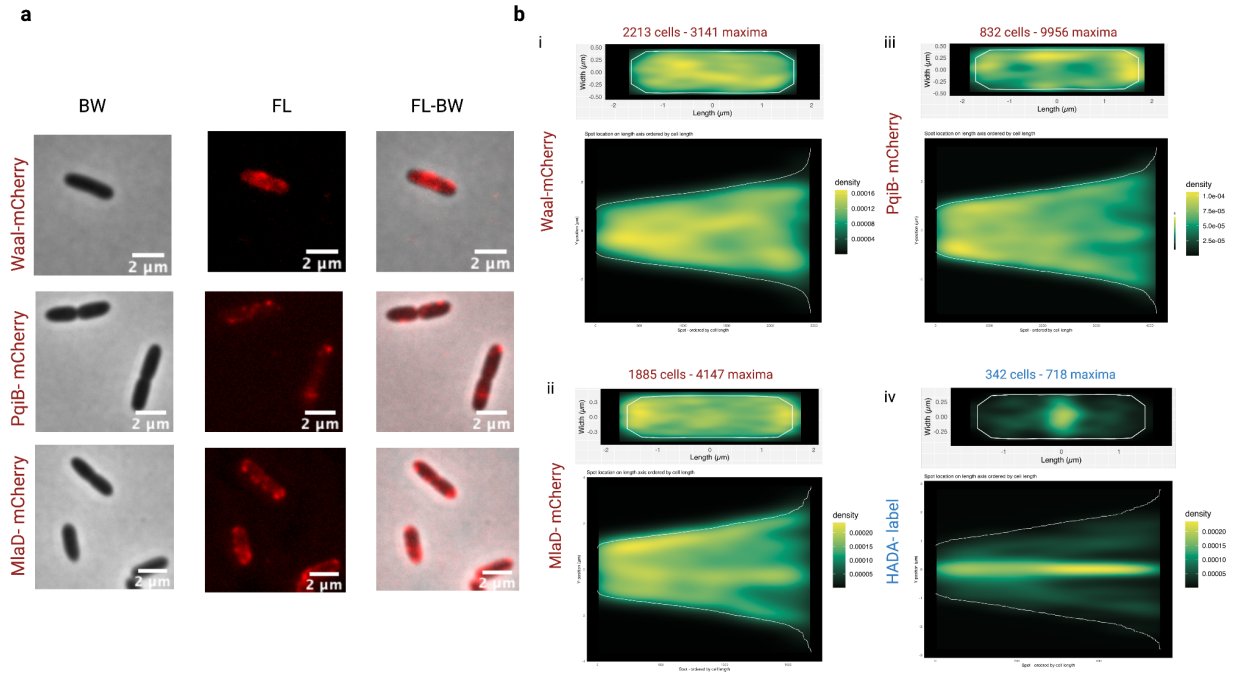


Figure 1. Localisations of MlaD, Waal and PqiB Systems in *Pseudomonas aeruginosa* determined by mCherry fluorescent localisation.

a Confocal and fluorescent images of *Pseudomonas aeruginosa PAO1* cells, labelled genetically at the chromosome (Waal/PqiB/MlaD) with an mCherry fusion BW- Brightfield, FL- Fluorescent channel, FL-BW- Merge. **b** BactMAP length sorted images (below) and qMean densities of each label. HADA label control included, indicating division and pre-division sites.

In Figure 1b we can see a comparative increase in the BactMap heatmaps(13) of length-sorted images, at the division site in the long cells in PqiB and MlaD, similar to our HADA peptidoglycan synthesis label division control(17), however this is likely due to poles after a division event, where both systems are in higher abundance. However across images we see the complexes and increases in fluorescence are also present across the cell, reflective of more than one localisation. This is in contrast to HADA labelling, which clearly shows a high increase of fluorescence at the division site, with a presence across the cell as well as pre-septal regions. Waal across our analysis had no specific localisation pattern detected. However in both PqiB and MlaD, in images and in our quantitative analysis, a slight preference for the poles was seen throughout. We repeated these experiments in *E. coli*, using a leaky plasmid system not linked to the genome, to show a similar localisation pattern for both PqiB and MlaD (SI Figure 1) here too.

In order to see if this fluorescence was coordinating with peptidoglycan insertion, we used HADA labelling as a control for division site detection in our mCherry fusion strains and looked for correlation in localisation(18), the mCherry labelled cells were pulse chased at 0.1OD for 5 minutes with HADA 100µM. An analysis of the fluorescence in HADA labelled cells again indicated there is no division-site related increase in PqiB or MlaD cells.(Figure 2) Figure 2ai, of a dividing *Pseudomonas aeruginosa* cell

during exponential growth shows a localisation of MlaD in foci throughout the cell, this is in contrast to the HADA localisation, which shows an increase during division, also during pre-division (Figure2ai). Whilst there is a cross-over of fluorescence of HADA insertion and MlaD localisation(Figure 2b), especially when division isnt yet occuring, we find that on average this fluorescence does not correlate (Fig2c). MlaD and PqiB therefore are not transported to the division site during division.

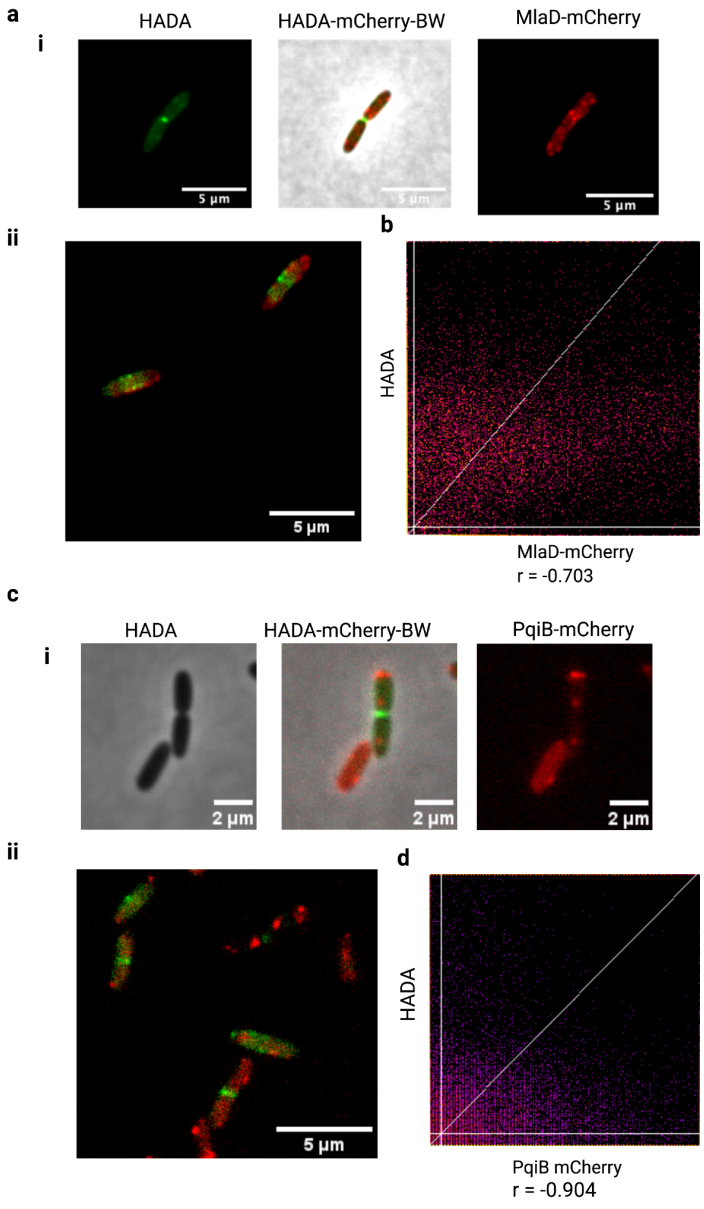


Figure 2. HADA labelling, and division shows no co-localisation with MlaD

a- MlaD Localisation of HADA(Green) and MlaD(Red) in example dividing cell **i** and pre-division **ii b** Scatterplot of localisation of post thresholded regions within cells, indicate no correlation, with a pearsons

correlation of -0.703 **c-PqiB** Localisation of HADA(Green) and PqiB(Red) in example dividing cell **i** and pre-division **ii d** Scatterplot of localisation of post thresholded regions within cells, indicate no correlation, with a pearsons correlation of -0.903

The polar localisation of the MCE domain containing proteins in *Pseudomonas aeruginosa* MlaD and PqiB in experiments was not expected, and at stationary phase of overnight growths the polar localisation was shown to be more pronounced. Therefore we analysed stationary phase cells, and recovered these cells to see changes in fluorescence. This revealed the polar localisation to be partially lost after incubation with new media, suggesting that this polar increase is associated with aging cells (Figure 3). This localisation preference was more accentuated in PqiB cells, with MlaD-mCherry fluorescent cells, also showing general periplasmic localisation after 60 minutes growth, indicating the polar role in MlaD and PqiB may be only part of their function in the cell.

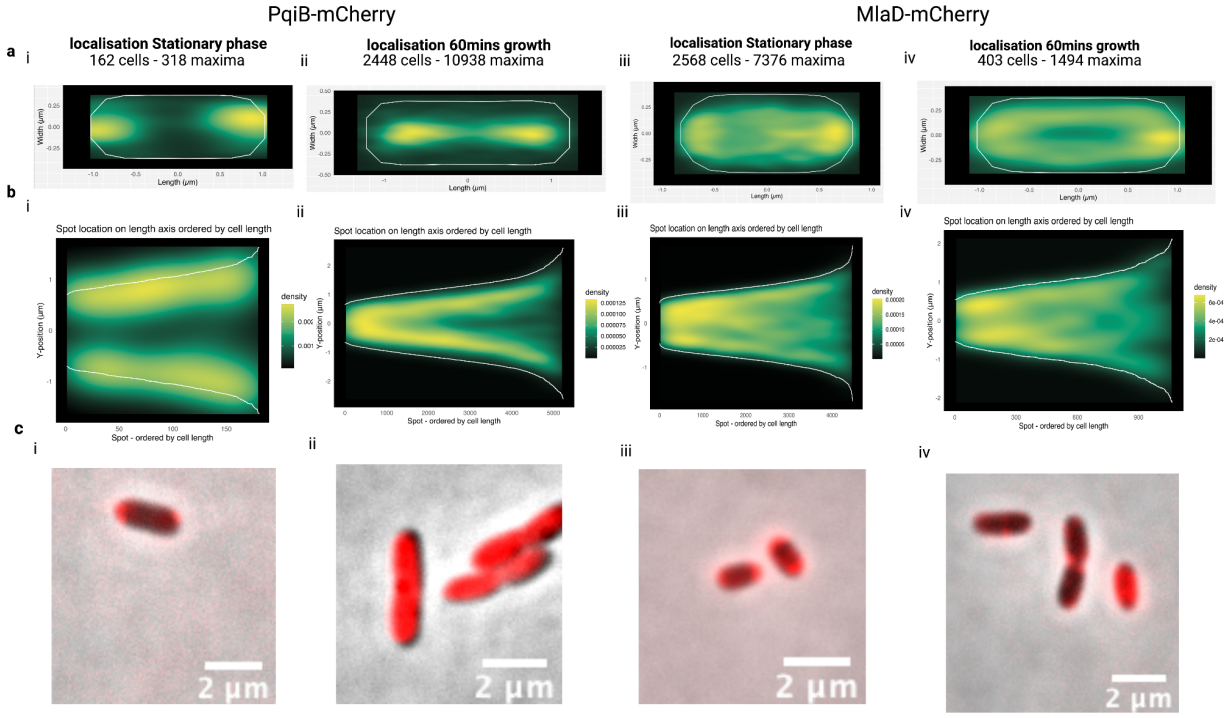


Figure 3. PqiB and MlaD during stationary phase have a higher polar preference, which is relieved on growth

Pseudomonas aeruginosa PAO1 chromosomal fluorescent fusion localisations at stationary phase and after recovery through additional media **a** Qmean distribution of fluorescent maxima visualisation by average location (green-yellow density) as analysed by MicrobeJ and BactMAP **b** longitudinal distribution based on cell length **c** Fluorescent and BW channels merged, showing localisation as seen through microscope.

Discussion

We labelled the MCE domain containing proteins with a fluorescent tag, expecting their localisation to correlate with division (Figure 2). In chromosomal fused fluorescently tagged strains shown to have a similar phenotype to wildtype cells (SI Figure 2) however, the fluorescence whilst showing during division does not significantly indicate any co-ordination with the peptidoglycan insertion/division proxy label HADA (Figure 3). There is in fact a small preference towards polar localisation in the MCE domain containing proteins targeted. We then determined this polar localisation is exaggerated in stationary phase cells, and reduced upon inoculation with new media and subsequent new growth. The MCE domain of MlaD has been associated with cardiolipin(10), therefore it could be speculated that these polar localisations, or increase in fluorescence at the poles is related to a role in promoting and maintaining polar curvature with cardiolipin and other lipids. Cardiolipin presence at the poles is also noted to increase during stationary phase(19), which would fit this narrative, however we have no evidence they are undertaking this role in this localised region.

However we do note, that there seems to be a localised role for both proteins in *Pseudomonas aeruginosa* cells, and also *Escherichia coli* cells. This specific localisation for maintenance of the inner leaflet of the outer membrane indicates there is likely an importance in the local lipid environment. Previous studies on the transcriptome of *Pseudomonas aeruginosa PAO1* have shown *mlaD* PA4454 and *pqiB* 3213 to be differentially expressed, with an increase in transcription during heatshock in the *mlaD* gene, and no such increase in *pqiB*, similar to the upregulation of LPS synthesis genes(20). Their deletion in *E.coli* has also shown to cause minor cell envelope defects, therefore our data, along with these previous studies suggests they may have a role in later-stage membrane stabilisation and homogenesis.

Our experiments therefore have some evidence to suggest that if inner leaflet formation of the outer membrane is in concert with division, the MCE domain-containing proteins PqiB and MlaD are not part of this process, and these roles are likely performed by another protein or set of proteins, such as the AsmA domain-containing proteins recently discovered. Therefore given the localisation shown by PqiB and MlaD indicating that lipid transport into the inner leaflet of the outer membrane is important, studies into the localisation of the AsmA domain-containing proteins during division and throughout the cell cycle could bring us closer into uncovering the full mechanisms of cell envelope formation at the outer membrane's inner leaflet.

Author Contributions

CG/MB/DR/LB - writing, Pseudomonas strain creation KG/HH/CG, Design/plasmids CG, PG experiment CG, Chemical genomics- MA/CG, Microscopy and labelling - CG Experimental design and planning - CG/JB/MB/LB/DR/MA, Grant acquisition DR/MB/LB/CG

Supplementary Data

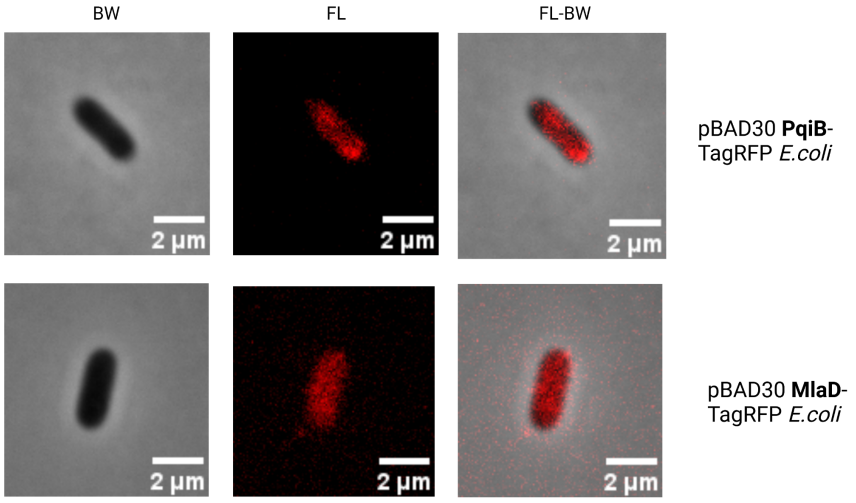
Supplementary Data is available at the DOI

References

1. Isom GL, Davies NJ, Chong ZS, Bryant JA, Jamshad M, Sharif M, et al. MCE domain proteins: Conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. Sci Rep. 2017;
2. Mika JT, Thompson AJ, Dent MR, Brooks NJ, Michiels J, Hofkens J, et al. Measuring the Viscosity of the Escherichia coli Plasma Membrane Using Molecular Rotors. Biophys J. 2016 Oct 4;111(7):1528–40.
3. Douglass MV, McLean AB, Trent MS. Absence of YhdP, TamB, and YdbH leads to defects in glycerophospholipid transport and cell morphology in Gram-negative bacteria. PLOS Genet. 2022 Feb 28;18(2):e1010096.
4. Graham CLB, Newman H, Gillett FN, Smart K, Briggs N, Banzhaf M, et al. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-Negative Bacteria. Int J Mol Sci. 2021 Jan;22(23):12831.
5. Consoli E, Luirink J, den Blaauwen T. The Escherichia coli Outer Membrane β-Barrel Assembly Machinery (BAM) Crosstalks with the Divisome. Int J Mol Sci. 2021 Jan;22(22):12101.
6. Lithgow T, Stubenrauch CJ, Stumpf MPH. Surveying membrane landscapes: a new look at the bacterial cell surface. Nat Rev Microbiol. 2023 Feb 24;1–17.
7. Horne JE, Brockwell DJ, Radford SE. Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. J Biol Chem. 2020 Jul 24;295(30):10340–67.
8. Isom GL, Rooke JL, Antunes CA, Sheehan E, Wells TJ, Icke C, et al. Mammalian Cell Entry domains are required for bile resistance and virulence in Salmonella [Internet]. bioRxiv; 2018 [cited 2023 Feb 28]. p. 263871. Available from: https://www.biorxiv.org/content/10.1101/263871v1
9. Ekiert DC, Bhabha G, Isom GL, Greenan G, Ovchinnikov S, Henderson IR, et al. Architectures of Lipid Transport Systems for the Bacterial Outer Membrane. Cell. 2017 Apr 6;169(2):273-285.e17.
10. Yero D, Díaz-Lobo M, Costenaro L, Conchillo-Solé O, Mayo A, Ferrer-Navarro M, et al. The Pseudomonas aeruginosa substrate-binding protein Ttg2D functions as a general glycerophospholipid transporter across the periplasm. Commun Biol. 2021 Apr 9;4:448.
11. Kuru E, Velocity Hughes H, Brown PJ, Hall E, Tekkam S, Cava F, et al. In situ Probing of Newly Synthesized Peptidoglycan in Live Bacteria with Fluorescent D-Amino Acids. Angew Chem Int Ed Engl. 2012 Dec 7;51(50):12519–23.
12. Ducret A, Quardokus EM, Brun YV. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol. 2016 Jun 20;1(7):16077.
13. van Raaphorst R, Kjos M, Veening JW. BactMAP: An R package for integrating, analyzing and visualizing bacterial microscopy data. Mol Microbiol. 2020;113(1):297–308.
14. Cacace E, Kritikos G, Typas A. Chemical genetics in drug discovery. Curr Opin Syst Biol. 2017 Aug 1;4:35–42.

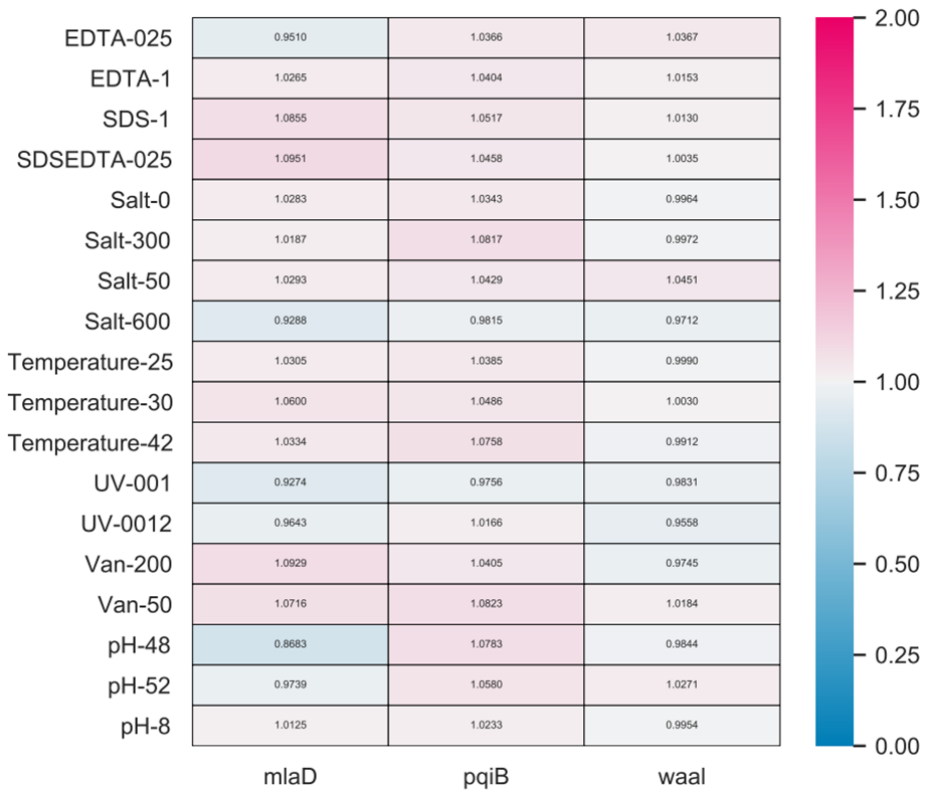
15. Kritikos G, Banzhaf M, Herrera-Dominguez L, Koumoutsi A, Wartel M, Zietek M, et al. A tool named Iris for versatile high-throughput phenotyping in microorganisms. Nat Microbiol. 2017 Feb 17;2:17014.
16. Doherty H, Kritikos G, Galardini M, Banzhaf M, Moradigaravand D. ChemGAPP; A Package for Chemical Genomics Analysis and Phenotypic Profilin [Internet]. 2023 [cited 2023 Feb 28]. Available from: https://europepmc.org/article/PPR/PPR593530
17. Silva AM, Otten C, Biboy J, Breukink E, Van Nieuwenhze M, Vollmer W, et al. The fluorescent D-Amino Acid NADA as a tool to study the conditional activity of transpeptidases in Escherichia coli. Front Microbiol. 2018;
18. Arena ET, Rueden CT, Hiner MC, Wang S, Yuan M, Eliceiri KW. Quantitating the cell: turning images into numbers with ImageJ. WIREs Dev Biol. 2017;6(2):e260.
19. Agrawal A, Rangarajan N, Weisshaar JC. Resistance of early stationary phase E. coli to membrane permeabilization by the antimicrobial peptide Cecropin A. Biochim Biophys Acta BBA - Biomembr. 2019 Oct 1;1861(10):182990.
20. Chan KG, Priya K, Chang CY, Rahman AYA, Tee KK, Yin WF. Transcriptome analysis of Pseudomonas aeruginosa PAO1 grown at both body and elevated temperatures. PeerJ. 2016 Jul 19;4:e2223.

Supplementary Data



SI Figure 1. *E.coli* plasmid based expression of MlaD and PqiB reveals similar localisation patterns to *Pseudomonas aeruginosa PAO1*

Fluorescent localisation of non-native 'leaky' expression of MlaD-tagRFP and PqiB-tagRFP fusion plasmids at exponential growth in *Escherichia coli BL21* using pBAD30 vector BW- Confocal, FL- Fluorescent channel, FL-BW- Merge.



SI Figure 2. Chemical screening of Genetic- mCherry *Pseudomonas aeruginosa* chromosomal knockins.

Fitness of mCherry-gene knock-ins compared to wildtype. Fitness 1 = wildtype fitness. Salt- mM concentrations, Temperature - ^{o}C, pH 4.8, 5.2 and 8.

Dunnett's multiple comparisons test	Predicted (LS) mean diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value
EDTA-025					
WT vs. mlaD	217.4	-1054 to 1489	No	ns	0.9562
WT vs. pqiB	-142.3	-1414 to 1129	No	ns	0.9868
WT vs. waal	-137.2	-1409 to 1135	No	ns	0.9882
EDTA-1					
WT vs. mlaD	20.55	-1251 to 1292	No	ns	>0.9999
WT vs. pqiB	-29.13	-1301 to 1243	No	ns	>0.9999
WT vs. waal	-209.7	-1481 to 1062	No	ns	0.9604
SDS-1					
WT vs. mlaD	-241.9	-1514 to 1030	No	ns	0.9414
WT vs. pqiB	-76.98	-1349 to 1195	No	ns	0.9978
WT vs. waal	-26.98	-1299 to 1245	No	ns	>0.9999
SDS-EDTA					
WT vs. mlaD	-115.3	-1673 to 1442	No	ns	0.9961
WT vs. pqiB	52.25	-1505 to 1610	No	ns	0.9996
WT vs. waal	66.16	-1491 to 1624	No	ns	0.9993
Salt-0					
WT vs. mlaD	-8.792	-1281 to 1263	No	ns	>0.9999
WT vs. pqiB	-35.47	-1307 to 1236	No	ns	0.9999
WT vs. waal	3.045	-1269 to 1275	No	ns	>0.9999
Salt-300					
WT vs. mlaD	59.93	-1212 to 1332	No	ns	0.999
WT vs. pqiB	-61.15	-1333 to 1211	No	ns	0.9989
WT vs. waal	-31.81	-1304 to 1240	No	ns	0.9999
Salt-50					
WT vs. mlaD	1.734	-1556 to 1559	No	ns	>0.9999
WT vs. pqiB	-33.96	-1592 to 1524	No	ns	>0.9999

WT vs. waal	-215	-1773 to 1343	No	ns	0.9759
Salt-600					
WT vs. mlaD	119.8	-2083 to 2323	No	ns	0.9984
WT vs. pqiB	31.05	-2172 to 2234	No	ns	>0.9999
WT vs. waal	48.51	-2154 to 2251	No	ns	>0.9999
Van-200					
WT vs. mlaD	-114.4	-1386 to 1157	No	ns	0.993
WT vs. pqiB	53.93	-1218 to 1326	No	ns	0.9993
WT vs. waal	177.9	-1094 to 1450	No	ns	0.975
Van-50					
WT vs. mlaD	-88.66	-1360 to 1183	No	ns	0.9967
WT vs. pqiB	-135.4	-1407 to 1136	No	ns	0.9886
WT vs. waal	-92.04	-1364 to 1180	No	ns	0.9963
pH4.8					
WT vs. mlaD	387	-1171 to 1945	No	ns	0.8822
WT vs. pqiB	-112.6	-1670 to 1445	No	ns	0.9963
WT vs. waal	102.1	-1455 to 1660	No	ns	0.9972
pH5.2					
WT vs. mlaD	161.7	-1396 to 1719	No	ns	0.9894
WT vs. pqiB	-162.3	-1720 to 1395	No	ns	0.9893
WT vs. waal	-117.1	-1675 to 1440	No	ns	0.9959
pH8					
WT vs. mlaD	9.441	-1262 to 1281	No	ns	>0.9999
WT vs. pqiB	-11.57	-1283 to 1260	No	ns	>0.9999
WT vs. waal	-40.86	-1313 to 1231	No	ns	0.9996

SI Table 1. Anova test of strain growth

Constructs and plasmids used in this study

Name	**Sequence**		**Notes**
Primers	**Primer 1**	**Primer 2**	**Description of primer function**
pEX18seq	GAT TAA GTT GGG TAA CGC CAG	GTG GAA TTG TGA GCG GAT AAC	Suicide Vector used for propagation in E.coli and genetic insertion in Pseudomonas sp.
Pseudogen eamplifier	AAG CTT GCA TGC CTG CAG G	GAA TTC GAG CTC GGT ACC CGG	A primer for the amplification of mCherry constructs
linpEX	CGG GTA CCG AGC TCG AAT TCG TAA TCA TGG	GTC GAC CTG CAG GCA TGC AAG CTT GG	Primers for amplification of the pEX18 plasmid
PqiBlin	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CGC TTC GGC GTG CTG TTC CAG CAG	CCA AGC TTG CAT GCC TGC AGG TCG ACG GCC TTG CCG AAG GCG CTG ACC AC	Primers to linearise and amplify PqiB-mCherry for gibson cloning
linRodA	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG GTA CCG CCC AGG TGG TGG CGA TCA AGC AG	CCA AGC TTG CAT GCC TGC AGG TCG ACC GGG TAG TCG AAG CCG AGG GTG GAC AG	Primers to linearise and amplify RodA-mCerulean for gibson cloning
linFtsW	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CGT GAC GCC GGG CTG ATT GCC CAG	CCA AGC TTG CAT GCC TGC AGG TCG ACC CTC TGG CAA CAG CTT GTT CAG CGG TTC CG	Primers to linearise and amplify FtsW-mCerulean for gibson cloning
linWaaI	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CCA ACT CGA CCG CAA CGA CCC GGA G	CCA AGC TTG CAT GCC TGC AGG TCG ACA TCG AGA TCA TCG ACT CCA ATG GCC GGG TCA TG	Primers to linearise and amplify Waal-mCherry for gibson cloning
linYebT	CCA AGC TTG CAT GCC TGC AGG TCG ACA TCG AGA TCA TCG ACT CCA ATG GCC GGG TCA TG	CCA AGC TTG CAT GCC TGC AGG TCG ACG CTT GCT GCT GGT CGC CTA TGC CCT G	Primers to linearise and amplify YebT-mCherry for gibson cloning
linMlaD	CCA TGA TTA CGA ATT CGA GCT CGG TAC CCG CAT TGG CAT GGT GCT GGC CCT GCA G	CCA AGC TTG CAT GCC TGC AGG TCG ACG GTC GTT GCG GTT CTT CTG CAT GGT GTC GG	Primers to linearise and amplify MlaD-mCherry for gibson cloning
Plasmid	**Sequence**		

pEX18	Gifted by Prof Lori Burrows
PqiB-mCherry_pUC57	GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTC GGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGAT AAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTT GCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACAT CGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTT CTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTT GGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATG AGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTG TAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATG GAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTG AGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAG TCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAA GTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCAT GACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCT TTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTAC CAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCAC CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTC GTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG GAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCT ATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTAT CCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGA GTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTG GCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCA GGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACC ATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGTGCGGGTGTTCGGCGAAGACCTGCTGCAACTGCCAGAGGAG CGCCGCTCCCTGGTCGAACGGCGCTTCGGCGTGCTGTTCCAGCAGGGCGCGCTGTTTTCCTCGCTTACCGTGGTGGAGAACG TCGCCCTGCCGCTGATCGAAAACGCCGGCCTGCCGCGCGCCGACGCCGAGCACCTGGCCCAGGTCAAGCTGGCCCTGGCCG GGCTACCGGCCAACGCCGGCGACAAGTATCCGGCCTCGCTTTCCGGCGGCATGATCAAGCGCGCCGCCCTGGCCCGCGCCC TGGCGCTGGACCCGGACATCCTGTTCCTCGACGAGCCCACCGCCGGCCTCGACCCGATCGGCGCGGCGGCCTTCGACAACC TGATCCGTACCCTGCGCGACGCCCTCGGCCTGACCGTGTTCCTGGTCACCCATGACCTGGACACCCTGTACACCATCTGCGA CCGGGTCGCCGTGCTGTCGCAGAAGAAGGTCCTGGTAGTCGATACCCTGGAACGGGTCGCCGCCACCGACGACGCCTGGGT CCGCGAATACTTCCACGGGCCGCGCGGGCGCGCCGCCCACCAAGCCGCCGCCGTGCCGGAGAATCACTGAGATGGAAACCC GAGCCCACCATGTGTTGATCGGCCTGTTCAGCGTGATCGTGATCGGCGCCGCCCTGCTCTTCGGCCTGTGGCTGGCCAAGTC CGGCTCGGAGGGCAAGTTCAACTACTACGACATCGTCTTCAACGAAGCGGTCAGTGGCCTTTCCCAGGGCAGCTCGGTGCA GTACAGCGGGATCAAGGTCGGCGACGTGGCCTTCCTCCGCCTCGATCCGAAGGACCCGCGCAAGGTCTGGGCACGCATCCG CGTGGTGGCCAGCGCGCCGATCAAGCAGGACACCACCGCCAAGCTGGCGCTCACCGGGATCACCGGCACCTCGATCATCCA GCTCAGCAGCGGCACGCCCGCCAGCCCGATGCTGGAGGGCAAGGACGGGAAGATCCCGGTGATCGTCGCCACGCCCTCGC CGCTGACCCAGTTACTGAGCAACGGCGAAGACCTGATGGGCAACATCAACCAGCTCATCGCGCGCTTCAGCAACCTGCTCT CCGAGGAAAACACAGCGCGCATCAGCCGCACCCTCGACCATCTCGACCAGGCCACCGGCGCGCTCTCCGCCGAGCGCGAG AACGTCAGCGCGGTGATGCAGCAGTTGGCCCAGGCGAGCAGGCAGGCCAACGCCGCCCTGGCCCAGGCCAGCGAGTTGAT GCGCAGCGCCAACGGCCTGCTCAACGAGCAAGGCAAGGGGATGCTGGAAAACGCCAACAAGACCATGGCTTCGCTGGAGC GCACCAGCGCCACCCTCGACCAGTTGATCAGCGAGAATCGCCACTCGCTGGACGGCGGCATCCAGGGACTCGCCGAGCTGG GCCCGGCGGTCAGCGAACTGCGCGATACCCTGGCGGCACTGCGCGGAATCAGCCGGCGCCTGGAAGAGAACCCGGCGAAC TACCTGCTGGGCCGGGAAAAAACCAAGGAGTTCACCCCATCGGGGTCCGGATCGGGGATGGTCTCGAAGGGGGAGGAAGA CAACATGGCCATCATCAAAGAGTTCATGCGGTTCAAGGTCCATATGGAAGGCAGCGTGAATGGCCACGAATTCGAGATCGAA GGGGAGGGGGAGGGGCGGCCGTATGAAGGGACCCAAACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCTGCCGTTCG CGTGGGATATCCTGTCCCCGCAATTCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACTACCTCAAG CTGTCCTTCCCCGAGGGCTTCAAATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGACGGTGACCCAGGATAGC AGCCTCCAGGATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGGACCAACTTCCCGTCCGATGGGCCCGTCATGCAAAAG AAGACGATGGGCTGGGAGGCCTCGAGCGAGCGGATGTATCCCGAAGACGGGGCCCTCAAAGGGGAGATCAAGCAACGCCT GAAGCTGAAAGATGGGGGCCATTATGATGCGGAAGTGAAGACGACCTACAAAGCCAAGAAACCGGTGCAACTGCCGGGCG CGTACAATGTCAATATCAAACTGGATATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGCCGAAGG GCGGCACTCGACGGGGGGGATGGATGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGGCTCGCCCTTCGCCCGCTGCGT CGCCTGTCCCTGGCCGCCGGCCTGGCCGCGCTCGCCACCCTCGGCGCCTGCTCGATCCTGCCGGAGGCGCAGGTCCTCCAG

	GTCTACCTGCTACCGGTGCACAACCCTCCGGCCAGCGCCGCCGCGCGGCCGGTCGACTGGTCGCTGCGGATCGCCCGGCCG CGTACCAGCCTGGTGCTGGAGAGCCCGCGCATCGCGGTGCGCCCGCACGGCGACGAAATCAGCGTGTACCAGGGCGCGCGC TGGAGCGATCCGGCGCCGTCGCTGCTGCGCGATCGTCTGATGCAGGCGTTCCAGGCCGACGGCCGGGTCCGCGGCCTGAGC AGCGACGACAGCAACCTGCAGGCCGATTTCGAACTGGGCGGCGACCTTCGCGCCTTCCAGACCGAGTACCCGAACGGCCA GGCCAGCGCGCTGATCCGCTACGACGCGCGCCTGGTGCGTACCGACGACAAGCGCGTGGTCGCCAGCCGGCGTTTCGAGGT CAGCCAGCCGGTGGATGGCAAGAAGGTCGCGGCGGTGGTCAGCGCCTTCGGCAAGGCCGGCGATACGCTGTCGGCCCAGG TTCTCGACTGGAAGGATCCCCGGGTACCGAGCTCGAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGG CGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTT CCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGC ATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGC CCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTC ACCGTCATCACCGAAACGCGCGA
Waal-mCherry_pUC57	GCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTAC TGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAG CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGG CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTAT AGGGCAAGCTTGCATGCCTGCAGGTCGACTCTAGGGATCCCACATATCCAGCGCGACGCGGTGCACTACGAGCCGCTACCGG GGCAGACCCTCCGCCAACTCGACCGCAACGACCCGGAGTGCGGCGACGCAGTCCGGACGCAACTGGGCGAGCTGATCGCC GCGCTGCACGCCAAGGGGGTGTACTTCCGCTCGCTGCACCTGGGCAACGTGGTGCGCACGCCGGAAGGCCGGATGGGCCTG ATCGACATCGCCGACCTGCGGGTCCAGCGCTCGGCGCTCAGCGCCGCCCAGCGCATTCGCAATTTCAAGCACCTGCTGCGCT ACGAACAGGATCGAACCTGGTTCCTCGACGACAGCGAGCACCTGGTGCTCAAGGCCTATCTCCGCAGCAGCCAGGTCCACT GGACACTCGAGCAGCTCCTCAAGCAACTGCAACTGGATTGAGCAAGGAATGTTCGCAGCTACCCGACTGTCCCGTCTTCGT CACGATACCAGCCGCATCCTCAGCCACTGGATCCTGCCGCTGGGCTGGCTGGCCCTGCTCACCGGAATGTTCTGGGTCGGCG ACCGTTCCGACTATCACCGGCTTTTCTATATCTTGCTGGCTGCACCGACGCTGCTGTACGTGATACTGCAACCGCGTCTGCTTC GCCCGCTAACGGGCTCGCCGCTGTTCATCGCCTTCCTCGCCTTCAGCAGCTACATGATGCTGAGCCTGTCCTGGTCGACCCCC GAGAACTCCACCGGCTCGCTGCTCAAGCGCCCCCTGTACATCGCCCTCCTGTTCTTCTGCGCCGCCATCCTGGCGCTCGAAG CTCCCCTCCGCCTCAAGACCGCGACCTGGCTGGCAGCCCTTGGCGCGGTGATCTCCGCGGCAGCCACCTTGCTGCGATACTA CTGGGACGCCAACCCGCTGCGCCTTACCGGTTATGGCGCTCTCTATAACCCATTACTGAGCGCCCATGTCTATGGCGCCTTCA CGGCACTCTGGCTGGCCTATTGGATGCAATCCCGCCCCATCCTGGCCCCCTTGCCACTGATATCTCTGGCGCTGCTCGGTGGC CTTCTCATCGCGACCGGTTCACGTACTCCCCTGGTAGGGCTCACAGCGGCCCTTATGTGGCTGGTCCTGGCCGGAGACAGGA AAAAAGCCCTCATCGCCCTGGCACTCGCCCTGGCTGGAGCGCTACTGGGCTACATCCTGTACCCGGAAGTGATCACCCAGAG AGGGGCATCGTTCCGCCCGGAAATCTGGGCCGACGCCCTACGCCAGATCAGCGAGCATCCATGGCTCGGACACGGCTACGAT CATCCGATGCGAATCGTCCTGAGCAACGGCATGCTGCTCGCCGACCCGCACAACATCGAACTGGGTGTGCTTTTCGCTGGAG GGATCATCGGGCTCCTGCTGTGGGTGGCGATCTACGCACTGGCCTTCGGCTTCTCCTGGAAAAACCGGAAGTCTCCAGCCGT TCTGCTAGCCTCGACCTGGCTGGTGTTCGGCCTGGCAGCCGGACTCACGGAGGGCAACGCCTTCCTGCCCCGTCCCAAGGA ACACTGGTTCCTGATCTGGATTCCCATGGCGCTGCTGTATGCGCTGTGGATCCAGCAAAGGTTCGCCGCCAGCAGGCGCGGT GAGGATATCGCCGCGCCTTCGGGGTCCGGATCGGGGATGGTCTCGAAGGGGGAGGAAGACAACATGGCCATCATCAAAGAG TTCATGCGGTTCAAGGTCCATATGGAAGGCAGCGTGAATGGCCACGAATTCGAGATCGAAGGGGAGGGGGAGGGGCGGCCG TATGAAGGGACCCAAACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCTGCCGTTCGCGTGGGATATCCTGTCCCCGCAA TTCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACTACCTCAAGCTGTCCTTCCCCGAGGGCTTCA AATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGACGGTGACCCAGGATAGCAGCCTCCAGGATGGCGAGTTCA TCTACAAGGTGAAGCTCCGGGGGACCAACTTCCCGTCCGATGGGCCCGTCATGCAAAAGAAGACGATGGGCTGGGAGGCCT CGAGCGAGCGGATGTATCCCGAAGACGGGGCCCTCAAAGGGGAGATCAAGCAACGCCTGAAGCTGAAAGATGGGGGCCAT TATGATGCGGAAGTGAAGACGACCTACAAAGCCAAGAAACCGGTGCAACTGCCGGGCGCGTACAATGTCAATATCAAACTG GATATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGCCGAAGGGCGGCACTCGACGGGGGGGATG GATGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGTGCGGCGATATCAGCGATCAGGAGAAAAACAGCCGCGCAATGCC GCGGAAGGTCTTCTCGTTCCAGTACTTCAGCGGCAAGCCGGAGAGCAGCTCCCGGGCCAGTACCTTGTCCCGGTTCGAGCA CTTGAGGAACATCGAATTGCGGAAGCGCATGATCACCTGCTCGTAATCGGGGTGATCGCTGAACTGGCTGTACGTCTTGAAT ACGTTGTCCACCATGAAACGGGCGTTCTTGTACGTATTGGTGGGATGCTTGCGGTACTGGGCGAGGACCTCGCCAAGAATGT CGATGACGTAGCCAGCCTTGGTCACCGCCAGTTCGATATAGACATCTTCCAGGCGTATGTCCGGATCGAAACCGCCGACCTTC TCCAGGGCTTCCCGCCGGAACATCAAGGTGGGCGCCATCGGGCCCGGCTTGCGGTCCAGGAACAGGTCATCGAAGTCGAGA CGGCGGAAAGGCAGGTCGCGTTTGCGCTGCTCCTTTCCCGGCATGACCCGGCCATTGGAGTCGATGATCTCGATATTGCCAG GATCCCCGGGTACCGAGCTCGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA CGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGACAGCTTATCATCGAATTTCTGCCATTCATCCGCT TATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACGCCCCGCCCTGCCAC TCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAG CGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGT

	TTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGC
	CAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCG
	ATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATT
	GCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTT
	TACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAAT
	GTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAA
	TCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGT
	CTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCC
	GTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGA
	AGTGACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTGTTCC
	GGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAAGTTCTGCAAAGCCTGA
	TGGGACATAAGTCCATCAGTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTT
	TGCGATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATATGGAAATGTGG
	AACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGAACGAAACAGTCGGGA
	AAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCATGATG
	CCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTTACGTTGAAGTGGAGC
	GGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCG
	CCGCAGTCGAGCGACAGGGCGAAGCCCTCGGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTATGGCCCTGCAAACGC
	GCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTCGCGGAAAACTTGGCC
	CTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGACAGATGAGGGGCAGGC
	TCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGCGAGTTTCCCACAGATG
	ATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAGATGAGGGGCGCGATCC
	TTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTCCACAGGCAGAAAATCC
	AGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAACCAATATTTATAAACCTT
	GTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTCTCGAACCCTCCCGGTC
	GAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAAAACTTCCCTTGGGGTTAT
	CCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCA
	TGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACAGTCCTGTCTGTGACA
	AATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTAT
	TTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGT
	AAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATCAAAAACGTATGCTGT
	ATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATAT
	GCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGGAAGTG
	GTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAACGTGCGCACAGTCC
	ATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCG
	GCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGTAAGCCGGATGGCT
	CAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGC
	CGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCC
	GCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTTATCTGTCACAGATTT
	GAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTC
	TCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTTCGTC
	ATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGAATTGGCTATCCGCGT
	GTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTC
	TTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACTGAGGTATGTG
	CTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGATTTTGTTGTTGCTTTG
	CAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAAACACTTAACC
	AGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCGGAAGCGAGGAAAAT
	AACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGG
	GCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATATGC
	GTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGCCCATAAAGGTGGCGT
	TTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACC
	CCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTTTCTATCTT
	GGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTTCCTGTCTGGCTCTGCA
	CCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTG
	AAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATTAATGTCGTATGTGCTG
	CTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGC
	TCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCC
	CCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAACGGATGAAGTTGGT
	AAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTGGAGAAATGCTCTTT

	CTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGCGCCTG TTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGATGGTGGATTCGTTA ATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGT GCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGCTTACTGAGGACGC ACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCTGGTGTCATAGAAA TTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGA GCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGTTATGCAAGCCGAT TGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACA CCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAAAAA GCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGGGGTGATATTTGAAG CTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTT GCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTGAGTGTATAGA GAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCACGTTTTAGTCTACGTTTATCTGTCTTTAC TTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGTCT GATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGG TCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGT CGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATC GTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGT ATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACT TGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTCGTAACC TGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCCACAACATTTTGCGC ACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGATGCAAGTGTGTCGCTG TCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAAGTTGTTTTTACGTTA AGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATAT GTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCGGGTTTTCGCTAT TTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAA GGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAA GTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTCGAT
MlaD_pUC57	GCGGCCGCAAGGGGTTCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTAC TGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAG CTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGG CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTAT AGGGCAAGCTTGCATGCCTGCAGGTCGACTCTAGGGCCTGTTCATTGGCATGGTGCTGGCCCTGCAGGGCTACAACATCCTG ATTTCCTATGGTTCCGAGCAGGCGGTCGGGCAGATGGTGGCCCTGACCCTGCTGCGCGAACTGGGGCCGGTGGTGACCGGC CTGCTGTTCGCCGGCCGTGCCGGCTCTGCGCTTACCGCGGAAATCGGCAACATGAAGGCCACCGAACAGCTTTCCAGCCTG GAAATGATCGGTGTCGACCCGCTCAAGTACATCGTCGCGCCGCGCCTGTGGGCCGGTTTCATCTCCATGCCGCTGCTCGCCG CGATCTTCAGCGTGGTCGGCATCTGGGGTGGGGCGATGGTCGCGGTCGACTGGCTGGGCGTATACGAGGGGTCGTTCTGGGC GAACATGCAGAACAGCGTGCAGTTCACCGAGGATGTGCTCAACGGCGTGATCAAAAGTATCGTATTCGCCTTCGTGGTGACC TGGATCGCGGTCTACCAAGGCTACGACTGCGAGCCGACCTCGGAAGGAATCAGCCGGGCGACGACCCGGACCGTGGTCTAT GCCTCCCTGGCGGTGCTGGGGCTCGACTTCATTCTGACTGCTTTGATGTTTGGAGATTTCTGAATGCAAACCCGCACCCTGG AAATCGGTGTCGGCCTGTTCCTCCTGGCCGGCCTGCTGGCCCTGTTGCTGCTGGCCCTGCGGGTCAGCGGCCTGAGCGTGGG CAACGCCGGCGATACCTACAAGGTCTACGCCTACTTCGACAACATCGCCGGTGTTACCGTGCGCGGCAAGGTCACCCTGGCC GGCGTGACGATCGGCAAGGTGACGGCGGTCGACCTGGATCGCGACAGCTACACTGGTCGCGTGACCATGGAGATCAACCAG AACGTGAACAACCTGCCGGTCGATTCCACGGCGTCGATCCTGACCGCCGGCCTGCTGGGCGAGAAATACATCGGCATCAGC GTCGGCGGCGACGAGGACGTTCTGAAGGACGGCAGCACCATCCACGACACCCAGTCGGCGCTGGTGCTGGAAGACCTGAT CGGCAAGTTCCTGCTGAACTCGGTTAACAAAGACGAAGCCAAAAAGTCGGGGTCCGGATCGGGGATGGTCTCGAAGGGGG AGGAAGACAACATGGCCATCATCAAAGAGTTCATGCGGTTCAAGGTCCATATGGAAGGCAGCGTGAATGGCCACGAATTCG AGATCGAAGGGGAGGGGGAGGGGCGGCCGTATGAAGGGACCCAAACCGCGAAACTGAAGGTGACCAAAGGGGGGCCGCT GCCGTTCGCGTGGGATATCCTGTCCCCGCAATTCATGTACGGCTCGAAAGCGTACGTGAAGCATCCGGCGGATATCCCGGACT ACCTCAAGCTGTCCTTCCCCGAGGGCTTCAAATGGGAGCGCGTCATGAACTTCGAAGATGGGGGGGTCGTGACGGTGACCC AGGATAGCAGCCTCCAGGATGGCGAGTTCATCTACAAGGTGAAGCTCCGGGGGACCAACTTCCCGTCCGATGGGCCCGTCAT GCAAAAGAAGACGATGGGCTGGGAGGCCTCGAGCGAGCGGATGTATCCCGAAGACGGGGCCCTCAAAGGGGAGATCAAGC AACGCCTGAAGCTGAAAGATGGGGGCCATTATGATGCGGAAGTGAAGACGACCTACAAAGCCAAGAAACCGGTGCAACTG CCGGGCGCGTACAATGTCAATATCAAACTGGATATCACGTCCCACAATGAAGACTATACCATCGTGGAACAGTATGAACGCGC CGAAGGGCGGCACTCGACGGGGGGGATGGATGAACTGTATAAAAGCGGGTCCGGGTCCGGATGAGGTTTTCTTCCATGCTG ACTCTTCTGCGTCGCGGCCTGCTGGTGTTCCTGGCGGCTTTCCCGCTGCTTTCCATGGCGGCGCCGACCCCGCAACAGGTGG

	TGCAGGGCACGGTCGACGAACTGCTTTCCGACATCAAGGCCAACAAGGCCGCCTACAAGGCCGATCCGCAAAAGCTCTACG
	CCACTCTCGACCGTATCCTTGGGCCGGTGGTCGATGCCGAAGGCATCGCCAAGAGCGTGATGACCGTCAAGTACTCGCGCCA
	GGCCTCACCCGAGCAGATCAAGCGCTTCGAGGAAGTGTTCAAGAACAGCCTGATGCAGTTCTACGGCAACGCGCTGCTCGA
	ATACGACAACCAGGACATCCGCGTGCTGCCTAGTTCGGCCAAGCCGAGCGACGATCGCGCCAGCGTCAACATGGAGATCCG
	TGACAGCAAGGGCACGGTCTATCCGGTCTCCTACACCATGACCAACCTGGCCGGTGGCTGGAAGGTCCGCAACGTGATCATC
	AACGGCATCAACATCGGCAAGCTGTTCCGCGACCAGTTCGCCGACACCATGCAGAAGAACCGCAACGACCTCGAGAAGAC
	CATCGCCGGCTGGGAGGATCCCCGGGTACCGAGCTCGAGTATTCTATAGTCTCACCTAAATAGCTTGGCGTAATCATGGTCAT
	AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG
	TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGC
	ATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGACAGCTTATCATCGAATT
	TCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTA
	CGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGAT
	GAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTT
	GTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAAC
	CCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTG
	GTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACC
	AGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAA
	ACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTG
	ACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGC
	TTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCT
	TACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATT
	CTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGCGATAAGCTCATGGAGCGGCGTAACCGTCGCACAGGAAGGACAGAGA
	AAGCGCGGATCTGGGAAGTGACGGACAGAACGGTCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGT
	GTGACGTTCTCTGTTCCGGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTATACCGCTGAAA
	GTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAAGTCTACACGAAGGTTTTTGCGCTGGATGTGGCTGCCC
	GGCACCGGGTGCAGTTTGCGATGCCGGAGTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCC
	TTTATATGGAAATGTGGAACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTTATCCACTGAGAAGCGA
	ACGAAACAGTCGGGAAAATCTCCCATTATCGTAGAGATCCGCATTATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTA
	GTGTTCTGTCATGATGCCTGCAAGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGTT
	ACGTTGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAACCATGATGTGGTCTGTCCTTTTAC
	AGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCGAAGCCCTCGGCTGGTTGCCCTCGCCGCTGGGCTGGCGGCCGTCTA
	TGGCCCTGCAAACGCGCCAGAAACGCCGTCGAAGCCGTGTGCGAGACACCGCGGCCGGCCGCCGGCGTTGTGGATACCTC
	GCGGAAAACTTGGCCCTCACTGACAGATGAGGGGCGGACGTTGACACTTGAGGGGCCGACTCACCCGGCGCGGCGTTGAC
	AGATGAGGGGCAGGCTCGATTTCGGCCGGCGACGTGGAGCTGGCCAGCCTCGCAAATCGGCGAAAACGCCTGATTTTACGC
	GAGTTTCCCACAGATGATGTGGACAAGCCTGGGGATAAGTGCCCTGCGGTATTGACACTTGAGGGGCGCGACTACTGACAG
	ATGAGGGGCGCGATCCTTGACACTTGAGGGGCAGAGTGCTGACAGATGAGGGGCGCACCTATTGACATTTGAGGGGCTGTC
	CACAGGCAGAAAATCCAGCATTTGCAAGGGTTTCCGCCCGTTTTTCGGCCACCGCTAACCTGTCTTTTAACCTGCTTTTAAAC
	CAATATTTATAAACCTTGTTTTTAACCAGGGCTGCGCCCTGTGCGCGTGACCGCGCACGCCGAAGGGGGGTGCCCCCCCTTC
	TCGAACCCTCCCGGTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACACGATGCCTGAAAAA
	ACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTATTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCC
	TTGTAGGCCTTTATCCATGCTGGTTCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGAC
	AGTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTTTTTCACAAAGTTATCCCTGCT
	TATTGACTCTTTTTTATTTAGTGTGACAATCTAAAAACTTGTCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATC
	AATCACAAGAAACGTAAAAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCGGGATC
	AAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCTACAGGAACATGACGGTATCTGCGAGATC
	CATGTTGCTAAATATGCTGAAATATTCGGATTGACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGC
	GGGGAAGGAAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTCCTTGGTTTATCAAA
	CGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCCATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGG
	TTTACGCAGTTTCGGCTTAGTGAAACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCG
	TAAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAGCTGCCTCAAAGTTACCAGCGT
	ATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTAATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGA
	GAAAAAGAAAGGCCGCCAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTGAGGGTT
	ATCTGTCACAGATTTGAGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTAATTTGTCACAGTTTTGCTGTTTCCTTCAG
	CCTGCATGGATTTTCTCATACTTTTTGAACTGTAATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTT
	CTCTTTCCCTTCGTCATGTGACCTGATATCGGGGGTTAGTTCGTCATCATTGATGAGGGTTGATTATCACAGTTTATTACTCTGA
	ATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGATATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGA
	CAGAACAGTTCTTCTTTGCTTCCTCGCCAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGT
	GACTGAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGTTTTTTGATGACTTTGCGAT
	TTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAAACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCAT

	GGAAACACTTAACCAGTGCATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGCCCG GAAGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGGGGTTTCTTCTCAGGCTATCAGAGAT GCCGAGAAAGCAGGGCGACTACCGCACCCGGATATGGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAA CAAATTAATCATATGCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGGGGTTGCTGC CCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGATCTGGCTCTGAAGGGGCTACGTGTTTTGCTCG TGGAAGGTAACGACCCCCAGGGAACAGCCTCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCT CCTGCCTTTCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGCTTGACATTATTCCTT CCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATTTGATGAAGGTAAACTGCCCACCGATCCACACCTGATG CTCCGACTGGCCATTGAAACTGTTGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATT AATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACACCTCCGCACTGCAGTTTTTCGAT ATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGGGTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGT AATGGCTCTCAGTCCCCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTACGTGAAA CGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCCATTGATCAACGCTCTTCAACTGGTGCCTG GAGAAATGCTCTTTCTATTTGGGAACCTGTCTGCAATGAAATTTTCGATCGTCTGATTAAACCACGCTGGGAGATTAGATAAT GAAGCGTGCGCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACACCAGCTGCCCCGA TGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAATGCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAG TTTACTCTTGAAGTGCTCCGGGGTGATAGTGTTGAGAAGACCTCTCGGGTATGGTCAGGTAATGAACGTGACCAGGAGCTGC TTACTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGCGTTCGGTCGAAGAGTATCT GGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGCTGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCG AGCTGGATGATGAGCAGATGGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCAGCGT TATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGCGGAAAATATTTCACGTAAGATTATTACC CGCTGTATCAACACCGCCAAATTGCCTAAATCAGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGA TGCACTTCAAAAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAGAAAAAAGCTGG GGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTAAAACGTCATCTGCATCAAGAACTAGTTTAAGCT CACGACATCAGTTTGCTCCTGGAGCGACAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCA ACTGAGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCGACCACGTTTTAGTCTACG TTTATCTGTCTTTACTTAATGTCCTTTGTTACAGGCCAGAAAGCATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCAC TTGTATCGTCGGTCTGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCC CACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATAATCAGACTGGGACCACGG TCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCAC GGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACC ACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCACTCGTGTTGTCGGTCTGATTATCGGTCTGGG ACCACGGTCCCACTTGTATTGTCGATCAGACTATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGAC ATGTCGTCGTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTGTGTCCTGCTTATCC ACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAGGCCGTGCCGGCACGTTAACCGGGCTGCATCCGATG CAAGTGTGTCGCTGTCGACGAGCTCGCGAGCTCGGACATGAGGTTGCCCCGTATTCAGTGTCGCTGATTTGTATTGTCTGAA GTTGTTTTTACGTTAAGTTGATGCAGATCAATTAATACGATACCTGCGTCATAATTGATTATTTGACGTGGTTTGATGGCCTCCA CGCACGTTGTGATATGTAGATGATAATCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTC GCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACTTAATGTTTTTATTTAAAATA CCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAGCGAGCTTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTG GAATGAACAATGGAAGTCCGAGCTCATCGCTAATAACTTCGTATAGCATACATTATACGAAGTTATATTCGAT
Ecoli PqiB	ATTCGAATTCTttgtttaactttaagaaggagatatacatATGGAGTCAAATAACGGAGAAGCTAAAATACAAAAAGTCAAAAACTGGAGCCC GGTGTGGATTTTCCCGATTGTGACGGCGCTGATTGGTGCTTGGGTTCTGTTTTACCACTACAGCCATCAGGGTCCAGAGGTTA CACTCATCACCGCCAATGCCGAAGGTATCGAGGGCGGTAAGACCACGATTAAATCCCGTAGCGTCGATGTCGGAGTGGTGGA GAGCGCTACCTTGGCGGATGACCTGACACATGTTGAGATCAAGGCGCGTTTGAACTCCGGTATGGAAAAGCTGCTGCACAA GGATACCGTGTTTTGGGTTGTGAAACCGCAAATCGGTCGTGAGGGTATCAGCGGCCTGGGCACGTTACTGAGCGGCGTATAT ATTGAACTGCAGCCAGGGGCGAAAGGCTCTAAGATGGATAAATACGATCTCCTTGACTCTCCACCGCTGGCCCCACCGGATG CAAAGGGCATACGCGTGATCCTGGACAGCAAGAAAGCGGGTCAGCTGAGCCCGGGCGATCCGGTTCTCTTCCGTGGTTATC GTGTCGGTTCCGTGGAAACCTCGACCTTCGATACCCAAAAACGCAACATTAGCTATCAACTCTTTATCAACGCTCCGTATGAT CGTCTGGTGACCAATAACGTTCGTTTTTGGAAAGATTCTGGCATCGCCGTTGATCTGACCTCGGCGGGTATGCGTGTTGAGAT GGGTTCTCTAACGACCCTGCTGTCTGGTGGCGTTAGCTTTGATGTGCCGGAGGGCTTGGACCTGGGCCAACCGGTAGCTCCG AAAACGGCGTTTGTTCTGTACGACGATCAGAAGAGCATCCAAGACTCACTCTATACCGATCACATTGATTACCTGATGTTTTT CAAGGACTCCGTGCGTGGTTTGCAGCCGGGTGCTCCGGTTGAGTTCAGAGGGATCCGTCTGGGCACCGTTAGCAAGGTGCC GTTTTTCGCACCGAACATGCGTCAGACTTTCAATGATGACTACCGCATTCCCGTCCTGATTCGCATCGAACCGGAACGTCTTA AGATGCAACTGGGTGAAAATGCGGACGTTGTTGAGCACCTGGGTGAGCTGTTGAAGCGCGGATTAAGGGGCTCCTTGAAGA CCGGCAACCTGGTCACCGGTGCGCTGTACGTTGACCTGGATTTTTACCCGAACACCCCGGCAATTACCGGTATCCGTGAGTT CAACGGTTACCAGATCATCCCGACCGTGTCCGGCGGTCTGGCGCAGATTCAGCAGCGCTTAATGGAAGCGTTAGACAAAATT AACAAGCTGCCGCTGAATCCGATGATTGAGCAGGCAACCAGCACCCTGTCTGAATCCCAACGCACGATGAAAAACCTGCAA

	ACCACGCTGGACAGCATGAACAAGATTTTGGCGAGCCAGAGCATGCAACAGCTGCCGACTGATATGCAAAGCACCCTGCGC GAGTTGAACCGTAGCATGCAGGGCTTTCAGCCGGGTTCTGCCGCGTATAACAAGATGGTGGCGGACATGCAGCGCTTGGACC AGGTTTTACGTGAGTTACAACCTGTGCTGAAGACCCTGAATGAAAAAAGCAATGCTCTGGTTTTTGAGGCGAAGGACAAGA AAGACCCGGAGCCGAAAAGAGCAAAGCAAAGCGGCAGTGGCAGCGGTATGAGCAAGGGTGAGGAACTGTTCACCGGTGT GGTTCCGATCCTGGTGGAGTTGGACGGTGACGTTAACGGCCACAAGTTCTCGGTCTCCGGTGAAGGTGAAGGCGACGCCAC GTACGGTAAACTGACTCTGAAGTTCATTTGTACCACTGGCAAACTGCCGGTGCCTTGGCCGACGTTGGTGACCACCTTCGGC TACGGCGTACAGTGCTTTGCGCGTTATCCGGACCACATGAAACAGCATGATTTCTTCAAGAGCGCTATGCCGGAGGGCTATGT TCAAGAGCGTACCATTTTCTTCAAAGACGATGGTAACTATAAGACCCGTGCAGAAGTGAAATTCGAAGGCGACACCTTGGTT AACCGTATTGAACTAAAAGGTATCGACTTCAAAGAAGACGGTAATATTTTGGGCCACAAGCTGGAATATAACTACAACTCCC ACAATGTTTATATCATGGCAGATAAACAAAAAAATGGCATCAAAGTGAACTTTAAGATCCGTCATAACATCGAGGACGGTTCT GTTCAGCTGGCGGACCATTATCAGCAGAATACCCCGATCGGTGACGGTCCGGTGCTCTTGCCGGATAATCATTACCTGTCAAC GCAGAGCGCTCTGTCTAAAGATCCAAATGAAAAACGCGACCACATGGTTCTTCTCGAGTTCGTGACTGCGGCAGGTATCACC CATGGTATGGATGAACTCTACAAAAAGCTTGGCTGTTTTGGCGG
Ecoli MlaD	ATTCGAATTCTttgtttaactttaagaaggagatatacatATGCAAACAAAAAAGAATGAAATATGGGTAGGCATTTTCTTGCTGGCTGCGCTGT TAGCGGCGCTGTTTGTGTGCCTGAAGGCTGCTAATGTGACTTCAATCCGCACCGAGCCGACCTACACCCTTTATGCGACGTTT GATAACATTGGCGGTCTGAAGGCGCGTAGCCCGGTCAGCATCGGTGGTGTTGTCGTAGGTCGTGTTGCAGATATTACCCTGG ATCCGAAAACCTACCTGCCGCGTGTGACGCTAGAAATTGAGCAGCGTTACAACCACATCCCGGACACCTCTAGCCTGAGCAT CCGCACTTCCGGTCTCCTGGGCGAGCAATATCTGGCGCTAAATGTCGGTTTTGAGGACCCAGAGCTGGGTACGGCGATTTTG AAGGACGGCGACACGATTCAGGATACCAAGAGCGCAATGGTGCTAGAGGACCTGATTGGTCAGTTTCTGTATGGTAGCAAG GGTGACGACAATAAAAATTCTGGAGACGCGCCGGCGGCCGCGCCAGGTAACAACGAAACCACCGAACCGGTGGGTACGAC GAAAAGCGGTTCCGGCTCGGGTATGAGCAAAGGCGAGGAACTGTTCACCGGCGTTGTTCCGATCTTGGTTGAATTAGATGGC GATGTTAATGGCCATAAATTTAGCGTGTCCGGCGAGGGCGAGGGTGACGCTACCTACGGTAAGCTGACCTTGAAGTTCATCT GCACCACGGGTAAGCTGCCTGTGCCGTGGCCGACGTTGGTTACCACCTTCGGTTACGGCGTGCAATGTTTTGCACGTTACCC GGATCACATGAAACAACATGATTTTTTCAAGTCTGCTATGCCGGAAGGCTACGTGCAAGAAAGAACCATTTTCTTCAAGGAC GATGGCAACTATAAGACCCGTGCAGAAGTTAAATTCGAGGGCGATACCCTCGTGAACCGCATTGAGTTGAAAGGTATCGATT TCAAGGAAGACGGCAACATCCTGGGTCATAAACTGGAGTATAACTACAACAGTCATAACGTTTATATCATGGCAGATAAACAG AAAAACGGCATCAAAGTTAACTTTAAGATCCGCCACAACATTGAAGATGGTTCTGTTCAACTGGCCGACCACTACCAGCAGA ATACCCCGATCGGCGACGGTCCGGTGTTGTTGCCGGACAACCACTATCTCTCGACCCAGAGCGCGTTGTCCAAGGACCCGAA TGAAAAACGTGATCACATGGTTCTGCTGGAGTTCGTCACTGCCGCGGGAATCACCCATGGTATGGATGAACTGTACAAAAAG CTTGGCTGTTTTGGCGG
pBAD30-KAN-tag RFP	ATCGATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTG ATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCC CGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCG GGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGG AAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGA TCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAG TAACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACA GGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTC ATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATC TCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAA TGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTT AAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCATACTTTTCATACTCCC GCCATTCAGAGAAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCA AACCGGTAACCCCGCTTATTAAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCT ATAATCACGGCAGAAAAGTCCACATTGATTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAG CGCATCCTACCTGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGGCTAGTCTtttgtttaactttaagaaggagaA TTCGaGGATCCatggtgagcaaaggcgaagaactgattaaagaaaacatgcatatgaaactgtatatggaaggcaccgtgaacaaccatcattttaaatgcaccagcgaaggcgaaggcaaac cgtatgaaggcacccagaccatgcgcattaaagtggtggaaggcggcccgctgccgtttgcgtttgatattctggcgaccagctttatgtatggcagccgcaccttattaaccatacccagggcattccgg attttttaaacagagctttccggaaggctttacctgggaacgcgtgaccacctatgaagatggcggcgtgctgaccgcgacccaggataccagcctgcaggatggctgcctgatttataacgtgaaaattc gcggcgtgaactttccgagcaacggcccggtgatgcagaaaaaaaccctgggctgggaagcgaacaccgaaatgctgtatccggcggatggcggcctggaaggccgcaccgatatggcgctgaaac tggtgggcggcggccatctgatttgcaactttaaaaccacctatcgcagcaaaaaaccggcgaaaaacctgaaaatgccgggcgtgtattatgtggatcatcgcctggaacgcattaaagaagcggataa agaaacctatgtggaacagcatgaagtggcggtggcgcgctattgcgatctgccgagcaaactgggccataaactgaacggcatggacgaactgtataaaGaAgctTaAGAGAAGATTTT CTAGAAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGG TCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAG GGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACG CTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCAGGACGC CCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTatcttttcta cggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacgggaaacgtcttgctcta

	ggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaaca tggcaaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgat ccccggcaaaacagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtttgtaattgtccttttaacagtgatcgcgtat ttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctcaccggatt cagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcgg tgagttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagtttttctaagaattaattcatgagcggatacatatttgaatgtatt tagaaaaataaacaaatagggttccgcgcacatttccccgaaaaCTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACGCGCCCTGTAGC GGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCG CTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTA GTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTTGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTT CGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTTGAACAACACTCAACCCTATCTCGGGCTA TTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTT TAACAAAATATTAACGTTTACAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTG AGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGC TGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAA CTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGC ACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTCAGGCATTTGAGAAGCACACGGTCACACTGCTTCCGGTAGTCAA TAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACCGGGTCGAATTTGCTTTCG AATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAA TTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGCAT GATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCGCTAGCGGAGTGTATACTGGCTTACT ATGTTGGCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATAT GTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACG AACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCC ATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATA CCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCC GCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCA GCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGAC TGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTT TTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTGCT CATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCG GTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCTGCTCATGTTTGACAGCTTATC

ACCESS MICROBIOLOGY
an open research platform

PEDAGOGY
Graham *et al.*, *Access Microbiology*
DOI 10.1099/acmi.0.000472.v3

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Education-based grant programmes for bottom-up distance learning and project catalysis: antimicrobial resistance in Sub-Saharan Africa

Chris L. B. Graham[1,2,*], Harry Akligoh[3], Joy King Ori[1], Gameli Adzaho[1], Linda Salekwa[4], Patrick Campbell[1], Courage K. S. Saba[5], Thomas E. Landrain[1] and Marc Santolini[1,6]

Abstract

International development and aid are often conducted through the allocation of funding determined by decisions of non-locals, especially in the west for those in the global south. In addition, such funding is often disassociated from local expertise, therefore providing little long-term developmental impact and generating distrust. This is particularly true for conservation, as well as environmental and educational programmes. We hypothesize that by granting local people the educational tools and the necessary funding to develop their own projects through the use of an applicant-driven peer-review approach, it is possible to relocalize the decision-making process to the programme participants, with the potential to generate and select more relevant projects with developmental outcomes of higher quality. Here we created an online curriculum for antimicrobial resistance (AMR) education that was followed by 89 participants across Ghana, Tanzania, Nigeria and Uganda. We then created an open research programme that facilitated the creation of eight *de novo* projects on AMR. Finally, we organized an applicant-driven grant round to allocate funding to the 'Neonatal Sepsis in Nigeria' project to conduct a pilot study and awareness campaign. This work opens perspectives for the design of frugal educational programmes and the funding of context-specific, community-driven projects aimed at empowering local stakeholders in the global South.

DATA SUMMARY

Supplementary Material files can be retrieved using the following link on figshare: https://doi.org/10.6084/m9.figshare.20418828.v1. This includes a pdf of the form used for applicant selection, supplementary information and method notes, the reviewer form .pdf file and the location data determined by Microbis[1] and Geocoder. Included in this Figshare is FigS1, which summarizes this data.

INTRODUCTION

Antimicrobial resistance (AMR)

Harmful bacteria cause more deaths per year worldwide than either AIDs or Malaria [2]. For most of these bacteria, effective antibiotic treatments are available. This has been true since the advent of penicillin, and the following decades uncovered many more antimicrobials capable of saving human life. However, globally bacteria have become resistant to many of our most effective and widely available antibiotics. This resistance is not yet adequately studied or monitored, with problem areas in the global south

Received 05 August 2022; Accepted 29 January 2023
Author affiliations: [1]Just One Giant Lab, Paris, France; [2]University of Warwick, Coventry, UK; [3]Duplex Bioscience LBG, Accra, Ghana; [4]Mbeya University of Science and Technology, Mbeya Tanzania; [5]University for Development Studies, Tamale, Ghana; [6]Université Paris Cité, INSERM, U1284, F-75004 Paris, France.
***Correspondence:** Chris L. B. Graham, christopherlbgraham@live.co.uk
Keywords: AMR; community science; curricula; funding; grant review; training.
Abbreviations: AMR, Antimicrobial Resistance; API , Application Programming Interface; JOGL, Just One Giant Lab; JPIAMR, Joint Programming Iniative on Antimicrobial Resistance; LMIC, Low-Middle Income Countries; MOOC, Massive Open Online Course; MUST, Mbeya University of Science and Technology; NGO, Non Governmental Organisation; ST, Sequence Types; UDS, University for Development Studies -Kumasi, Ghana.
A previous version of this paper is available at AfricRxiv's ScienceOpen repository: https://www.scienceopen.com/hosted-document?doi=10.14293/111.000/000031.v1.
A supplementary figure and supplementary information are available with the online version of this article.

especially comprising the overuse of antibiotics in food crops and livestock, and their overspill into water sources and effect the local environment and peoples, and allowing for spread of resistance, constituting a *One Health* problem. Natural selection of antibiotic-resistant strains in these antibiotic-rich environments gives an advantage to new and emerging resistant strains in rural and urban settings globally, as well as clinical settings where antibiotics are more commonly used and required for medical treatment [3]. Soon this will inevitably lead to an antibiotic shortage.

LMIC contextualized by AMR

Low and middle-income countries (LMICs) bear the highest burdens of infectious diseases with potentially the least resources, and also have limited data on the epidemiology and burden of antimicrobial resistance [4, 5]. A 2017 WHO report highlighted the gaps in information on pathogens of major public health threats. The report also emphasized the lack of high-quality data and how it limits our ability to assess and monitor trends of resistance worldwide [6]. Congruent to earlier reports, the laboratory identification of AMR and bacteria still rely on the use of substandard phenotypic techniques coupled with poor laboratory information management systems seen mostly in LMICs such as within Africa. Aside from this, other barriers such as a poor diagnostic infrastructure, limited staff capacity and training as well as questionable quality-management systems compound the problems of tracking down resistance [7, 8].

The COVID-19 pandemic has underscored the importance of new paradigms for testing and researching infectious pathogens. The use of pathogen genomics has provided information to scientists in record time to help guide public health strategies. Hence recent efforts in Nigeria saw researchers use sequencing to study and characterize sequence types (STs) for AMR bacteria causing hospital-acquired infections [9]. Datasets collected were made available to their AMR National Coordinating Centre (NCC) in an effort towards overcoming the current knowledge and fragmented data gap. Gains made by countries in Europe in the fight against antimicrobial resistance have largely been due to the availability of data at several levels and education, which is mostly absent in LMICs [10]. Therefore, in the face of growing global health threats, where individualism is a danger to gains made, LMICs particularly those in Africa are optimistic about initiatives that aid their developmental agenda in strengthening existing but weak or non-existent infrastructure to tackle health threats like AMR.

Education and grassroots project creation programmes

There have been efforts in the past to improve the infrastructure of LMICs, especially in Africa, however the majority of these programmes have been initiated by projects coming from a western nation with control over project management at all levels [11]. This has recently begun to change, with an emphasis towards projects that give natives project ownership and motivation, creating sustainable projects without outside interference.

Therefore, similar to this emphasis on bottom-up – as opposed to top-down – change, a concept of 'smart people before smart cities' has become dominating sentiment, with the creation of opportunities for residents, in particular educational ones, being a priority that must go hand in hand with infrastructure improvements [12]. This capacity building in people and facilitation of independent work therefore must take priority and allow for new decentralized teams on the continent to work on global problems using their own motivation and methods. This has been shown to be more effective at fostering long-term change [13].

Online learning and capacity building are a priority over travel

During the COVID-19 pandemic, international programmes of educational support were encouraged to make educational resources available online and disseminated locally where possible. This has shown to work well in some settings such as high schools [13, 14] as well as within new online infrastructure, such as the open source and African built 'Voltschool' [15]. Similarly the funding granted for these international programmes can be more efficiently spent on capacity building rather than on flights. Prior to this study, the researchers conducted another study focusing on an agile funding allocation scheme to projects determined by applicants, and found that the correlation between reviews, with reviewers as applicants was similar to non-applicant reviewers in other schemes, and of high efficiency [16]. Following this scheme, here we used the travel spending funds towards a microgrant allocation to allow the applicants within our curriculum cohort to determine a project within their cohort that should receive seed funding for capacity building beyond education, as well as motivate their own ideas with *de novo* project creation.

Goal of study

Just One Giant Lab (JOGL) is an NGO for collaborative science. In partnership with Hive Biolab, Kumasi, Ghana, University for Development Studies (UDS), Tamale, Ghana and Mbeya University of Science and Technology (MUST) Tanzania, and with the support from the Microbiology Society, JOGL hosted a grant round to catalyse project creation by participants in response to the need for increased capacity building and a focus on education organized a virtual training series on antimicrobial resistance (AMR). We also provided a training on the use of an AMR tracking app 'microBIS.io' for a pilot result. The goal of the workshop was to build the capacity of university students and laboratory technicians on the latest trends in AMR screening and testing, as well as increase AMR stewardship and empower them with the resources to become agents of change in reversing the spread and negative effects of AMR in Africa.

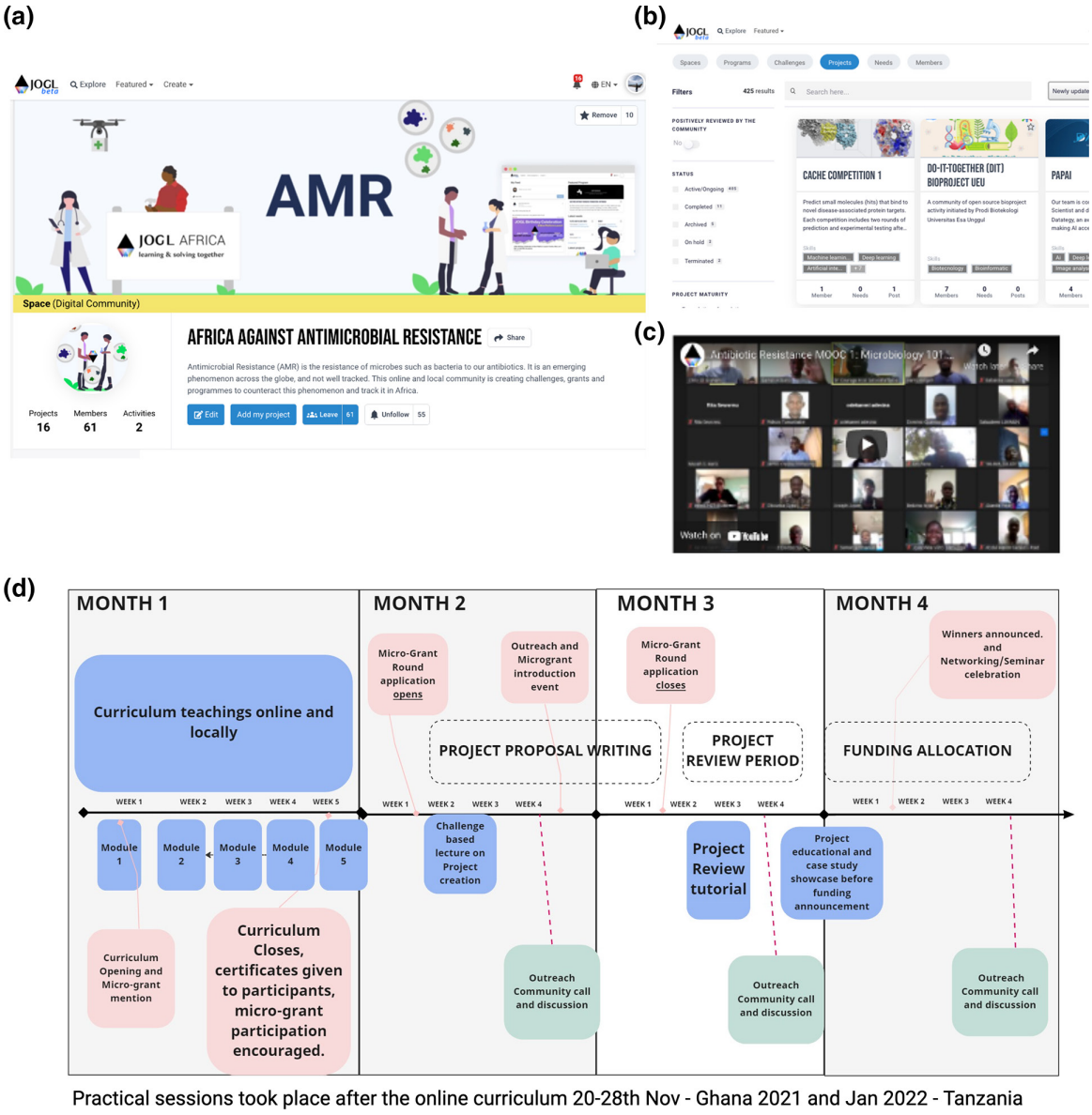


Fig. 1. Online project creation and networking. (a) Programme page in the JOGL social innovation network used to allow for project creation in an open-source accessible website. (b) Project search and display function. (c) Videos of modules were recorded and displayed on the platform. (d) Timeline of initiative: pink – microgrant related deadlines; green/light green – community call; blue – curriculum and lectures.

METHODS

All participants in the AMR workshop were invited to join the 'Africa against AMR Community' space on the JOGL collaborative network platform [17, 18] as well as a special Africa AMR channel within a community workspace on instant messaging service 'Slack' [19] to enhance networking and collaboration.

The project team developed a curriculum to cover important aspects of microbiology, AMR, and innovative approaches to tackle the challenge Finally, at the end of the programme a grant round was conducted. In person practical sessions were held separately from the main programme, to coordinate with term times for participating academics, in Tanzania and Ghana, asynchronously (Fig. 1).

Geotagging and surveying

Geotags of the coarse-grained information (town level) participant locations were pulled from the JOGL website and converted using Geo-Coder [17, 20], and the names and identifiable aspects of members were removed before geotagging. Upon signup to the platform, consent is given for location tracking, in the user terms and conditions.

Outreach and recruitment of participants

Participants were contacted through social media channels such as Twitter, Facebook and LinkedIn to advertise a free curriculum and grant round, in addition to on our constructed site. Existing media accounts with 4000+ followers allowed for sharing between individuals through connections.

Curriculum

The course was composed of two parts – an online workshop series and onsite practical sessions. Each lecture was created 1 week ahead of schedule. The online workshop consisted of five modules delivered as weekly Zoom webinars from 21 September to 19 October. These sessions were recorded and the videos were uploaded on YouTube [21]. Practical sessions on microbial identification and sensitivity testing [22] were scheduled for interested participants in Kumasi, Ghana (Hive Biolab), Ghana, and Mbeya, Tanzania (MUST). The details of the curriculum can be found in the Supplementary Material, available in the online version of this article. No formative assessment was made, however attendance was kept each session through a form.

Online web-space

For presenting the modules within the 'Massive Open Online Course' (MOOC), and the grant review round, we used a citizen science and project creation network called 'Just One Giant Lab' [17]. This allowed participants to create their own projects in addition to taking part in our syllabus available here [21]. The social network aspect in a focused programme allowed for sharing and openly disseminating the curriculum internationally. These were all created with the help of the design team at JOGL.

Local collaborator selection

Coordinators of the programme local to Ghana and Tanzania were found through the JPIAMR portal for collaborator finding, as well as through existing shared contacts. They were reached out through email and the collaboration was initiated at the grant writing stage, with further participants and partners invited after the grant award by our funders.

Administration

Administration was conducted first through a google form (Data Summary) for the collection of emails. The resulting email spreadsheet, which included demographic information, was then used to liaise for the rest of the course.

The grant round leveraged the JOGL interface to facilitate contact with applicants. Using the JOGL application programming interface (API), applicant data was collected so that anyone who joined the grant round were then emailed with any updates through the course of the endeavour. Administration within the consortium of collaborators between MUST, University of Warwick, Hive Biolab and JOGL was conducted over Slack channels, emails and Zoom meetings with no in-person meeting. There was one public slack channel, a private administrative channel, and a questions and answers channel. Participants were invited to the programme through social media posts, word of mouth, and by emails to admins of universities and makerspace laboratories in the region, in a campaign that began 2 months prior to the beginning of the course, and involved a sign-up process.

Network of skills creation

Similar to Masselot *et al.* [23], participants filled in their professional background, skills and employment status on signing up to the JOGL platform. In order to better understand how skills were related across participants, we used a network approach to assess similarity between skills and got further insights about the global diversity of the community. In this network approach, each declared skill is a node and the considered skills as linked if they co-occur in a participant. Links are then weighted by the number of participants they co-occur in. Gephi 0.9.3 was used to represent the network in Fig. 2 [24], its modularity algorithm was used with default parameters to compute communities.

AMR in africa grant round

JOGL organized the 'AMR in Africa Grant Round' to support one research, innovation or education project tackling AMR in Africa. The call for proposals ran for 1 month, from 18 November to 18 December 2021. We used a proposal template as attached (Data Summary) with a focus on the antimicrobial resistance outcomes to encourage a formal application approach, then used the following format for administration (Fig. 3). Applicants were required to mark at least four other applicants, therefore giving

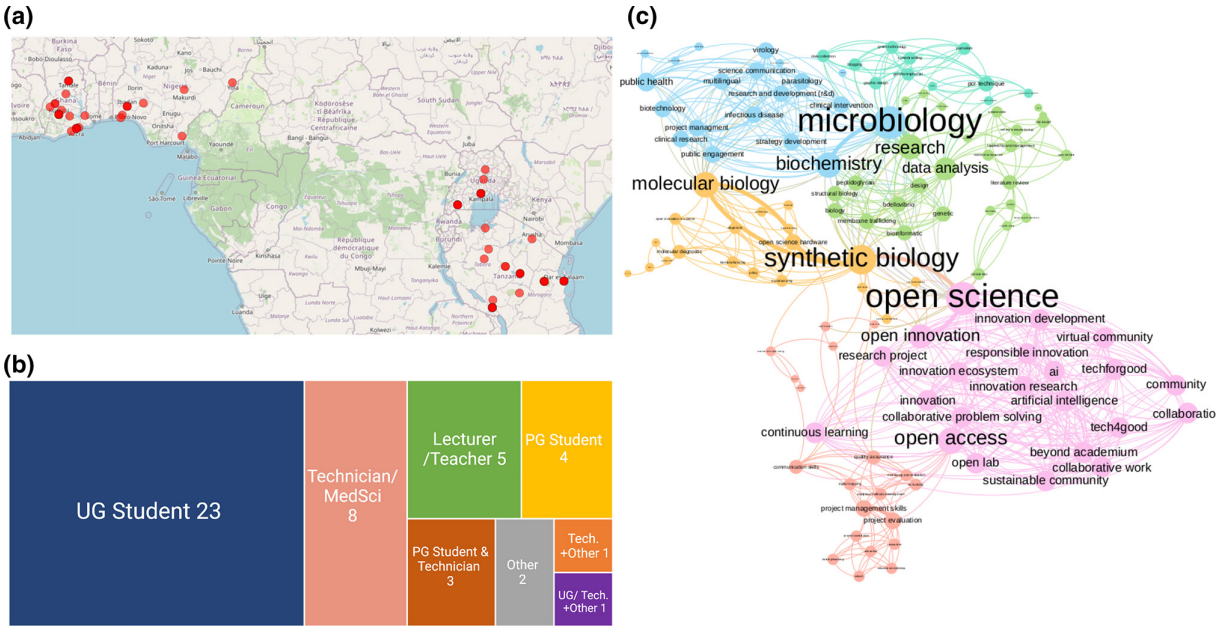


Fig. 2. Location and profession of participants of virtual curriculum in Africa. (a) Locations derived from geocoded data from form input. All participants are shown as red dots by city. (b) A visualization of the professions who filled in the post-curriculum form. (c) Network of skills in the cohort of participants, taken information given upon signup to the JOGL platform. Skills are connected if a participant has both of them in their profile. The thickness of connections indicates number of participants sharing two skills and the size of nodes indicates the number of neighbouring nodes. Gephi 0.9.3(22) was used to represent the network, as described in Methods. Colour indicates the community, assessed using modularity.

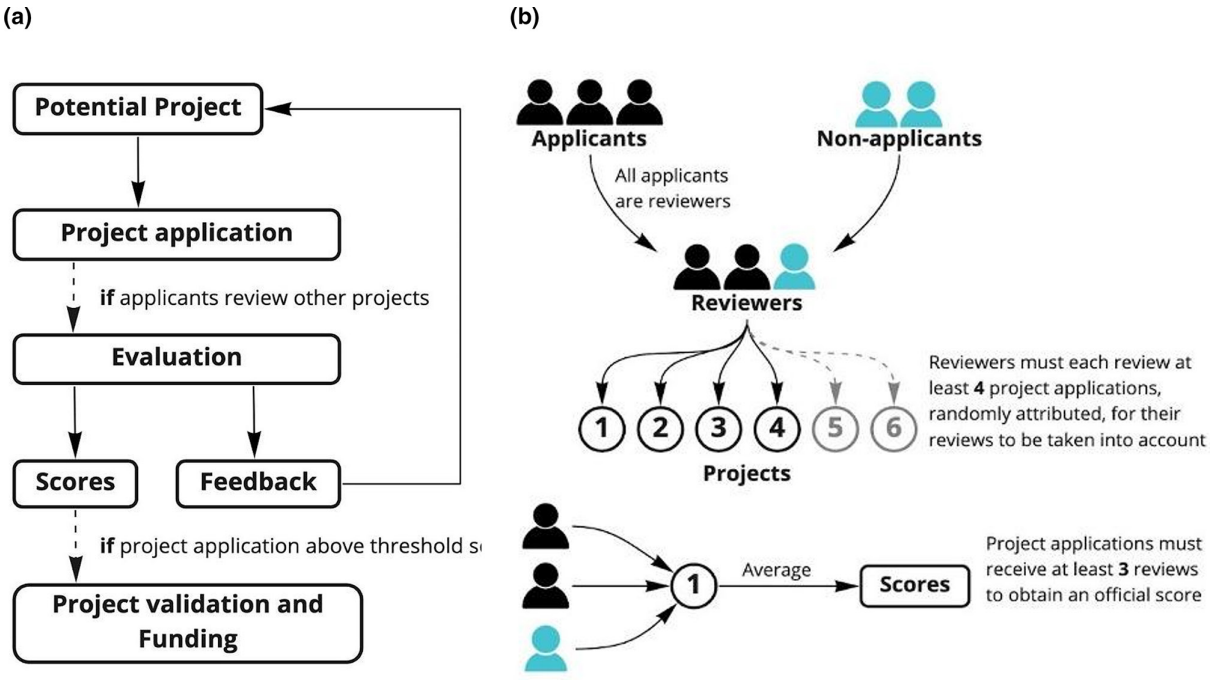


Fig. 3. Grant application process. (a) Flow of community review and grant application. (b) Applicants review one another allowing full scalability. (Adapted from Graham, Landrain and Santolini *et a*. 2022) [16].

a democratic and local outlook on the most relevant and innovative solutions by curricula through experts and existing creators. This followed a community review methodology as used in the OpenCOVID19 programme 2020-2021 [25].

RESULTS

Background of participants and trainers

The virtual training attracted 89 participants (whom attended at least two online sessions) across the African continent from diverse backgrounds; such as laboratory technicians, students, researchers and innovators (Fig. 2). More than 50% of the participants were undergraduate students, while the other 50% comprised lab technicians/scientists, postgraduate students, and teachers/lecturers also. Most of the participants came from Ghana, Tanzania, Nigeria and Uganda as per the details filled in the application form. The results of the form indicated our shared connections were greatest in Ghana, Nigeria, Uganda and Tanzania as visualized in Fig. 2. This is likely as our collaborating partners were from Tanzania and Ghana, who also shared the opportunity within their networks, notably from a range of cities across the aforementioned countries (Fig. 2a). The distance between participants, normally inhibitory to practical or in-person instruction, was overcome using online teaching and community building. The community leveraged through the creation of a curriculum was that of a network of participants with some relevant skills already (Fig. 2c) who could complement each other's strengths during project creation: those with more experience with open science in general, creating a community of open science advocates (including JOGL staff) (pink), but also project developers (orange) and biologists (green and blue), with some having multiple of these skills, indicated by degree of connection between groups. Most notably synthetic biologists who joined the grant and education programme had the most experience in open science.

Participants engagement and post-MOOC thoughts

Whilst 98 attendees attended the first online seminar, this over-time reduced to 61 participants by MOOC5 who could attend the seminars live, with the recordings later available on YouTube, however 119, 89, 69, 50 and 34 of the attendees attended at least one, two, three, four and five of the sessions live (Fig. 4c), respectively, suggesting attendees joined various lectures live depending on relevance or convenience. Despite the drop in attendees in the Zoom call, the overall feedback shared by the participants in a post-workshop survey was positive (Fig. 4). Roughly 46 of the 47 survey respondents rated the workshop 4 or 5 (1–5 scale) (Fig. 4a). Similarly, over 44 of the respondents indicated that the training had a high positive impact on their current work (ratings of 4 or 5) (Fig. 4d). Of the modules taught during the online sessions, less specialist introductory aspects such as MOOC 2, MOOC 3 and MOOC 1 (Supplementary Material – Methods and notes) were deemed to be of the highest relevance to the participants trained. Importantly, there was a large number of applicants who continued to apply for the grant round in AMR project creation, suggesting curricula are a positive means for *de novo* project creation and onboarding (Fig. 4b).

Some participants shared additional content about their experiences during the five-part online training:

> "I had 'locked' myself but the course now opened my mind. It woke me up to see the ever increasing challenge of AMR."
>
> "The training was well organised, precise and straight to the main point."
>
> "The modules were very relatable to my work and also improved my knowledge on AMR"
>
> "It was lively at all times and encouraged participatory discourse".

They also made some suggestions for improvement, some of which have been captured below:

> "You may please consider a session to educate non-specialists because, I am not working with a medical institution and having difficulties engaging fully"
>
> "Engage more stakeholders for example political policymakers and other African country health ministries"
>
> "Involve practical sessions for participants from other countries outside Ghana and Tanzania"
>
> "I think you are doing a great job but I think what needs to be improved should be the technical aspect of it because we lost connection at a point during one module.'

Participants' comments on modules they found relevant (related to informational content and not developmental or project-related content) may be due to the fact that the workshop was dominated by undergraduate students and early career lab technicians, who identified concepts they could apply in their day-to-day work. However, the modules on antimicrobial stewardship and digital mapping are equally relevant in developing additional leadership, coordination and digital literacy skills required for the next level of careers. Comments in particular on ministerial and governmental cooperation to improve outcomes and network are particularly significant changes that future programmes could integrate.

In summary, the AMR workshop and curriculum as a means to build new projects were carried out to a satisfactory level with future iterations likely to alter the curriculum based on signup and topic interest, based on the team's experience organizing

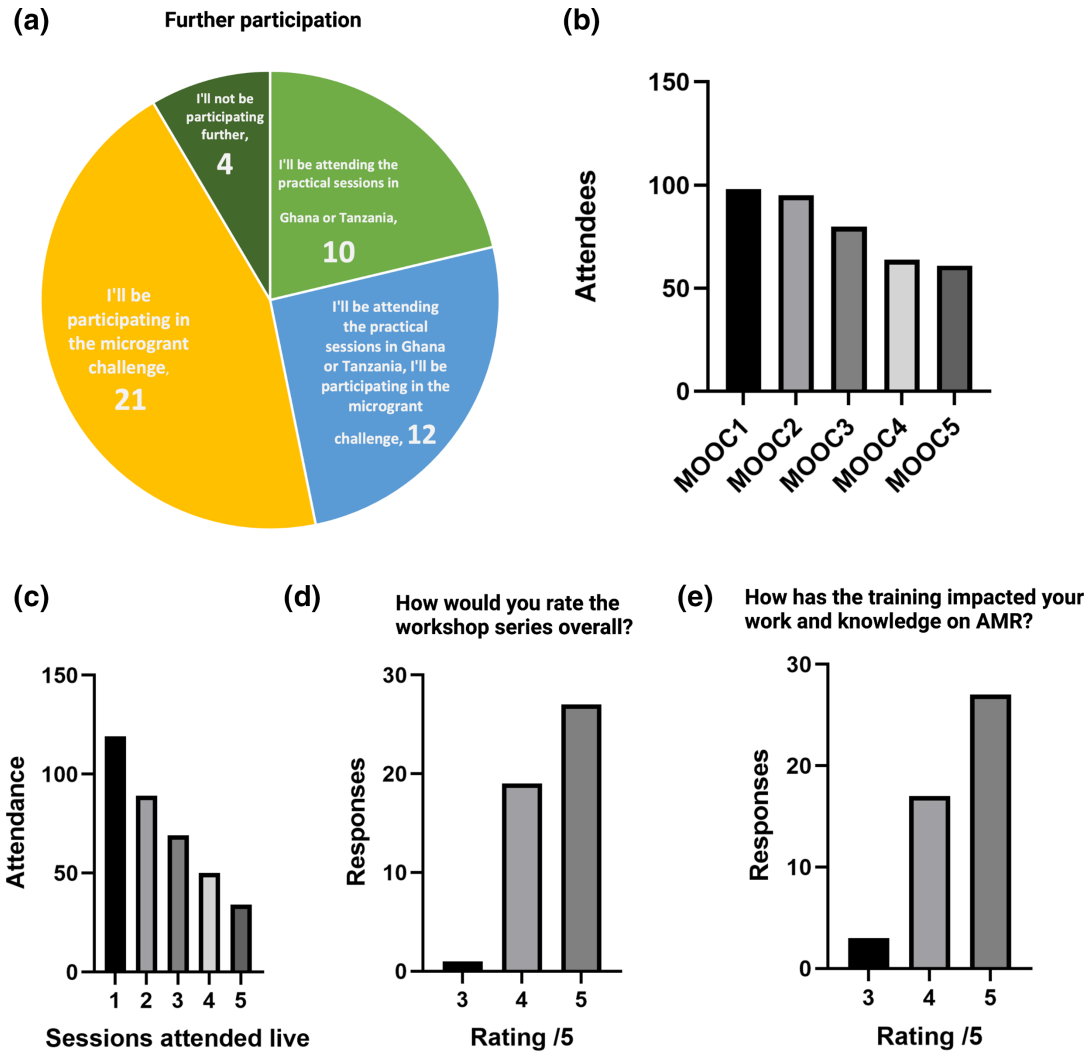


Fig. 4. Participant feedback of pre-grant curriculum. (a) Participation, practically, in grant challenge or no further participation. (b) Attendance of online MOOC lectures. (c) Attendance of live lectures. (d) Rating of curriculum [2–6], according to participant feedback. (e) Impact of programme on future work [2–6].

and delivering the programme, and the feedback received from the participants (Fig. 4) . There is a great potential to scale the workshop and its associated activities to reach more students, young professionals, and general science enthusiasts in Africa due to its mediation through online platforms.

Grant and project creation outcomes

After the curricula, in total eight projects [17] applied to the grant round from a subset of the curriculum takers, but also from individuals who saw the grant round advertised in their network, with a geographical distribution similar to the participant distribution (Fig. 3/Fig. 5a). In order to ensure fairness after submission and review and prevent gamification, the reviewers' scores across the questions in each form were normalized to their overall average review mark across their five reviews using the community review method. The distribution of scores per question shown by the heatmap indicates independence of questions asked in the form, and allowed for detection of odd reviewer behaviour [16]. (Fig. 5b). The winning team with the highest average review score for the 1000 euro grant was 'Neonatal Sepsis in Nigeria' [26]. The other seven teams that applied to the grant, or formed as a consequence of its existence were encouraged to apply to later or other grants. As a result of the programme there is still an active community of collaborators on the community 'slack' channels whom conduct meetings on science policy on a monthly basis in preparation for future programmes, and this initiative acted as a catalyst for its community growth, we are currently writing collaborative grants with many of the participants. This shows that with little

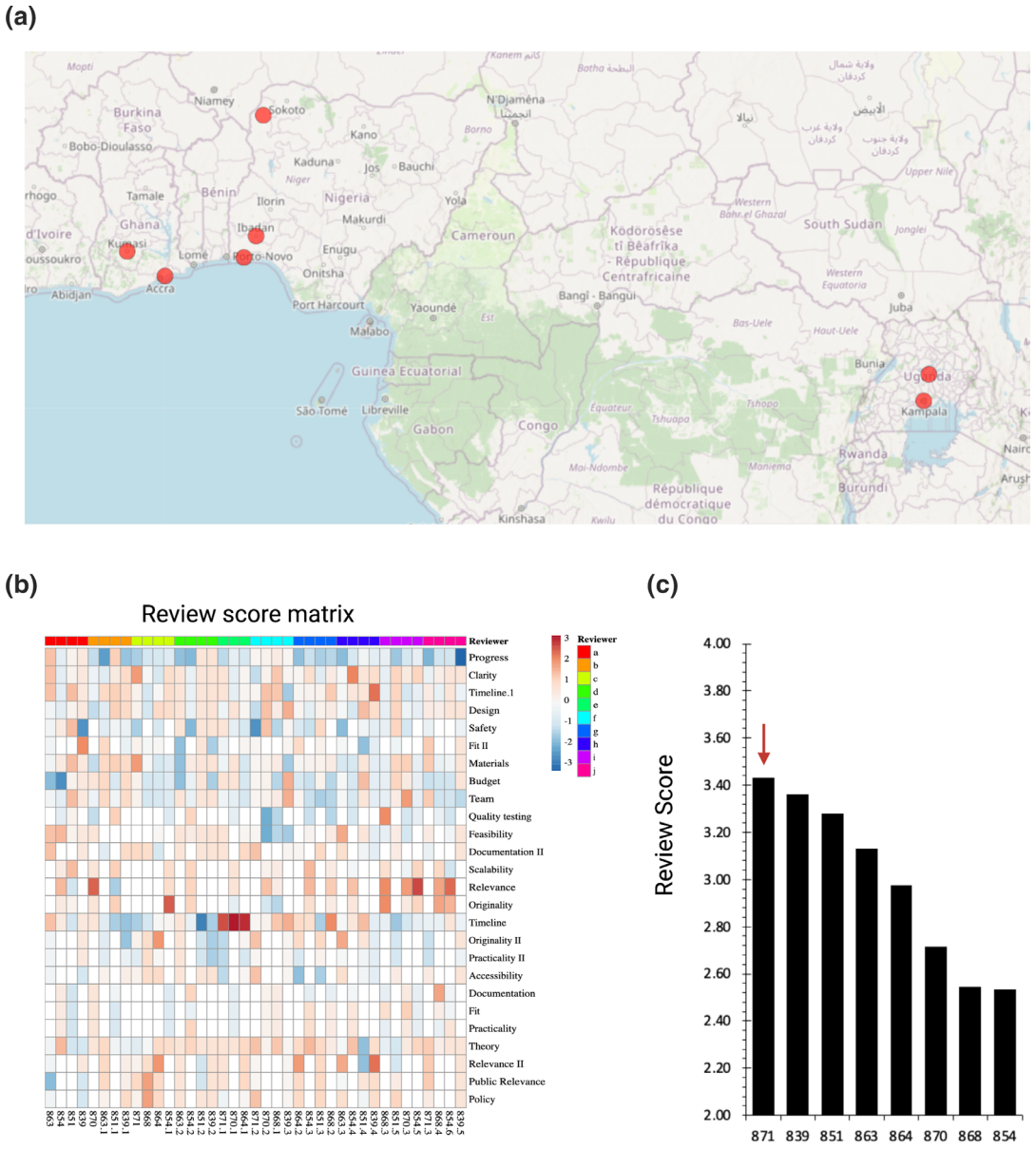


Fig. 5. Project catalysis and scoring by applicant-based peer review. (a) Location of new projects created [17, 20], (b) Review behaviour as visualized by heatmap of scores per question red to blue, high scoring to low scoring respectively, determined by Z score, tree of review similarity adjacent. (c) Review score rankings by applicants. Winning project is denoted by a red arrow.

organization and a small prize pot, a significant collaborative output can be achieved with micro-grants if accompanied by training and encouragement.

Network building and community science considerations

In addition to the projects created, by creating a community internationally through the web, the collection of information for database-related programmes can be improved. A collaborator who took part in the education aspect of the programme, asked participants working in hospitals to gather data of location, and resistance of bacterial strains, which allowed the team to begin the creation of an antimicrobial resistance location database. Further success in reaching out to participants would allow an international map, and service not possible in on the ground programmes. We have supplied a sample of this data for illustrative purposes (Fig. S1). In future programmes, such communities can be leveraged for similar community science outputs such as data gathering.

Fostering of collaboration among organizers and participants

Some of the participants met virtually for the first time because of the project. Throughout the development of the courses and practical sessions, there has been exchanges of good practices among participants who are from different countries and continents. The high level of collaboration shown throughout the project paved the way for future joint proposal writing for a bigger grant to curb the menace of AMR worldwide, and an active online science sharing community. Some participants have networked during the workshop and we hope this project provided the right platform for early career scientists in AMR to share their experiences with one another to tackle issues of AMR in Africa.

CONCLUSION

Through the use of a local network of contacts and the 'Just One Giant Lab' web platform, we were able to create an online curriculum for AMR and deliver this to a diverse audience-based predominantly around Uganda, Nigeria, Tanzania and Ghana. We hosted a programme fostering the creation of eight *de novo* projects and organized a community review round that funded one project selected by applicants and participants of the online curriculum. As such, we performed a hybrid model of online curriculum and grant allocation. We worked with local partners and participants as community drivers, with democratic and local values upheld by decisions of applicants themselves, whilst encouraging a new collaborative project-based perspective in antimicrobrial resistance for medical staff, technicians, students and others.

The techniques of applicant-driven grant decision-making, and online delivery of free educational resources as a pathway for science and awareness funding, provided some success across multiple sub-saharan African nations on a small scale. This method with further funding could help foster development in a cost-efficient manner and at an international scale. This programme and its scale would not have otherwise been possible in person due to travel costs, yet through an online medium, it allowed for capacity building with an educational programme as well as local project creation and curation. This technique opens perspectives to design frugal approaches, allowing a programme team to locally empower individuals in context-specific project catalysis and science education. The authors would encourage other similar initiatives, and also find it likely that if the initiative centred on antimicrobial resistance was possible in areas as culturally diverse as the regions presented it would be replicable in other consortiums nationally and internationally elsewhere, especially the global south.

Funding information
The project was funded by the Microbiology society under GA002608. CG is supported by by the BBSRC through studentships [BB/M01116X/1] and by the UK antibiotic research charity ANTRUK ANTRUK [ANTRUK_SRG_05–2021].

Acknowledgements
The facilitators for the workshop were Professor Courage Kosi Setsoafia Saba (University for Development Studies, Ghana), Dr Linda Salekwa (Mbeya University of Science and Technology, Tanzania), Dr Beverly Egyir (Noguchi Memorial Institute for Medical Research, Ghana), Harry Akligoh (Kumasi Hive and Duplex Bioscience LBG, Ghana), Chris Graham (University of Warwick, UK and Just One Giant Lab), and Gameli Adzaho (Just One Giant Lab). Dr John Moat of University of Warwick also made inputs into the content of one of the modules. We thank the Microbiology Society for the International Development Fund grant (GA002608) and IFD for funding JOGL Africa.

Author contribution
C.G., H.A., G.A., L.S. and C.S. were behind the initial conception of the project and funding acquisition, T.E.L. provided supervision and advice to the team during pedagogy . H.A. designed microbis.io a data tracker for hospital technicians. C.G. and J.K. wrote the manuscript. Data analysis and Microgrant administration was provided by C.G. Workshops were created and delivered by C.G., H.A., G.A., L.S. and C.S. The online environment was designed by C.G. with support of T.E.L. Networking was led by G.A. G.A., P.C., T.E.L., M.S. and H.A. provided editorial and background information for the final manuscript. M.S. advised CG throughout on manuscript preparation, quality control and submission.

Conflicts of interest
The authors declare no conflict of interest

References

1. Digital Mapping and Using MicroBIS.io; (n.d.). https://app.slidebean.com/p/5is190kwng/Digital-Mapping-and-Using-MicroBISio#1 [accessed 6 January 2023].
2. Murray CJ, Ikuta KS, Sharara F, Swetschinski L, Aguilar GR. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* 2022;399:629–655.
3. Hendriksen RS, Munk P, Njage P, van Bunnik B, McNally L, *et al*. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat Commun* 2019;10:1124.
4. Iskandar K, Molinier L, Hallit S, Sartelli M, Hardcastle TC, *et al*. Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. *Antimicrob Resist Infect Control* 2021;10:63.
5. O'NeilJ. Antimicrobrial resistance. tackling crisis health wealth nations. *AMR Rev* 2014.
6. World Bank. Drug-resistant infections: a threat to our economic future. *World Bank* 2017.
7. Seni J, Mapunjo SG, Wittenauer R, Valimba R, Stergachis A, *et al*. Antimicrobial use across six referral hospitals in Tanzania: a point prevalence survey. *BMJ Open* 2020;10:e042819.
8. Sneddon J, Afriyie D, Sefah I, Cockburn A, Kerr F, *et al*. Developing a Sustainable Antimicrobial Stewardship (AMS) programme in Ghana: replicating the scottish triad model of information, education and quality improvement. *Antibiotics* 2020;9:636.
9. Okeke IN, Aboderin AO, Egwuenu A, Underwood A, Afolayan AO, *et al*. Establishing a national reference laboratory for antimicrobial

resistance using a whole-genome sequencing framework: Nigeria's experience. *Microbiology* 2022;168.

10. Gabriel S, Manumbu L, Mkusa O, Kilonzi M, Marealle AI, *et al*. Knowledge of use of antibiotics among consumers in Tanzania. *JAC Antimicrob Resist* 2021;3:dlab183.
11. Yoshida K, van der Walt JL. The policy-implementation-results linkage for education development and aid effectiveness in the *Education 2030* era. *Com J Comp Int Educ* 2018;48:39–55.
12. Tonar R. Smart Cities Are Built By Smart People, Not Smart Things. Forbes [Internet]; (n.d.). www.forbes.com/sites/ellistalton/2019/07/09/smart-cities-are-built-by-smart-people-not-smart-things/
13. Narayanan S, Rajan AT, Jebaraj P, Elayaraja MS. Delivering basic infrastructure services to the urban poor: a meta-analysis of the effectiveness of bottom-up approaches. *Utilities Policy* 2017;44:50–62.
14. Joseph Paschal M, Mkulu DG. Online classes during COVID-19 pandemic in higher learning institutions in Africa. *GRHE* 2020;3:1.
15. Voltschool Education. Internet; (n.d.). https://voltschool.com/
16. Graham CLB, Landrain TE, Vjestica A, Masselot C, Lawton E, *et al*. Community review: a robust and scalable selection system for resource allocation within open science and innovation communities. *F1000Res* 2022;11:1440.
17. Jogl T. Africa Against AMR Community Space [Internet]; 2021. https://app.jogl.io/space/amrafrica
18. Masselot CM, Tzovaras BG, Graham CLB, Finnegan G, Jeyaram R, *et al*. Co-Immune: a case study on open innovation for vaccination hesitancy and access. *Public and Global Health* 2021. DOI: 10.1101/2021.03.29.20248781.
19. Slack. Slack - About; (n.d.). https://slack.com/intl/en-gb/about [accessed 6 January 2023].
20. Geocoder. GPS Visualiser Geo-Coder; 2022. www.gpsvisualizer.com/geocoding.html
21. Graham C. Africa MOOC Playlist; 2021. www.youtube.com/watch?v=r8fJ9bHXIgs&list=PLGNQcqWTh7vgqBockgdGnwPxI3NH9bnkG
22. EUCAST guidelines website. Eur Comm Antimicrob Susceptibility Test; (n.d.). www.eucast.org/
23. Masselot C, Greshake Tzovaras B, Graham CLB, Finnegan G, Jeyaram R, *et al*. Implementing the co-immune open innovation program to address vaccination hesitancy and access to vaccines: retrospective study. *J Particip Med* 2022;14:e32125.
24. Home gephi/gephi Wiki. GitHub; (n.d.). https://github.com/gephi/gephi [accessed 6 January 2023].
25. OpenCOVID19 Project list (JOGL). JOGL, Online; 2021. https://app.jogl.io/program/opencovid19?tab=projects
26. Kemiki O. Carbapenem resistance in Neonatal sepsis in Nigeria; 2021. https://app.jogl.io/project/871

Peer review history

VERSION 2

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000472.v2.1

Georgios Efthimiou; University of Hull, Biomedical Science, Hardy Building, Cottingham Road, UNITED KINGDOM, Hull

Date report received: 29 January 2023
Recommendation: Accept

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. All comments by the reviewers were satisfactorily addressed.

Author response to reviewers to Version 1

Dear Reviewers and Editor,

Thank you for your kind comments and for your time reviewing this work, we hope you had a good new year. We agree with your comments and have changed the manuscript and figures to align as much as possible with your own thoughts and attach our responses to each point made.

Reviewer 1: Morgane Feeney

Major points:

1. It is not clear from the text whether or not there was any formative or summative assessment associated with the MOOC (apart from the grant round that followed). Perhaps the authors could make this point clearer. - We have added that there was no formative or summative assessment associated with the MOOC.

2. It is not entirely clear to me from the timeline in Figure 1, when the in-person practical sessions were held - I think it might be helpful to show this. - We have added the dates for this on the figure, they were held asynchronously due to term time constraints.

3. If the authors have this data, it would be perhaps helpful to show the number of participants attending each session, and/or accessing the YouTube recordings afterwards. Was engagement sustained throughout all sessions of the course? - We have added the attendees list for each MOOC session, which shows that we sustained attendance, although typically this dropped over time.

Minor points:

1. line 65, "Recent efforts in Nigeria…." - a citation would be helpful here for readers who are unfamiliar with this work -Thank you we have added this reference

2. minor editing for grammar needed - for example, line 75 "There has been efforts" should be "There have been efforts" -added

3. lines 93-95, perhaps briefly summarise the results of the other study? - We have now summarised the study results very briefly here.

4. Figure 1, the colour code being used in panel d is not immediately self-evident, perhaps this could be explained in the legend or with a key. - We have altered the colour code to be more self-evident, and labelled the diagram in the figure legend.

5. Figure 1, much of the text is too small to read easily when the figure is viewed at 100% magnification. Perhaps this figure could be made a little more readable. - We have made some text larger, however in the case of the screenshots of the app, we cant make these any bigger without taking the focus away from the science and onto the app, which we would prefer not to do. We do hope they serve as a reference to see what the platform looks like generally for a user, so other initiatives may also use a platform similar.

6. Methods - use of Gephi 0.9.3 could be a little clearer and more detailed, to allow researchers to reproduce these results - Apologies, we have now added a reference here, we have also elaborated further. Please see 'Network in skills creation' within the methods section.

7. Fig 3a - "intensity indicating number of participants from the region" - the intensity differences are not very easily distinguishable on the map as presented, perhaps these could be made clearer and/or a scale added to help readers understand these data. - We have removed reference to intensity, as infact its an overlay and therefore no key could indicate number dependent on intensity.

8. Fig 3c - some of the smaller nodes are unreadable, perhaps the font sizes could be made larger - We have made the diagram larger, so the majority of text is readable.

9. Fig 4 - the use of similar colours (blue and yellow) across the pie chart and the bar graphs may imply connections between these data/groups - however, they don't appear to be connected. Perhaps a different colour palette for the bar charts would be helpful. - In reformatting this figure to add attendance, we have reformed this bar chart, therefore there is no colour connection any longer.

10. Fig 5b - unclear what is meant by the "tree" on the left of the heatmap, or how this was generated? Is the tree necessary at all? - The tree was a similarity between questions tree, however was not relevant, therefore we have cropped this off the heatmap.

11. Fig S1 - tryptophan resistance? Perhaps the authors could explain what is meant by this, or provide a little more detail about the strains being shown? - This is sample data from the participation of hospitals on Harry Akligoh's Microbis platform. It is illustrative of a larger collaboration or network allowing for a larger and more meaningful dataset.

12. For the parts i-iii in panel A, it would be helpful to explain what these are in the figure legend. - These are now labelled.

13. A little more detail on the community building (i.e. how the Slack channels were organised, to promote effective communication) might be helpful for readers wishing to start similar initiatives. - We have added a little on this, however it is difficult to communicate effectively how this was dealt with. We have now included the use of social media to advertise the grant round and curricula, as well as the timeline of events as references in this community building, as well as networking from our team of collaborators.

Reviewer 2: Gilbert Miller

1. Methods: What was the number of the participants? Was there good level of engagement and satisfaction? - We have now added this further

2. Methods: Provide more detail about Gephi 0.9.3. What exactly was done and how? - Apologies, we have now added a reference here, we have also elaborated further. Please see 'Network in skills creation' within the methods section.

3. Methods: Was there any assessment linked to the MOOC? - There was no assessment therefore we have added this to the text as above.

4. Fig 3a: Provide a key of the dot intensities and the number of participants, as it is hard to understand this figure - We have now removed reference to dot intensity, as it is an overlay of partially transparent dots that creates thai intensity

5. Figs 3a and 5a: The map images are a bit blurry, please provide clearer images. - Unfortunately we are unable to obtain clearer map images without new software.

6. Fig 3c: It is difficult to read some of the letters in the small nodes, increase the font size"intensity indicating number of participants from the region" - the intensity diff - We have increased the size of this image to make the text easier to read, the purpose of the nodes, in seeing a diversity of skills can now for the most part be fulfilled. We hope the size included is large enough. The text size is linked to the node size, therefore this limits us in the clarity of these nodes.

7. Conclusions: Can this approach be applied in other parts of the world? - We think so, at least in similar demographic regions such as the global south, this has been included in the conclusive text now.

8. Conclusions: How will your approach help raising AMR awareness in the area? - We have now included mentions of improving AMR awareness directly in the conclusion. We believe this method is a responsible and highly scalable method of democratic science teaching, and it can be applied to AMR again successfully.

Best wishes

Chris Graham

VERSION 1

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000472.v1.5

Georgios Efthimiou; University of Hull, Biomedical Science, Hardy Building, Cottingham Road, UNITED KINGDOM, Hull

Date report received: 02 January 2023
Recommendation: Major Revision

Comments: The work presented is clear and the arguments well formed. This study would be a valuable contribution to the existing literature. This is a study that would be of interest to the field and community. Please address all comments made by the reviewers.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000472.v1.3

Gilbert Miller; University of Cape Town, SOUTH AFRICA

Date report received: 02 January 2023
Recommendation: Minor Amendment

Comments: The authors have described an innovative initiative to tackle the problem of AMR in Ghana, although this approach can also be used elsewhere too. The manuscript is well-written, with good organisation, methodology and presentation of results/discussion. Some minor points that need to be addressed: 1. Methods: What was the number of the participants? Was there good level of engagement and satisfaction? 2. Methods: Provide more detail about Gephi 0.9.3. What exactly was done and how? 3. Methods: Was there any assessment linked to the MOOC? 4. Fig 3a: Provide a key of the dot intensities and the number of participants, as it is hard to understand this figure 5. Figs 3a and 5a: The map images are a bit blurry, please provide clearer images. 6. Fig 3c: It is difficult to read some of the letters in the small nodes, increase the font size"intensity indicating number of participants from the region" - the intensity diff 7. Conclusions: Can this approach be applied in other parts of the world? 8. Conclusions: How will your approach help raising AMR awareness in the area?

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000472.v1.4

Morgan Feeney; University of Strathclyde, UNITED KINGDOM
https://orcid.org/0000-0003-4841-0706

Date report received: 16 December 2022
Recommendation: Minor Amendment

Comments: This paper by Graham and colleagues is a very nice description of an excellent initiative, which should hopefully encourage other teams to follow their example in developing initiatives to tackle the problems of AMR in this way. It is clear from their results that pairing a MOOC with a grant round is an effective way of teaching, and of building solutions to AMR. Moreover, it is encouraging to see that this appears to have sparked lasting networks and collaborations amongst the participants as measured by their continued engagement via Slack etc. Major points: 1. It is not clear from the text whether or not there was any formative or summative assessment associated with the MOOC (apart from the grant round that followed). Perhaps the authors could make this point clearer. 2. It is not entirely clear to me from the timeline in Figure 1, when the in-person practical sessions were held - I think it might be helpful to show this. 3. If the authors have this data, it would be perhaps helpful to show the number of participants attending each session, and/or accessing the YouTube recordings afterwards. Was engagement sustained throughout all sessions of the course? Minor points: 1. line 65, "Recent efforts in Nigeria…." - a citation would be helpful here for readers who are unfamiliar with this work 2. minor editing for grammar needed - for example, line 75 "There has been efforts" should be "There have been efforts" 3. lines 93-95, perhaps briefly summarise the results of the other study? 4. Figure 1, the colour code being used in panel d is not immediately self-evident, perhaps this could be explained in the legend or with a key. 5. Figure 1, much of the text is too small to read easily when the figure is viewed at 100% magnification. Perhaps this figure could be made a little more readable. 6. Methods - use of Gephi 0.9.3 could be a little clearer and more detailed, to allow researchers to reproduce these results 7. Fig 3a - "intensity indicating number of participants from the region" - the intensity differences are not very easily distinguishable on the map as presented, perhaps these could be made clearer and/or a scale added to help readers understand these data. 8. Fig 3c - some of the smaller nodes are unreadable, perhaps the font sizes could be made larger 9. Fig 4 - the use of similar colours (blue and yellow) across the pie chart and the bar graphs may imply connections between these data/groups - however, they don't appear to be connected. Perhaps a different colour palette for the bar charts would be helpful. 10.Fig 5b - unclear what is meant by the "tree" on the left of the heatmap, or how this was generated? Is the tree necessary at all? 11. Fig S1 - tryptophan resistance? Perhaps the authors could explain what is meant by this, or provide a little more detail about the strains being shown? 12. For the parts i-iii in panel A, it would be helpful to explain what these are in the figure legend. 13. A little more detail on the community building (i.e. how the Slack channels were organised, to promote effective communication) might be helpful for readers wishing to start similar initiatives.

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Very good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

SciScore report

https://doi.org/10.1099/acmi.0.000472.v1.1

iThenticate report

https://doi.org/10.1099/acmi.0.000472.v1.2

RESEARCH ARTICLE

Community review: a robust and scalable selection system for resource allocation within open science and innovation communities [version 1; peer review: 1 approved with reservations]

Chris L.B. Graham[1,2*], Thomas E. Landrain[2*], Amber Vjestica[3], Camille Masselot[4], Elliot Lawton[2], Leo Blondel[2], Luca Haenal[2], Bastian Greshake Tzovaras[4], Marc Santolini[2,4*]

[1]University of Warwick, Coventry, UK
[2]Just One Giant Lab, Paris, France
[3]University of Nottingham, Nottingham, UK
[4]Learning Planet Institute, Université de Paris, Paris, France

* Equal contributors

v1 **First published:** 06 Dec 2022, **11**:1440
https://doi.org/10.12688/f1000research.125886.1
Latest published: 06 Dec 2022, **11**:1440
https://doi.org/10.12688/f1000research.125886.1

Abstract

Resource allocation is essential to selection and implementation of innovative projects in science and technology. Current "winner-take-all" models for grant applications require significant researcher time in writing extensive project proposals, and rely on the availability of a few time-saturated volunteer experts. Such processes usually carry over several months, resulting in high effective costs compared to expected benefits. We devised an agile "community review" system to allocate micro-grants for the fast prototyping of innovative solutions. Here we describe and evaluate the implementation of this community review across 147 projects from the "Just One Giant Lab's OpenCOVID19 initiative" and "Helpful Engineering" open research communities. The community review process uses granular review forms and requires the participation of grant applicants in the review process. Within a year, we organised 7 rounds of review, resulting in 614 reviews from 201 reviewers, and the attribution of 48 micro-grants of up to 4,000 euros. The system is fast, with a median process duration of 10 days, scalable, with a median of 4 reviewers per project independent of the total number of projects, and fair, with project rankings highly preserved after the synthetic removal of reviewers. Regarding potential bias introduced by involving applicants in the process, we find that review scores from both applicants and non-applicants have a similar correlation of r=0.28 with other reviews

Open Peer Review

Approval Status ?

	1
version 1 06 Dec 2022	? view

1. **Ferdinando Patat**, European Southern Observatory, München, Germany

Any reports and responses or comments on the article can be found at the end of the article.

within a project, matching traditional approaches. Finally, we find that the ability of projects to apply to several rounds allows to foster the further implementation of successful early prototypes, as well as provide a pathway to constructively improve an initially failing proposal in an agile manner. Overall, this study quantitatively highlights the benefits of a frugal, community review system acting as a due diligence for rapid and agile resource allocation in open research and innovation programs, with implications for decentralised communities.

Keywords
community science, decision making,

Corresponding authors: Chris L.B. Graham (christopherlbgraham@live.co.uk), Marc Santolini (marc@jogl.io)

Author roles: Graham CLB: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Landrain TE**: Conceptualization, Data Curation, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Vjestica A**: Data Curation, Formal Analysis, Investigation; **Masselot C**: Methodology, Project Administration; **Lawton E**: Data Curation, Project Administration; **Blondel L**: Data Curation, Methodology, Software; **Haenal L**: Software; **Greshake Tzovaras B**: Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Santolini M**: Data Curation, Formal Analysis, Investigation, Methodology, Supervision, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the AXA Research Fund. This work was partly supported by the French Agence Nationale de la Recherche (ANR), under grant agreement ANR-21-CE38-0002-01.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Graham CLB, Landrain TE, Vjestica A *et al.* **Community review: a robust and scalable selection system for resource allocation within open science and innovation communities [version 1; peer review: 1 approved with reservations]** F1000Research 2022, **11**:1440 https://doi.org/10.12688/f1000research.125886.1

First published: 06 Dec 2022, **11**:1440 https://doi.org/10.12688/f1000research.125886.1

Introduction

The distribution of scientific funding through grants requires the identification of novel, feasible and potentially impactful projects. However, the traditional scientific grant allocation system involving a closed panel of experts in the field, or in similar fields,[1] is notoriously slow,[2] time consuming and expensive, often taking months and occurring in timescales of yearly rounds or grant calls. In extreme cases, the grant review program can be more costly than simply allocating small grants to each applicant, as in the case of the NSERC grant system of 2008.[3] In addition, the allocation of grants has shown to suffer from various biases, such as the composition of the grant panel,[4] gender and geographical location,[5] group based dynamics personality triumphing over other qualitative factors,[6–8] and socio-psychological factors such as group dynamics and personality traits triumphing over other qualitative factors.[8,9] Overall, selection results are only weakly predictive of future performance.[10]

Often, the reason to conduct grant allocations in a 'closed' setting is to protect the intellectual property of the grant applicants. As a result, the majority of unsuccessful grant applications, which contain a large amount of research effort, are inevitably lost, unavailable to the public after the fact.[11] The recent emergence of the open science movement[12–14] has reversed this incentive, with open access practices and early sharing of results such as pre-registration now becoming normalised by institutions and journals.[15]

Beyond the allocation of funding, the review of early-stage, unpublished work by community peers has been leveraged to allocate other types of resources. For example, conferences often need to allocate time for their participants to showcase their work to other members of the community during a usually short amount of time, thereby providing a platform for promoting the work, building novel collaborations, and getting feedback to improve a manuscript. In such cases, peer reviewing is needed to decide in a collegial fashion whether a work is worth a full oral presentation, a shorter lightning talk, a poster, or is not of a high enough standard to be showcased to participants. For example, the EasyChair online platform has been used by close to 100k conferences for handling such review processes.[16] Often, participants to a conference are also part of the "program committee" reviewing the proposed abstracts and papers of peer applicants, alongside external members of the scientific community. This allows for a rapid process usually lasting less than a few weeks.

This suggests there is a potential for a new, more agile route for community-driven grant allocation bypassing pre-selected grant panels that handle funds and introduce barriers,[8] and relying instead on peer applicants to handle a large-scale application process in a short timescale. In this study, we present the design, implementation, and results of a community-driven, open peer-review system to support two open research communities during the coronavirus disease 2019 (COVID-19) pandemic across seven selection rounds (Figure 1): the "OpenCOVID19" initiative from Just One Giant Lab (JOGL)[14,17] and the COVID relief charity Helpful Engineering.[18] We show that this system is robust (unaffected by reviewer removal), agile (fast timeline), iterative (covering multiple grant rounds), decentralised (driven by the community), and scalable. Finally, we discuss these results and the perspectives they offer for the design of future community-driven review systems.

Methods

Context

The implementation of a crowd-based, open peer-review system followed the need to support two nascent community efforts, first by allocating volunteers to projects in the COVID relief charity Helpful Engineering,[18] then by allocating funding to projects in the JOGL "OpenCOVID19" initiative.[17] The method was developed as an open access grant review and funding allocation system, meaning that it was open to anyone willing to review. It was implemented using the Just One Giant Lab platform (app.jogl.io) as the project proposal host, and free-to-use tools and forms to conduct the review process (Extended Data:FigS2[19]). The implementation was applied and refined over 7 rounds across 1 year.

General process of review

The peer review system was conducted on early phase projects within both JOGL and Helpful Engineering. These projects were submitted by project leaders to a grant review process in order to allocate volunteers in the case of Helpful Engineering, and funding in the context of OpenCovid19. Reviews of these projects (see Figure 1b) were initially conducted by members of the community and included members of other projects who also submitted their project for review.

As a consequence of the process being experimental and serving an urgent need, the process was altered over time. However, it followed the same general pattern (Figure 1, Extended Data:FigS1[19]). First, a template for the grant proposal was created by the community and was iteratively edited (Extended Data[19]) template followed typical grant application templates,[20] with sections on team composition, the project general hypothesis and its timeline. The proposal was then

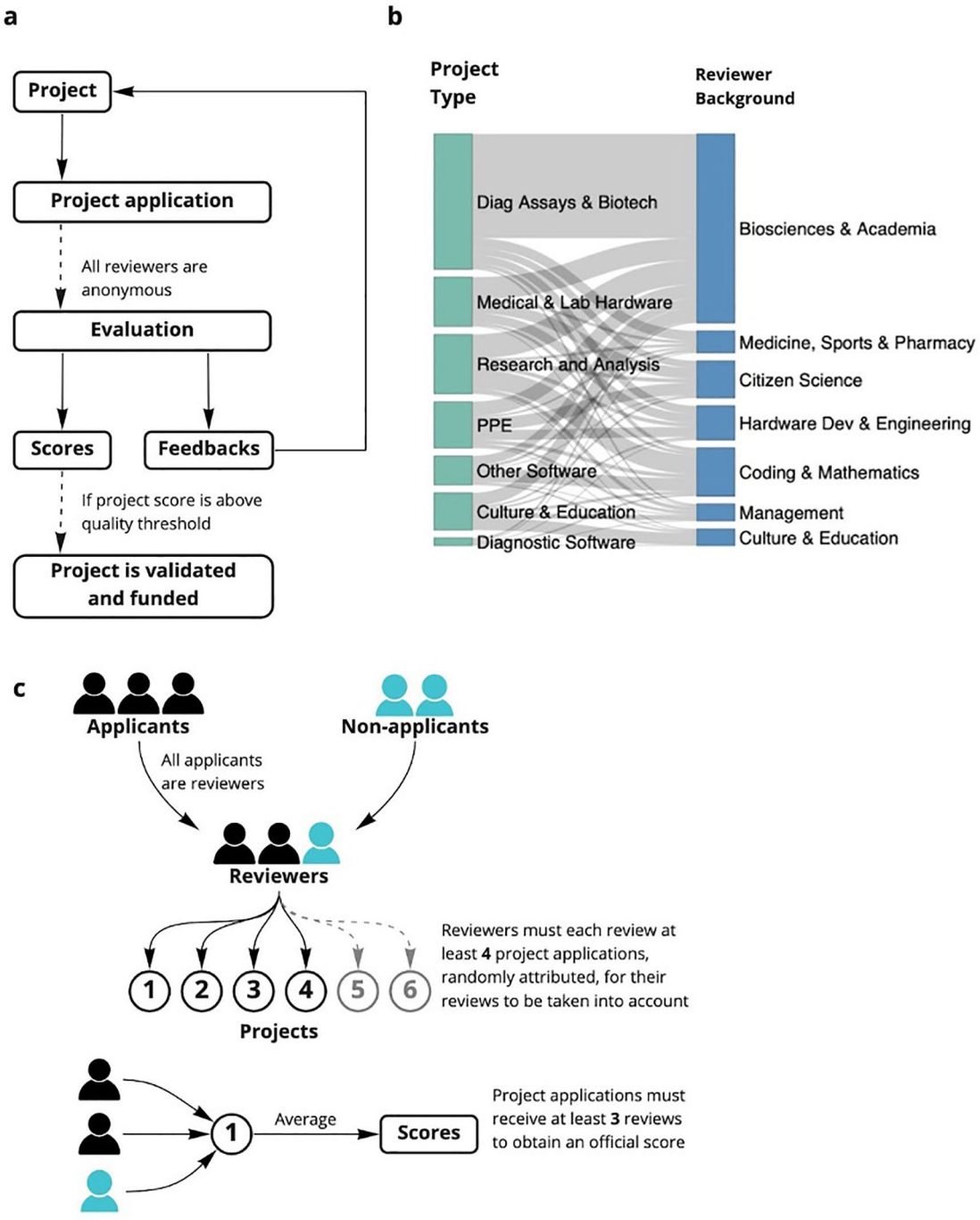


Figure 1. Overview of the open peer review process. (a) Stages of the open peer review process JOGL rounds 3-5. The online review forms and templates are found in supplementary data. (b) community review method JOGL rounds 3-5 (c) distribution of project type to expertise across rounds.

submitted using a google form, which requested an email address and required only one application per project (Extended Data:FigS2a[19]). In Helpful Engineering rounds this included a link to their proposal hosted in editable google documents, while in JOGL rounds this included instead a link to their open access JOGL page proposal. The project links were manually formatted into a google sheet with a link to a review form for convenience, along with descriptions of desirable reviewer skills by the applicants in the proposal submission form to help reviewers find relevant projects (Extended Data: FigS2B[19]). A technical evaluation form scoring various criteria (eg: proposal efficacy, team composition, impact) on a

scale from 1-5 (Extended Data[19]) was created by the designers of the program and iteratively changed following feedback from the community (Extended Data:FigS2c[19]). This form separated questions on projects into two areas centred around Impact and Feasibility for ease of identifying the problems and/or strengths in their grant application. A message with a link to the reviewer form for use in review, along with a nested google sheet containing project proposal links was spread among the community through announcements and email. In later rounds (JOGL 3-5) all applicants were asked to review at least three other projects and the process was randomised, removing the need for a sheet. The review process was given between 4 days HE 1, 8 days HE 2, 7 Days - JOGL 1, 10 days - JOGL 2, 16 days - JOGL 3, 21 days - JOGL 4 and 28 days JOGL 5, (Extended Data:FigS1b[19]) to allow reviews to occur and be collected via a google form into a google sheet automatically (Extended Data:FigS2d[19]). No reviewer selection was performed, however usernames (Slack handles or JOGL user names depending on the round) and emails were collected for conducting further analyses. The average reviewer scores were then composed into a presentation to the community, and those projects with a score above a given impact/feasibility threshold (Extended Data:FigS2e[19]) were chosen for grant funding. Due to the community aspect of our study, members from the JOGL HQ participated in the process, and their removal from the analysis does not change the observations (Extended Data:FigS10[19]), we therefore retain these in our analysis.

Iterative changes to the review process

As mentioned in the previous section, the method of review was iteratively changed throughout the programme, elongating from an initial "emergency style" four day period of review and allocation (HE round 1) to 21 and 28 days in JOGL rounds 4 and 5 as the need for rapid response reduced, with an overall median average of 10 days per round (Extended Data:FigS1b[19]). As such, the design of the general process described in Figure 1 and Extended Data:FigS1[19] had some variations. For example, initially applicants were not required to review applications (Figure 1b). Upon scaling up of the programme, the process was adapted to be less dependent on volunteer reviewers, (Extended Data:Fig S1b,A-D[19]) and more dependent on the applicant's reviews of their competing peers (Figure 1c). In JOGL rounds 3, 4 and 5 (Extended Data:FigS1b[19]) teams depositing a proposal could only be eligible after having reviewed at least three other teams. The changes in the process and differences in the rounds are summarised in Extended Data:FigS1c.[19] The major changes between Helpful Engineering (HE) and JOGL rounds (Extended Data:FigS1c[19]) occurred through changes in the nature of proposal submission from google document links to an online project repository. In addition, HE rounds offered no grants, but instead publicity and allocation of members to projects, while JOGL offered microgrants worth up to 4000 euros per team (Extended Data:FigS2c[19]).

Final selection process

In Helpful Engineering, this review method allowed 54 projects to be reviewed and ranked by score for community recruitment purposes, with no official threshold, but instead an arbitrary set of "Highlighted projects". Within JOGL this grant system reviewed 96 eligible applications (Figure 2) and allocated requested funds to 36 of these. Once the review process had taken place, the cut-off threshold of scores given by reviewers to projects for funding by JOGL was decided by an absolute threshold (above 3.5/5 average reviewed score) rather than a rejection rate. The absolute 3.5/5 threshold was chosen due to the gap in project scores in the first JOGL round, and maintained at this standard for consistency. Those with a score above the threshold were funded.

Detection of fraudulent reviewer behaviour

The results of each round, and number of reviews per reviewer were closely monitored through simple email handle tracking by a data handling administrator. If a number of emails were found to be grading a particular project and not others this was suggestive of fraudulent behaviour and self-grading. These reviews were then removed, and teams that were found responsible for this bad behaviour were removed from the review process, as described in grant round participation rules. This was performed only one time across all rounds prior to the rule of each reviewer having a minimum review count for their scores to be counted, which was created after this event.

Computation of inter-review correlations

In order to compute the correlation between reviews within a project, we first proceeded with data cleaning. Indeed, in several rounds, reviewers had to answer only a subset of questions from the review form that corresponded to the topic of the project (e.g. data project vs bio project). However, in some cases, projects were assigned to one or the other category by the different reviewers, leading them to answer to different sets of questions, making the correlation only partial. To mitigate this effect, for each project we kept only the reviews that corresponded to the choice of topic that was most expressed among reviewers. If no majority could be found, the project was removed from analysis. We then converted review scores into vectors of length the number of grades in the form. A Spearman's rho correlation was then computed between all pairs of reviews within a project. Finally, for each review we computed the average correlation with the other reviews in the project. This number was then associated with the features of the reviewer who produced the review (Figure 4 and Extended Data:FigS7[19]).

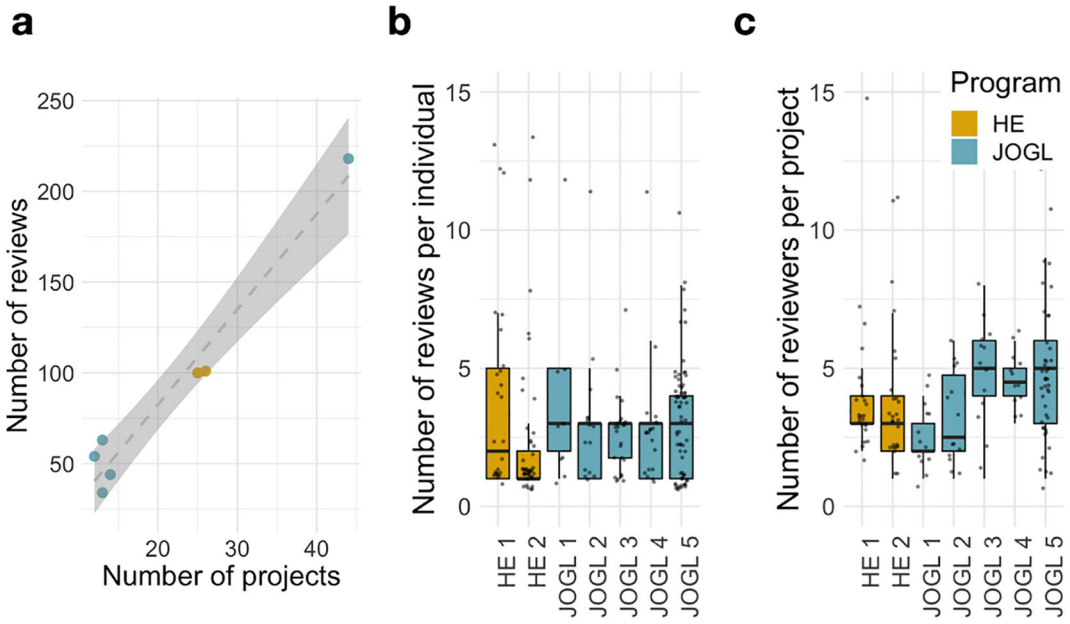


Figure 2. Scalability of the community review methodology. (a) Number of Reviewers and projects during each round of peer/grant review. HE- Helpful Engineering Crowd reviews, JOGL- Just One Giant Lab funded projects. (b) Number of reviews per individual reviewer. (c) Number of reviewers per project. Despite a scale-up in the number of projects, the number of reviews per round scales linearly with the number of projects applying.

Reviewer feasibility and impact scores

For JOGL rounds 1-5, we categorised the 23 to 29 questions from the review forms into either impact or feasibility related questions (see Underlying Data Review forms). The feasibility and impact categories were used to provide two dimensional projections of project scores during the result presentation.

Reviewer professions and project types

For all JOGL rounds, reviewer responses of the "What is your expertise relevant to this project" question were manually coded into simple categories per review (see Table S1 in the Extended Data[19]). This data was then used as a proxy for expertise distribution across rounds (Figure 1b).

In addition, reviewer responses to the "Which category would you say the project falls under?" question were manually coded into a set of simple categories, representing a summary of the project types across rounds per review (see Extended Data conversion table[21]). The data, due to suggested categories provided by the form, needed little manual coding, but was formatted into a list, then concatenated into similar project types for simplicity. This data was used to assess project type distribution across rounds (Figure 1b).

Bootstrap analysis

In order to perform the bootstrap analysis of Figure 3d, we first ranked all projects using their average review score across reviewers. We then selected a review at random. If the corresponding project had at least another review, we removed the selected review and recomputed the average scores and final ranking. We then computed the Spearman correlation between the obtained scores and the original scores. This process was repeated until each project had only one review. Finally, we reiterated this analysis 50 times. The analysis code can be found as Extended Data.[19]

Ethics/Consent

We confirm all ethical guidelines have been followed, using the same ethical procedures described in Commission Nationale de l'Informatique et des Libertes registration number “2221728” for user research and “2227764” for grant administration. Consent was granted by a user agreement on JOGL’s website upon signup (https://app.jogl.io/data), and on the google forms used during the study.

Results

Scalability of the review process

We describe in Figure 2 the reviewing activity across the seven rounds implemented. Despite the large differences in number of projects between rounds, we find that the number of reviews per round scales linearly with the number of

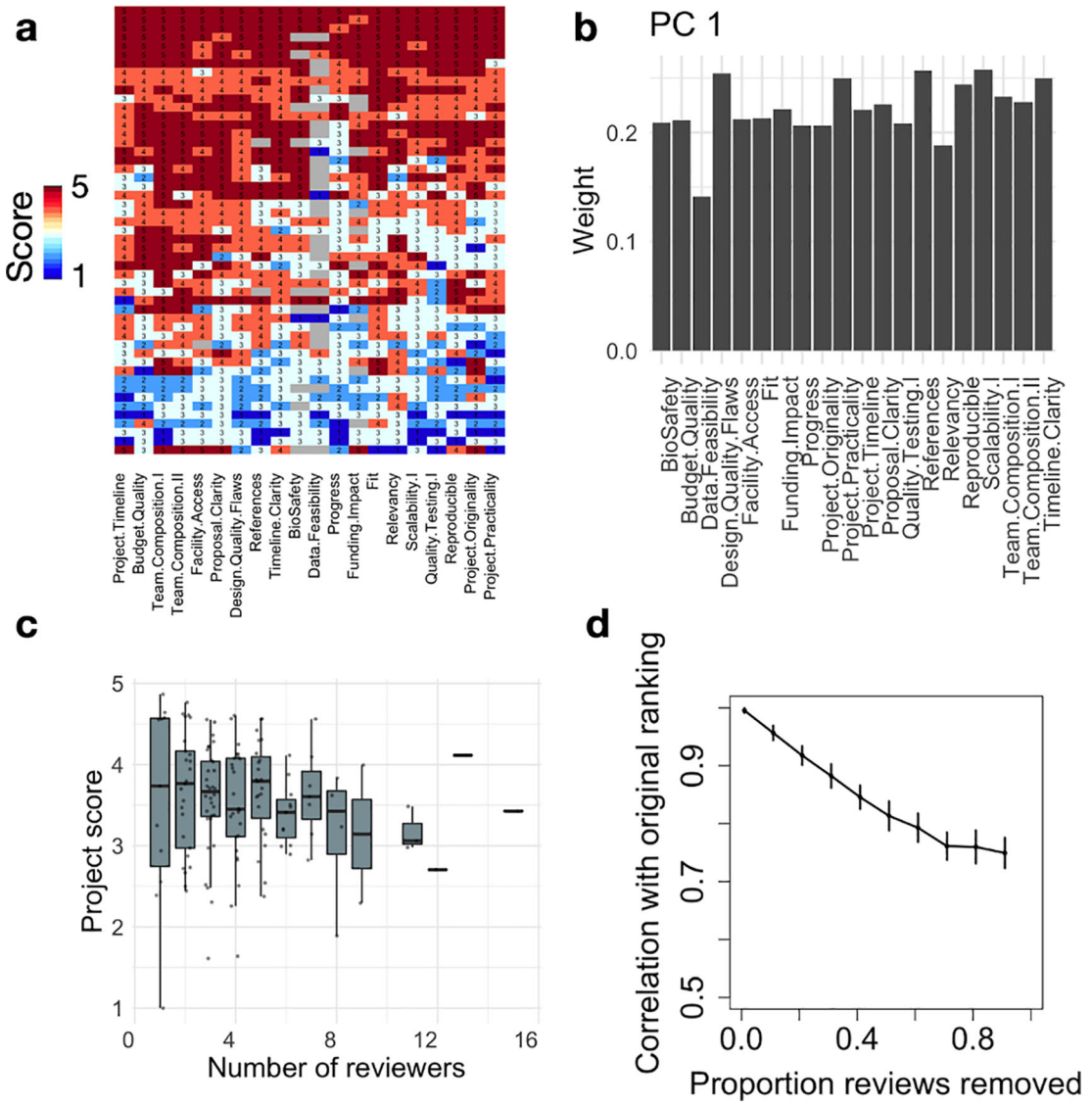


Figure 3. Robustness of the Community review process. (a) Heatmap showing review scores (rows) across questions (columns) for the JOGL round 4. Row and column clustering was performed using correlation distance and average linkage. (b) We show for PC1 (53% variance) the weights of the questions from the original question space. PC1 has near uniform weights across dimensions, indicating that it corresponds to an average score. (c) Project average score across reviewers as a function of number of reviewers. (d) Bootstrap analysis showing the Spearman correlation between the final project ranking and simulated project rankings with increasing proportion of reviews removed from the analysis (see Methods).

projects applying (Figure 2a). In addition, the number of reviews per individual and number of reviewers per project have relatively stable distributions across rounds, independent of scale (Figure 2b-a). For example, despite the substantial growth in reviewers and projects in JOGL round 5, we find that the distributions of number of reviews per reviewer and number of reviewers per project are comparable to those observed in the previous rounds, highlighting the scalability of this review system to different systems. Finally, we note that the number of reviewers per project show a sustained increase from JOGL round 3 onwards, corresponding to the change in review process, where applicants were required to review at least 3 other projects (see Methods). This highlights the benefits of this requirement in promoting sustained engagement.

Robustness of the final project ranking

In order to obtain a granular score for each project, the reviewers had to grade between 23 (JOGL 1-2) and 29 (JOGL 3-5) criteria in the review form.[21] We first investigate whether these questions would cover different dimensions of project

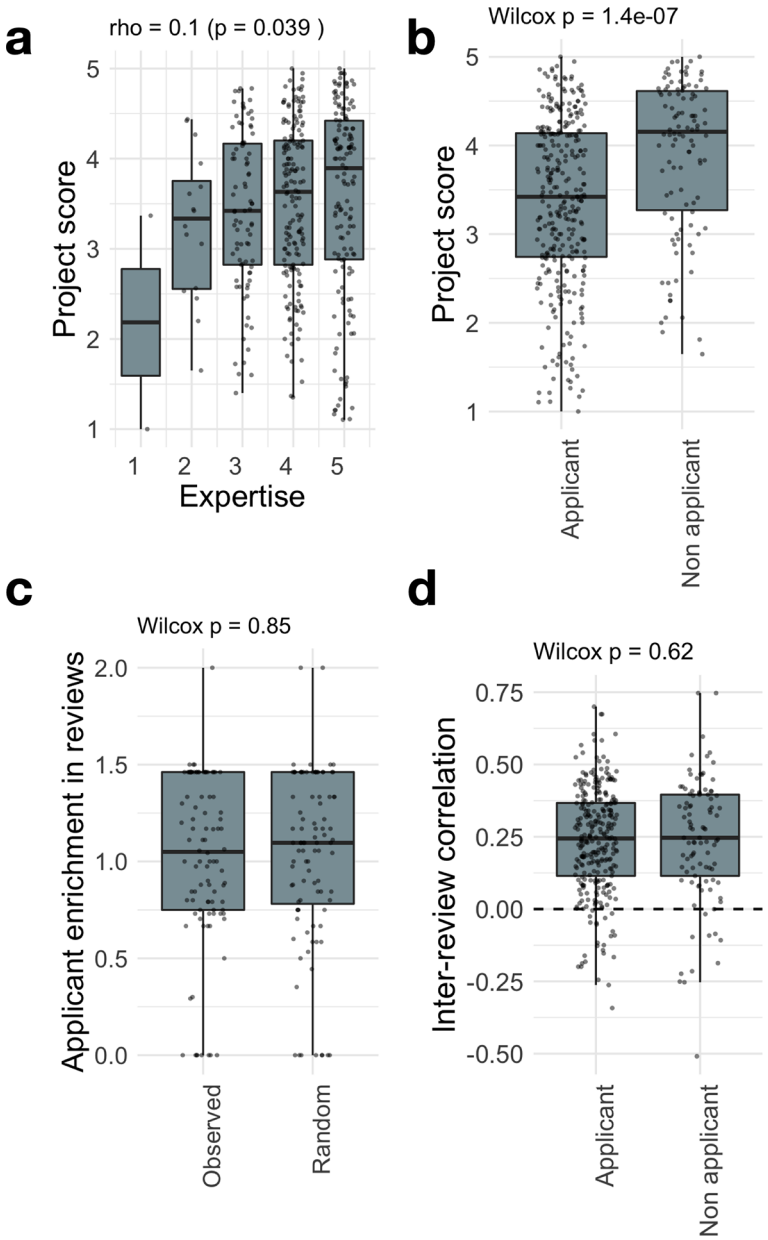


Figure 4. Questionnaire granularity allows to measure and mitigate reviewer biases. Breakdown of project score as a function of (a) self-assessed expertise, (b) applicant status (i.e. the reviewer is also an applicant in the round). See Fig S4 for a breakdown by review round. (c) For each project, we compute the ratio between the proportion of applicant reviewers to the average proportion of applicant reviewers observed in the round. The boxplot compares the computed enrichments to the ones obtained for randomly assigned reviewers to projects, showing that applicants are evenly distributed across projects. (d) For each project, we compute the ratio between the proportion of applicant reviewers to the average proportion of applicant reviewers observed in the round. The boxplot compares the computed enrichments to the ones obtained for randomly assigned reviewers to projects, showing that applicants are evenly distributed across projects.

quality. We show in Figure 3a a heatmap of reviewer scores in JOGL round 4 across 20 questions (removing questions only representing a minority of projects), visually showing a greater inter-review variability (rows) than inter-questions variability (columns). As such, respondents seem to assign a project with either low scores or high scores throughout their review. To quantify the number of dimensions of variation across grades, we conduct a Principal Component Analysis (PCA) on the questions correlation matrix, i.e correlations between pairs of questions across reviews (see Extended Data: Fig S2a[19]). We find that the first principal component (PC1) explains most of the variance (53%), with the next largest PC

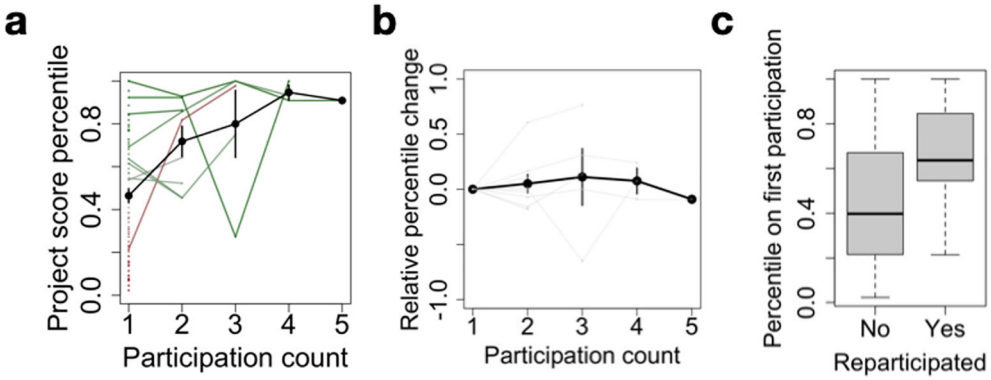


Figure 5. Multiple participations foster long-term project sustainability. (a) Project score percentile as a function of participation count. For each project, a score percentile is computed to quantify their relative rank within a specific application round, allowing to compare multiple projects across rounds. Participation count refers to the successive number of rounds a project has applied to. The black line denotes the average across projects, error bars represent standard error. Dots correspond to projects with only one participation, and lines to re-participating projects. Finally, the color gradient indicates relative score at first participation, from red (low) to green (high). (b) Same as a., after subtracting the percentile at first participation. (c) Score percentile at first participation as a function of whether or not a project has re-participated.

explaining less than 6% of the variance (Extended Data:Fig S3[19]). When examining the weights of the various questions in PC1, we find that they all contribute to a similar level (Fig 3b), meaning that the PC1 is close to the average over all questions, confirming the visual insight from Fig 3a. This shows that scores are highly correlated, and that the average score across the review form is a reasonable operationalisation of project quality. In addition, we find that the top 10 PCs explain ~90% of the variance, indicating that review forms could be reduced in complexity using only half of the number of questions to obtain a similar outcome.

We next investigate the reliability of the review scores obtained across reviewers. As suggested by the previous section, for each review we compute the average score across all criteria from the review form. In the following, we refer to this average score as the review score. We observe a generally good discrimination of review scores between projects, with intra-project variation smaller than inter-project variation (Extended Data:FigS4[19]).

Finally, we investigate the robustness of the final project ranking as a function of the number of reviews performed using a bootstrap analysis (see Methods). For each project, a project score is computed by averaging its review scores, and projects are then ranked by decreasing score. We show in Figure 3d the Spearman correlation between the original project ranking and the ranking obtained when removing a certain proportion of reviews. We find that even with only one review per project, the final ranking is strongly conserved (rho=0.75 and see Extended Data:FigS5[19]), confirming that intra-project variability is much smaller than the range of inter-project variability. This supports our design strategy, showing that the use of a granular form allows us to differentiate between projects whilst minimising the impact of individual reviewers' variability.

Measuring reviewer biases

The previous results show the existence of variability between reviews from different reviewers, yet with limited impact on final rankings (Figure 3d). Here we investigate the source of review variability: is it due to inherent grading variability between individuals, or can it be attributed to other factors? To evaluate this question, we analyse how review score varies with reviewer attributes. We explore in particular two possible sources of bias for which we could gather data: expertise and application status. First, reviewer expertise might be important in determining an accurate project score. This feature is operationalised using the self-reported expertise grade (1 to 5) present in the review forms of JOGL rounds. Second, a majority of reviewers (65%) were applicants of other competing projects, which could lead to a negative bias when reviewing other competing projects.

We show in Figure 4 how the review score varies as a function of these reviewer characteristics. We find that review score increases slightly with expertise (Figure 4a, Spearman's rho=0.1, p=0.039). However, the strongest effect is found when looking at applicant bias: review scores from applicants are significantly lower than those from non-applicants (Figure 4b, p=1.4e-7). Given the fact that in JOGL rounds 3-5, applicants were required to score at least 3 projects, they are found to have a lower expertise towards other projects (Extended Data:Fig S6[19]), which could explain the lower scores as

suggested by Figure 4a. Yet, when controlling for review expertise, we find that application status is the main contributing factor, with a score difference between applicants and non-applicants of -0.52 points (p=1.61e-6, Extended Data: Supplementary Table 1[19]). This supports that application status is a significant source of bias in the final score.

Such differences could be due to unfair grading, with reviewers from a certain category (applicants or non-applicants) grading more "randomly" than others. To analyse this effect, we need to look beyond average score into correlations. Indeed, two similar average scores could stem from highly different fine-grain score vectors. Imagine two reviewers grading 3 questions from 1 to 5. The first reviewer gives the grades 1, 2, and 5, while the second gives 5, 1, and 2. These reviews produce the same average score (2.67). However, their fine-grain structure is anti-correlated, with a Pearson correlation r = -0.69. In our context, we find that review scores are positively correlated, with a median Pearson correlation between their reviews of r = 0.28 across rounds (Figure 4d), in line with previous observations in traditional funding schemes [35]. More importantly, we find no difference between applicants and non-applicants in their correlation with other project reviews (Figure 4c). This indicates that the variability between grades within a review form are conserved across reviewer characteristics (see Fig S7 and Extended Data:Fig S9 for the other characteristics[19]). As such, if applicants are uniformly distributed across projects, one will not expect a difference in the final rankings.

A framework for iterative project implementations

In the JOGL implementation of the community review system, projects can apply to any number of rounds, irrespective of whether or not they have already successfully obtained funding in a previous round. We found 9 projects that applied to multiple rounds. On average, the relative performance of the projects in a grant round increases as a function of the number of participations (Figure 5a). We find that this effect is explained by re-participation being associated with early success, with initially lower performing projects eventually dropping out (Figure 5b-c). As such, the multiple round scheme supports projects with a high initial potential in the long-term through repeated micro-funding allocations. We also note that in the case of 2 projects, re-participation after an initial failure allowed them to pass the acceptance threshold. This highlights how constructive feedback allows for a rapid improvement of a project and its successful re-application in the process.

Discussion

In this manuscript we describe the "community review" method for the identification of novel, feasible and potentially impactful projects within two communities of open innovation: Helpful Engineering and OpenCovid19. This process was leveraged for the attribution of volunteers as well as micro-grants to projects over a year, in an agile and iterative framework.

Key to the system is the requirement of applicants to take part in the reviewing process, ensuring its scalability. As such, the number of reviews is proportional to the number of projects applying (Figure 2), with a fast median process duration of 10 days. This requirement comes at a risk, since applicants might be negatively biased towards other projects they are competing against. Accordingly, we found that applicants consistently give a lower score to projects when compared to non-applicants (-0.52 points). This bias cannot be explained solely by the lower expertise of applicants towards the randomly assigned projects. Indeed, we found that self-reported expertise has only a limited impact on the final score (Figure 4c). The effect is most stringent for rare cases of self-reported expertise of 1 and 2 out of 5, suggesting that a threshold of 3 might be implemented to remove non-expert bias. It is on the other hand possible that non-applicants are positively biased towards projects from which they might have personally been invited to review. We however noted no such report in the conflict of interest question in the review form.

Despite these biases, we found that applicants and non-applicants have a similar behaviour when grading questions in the form, with a stable Pearson correlation between their reviews of r = 0.28 (Figure 4/Extended Data:Fig S8[19]). This is slightly higher than the correlation of 0.2 observed in an analysis of the ESRC's existing peer review metrics,[22] suggesting comparable outcomes when compared to existing institutional methods. The similarity of their correlation profiles means that such biases contribute a similar "noise" to the system: they might change the overall average scores, but not their ranking as long as applicants are well distributed across projects. Accordingly, we found that the community review system is robust to the removal of reviewers, with an average ranking Spearman correlation of 0.7 in the extreme case of one reviewer per project.

Finally, we showed that some projects apply multiple times to the application rounds. While the number of such projects of this type is small (9 projects), we find that it had two benefits. First, we found two projects that re-applied after an unsuccessful application, allowing them to pass the acceptance threshold on the second application. This showcases the ability of the feedback system to benefit projects in constructively improving their application. Furthermore, we found that the number of applications of a project is strongly dependent on its performance on the first application. This means

that the iterative process allows to select highly promising projects and sustain their implementation in the mid- to long-term. This is of particular importance when considering traditional hackathon systems, where promising projects are usually not supported over longer periods of time.

The speed and cost-efficiency of the community review process has allowed for a reactive response to the high-pressure environment created by the pandemic. This agility has meant that within the short time frame given, projects have been able to produce literature, methods and hardware and put them to use.[23–28] Overall, the community review system allows for a rapid, agile, iterative, distributed and scalable review process for volunteer action and micro-grant attribution. It is particularly suited for open research and innovation communities collaborating in a decentralized manner and looking for ways to distribute common resources fairly and swiftly. Finally, community review offers a robust alternative to institutional frameworks for building trust within a network and paves the way for the installation of community-driven decentralized laboratories.

Data availability

Underlying data

Open Science Framework: DATA FOR: Community review: a robust and scalable selection system for resource allocation within open science and innovation communities. https://doi.org/10.17605/OSF.IO/CAZ4N.[21]

This project contains the following underlying data:

- Review Data (the raw responses of the review rounds analysed by the paper, and the raw data used in the study.)
- Project round progress.csv (aggregated data and is based on data post analysis for our final figure, and the scores of each project over time, however we have aggregated this for ease of viewing.)
- Grant Review forms (the forms used to assess each proposal)
- Peer Review protocol (the protocol used to analyse the raw data, giving the correlation values we refer to in the paper)
- Coded expertise (the simplified version of project and reviewer type collected during review)

Extended data

Open Science Framework: EXTENDED DATA FOR: Community review: a robust and scalable selection system for resource allocation within open science and innovation communities, https://doi.org/10.17605/OSF.IO/W5Q9B.[19]

This project contains the following extended data:

- Supplementary figures 1-10
- Supplementary table 1
- Analysis code

Acknowledgements

An earlier version of this article can be found on biorxiv at doi (https://doi.org/10.1101/2022.04.25.489391). We acknowledge the work of the projects involved and Helpful Engineering as well as Just One Giant Lab volunteers for their logistical and communications help throughout the pandemic.

References

1. Langfeldt L: **The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome.** *Soc. Stud. Sci.* 2001 Dec 1; **31**(6): 820–841.
Publisher Full Text

2. Herbert DL, Barnett AG, Graves N: **Australia's grant system wastes time.** *Nature.* 2013 Mar; **495**(7441): 314–314.
PubMed Abstract | Publisher Full Text

3. Gordon R, Poulin BJ: **Cost of the NSERC Science Grant Peer Review System Exceeds the Cost of Giving Every Qualified Researcher a Baseline Grant.** *Account. Res.* 2009 Feb 27; **16**(1): 13–40.
PubMed Abstract | Publisher Full Text

4. Roumbanis L: **Peer Review or Lottery? A Critical Analysis of Two Different Forms of Decision-making Mechanisms for Allocation of Research Grants.** *Sci. Technol. Hum. Values.* 2019 Nov 1; **44**(6): 994–1019.
Publisher Full Text

5. Severin A, Martins J, Delavy F, *et al.*: **Gender and other potential biases in peer review: Analysis of 38,250 external peer review reports.** *PeerJ. Inc.* 2019 Jun [cited 2021 Jul 9]. Report No.: e27587v3.
Reference Source

6. Coveney J, Herbert DL, Hill K, *et al.*: **'Are you siding with a personality or the grant proposal?': observations on how peer review panels function.** *Res. Integr. Peer. Rev.* 2017 Dec 4; **2**(1): 19.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Pier EL, Raclaw J, Kaatz A, *et al.*: **'Your comments are meaner than your score': score calibration talk influences intra- and inter-panel variability during scientific grant peer review.** *Res. Eval.* 2017 Jan; **26**(1): 1–14.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Fogelholm M, Leppinen S, Auvinen A, *et al.*: **Panel discussion does not improve reliability of peer review for medical research grant proposals.** *J. Clin. Epidemiol.* 2012 Jan; **65**(1): 47–52.
Publisher Full Text

9. Coveney J, Herbert DL, Hill K, *et al.*: **'Are you siding with a personality or the grant proposal?': observations on how peer review panels function.** *Res. Integr. Peer. Rev.* 2017; **2**: 19.
PubMed Abstract | Publisher Full Text | Free Full Text

10. Fang FC, Casadevall A: **NIH peer review reform--change we need, or lipstick on a pig?** *Infect. Immun.* 2009 Jan 21;2009 Mar; **77**(3): 929–932.
PubMed Abstract | Publisher Full Text | Free Full Text

11. von Hippel T, von Hippel C: **To apply or not to apply: a survey analysis of grant writing costs and benefits.** *PLoS One.* 2015; **10**(3): e0118494.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Landrain T, Meyer M, Perez AM, *et al.*: **Do-it-yourself biology: challenges and promises for an open science and technology movement.** *Syst. Synth. Biol.* 2013 Sep; **7**(3): 115–126.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Masselot C, Tzovaras BG, Graham CLB, *et al.*: **Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study.** *J. Particip. Med.* 2022 Jan 21; **14**(1): e32125.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Santolini M: *Covid-19: the rise of a global collective intelligence?* The Conversation; 2020 [cited 2021 Apr 2].
Reference Source

15. eLife Journal Policy: **eLife.** 2022.
Reference Source

16. Easy Chair: 2021.
Reference Source

17. Kokshagina O: **Open Covid-19: Organizing an extreme crowdsourcing campaign to tackle grand challenges.** *RD Manag.* 2021 Mar 23; **52**: 206–219.
Publisher Full Text

18. Helpful Engineering: **Helpful.** [cited 2021 Jul 9].
Reference Source

19. Graham CLB: **EXTENDED DATA FOR: Community review: a robust and scalable selection system for resource allocation within open science and innovation communities.** 2022, November 26.
Publisher Full Text

20. Je-S electronic applications - Economic and Social Research Council: [cited 2021 Jul 9].
Reference Source

21. Graham CLB:DATA FOR: Community review: a robust and scalable selection system for resource allocation within open science and innovation communities. [Dataset]. 2022, November 26.
Publisher Full Text

22. Jerrim J, de Vries R: **Are peer-reviews of grant proposals reliable? An analysis of Economic and Social Research Council (ESRC) funding applications.** *Soc. Sci. J.* 2020 Mar 6; 1–19.
Publisher Full Text

23. Aidelberg G, Aranoff R, Javier Quero F, *et al.*: *Corona Detective: a simple, scalable, and robust SARS-CoV-2 detection method based on reverse transcription loop-mediated isothermal amplification.*ABRF; 2021.
Reference Source

24. Bektas A: **Accessible LAMP-Enabled Rapid Test (ALERT) for detecting SARS-CoV-2.** *Viruses.* 2021.
Publisher Full Text | Reference Source

25. Cheng C: **COVID-19 government response event dataset (CoronaNet v. 1.0).** *Nat. Hum. Behav.* 2020; **4**(7): 756–768.
PubMed Abstract | Publisher Full Text

26. Greshake Tzovaras B, Rera M, Wintermute EH, *et al.*: **Empowering grassroots innovation to accelerate biomedical research.** *PLoS Biol.* 2021 Aug 9; **19**(8): e3001349.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Monaco C, Jorgensen E, Ware S: *The One Hour COVID Test: A Rapid Colorimetric Reverse-Transcription LAMP–Based COVID-19 Test Requiring Minimal Equipment.*ABRF; 2021 Sep.
Reference Source

28. Tzovaras BG: **Quantified Flu: an individual-centered approach to gaining sickness-related insights from wearable data.** *medRxiv.* 2021.

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 17 February 2023

https://doi.org/10.5256/f1000research.138239.r161990

?

Ferdinando Patat
European Southern Observatory, München, Germany

The paper reports on the design, deployment and evaluation of what the authors describe as an 'agile community review system' for allocating micro-grants. As opposed to classical peer-review performed by pre-selected expert panels, the community review involves the participation of applicants themselves. It therefore provides as scalable solution, in the sense that the number of reviewers increases with the number of applications, hence keeping the review load manageable and the duty-cycle faster than in the classical panel case. The data show that the inter-reviewer correlation is similar to what has been reported for the classical paradigm. Also, the final rankings are shown to be robust against the randomised removal for reviewers.

The paper presents the case very clearly and accurately. One aspect which is weak (honestly this is probably the only weakness of the paper) is that the casual reader is left under the impression that the idea of a community review is new and proposed here for the first time. And that is not the case. This concept, sometimes indicated as *distributed peer review* (DPR), has been around for many years. As far as I know the idea appeared first in 2009 in a paper by Saari & Merrifield (2009) [1]. Although it was referring to applications for telescope time, the concept was of general interest. This mechanism has been implemented at least in three major, ground-based astronomical facilities (GEMINI, ALMA, ESO). Also, a similar distributed process has been deployed in the field of computer sciences for the selection of conference papers[5]. Other interesting examples are those of the US National Science Foundation[2,3] and the US National Institute of Food and Agriculture (NIFA), which has deployed DPR in 2016[6,7].

The authors should therefore provide a short recap of these initiatives and cite the applicable references. In addition to the papers published by the above organisations, there is an article on Physics Today[4] which gives a nice summary for the specific cases in astronomy.

Other minor points, which the authors should consider, are listed here:

1. Scalability. The authors correctly state that the approach is scalable, in the sense that increasing the number of projects linearly increases the number of reviewers. They show this in Fig. 2b. At a first read it looks like this is an unexpected results and, somehow, the

data show that this is the case. However, I argue that this is completely expected, since the number of potential reviewers, on average, is directly proportional to the number of submitted projects. This would not be necessarily the case if the system would not force the applicants to review proposals. Maybe the authors should be more explicit on this.

2. In the section about fraudulent behaviour, the authors state that this was removed from the review process. However, they do not explain what this means in practice in terms of actions taken against those reviewers. Is their project simply rejected? Are they warned? Or is just their evaluation removed from the data?

3. The caption of Figure 4 for panels c) and d) does not match what is presented in the figure and also its description in the main text section *Measuring reviewer biases*:

   "For each project, we compute the ratio between the proportion of applicant reviewers to the average proportion of applicant reviewers observed in the round. The boxplot compares the computed enrichments to the ones obtained for randomly assigned reviewers to projects, showing that applicants are evenly distributed across projects. (d) For each project, we compute the ratio between the proportion of applicant reviewers to the average proportion of applicant reviewers observed in the round. The boxplot compares the computed enrichments to the ones obtained for randomly assigned reviewers to projects, showing that applicants are evenly distributed across projects."

   The authors never explain what they mean by "enrichment". Also, while panel c) may be showing a ratio, this is not the case for panel d), which on the y-axis has the label "inter-review correlation", which sounds like a Pearson correlation, and not a ratio. In general, I am rather confused about panels c) and d) in Fig. 4, which I reckon requires a better explanation also in the main text (which, by the way, never includes the word "enrichment").

4. As shown in Fig. 3a, there are two evaluation criteria related to the Team composition. Given that many organisations are now moving to a dual anonymisation of the applications, it would be interesting to see how the scores change if one removes these two indicators from the final score. I guess that, given the large number of indicators (and the results shown by the authors about the slight change when 50% of them are removed), no measurable effect is going to be seen. Also because, since the team's identity is known to the reviewers, there is certainly a cross-talk between this indicator and all others. It is probably therefore impossible to disentangle its effect from the available data. Nevertheless, I suggest the authors mention this aspect.

The study is well designed, the analysis was conducted in an appropriate way and included a satisfactory level of checks and controls. The methods are well explained and the necessary data are made available so that the analysis can be repeated and validated. The statistical analysis and its presentation meet the required standards.

The conclusions are well supported, interesting and useful for other organisations which may consider adopting a similar schema. I therefore recommend the article for indexing after the point on the missing citations and short description of existing cases is addressed.

References

1. Merrifield M, Saari D: Telescope time without tears: a distributed approach to peer review.

Astronomy & Geophysics. 2009; **50** (4): 4.16-4.20 Publisher Full Text
2. Naghizadeh P, Liu M: Incentives, Quality, and Risks: A Look Into the NSF Proposal Review Pilot. *arXiv*. 2013. Publisher Full Text | Reference Source
3. Mervis J: A radical change in peer review. *Science*. 2014; **345** (6194): 248-249 Publisher Full Text
4. Singh Chawla D: Distributed peer review passes test for allocating telescope slots. *Physics Today*. 2020. Publisher Full Text | Reference Source
5. Shah N: Challenges, experiments, and computational solutions in peer review. *Communications of the ACM*. 2022; **65** (6): 76-87 Publisher Full Text
6. NIFA - FAQs for Distributed Peer Review (DPR). *NIFA*. 2016. Reference Source
7. distributed peer-review (DPR). *NIFA*. 2016. Reference Source

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Astrophysics: Supernovae. Peer-review: proposal selection and time allocation processes.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 08 Mar 2023

Chris Graham

Dear Ferdinando,

Thank you for a thorough and well-rounded review of our article. We are glad that you found the article interesting and our methods to be robust and the data well described. You are correct, our method is very similar to distributed peer review (DPR) and we should have

included citations and mentions of this in our literature review. Its use in astronomy, and especially by the NSF are particularly interesting, as are the articles which cite its drawbacks. Thank you for improving the paper with mention of this. We have incorporated the mention of our use of DPR throughout the paper.

We have now expanded the introduction to speak about distributed peer review, and throughout conclusions in the article have also made reference to distributed peer review, although we did not initially use DPR, and were instead adopting a crowd-sourcing approach, and then combined the two methods. These changes have been made especially to the Introduction, where we have incorporated your suggested references, as well as others not mentioned

In terms of your specific comments, for Figure 2d, the wording has now been changed to reflect the reality that the community review method was expected to be scalable. The use of language was through the assumption of a null hypothesis and was innappropriate.

We have also changed the Figure legend of Figure 4c, as pointed out to be reflective of the figure. This was a copy-editing error. The use of the word ' enrichment' has also been specified for Figure 4d.

Your final point, on removing team composition questions is very interesting especially when thinking about double anonymisation of reviewers. We hope, based on the data of the PCAs which separated the questions by their power and difference to other questions (Figure 3b), that this was not a big influence on the final scores therefore would agree with your own assumption that this wouldn't effect the rankings greatly. Unfortunately although your hypothesis for the inclusion of a further figure to analyse this is sound, our team would prefer to keep the paper in its current state in terms of new hypotheses, but think this is an excellent question and in future publications will ask this very question if given the chance.

Overall your review has identified some important literature and potential issues, as well as a new hypothesis and we are very grateful for it.

Thank you for taking the time to analyse our article in the way you have, you've enlightened us specifically on the astronomy field's use of DPR and shown how there is precedent for such methods, with the example of the National Science Foundation's pilot study. Hopefully there will be further attempts with such large organisations in the future, perhaps with safeguarding measures to counteract any collusion, as well as controls studies for grant review with existing techniques.

Competing Interests: No competing interests.

Original Paper

Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study

Camille Masselot[1*], MPH; Bastian Greshake Tzovaras[2*], PhD; Chris L B Graham[1], MSc; Gary Finnegan[3], MSc; Rathin Jeyaram[2], MSc; Isabelle Vitali[4], MBA; Thomas Landrain[1], MSc; Marc Santolini[1,2], PhD

[1]Just One Giant Lab Association, Paris, France

[2]Center for Research and Interdisciplinarity, INSERM U1284, Universite de Paris, Paris, France

[3]IXL Editorial, Kildare, Ireland

[4]Sanofi, Gentilly, France

[*]these authors contributed equally

Corresponding Author:
Marc Santolini, PhD
Center for Research and Interdisciplinarity
INSERM U1284
Universite de Paris
8bis rue Charles V
Paris, F-75006
France
Phone: 33 679751700
Fax: 33 632377297
Email: marc.santolini@cri-paris.org

Abstract

Background: The rise of major complex public health problems, such as vaccination hesitancy and access to vaccination, requires innovative, open, and transdisciplinary approaches. Yet, institutional silos and lack of participation on the part of nonacademic citizens in the design of solutions hamper efforts to meet these challenges. Against this background, new solutions have been explored, with participatory research, citizen science, hackathons, and challenge-based approaches being applied in the context of public health.

Objective: Our aim was to develop a program for creating citizen science and open innovation projects that address the contemporary challenges of vaccination in France and around the globe.

Methods: We designed and implemented Co-Immune, a program created to tackle the question of vaccination hesitancy and access to vaccination through an online and offline challenge-based open innovation approach. The program was run on the open science platform Just One Giant Lab.

Results: Over a 6-month period, the Co-Immune program gathered 234 participants of diverse backgrounds and 13 partners from the public and private sectors. The program comprised 10 events to facilitate the creation of 20 new projects, as well as the continuation of two existing projects, to address the issues of vaccination hesitancy and access, ranging from app development and data mining to analysis and game design. In an open framework, the projects made their data, code, and solutions publicly available.

Conclusions: Co-Immune highlights how open innovation approaches and online platforms can help to gather and coordinate noninstitutional communities in a rapid, distributed, and global way toward solving public health issues. Such initiatives can lead to the production and transfer of knowledge, creating novel solutions in the public health sector. The example of Co-Immune contributes to paving the way for organizations and individuals to collaboratively tackle future global challenges.

(J Particip Med 2022;14(1):e32125) doi: 10.2196/32125

KEYWORDS

open science; open innovation; programmatic research; collective intelligence; web based; immunization; vaccination access; vaccine hesitancy; innovation; vaccine; public health; access; framework; participatory; design; implementation

Introduction

Background

As the world faces a rise in the number of complex challenges that threaten the resilience of our economic, environmental, social, and health systems, we observe a shift toward more collaboration and openness in the way science and innovation is performed [1-3], bringing governments, civil society, and the private sector closer. Examples of this include the efforts made to accelerate society's progress toward the Sustainable Development Goals (SDGs) [4] and the fight against pandemics, such as COVID-19 [5]. Yet, access to vaccines and vaccination hesitancy remain as some of the complex challenges to be addressed in order to achieve universal health coverage [6].

Immunization is one of the most cost-effective interventions to protect oneself and others from infectious diseases [7] and saves between 2 million and 3 million lives per year [8].

Yet, the annual death toll for vaccine-preventable diseases stands at 1.5 million, and large gaps in coverage persist, not only between countries but also within their territories [7]. In particular, the World Health Organization (WHO) listed vaccine hesitancy among the top 10 global health threats for 2019 [9]. Continuing global efforts to leave no one behind may be a long-standing challenge [10] when new information technologies and social media platforms are both part of the problem [11] and the solution. More recently, the COVID-19 pandemic demonstrated the repertoire of logistical and administrative challenges to the deployment and administration of vaccines, especially in low-resource settings [12].

In response, the WHO Global Vaccine Action Plan 2011-2020 [7] committed 140 countries and 290 organizations to promoting and prioritizing greater collaboration between governments, nongovernmental organizations, the private sector, and all citizens to address outbreaks of vaccine-preventable diseases. Additionally, a number of new digital and open innovation initiatives have been launched: the WHO has developed the Vaccine Safety Net [13], a network of websites about vaccination; health authorities in Canada have developed a school-based quiz to educate children about immunology and vaccines [14]; Finland is testing a computer game to communicate the benefits of human papillomavirus vaccination [15]; a project in India uses digital necklaces to record children's immunization history [16]; and the global Vaccination Acceptance Research Network has been established [16].

Global health guidelines showcase the positive outcomes of social participation for universal health coverage [17], which include more meaningful dialogue, more sustainable solutions, and more trust from citizens in health system institutions or in the decisions that are made. Indeed, there is room for more initiatives that allow people to genuinely co-design solutions in a multidisciplinary manner during and following pandemics [18]. Hence, the number and sustainability of these types of initiatives could be amplified by fostering increased collaboration with nonacademic citizens in the creation and development of solutions in an open innovation framework [19]. This is the gap that Just One Giant Lab (JOGL) is proposing to fill with the Co-Immune program.

Citizen science is an emerging and highly diverse practice that can be broadly defined as the general public being involved in the process of doing research [20]. Research has demonstrated that intensity and diversity of collaboration positively affect the quality [21] and productivity [22] of research, while positively impacting the knowledge integration from participants [23]. Likewise, participant transdisciplinarity [24] seems critical to generating innovative outcomes [25] and dealing with complex real-world problems [26]. Such mechanisms are often at play in the field of citizen science, promising to transform the knowledge generation landscape by tapping into networks of nonacademic citizens [26,27] in a new social contract for this kind of research [28]. Citizen science has the potential to expand the number of individuals contributing knowledge and ideas, transform how hypotheses are generated, and transform how data sets are analyzed. Such approaches have already been applied to investigate individual diseases through patient-led research [29,30] and public health challenges, such as the epidemiology of cancer [31-33].

Other approaches to create and develop knowledge and solutions to complex challenges are slowly entering the mainstream. In particular, hackathons, challenge-based approaches, and the participation of citizens in science have been flourishing over the last two decades [34], especially within the natural sciences [35] and, more recently, within medical sciences, public health, and population-health research [36,37].

Hackathons are short, intensive, and collaborative events that are designed to prototype solutions addressing a specific problem. They originated in the early 2000s in digital and tech fields and have been adapted to address more complex challenges in global health [38-40]. Such initiatives are not without pitfalls: they suffer, by design, from the lack of paths to sustainability for the projects they launch [41]. In response to such criticisms, there are increasing efforts, such as the "Make the Breast Pump not Suck" hackathon and "Trans*H4CK," to improve hackathon methodology by working directly with affected communities [41]. Several initiatives, such as a Massachusetts Institute of Technology collaborative design studio, provide insights into hackathon methods [42] to facilitate better hackathons [43,44]. More recently, multiple entities have engaged in organizing hackathons to address the COVID-19 crisis [45,46].

Challenge-based approaches, which provide frameworks for learning while solving real-world issues, have also been on the rise in global health and have proven to be efficient for generating innovative solutions and for incentivizing mass community engagement [45]. For example, the potential of participative models to address complex questions, along with

the power of contests to offer a structure that catalyzes this work, has been exhibited by the Epidemium initiative on cancer epidemiology [46].

Despite the numerous tools and technologies created to facilitate collaboration in citizen science projects, challenges remain. These include the issues of the complementarity, coherence, and diffusion of these initiatives [34] to efficiently address international policies and local needs, as the local adoption of hackathon solutions often remains low [47].

Therefore, the promotion of transdisciplinarity and citizen science in an open innovation framework, coupled with methods such as hackathons, and a challenge-based approach represent an opportunity to address current complex challenges of vaccination that would overcome the limits of either solution alone. In this paper, we describe the design, implementation, and outputs of Co-Immune, a collaborative open innovation program that was run in 2019 to address vaccination hesitancy and access to vaccination.

Objectives

Co-Immune's aim was to develop an environment that favors the creation and development of citizen science and open innovation projects addressing the contemporary challenges of vaccination in France and around the globe. This program had four specific objectives: (1) to foster a collaborative, open, and transdisciplinary dynamic; (2) to promote the emergence of accessible knowledge and innovative solutions; (3) to support participants in the elaboration and development of their project; and (4) to disseminate the outputs and results in an open science framework. In this study, we describe the methodology of Co-Immune and its implementation, and we present its key outcomes.

Methods

Design

The overall program duration was 10 months (March 2019 to January 2020), divided into 6 months of preparation and 4 months of rollout of activities that included offline and online events, support for the development of citizen science projects, and assessment and awards for projects participating in the challenge-based competition. The main outputs of the program were projects, categorized as leading to (1) knowledge production, if they performed data analysis or generated new knowledge, whether it was context specific, generic [48], or knowledge transfer [49]; or (2) solutions, such as hardware, software, and interventions.

Co-Immune was coordinated online through the JOGL platform [50] and supported by 13 partners from the public and private sector (Table S1 in Multimedia Appendix 1). The challenge-based nature of the program was designed to be an incentive for teams and participants to continue developing their projects after hackathon events or to create their project on JOGL at any other time.

The governance of Co-Immune was designed to provide freedom for projects to develop innovative solutions while ensuring their compliance with local and international regulations and consideration of ethical and scientific integrity. To this end, we constituted the independent Committee for Ethics, Science and Impact (CESI), which issued an opinion on the rules of participation in the program and validated the strategic orientation of the program. Public health priorities were identified based on a literature review and divided between two main challenges to streamline participants' work: vaccination coverage and vaccination hesitancy. They were then validated by the CESI. In addition, through a series of semistructured interviews, experts at the 7th Fondation Merieux Vaccine Acceptance conference [51] identified eight specific issues to address and potential room for solutions. The CESI also participated in the co-elaboration of the assessment grid, which was used as a base to grant nonmonetary prizes to projects in December 2019.

To be eligible for a prize, a project was required to have created a comprehensive description of their initiative on the JOGL platform and a video pitch. This material was provided to experts in charge of the assessment.

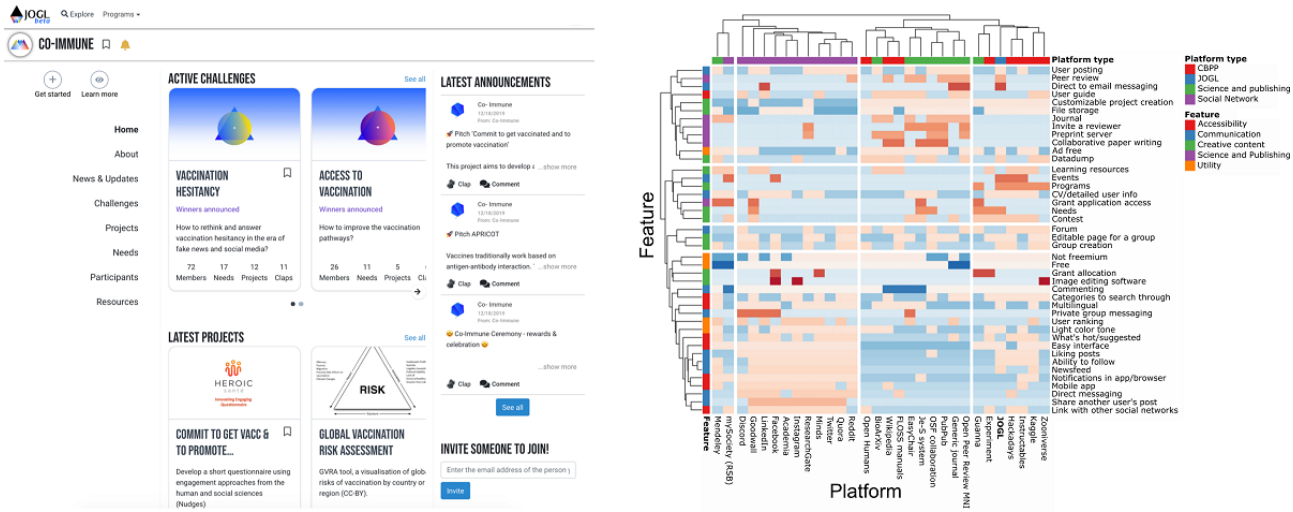
Participant Recruitment

Participants were recruited through our network of partners from around the globe and social media communication. Participation was open to everyone above the age of 18 years, if they had agreed to follow the participation rules validated by the CESI. Participants could take the role of "project leader" or "contributor."

JOGL Platform

Co-Immune participants used the JOGL platform to document their projects and recruit collaborators throughout the course of the program. JOGL is a decentralized mobilization platform designed for use in collaborative research and innovation (Figure 1). Within the JOGL platform, users can create a profile and declare their skills. Once registered, they can create or join projects, follow the activity of other members, post on their project feed, and comment on other posts. They can also highlight needs for a project they are part of, specifying skills that can help to solve project problems. We compared the JOGL features to those of other online platforms for citizen science, social networking, and science and publishing through a cluster analysis (see Figure 1 as well as the supplementary method and Figure S1 in Multimedia Appendix 1), indicating that the platform is functionally similar to other platforms in the space and is suitable to hold a citizen science program such as Co-Immune.

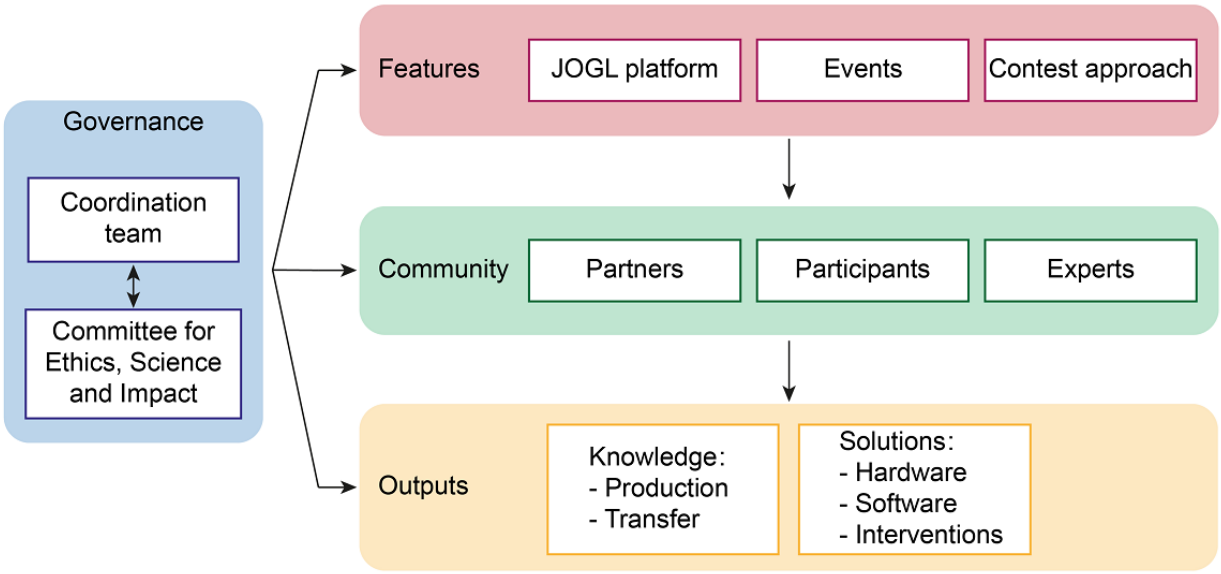
Figure 1. Overview of the Just One Giant Lab (JOGL) platform. The image on the left is a screenshot of the JOGL platform. The right-hand image is a heatmap of feature presence across popular online tools. For each platform (columns), we numerically encoded the presence (1) or absence (0) of each feature (rows). Then, for each element, we computed a Z score by standardizing values across platforms, represented here by the color spectrum: blue (low) to red (high). CBPP: citizen-based peer production network (ie, citizen science platform); CV: curriculum vitae; Je-S: Joint Electronic Submissions; MNI: Montreal Neurological Institute; OSF: Open Science Framework; RSB: Royal Society of Biology.



Implementation

The Co-Immune program was realized through an interrelated and interacting set of technological and social features (Figure 2). Our coordination team implemented the larger program (ie, events, online platform, and contest approach) and helped to recruit a community of partners and participants who interacted with each other and were supported in their efforts through the high-level design features. With support from the governance structure of the Co-Immune program, the individual projects managed to provide outputs that included knowledge production and transfer as well as solutions, such as hardware, software, and interventions.

Figure 2. Workflow of the Co-Immune program design. JOGL: Just One Giant Lab.



Building an Open Community

To build the community, we contacted organizations involved in a wide range of domains before the launch of the program, thereby creating a first pool of contributing professionals and students. We also recruited participants via the organization of events, typically in the evening, aimed at creating projects, fostering collaboration among participants to address project needs, and providing mentorship. To facilitate the coordination of the community, all participants were required to use the JOGL platform to describe their projects, form teams, list their needs, and initiate collaboration.

In order to create a supportive and collaborative environment for the participants, we reached out to various organizations to establish partnerships. Our intention was two-fold: (1) to

facilitate the participation of the organizations' students and employees as participants or mentors by involving their institution and (2) to enhance the sustainability of projects after the course of the program by connecting them with potential partners at the early stage of their development.

The 13 partners operated in the health, technology, and social sectors, and included research, innovation, and education organizations, as well as professional networks, incubators, and communication specialists (Figure 3). The number of partners grew over the life span of the initiative and were often suggested by existing partners or through connections made during events.

We organized 10 offline and online events between October and December 2019 (Table 1). Participants for events were recruited through social media and mailing lists leveraging our network of partners. Among the four on-site events that were organized, two were hackathons aimed at motivating participants to join the program, while the other two were aimed at fostering collaboration around the most advanced projects. Their median duration was 2.25 (IQR 2) hours.

The facilitation of the hackathon-style events relied on the use of participatory and collective intelligence design and problem-solving techniques [52]. In particular, participants were encouraged to form multidisciplinary teams including both professionals and students.

Three partners in Paris—Epitech, the Wild Code School, and the Center for Research and Interdisciplinarity (CRI)—co-organized and hosted events for their students, respectively, in their engineering, coding, and life science and education schools. Other partners—Kap Code, Excelya, and CorrelAid—mobilized their teams to act as mentors during these events. A total of 14 mentors attended events, and five came to more than one event.

In addition, we organized four 1-hour online events. The first was an opportunity to share information about Co-Immune with people around the globe. Another event discussed best practices to document open science projects. Finally, two events focused on the resolution of needs of single projects (Table 1 [53,54]).

Figure 3. Treemap representing the domains of action of the 13 Co-Immune partners.

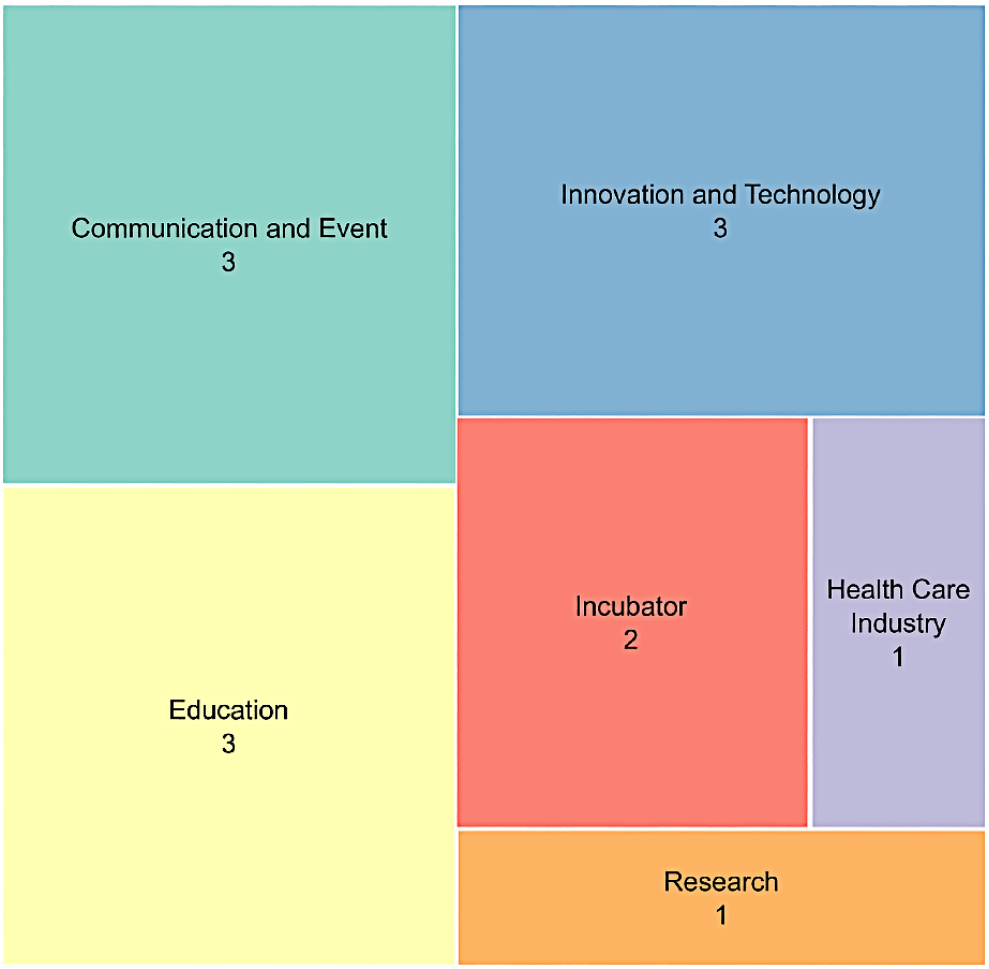


Table 1. Co-Immune events.

Name	Mode; type; location	Duration (hours), n	Objective	Design; supporting partners (if applicable)	Participants, n
Launch	Offline; ceremony; CRI[a], Paris	3	Gather the initial community	Presentation of the program design, features, timeline, and partners, as well as networking	60
OpenJOGL[b]; Co-Immune	Online	1	Q&A[c] session on the program	Presentation of Co-Immune and questions and answers	3
Sprint; open data	Offline; hackathon; CRI, Paris	2.5	Build community, create projects, and create data repositories	Statement of the problem (videos of experts), team formation and effort, mentoring, and publication of results on the JOGL platform; supported by CRI and CorrelAid	25
OpenJOGL; Vaccination Awareness Escape Game [54]	Online	1	Foster collaboration around single projects	Pitch of the project and its needs, feedback from experts, and questions and answers	7
Sprint; project creation	Offline; hackathon; CRI, Paris	4	Build community and create multidisciplinary projects	Statement of the problem (videos of experts), ice breaker, multidisciplinary team formation and effort, mentoring, presentation of results, vote for the most promising projects, publication of results on the JOGL platform, and networking; supported by CRI, Epitech, Wild Code School, CorrelAid, and Excelya	22
Sprint; open data	Offline; hackathon; Wild Code School, Paris	3	Accelerate the development of projects related to data science	Selection of a project by participants among the two choices available, team formation and effort, mentoring, presentation of results, publication on the JOGL platform, and networking; supported by Wild Code School, CorrelAid, and Excelya	15
Sprint; open data	Offline; hackathon; Epitech, Paris	3	Build the community, create projects, and accelerate the development of one project using Twitter data	Statement of the problem, selection of a project by participants among the four choices available (including one already existing project), team formation and effort, mentoring, presentation of results, vote for the most promising project, publication of results on the JOGL platform, and networking; supported by Epitech, Kap Code, Excelya, and CorrelAid	35
OpenJOGL; HERA[d]: A Health Platform for Refugees [53]	Online	1	Foster collaboration around single projects	Pitch of the project and its needs, feedback from experts, and questions and answers	7
OpenJOGL; better documentation for better collaboration	Online	1	Help teams document their projects in the most open and reproducible way	Expert presentation on best practices for documenting open science projects, presentation of Co-Immune expectations for documentation, and questions and answers	13
Closing ceremony	Offline; ceremony; CRI, Paris	2	Close the Co-Immune program	Presentation of the main outputs of the program and awards for the best projects	70

[a]CRI: Center for Research and Interdisciplinarity.

[b]JOGL: Just One Giant Lab.

[c]Q&A: question and answer.

[d]HERA: Health Recording App.

Co-Immune Experts: CESI Members, Mentors, and Interviewees

Individuals who were considered "experts" included all the CESI members as well as experienced professionals of a certain field who attended events and provided technical guidance to teams as "mentors."

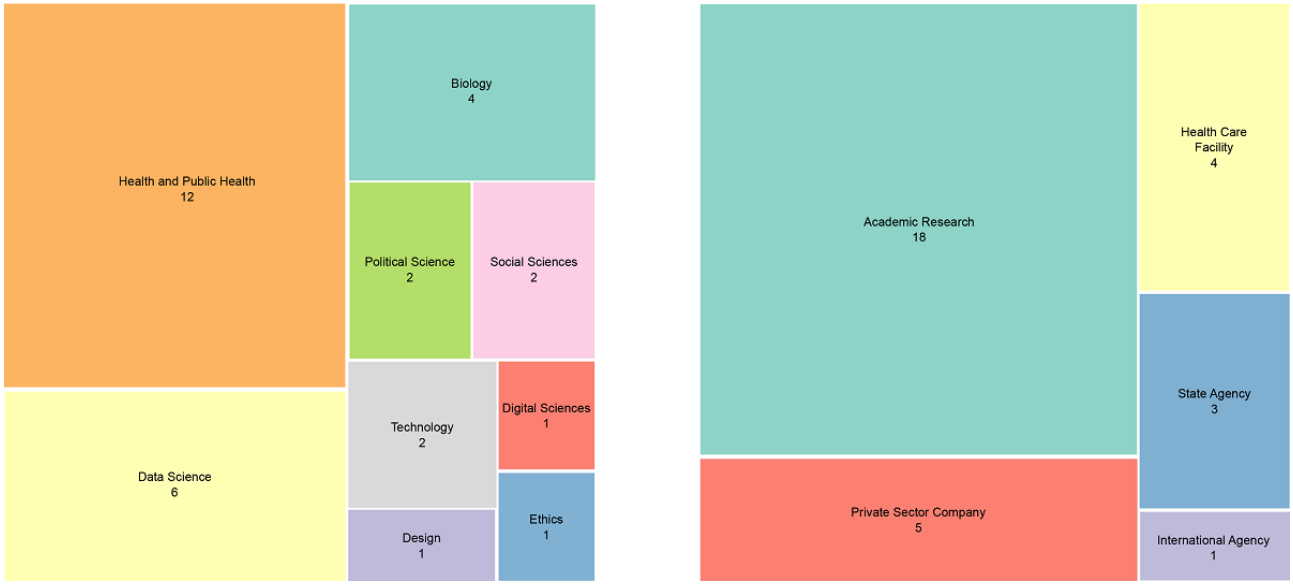
The CESI members were sought to represent the diversity of stakeholders involved in advancing access to vaccines and reducing vaccine hesitancy. By choosing interviewees who were researchers specializing in the challenges of access to vaccines and vaccination hesitancy, we aimed at benefiting from their expert understanding of the issues and of the priorities to be addressed to streamline the work of participants around particular problems. Finally, we grew the pool of mentors over the span of the program to best match their expertise with the needs of the projects in an agile manner.

Overall, the mentors' domains of expertise ranged from biology to social sciences, design, technology, and data science (Figure

4). One-third of them were working as health or public health professionals.

The CESI consisted of eight volunteer members and included virologists, pharmacists, health economists, experts in the digital sciences and ethics fields, and biologists; members were working at international, national, and local levels of the health system. All of them worked for public or nonprofit organizations. Interviewees were mostly researchers in social sciences and medical practitioners.

Figure 4. Treemap of the 31 Co-Immune experts: domains of expertise (left) and affiliations (right).



Co-Immune Project Assessment

The assessment of projects by experts was designed to be an opportunity for learning and growth. To be assessed, teams were asked to provide a video pitch summarizing their project and detailed documentation on their project page on the JOGL platform, including links to their open access data and code. Project assessment was performed through a grid that was codeveloped by JOGL and the CESI. In addition to grades, teams received detailed feedback on their projects.

The assessment grid was based on a literature review of project evaluation standards and consisted of 10 questions graded from 0 to 5 (Multimedia Appendix 2). Three areas were assessed: the approach, the implementation strategy, and the impact. First, the assessment of the approach included the following: (1) clarity and relevance of the problem and alignment with the program scope, (2) fit between the approach and methodology and the problem statement, and (3) innovation potential (ie, the project introduces groundbreaking objectives, novel concepts, or approaches). Second, the implementation strategy was assessed using following the criteria: (1) state of progress toward set goal (ie, state of advancement), (2) clarity and relevance of the timeline and needs for the future (ie, major tasks and milestones), and (3) project actively engages and aligns with all relevant stakeholders. Finally, the assessment of the impact covered the following: (1) clarity and relevance of the criteria used to measure impact, (2) the extent to which the project considers its ecosystem (ie, ecological, environmental, ethical, and social considerations), (3) sustainability and scalability of the project in the long term, and (4) open and reproducible dissemination strategy. For each of these three categories, JOGL awarded a prize to the project with the best score based on the grades given by reviewers. Additionally, a grand prize was given to the project with the overall highest score. JOGL provided visibility, while two partners also provided an award to a project of their choosing.

JOGL Platform Data Collection and Analysis

Participants added their professional background, skills, and employment status to the JOGL platform. These data were used to evaluate the composition of the community. All users who joined JOGL during the span of the program were considered to be participants of Co-Immune, as it was the only ongoing program, and all outreach activities were related to it.

To better understand how skills were related across participants, we used a network approach to assess similarity between skills and to get further insights about the global diversity of the community. In this network approach, each declared skill was a node and the skills were considered linked if they co-occurred in a participant. Links were then weighted by the number of participants within which they co-occurred. Gephi 0.9.2 was used to represent the network shown in the skill map of the Co-Immune community, and the modularity algorithm was used with default parameters to compute communities representing the sets of skills that tend to co-occur more together than with other skills. Since these skills are linked through the participants who share them, they can be understood as "participant types" constitutive of the Co-Immune community.

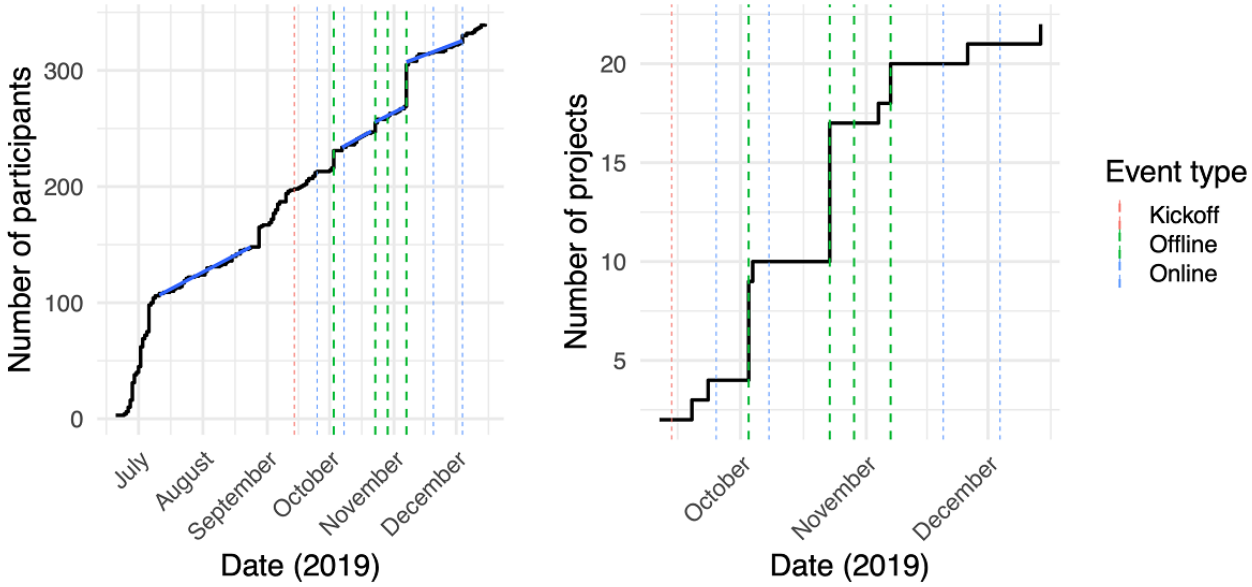
We provide the data related to this study on Zenodo [55]. These data include (1) the link, description, and assessment scores of projects; (2) the profiles of platform users; (3) the description of events; (4) the profiles of experts; and (5) the list and types of partners.

Results

Community Growth Through Events

During the program, 234 participants signed up to the platform (Figure 5). The participant growth was mostly linear over the life span of the program (July 10 to December 18, 2019), suggestive of the potential for continued growth if the program had continued. The growth rate outside of events, at around one per day (between 0.86 and 0.98 users/day), was consistent with the prekickoff growth rate (0.94 users/day). This highlights the importance of events (dashed lines in Figure 5) for driving participant enrollment, with the four offline events accounting for 45% of the growth. In total, offline events were responsible for the generation of 82% (18/22) of the projects. The rest consisted of 4 out of 22 (18%) projects created on the platform outside of events and 2 already-existing projects prior to the program.

Figure 5. Growth of the number of participants (left) and number of projects (right) over the life span of the program. Dashed bars show when events for community facilitation where held (green: offline events; blue: online events; red: kickoff meeting). Blue lines give a linear fit during the corresponding periods, showing stable growth pre- and postkickoff.



Participant Skills and Backgrounds: A Transdisciplinary Community

Out of the 234 participants, 187 (79.9%) declared their job category. The community was composed of a mix of students (67/187, 35.8%) and workers (94/187, 50.3%), most of whom worked full time (81/94, 86%; Figure 6). Other categories included "between jobs" (n=11), "nonprofit" (n=12), and "for profit" (n=3). Out of the 75 participants who declared their country in their JOGL profile, 57% (n=43) were based in France, with the rest coming from other regions, including the rest of Europe, the Americas, Africa, and Asia.

The 234 participants specified a total of 492 unique skills (median 3 [IQR 4.5] skills per participant). We observed a high representation of data science and coding alongside biology, which, altogether, related to the technical skills emphasized during the program (Figure 6). The skill network shows that the community spanned a vast interdisciplinary landscape, from open science to open data and coding, and from project management to biology. The network exhibited the largest connected component of 416 interconnected skills (84.6% of all skills; Figure 7). The modularity maximization (see the Methods section) resulted in the identification of 12 modules corresponding to "participant types" constitutive of the Co-Immune community.

Figure 6. An overview over the Co-Immune community: participant categories (left) and the 20 most represented skills (right) in the Co-Immune community.

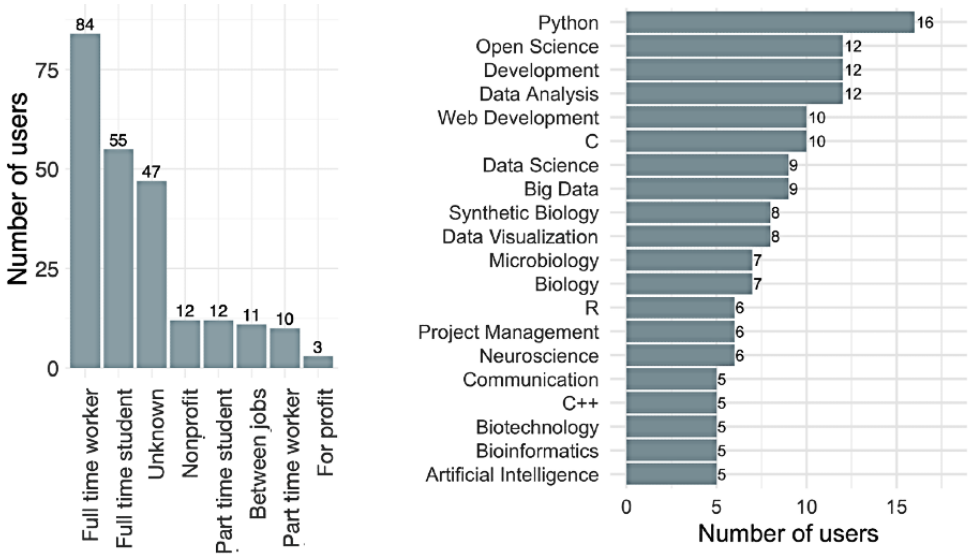
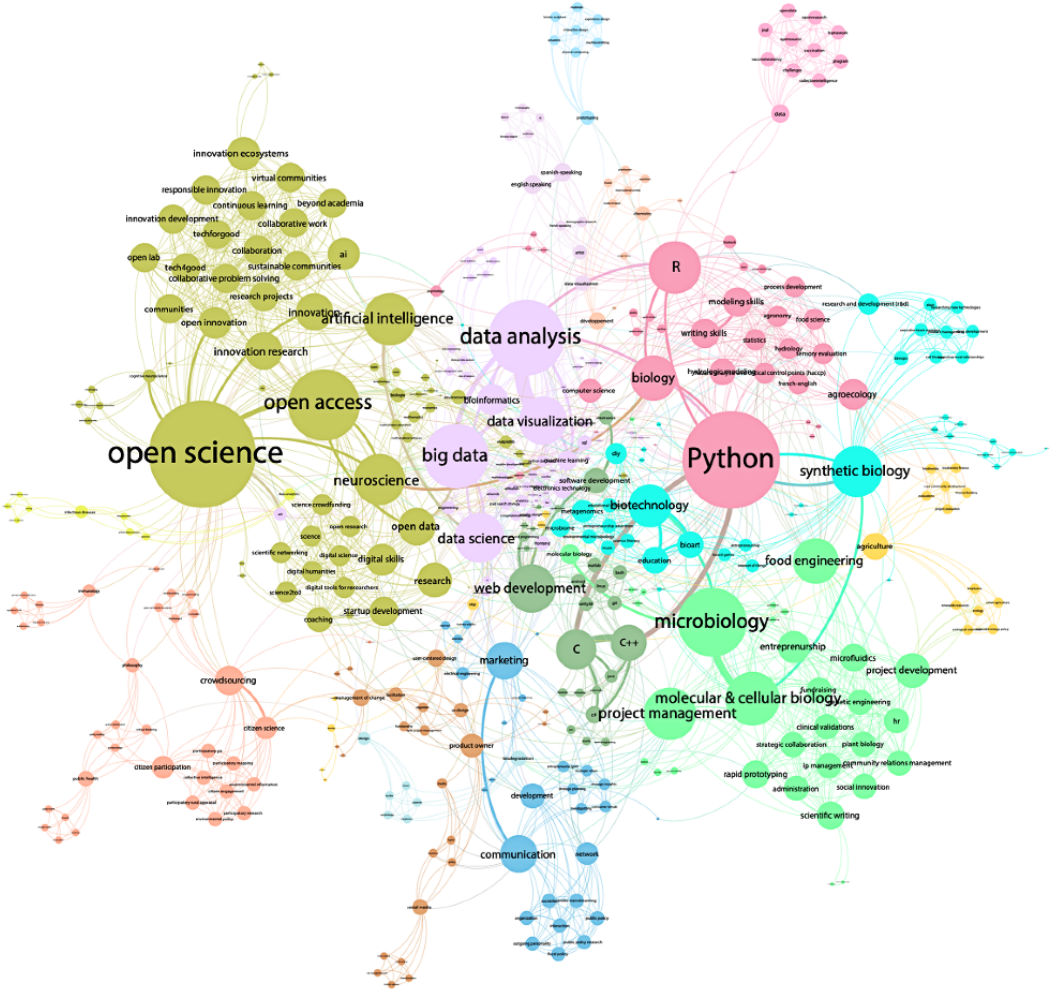


Figure 7. Skill map of the Co-Immune community. Skills are linked if they appear in the profile of the same participant. Link weight indicates the number of participants sharing the skills. Node size indicates weighted degree.



Co-Immune Project Description

A total of 22 projects were created by 20 project leads, with teams of up to 11 members (Table 2 [52,53,56-75]). Among these, 15 (68%) projects proposed to develop software covering web technologies, mobile apps, algorithms, data lakes, data modeling and analysis, and visualization tools. The other 7 (32%) projects included hardware development and interventions involving biotechnologies, game design, behavioral

sciences, education, and communication. Overall, one-third of the projects focused on knowledge transfer.

Among the 15 projects relying on software technology, 11 (73%) aimed at contributing to the production of knowledge by facilitating the analysis of publicly available data; they did this via the use of parsing tools and the creation of repositories (n=3), the analysis of open data (n=3), the development of machine learning tools to extract and analyze Twitter data related to vaccination hesitancy (n=2), and the production of data visualizations (n=3). In particular, more than 40 data sets were identified and collected by 4 projects that were created during the data-centered events. In addition, a database of 2464 tweets, in French, posted over a period of 7 years was made available by a partner, and another data set of 89,979 tweets was gathered by the project Qualitative Analysis of Tweets on Vaccination [56].

Out of the 15 projects above, 4 (27%) used software for knowledge transfer; for instance, the HERA (Health Recording App) project [52] provided educational content and health data storage through its mobile app to improve the monitoring of vaccination and perinatal health among Syrian refugees in Turkey. The Pass It On project [60] focused on role-playing video games directed at health professionals as another method of knowledge transfer. The Neutralizing Information About Vaccines project [70] implemented an algorithm for parsing web pages, helping citizens identify trustworthy content related to vaccines.

A total of 5 projects out of 22 (23%) focused on different interventions (Table 2), including raising awareness about vaccination through an escape game (ie, Vaccination Awareness Escape Game [54]) and communication campaigns on social media (ie, Go Viral! [71]). The HEROIC Santé project [57] developed and tested a short questionnaire using engagement approaches from the social sciences to engage health care professionals and users around the question of flu vaccination. Finally, one team proposed applying synthetic biology methods to tuberculosis vaccines (ie, Project APRICOT [Antigen Presentation Using Crispr for TB] [58]).

Table 2. Co-Immune project descriptions.

Project name	Project status	Solution category	Summary description
HERA[a]: A Health Platform for Refugees [53]	• Assessed[b] • Awarded • Grand prize • Best approach prize • Best impact strategy prize	• Software • Knowledge transfer	A mobile health app designed for improving the monitoring of vaccination and perinatal health of Syrian refugees in Turkey; it provides recall of vaccines, storage of health data, health promotion (educational content), and financial incentives for immunization
Qualitative Analysis of Tweets on Vaccination [56]	• Assessed • Awarded • Partner prize	• Software • Knowledge production	A web-based platform providing real-time visualization and analysis of tweets related to vaccination and vaccination hesitancy; data analysis included sentiment analysis and network analysis; an area of development was the development of predictive models of epidemic occurrence based on Twitter data
Commit to Get Vacc & to Promote Vaccination – HEROIC Santé [57]	• Assessed • Awarded • Best implementation strategy prize	• Intervention • Knowledge transfer	A short questionnaire (7 minutes) using engagement approaches from the human and social sciences, such as "the importance of the source," "voluntary consent," or "fear and danger management," to engage health care professionals and users, not only to be vaccinated against the flu, but also to promote flu vaccination
Project APRICOT[c] [58]	• Assessed • Awarded • Partner prize	• Hardware	Development of a synthetic biology–based methodology that addresses the evasion mechanisms adopted by the mycobacterium tuberculosis and induces the acceleration of lysosomal biogenesis to improve antigen presentation
Vaccination Awareness Escape Game [54]	• Assessed • Not awarded	• Intervention • Knowledge transfer	An escape game to raise vaccination awareness among the general population
Harmonize Vaccination [59]	• Assessed • Not awarded	• Software • Knowledge production	A tool for parsing various formats of vaccination coverage data sets and for visualizing them on a common platform
Pass It On: A Game About Vaccine Hesitancy [60]	• Assessed • Not awarded	• Software • Knowledge transfer	A role-play video game aiming to improve the capacity of health professionals to respond to their patients' hesitation to be vaccinated
Global Vaccination Risk Assessment [61]	• Assessed • Not awarded	• Software • Knowledge production	A tool to create an overview of risk factors of "not getting vaccinated," by country, while looking at the more comprehensive picture; the methodology of this project is based on fuzzy logic, multi-criterion analysis, and the risk triangle
Immuno [62]	• Not assessed[d]	• Hardware • Knowledge transfer	A board game providing access to the general public's understanding of medical sciences related to immunization
Vaccine DataDump [63]	• Not assessed	• Software • Knowledge production	A vaccination-related data repository and analysis tool for quick analysis of vaccine-related issues
Measuring Vaccination Hesitancy From Social Media [64]	• Not assessed	• Software • Knowledge production	Data analysis of social media (ie, Twitter) to examine whether negative sentiment related to vaccination precedes declaration of symptoms and to study the relationship between vaccination hesitancy and epidemiological outbreaks
Mortality According to Access to Vaccines [65]	• Not assessed	• Software • Knowledge production	Data analysis exploring the link between immunization coverage, mortality rate, and distance from health centers
The Health System Matrices [66]	• Not assessed	• Software • Knowledge production	Exploratory analysis of the various parameters influencing vaccination coverage over time
Meta Immune – Data Exploration of Existing DB [67]	• Not assessed	• Software • Knowledge production	A data lake on immunization data

Project name	Project status	Solution category	Summary description
Biloba[e] [68]	• Not assessed	• Intervention	An intervention incentivizing people to increase vaccine uptake through vouchers, supporting the existing mobile app Biloba
Wakuchin Senshi [69]	• Not assessed	• Intervention • Knowledge transfer	An interactive role-play board game to increase awareness about vaccination among the general population
Neutralizing Information About Vaccines [70]	• Not assessed	• Software • Knowledge transfer	An algorithm for parsing web pages, identifying misinformation, and identifying trustworthy content to help users in their health decisions related to vaccines; this also aims to be used by search engines in their recommender systems
Go Viral! [71]	• Not assessed	• Intervention • Knowledge transfer	A communication campaign on social media using gamification methods to illustrate contagion among users and, thereby, increase awareness of the importance of vaccines
Make Vaccines Affordable [72]	• Not assessed	• Software • Knowledge transfer	A web-based portal with data related to population demand for care in order to negotiate prices of vaccines with suppliers
Identify Topics of Discussion in Vaccination Posts [73]	• Not assessed	• Software • Knowledge production	Analysis of discussion in vaccination-related posts on Twitter and their evolution over time
Detect Vaccine Administration in Social Media Patient Data [74]	• Not assessed	• Software • Knowledge production	A classifier able to detect vaccine administration in tweets related to vaccination
Detect Vaccine Hesitancy in Social Media Patient Data [75]	• Not assessed	• Software • Knowledge production	A classifier able to detect vaccine hesitancy in tweets related to vaccination

[a]HERA: Health Recording App.

[b]These were projects that were assessed by experts at the end of the program. To be assessed by a pool of experts, the project team needed to provide detailed documentation of their project, provide a short video pitch, and deposit their data and code on the Just One Giant Lab (JOGL) platform.

[c]APRICOT: Antigen Presentation Using Crispr for TB.

[d]These were projects that were not assessed by experts at the end of the program because they did not provide sufficient documentation.

[e]The Biloba project, which was not part of Co-Immune, was used as a base to create the team's own project, as the Biloba founder was a mentor during this event.

Co-Immune Project Assessment

Out of 22 projects, 7 (32%) provided sufficient documentation on JOGL to be assessed by the pool of independent experts. In total, 27 reviews were performed, yielding scores ranging from 18 to 32.8 out of a possible total of 45 across the different dimensions that were assessed (ie, approach, implementation strategy, and impact). The average score was 25.1 (SD 6.4).

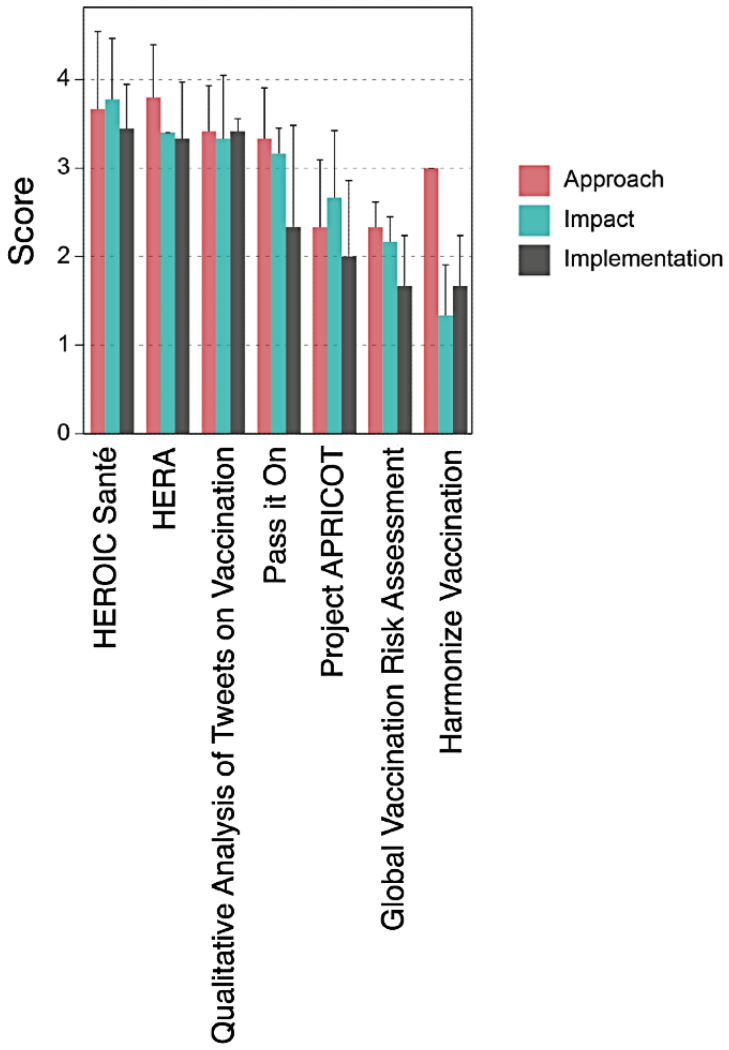
HERA: A Health Platform for Refugees [53] was awarded with prizes, based on a total score of 15, for best approach (mean score 11.4, SD 2) and impact (mean score 14.6, SD 3.2). Commit to Get Vacc & to Promote Vaccination – HEROIC Santé [57] was awarded the best implementation strategy prize (mean score 10.33, SD 2.5).

The projects were more successful, globally, in terms of approach, with a mean score of 9.37 (SD 1.79) out of 15 points. Out of 7 projects that were assessed, 4 (57%; Figure 8) had a score higher than 4 out of 5 for clarity, relevance, and alignment of their problem statement with the program objectives. For 6 projects (86%), the fit between the methods and the projects' objectives was scored highly by reviewers, with a score of at least 3 out of 5.

The implementation strategy score of projects was low, overall, given the early stage of the projects at the time of review. As such, only projects that existed prior to the program—HERA [52] and HEROIC Santé [57]—got a score of at least 3 out of 5.

For winners in each category, JOGL awarded them physical space for showcasing their project during the 2020 ChangeNOW forum at the Grand Palais in Paris as well as tickets for the Maddy Keynote, a major innovation event in Paris. Two partners—Excelya and the Wild Code School—also provided awards to the projects of their choice. Additionally, the Qualitative Analysis of Tweets on Vaccination [56] project was chosen to be the focus of a hackathon by the Wild Code School, and Project APRICOT [58] was offered technical support for data science and legal and regulatory affairs by Excelya.

Figure 8. Bar plot of review scores per category for all reviewed projects. Bars show average values for all questions related to each category, and error bars represent SDs. Projects are shown by decreasing global score. APRICOT: Antigen Presentation Using Crispr for TB; HERA: Health Recording App.



Discussion

Principal Findings

The Co-Immune program was designed to foster the creation and development of citizen science and open innovation projects addressing the contemporary challenges of vaccination in France and around the globe by reaching four specific objectives: (1) to foster collaborative, open, and transdisciplinary dynamics; (2) to promote the emergence of accessible knowledge and innovative solutions; (3) to support participants in the elaboration and development of their projects; and (4) to disseminate the outputs and results in an open science framework. Below, we discuss to what extent Co-Immune reached these objectives and highlight the challenges and facilitators in implementing such a program.

First, the program succeeded in creating a collaborative and transdisciplinary environment through its three core features: the JOGL platform, the organization of events, and the contest approach. This led to forming partnerships with 13 different organizations and recruiting over 230 participants, who displayed 492 unique skills and were engaged in creating 22 projects. The use of on-site hackathons was beneficial in gathering nonacademic participants from various backgrounds. Our data show that in-person events and local outreach played a significant role in growing the community around Co-Immune. These offline events recruited 45% of the total community. Local enrollment was further strengthened by local partnerships, such as higher education organizations. However, the localization of our on-site events in Paris did not allow for the participation of people living in other parts of France or the rest of the world. Additionally, our online communication restricted the access of the online events to our realm of influence and to people with an internet connection. More inclusive participation geared toward people with diverse socioeconomic statuses and geographic situations is desirable in the future to give them agency over solving the problems that affect them. The development of new communities is usually a slow process in the absence of exogenous shocks, such as the surge in collaborative communities created by the COVID-19 pandemic [76]. Tapping into existing projects and networks for events has proven to be fruitful in our case, allowing for a steady growth of the Co-Immune community up until the end of the program. However, we did not observe further growth of the community after the end of the program. This highlights that in order to build a sustainable community using open innovation to tackle

global health challenges, one needs to facilitate the entry and exit of members, provide resources to support the current ones, focus on building on existing communities and projects, design inclusive environments for collaboration, and empower members to run their own activities.

Second, two design elements of the program converged to promote the emergence of knowledge and solutions to address aspects of access to vaccines and vaccination hesitancy: (1) the identification of challenges by experts in the field and (2) the alignment of the program strategy with national and international policies by frequent consultation with public health bodies and mobilization of members of public institutions in the CESI. Yet, greater representation of people affected by poor access to vaccines and people who are hesitant would be desirable to strengthen the alignment between the solutions developed and the most pressing needs at the local level.

Recently, online events have been used widely during the COVID-19 pandemic [76-78], supporting our initial assumption that forming and animating a distributed online community for public health programs is a relevant approach.

Third, the use of the JOGL platform, the mentorship during events, the assessment and feedback from experts, and the connection with a wide range of partners supported participants in the elaboration of their project in an efficient way. The use of the JOGL platform enabled projects to gain visibility, list their needs to create interfaces for collaboration, and share open data sets, code, and tools. Indeed, online platforms can offer projects that started at hackathons a pathway to pursue their development, potentially alleviating one of the main drawbacks of such short temporal interventions [43]. In this case, it also enabled the program coordinators to connect participants with project leaders based on a match between needs and skills. Yet, this approach was time-consuming, and scaling up our efforts proved to be challenging. The automation of such matchmaking tasks through a recommender system would help to minimize these efforts and increase the impact of projects through accelerated development [79]. In addition, mentoring is a known strategy that is used by open, online communities [80,81] and was leveraged by the Co-Immune program. Given the diversity of backgrounds and level of expertise across the participants, it was necessary to engage a similar diversity among the mentors. In our context, the highly rated projects that eventually received awards did not originate or participate in hackathons, but rather benefited from Co-Immune as a platform for further growth. Several of these projects already existed before the start of Co-Immune and had a higher maturity level than the projects created during the short span of the program. In addition, these projects were launched and run by people outside the larger Paris region. Thus, we stress the potential of online platforms and open innovation to build on existing projects and to replicate, adapt, and scale their activities in other contexts. Additional support consisted of promoting visibility on social media by the organization team as well as opportunities for networking during events. Although no financial compensation was provided as part of this program, partners, through their own experts and co-organizing events, engaged in close relationships with JOGL and the individual projects. This was favorable for sustaining collaborations and projects after the end of the program. In the future, the sustainability of the newly created project efforts could potentially be improved by using incentives, such as microgrants or fellowship programs, for continuing projects in the postprogram period [79]. While the short time frame and limited resources allocated to the program did not allow us to implement a strong monitoring and evaluation strategy, future implementations should ensure that they conduct a minimum of pre- and postprogram data collection for assessing the full impact of the program.

Finally, the open science environment of this program was not only an asset for disseminating the outputs and results of the projects developed, but it also enabled them to replicate initiatives and, thereby, accelerate the resolution of the global health challenges they address. An example of this was given by the team from the project HERA: A Health Platform for Refugees [52], who opened its code, enabling any individual to replicate it. However, the lack of a thorough evaluation strategy prevents us from reaching a more definitive conclusion on the effective replication of projects carried out in Co-Immune.

Co-Immune showcases that short, focused programs can be efficient at mobilizing diverse communities in a rapid manner and harvesting ideas from various domains to address global health challenges. Yet, more case studies and evaluation work on similar programs are necessary to assess the full relevance of their design and the impact of the projects that are developed within them.

Conclusions

Co-Immune highlights how open innovation approaches and online platforms can help to gather and coordinate noninstitutional communities in a rapid, distributed, and global way toward solving SDG-related issues. The Co-Immune program gathered participants and partners from various backgrounds in a newly formed community to facilitate the creation of new projects as well as the continuation of existing projects to address the issues of vaccination hesitancy and access. In an open framework, the projects made their data, code, and solutions publicly available.

Through hackathons and other contest approaches, such initiatives can lead to the production and transfer of knowledge, creating novel solutions in the public health sector. The example of Co-Immune contributes to paving the way for organizations and individuals to collaboratively tackle future global challenges.

Acknowledgments

First, we would like to thank all the Co-Immune participants that made the program possible by bringing their creativity, skills, and insights to address contemporary public health challenges. We thank Sanofi for funding this program, especially Diane Brément and Nansa Burlet for their assistance. We thank the Just One Giant Lab (JOGL) team for their work on coordinating the

Co-Immune program, with special efforts from Lola Casamitjana and Marine Vouard. We thank the Center for Research and Interdisciplinarity (CRI), Paris; Epitech Paris; Sup Biotech; and the Wild Code School for their support in organizing events with their students; Kap Code, Excelya, Girls in Tech, CorrelAid, and Data for Good for their guidance and technical assistance to participants and projects; and S^{3}Odéon, TUBÀ, ChangeNOW, and the Maddy Keynote for the visibility they provided to this program. We thank the interviewees at the 7th Fondation Merieux Vaccine Acceptance conference for highlighting the key issues to address and potential solutions on which participants could build. We thank the mentors for the support they provided to projects and participants throughout the duration of the Co-Immune program. We thank Enric Senabre Hidalgo for insightful comments during the final stages of writing this manuscript. Finally, we thank the members of the Committee for Ethics, Science and Impact (CESI)—Gilles Babinet, Jérôme Béranger, Anshu Bhardwaj, Liem Binh Luong Nguyen, Mélanie Heard, Ariel Lindner, Juliette Puret, and Olivier Rozaire—for their valuable input that allowed the creation and implementation of a framework for ethics, science, and impact for the Co-Immune program and the independent assessment of its projects.

Authors' Contributions

TL and IV co-designed the theme and the scope of the program at the early stage of the initiative. CM and TL conceived the program. CM led the coordination team of the program. CM, BGT, TL, CLBG, and MS participated in the program implementation. BGT, RJ, and MS analyzed data. CM, BGT, GF, CLBG, and MS wrote the paper.

Conflicts of Interest

GF was paid by Just One Giant Lab (JOGL) to support the elaboration of the manuscript; provided consulting services to Vaccines Europe, a trade association based in Belgium; and is a volunteer board member of the Coalition for Life-Course Immunisation, a UK-based charity. CM worked as the Co-Immune program coordinator for JOGL. BGT was an unpaid advisor to JOGL. JOGL received funding from Sanofi to develop and implement the program, which included publication fees. Sanofi respected the strict independence of JOGL, which administers its platform and the Co-Immune page in complete autonomy. Similarly, the Committee for Ethics, Science and Impact (CESI) was independent of Sanofi and decided alone on the strategic and scientific orientations of the program and the best projects to be rewarded.

Multimedia Appendix 1

Co-Immune partners and supplementary method for comparing Just One Giant Lab (JOGL) with other platforms.
[PDF File (Adobe PDF File), 196 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Co-Immune project assessment grid.
[PDF File (Adobe PDF File), 213 KB-Multimedia Appendix 2]

References

1. Nosek B, Alter G, Banks G, Borsboom D, Bowman SD, Breckler SJ, et al. SCIENTIFIC STANDARDS. Promoting an open research culture. Science 2015 Jun 26;348(6242):1422-1425 [FREE Full text] [doi: 10.1126/science.aab2374] [Medline: 26113702]
2. Burgelman J, Pascu C, Szkuta K, Von Schomberg R, Karalopoulos A, Repanas K, et al. Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. Front Big Data 2019;2:43 [FREE Full text] [doi: 10.3389/fdata.2019.00043] [Medline: 33693366]
3. Elliott K, Resnik D. Making open science work for science and society. Environ Health Perspect 2019 Jul;127(7):075002 [FREE Full text] [doi: 10.1289/ehp4808]
4. Fraisl D, Campbell J, See L, Wehn U, Wardlaw J, Gold M, et al. Mapping citizen science contributions to the UN sustainable development goals. Sustain Sci 2020 Jul 02;15(6):1735-1751 [FREE Full text] [doi: 10.1007/s11625-020-00833-7]
5. Auffray C, Balling R, Blomberg N, Bonaldo MC, Boutron B, Brahmachari S, et al. COVID-19 and beyond: A call for action and audacious solidarity to all the citizens and nations, it is humanity's fight. F1000Res 2020 Sep 14;9:1130 [FREE Full text] [doi: 10.12688/f1000research.26098.1]
6. United Nations Statistics Division. Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development. New York, NY: United Nations; 2017 Jul. URL: https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%202021%20refinement_Eng.pdf [accessed 2021-12-21]
7. Immunization, Vaccines and Biologicals Team. Global Vaccine Action Plan 2011-2020. Geneva, Switzerland: World Health Organization; 2013. URL: https://www.who.int/publications/i/item/global-vaccine-action-plan-2011-2020 [accessed 2021-12-21]
8. Immunization coverage. World Health Organization. 2021 Jul 15. URL: https://www.who.int/news-room/fact-sheets/detail/immunization-coverage [accessed 2021-12-21]

9. Ten threats to global health in 2019. World Health Organization. URL: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 [accessed 2020-10-08]
10. Immunization Agenda 2030: A Global Strategy to Leave No One Behind. Geneva, Switzerland: World Health Organization; 2020. URL: https://tinyurl.com/4hhuzvru [accessed 2021-12-21]
11. Chan M, Jamieson K, Albarracin D. Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the United States. Vaccine 2020 Sep 11;38(40):6236-6247 [FREE Full text] [doi: 10.1016/j.vaccine.2020.07.054] [Medline: 32792251]
12. Wouters OJ, Shadlen KC, Salcher-Konrad M, Pollard AJ, Larson HJ, Teerawattananon Y, et al. Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment. Lancet 2021 Feb 12:1023-1034 [FREE Full text] [doi: 10.1016/S0140-6736(21)00306-8] [Medline: 33587887]
13. Vaccine Safety Net. Geneva, Switzerland: World Health Organization URL: https://www.vaccinesafetynet.org/ [accessed 2020-10-08]
14. Boost Community. URL: https://boostcommunity.org/ [accessed 2020-10-08]
15. Finnegan G. Has Finland found an antivax antidote? Vaccines Today. 2019 Jul 08. URL: https://www.vaccinestoday.eu/stories/has-finland-found-an-antivax-antidote/ [accessed 2020-10-09]
16. Vaccine Acceptance & Demand. Washington, DC: Sabin Vaccine Institute URL: https://vaccineacceptance.org/ [accessed 2020-10-09]
17. Voice, Agency, Empowerment - Handbook on Social Participation for Universal Health Coverage. Geneva, Switzerland: World Health Organization; 2021. URL: https://www.who.int/publications-detail-redirect/9789240027794 [accessed 2021-08-14]
18. Marston C, Renedo A, Miles S. Community participation is crucial in a pandemic. Lancet 2020 May 30;395(10238):1676-1678 [FREE Full text] [doi: 10.1016/S0140-6736(20)31054-0] [Medline: 32380042]
19. West J, Lakhani K. Getting clear about communities in open innovation. Ind Innov 2009 Sep;15(2):223-231 [FREE Full text] [doi: 10.1080/13662710802033734]
20. Haklay M, Fraisl D, Greshake Tzovaras B, Hecker S, Gold M, Hager G, et al. Contours of citizen science: A vignette study. R Soc Open Sci 2021 Aug;8(8):202108 [FREE Full text] [doi: 10.1098/rsos.202108] [Medline: 34457323]
21. Hsiang Liao C. How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. Scientometrics 2010 Nov 2;86(3):747-761 [FREE Full text] [doi: 10.1007/s11192-010-0309-2]
22. Juan AA, Daradoumis T, Roca M, Grasman SE, Faulin J, editors. Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications. Hershey, PA: IGI Global; 2012.
23. Polk M. Transdisciplinary co-production: Designing and testing a transdisciplinary research framework for societal problem solving. Futures 2015 Jan;65:110-122. [doi: 10.1016/j.futures.2014.11.001]
24. Leavy P. Essentials of Transdisciplinary Research: Using Problem-Centered Methodologies. Walnut Creek, CA: Left Coast Press, Inc; 2011.
25. Stokols D. Toward a science of transdisciplinary action research. Am J Community Psychol 2006 Sep;38(1-2):63-77. [doi: 10.1007/s10464-006-9060-5] [Medline: 16791514]
26. Nielsen M. Reinventing Discovery: The New Era of Networked Science. Princeton, NJ: Princeton University Press; 2011.
27. Franzoni C, Sauermann H. Crowd science: The organization of scientific research in open collaborative projects. Res Policy 2014 Feb;43(1):1-20 [FREE Full text] [doi: 10.1016/j.respol.2013.07.005]
28. Vayena E, Brownsword R, Edwards S, Greshake B, Kahn JP, Ladher N, et al. Research led by participants: A new social contract for a new kind of research. J Med Ethics 2016 Apr;42(4):216-219 [FREE Full text] [doi: 10.1136/medethics-2015-102663] [Medline: 25825527]
29. Greshake Tzovaras B, Angrist M, Arvai K, Dulaney M, Estrada-Galiñanes V, Gunderson B, et al. Open Humans: A platform for participant-centered research and personal data exploration. Gigascience 2019 Jun 01;8(6):1-13 [FREE Full text] [doi: 10.1093/gigascience/giz076] [Medline: 31241153]
30. Lewis DM, Leibrand S, Street TJ, Phatak SS. Detecting insulin sensitivity changes for individuals with type 1 diabetes. Diabetes 2018 Jul;67(Supplement 1):79-LB [FREE Full text] [doi: 10.2337/db18-79-lb]
31. Afuah A, Tucci C. Crowdsourcing as a solution to distant search. Acad Manage Rev 2012 Jul;37(3):355-375. [doi: 10.5465/amr.2010.0146]
32. Sitruk Y, Kazakçi A. Crowd-based data-driven hypothesis generation from data and the organisation of participative scientific process. In: Proceedings of the DESIGN 2018 15th International Design Conference. 2018 Presented at: DESIGN 2018 15th International Design Conference; May 21-24, 2018; Dubrovnik, Croatia p. 1673-1684 URL: https://doi.org/10.21278/idc.2018.0510 [doi: 10.21278/idc.2018.0510]
33. Anderson C. The end of theory: The data deluge makes the scientific method obsolete. Wired. 2008 Jun 23. URL: https://www.wired.com/2008/06/pb-theory/ [accessed 2021-12-21]
34. Bosman J, Bruno I, Chapman C, Greshake Tzovaras B, Jacobs N, Kramer B, et al. The Scholarly Commons - Principles and practices to guide research communication. OSF Preprints. 2017 Sep 15. URL: https://osf.io/6c2xt/ [accessed 2021-12-21]
35. Hecker S, Haklay M, Bowser A, Makuch Z, Vogel J, Bonn A, editors. Citizen Science: Innovation in Open Science, Society and Policy. London, UK: UCL Press; Oct 15, 2018.

36. Katapally T. The SMART framework: Integration of citizen science, community-based participatory research, and systems science for population health science in the digital age. JMIR Mhealth Uhealth 2019 Aug 30;7(8):e14056 [FREE Full text] [doi: 10.2196/14056] [Medline: 31471963]
37. Rowbotham S, McKinnon M, Leach J, Lamberts R, Hawe P. Does citizen science have the capacity to transform population health science? Crit Public Health 2017 Nov 09;29(1):118-128. [doi: 10.1080/09581596.2017.1395393]
38. Ramatowski JW, Lee CX, Mantzavino A, Ribas J, Guerra W, Preston ND, et al. Planning an innovation marathon at an infectious disease conference with results from the International Meeting on Emerging Diseases and Surveillance 2016 Hackathon. Int J Infect Dis 2017 Dec;65:93-97 [FREE Full text] [doi: 10.1016/j.ijid.2017.09.025] [Medline: 29017856]
39. Angelidis P, Berman L, Casas-Perez MDLL, Celi LA, Dafoulas GE, Dagan A, et al. The hackathon model to spur innovation around global mHealth. J Med Eng Technol 2016;40(7-8):392-399 [FREE Full text] [doi: 10.1080/03091902.2016.1213903] [Medline: 27538360]
40. Kienzler H, Fontanesi C. Learning through inquiry: A Global Health Hackathon. Teach High Educ 2016 Sep 23;22(2):129-142 [FREE Full text] [doi: 10.1080/13562517.2016.1221805]
41. D'Ignazio C, Klein LF. Data Feminism. Cambridge, MA: The MIT Press; Mar 2020.
42. Costanza-Chock S. Design Justice: Community-Led Practices to Build the Worlds We Need. Cambridge, MA: The MIT Press; Mar 2020.
43. DePasse J, Carroll R, Ippolito A, Yost A, Santorino D, Chu Z, et al. Less noise, more hacking: How to deploy principles from MIT's hacking medicine to accelerate health care. Int J Technol Assess Health Care 2014 Aug 06;30(3):260-264 [FREE Full text] [doi: 10.1017/s0266462314000324]
44. Day K, Humphrey G, Cockcroft S. How do the design features of health hackathons contribute to participatory medicine? Australas J Inf Syst 2017 Mar 08;21:1-20 [FREE Full text] [doi: 10.3127/ajis.v21i0.1383]
45. Pan S, Stein G, Bayus B, Tang W, Mathews A, Wang C, et al. Systematic review of innovation design contests for health: Spurring innovation and mass engagement. BMJ Innov 2017;3:227-237 [FREE Full text] [doi: 10.1136/bmjinnov-2017-000203] [Medline: 29576873]
46. Benchoufi M, Fournier M, Magrez D, Macaux G, Barué V, Mansilla Sanchez A, et al. Epidemium: A multidisciplinary community to tackle cancer using big and open data. In: Proceedings of the American Society of Clinical Oncology (ASCO) Annual Meeting. 2018 Presented at: American Society of Clinical Oncology (ASCO) Annual Meeting; June 1-5, 2018; Chicago, IL p. e13604 URL: https://ascopubs.org/doi/10.1200/JCO.2018.36.15_suppl.e13604 [doi: 10.1200/jco.2018.36.15_suppl.e13604]
47. Badger E. Are civic hackathons stupid? Bloomberg. 2013 Jul 05. URL: https://www.bloomberg.com/news/articles/2013-07-05/are-civic-hackathons-stupid [accessed 2021-12-21]
48. Tress B, Tress G, Fry G, Opdam P, editors. From Landscape Research to Landscape Planning: Aspects of Integration, Education and Application. Dordrecht, The Netherlands: Springer; 2006.
49. Argote L, Ingram P. Knowledge transfer: A basis for competitive advantage in firms. Organ Behav Hum Decis Process 2000 May;82(1):150-169. [doi: 10.1006/obhd.2000.2893]
50. Just One Giant Lab (JOGL). URL: https://app.jogl.io/ [accessed 2021-12-22]
51. Attwell K, Betsch C, Dubé E, Sivelä J, Gagneur A, Suggs LS, et al. Increasing vaccine acceptance using evidence-based approaches and policies: Insights from research on behavioural and social determinants presented at the 7th Annual Vaccine Acceptance Meeting. Int J Infect Dis 2021 Apr;105:188-193 [FREE Full text] [doi: 10.1016/j.ijid.2021.02.007] [Medline: 33578012]
52. Peach K, Berditchevskaia A, Bass T. Collective Intelligence Design Playbook (beta). Nesta. URL: https://www.nesta.org.uk/toolkit/collective-intelligence-design-playbook/ [accessed 2021-12-21]
53. HERA: A Health Platform for Refugees. URL: https://app.jogl.io/project/18 [accessed 2021-12-22]
54. Vaccination Awareness Escape Game. URL: https://app.jogl.io/project/22 [accessed 2021-12-22]
55. Masselot CM, Greshake Tzovaras B, Graham CLB, Jeyaram R, Finnegan G, Vitali I, et al. Co-Immune: A case study on open innovation for vaccination hesitancy and access [Data set]. Zenodo. 2021. URL: https://zenodo.org/record/4560273#.YVYlGZpBzb2 [accessed 2021-09-30]
56. Qualitative Analysis of Tweets on Vaccination. URL: https://app.jogl.io/project/20 [accessed 2021-12-22]
57. Commit to Get Vacc & to Promote Vaccination - HEROIC Santé. URL: https://app.jogl.io/project/115 [accessed 2021-12-22]
58. Project APRICOT. URL: https://app.jogl.io/project/117 [accessed 2021-12-22]
59. Harmonize Vaccination. URL: https://app.jogl.io/project/62 [accessed 2021-12-22]
60. Pass It On: A Game About Vaccine Hesitancy. URL: https://app.jogl.io/project/70 [accessed 2021-12-22]
61. Global Vaccination Risk Assessment. URL: https://app.jogl.io/project/73 [accessed 2021-12-22]
62. Immuno. URL: https://app.jogl.io/project/9 [accessed 2021-12-22]
63. Vaccine DataDump. URL: https://app.jogl.io/project/26 [accessed 2021-12-22]
64. Measuring Vaccination Hesitancy From Social Media. URL: https://app.jogl.io/project/59 [accessed 2021-12-22]
65. Mortality According to Access to Vaccines. URL: https://app.jogl.io/project/60 [accessed 2021-12-22]
66. The Health System Matrices. URL: https://app.jogl.io/project/61 [accessed 2021-12-22]
67. Meta Immune - Data Exploration of Existing DB. URL: https://app.jogl.io/project/63 [accessed 2021-12-22]

68. Biloba. URL: https://app.jogl.io/project/67 [accessed 2021-12-22]
69. Wakuchin Senshi. URL: https://app.jogl.io/project/68 [accessed 2021-12-22]
70. Neutralizing Information About Vaccines. URL: https://app.jogl.io/project/69 [accessed 2021-12-22]
71. Go Viral!. URL: https://app.jogl.io/project/71 [accessed 2021-12-22]
72. Make Vaccines Affordable. URL: https://app.jogl.io/project/72 [accessed 2021-12-22]
73. Identify Topics of Discussion in Vaccination Posts. URL: https://app.jogl.io/project/74 [accessed 2021-12-22]
74. Detect Vaccine Administration in Social Media Patient Data. URL: https://app.jogl.io/project/78 [accessed 2021-12-22]
75. Detect Vaccine Hesitancy in Social Media Patient Data. URL: https://app.jogl.io/project/79 [accessed 2021-12-22]
76. Luo EM, Newman S, Amat M, Charpignon M, Duralde ER, Jain S, et al. MIT COVID-19 Datathon: Data without boundaries. BMJ Innov 2021 Jan;7(1):231-234 [FREE Full text] [doi: 10.1136/bmjinnov-2020-000492] [Medline: 33437494]
77. Kinder F, Harvey A. COVID-19: The medical students responding to the pandemic. BMJ 2020 Jun 15;369:m2160. [doi: 10.1136/bmj.m2160] [Medline: 32540952]
78. Brereton B. EUvsVirus Hackathon project: A case study from a mentor's perspective. All Irel J Teach Learn Higher Educ 2020 Jun 30;12(2):1-8 [FREE Full text]
79. Kim J, Kim H, Oh H, Ryu Y. A group recommendation system for online communities. Int J Inf Manage 2010 Jun;30(3):212-219. [doi: 10.1016/j.ijinfomgt.2009.09.006]
80. Fagerholm F, Guinea A, Münch J, Borenstein J. The role of mentoring and project characteristics for onboarding in open source software projects. In: Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. New York, NY: ACM Press; 2014 Presented at: 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement; September 18-19, 2014; Torino, Italy p. 1-10. [doi: 10.1145/2652524.2652540]
81. Schilling A, Laumer S, Weitzel T. Train and retain: The impact of mentoring on the retention of FLOSS developers. In: Proceedings of the 50th Annual Conference on Computers and People Research. New York, NY: ACM Press; 2012 Presented at: 50th Annual Conference on Computers and People Research; May 31-June 2, 2012; Milwaukee, WI p. 79-84. [doi: 10.1145/2214091.2214112]

Abbreviations

APRICOT: Antigen Presentation Using Crispr for TB
CESI: Committee for Ethics, Science and Impact
CRI: Center for Research and Interdisciplinarity
HERA: Health Recording App
JOGL: Just One Giant Lab
SDG: Sustainable Development Goal
WHO: World Health Organization

Edited by G Eysenbach, S Woods; submitted 18.07.21; peer-reviewed by S Rowbotham, M Das; comments to author 17.08.21; revised version received 30.09.21; accepted 03.10.21; published 21.01.22

Please cite as:
Masselot C, Greshake Tzovaras B, Graham CLB, Finnegan G, Jeyaram R, Vitali I, Landrain T, Santolini M
Implementing the Co-Immune Open Innovation Program to Address Vaccination Hesitancy and Access to Vaccines: Retrospective Study
J Particip Med 2022;14(1):e32125
URL: https://jopm.jmir.org/2022/1/e32125
doi: 10.2196/32125
PMID:

Membrane staining and phospholipid tracking in *Pseudomonas aeruginosa* PAO1 using the phosphatidylcholine mimic propargyl-choline

Author names

Chris LB Graham[1]. https://orcid.org/0000-0002-4271-6731 , Jack A. Bryant[2,3] https://orcid.org/0000-0002-7912-2144 , Manuel Banzhaf[2,4], https://orcid.org/0000-0002-4682-1037 , David I Roper[1].https://orcid.org/0000-0003-0009-621X

Affiliation(s)

1. School of Life Sciences, University of Warwick, Coventry, UK
2. School of Biosciences, University of Birmingham, Birmingham, UK
3. School of Life Sciences, University of Nottingham, Nottingham, UK
4. Newcastle University Biosciences Institute, Newcastle University, Newcastle, UK

Corresponding author and email address

Corresponding author: Chris L.B Graham email: Chris.L.B.Graham@warwick.ac.uk

Keywords

Propargyl-choline, Cy3, Dye, Pseudomonas aeruginosa, membrane label, lipid tracking, phosphatidylcholine

Abstract

The use of membrane-specific dyes for *in vivo* fluorescent microscopy is commonplace. However, most of these reagents are non-specific and cannot track specific lipid species movement, instead often acting as non-covalent lipid associated probes or requiring uptake of whole lipids and acyl tails into the membrane. This issue has been solved in eukaryotic cell biology by use of click-chemistry liable phospholipid headgroup pulse-labels. Here we describe a method for *in vivo* phospholipid labelling by fluorescent imaging in *Pseudomonas aeruginosa* using a phosphatidylcholine (PC) mimic, “propargyl-choline”(PCho). This click-chemistry liable headgroup mimic is visible by microscopy and allows the covalent labelling of lipids. Fluorescence of the cell membranes, visible in heterogenous patches, is dependent on PCho concentration and is localised in the membrane fraction of cells, demonstrating that it is suitable for membrane labelling and cell imaging.

Impact statement

Pseudomonas aeruginosa is an opportunistic pathogen. Pathogenicity and antibiotic resistance of the organism can be partly attributed to the presence of phosphatidylcholine lipids and more broadly the cell envelope. In 2019 more than 82,000 people died due to *P. aeruginosa* strains with resistance to one or more antibiotic treatments (1). In order to enable better study and understanding of *Pseudomonas sp.* lipids we describe an *in vivo* method to label *Pseudomonas aeruginosa* PC lipids and describe their subsequent visualisation by click-labelling and fluorescent microscopy. The phospholipid headgroup mimic propargyl-choline(PCho), substituting for a phosphatidyl choline headgroup(PC), has previously been used in mammals as a click-able microscopy label for use in membrane stability assays in engineered bacteria. Here we show its use in 'wild-type' bacterial cells, as a method to visualise the movements and localisation of membranes, similar to FM4-64 and applicable in situations where the tracking of a specifically labelled membrane lipid is useful. The ability to image a lipid mimic such as PCho in a model species such as *Pseudomonas aeruginosa* using PCho in bacteria could also, as in eukaryotes provide insight on lipid related organisations, growth and replication stages of bacteria in general not yet touched on.

Data summary

The microscope images, code and GIFs of lipid movement can all be found at the Figshare X. – To be uploaded once a link is available to upload.

SI Figure 1. Cell Viability and propargyl-choline concentration - at end of document in this file.

SI Figure 2. GIF of lipid movement

https://drive.google.com/file/d/1c77Fqm3cbA5RYsM94975deZDZDoUE0mc/view?usp=sharing and

SI Figure 3. GIF of lipid tracking

https://drive.google.com/file/d/1ArmTBPJKB4bpIuFkKCnvepX-5EVovovY/view?usp=sharing

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Introduction

The insertion and maintenance of lipids in the inner membrane and inner leaflet of the outer membrane of Gram-negative bacteria is not yet fully understood. Whilst labelling of cells with lipid probes has revealed lipid raft localisation, cardiolipin localisation (2,3,4), and changes in phospholipid abundance over time, few methods can track covalently modified lipids movement. Fluorescence-labelling techniques for microscopy typically monitor lipid movement and localisation using probes able to detect specific lipid headgroups (3) or by labelling the lipid tail (5). The phosphatidylcholine (PC) mimic, propargyl-choline (PCho), has recently been used to label phospholipids in *E. coli* cells, which were modified to include phosphatidylcholine metabolism. PCho labelled cells were suggested to have PCho in the inner and outer membranes, however PC biosynthesis is not native *to E. coli* and the study did not focus on the labelling and imaging process (6). Here we chose to study the distribution of the phosphatidylcholine mimic in wildtype *Pseudomonas aeruginosa PAO1* due to this organism natively having phosphatidylcholine in the cell envelope.

Pseudomonas aeruginosa is an opportunistic pathogen, with pathogenicity as well as antibiotic resistance associated with its cell envelope structure (6). PC has been shown to be required for efficient infection, through use of PC as an inflammation facilitator and pre-cursor for lung damage (8). Similar pathogens also scavenge PC from the host (9) which may play a role in growth, and PC has been shown to be required for cytotoxin production in related *Pseudomonas sp* (10). Therefore labeling technologies that enable analysis of exchanges and dynamics of lipids, especially phosphatidylcholine, would be a valuable tool.

Mimics for phosphatidylcholine have been developed for mammalian studies. In the mammalian system, phosphatidylcholine metabolism can be directly supplemented with soluble choline mimics, which do not affect other aspects of core cell metabolism (11). PCho has yet to be used in wildtype Gram-negative bacteria for PC pulse-chase labelling. However, the capacity to label, image and track phospholipids in Gram-negative bacteria would enable this process of lipid insertion and dynamics to be studied simultaneously with peptidoglycan biogenesis through the simultaneous use of fluorescent *D*-amino acid mimics to address questions of how these processes are coordinated in the cell (12).

Therefore, in this study we establish the use of PCho (11) to determine localisation of the phospholipid phosphatidylcholine in *P. aeruginosa* after chemical crosslinking to an azide group. We found PCho to be of similar technical use to existing membrane labels such as FM464, however with potential for tracking this specific lipid species distribution and behaviour rather than displaying an affinity for detecting or binding to lipids in general (13). The use of PCho, a soluble headgroup rather than lipid tail addition also removes the need for membrane perturbation when using full phospholipid addition (14) or use of mutant cells with alternative enzyme pathways to wildtype (6).

Results

The phosphatidylcholine headgroup mimic propargyl-choline localises to *Pseudomonas* membranes

In order to determine the efficacy of PCho as a membrane localising PC mimic, the compound was incubated with *Pseudomonas aeruginosa PAO1* cells during exponential growth for 5 minutes. This growth period equates to approximately 0.25 of the generation time in these growth conditions (15). Cells were then fixed and labelled with a fluorescent Cy3-azide, as established in mammalian cells (11). The cells were also labelled for teichoic acid as previously described (16)(Figure 1a). Cells incubated with PCho that underwent click-labelling demonstrated fluorescence at the membrane after click-labelling for 5 minutes. This indicates that the lipid head group mimics are localised to the membrane specifically (Figure 1b).

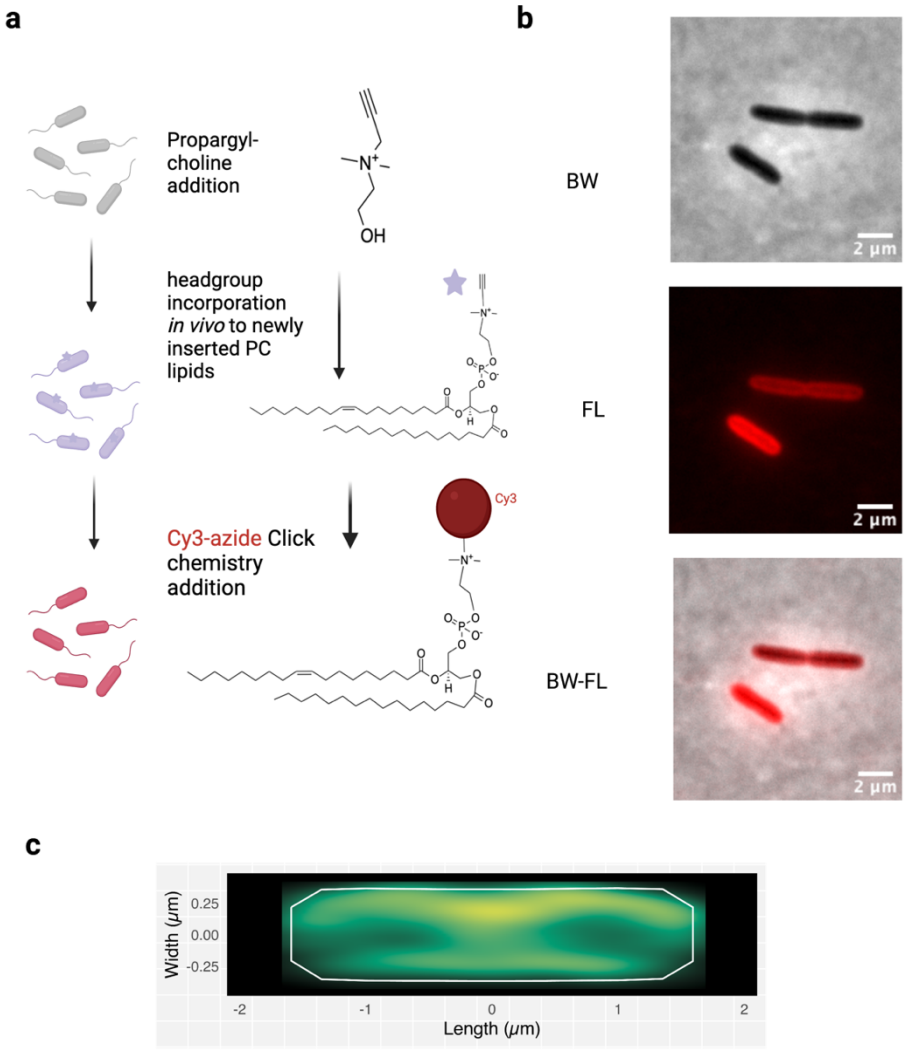


Figure 1. Phosphatidylcholine Lipid insertion can be visualised by Propargyl- Choline

a Propargyl-choline mimics phosphatidyl choline and can be labelled by click chemistry **b** Membrane localisation of propargyl-choline-cy3 fluorescence. (100µM) **c** Average localisation of cell fluorescence maxima 1119 cells, 0.1 OD exponential phase by BACTmap (17)

We then optimised PCho staining across a variety of concentrations (SI Fig1), with 100 µM being the optimum concentration for membrane visualisation. However averaging of fluorescence distribution using BACTmap software (17) (Figure 1c) revealed no specific increase in fluorescence at any region except for the membrane and membranous cell division site. Titration of PCho revealed no fluorescence above background in the absence of PCho and visible fluorescence in cells from 1 µm to 100 µM PCho (SI Figure 1b).

The optimum insertion time was 5 minutes at 100 µM PCho, which was sufficient to visualise membranes after washing. However, lower concentrations were also sufficient for visualisation

(SI Figure 1b). In order to determine whether PCho, and its storage buffer DMSO had a detrimental effect on growth, we measured the growth cycles of cells with higher PCho concentrations (SI Figure 1b). Only concentrations above 150µm PCho, corresponding to 2.5% DMSO, affected growth rate. This effect was marginal and was insignificant during growth phase. This indicated that at the PCho concentrations used for microscopy (1 µm – 100 µM) the growth defect was insignificant.

We then confirmed incorporation of PCho into the membrane by thin layer chromatography (TLC) of the lipid fraction prepared from whole *Pseudomonas aeruginosa PAO1* cells. The fluorescent headgroup of propargyl-choline-cy3 was only incorporated in cells that had both been labelled with PCho and Cy3 after click-chemistry labelling. This indicates that PCho was incorporated into the lipid fraction of cells (Figure 2). Phosphomolybdic acid staining for lipid species revealed an additional spot by TLC for cells without Cy3 addition. This could potentially represent an unlabelled PCho phosphatidylcholine mimic (Black line).

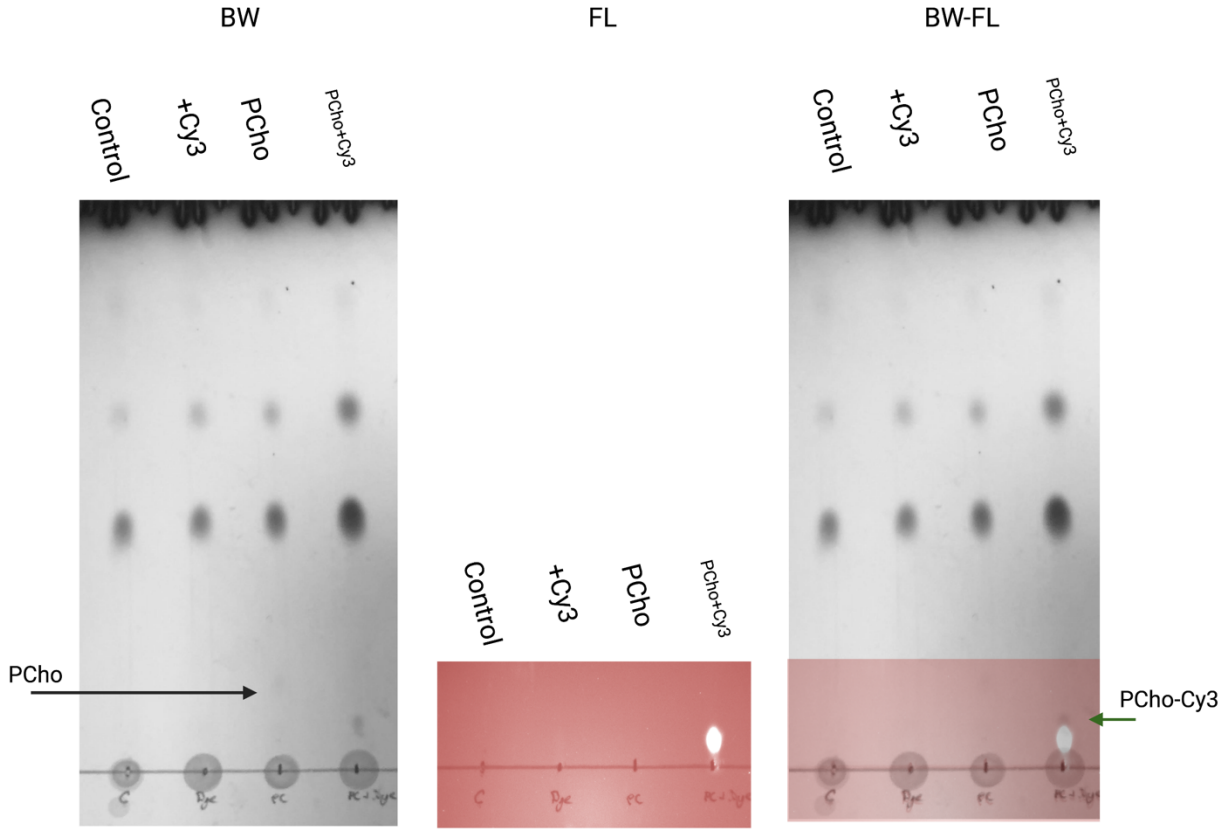


Figure 2 Insertion of propargyl-choline and Cy3 in *P. aeruginosa* membranes.

TLC plate of *Pseudomonas aeruginosa PAO1* lipid fractions. All cells were subject to conditions used for click chemistry described in methods. Control - Wildtype *P. aeruginosa* lipid extract, +PCho 100 µM propargyl-choline addition, +Cy3 with Cy3 addition, PCho and Cy3, both propargyl-choline and Cy3 added. BW- Visible light photo of

stained TLC plate, FL- Cy3 fluorescence indicating Cy3 lipid attachment to lipid fractions Black arrow - propargyl-choline labelled lipid peak, Green arrow– Cy3-Propargyl-choline-lipid peak.

In order to investigate fluorescence distribution over larger cell stretches we labelled elongated cells. Cell elongation was achieved by exposure to the β-lactam antibiotic aztreonam, which is highly selective for inhibition of cell division associated PBP3 (FtsI) (18). We hypothesised that if PC is incorporated into the membrane at distinct sites then this could be visible in elongated cells. We found that labelled cells have regions of higher intensity fluorescence, which could indicate sites of increased PC insertion compared with no insertion over this same time period (Figure 3). Whilst we found differences in fluorescent labelling, with banded regions of higher PCho concentration along the non-dividing cells, this banding was not observable for untreated cells.

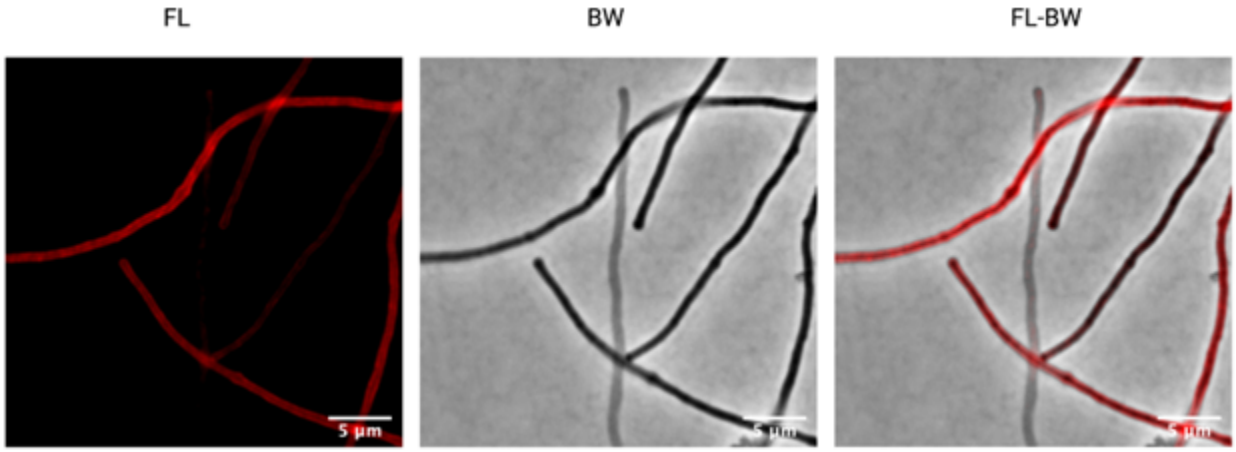


Figure 3. Aztreonam elongated cells show localised fluorescent incorporation

BW- Confocal, FL- fluorescent wavelength, BW-FL - merged Confocal Fluorescence images. All images were brightness and contrast adjusted manually for the highest-level detail observable in fluorescence channels. 5-minute incubation at 100 µM propargyl-choline

Propargyl-choline-Cy3 fluorescence has a similar localisation pattern to FM-464X

In order to evaluate Cy3-mediated fluorescence of labelled PCho as a general membrane stain, we compared it with a widely used membrane localisation method, the membrane dye FM-464X (19) (Figure 4). We found that the FM-464X membrane localisation was a visually clearer membrane label than PCho (Figure 4). However, fluorescence was heterogeneous among cells that were stained with either FM-464X or PCho. This suggests that differences in fluorescent intensity patterns between cells could be dependent on preparation/visual depth as indicated

by maps of fluorescent maxima, which have a peak at the cell centre. However, heterogeneity in PCho labelled cells also occurred at the cellular membrane level. Heterogenity of fluorescence in individual cells was suggestive of a varying distribution of label incorporation, suggesting a real fluorescence heterogeneity among lipids.

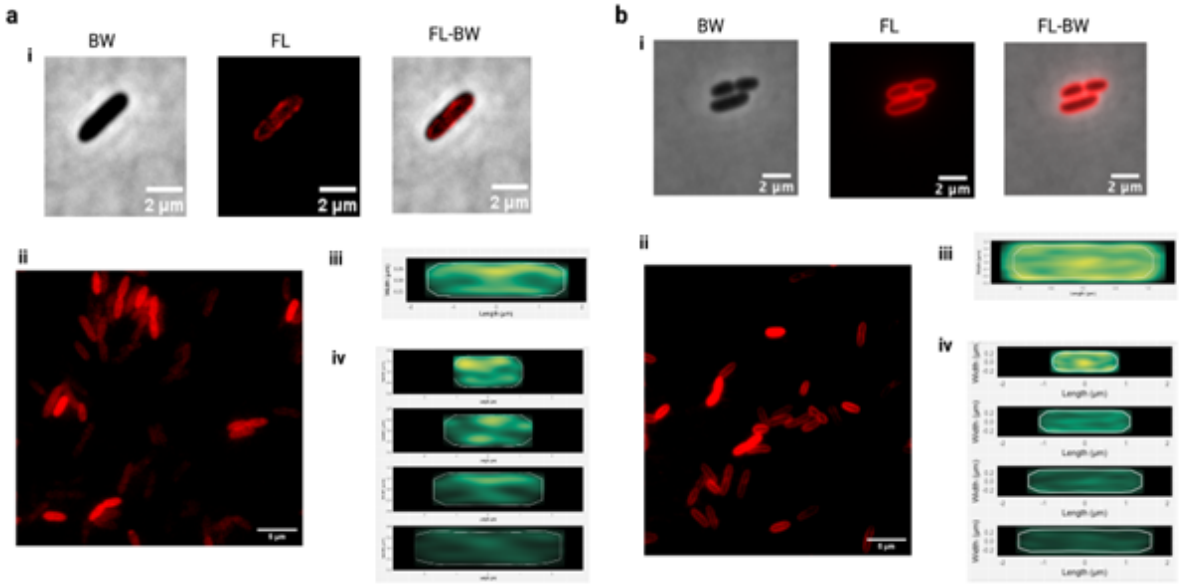


Figure 4. Membrane localisation and preference of Propargyl-choline labelled cells.

a Localisation of Cy3 fluorescence in propargyl-choline time-pulsed cells after 5 minutes incubation with 100µM propargyl-choline on individual cell i, groups of cells ii, and analysed as a population of 1119 cells across length classes iii/iv. **b** FM4-64X localisation on individual cells i, groups of cells ii and analysed as a population of 224 cells iii/iv.

The propargyl-choline click labelled lipids can track lipid movements.

We hypothesised that the PCho labelling methodology could be used to track lipid domains in live cells. The Cy3 click-labelled PCho photobleached over time, however the heterogenous patterns of fluorescence localisation changed over the course of miliseconds, independent of the bleaching effect. This suggested that the ~3% phospholipids that were potentially labelled could be changing location over the course of the imaging experiment. These groups of lipid particles could be tracked in ImageJ (20) using Trackmate (21), to give a population of potential lipid speeds over time. Lipid group track speeds revealed movement around a periplasmic track between 0.25-0.35 µm/s, which is similar to the speed of these lipids in mammalian cells (22) (Figure 5). Speed of group movement increased following exposure to 1 mM octanol, which is known to increase membrane fluidity (23) (Figure 5a). This indicates that tracking the lipid particle groups is indicative of lipid movement within the membrane. These results suggest that regions of increased fluorescence labelling could indicate lipid microdomains and lipid

movements. Therefore this method could have potential uses in TIRF as a means of studying lipid movement phenomena.

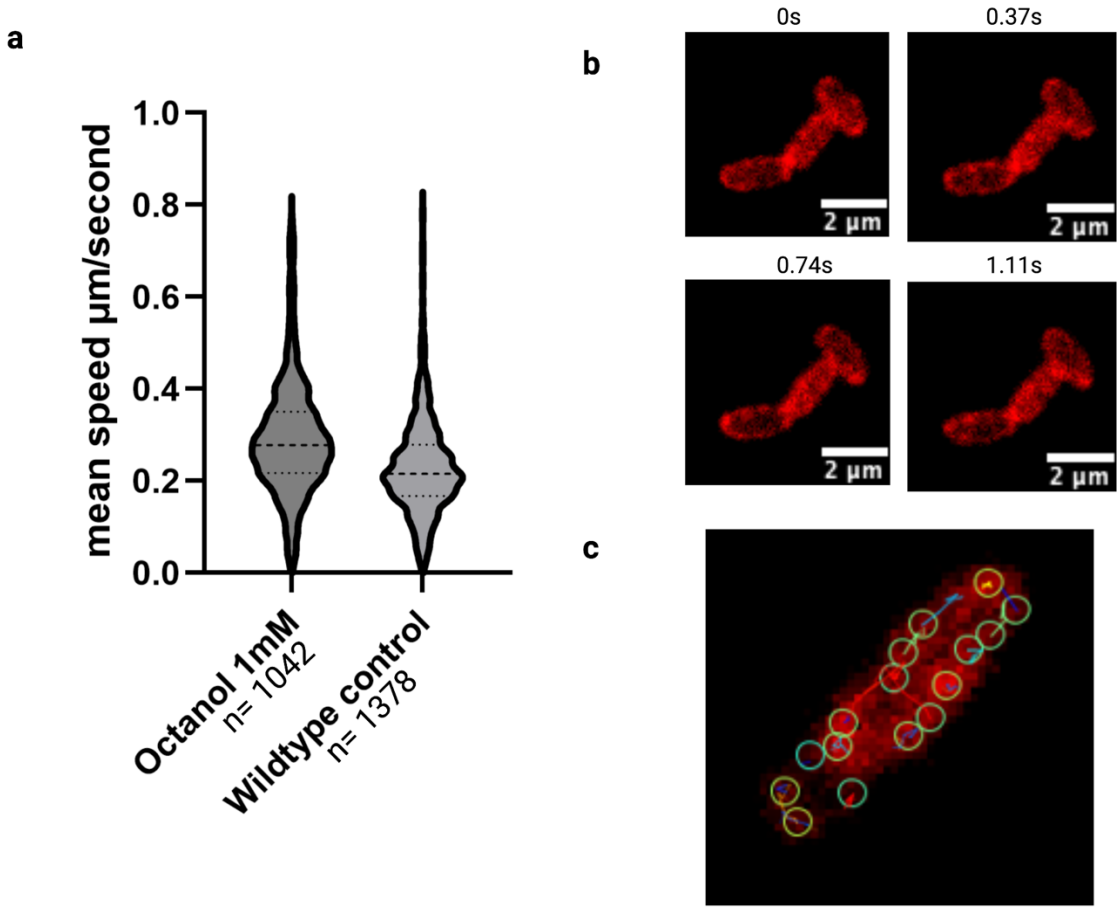


Figure 5. Use of Propargyl-choline as a potential lipid domain visualisation tool.

a Trackmate (21) mean track speed µm/s and fluorophore counts in a *Pseudomonas aeruginosa* PAO1 propargyl-choline and Cy3 labelled cells. T test difference in population, $p < 0.0001$ **b** 1.11s Timecourse of PCho-Cy3 fluorescence movements **c** attached GIF in supplementary data showing fluorescence change over time in periplasm. Red indicating fluorescence (SI Data https://drive.google.com/file/d/1t2L4Jn01psC8yq3H44j4XlmXO4NtEPLk/view?usp=drive_link), 100ms exposure Cy3 labelled PCho *Pseudomonas aeruginosa* cells **c** Trackmate fluorescence domains capture, and tracks shown

Conclusion

In this work, we have adapted an existing technique for phosphatidylcholine localisation in eukaryotic membranes for use in wild-type bacterial cells. Pulse-labelling *P. aeruginosa* cells

with propargyl-choline (PCho) allowed for visualisation of new phosphatidylcholine mimic insertion in the cell membrane of living cells. (Figure 1). We confirmed insertion of the label using extracted lipids from *P. aeruginosa* PAO1, which revealed a fluorescently labelled group visualised by TLC, only when both Cy3-azide and propargyl-choline were present during labelling.

This visualisation of a specific lipid, as opposed to other lipid labels which label lipids dependent on tail or have variant lipid affinities, allows for a specific and titratable bacterial membrane label. Analogous labels for use in cell envelope studies, such as peptidoglycan labelling with fluorescent d-amino acid mimics, have in the past been used in fluorescence microscopy allowing for significant advancements in the study of bacterial lifecycles(12). It should be noted that PCho incorporation could also be used for teichoic acid visualisation(16). Therefore, choice of a species without teichoic acids, such as *P. aeruginosa*, is necessary to allow for a direct lipid visualisation without potential crossover. Related species such as *Brucellus abortus* have been shown to use PC in their own membranes, after scavenging from the host eukaryotic cells, (9) therefore a marker for PC in bacteria would be useful for understanding these processes and the nuances of pathogenicity more clearly.

In this study, having tested this in *P. aeruginosa* PAO1*, we* propose this PCho incorporation and visualisation method as a phospholipid fluorescent labelling technique for compatible bacteria when other general labels are not suitable. We also propose the covalent label may have further use due to its titratable nature, in single molecule tracking techniques for study of lipid dynamics in bacteria.

Methods

Imaging of fluorescence in Pseudomonas aeruginosa

Strains were streaked from glycerol stocks onto LB Agar plates, and incubated at 37°C overnight. One PA colony was grown overnight at 37°C, at 180 rpm in 2 ml LB. The following morning a 1/10 dilution of the samples were then grown at 37°C, 180 rpm until the samples had all reached an OD_{600} of 0.3. 1% agarose in PBS was heated in a microwave until piping hot. Microscope slides were topped with a solution of 1% agarose in PBS which was flattened with a coverslip and left to cool. The coverslip was removed and 10 µl of sample was added to the slide, the coverslip was placed on top to spread the sample across the slide. Samples were then analysed using confocal microscopy specifically to identify the fluorescence, with corresponding filters dependent on expected fluorescence. Cy3-Propargyl-choline fluorescence was detected using a TXR filter set on a Leica DMi8 confocal microscope. The resulting images were taken in clusters of 15 across the sample at random to reduce bias, and allow for quantitive cell measurements.

Analysis of fluorescence localisation

Images were imported at LIFs or TIF libraries, which were then analysed by MicrobeJ software to determine determine cell contouring and maxima points. The points of increased

fluorescence to background, were then tracked across the cell and mapped per individual point across thousands of cells, dependent on cell length. The points were then mapped for each strain. Automatic, cellular counting and size determination by MicrobeJ (24) and BactMAP (17) allowed for quantitative analysis. Fluorescent points were tracked using custom tolerance and intensity filters maintained throughout study. Scripts and conditions for image analysis attached in the Supplementary Data file.

Propargyl-Choline click labelling

0.1 OD cells were incubated with alternative concentrations of propargyl-choline (dissolved in DMSO) (1µM to 2800µM) for 5 minutes. These cells were then concentrated by centrifugation for 10mins at 5000xg and (Figure 1/3/4) fixed by 4% paraformaldehyde PBS for 30 minutes, however this is not necessary for labelling (Figure 2/5). Cells were washed by centrifugation at 5000xg by pelleting, and resuspended in 100mM Tris-HCL pH 8.8, 1mM CuS04 50µl and ascorbic acid 0.1M . The cells were reacted with 100µM Cy3-azide, for a click chemistry reaction, then after 30mins at room temperature washed with TBS 1% Tween solution by centrifugation four times to remove the fluorescent azide remaining, before imaging.

Thin layer chromatography and lipid extraction

Lipid extraction used the Folch method (25) of lipid extraction, with 1:2:1 Chloroform:Methanol:Water at 55°C for 30mins, and vortexing, followed by extraction of the chloroform lipid phase after centrifugation. The thin layer chromatography was conducted on 60A Sepharose 254nm TLC plates, using 65:25:10 Chloroform, ethanol and acetic acid. The TLC plate fluorescence was recorded using Cy3 fluoremetric TXR filters on a 5x Zeiss Axio zoom microscope, and posed adjacently.

Tracking lipid foci

Trackmatev6.02 , implemented through a FIJI package was used to identify foci and track their movement over time (21). Cells observed in octanol 1mM were compared to H_20 and imaged at 100ms intervals using TXR filters. Trackmate parameters: Cell Threshold 60,000, foci diameter 0.1µm. linking distance 0.3µm, gap closing 0.3µm, Simple LAP tracking, LoG detector, subpixel localisation= True.

Author statements

Author contributions

CG performed click-labelling, culture, microscopy and analysis. CG and JB performed Thin Layer Chromatography of samples. The paper was written by CG, MB, JB and DR. The work was supervised by JB, MB, and DR.

Conflicts of interest

The authors declare no conflicts of interest

Funding information

Antibiotic Research UK ANTRUK_SRG_05-2021

BBSRC Award BB/M01116X/1

Acknowledgements

We would like to thank Prof Lori Burrows for the *Pseudomonas aeruginosa* PAO1 cells. We would also like to thank Antibiotic Research UK ANTRUK_SRG_05-2021 for funding this research, as well as the BBSRC Award BB/M01116X/1 for funding CGs studentship. We would like to thank Dr Sam Benedict from that laboratory of Dr Patrick Moynihan for their assistance in visualisation of the TLC plates.

References

1. Murray, Christopher J L et al., Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis, The Lancet, Volume 399, Issue 10325, 629 - 655

2. Horne JE, Brockwell DJ, Radford SE. Role of the lipid bilayer in outer membrane protein folding in Gram-negative bacteria. J Biol Chem. 2020 Jul 24;295(30):10340–67.

3. Vanounou S, Parola AH, Fishov I. Phosphatidylethanolamine and phosphatidylglycerol are segregated into different domains in bacterial membrane. A study with pyrene-labelled phospholipids. Mol Microbiol. 2003;49(4):1067–79.

4. Hou S, Johnson SE, Zhao M. A One-Step Staining Probe for Phosphatidylethanolamine. Chembiochem. 2015 Sep 7;16(13):1955-1960. doi: 10.1002/cbic.201500127. Epub 2015 Jul 23. PMID: 26224023..

5. Zhang J, Nie J, Sun H, Li J, Andersen JP, Shi Y. De novo labeling and trafficking of individual lipid species in live cells. Mol Metab. 2022 Jul 1;61:101511.

6. Nilsson I, Lee SY, Sawyer WS, Baxter Rath CM, Lapointe G, Six DA. Metabolic phospholipid labeling of intact bacteria enables a fluorescence assay that detects compromised outer membranes. J Lipid Res. 2020 Jun 1;61(6):870–83.

7. Pang Z, Raudonis R, Glick BR, Lin TJ, Cheng Z. Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. Biotechnol Adv. 2019 Jan-Feb;37(1):177-192. doi: 10.1016/j.biotechadv.2018.11.013. Epub 2018 Nov 27. PMID: 30500353.

8. *Wargo MJ, Gross MJ, Rajamani S, Allard JL, Lundblad LK, Allen GB, Vasil ML, Leclair LW, Hogan DA. Hemolytic phospholipase C inhibition protects lung function during Pseudomonas aeruginosa*

infection. Am J Respir Crit Care Med. 2011 Aug 1;184(3):345-54. doi: 10.1164/rccm.201103-0374OC. Epub 2011 May 11. PMID: 21562128; PMCID: PMC3175536.

9. Comerci DJ, Altabe S, de Mendoza D, Ugalde RA. Brucella abortus synthesizes phosphatidylcholine from choline provided by the host. J Bacteriol. 2006 Mar;188(5):1929-34. doi: 10.1128/JB.188.5.1929-1934.2006. PMID: 16484204; PMCID: PMC1426538.

10. Cao F, Xiong M, Li S, Cai H, Sun Y, Yang S, Liu X, Zhu R, Yu X, Wang X. Phosphatidylcholine absence affects the secretion of lipodepsipeptide phytoxins in Pseudomonas syringae pv. syringae van Hall CFCC 1336. Microbiol Res. 2018 Jan;206:113-120. doi: 10.1016/j.micres.2017.10.001. Epub 2017 Oct 16. PMID: 29146248.

11. Jao CY, Roth M, Welti R, Salic A. Metabolic labeling and direct imaging of choline phospholipids in vivo. Proc Natl Acad Sci. 2009 Sep 8;106(36):15332–7.

12. Silva AM, Otten C, Biboy J, Breukink E, Van Nieuwenhze M, Vollmer W, et al. The fluorescent D-Amino Acid NADA as a tool to study the conditional activity of transpeptidases in Escherichia coli. Front Microbiol. 2018;

13. Dietrich C, Yang B, Fujiwara T, Kusumi A, Jacobson K. Relationship of lipid rafts to transient confinement zones detected by single particle tracking. Biophys J. 2002 Jan;82(1 Pt 1):274-84. doi: 10.1016/S0006-3495(02)75393-9. PMID: 11751315; PMCID: PMC1302468.

14. Graham CLB, Newman H, Gillett FN, Smart K, Briggs N, Banzhaf M, et al. A Dynamic Network of Proteins Facilitate Cell Envelope Biogenesis in Gram-Negative Bacteria. Int J Mol Sci. 2021 Jan;22(23):12831.

15. Ma L, Conover M, Lu H, Parsek MR, Bayles K, Wozniak DJ. Assembly and Development of the Pseudomonas aeruginosa Biofilm Matrix. PLOS Pathog. 2009 Mar 27;5(3):e1000354.

16. Guilmi AMD, Bonnet J, Peißert S, Durmort C, Gallet B, Vernet T, et al. Specific and spatial labeling of choline-containing teichoic acids in Streptococcus pneumoniae by click chemistry. Chem Commun. 2017;53(76):10572–5.

17. van Raaphorst R, Kjos M, Veening JW. BactMAP: An R package for integrating, analyzing and visualizing bacterial microscopy data. Mol Microbiol. 2020;113(1):297–308.

18. Karlowsky JA, Kazmierczak KM, de Jonge BLM, Hackel MA, Sahm DF, Bradford PA. In Vitro Activity of Aztreonam-Avibactam against Enterobacteriaceae and Pseudomonas aeruginosa Isolated by Clinical Laboratories in 40 Countries from 2012 to 2015. Antimicrob Agents Chemother. 2017 Aug 24;61(9):e00472-17.

19. FMTM 4-64FX, fixable analog of FMTM 4-64 membrane stain [Internet]. [cited 2023 Jan 3]. Available from: https://www.thermofisher.com/order/catalog/product/F34653

20. *Vischer NOE, Verheul J, Postma M, van den Berg van Saparoea B, Galli E, Natale P, et al. Cell age dependent concentration of Escherichia coli divisome proteins analyzed with ImageJ and ObjectJ. Front Microbiol. 2015;*

21. *Ershov D, Phan MS, Pylvänäinen JW, Rigaud SU, Le Blanc L, Charles-Orszag A, et al. TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. Nat Methods. 2022 Jul;19(7):829–32.*

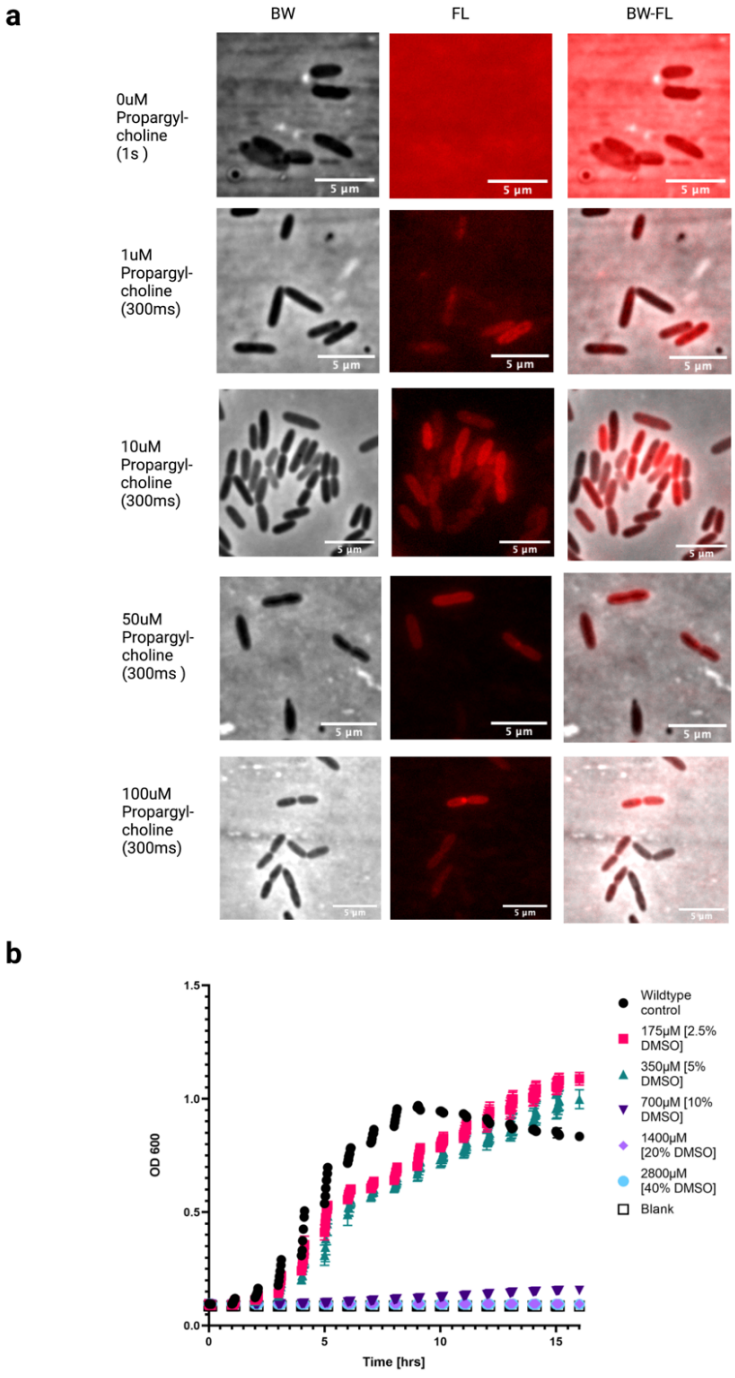
22. Chang JC, Rosenthal SJ. Visualization of Lipid Raft Membrane Compartmentalization in Living RN46A Neuronal Cells Using Single Quantum Dot Tracking. ACS Chem Neurosci. 2012 Oct 17;3(10):737–43.

23. Vaňousová K, Beranová J, Fišer R, Jemioła-Rzemińska M, Matyska Lišková P, Cybulski L, et al. Membrane fluidization by alcohols inhibits DesK-DesR signalling in Bacillus subtilis. Biochim Biophys Acta BBA - Biomembr. 2018 Mar 1;1860(3):718–27.

24. Ducret A, Quardokus EM, Brun YV. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol. 2016 Jun 20;1(7):16077.

25. Breil C, Abert Vian M, Zemb T, Kunz W, Chemat F. "Bligh and Dyer" and Folch Methods for Solid–Liquid–Liquid Extraction of Lipids from Microorganisms. Comprehension of Solvatation Mechanisms and towards Substitution with Alternative Solvents. Int J Mol Sci. 2017 Apr;18(4):70

Supplementary Data



SI Figure 1. Concentrations of propargyl-choline and affect on image quality and survival

a Propargyl-choline and effect on visualisation in cells. **a** Propargyl-choline addition and effect on survivability

Please note above is for referee use and will be moved in later manuscripts to a DOI.