

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/173595>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

An interdisciplinary approach to
developing tools to study antibiotic
permeability in Gram-negative
bacteria

Abby Mae Henney

B.Sc. (Hons)

A thesis submitted in partial fulfilment of the requirements for
the degree of

Doctor of Philosophy in Biological Sciences

University of Warwick, School of Life Sciences

May 2022

Table of Contents

Table of Contents	1
List of Tables	6
List of Figures	7
Acknowledgements	12
Declaration	13
Abstract	14
Abbreviations	15
Chapter 1. Introduction	17
1.1 Antibiotics	17
1.1.1 Antibiotic mechanism of action	18
1.1.2 Antimicrobial resistance	20
1.1.3 Mechanisms of AMR	20
1.2 Tackling AMR	22
1.3 Outer Membrane permeability	24
1.4 Permeability assays	27
1.4.1 Use of enzymatic reactions to measure permeability	28
1.4.2 Use of radiolabeled compounds to measure permeability	29
1.4.3 Use of fluorescence to monitor permeability	30
1.4.4 Use of mass spectrometry to measure permeability	30
1.4.5 Use of drug susceptibility to measure permeability	31
1.4.6 Use of chemoinformatics and simulations in permeability studies	32
1.4.7 Use of outer membrane vesicles to model permeability	33
1.5 Identified rules of Gram-negative permeability	34
1.6 Thesis aims	36
Chapter 2. Material and methods	37
2.1 General materials and methods	37
2.1.1 Reagents and materials	37
2.1.2 Bacterial growth reagents and strains	37
2.1.2 Bacterial growth media	37
2.1.2.1 Lysogeny broth	38
2.1.2.2 LB agar plates	38
2.1.2.3 Super Optimal broth with Catabolite repression	38
2.1.2.4 Low salt LB	38
2.1.2.5 Cation-adjusted Mueller-Hinton Broth	39
2.1.2.6 Cell Free Auto Induction Media (CFAI)	39
2.1.2.7 <i>E. coli</i> strains used in this work	39
2.1.3 Preparation of competent Cells	40

2.1.4 DNA quantification	41
2.1.5 DNA transformation of competent cells	41
2.1.6 Glycerol stocks	42
2.1.7 Sequencing	42
2.1.8 Minimal Inhibitory Concentration (MIC) microbroth dilution	42
2.2 Protein purification	43
2.2.1 Purifying <i>E. coli</i> PBP3	43
2.2.2 PBP4	43
2.2.3 Purifying deGFP	44
2.2.4 Protein Quantification	45
2.2.4.2 Bradford (BioRad) Protein Assay	45
2.2.5 Visualising proteins in SDS-PAGE gels	45
2.2.6 Bocillin gels	46
2.3 Chapter 3 specific techniques - iAmp	47
2.3.1 iAmp synthesis	47
2.3.2 HPLC	47
2.3.4 Mass Spectrometry	47
2.3.5 D-ala release coupled assay	47
2.3.6 Imaging iAmp	48
2.4 Chapter 4 specific techniques - mass spectrometry and β-lactam acylation	49
2.4.1 Proteomic preparation	49
2.4.2 Proteomic gel digestion	49
2.4.3 In-solution and FASP protocol	50
2.4.4 C18 Stage tip protocol	50
2.4.5 Proteomic mass spectrometry	50
2.4.6 Proteomic search: Maxquant and Scaffold	51
2.4.7 FT-ICR MS	51
2.4.8 MALDI-TOF	51
2.5 Chapter 5 specific techniques - synthetic cells	52
2.5.1 Creating cell-free extract	52
2.5.2 Cell-free protein synthesis reaction	52
2.5.2.1 DNA preparation	53
2.5.3 Creating synthetic cells	54
2.5.4 Preparing synthetic lipids	54
2.5.5 Feeding Solutions	54
2.5.6 Preparing Synthetic Cells in Eppendorf method	55
2.5.7 Preparing Synthetic cells and GUVs bulk production using microplate method	55
2.5.8 Imaging synthetic cells and GUVs in microplate	56
2.5.9 Calcein leakage assay	57
2.6 Chapter 6 specific techniques - machine learning	57
2.6.1 The initial permeability dataset (generated in-house)	57
2.6.2 Second permeability dataset - from open-source platform	58
2.6.3 Software packages used in Chapter 6	59

2.6.4 Molecular representations of compounds	59
2.6.5 The models used to predict permeability	60
Chapter 3. iAmp – a quantifiable β-lactam?	62
3.1 Background	62
3.1.1 β -lactam Permeability	62
3.1.2 Fluorescent antibiotics	63
3.1.3 iAmp	65
3.1.4 Research aims	66
3.2 Results and Discussion	66
3.2.1 Synthesis of iAmp	66
3.2.2 Purification of iAmp	67
3.2.3 Electrospray ionisation mass spectrometry of key HPLC Peaks	70
3.2.4 Real-time analysis of iAmp synthesis	77
3.3 - Investigating the activity of iAmp	80
3.3.1 Minimal Inhibitory Concentration of iAmp	80
3.3.2 iAmp activity investigated using bocillin binding gels	80
3.3.3 D-ala release in DacB carboxypeptidation with iAmp and ampicillin	81
3.3.4 Imaging iAmp in <i>E. coli</i>	82
3.4 Structural Analysis of iAmp	84
3.4.1.1 ESI MS of iAmp	84
3.4.1.2 NMR Structural Analysis	85
3.4.2 Solved structure of Compound 8	86
3.4.3 Proposed Mechanism of Compound 8 synthesis	86
3.5.1 Modelling the binding of Compound 8 and ampicillin in porins	90
3.6 Future work	90
3.7 Conclusions	92
Chapter 4. Monitoring β-Lactam acylation using mass spectrometry techniques	93
4.1 Background	93
4.1.1 Penicillin-Binding Proteins	93
4.1.2 Mass Spectrometry	94
4.1.3 Ionisation	95
4.1.5 FT-ICR MS	96
4.1.6 Orbitrap Mass Analyser	96
4.1.7 MS permeability studies	96
4.1.8 Proteomics	97
4.1.9 Research aims	97
4.2 Results	98
4.2.1 Protein purifications	98
4.2.2 Proteomic Results	98
4.2.3 <i>E. coli</i> PBP4 acylated with ampicillin	99
4.2.4 <i>E. coli</i> PBP1a acylated with Ampicillin	100
4.2.5 Method Development	100
4.2.6 PBP3s	102

4.2.7 Nitrocefin and PBP3s	102
4.2.8 <i>E. coli</i> PBP3 acylated with nitrocefin	102
4.2.9 <i>Burkholderia pseudomallei</i> PBP3 acylated with nitrocefin	104
4.2.10 <i>Pseudomonas aeruginosa</i> PBP3 acylated with a range of β -lactams	105
4.2.11 <i>Pseudomonas aeruginosa</i> PBP3 acylated with Nitrocefin	106
4.2.12 <i>P. aeruginosa</i> PBP3 acylated with aztreonam	108
4.2.13 Ceftazidime PBP3 adducts <i>P. aeruginosa</i>	109
4.3 Acylation of <i>E. coli</i> PBPs by ampicillin	109
4.3.1 <i>E. coli</i> PBPs mixture treated with ampicillin	109
4.3.2 <i>E. coli</i> PBP1b with ampicillin	110
4.3.3 <i>E. coli</i> PBP1a treated with ampicillin	111
4.3.4 <i>E. coli</i> PBP4 and ampicillin	111
4.3.5 <i>E. coli</i> PBP3 ampicillin	112
4.3.2 Discussion of bottom-up approach	112
4.4 Top-Down Proteomics MS	113
4.4.1 CID of PBP3-ceftriaxone acyl enzyme	116
4.4.2 CID ECD of PBP3-ceftriaxone acyl enzyme	117
4.5 Using MALDI-TOF MS to monitor the acylation of PBPs	118
4.5.1 MALDI-TOF of other PBP4	119
4.6 Discussion and Future prospects	120
4.7 Conclusion	121
Chapter 5. Creating synthetic cells to observe antibiotic permeability	122
5.1 Background	122
5.1.1 Liposomes	122
5.1.2 Synthetic cells	123
5.1.3 Outline of work	125
5.1.4 Research aims	126
5.2 Results	127
5.2.1 Creating cell free protein synthesis system	127
5.2.2 Optimising CFPS system	127
5.2.3 Purifying deGFP	128
5.2.4 Inhibiting CFPS with antibiotics	130
5.3 Creating Synthetic Cells	131
5.3.1 Creating Synthetic Cells using Eppendorf technique	131
5.3.2 Creating GUVs using the microplate technique	132
5.3.3 Creating synthetic cells using the microplate technique	133
5.3.4 Monitoring the synthetic cells creating deGFP over time	135
5.3.5 Using antibiotics to inhibit synthetic cells producing deGFP	135
5.3.6 Optimisation of microplate method	137
5.3.7 Creating GUVs with asymmetric bilayers	140
5.3.8 GUV calcein leakage	141
5.4 Discussion and Future Prospects	142
5.5 Conclusion	145

Chapter 6. Machine learning to predict antibiotic permeability	146
6.1 Background	146
6.1.1 Machine learning in medicinal chemistry	146
6.1.2 Key concepts in machine learning	147
6.1.3 Algorithms	149
6.1.4 Deep learning	152
6.1.4.1 Graph Convolutional Networks (GCN)	152
6.1.5 Data engineering	152
6.1.5.1 Feature engineering	153
6.1.6 Evaluating models	156
6.1.6.1 Metrics	157
6.1.6.2 Bias – Variance Trade off	158
6.1.6.3 Overfitting	159
6.1.7 Research aims	160
6.2 Results and discussions	161
6.2.1 Using MICs as a permeability indicator	161
6.2.2 Feature engineering and extraction of compounds in permeability dataset	164
6.2.3 Results of permeability dataset	165
6.3 Creating a new dataset from SPARK database	167
6.3.1 Data preprocessing of new SPARK derived database	167
6.3.2 Featurising the SPARK derived dataset using Mol2Vec	168
6.3.3 Results of new dataset	169
6.4 Using learnt molecular representations	169
6.4.1 Featurising the new SPARK derived dataset	169
6.4.2 Results of GCN model predicting permeability from SPARK derived dataset	169
6.5 Comparing graph convolutional model to a random forest model	171
6.5.1 Creating random forest model to predict permeability barriers	171
6.5.2 Random forest prediction results on SPARK derived dataset	172
6.5.3 Approaching permeability predictions from a binary perspective	174
6.6 Predictions and atomic contributions	176
6.7 Discussion and Future Prospects	180
6.8 Conclusions	183
7. Conclusions and future work	184
8. Bibliography	188
9. Appendix	206
9.1 Compound 8 NMR	206
9.2 Compound 8 MS	207
9.3 SDS-PAGE gels used in in-gel digest proteomic preparations	208
9.4 Sequenced <i>E. coli</i> PBP3	209
9.5 MIC results	210
9.6 Review: Quantifying Antibiotic Permeability	214

List of Tables

Table 2.1 <i>The <u>E. coli</u> strains used in this thesis, along with their genotype and source.</i>	40
Table 2.2 <i>Components of 12% SDS-PAGE gels.</i>	46
Table 2.3 <i>Components in the D-ala release assay.</i>	48
Table 2.4 <i>The components of the CFPS reaction.</i>	53
Table 2.5 <i>The fluorophores used in this work, and their excitation and emission values.</i>	56
Table 3.1 <i>Predicted monoisotopic masses of ampicillin and related structures.</i>	71
Table 3.2 <i>Permeability properties of KiAmp, iAmp and ampicillin.</i>	88
Table 4.1 <i>Predicted peptide mass of active site serine of <u>E. coli</u> PBP4, with and without ampicillin treatment.</i>	99
Table 4.2 <i>Predicted masses of <u>E. coli</u> PBP3 active-site serines with nitrocefin adducts</i>	103
Table 4.3 <i>The predicted/calculated fragments of <u>B. pseudomallei</u> PBP3 active site peptides acylated by nitrocefin.</i>	104
Table 4.4 <i>The predicted m/z ratios of <u>P. aeruginosa</u> PBP3 peptide fragments with the antibiotics; nitrocefin, aztreonam, and ceftazidime</i>	106
Table 4.5 <i>Predicted masses of <u>E. coli</u> PBP3 treated with ceftriaxone.</i>	114
Table 6.1 <i>A Glossary of key terms.</i>	148
Table 6.2 <i>Metric scores from the MLP MIC model predicting permeability.</i>	166
Table 6.3 <i>Metric results from the multiclass GCN model predicting permeability barriers.</i>	170
Table 6.4 <i>Metrics calculated for each output of the random forest model generating predictions for three permeability outputs.</i>	173
Table 6.5 <i>Metric scores of binary GCN model of predicting permeability issues of the SPARK derived dataset.</i>	175
Table 6.6 <i>The metric results of the binary random forest model predicting permeability barriers of compounds.</i>	176
Table 6.7 <i>The results of the predictions of binary GCN and binary RF model on nine unseen compounds.</i>	177

List of Figures

Figure 1.1 <i>The WHO priority pathogens.</i>	18
Figure 1.2 <i>The targets of antimicrobial agents.</i>	19
Figure 1.3 <i>Differences in Gram-negative (left) and Gram-positive bacteria (right) cell walls</i>	20
Figure 1.4 <i>How resistance emerges in bacteria.</i>	22
Figure 1.5 <i>A diagram of the cell envelope of Gram-Negative bacteria</i>	24
Figure 1.6 <i>Key components of the outer membrane, LPS (A) and porins (B).</i>	25
Figure 1.7 <i>Antimicrobials have different entry pathways through the Gram-negative outer membrane.</i>	27
Figure 1.8 <i>Micro-iodometric assay used as an enzyme readout for permeability.</i>	29
Figure 1.9 <i>Antimicrobial susceptibility as a tool to measure permeability in two ways.</i>	32
Figure 1.10 <i>The formation of OMV and their use in permeability assays.</i>	34
Figure 3.1 <i>A depiction of <u>E. coli</u> cells where porins are present in the outer membrane</i>	63
Figure 3.2 <i>Structure of bocillin.</i>	64
Figure 3.3 <i>Structure of ampicillin and iAmp as described by Kotagiri et al.¹²⁸</i>	65
Figure 3.4 <i>Eppendorfs containing ampicillin on the left, and iAmp on the right, visualised in gel imager</i>	66
Figure 3.5 <i>Fluorescence spectrum of the compound purified by HPLC</i>	67
Figure 3.6 <i>HPLC elution profile of ampicillin and iAmp from a reverse phase C18 column monitored using fluorescence detection Ex. 350nm Em. 450 nm.</i>	68
Figure 3.7 <i>HPLC trace of iAmp and ampicillin purified by reverse phase C18 monitored at 218 nm</i>	69
Figure 3.8 <i>HPLC trace of iAmp and ampicillin purified by reverse phase C18 monitored at 254 nm.</i>	69
Figure 3.9A-B <i>ESI TOF MS⁺ of predominant ampicillin peaks from predominant two peaks of ampicillin as detected in HPLC separation.</i>	74
Figure 3.10A-B <i>ESI TOF MS of iAmp fractions from peaks collected at the first (A) and second (B) peak of HPLC purification, observed at 254 nm.</i>	75

Figure 3.11 The positive mode ESI MS spectra of major fluorescent peak from HPLC separation of iAmp.	76
Figure 3.12 The synthesis of iAmp was followed using RT-qPCR equipment monitored at Ex./Em. 350/450 nm.	77
Figure 3.13A-C Time point investigation into the production of iAmp.	79
Figure 3.14A-B SDS-PAGE gels used to assess iAmp binding to PBP1b	81
Figure 3.15 D-ala release by PBP4 D-D carboxypeptidase activity in the presence and absence of Ampicillin and iAmp	82
Figure 3.16A-D Staining <i>E. coli</i> with iAmp, and bocillin.	83
Figure 3.17 This is the observed mass spectrum for iAmp	84
Figure 3.18 Ampicillin sample ESI+ TOF MS	85
Figure 3.19 The chemical structure of Compound 8	86
Figure 3.20 Proposed mechanism for the formation of Compound 8 (8) from ampicillin (1).	87
Figure 3.21 Comparing the structures of; KiAmp, Compound 8, ampicillin.	89
Figure 4.1 Penicillin covalently acylating PBP inhibiting the polymerisation of lipid II.	94
Figure 4.2 The workflow of mass spectrometry	95
Figure 4.3 SDS-PAGE of <i>E. coli</i> PBPs used in this chapter.	98
Figure 4.4 <i>E. coli</i> PBP4 treated with ampicillin.	100
Figure 4.5 Effect of heat degradation on PBP-antibiotic adduct stability.	101
Figure 4.6 The active-site serine of <i>E. coli</i> PBP3 treated with nitrocefin.	103
Figure 4.7 <i>B. pseudomallei</i> PBP3 treated with nitrocefin.	105
Figure 4.8 MS spectrum of <i>P. aeruginosa</i> PBP3 acylated by nitrocefin.	107
Figure 4.9 MS spectrum of <i>P. aeruginosa</i> PBP3 treated with nitrocefin, acylation observed on an unexpected serine.	107
Figure 4.10 <i>P. aeruginosa</i> PBP3 in complex with ceftriaxone, with SxxK and SxN sites highlighted.	108
Figure 4.11 <i>P. aeruginosa</i> PBP3 treated with aztreonam.	109

Figure 4.12 <i>P. aeruginosa</i> PBP3 treated with ceftazidime.	109
Figure 4.13 <i>E. coli</i> PBP1b treated with ampicillin.	110
Figure 4.14 <i>E. coli</i> PBP1b treated with ampicillin, a serine with an ampicillin adduct is located in the second conserved serine of PBPs at SxN.	110
Figure 4.15 <i>E. coli</i> PBP1b treated with ampicillin, an ampicillin adduct is located on the SxxK active-site serine.	111
Figure 4.16 Mechanism of ceftriaxone acylating the active site serine.	114
Figure 4.17 FT-ICR MS Spectrum of <i>E. coli</i> PBP3, top is the control (no antibiotic).	115
Figure 4.18 The FT-ICR MS CID spectrum of <i>E. coli</i> PBP3.	117
Figure 4.19 FT-ICR MS spectra of ECD CID PBP3-ceftriaxone sample.	118
Figure 4.20 MALDI spectrum of <i>E. coli</i> PBP3 treated with ceftriaxone.	119
Figure 5.1 Size range of liposomes	122
Figure 5.2 Components of CFPS.	124
Figure 5.3 The inverted emulsion transfer technique used to form GUVs	126
Figure 5.4 Troubleshooting CFPS.	128
Figure 5.5A-B Purified deGFP and the accompanying concentration curve created	129
Figure 5.6 The CFPS deGFP production over time.	129
Figure 5.7 Visualising deGFP produced from CFPS reactions.	130
Figure 5.8. The deGFP produced by CFPS reaction in the presence of different antibiotics.	131
Figure 5.9A-C A synthetic cell synthesised using Eppendorf technique	132
Figure 5.10A-B GUVs created by the microplate method that encapsulates deGFP.	133
Figure 5.11 Synthetic cells formed using 96-well microplate method	134
Figure 5.12A-C Synthetic cells producing deGFP over two hours.	135
Figure 5.13A-B Gentamicin inhibiting deGFP produced by synthetic cell	136

Figure 5.14A-D Tile scans from the optimised microplate method	138
Figure 5.15A-C The vesicles formed from the optimised protocol at 10x magnification.	139
Figure 5.16A-B GUVs that have fluorescent lipids integrated into the asymmetric bilayer	140
Figure 5.17A-C A closer look at GUVs that have fluorescent lipids integrated into the asymmetric bilayer.	141
Figure 5.18 Calcein leakage assay performed on GUVs created using the 96-well microplate method.	142
Figure 6.1 The boom of machine learning in relation to antibiotics.	147
Figure 6.2 Contextualising the glossary terms used in ML.	149
Figure 6.3A-B An overview of two of the machine learning algorithms used in this chapter.	151
Figure 6.4 β -lactam ring featurised using ECFP.	154
Figure 6.5 Mol2Vec Featurisation of the β -lactam ring.	155
Figure 6.6 Graph featurisation of a benzene ring.	156
Figure 6.7 Overview of Confusion matrices and ROC-AUC curves.	158
Figure 6.8 Building a generalisable mode	158
Figure 6.9A-C Model Underfitting vs overfitting.	159
Figure 6.10A-B: Validation curves and learning curves.	160
Figure 6.11 - An overview of the general workflow of this work.	161
Figure 6.12 MIC results of nine bacteria against a broad spectrum of bacteria.	162
Figure 6.13 Distribution of classes in the MIC dataset.	165
Figure 6.14A-B Training curves of the permeability prediction made by the MLP model of permeability.	166
Figure 6.15 Feature engineering of the SPARK derived dataset.	167
Figure 6.16 Class distribution in the new SPARK derived dataset.	168
Figure 6.17A-D Results of the multiclass GCN model predicting permeability	170

Figure 6.18 Confusion matrices and ROC curve scores for each class output of RF model.	173
Figure 6.19 Confusion matrix of results from binary GCN model predicting permeability issues on the SPARK derived dataset.	175
Figure 6.20 Confusion matrix of the results from the binary random forest model used to predict permeability barriers on the SPARK derived dataset.	176
Figure 6.21 Nine unseen compounds used to test the efficacy of binary models.	178
Figure 6.22A-D Two compounds used to probe the accuracy of permeability predictions.	179

Acknowledgements

I would like to thank my supervisors, Prof. Chris Dowson and Dr Adrian Lloyd, for their support and supervision throughout this project. I want to thank Chris for the opportunity of this project and for allowing me to approach it creatively, and for encouraging my growth as a researcher. I would like to thank Adrian for his guidance during this process, especially with the writing. Adrian's kindness and patience for others is remarkable; I'm truly grateful to him for pushing me past my self-doubt and encouraging me to be proud of my work.

Thank you to lab friends past and present for their encouragement. I'm lucky to have shared this experience with many wonderful people and great scientists who offered enthusiastic support and training. In particular, I would like to thank Jessie, Megan and Hector for their friendship and motivation through the highs and lows of lab work and writing.

A big thank you to all my collaborators and other experts who made this work possible. In particular, Cleidi and Andrew (Warwick Proteomics Facility), Ian (Warwick imaging suite), John (microbiology support), Yuko (FTICR-MS) and Sahan (synthetic biology). I would also like to thank my advisory board members, Prof. Roper and Dr Crow, for their guidance throughout this project.

Finally, I would like to thank my friends and family for their encouragement throughout, especially for their reassurance in the difficult parts. It would not have been possible without your support. I'm grateful to everyone who believed I could finish this, especially when I didn't. To Andrea, thank you for your support, kindness and encouragement.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I hereby declare that I have personally carried out the work submitted in this thesis under the supervision of Professor Christopher Dowson and Dr Adrian Lloyd at the School of Life Sciences, University of Warwick. Where other individuals have contributed to work, it has been specifically stated in the text. I have composed this work, and no part of this work has previously been submitted to be considered for a degree or qualification. All sources of information have been specifically acknowledged in the form of references.

Abstract

Rising antibiotic resistance is an imminent threat to modern healthcare. Most new antibiotics have a narrower spectrum of activity and work only in Gram-positive bacteria. The World Health Organisation published a list of high priority bacteria with startling emerging resistance, including Gram-negative bacteria such as *Pseudomonas*, *Klebsiella*, and *Escherichia*. Gram-negative bacteria are inherently more resistant to antimicrobials, due to their outer membrane, which reduces the permeability of antimicrobials. This innate resistance presents challenges in developing novel antibiotics. A high-throughput universal Gram-negative permeability assay would streamline the drug development process, enabling new broad-spectrum antibiotics to reinvigorate the drug pipeline.

This thesis will explore potential avenues for monitoring and predicting permeability. The following work is an interdisciplinary investigation into Gram-negative permeability and toolkit development. This project has utilised tools from biochemistry such as enzyme assays and mass spectrometry to synthetic biology and computer science. This work approaches permeability in four ways. Firstly a potential fluorescent derivative of ampicillin was synthesised and analysed. Secondly, various proteomic techniques were utilised to monitor the covalent binding of β -lactams to penicillin-binding proteins, as a proxy of β -lactam permeability. Thirdly, bottom-up synthetic cells were created to model permeability in bacteria. Finally, machine learning algorithms consisting of supervised and unsupervised techniques were used to predict permeability. In this work alternative methods to quantifying β -lactam permeability were trialled in chapters 3 and 4. Whereas in chapters 5 and 6, the foundations to a new permeability assay and permeability predictions were laid. Ultimately, the creation of tools to predict and quantify permeability will come from many fields. Without a better understanding of permeability, antimicrobial drug development will stagnate. Therefore, to deter the impending antimicrobial resistance crisis, drug development needs to be faster, more intelligent, and better thought out; this could be achieved with a permeability assay used in conjunction with activity screens.

Abbreviations

AI	Artificial intelligence
AMP	Antimicrobial peptides
AMR	Antimicrobial resistance
ATP	Adenosine 5'triphosphate
BCA	Bicinchoninic acid
CAA	Chloroacetic acid
cAMP	Cyclic adenosine monophosphate
CFAI	Cell-free auto-induction
CFPS	Cell-free protein synthesis
CL	Cardiolipin
CTP	Cytidine-5'-triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionisation
FASP	Filter-aided sample preparation
FTICR	Fourier Transform Ion Cyclotron Resonance
GCN	Graph Convolutional Networks
GFP	Green fluorescent protein
GTP	Guanosine-5'-triphosphate
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HPLC	High performance liquid chromatography
IMAC	Immobilised metal affinity chromatography
kDa	Kilo Dalton
KDO	3-deoxy- α -D-manno-octulosonic acid
LB	Lysogeny broth
LC-MS	Liquid-chromatography MS
LPS	Lipopolysaccharide
MALDI	Matrix assisted laser desorption ionisation
MBC	Minimal bactericidal concentrations
MIC	Minimal inhibitory concentrations
MLP	Multi layer perceptron

MS	Mass spectrometry
MWCO	Molecular weight cut-off
NAD	Nicotinamide adenine dinucleotide
OMV	Outer membrane vesicles
PBP	Penicillin-binding proteins
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine (PE)
PG	Phosphatidylglycerol
POPC	Phosphatidylcholine
RF	Random forest
RFU	Relative fluorescence units
RNA	Ribonucleic acid
ROC-AUC	Receiver operator curve – area under curve
SAR	Structure-activity relationship
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SOC	Super Optimal broth with Catabolite repression
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetramethylethanediamine
TFA	Trifluoroacetic acid
TOF	Time-of-flight
UTP	Uridine-5'-triphosphate
v/w	Volume to weight ratio
w/w	Weight to weight ratio
WHO	World Health Organisation

Chapter 1. Introduction

1.1 Antibiotics

Bacteria are the causative agents of many communicable diseases such as pneumonia and endocarditis, these diseases when left untreated are often deadly. Antibiotics are antimicrobial compounds capable of killing or inhibiting the growth of bacteria. Prior to the discovery and commercialisation of antibiotics, infectious diseases caused early mortality. The discovery of antibiotics, coupled with vaccination programs, has dramatically increased average life expectancies from 56 years in the 1920s (US) to 80 years (US)¹. Antibiotics have enabled us to push the boundaries of modern medicine, allowing safe labour, transplant surgeries, and chemotherapy^{1,2}.

Flemming serendipitously discovered the first recognised and clinically used antibiotic penicillin in 1928³; Florey and Chain purified penicillin in 1940⁴, making it commercially available. Sulfonamides and sulfones were also developed in the 1930s and 40s⁵. Unlike penicillin, which was discovered as a natural product of mould^{3,4}, sulfonamides and sulfones were made synthetically. Since then, many other classes of antibiotics have been discovered namely; aminoglycosides (1943)⁶, chloramphenicol (1947)⁷, polymyxins (1949)⁸ tetracyclines (1948)⁹, macrolides (1952)¹⁰, glycopeptides (1953)¹¹, rifamycins (1957)¹², quinolones (1962)¹³, lipopeptides (1980)¹⁴ and oxazolidinones (1986)¹⁵. Shortly after each class was brought to market, bacteria evolved resistance to survive in the presence of antibiotics, limiting their efficacy. As resistance to antibiotics is constantly spreading and increasing, the risk of a post-antibiotic era grows. The challenge of rising levels of antimicrobial resistance has been exacerbated by the limited success of antibiotic development in the last twenty years.

The majority of antibiotics approved in the last five years (2017-2021) belong to established antibiotic classes like fluoroquinolones and β -lactams, with only one compound belonging to different chemical classes (Lefamulin)¹⁶. Prior to this, teixobactin was developed, which polymerises lipid II, preventing the production and maintenance of peptidoglycan. However, this drug is only active in Gram-positive organisms. The antimicrobial drug development pipeline needs novel classes of antibiotics, particularly

those that target the World Health Organisation (WHO) priority pathogens (**Figure 1.1**). These are pathogens that the WHO has recognised and defined as causing the most significant threat globally and are split into three levels: critical, high, and medium (**Figure 1.1**)¹⁶.

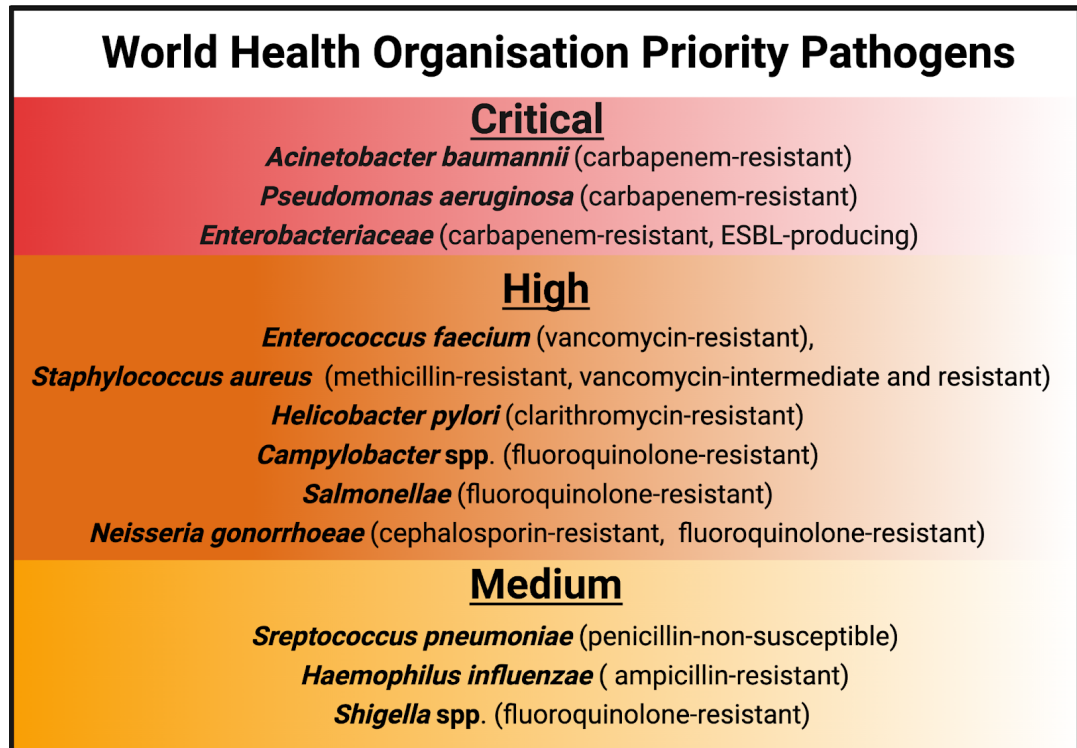


Figure 1.1 The WHO priority pathogens. WHO priority pathogens are split into three levels; critical, high and medium. In each category is the organism and the AMR of concern.

1.1.1 Antibiotic mechanism of action

Antibiotics fight bacterial infections by inhibiting bacterial-specific pathways, leading to either stasis (bacteriostatic) or death (bactericidal). Different antibiotic classes target specific areas or pathways in bacterial cells, such as; the cell wall, protein synthesis, nucleic acid synthesis, or inhibiting metabolic pathways (**Figure 1.2**)¹⁷.

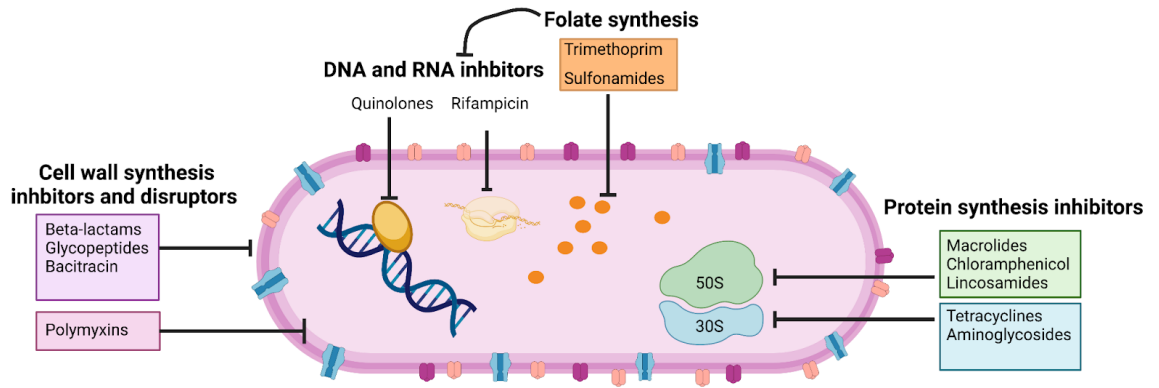


Figure 1.2 The targets of antimicrobial agents. Antibiotics target a variety of essential pathways in bacteria. This figure provides an overview of the general targets of different classes of antimicrobials.

The cell wall is a significant target for many antibiotics, including β -lactams and glycopeptides. Peptidoglycan is a complex sugar peptide polymer that forms a mesh around the cell; it is a crucial cell wall component¹⁸. The production of peptidoglycan is complex and requires many different enzymes, spanning multiple cellular compartments¹⁸. β -lactams and glycopeptides specifically target and inhibit the activity of the final enzymes involved in peptidoglycan synthesis¹⁸. These enzymes are named penicillin-binding proteins and polymerise lipid II and the cross-linking of the glycan strands¹⁸. Peptidoglycan is an ideal antibiotic target because it is essential in conferring structural shape and rigidity in bacteria, and without it, bacteria are susceptible to osmotic pressure, subsequent cell lysis and death^{17,19}.

Bacteria are divided into two categories; Gram-positive and Gram-negative; based on the response of the bacteria to Gram staining^{20,21} (**Figure 1.3**). Both categories have different cell wall compositions, but both have peptidoglycan (**Figure 1.3**)²².

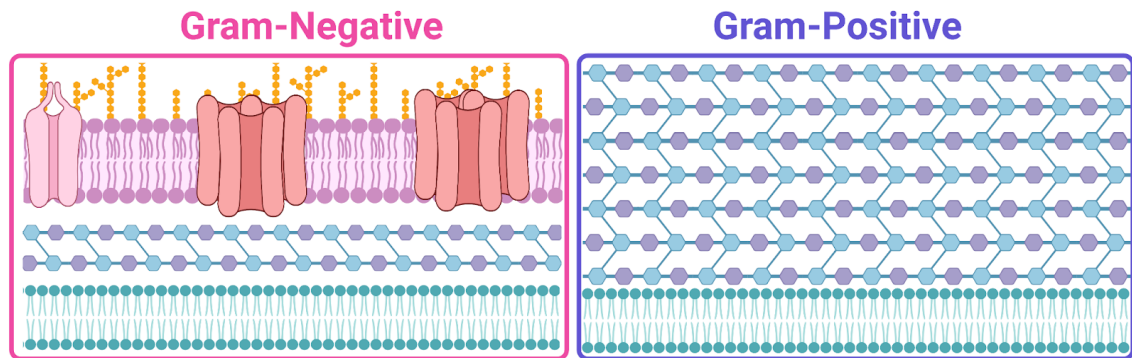


Figure 1.3 Differences in Gram-negative (left) and Gram-positive bacteria (right) cell walls. All bacteria have a phospholipid bilayer (green, bottom of both diagrams) coated in peptidoglycan (blue and purple cross-linked structure). The thickness of this peptidoglycan, however, varies between the two groups. Gram-negative bacteria (left) have a much thinner peptidoglycan layer than Gram-positive bacteria, which are coated in a thick peptidoglycan layer (right). Gram-negative cells also have an outer membrane on top of the peptidoglycan layer, which is absent in Gram-positive cells

Multiple antibiotics target the protein synthesis pathway. Examples of drugs targeting this pathway include aminoglycosides and tetracyclines, which target the 30S ribosome^{23,24}. Other antibiotics such as macrolides and chloramphenicol target the 50S subunit²⁴. Multiple antibiotics target the DNA synthesis pathways; some antibiotics directly interfere with essential DNA transcription by inhibiting topoisomerases, such as quinolone antibiotics²⁵. Drugs like trimethoprim interfere with folic acid metabolism, which is essential for normal nucleic acid synthesis²⁶.

1.1.2 Antimicrobial resistance

Antibiotic treatment subjects bacteria to selective pressure; in response, bacteria develop resistance mechanisms to survive. Bacteria have become resistant to most antibiotics, and super-resistant strains are emerging²⁷. Antimicrobial resistance (AMR) claims the lives of up to 1.27 million people globally, with resistant bacterial infections being untreatable²⁸. Predictions show that by 2050 up to 10 million people a year could die from rising levels of AMR, and the global financial impact of AMR could amount to 100 trillion dollars²⁷.

1.1.3 Mechanisms of AMR

There are two genetic mechanisms by which bacteria can acquire resistance, first by mutational resistance and secondly by horizontal gene transfer^{1,29}. Antibiotics act as a

selective pressure on bacteria; as bacteria multiply and genetic diversity increases, subpopulations of bacteria that can survive the antibiotic due to spontaneous mutations emerge²⁹. As spontaneous mutations occur, those mutations that provide a selective advantage against the antimicrobial are carried forward to the next generation, leading to populations with resistance²⁹. Rather than by spontaneous mutation, bacteria can also gain resistance by horizontal gene transfer. Horizontal gene transfer involves the uptake of genetic information from another cell; it usually happens by conjugation^{29,30}. There are also phenotypic mechanisms of resistance, such as; biofilm formation, and persistent populations³¹. Biofilms are formed by bacteria adhering to a surface and producing an extracellular matrix. Biofilms can protect bacteria from antibiotics and often complicate hospital acquired infections by forming on catheters and breathing tubes. Persisting bacteria are subpopulations of bacteria that are in a transitive dormant state; these persistent cells are thought to be responsible for recurrent infections³².

Genetic resistance to antibiotics arises in the following ways: alterations or destruction of the antibiotic, reduced influx or increased efflux of the drug, or alterations of the target of the antibiotic (**Figure 1.4**). In the case of antibiotic alteration or destruction, many bacteria can produce enzymes that modify antibiotics, such as aminoglycoside modifying enzymes that can acetylate, adenylate, or phosphorylate aminoglycosides rendering them inactive³⁰. β -lactamases are hydrolysing enzymes that cleave the β -lactam from the penicillin-binding proteins (PBP); the expression of these enzymes can confer resistance to β -lactams³⁰. Modifications of the target of antibiotics can occur for many different antibiotics. An example of this type of resistance has been observed in β -lactams³⁰. Bacteria have a range of PBPs that are targeted by β -lactams; bacteria can acquire a resistant PBP (PBP2a) through horizontal gene transfer. PBP2a can function in the presence of β -lactams, rendering the cell resistant to β -lactams.

Gram-negative bacteria are more intrinsically resistant to treatment due to their impermeable outer membrane. Gram-negative bacteria, as well as their Gram-positive counterparts also express efflux pumps that actively push out antimicrobials. Low compound accumulation is a result of both an impermeable outer membrane and active efflux³³. The combination is commonly seen in clinical strains and is especially hard to treat.

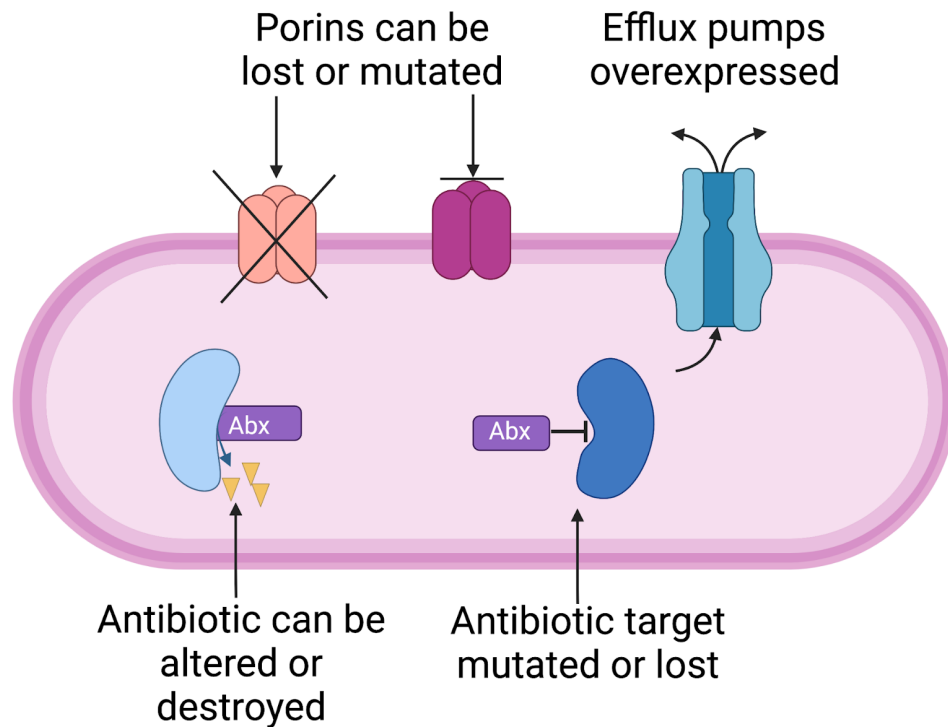


Figure 1.4 How resistance emerges in bacteria. Resistance in bacteria can occur due to decreased accumulation by; the loss or modification of porins, modification or loss of LPS, or overexpression of efflux pumps. The target of antibiotics can be lost or modified. Alternatively, antibiotics can be altered or destroyed by enzymes in the cell.

1.2 Tackling AMR

The rise of antimicrobial resistance has two key drivers; the misuse of antibiotics and the stagnation of antimicrobial drug development. Antibiotics have been misused and overused in healthcare, agriculture and veterinary practices, driving AMR rates¹. To address this, the WHO has suggested numerous approaches to antibiotic stewardship at a personal level, in healthcare settings and agricultural settings, to preserve the antibiotics in our arsenal^{27,34}.

During the last 20 or so years, there has been a mass exodus by big pharmaceutical companies from antibiotics drug discovery³⁵⁻³⁸. Big pharma has left for two main reasons; developing antimicrobials is not profitable, and secondly, in a practical sense, it is challenging^{35,36}. Developing antimicrobials is not profitable because infections are acute, so treatments are short, and resistance emerges quickly, limiting financial returns^{36,38}. The speed at which bacteria acquire resistance is increased further by physicians' willingness to prescribe novel antibiotics rather than saving novel antibiotics as a last resort³⁸. The WHO (2021) reports that small pharmaceutical companies have developed most

antibiotics in the preclinical pipeline. These small pharmaceutical companies do not have the funds for the entire clinical trial process^{16,38}. If smaller therapeutic companies bring a drug through the clinical trials process, they are at risk of going bust, as evidenced by Achaogen³⁹, which went bust after attempting to bring 'Zemdri' to market³⁷. There have been some attempts to mitigate this with 'push' and 'pull' incentives aimed to bring big pharma back into antibiotic development and to support smaller biotech companies throughout the crippling expensive development process. However, this has not been enough to increase the entry of new antibiotics to the market⁴⁰.

Aside from the financial practicalities impeding antibiotic development, developing antibiotics is inherently difficult due to the intrinsic properties of bacteria⁴¹. There are barriers to developing successful antimicrobials, such as accumulation, resistance rates and cytotoxicity^{41,42}. Accumulation refers to achieving inhibitory concentrations inside bacteria; the main barriers are efflux and permeability. A potential antimicrobial exhibiting a target-specific effect can have its effect significantly reduced *in vivo* due to attenuation of accumulation, either rendering the compound ineffective or requiring high levels of a compound that would be toxic^{43,44}. Therefore, understanding permeability issues is key to reducing the cost of developing these drugs and enabling the reinvigoration of the antimicrobial pipeline^{44,45}.

Alternatives to traditional antibiotic therapies are also being developed to treat infectious diseases. Vaccines against *Streptococcus pneumoniae*, a leading cause of pneumoniae, were successful in drastically reducing pneumoniae infections amongst children and elderly populations⁴⁶. However, vaccines have a narrow spectrum of activity, targeting not just a specific bacterial species, but also specific serotypes⁴⁶. The causative agents of infections like pneumoniae vary throughout different age groups, genders and locations⁴⁶. Therefore, targeting one pathogen will not give lasting protection against pneumoniae. Additionally, by targeting a specific serotype will enact a selective pressure on bacteria, causing populations of different serotypes to thrive. Alternatively, there has been a push to use other therapies such as; bacteriophages, probiotics, antibodies, and CRISPR-Cas9 treatments⁴⁷⁻⁵⁰. By having a diverse anti-infective arsenal the rise of AMR could be minimised.

1.3 Outer Membrane permeability

Both Gram-negative and Gram-positive bacteria have worrying rates of AMR. However, Gram-negative bacteria pose a more significant challenge to drug development because of the outer membrane^{45,51}. This factor is reflected in the critical pathogens selected by the WHO (**Figure 1.1**), which are all Gram-negative¹⁶.

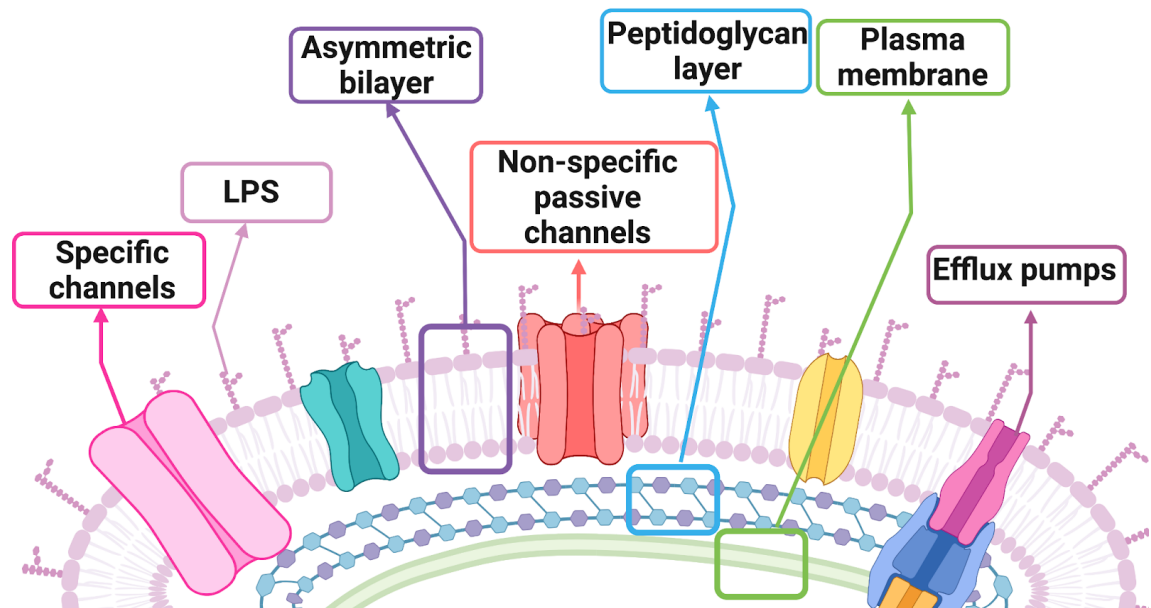


Figure 1.5 A diagram of the cell envelope of Gram-Negative bacteria. The major constituents of the outer membrane are LPS and phospholipids, which form an asymmetric bilayer. Within this outer membrane are transport proteins, non-specific (OmpF), specific (FhuA, LamB), channel proteins, and efflux pumps. Beneath this is a thin layer of peptidoglycan and a plasma membrane.

The outer membrane of Gram-negative bacteria is an asymmetric lipid bilayer (**Figure 1.5**)⁴². The inner leaflet of the outer membrane is composed mainly of phospholipids⁵²; the composition of phospholipids varies between species and varies due to conditions such as temperature and growth phase⁵². In *E. coli*, the inner phospholipid leaflet is generally regarded to have a composition of 75% phosphatidylethanolamine (PE), 20% phosphatidylglycerol (PG), and 5% cardiolipin⁵². The ratio of these phospholipids can be altered by the bacteria, as can the degree of saturation of their fatty acyl chains⁵². In low temperatures, the fluidity of the membrane is maintained by an increase in unsaturated fatty acids⁵². At high temperatures, the fluidity is maintained by an increase in saturated fatty acids⁵².

The outer leaflet of the outer membrane is primarily composed of lipopolysaccharide (LPS) (**Figure 1.6**). Lipid A makes up the lipid tail of LPS; appended to which are two sugar 'cores' ⁵¹(**Figure 1.6**). The disaccharide sugar core proximal to lipid A is composed of KDO (3-deoxy- α -D-manno-octulosonic acid) sugars, whilst the outer core is composed of hexoses. Attached to the core sugars is another oligosaccharide - known as the 'O-antigen'; this is the outermost component of the LPS, and it is immunogenic⁵¹ (**Figure 1.6**). The LPS in the outer leaflet is anionic; divalent cations stabilise the LPS, forming a rigid and stable layer⁴¹.

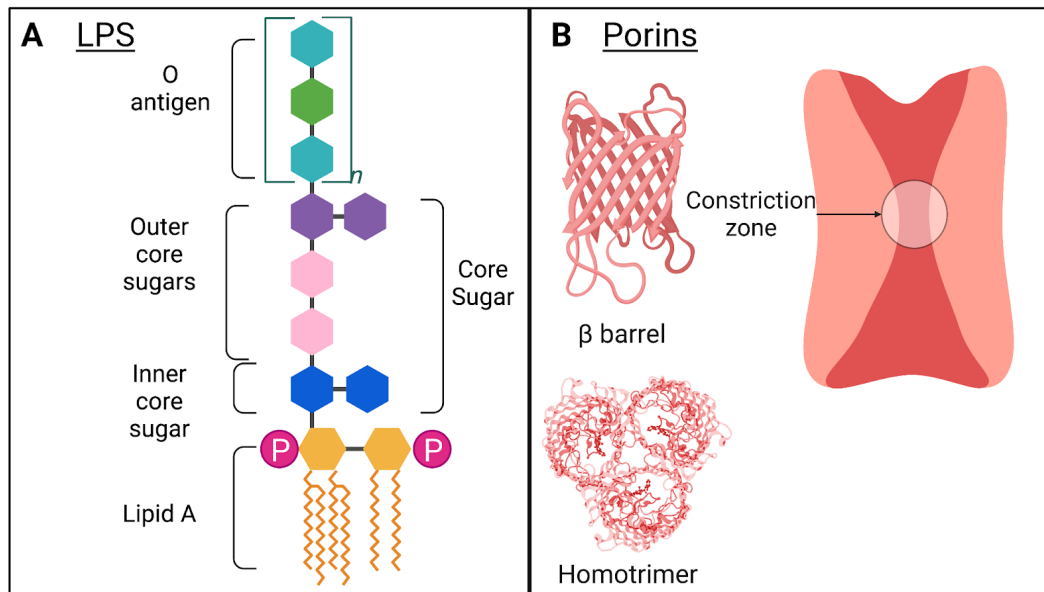


Figure 1.6 Key components of the outer membrane, LPS (A) and porins (B). LPS is a membrane anchored polysaccharide, made of lipid A, the inner sugar core (KDO2), the outer sugar cores (hexoses), and finally the O antigen. Porins are β -barrel homotrimers that have a constriction zone which is thought to confer some specificity.

The outer lipid layer occludes the entry of many key hydrophilic and hydrophobic molecules that are necessary for bacterial survival⁵¹. The outer membrane has many membrane channels that vary in their substrate selectivity to compensate for this. In *E. coli*, for example, there are selective transport channels such as LamB for sugars and FhuA for iron and non-selective transport channels such as OmpF and OmpC (**Figure 1.6**)^{41,42}.

OmpF and OmpC are homotrimeric β -barrel porins^{42,53} allowing solutes with molecular weights under 600 Da entry^{42,54} (**Figure 1.6**). Structural orthologues of OmpF and OmpC are found in other bacteria, such as OmpK35 and OmpK36 in *Klebsiella sp.*⁴². Hydrophilic antibiotics enter *E. coli* primarily through OmpF and OmpC^{42,55}. These porins are primarily composed of β -sheets; they have small extracellular turns and intracellular loops⁵⁶. One of

these loops (L3), folds into the channel creating a narrower region known as ‘the constriction zone’⁵⁶ (**Figure 1.6**).

Even though OmpF and C are generally considered non-selective, the constriction zone may result in some specificity in transport due to the electrostatic interactions between the acidic amino acids in L3 and antibiotics⁵⁷⁻⁶¹. Modifications to this region can make it harder for molecules such as β -lactams and chloramphenicol to travel through^{58,60,61} (**Figure 1.7**). Other species are less permeable to antibiotics than *E. coli* due to the porins they express; *Pseudomonas* species, for example, have monomeric porins that can only allow solutes with a mass smaller than 200 Da entry and have more electrostatic requirements⁴¹. The porins in *Pseudomonas* contribute to the relative impermeability of the bacteria, which, in general, is thought to be 80% less permeable than *E. coli*⁴¹. It is believed that most antibiotics permeate slowly through the lipid layer of *Pseudomonas*, because of the selectivity of their porins⁶².

When exposed to antibiotics, Gram-negative bacteria can further reduce permeability by altering the composition of the outer membrane⁵⁵. When exposed to colistin, bacteria can alter the phosphate head of the lipid A in LPS to reduce its negative charge, therefore reducing the electrostatic interactions between LPS and colistin^{63,64}. Bacteria can also become LPS deficient in order to evade colistin treatment^{63,64}. Changes to the lipid layer can affect antibiotics that translocate through the lipid layer: aminoglycosides, macrolides and rifamycins (**Figure 1.7**)⁵⁵. For small hydrophilic compounds like β -lactams, modifications or loss of porins can induce resistance⁵⁵. Mutations to the constriction zone of porins can cause resistance; by altering the amino acids in this zone, the selective electrostatic interactions that confer specificity are lost^{58,65}.

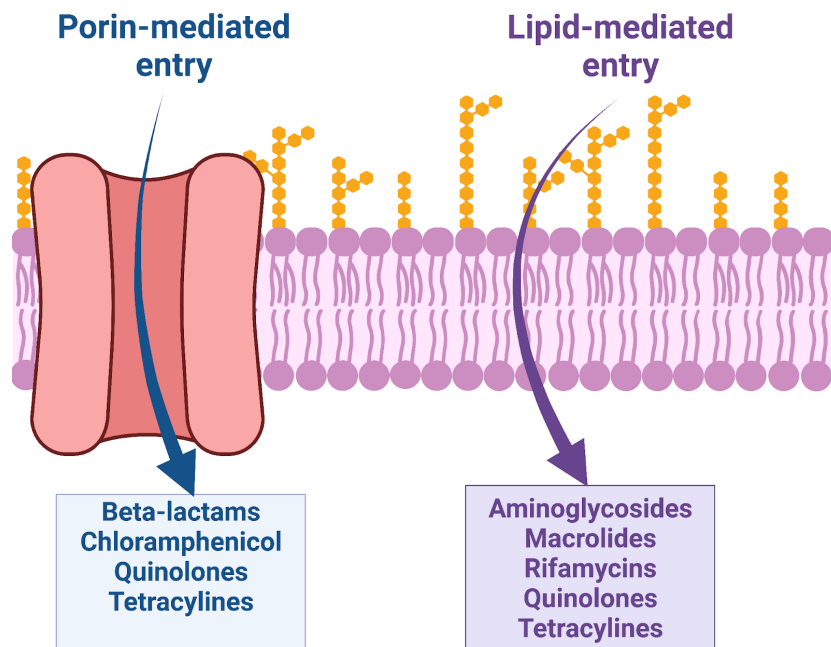


Figure 1.7 Antimicrobials have different entry pathways through the Gram-negative outer membrane. β -lactams and chloramphenicol primarily enter the cell through porins, whereas aminoglycosides, macrolides and rifamycins have lipid-mediated pathways. Quinolones and tetracyclines are capable of translocating the outer membrane via both pathways.

1.4 Permeability assays

Antibiotics are typically identified in two ways; phenotypic screening or target-based screening⁴³. In phenotypic based screens bacteria are exposed to large libraries of compounds and are identified based on antimicrobial susceptibility measured by minimal inhibitory concentrations (MIC) and minimal bactericidal concentrations (MBC). Whereas in target-based screening, a target of interest is identified, and potential inhibitors are screened for affinity and inhibitory activity. Both screening methods are expensive and yield minimal hits⁴³. To streamline both methods, a permeability assay could be used in tandem to either eliminate impermeable molecules or to define rules surrounding permeability for rational drug design.

Permeability assays either analyse individual factors influencing permeability, such as porins or the lipid layer or take a holistic approach by measuring accumulation in whole cells. Proteoliposomes, liposomes, and artificial bilayers have commonly been used to monitor permeability^{42,66-68}. Using these simplified models of the outer membrane allows for precise experimental manipulation to examine specific parts of permeability.

Whole-cell permeability assays provide a more realistic account for permeability *in vivo*. However, it is difficult to discern the individual contributions of each variable.

1.4.1 Use of enzymatic reactions to measure permeability

Permeability has been monitored using; enzymatic reactions, radiometric compounds, via compound fluorescence, mass spectrometry, and antimicrobial activity^{41,42,69}. Enzymatic reactions to monitor permeability were amongst the first means of measuring permeability⁷⁰. Of this group of assays, perhaps the most widely used is the micro-iodometric assay⁷⁰ (**Figure 1.8**). The micro-iodometric assay is centred around the use of triiodide starch solution, which is a deep blue colour^{71,72} (**Figure 1.8**). When cells are treated with penicillin, the penicillin is hydrolysed to penicilloic acid by β -lactamases⁷⁰⁻⁷² (**Figure 1.8**). The build up of penicilloic acid then decolourises the starch-iodide solution, showing that penicillin has successfully translocated the outer membrane^{71,72} (**Figure 1.8**). As β -lactamases rapidly turn over β -lactams, the rate-limiting step in this assay is the translocation of penicillin across the outer membrane. Thus decolourisation of the starch-iodide solution can be used as a readout of β -lactam permeability⁷⁰. This assay has been well categorised and does not require washing. However, this technique is not applicable to other antibiotics, and is therefore unsuitable for broad drug development screening.

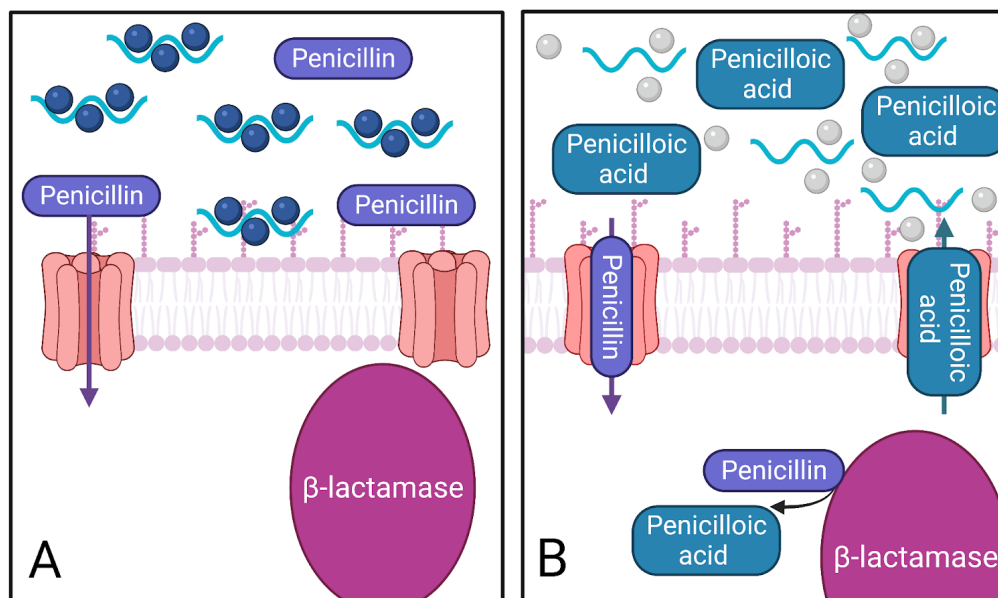


Figure 1.8 Micro-iodometric assay used as an enzyme readout for permeability. The triiodide-starch complex is a deep blue colour; this complex is decolourised rapidly in the presence of penicilloic acid. Before the breakdown of penicillin into penicilloic acid, the triiodide-starch complex is a deep blue colour (A). However, this complex is decolourised by the build-up of penicilloic acid by β -lactamase (B)

1.4.2 Use of radiolabeled compounds to measure permeability

Another common technique for measuring accumulation is by using radioactively labelled antibiotics. Here cells are incubated with the radioactivity labelled antibiotic, the excess antibiotic is washed away, and the radioactivity is measured. Radioactivity has been used to monitor ^{14}C rifampicin⁷³, ^{14}C azithromycin⁷⁴, ^{14}C erythromycin⁷⁴, ^3H tetracycline^{75,76}, ^3H chloramphenicol⁷⁵, ^{14}C norfloxacin⁷⁵, ^3H ciprofloxacin⁷⁷. It is a sensitive technique with a low background signal, and the data is directly quantifiable⁶⁹. The technique has been used with a range of bacteria, including; *E. coli*, *S. aureus*, *P. aeruginosa*, and in combination with other permeability techniques for greater confidence⁷⁷. However, this technique includes wash steps, which prevent the immediate measurement of permeability. Furthermore, radiolabelling thousands of compounds for permeability screens would be expensive, impractical and carries too many risks⁶⁹.

1.4.3 Use of fluorescence to monitor permeability

Another frequently used technique is to monitor compounds using their innate fluorescent qualities, or by fluorescently modifying antibiotics. Fluoroquinolones^{68,78,79} and tetracyclines^{80,81} have innate fluorescent qualities that allow permeability to be monitored based on this. Innately fluorescent antibiotics have been used to quantify permeability into liposomes^{25,68,79} through lipid layers and porins and through whole-cell analysis, looking at single cells or general populations of bacteria^{82,83}. The permeability of norfloxacin was quantified using liposomes and proteoliposomes trapped on a microfluidic chip^{68,79}.

There are two approaches in whole-cell fluorescent permeability assays: the drug is added, the cell is washed and lysed, and the fluorescence is quantified, giving permeability values for a population of bacteria^{82,84}. Alternatively, the permeability of fluorescent compounds in individual cells can be quantified, either by a microfluidic method^{83,85}, or by using deep UV micro spectrofluorimetry^{82,84}. These techniques can also be adapted for use with fluorescently labelled antibiotics, increasing their applicability⁸⁶⁻⁸⁸. The microfluidic techniques, in particular, allow for rapid quantification of permeability, which is often missed by other approaches due to wash steps⁸³.

1.4.4 Use of mass spectrometry to measure permeability

Mass spectrometry (MS) techniques are popular in quantifying the accumulation of antimicrobials^{69,77,89-96}. MS is a sensitive label-free technique that applies to a broad range of targets⁶⁹. The general workflow of this technique involves growing a dense bacterial culture, treating the bacteria with antibiotics, harvesting the cells, and then washing and lysing cells before analysing and quantifying the cell lysate⁶⁹. Some authors have proposed measuring the supernatant and measuring the decrease in the concentration of antibiotics as opposed to the accumulated antibiotic⁹⁰. MS techniques have been used in conjunction with radiometric measurements⁷⁷ and spectrofluorometric measurements to ratify and reinforce measurements⁹⁷. MS techniques have been used for large-scale accumulation assays^{89,96}. Richter *et al.*⁸⁹ generated some 'rules' surrounding permeability, leading to the authors converting narrow-spectrum drugs to broader spectrum drugs⁸⁹. More recently, a high-throughput solid-phase extraction MS technique was established; compatible with 96-well plates⁹⁶. Another solid-phase extraction MS technique was used in conjunction

with liquid-chromatography MS (LC-MS) to monitor the inhibition of their target enzyme CoaD via the concentration of metabolite ions, this led to a sensitive assay that aided in the optimisation of an inhibitor of CoaD⁹⁸. Criticisms of general MS techniques include the use of multiple wash-steps and non-specific binding, both of which lead to variability in results and initial rates or missed permeability measurements due to preparation steps. MS permeability assays also introduce bias into compound screening, as these techniques require the compound to be ionisable. Furthermore most MS techniques do not account for subcellular localisation^{43,69}. There have, however, been some developments to measure subcellular accumulation by MS, either by subcellular fractionation⁹⁵, or by imaging MS (TOF-SIMS)⁹⁹.

1.4.5 Use of drug susceptibility to measure permeability

Another method used to measure whole-cell permeability is via drug susceptibility testing. Here, bacteria are treated with antibiotics to find the minimal inhibitory concentration (MIC)⁴³. By using knockouts or mutants, an idea of the permeability properties of a drug can be built (**Figure 1.9**). This general method has been used by Iyer *et al.*¹⁰⁰ to create a permeability assay called 'titratable outer membrane permeability assay system' (TOMAS) that uses an *E. coli* knockout strain transformed with a OprD from *P. aeruginosa*, the predominant porin in this bacteria. The OprD gene was expressed in an arabinose expression vector, so the concentration of OprD was titratable, meaning, increasing the concentration of arabinose (the inducer of promoter on the plasmid), resulted in the expression of more OprD^{100,101} (**Figure 1.9**). The difference in MIC between the non-induced strains and the various titrations of arabinose was used to inform on porin permeability and preference¹⁰⁰. This assay was used to aid in the development of a PBP3 inhibitor¹⁰¹. This assay could be applied to other porins or specific mutations in porins. However, the parent knockout strain in the TOMAS assay struggles when efflux pumps are also knocked-out, limiting the ability of this assay to investigate both effects simultaneously.

Zgurskaya *et al.*^{33,102,103} also used a drug susceptibility technique as a read-out for permeability; they used strains of *E. coli* and *P. aeruginosa* which, overexpressed FhuA channels that had been mutated to remain 'open'^{33,102} (**Figure 1.9**). The authors also created efflux knockout versions of these strains to investigate the interplay between permeability and efflux³³. Zgurskaya *et al.*^{103,104} used this assay to predict the permeability

properties of *E. coli* and *P. aeruginosa*^{103,104}. These assays are sensitive to the porin used and are applicable to large drug screening techniques. However, using these drug susceptibility assays does not allow for resolution of subcellular localisation, nor do they allow accurate kinetics measurements.

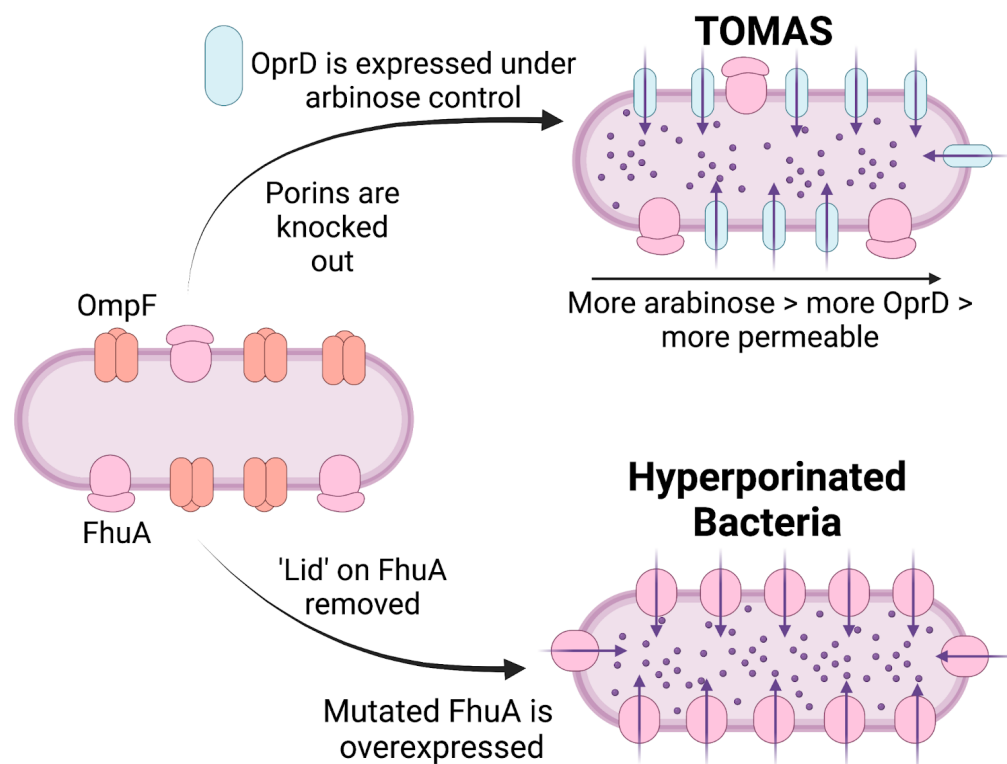


Figure 1.9 Antimicrobial susceptibility as a tool to measure permeability in two ways. Whole-cell susceptibility assays have been used to assess permeability, this has been adapted in two ways to look at permeability; TOMAS assay and hyperporinated FhuA strains.

1.4.6 Use of chemoinformatics and simulations in permeability studies

Computational techniques have been implemented to look at permeability, in the form of simulations of permeability or using cheminformatics to predict permeability. Molecular dynamics and Brownian motion simulations have been used to model antibiotics passage through porins and the lipid layer^{42,57,60,61,65,105,106}. These simulations have highlighted the importance of the electrostatic interactions in porins and the necessity to explore these interactions further to understand permeability better^{42,57,60,61,65,107,108}. Simulation studies have highlighted the differences in permeability between artificial membranes made of

lipids like phosphatidylcholine (POPC) compared to membrane mimetics, showing that different lipids affect the permeability of lipid-mediated drug uptake¹⁰⁹. Molecular dynamic studies have been used to aid in the development of new antibiotics¹¹⁰; using a combination of biochemical (target assays), molecular dynamics and the TOMAS permeability assay, diazabicyclooctane, an inhibitor of PBPs was developed^{100,101,110}.

Cheminformatic techniques determine the chemical space of compounds; this is then used to inform machine learning techniques to make predictions on compound activity. This approach was used by Stokes *et al.* to identify Halicin as a novel antibiotic¹¹¹. Zgurskaya *et al.* have been using cheminformatic techniques to elucidate the rules surrounding accumulation in *E. coli* and *P. aeruginosa*^{103,104}.

1.4.7 Use of outer membrane vesicles to model permeability

There have also been permeability assays that utilise outer membrane vesicles (OMVs) as models of permeability^{105,109} (**Figure 1.10**). OMVs are small (25-250 nm) vesicles created from the blebbing of the Gram-negative cell wall^{112,113}. OMVs contain an asymmetrical lipid bilayer, membrane proteins, and some soluble proteins from the original cell, such as β -lactamases¹¹³⁻¹¹⁵ (**Figure 1.10**). These OMVs have been used in swelling assays¹⁰⁵ and OMPA¹⁰⁹ (Outer Membrane vesicle based Permeation Assay). In OMPA, OMVs are reconstituted into a bilayer (using filtered 96-well plates, pegylated lipids and agarose), and the concentration of the drug through the bilayer is monitored via absorbance measurements and LC-MS¹⁰⁹ (**Figure 1.10**). In the OMV swelling assay, the size of OMVs is monitored by dynamic light scattering (DLS) in response to antibiotics over time¹⁰⁵ (**Figure 1.10**). The results of the swelling assay were used to compare predictions of drug permeabilities generated by molecular dynamic experiments.

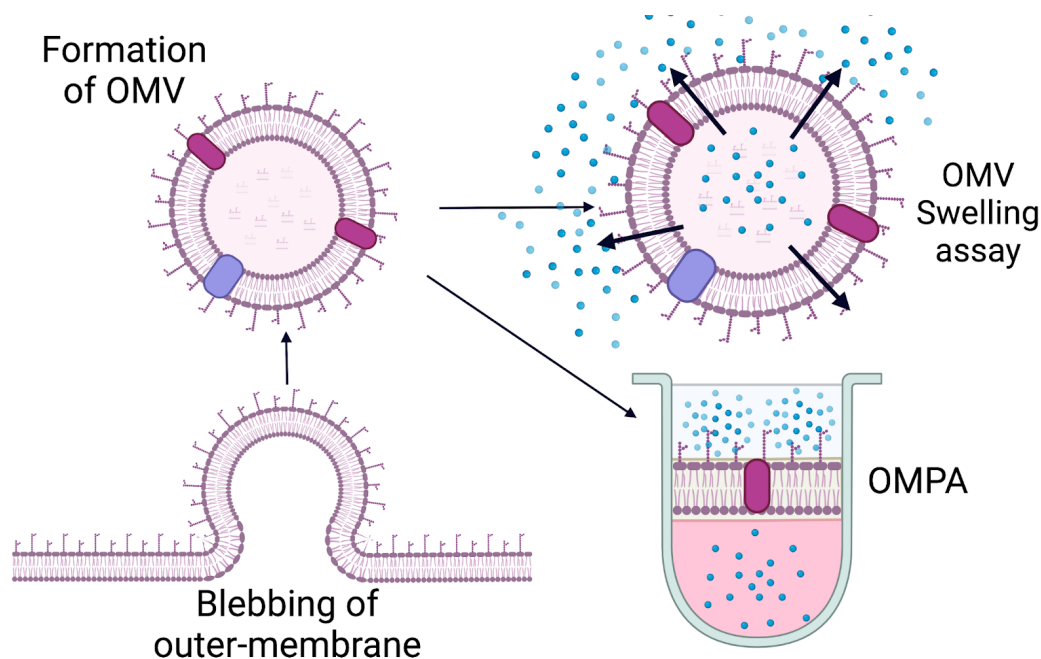


Figure 1.10 The formation of OMV and their use in permeability assays. OMV are formed from blebbings of the outer membrane. These OMVs have been used in swelling assays - where drugs were added, causing the OMV to swell increasing the size of OMVs. OMVs have been used in OMPA assays which constitute the vesicle into a bilayer and the accumulation of drugs at the other side of the bilayer is quantified by MS and spectrophotometry.

1.5 Identified rules of Gram-negative permeability

Some general rules surrounding permeability have been identified by utilising some of the permeability assays mentioned above in combination with large-scale physicochemical analysis of known antimicrobials. In 2008 O'Shea and Moser concluded that porin limitations control entry into Gram-negative organisms, finding that compounds <600 Da and highly polar compounds were more permeable¹¹⁶. They found that compounds in high throughput screens were generally less polar and smaller than antimicrobials on the market^{43,116}. A retrospective study of 3200 antimicrobials by AstraZeneca in 2014 found that hits from high-throughput screens tend to be more hydrophobic than active antimicrobials, but simply reducing hydrophobicity was not enough to encourage permeability^{43,117}. This is presumably, in part, due to the various entry points into bacteria, *i.e.* via porins or through the lipid layer, and due to efflux.

Another 2014 study by Davis *et al.*⁹¹ used an LC-MS accumulation assay to identify physicochemical properties affecting accumulation in *E. coli*, *B. subtilis*, and *M. smegmatis*⁹¹. They found variability in the physicochemical descriptors affecting each bacterium⁹¹. In *E.*

coli hydrophobicity (in agreement with O'Shea and Moser and AstraZeneca), ring content and size were positively correlated to accumulation⁹¹. In *B. subtilis*, flexibility and hydrophobicity were positively correlated to accumulation, whereas in *M. smegmatis*, ring content was positively correlated to permeability⁹¹. This study also noted the importance of efflux⁹¹. This study was done using one class of compounds (sulfonyladenosines) and noted the variability between bacteria⁹¹.

In 2017 Richter *et al.*⁸⁹ used an LC-MS assay to measure the accumulation of 180 compounds in *E. coli*⁸⁹. Two hundred ninety-seven physicochemical descriptors of the 180 compounds were used to identify key predictors of permeability using a random forest model⁸⁹. The key features identified were; an ionisable Nitrogen (primary amine), high Rigidity, and low Three dimensionality (globularity) - they termed these 'eNTRy rules'⁸⁹. Richter *et al.*⁸⁹ used these rules to convert antimicrobials specifically targeting Gram-positive organisms to broad-spectrum antimicrobials⁸⁹. In 2018 Iyer *et al.*⁹² monitored the accumulation of LigA inhibitors using an LC-MS accumulation assay, finding that positive charge and hydrophobicity increase intracellular accumulation^{42,92}.

Zgurskaya *et al.* 2018 used MICs and MIC ratios of hyperporinated (FhuA) *E. coli* and *P. aeruginosa* PAO1 mutants compared to wild-type and efflux mutants to identify the molecular properties of accumulation¹⁰³. They used 142 molecular descriptions that accounted for the charge, connectivity, molecular topology, number of atoms, number of bonds, physical properties, potential energy, shape and surface area of fluoroquinolones and β -lactams to inform a random forest model (like Richter *et al.*) and learn molecular properties of accumulation¹⁰³. They identified that permeability and efflux were positively correlated with different properties, in *E. coli* relative to *P. aeruginosa*¹⁰³. Molecular descriptors that were positively correlated with permeability in *E. coli* were rigidity and many electrostatic interactions^{42,103}. In *P. aeruginosa*, shape, flexibility, and negative partial charges of compounds were positively correlated with permeability^{42,103,104}. In comparison, the qualities that were positively correlated to efflux were partial positive charges, number of rotatable bonds, and lipophilicity^{42,103,104}. Zgurskaya *et al.* went on to use a machine learning algorithm to predict permeability from MICs of hyperporinated strains and molecular dynamics results¹⁰⁴.

There is some consensus around the rules of permeability; hydrophobicity, charge and shape. However, there is variability between the rules across different types of bacteria and across the rules that govern efflux. Further variability in these results could arise from

the method and conditions used in the permeability assay, variability in wash stages, or growth media used, which could alter the environment of the bacterium leading to differences in accumulation. Therefore, developing strict rules that define permeability may not be possible, especially until there is a better understanding of antimicrobial accumulation in bacteria.

To develop new antibiotics, more tools to understand Gram-negative permeability are essential. These tools could be in the form of a permeability assay used for high throughput screening that does not perturb the natural permeability process of bacteria. Or by creating models of the outer membrane that are representative and controllable to dissect the individual permeability components, or by fully elucidating the path of each antibiotic class to extrapolate these findings to new antimicrobials.

1.6 Thesis aims

- To develop tools and methodology to monitor β -lactam permeability in a label-free manner.
- To create a model of the *E. coli* outer membrane using synthetic biology, that can be used to measure permeability in a high throughput format.
- To use machine learning tools to create a model that can predict permeability.

Chapter 2. Material and methods

This chapter begins with general methods, followed by more specific methods relating to individual results chapters.

2.1 General materials and methods

2.1.1 Reagents and materials

All chemicals purchased for these experiments were of analytical grade, the majority of which came from Merck (Sigma), Fisher scientific, Melford, and Avanti polar lipids unless otherwise stated. The oligonucleotides used for molecular biological techniques were purchased from IDT.

All buffers were prepared using Ultrapure MilliQ water, buffers for molecular biology techniques, chromatography, and cell-free expression systems were all filtered through 0.22 μm MF-Millipore membrane filters (Merck). Buffer pH was adjusted using a SevenEasy pH meter (Mettler Toledo), calibrated at room temperature with pH 4.0, 7.0 and 10.0 buffer standards. Buffers for long term storage were kept at 4 °C.

2.1.2 Bacterial growth reagents and strains

2.1.2 Bacterial growth media

All media was sterilised by autoclaving for twenty minutes at 121°C. Antibiotic resistance genes were used to select the correct colonies. Antibiotic stocks at either 100 $\mu\text{g.mL}^{-1}$ ampicillin, 50 $\mu\text{g.mL}^{-1}$ kanamycin or 35 $\mu\text{g.mL}^{-1}$ chloramphenicol were added to media when necessary. Agar plates were poured in microbiological safety cabinets.

2.1.2.1 Lysogeny broth

Lysogeny broth (LB) media was used as the standard for most bacterial growth. Its composition, unless stated otherwise, was 1%(w/v) tryptone, 1%(w/v) sodium chloride, 0.5% (w/v) yeast extract¹¹⁸. For some protein expression trials this was supplemented with 1% (w/v) glucose to protect from leaky expression. Glucose solution was prepared and sterilised separately and was added when appropriate¹¹⁹.

2.1.2.2 LB agar plates

LB media was solidified by supplementation with 1.5% (w/v) bactoagar. The suspension was autoclaved for twenty minutes at 121°C, allowed to cool to 50°C, at which point appropriate antibiotics were added. 25 mL of this media was poured per petri dish and allowed to set in a microbiological safety cabinet.

2.1.2.3 Super Optimal broth with Catabolite repression

Super Optimal broth with Catabolite repression (SOC) media was used in the outgrowth step of chemical transformations¹²⁰. SOC was composed of 2% (v/v) peptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulfate and 20 mM glucose¹²⁰.

2.1.2.4 Low salt LB

For strains of *E. coli* lacking porins, low salt LB¹²¹ was used for growth and recovery. As it was observed that both BL21ΔABCF and BL21ΔACF grew poorly in high salt concentrations¹²². Low salt LB was composed of 1% (w/v) tryptone, 0.5% (w/v) sodium chloride and 0.5% (w/v) yeast extract. For low salt LB plates 1.5% (w/v) bacto agar is added¹²¹.

2.1.2.5 Cation-adjusted Mueller-Hinton Broth

Cation-adjusted Mueller-Hinton¹²³ broth was used for determination of minimal inhibitory concentrations (MIC). The medium was composed of 17.5 g.L⁻¹ casein acid hydrolysate, 3.0 g.L⁻¹ beef extract, 1.5 g.L⁻¹ starch, 20-25 mg.L⁻¹ Ca²⁺ and 10-12.5 mg.L⁻¹ Mg²⁺ (both from chloride salts) adjusted to pH 7.3¹²³.

2.1.2.6 Cell Free Auto Induction Media (CFAI)

CFAI broth¹²⁴ is used as a medium to perform protein over-expression. The medium was composed of 100 mM sodium chloride 2% (w/v) tryptone, 0.5% (w/v) yeast, 80 mM dipotassium hydrogen phosphate, and 44mM potassium dihydrogen phosphate¹²⁴. This was prepared with MilliQ water, buffered with 5.0 M potassium hydroxide to pH 7.2 and autoclaved¹²⁴. This was supplemented with a filter sterilised 40 mL sugar solution; 1.25% (w/v) glucose, 10% (w/v) lactose and 15% (v/v) autoclaved glycerol¹²⁴.

2.1.2.7 *E. coli* strains used in this work

A range of bacterial species and strains were used in this work. *E. coli* strains in particular were used for molecular biology techniques, protein purification, and compound testing. The strains used, their genotype and their use is described below in **Table 2.1**.

Strain	Genotype	Use
NEB-5a	<i>fhuA2</i> Δ (<i>argF-lacZ</i>)U169 <i>phoA glnV44</i> Φ 80 Δ (<i>lacZ</i>)M15 <i>gyrA96 recA1 relA1</i> <i>endA1 thi-1 hsdR17</i>	Plasmid overproduction
BL21 (λDE3)	<i>F- ompT hsdSB (rB- mB-)</i> <i>gal dcm rne131</i> (λ DE3)	Protein purification overexpression strain
BL21 pLysS	<i>F-, ompT, hsdSB (rB-, mB-), dcm, gal,</i> λ (DE3), pLysS, Cmr	Protein purification overexpression strain
Rosetta (DE3) cells	Genotype: <i>F- ompT hsdSB(rB- mB-)</i> <i>gal dcm</i> (DE3) pRARE (CamR)	Cell-free protein synthesis lysate
BL21ΔACF¹²² A gift from Jack Leo (Addgene plasmid # 102268)	Genotype = Δ <i>ompA</i> Δ <i>ompC</i> Δ <i>ompF</i> Precursor strain = BL21Gold(DE3) [genotype <i>F- ompT hsdS(rB- mB-)</i> <i>dcm+</i> Tetr <i>gal</i> λ (DE3) <i>endA the</i>]	Permeability control
BL21 ΔACFB¹²² A gift from Jack Leo (Addgene plasmid # 102270)	Genotype = Δ <i>ompA</i> Δ <i>ompC</i> Δ <i>ompF</i> Δ <i>lamB</i> Precursor strain = BL21Gold(DE3) [genotype <i>F- ompT hsdS(rB- mB-)</i> <i>dcm+</i> Tetr <i>gal</i> λ (DE3) <i>entheHte</i>]	Permeability Control
ΔOMPA¹²⁵ Strain: JW0940-6 CGSC: 8942	<i>F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3),</i> λ , <i>ΔompA772::kan, rph-1, Δ(rhaD-rhaB)568,</i> <i>hsdR514</i>	Permeability Control
ΔOMPC¹²⁵ Strain: JW2203-1 CGSC Number: 9781	<i>F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3),</i> λ , <i>ΔompC768::kan, rph-1, Δ(rhaD-rhaB)568,</i> <i>hsdR514</i>	Permeability Control
ΔOMPF¹²⁵ Strain: JW0912-1 CGSC Number: 8925	<i>F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3),</i> λ , <i>ΔompF746::kan, rph-1, Δ(rhaD-rhaB)568,</i> <i>hsdR514</i>	Permeability Control

Table 2.1 The *E. coli* strains used in this thesis, along with their genotype and source.

2.1.3 Preparation of competent Cells

Chemically competent cells were prepared via the Hanahan^{120,126} method, following this adapted protocol by Green and Rogers¹²⁷. A small overnight growth of 5 mL was set up,

containing LB and antibiotic if required, and inoculated with whichever strain was to be made competent. The following day this culture was diluted 1 in 10 into sterile LB (usually 2.5 mL of overnight growth into 250 mL of fresh LB). This was supplemented with any required antibiotic and 20 mM magnesium sulphate. When the cell density reached 0.6 as measured at an absorbance of 600nm, the culture was harvested by centrifugation at 4,500 × g for 10 minutes at 4 °C in a JA 25.50 rotor. The following steps were then performed on ice in microbiological safety cabinet: the cells were gently resuspended and washed with 100 mL of ice cold TFBI buffer (30 mM potassium acetate, 10 mM calcium chloride, 50 mM manganese chloride, 100mM rubidium chloride, 15% (v/v) glycerol, pH 5.8). This suspension is incubated for 5 minutes on ice, and then pelleted again as before. The cells were then resuspended in 10 mL of TFBII (10 mM MOPS pH 6.5, 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol). The resulting suspension was incubated on ice for 1 hour, before dispensing in 50 µL aliquots, flash freezing in liquid nitrogen and storing at -80°C. The competency of these cells was determined using pUC19 plasmid the following day.

2.1.4 DNA quantification

DNA samples were quantified by using an N60 Nanophotometer (Implen), measuring absorbance at 260 nm with either water or eluent buffer as blank.

2.1.5 DNA transformation of competent cells

DNA transformation of competent cells was done according to NEB supplied protocol. 50 µL of the desired competent cells were thawed on ice. 2 µL of plasmid was added to the competent cells, and mixed by flicking the tube 4 times. This mixture incubated on ice for 30 minutes. After the incubation on ice the cells were heat shocked at 42°C for 30 seconds, and then left to cool down on ice. 950 µL of SOC media was added to the tube, which is then incubated at 37°C for 1 hour at 180 rpm. After incubation, 150 µL of cells were plated on solid media supplemented with the relevant antibiotic and grown overnight at 37°C .

2.1.6 Glycerol stocks

Following successful transformations, one colony was inoculated into 10 mL LB and incubated overnight at 37°C with shaking at 180 rpm. A 250 µL aliquot of the culture was mixed with 250 µL of 50% (v/v) sterile glycerol, and flash frozen with liquid nitrogen before storage at -80°C.

2.1.7 Sequencing

Samples were sent to GENEWIZ (Leipzig, Germany) for sequencing. The samples were prepared as per the company guidelines. Once the result was received, the sequence data was translated with the ExPASy translate tool (SIB). The translated protein sequence was then analysed using the protein specific Basic Local Alignment Search Tool (BLASTP) against the theoretical sequence of the protein to check for alignment and mutations.

2.1.8 Minimal Inhibitory Concentration (MIC) microbroth dilution

MICs were used in both the iAmp Chapter (Chapter 3) and in the machine learning (Chapter 6). MICs were carried out in accordance with The Clinical and Laboratory Standards Institute (CLSI) guidelines¹²⁸. Briefly, bacteria were diluted in phosphate-buffered saline (PBS) to a cell density of 1.5×10^8 CFU/mL as determined by 0.5 McFarland standards, and further diluted 1 in a 100 in cation-adjusted Mueller-Hinton broth. In 96 well plates (GREINER), 200 µL of cation-adjusted Mueller-Hinton broth was supplemented with the double the highest concentration of antibiotic to be tested. From this first well 100 µL was carried across in a serial dilution. The wells were then inoculated with 100 µL of the working stock of bacteria. The plate was then incubated for 18-22 hours at 37 °C, after which the results were interpreted. For the polymyxin control, *E. coli* was supplemented with a sub inhibitory concentration of polymyxin (0.0625 µg/mL).

2.2 Protein purification

The proteins purified in this work followed a general workflow; transforming gene into expression line, large scale growth of expressing line expressing gene of interest, cells are pelleted, lysed by sonication, and soluble fraction is kept for purification. The soluble fraction was then purified by IMAC chromatography and analysed for purity on an SDS-PAGE gel. More in depth protocols for PBP3 and PBP4 are described below.

2.2.1 Purifying *E. coli* PBP3

A pET46b plasmid (kind gift from Dr D. Bellini) containing *E. coli* PBP3 (without the transmembrane helix) was expressed in *E. coli* BL21 (DE3). 3 L of culture were grown at 37°C with shaking at 180 rpm until an O.D.₆₀₀ reached 0.6, at which point protein expression was induced with 0.5 mM IPTG at which point, cells were grown overnight at 25°C. Cells were pelleted at 10 000 x g for 15 mins. Cell pellets were resuspended in 150 mL of buffer A (100 mM Tris, 500 mM NaCl, 20 mM imidazole, 2% glycerol (v/v), pH 8.0) supplemented with 20 µg mL⁻¹ deoxyribonuclease I from bovine pancreas and 1mg mL⁻¹ lysozyme from chicken egg white. Cells were lysed by sonication at 70% power on salted ice bath, ten times for 20 seconds interspersed by a minute in between to maintain cooling. Lysate was centrifuged at 16 000 RPM for 30 minutes at 4°C. A 5 mL HisTrap column was charged with nickel and equilibrated with 50 mL of buffer A. following this the supernatant was loaded onto the column with a peristaltic pump at ambient temperature and at a flow rate of 1 mL per minute, unbound contaminants were washed off with 30 mL of buffer A. Target protein was eluted on the AKTA 10/100 system (GE Healthcare) at a flow rate 2 mL/minute, with a gradient of 0-100 buffer B (100 mM Tris, 500 mM NaCl, 500 mM imidazole, 2% glycerol (v/v), pH 8.0) over 50 minutes where 2 mL fractions were collected. Pooled fractions were analysed by SDS-PAGE gels, then concentrated and buffer exchanged into a storage buffer (10 mM Tris, 500 mM NaCl, 10% (v/v) glycerol (pH 8.0).

2.2.2 PBP4

pET28b containing *E. coli* *dacB* gene (kind gift from J. Tod) was transformed into 50 µL stock of BL21 pLysS cells. From this a colony was selected and grown overnight at 37°C. 1

mL of overnight starter culture (37°C with shaking at 180 rpm) was used to inoculate 1 L of LB media supplement with 20% (w/v) glucose. This was grown at 37°C with shaking at 180 rpm until an OD₆₀₀ of 0.6 was reached. The culture was then induced with 0.5 mL of 1 M isopropyl β-d-1-thiogalactopyranoside (IPTG) and further grown for 4 hours. The culture was then pelleted at 10 000 x g for 15 mins. The pellet was resuspended in 100 mM Tris pH 8 at 3 mL buffer/g wet weight cells, and frozen at -80°C until use.

For purification 2.5 mg/mL lysozyme and 1 mg/mL of DNase was added to the pelleted cells which were thawed at 4°C on the roller. The cell suspension was sonicated at 70% pulses for ten 15 second bursts interspersed by 1 minute cooling on ice per round. The lysate was centrifuged at 50 000 x g for 45 minutes at 4°C, and the subsequent supernatant was loaded onto a 5 mL nickel-loaded HisTrap HP column (GE healthcare) equilibrated with buffer A. This was eluted with 100% buffer B over 50 minutes flowing at 1 mL/min) 2 mL fractions collected.

PBP4 proteins were further purified using a Cibacron brilliant red dye 3BA gravity flow column. The column was first washed with water then equilibrated with 100 mL buffer C (50 mM N-(2-hydroxyethyl) piperazine N'-(2-ethanesulfonic acid) (HEPES), 5mM dithiothreitol (DTT), 10% (v/v) glycerol, 2mM phenylmethanesulfonyl fluoride). Then fractions containing PBP4 were exchanged into buffer C via a 30 kDa molecular weight cut-off centrifugal concentrator (Amicon). PBP4 in buffer C was then added to the column, which was washed with 0-100% buffer D (buffer C supplemented by 0.5 M NaCl). Eluted fractions containing the desired protein were identified through analysis on 12% SDS PAGE gels.

2.2.3 Purifying deGFP

p70a-deGFP plasmid¹²⁹ was transformed into BL21 (DE3) cells. A colony of the successful transformation was used to inoculate 1 L of LB, which was grown overnight at 37°C 180 rpm shaking, in 2.5 L Erlenmeyer flasks with Tunair caps. The following day this culture was pelleted by centrifugation at 10 000 x g for 15 mins (Beckman JLA 8.1000). The pellet was resuspended in buffer A (3 mL per gram of pellet). This was then supplemented with 20 μg mL⁻¹ DNase and 1mg mL⁻¹ lysozyme. Cells were lysed by sonication at 70% power on salted ice bath, ten rounds of 20 seconds sonication, interspersed by a minute cooling. The lysate was centrifuged at 50 000 x g for 30 minutes at 4°C (Beckman JA 25.5). The

supernatant was loaded onto a 5 mL nickel column as before (2.2.1) and the same purification steps were taken as for PBP3.

2.2.4 Protein Quantification

Soluble proteins purified without detergents were quantified using either absorbance at 280 nm or the BioRad protein assay based on the method of Bradford¹³⁰ (1976). Membrane proteins were quantified using the bicinchoninic acid (BCA) protein assay (Smith *et al.* 1985¹³¹).

2.2.4.1 Absorbance at 280 nm

A nanophotometer (Implen) was used to measure absorbance at 280 nm. The protein buffer was used as a blank. The absorbance was then used to calculate the protein concentration. Here, protein concentration ($\text{mg}\cdot\text{mL}^{-1}$) was equal to absorbance at 280 nm multiplied by a protein factor ("Eprot") where Eprot was the Molecular weight (Da) divided by Molar extinction coefficient at 280 nm ($\text{M}^{-1}\cdot\text{cm}^{-1}$) multiplied by path length (1). The Eprot value was calculated by ExPASy ProtParam.

2.2.4.2 Bradford (BioRad) Protein Assay

10 μL of soluble protein was diluted into 800 μL of water following this 200 μL of Bio-Rad protein assay dye was added. This was incubated for five minutes at room temperature. The absorbance was then measured at 595 nm in a Jenway UV-Vis spectrophotometer. The concentration is derived from the following formula (calibration curve and formula for these calculations were by Dr. A. Lloyd): $\mu\text{g}\cdot\text{mL}^{-1}$ protein = mean absorbance 595nm/0.1 x 1.95×1000 / sample volume (μL)

2.2.5 Visualising proteins in SDS-PAGE gels

To visualise the protein, and tentatively identify purified proteins SDS-PAGE gels¹³² were used. Protein samples were run on SDS-PAGE gels in a TRIS-glycine buffer system^{132,133}. Gels were made using a Mini-Protean Tetra System gel kit (Bio-Rad). Gel components are outlined in **Table 2.2**. 10 μg of protein was mixed with a 4x (Laemmli¹³³) loading buffer

composed of 0.2M Tris-HCl pH 6.8, 0.4M DTT, 8% (w/v) SDS, 4.3 M glycerol. The samples were briefly heat denatured, then loaded in 12% SDS-PAGE gels (**Table 2.2**) and run at 180 V for ~50 minutes. The gels were stained with InstantBlue (Exedeeon) and imaged with an E-box illuminator (Vilber).

Component	Resolving 5 mL	Stacking 1.5 mL
Tris pH 8.8	375 mM	
Tris pH 6.8		125 mM
Acrylamide (29:1)	12% (w/v)	5% (w/v)
SDS (w/v)	0.1% (w/v)	0.1% (w/v)
Ammonium persulfate (APS)	37.5 μ L of 10% (w/v)	7.5 μ L of 10% (w/v)
Tetramethylethanediamine (TEMED)	7.5 μ L	1.5 μ L

Table 2.2 Components of 12% SDS-PAGE gels.

2.2.6 Bocillin gels

SDS-PAGE gels that were used to visualise bocillin-binding penicillin binding proteins were prepared according to previously reported method¹³⁴, this technique was used in Chapter 3 (iAmp), and Chapter 4 (MS β -lactam). Bocillin is a BODIPY fluorophore couple to penicillin¹³⁴. In this assay 16 μ g of PBPs were incubated with β -lactam antibiotic of interest, or moenomycin or water, for thirty minutes at room temperature. The PBP was then incubated with a 5:1 molar ratio of bocillin at 37°C for 1 hour (in the dark). Following this, the samples had 4x loading dye added, and they were analysed by SDS-PAGE (see methods above). To detect fluorescence, Typhoon FLA 9500 laser was used then gels were stained with InstantBlue (Exedeeon) and imaged with an E-box illuminator (Vilber).

2.3 Chapter 3 specific techniques - iAmp

For the completion of iAmp Chapter (Chapter 3) the following techniques were used.

2.3.1 iAmp synthesis

iAmp was synthesised by a two-hour incubation at 99°C in a thermocycler as previously reported¹³⁵. It was then purified by high performance liquid chromatography (HPLC).

2.3.2 HPLC

To purify the initial sample of iAmp 10 µL of 10 mg/mL sample and ampicillin were injected by the autosampler into HPLC (Agilent). This was chromatographed through an analytical reverse phase C18 column (Zorbax), 4.6 mm inner diameter, 150 mm length, 5µm particle size, and a pore size of 80 Å. Samples were eluted by a 5-95% (v/v) methanol gradient over 35 minutes, at ambient temperatures. Both fluorescence and absorbance detectors were used. The fluorescent detector was set to excitation 350 nm and emission 450 nm. The absorbance was monitored at 218, 254, 280 and 350 nm. To purify greater amounts of iAmp a preparatory reverse phase C18 column was used, and multiple injections were made and samples were taken and collated from the same time points.

2.3.4 Mass Spectrometry

The peaks from iAmp HPLC chromatography were analysed by direct injection nanoflow ESI in positive mode (+5.5 kV) on Waters G2 Mass Spectrometry. The samples were prepared by 1 in 10 dilution into water and 0.1% formic acid. The chromatograms were processed into spectra and interpreted using MassLynX software (Waters)

2.3.5 D-ala release coupled assay

PBP4 Carboxypeptidase activity led to the release of the C-terminal D-Ala from UDP-MurNac-pentapeptide which in this assay (*components outlined in Table 2.3*) was

coupled to the production of resorufin leading to a measurable absorbance increase at $A_{555\text{ nm}}$ (Catherwood et al.¹³⁶, 2020). DAAO catalysed the oxidation of D-alanine (D-ala) to pyruvate, ammonia, and hydrogen peroxide. The latter was reacted by horseradish peroxidase (HRP) with amplex red to generate resorufin. D-ala release was monitored on Varian Cary 100 UV/Vis double beam spectrophotometer at 30°C.

Component	Final Concentration/volume
Water	to a final volume of 0.2 mL
Buffer Bis-Tris Propane pH 8.5	50 mM
MgCl ₂	20 mM
D-amino acid oxidase (DAAO)	36.1 U mL ⁻¹
Horseradish peroxidase (HRP)	14.3 U mL ⁻¹
10-acetyl-3,7-dihydroxy phenoxazine (Amplex Red)	50 μM
PBP4	50 nM
UDP-MurNac-pentapeptide (DAP)	20 μM

Table 2.3 components in the D-ala release assay. The following assay components were kind gifts from J. Tod and Dr. A. Lloyd; DAAO, HRP, Amplex red, UDP-MurNac-pentapeptide. DAAO, and HRP U mL⁻¹ was determined by Dr. A. Lloyd as μmol.min⁻¹ of 1mM D-alanine for DAAO, and 20 μM H₂O₂ for HRP.

2.3.6 Imaging iAmp

An overnight growth of *E. coli* was prepared from a single colony, by growing in 10 mL LB at 37°C overnight, the following day this was diluted with LB to OD₆₀₀ 0.6, and grown for 2 hours. From this growth three samples were prepared in 1.5 mL eppendorfs, they were treated with either 200 μM bocillin, 200 μM iAmp or equal volumes water, for 30 minutes at 37°C. The excess compound was removed by washing the cells with PBS in 3 kDa MWCO spin filters. A 1.5% (w/v) agarose pad was made to immobilise bacteria. 2 μL of bacteria was placed onto a cover slip, and the agarose pad was overlaid, on top of this the glass slide was added. The samples were visualised using confocal microscopy (ZEISS LSM 880

airyscan). iAmp was visualised with diode laser using the set-up for ALEX 350 fluorophore, as it was the most similar possible, absorbance 346 nm, emission 442 nm. Bocillin treated samples were visualised with the argon laser and filters for excitation and emission of 485 and 520 nm respectively. Both samples were visualised using 100x oil lens. Images were analysed using FIJI ImageJ¹³⁷ software. Imaging was done with the help of Warwick Imaging Suite (Ian Hands-Portman) who facilitated sample preparation and training on the microscope.

2.4 Chapter 4 specific techniques - mass spectrometry and β -lactam acylation

2.4.1 Proteomic preparation

Peptides were either prepared using in-gel or in-solution technique followed by the filter-aided sample preparation (FASP) protocol and C18 stage tips. Protocols used were provided by Warwick Proteomics Research Technology Platform.

2.4.2 Proteomic gel digestion

20 μ g of PBPs incubated with 1 mM antibiotic were fractionated on an SDS-PAGE gel as above, and visualised as above but with the exception of a shorter staining duration. Gel slices containing the protein were excised and de-stained in 50% (v/v) ethanol and 50 mM ammonium bicarbonate. Once de-stained, the gel fragments were dehydrated with 100% (v/v) ethanol. The samples were then reduced by the addition of 10 mM (tris(2-carboxyethyl)phosphine) (TCEP) and acylated with 40 mM chloroacetic acid (CAA), for 30 minutes at ambient temperatures. The gel slices were washed three times with 50% (v/v) ethanol 50mM ammonium bicarbonate. This was again dehydrated with 100% (v/v) ethanol. The dehydrated gel slice was then rehydrated with 2.5 ng/ μ l trypsin and 50 mM ammonium bicarbonate and left to digest overnight at ambient temperature. Various incubation temperatures of this step were trialled, to probe the fragility of the covalent bond upon processing. The following day the peptides were extracted by sonication (bath) with the addition of 25% (v/v) acetonitrile and 5% (v/v) formic acid, and the extracted

solution was then concentrated under vacuum to 20 μl , to be resuspending in 2% (v/v) acetonitrile, 0.1% (v/v) trifluoroacetic acid (TFA) to 50 μl .

2.4.3 In-solution and FASP protocol

100 μg of protein incubated with 1 mM antibiotic was diluted with 8 M urea into 400 μl , and buffer exchanged into urea using 10 kDa molecular weight cut-off (MWCO) microcentrifuge ultrafiltration units, three times. All centrifugation steps were performed at 8 000 x g. This buffer was then exchanged into 50 mM ammonium bicarbonate three times. The sample was then reduced with 10 mM TCEP and alkylated with 40 mM CAA in 50 mM ammonium bicarbonate for thirty minutes at room temperature. The sample was washed three more times with 50 mM ammonium bicarbonate to remove TCEP and CAA. Sample was then digested with trypsin 2 $\mu\text{g}/100\mu\text{g}$ overnight at room temperature. The following day the peptide was separated from the proteases via centrifugal ultrafiltration as above. Peptides were then prepared using the stage tip protocol.

2.4.4 C18 Stage tip protocol

A C18 membrane was loaded into a 200 μl tip, conditioned with 100% (v/v) methanol for 2 minutes and centrifuged in a microfuge at 2000 rpm. The membranes were then equilibrated with 100% (v/v) acetonitrile and re-centrifuged. The membranes were then equilibrated with 2% (v/v) acetonitrile and 0.1% (v/v) TFA for 4 minutes at 2000 rpm. Then 10 μg of peptides were applied to the stage tips. The sample was then washed with acidified ethyl acetate (1% (v/v) TFA). Peptides were then eluted with 60-80% (v/v) acetonitrile. The acetonitrile was removed under vacuum, and the peptides were resuspended in 50 μL ammonium bicarbonate 2% (v/v) acetonitrile and 0.1% (v/v) TFA.

2.4.5 Proteomic mass spectrometry

Bottom-up proteomic samples were run by Warwick Proteomic Facility. The protocol used is as follows. A 5 μL aliquot of extracted peptides (total sample volume 50 μl) from each sample was analysed by means of nanoLC-ESI-MS/MS using an Ultimate 3000/Orbitrap

Fusion mass spectrometer (Thermo Scientific) using a 60 minute LC separation on a 50 cm column.

2.4.6 Proteomic search: Maxquant and Scaffold

MaxQuant¹³⁸ was used to perform a database search of LC-MS/MS data to identify protein Scaffold¹³⁹ was used to visualise and analyse proteomics data sets. Proteome databases for *Escherichia coli* (*E. coli*), *Pseudomonas aeruginosa* (*PSEAE*), *Burkholderia pseudomallei* (*BURKP*), and *Homo sapiens* were acquired from (UNIPROT).

2.4.7 FT-ICR MS

20 μM *E. coli* PBP3 was diluted into 50 mM ammonium acetate and washed three times with 50 mM ammonium acetate in 30 kDa MWCO 1.5 mL centrifugal concentrators. For the ceftriaxone PBP3 sample the *E. coli* PBP3 was preincubated with 100 μM ceftriaxone sodium salt. The 20 μM PBP3 in 50 mM ammonium acetate was diluted to 5 μM in water. All FT-ICR MS experiments were run and analysed by Dr Y. Lam (University of Warwick). The samples were injected through nano electrospray ionisation and run on Bruker 12 tesla Solarix FT-ICR MS.

2.4.8 MALDI-TOF

MALDI-TOF samples were run and analysed by Dr Y. Lam, who had also developed this protocol. Samples for MALDI-TOF (Bruker Microflex) were diluted to 1 μM in 1:1 water with 0.1% TFA:acetonitrile, 5 μL of the sample was mixed with 5 μL of matrix. 1 μL of this mixture was spotted onto the grid. Matrix was 40 mg/mL sinapinic acid dissolved in water with 0.1% TFA:acetonitrile 1:1. The Instrument was calibrated with bovine serum albumin (BSA), in positive ionisation mode. A mass range of 10-200 kDa was collected using a 60HZ nitrogen laser set at 82% strength. 2000 scans were accumulated. Data was analysed using Bruker software.

2.5 Chapter 5 specific techniques - synthetic cells

These are techniques specifically used in the creation and visualisation of synthetic cells as described in Chapter 5.

2.5.1 Creating cell-free extract

A colony of Rosetta (DE3) cells was picked and used to inoculate 1 L of cell-free auto-induction (CFAI) media in a 2.5 L Erlenmeyer flask, with tunair cap¹²⁴. The culture was grown at 30°C shaking at 200 rpm for 16 hours overnight. The protocol followed was adapted from Levine *et al.*¹²⁴. The following day the culture was pelleted at 5000 x g for 10 minutes at 4°C. It was then washed and resuspended with 50 mL ice cold filtered and autoclaved S30a buffer (14 mM magnesium glutamate, 60 mM potassium glutamate, 50mM TrisHCl, 2mM DTT). The resuspended cells were centrifuged again at 5000 x g for 10 minutes at 4°C. Pellet was then flash frozen with liquid nitrogen (even if lysate was being prepared the same day). The frozen pellet was resuspended with 1 mL per 1 g of pellet with S30a buffer (resuspension via vortexing and resting on ice). The cell solution was then sonicated at 50% in an ice-salt bath. Three cycles of 30 seconds on and 1 minute off were used. The lysate was centrifuged at 18 000 x g for 15 minutes at 4°C. The supernatant was collected, and concentrated in a 3000 Da MWCO centrifugal concentrator. The concentrated supernatant was aliquoted, flash frozen, and stored at -80°C till use.

2.5.2 Cell-free protein synthesis reaction

Cell-free protein synthesis (CFPS) reactions were set up according to the work of Noireaux *et al.*^{129,140-144}. Prior to full-scale CFPS reactions, each cell-free lysate extract required adjustment to the correct magnesium glutamate, PEG-8000 conditions, and plasmid concentrations. Therefore trial CFPS reactions were performed in small-scale plate reader assays in a PHERAstar plate reader (BMG) in 394 well plates (GREINER). p70_a-deGFP^{129,140} plasmid (Arbour biosciences) was used in all GFP work. For all CFPS reactions a master mix was always prepared to avoid pipetting errors as the reaction contained such small volumes. The reactions were incubated for 2 hours at 30°C and continuously monitored in the PHERAstar (BMG).

The components of the CFPS reaction are described in **Table 2.4**. The energy solution in this system was made as a 14x stock, composed of; 700 mM HEPES (pH8), 21 mM ATP and GTP, 12.5 mM UTP and CTP, 3.64 mM coenzyme A, 4.62 mM NAD, cAMP 10.5 mM, 0.95 mM folinic acid, 14 mM spermidine, and 420 mM 3-phosphoglyceric acid.

Component	Stock concentration	Final Concentration
ATP NAD mix (additional supplement)	100 mM	1.5 mM
Green fluorescent protein (deGFP) plasmid nM	481nM	15 nM
Crude extract	110 mg/mL	14 mg/mL
Magnesium glutamate	100 mM	5 mM
Potassium glutamate	1000 mM	60 mM
DTT	1000 mM	60 mM
Amino Acid solution	100 mM	1.5 mM
Energy solution	14x	1x
PEG x8000	40% (w/v)	2 %(w/v)
tRNA	100 mg/mL	0.2 mg/mL
Maltodextrin	500 mM	35 mM

Table 2.4 The components of the CFPS reaction. The energy solution is composed of HEPES pH 8, ATP, UTP, CTP, GTP, Co-enzyme A, NAD, cAMP, folinic acid, spermidine and 3-phosphoglyceric acid.

2.5.2.1 DNA preparation

For all cell-free protein synthesis reactions the DNA used was prepared using MAXI prep (Qiagen) to prevent contamination with RNase. The DNA was quantified using the Nanodrop (as mentioned earlier).

2.5.3 Creating synthetic cells

2.5.4 Preparing synthetic lipids

The lipids used in this work were; palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) (Avanti), *E. coli* polar lipid extract (Avanti), KDO2 - lipid A (Avanti), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DOPG), 1-oleoyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]dodecanoyl}-sn-glycero-3-[phospho-rac-(1-glycerol)] (NBD-PG) (Avanti), 1,2-Dihexadecanoyl-sn-Glycero-3-Phosphoethanolamine (Marina Blue™ DHPE) (Invitrogen). When the lipid was purchased it was resuspended in chloroform (to the appropriate concentration), and aliquoted into sealed glass vials that were coated in foil. Lipids were prepared in different ways depending on using the *ependorf technique* or the *microplate technique*.

Eppendorf technique was adapted from^{141,145}; for this technique this lipid was dissolved to a concentration of 20 mg/mL in chloroform. 40 µL of this was mixed into 400 µL of mineral oil, this was incubated at 65°C for 1 hour for the chloroform to evaporate. The lipid-oil solution was then desiccated overnight to remove residual chloroform.

Microplate technique was adapted from^{146,147}; for this technique the lipids were made up to a concentration of 8 mM. The *E. coli* lipid extract had a reported phospholipid profile of 67% (w/w%) phosphatidylethanolamine, 23.2 (w/w%) phosphatidylglycerol, and 9.8% cardiolipin, as such the weighted average of this (798.64 g/mol) was used to calculate the molarity. To a glass 5 mL vial, 75 µL of lipid stock and 1.5 µL of 500 µM fluorescent membrane stain Dil iodide [1,1-Dioctadecyl-3,3,3,3-tetramethylindocarbocyanine iodide] (AATBio) was added and dried under nitrogen stream. Then in an ATMOS dry bag (Sigma) 1.5 mL of mineral oil was added to the lipid, giving a final concentration of 400 µM lipid in mineral oil, and 0.5 µM Dil iodide. The lipid mineral oil mix was sonicated in a 30°C water bath for 1 hour, and then was left incubated at room temperature overnight.

2.5.5 Feeding Solutions

The solution surrounding the formed GUVs is known as the outer solution or as the 'feeding solution' (Nouireaux *et al.*¹⁴¹). It can be formed in three ways. The simplest is F1

which consists of; PBS and 200-400 mM glucose (depending on the osmolarity of the solution encapsulated by GUVs). F2 is the S30a buffer (*mentioned in section 2.5.1*) supplemented with PEG x8000 2% (w/v), and 35mM maltodextrin. The most complex feeding solution F3 contains all of the components of the CFPS reaction without the DNA or lysate; S30a buffer, 1.5 mM amino acids, 1x energy solution (*composition mentioned in section 2.5.2*), 35mM maltodextrin, 0.2mg/mL tRNA.

2.5.6 Preparing Synthetic Cells in Eppendorf method

To create synthetic cells in eppendorfs, a 20 μ L reaction of CFPS was supplemented with 80 mM glucose (final volume). This mixture was added on top of the lipid-mineral oil mixture and vortexed for 30 seconds, the CFPS-lipid in oil-and glucose solution was then left to equilibrate on ice for 10 minutes. This solution was layered onto 200 μ L of a feeding solution. The feeding solution contained the same components as the regular CFPS set up, without crude extract and plasmid. This mixture was then centrifuged for 30 minutes at 16 000 x g which resulted in the pelleting of giant unilamellar vesicles (GUVs) which contained the cell free mixture. This solution was incubated at 30°C for 2 hours, with DNase before imaging under argon laser confocal microscopy.

2.5.7 Preparing Synthetic cells and GUVs bulk production using microplate method

Glass bottom 96-well plates (Corning) were used to bulk produce liposomes. Plates were coated with 30 μ L of 2 mg/mL BSA, which was allowed to incubate for an hour. The excess BSA was removed, and the wells were washed with 100 μ L of outer solution also known as feeding solution. The outer solution or feeding solution used depending on the experiment, a mixture of F1, F2, and F3 were used in different applications. The optimised method is as followed; 50 μ L of the chosen outer solution was added to each coated well. Then the plate was transferred to a dry bag (ATMOS Sigma) flushed three times and filled with nitrogen for the rest of the protocol. From here 20 μ L of 400 μ M lipid in mineral oil was added carefully on top of the outer/feeding solution to form an interface. This was incubated for thirty minutes. An emulsion was made of 5 μ L inner solution (either CFPS or purified protein) in 250 μ L of 400 μ M lipid in mineral oil. This was mechanically agitated by dragging the eppendorf across an eppendorf rack. 50 μ L of this emulsion was added on top

of the interface. The plate was removed from the dry bag, and spun at 300 x g for 3 minutes to form the GUVs.

2.5.8 Imaging synthetic cells and GUVs in microplate

Synthetic cells generated by the microplate technique were visualised in the microplate due to the optimised glass bottom of the plate. All confocal fluorescence microscopy was done using ZEISS LSM 550 microscope (Warwick Imaging Suite). The fluorophores and their relevant wavelengths are supplied in the table (**Table 2.5**) below. The Zeiss 880 has laser lines for 405 (Diode), 458 (Argon), 488 (Argon), 514 (Argon), 561 (diode), 633 nm (HeNe). The relevant lasers and filters were used for each fluorophore. When multiple fluorophores were used, the 'best scan' option was used for visualising both fluorophores. In general the following options were used; 1024x1024, bit depth of 12, max speed, laser at 2%, max speed, 1 AU (airy unit). Different experiments called for different types of images, as such tile scans, z-stacks, and time series were used. When a tile scan was called for, a bounding grid method was employed. For a time series experiment, a scan was taken every 3 minutes for 2 hours. Data from microscopy experiments were analysed using FIJI ImageJ¹³⁷ software to create the final images shown. In general most images had brightness and contrast auto corrected. Z-stacks were merged, colour channels were overlaid.

Fluorophore	Absorbance/ Excitation (nm)	Emission (nm)
deGFP protein	488	510
mPlum protein (a kind gift from M. Sambrook)	590	648
Dil Iodide (AATBio)	550	564
Marina Blue DHPE (Invitrogen)	365	460
16:0-12:0 NBD PG (Avanti)	461	534
Calcein	495	515

Table 2.5 The fluorophores used in this work, and their excitation and emission values.

2.5.9 Calcein leakage assay

To test the effect of antimicrobial peptides on GUVs a calcein leakage assay was done following protocols adapted from Tamba *et al.* 2005¹⁴⁸, and Moniruzzaman *et al.* 2015¹⁴⁹. In this work 1 mM calcein was encapsulated (inner solution) in GUVs (using microplate GUV method). This experiment was done using an asymmetric lipid bilayer that consisted of KDO2-lipid A on the outer membrane, and *E. coli* phospholipid on the inner membrane (400 μ M). The polymyxin 50 μ M was added to the outer solution, so after the GUVs were formed they were in contact with the polymyxin.

2.6 Chapter 6 specific techniques - machine learning

In **Chapter 6** (using machine learning to predict permeability) there were two sources of data; one from MICs generated according to section 2.1.9 and the second from an open-source database of permeability results (SPARK). The compounds in the databases were converted into a molecular (mathematical) representation in four ways; physiochemical descriptors, Mol2Vec algorithm¹⁵⁰, extended-circular fingerprints¹⁵¹ (ECFPs), and by graph representations¹⁵². These molecular representations were then used to train three different types of models; a multi-layer perceptron, random forest model, and a graph convolutional model. More details describing these concepts can be found in **Chapter 6**.

2.6.1 The initial permeability dataset (generated in-house)

The initial permeability database used for the machine learning algorithm was based on MIC results. Methods of undertaking MICs were outlined in section 2.1.9 and the *E. coli* strains used were mentioned in 2.1.2. Other strains that were also used were *Pseudomonas aeruginosa* PAO1 and *Staphylococcus aureus* ATCC 29213 were also used in this work. Below is a list of antibiotics used in this work. All antibiotics were made up to a starting concentration of 3.2 mg/mL in water or DMSO and stored at -20°C until use.

β-lactams

1. Flucloxacillin
2. Methicillin
3. Penicillin G
4. Ampicillin
5. Carbenicillin
6. Amoxicillin
7. Piperacillin
8. Azlocillin
9. Nafcillin
10. Cefsulodin
11. Cefaclor
12. Cefoxitin
13. Cephalexin
14. Cephalothin
15. Cefepime
16. Cefoperazone
17. Cefotetan
18. Ceftazidime
19. Cefotaxime
20. Ceftriaxone
21. Cefixime
22. Aztreonam

Macrolides

1. Clarithromycin
2. Dirithromycin
3. Azithromycin

4. Clindamycin
5. Erythromycin
6. Telithromycin

Glycopeptide

1. Vancomycin
2. Dalbavancin
3. Oritavancin
4. Teicoplanin
5. Telavancin

Oxazolidinones

1. Linezolid
2. Tedizolid

Quinolones

1. Oxolinic acid
2. Levofloxacin
3. Fleroxacin
4. Ciprofloxacin
5. Sparfloxacin

Aminoglycoside

1. Amikacin
2. Neomycin
3. Gentamicin
4. Spectinomycin

Tetracycline

1. Minocycline
2. Doxycycline

Polymyxins

1. Polymyxin B
2. Colistin

Other

1. Metronidazole
2. Fusidic acid
3. Novobiocin
4. Chloramphenicol
5. Nitrofurantoin
6. Phosphomycin
7. Rifampicin
8. Trimethoprim
9. Daptomycin
10. Triclosan
11. d-Cycloserine
12. 6-Aminopenicillanic acid
13. Quinidine
14. Sulbactam
15. β-lactamase (*E. faecalis*)

The relative change between BL21 *E. coli*, and BL21 *E. coli* treated with 0.0625 µg/mL polymyxin B determined the permeability classes assigned to the compounds. With compounds whose MICs stayed the same with polymyxin treatment being categorised as Class 0 - whereby permeability is not a major issue. Whereas compounds that exhibited a large difference between wildtype and hyperpermeable strain were classified as either 1, 2 or 3 depending on the degree of difference.

2.6.2 Second permeability dataset - from open-source platform

Using the CO-ADD SPARK platform a new dataset was generated by searching their available platform for compounds with MIC results for wild type and hyperpermeable strains of *E. coli*. The newly curated dataset was checked for null or repeated compounds which were removed.

2.6.3 Software packages used in Chapter 6

Both datasets were created on Google Collaboratory notebooks, using Pandas and Numpy packages. All the graphs were created using Matlab Plot and Seaborn libraries. The molecular descriptions of compounds were created using RDKit, Mol2Vec, and DeepChem¹⁵³ packages. The models were created using primarily DeepChem¹⁵³ and also Scikit learn, WandB, Keras, and tensorboard.

2.6.4 Molecular representations of compounds

The compound structural information is denoted in a SMILES format in both datasets. These SMILES had to be converted to mathematical molecular representations to train the models. This was done in four ways; physicochemical descriptors, Mol2Vec¹⁵⁰, Extended connectivity fingerprints¹⁵¹ (ECFP) and Graph representations¹⁵².

To convert the compounds to physicochemical descriptors, RDKit¹⁵⁴ descriptor function and fragment functions were used to return the values. These physicochemical descriptors were used; the number of rotatable bonds, molecular weight, number of carboxylic acids, number of amines, number of hydrogen bond acceptors, number of hydrogen bond donors, LogP, globularity, number of heterocycles and number of pyridines. This means for each compound there were numerical features generated for each descriptor, and these were the inputs for the model.

Extended connectivity fingerprints¹⁵¹ were generated using DeepChem's 'featurising' tool¹⁵³. In each compound, the neighbouring atoms were mapped with a radius of 3; the mapping gives each substructure a numerical value. The unique numerical values from a compound are transformed into a bit vector string of 1024 values.

Mol2Vec¹⁵⁰ was used to create molecular representations using a Morgan fingerprint¹⁵⁵ (similar to ECFPs) with a radius of 1. The substructures generated from Morgan fingerprints have a numerical 'identifier'. Mol2Vec model was pretrained on 20 million compounds and has embedded the chemical space between compounds¹⁵⁰. The numerical

identifiers are fed into the pre-trained Mol2Vec model, which returns a 300 dimension matrix for each compound¹⁵⁰. This makes the input for models a 300 dimension vector.

For the graph convolutional model¹⁵², molecular representations are learnt during the training. In the graphs, the atoms become nodes, and bonds become edges; information about the molecule is learnt by aggregating information from surrounding atoms and updating the atom. The graph convolutional model comes from work by Duvenaud 2015¹⁵² and is implemented by DeepChem¹⁵³.

2.6.5 The models used to predict permeability

The **first model** used to predict permeability was a **small multi-layer perceptron**. It was built using SKLearn. The dataset had 63 compounds, four permeability classifications, and 12 physicochemical descriptors. First, the data was split into three stratified *K*-folds. Following this, the features are scaled using 'MinMaxScaler' to a value between 0 and 1. The model was a 'Sequential' model, with a dense input layer of 12 with a 'rectified linear use' activation, followed by a dense hidden layer of 8 with a 'rectified linear use' activation function, with a final dense layer of 4 neurons again with a 'softmax cross entropy' activation. The model was compiled with an 'SGD' optimiser, and the loss function used was 'categorical cross entropy'. A validation callback that monitored 'value loss' was used to prevent overfitting.

The **second model** created was a **multilayer perceptron**. The dataset used was the 750 compounds from the SPARK dataset. The compounds were featurised into a 300 dimension vector; the values returned by physicochemical descriptors were also used to increase the dataset's features to 317. The model used was a 'sequential' Keras model; the first input layer was a dense layer with 317 neurons and a 'rectified linear use' activation function. This fed into a dense layer of 300 neurons and 'rectified linear use' activation function, followed by a dropout layer of 0.2. This fed forward into another dense hidden layer of 300, with the same activation function. This fed into another dropout layer of 0.2, which finally fed into a dense output layer with three neurons (3 classes), with a 'soft max' activation function. The model was compiled with 'sparse cross entropy' loss, an 'SGD' optimiser, and a validation callback was used. The average results of 5 stratified *K* folds were reported.

The **third type of model** used was the **graph convolutional model**. The model was built in DeepChem¹⁵³, using the model architecture described by Duvenaud *et al.*¹⁵². In this model, the input for each compound is an input of 75 atomic features, and adjacent atoms are used to generate graph representations for each molecule¹⁵². The model has two convolutional blocks, one gathering block and one final output layer. Molecules are fed into the model in 'batches', and a batch size of 11 was used. Each convolutional block has graph convolutional layers with 128 neurons, using a 'tanh' activation function; this is fed into a batch normalisation layer, which is then fed into a graph pooling layer. After the two convolutional blocks, there is a gathering block which consists of one dense layer of 256 neurons with a 'tanh' activation function, followed by a batch normalisation layer, into a graph gathering layer using a 'tahn' activation function. This is fed into a final output layer which consists of three neurons for the multiclass and two neurons for the binary class. The loss function used was 'Categorical Cross Entropy'. The dataset was split into train, test and validation datasets (80:10:10).

The **fourth type of model** used was a **Random Forest** model; it was created using the DeepChem program, which had an SKlearn wrapper, meaning the root of functions used in this model were from Sklearn. Before using this tree, the dataset was split into train test and validation datasets (80:10:10). The datasets were transformed with a normalisation transformation function on the y data. For this ensemble model, 100 estimators (number of trees in the model) were used. The 'max feature' function determined the split in the trees, and the 'auto' function was used; this uses the square root of the input features ($\sqrt{1024}$ in this case), meaning random 32 features were used to determine the split.

The database and the successful models (binary GCN and RF) were uploaded to github and are available at: https://github.com/amhen/abx_perm_ML

Chapter 3. iAmp – a quantifiable β -lactam?

3.1 Background

3.1.1 β -lactam Permeability

Penicillins make up 50% of the prescribed antibiotics in UK primary care; it is, therefore, essential to better understand the mechanisms that allow drug permeation into Gram-negative bacteria¹⁵⁶, which are intrinsically more resistant to antibiotics due to their outer membrane. This outer membrane is both selective and restrictive, and as such, it has channels known as porins that allow essential nutrients to enter. Porins allow small (under 600 Da) hydrophilic compounds to cross the membrane, so they are a key entry point for antibiotics, including β -lactams (**Figure 3.1**)^{157,158}. The most abundant proteins in the *E. coli* outer membrane are the porins OmpF and OmpC, both are β -barrel homotrimers^{55,56}. Porins allow the passage of hydrophilic molecules as the inside of the channel is lined with hydrophilic residues⁵⁵ (**Figure 3.1.B.i**).

Porin channels have extracellular loops and transmembrane turns. The third extracellular loop folds into the centre of the porin and is known as the constriction zone¹⁵⁹ (**Figure 3.1.C**). It is thought that there are molecular interactions in the porin that confer selectivity in this constriction zone⁵⁹. As a mechanism of antibiotic resistance, porins can be lost or mutated to achieve resistance through impeding permeability^{65,160}. There are mutations in the constriction zone that are thought to reduce the permeability of β -lactams^{59,65,160}. Establishing a universal, sensitive and specific method that could measure β -lactams travelling through the porins would benefit our understanding of penicillin permeability.

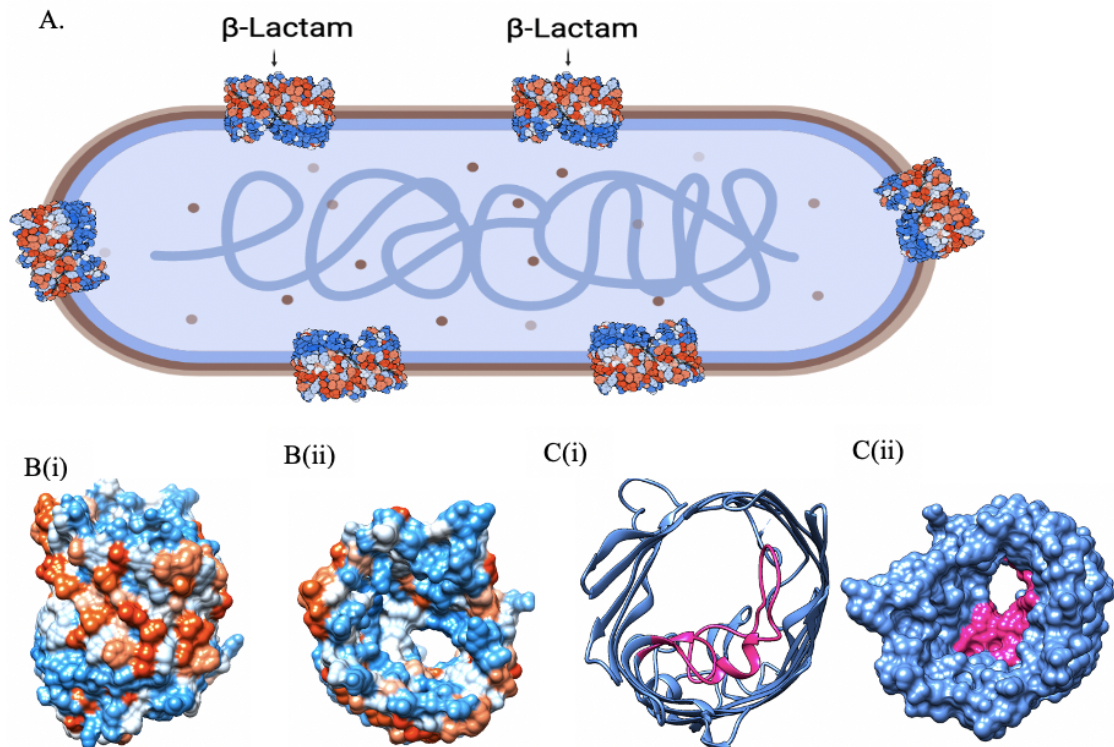


Figure 3.1 A depiction of *E. coli* cells where porins are present in the outer membrane. They have hydrophobic residues (in reds and oranges) on the outside of the channels to interact with the lipid environment of the membrane and hydrophilic residues (in blue) lining the porin channel which enable the transit of hydrophilic molecules such as β -lactams. **Figure 3.1.B**(i and ii) illustrates a monomer of OmpF with a single channel, though typically found as a homotrimer. **Figure 3.1.Ci** highlights (in pink) an extracellular *E. coil*-loop structure that extends into the channel and functionally constrict the channel. The narrowness of the constriction zone is exemplified in space filling model **3.1.C(ii)**. Figure was created using Biorender and Chimera¹⁶¹, the protein structure used was accession number 2ZFG¹⁶².

3.1.2 Fluorescent antibiotics

Fluorescent techniques allow biologists to visualise both biological and chemical processes. Fluorescent antibiotics allow the study of the interactions between drugs and bacteria. Consequently, they are conducive to observing drug permeability¹⁶³. Some antibiotics, such as fluoroquinolones, have weak intrinsic fluorescence, which means the drug can be monitored^{68,79}. Fluoroquinolones such as; ofloxacin, ciprofloxacin and norfloxacin contain quinolone groups, which exhibit fluorescence qualities when excited at the appropriate wavelength. As such, fluoroquinolones have been used to study permeability. Fluoroquinolones have been used in microfluidic chips to monitor the permeability across lipid layers and through porins^{68,79}, and more recently to monitor permeability across single cells^{83,85}. Fluorescent antibiotics have allowed the authors to

monitor the accumulation and subcellular localisation of antibiotics in a label-free method⁸³.

Unlike fluoroquinolones, most β -lactams are not innately fluorescent; this impedes the monitoring of β -lactam permeability. Antibiotics can be modified with fluorescent groups, which can help visualise accumulation. However, fluorescent analogues tend to have lowered efficacy; the addition of fluorophores will impact structure-activity relationship (SAR) studies^{86,164}. The reduced efficacy could be due to either a reduced affinity to the target or an alteration in permeability. Permeability can be reduced by an increase in negative charge¹⁶⁴. Alternatively, the changes to the molecule could alter the natural permeability process for a compound. Fluorescent groups tend to have lipophilic characteristics, meaning the addition of the fluorophore could change the properties of compounds, thus affecting the permeability of the drug¹⁶³. This prohibits the investigation of the permeability of these compounds. Bocillin is a popular fluorescent β -lactam often used in SAR studies; this is a bodipy-fluorophore conjugated onto penicillin (**Figure 3.2**)¹³⁴. Bocillin has enabled many drug interactions to be studied and is frequently used to examine the affinity of PBPs for different antibiotics via inhibition competition assays¹³⁴. It is also often used in visualisation. However, having the bodipy group on the β -lactam adds lipophilicity, meaning it may bypass the porins and slowly translocate through the lipid layer. As a result, it cannot be used to study porins. Therefore having a β -lactam with demonstrable fluorescent properties would be helpful to understand Gram-negative permeability.

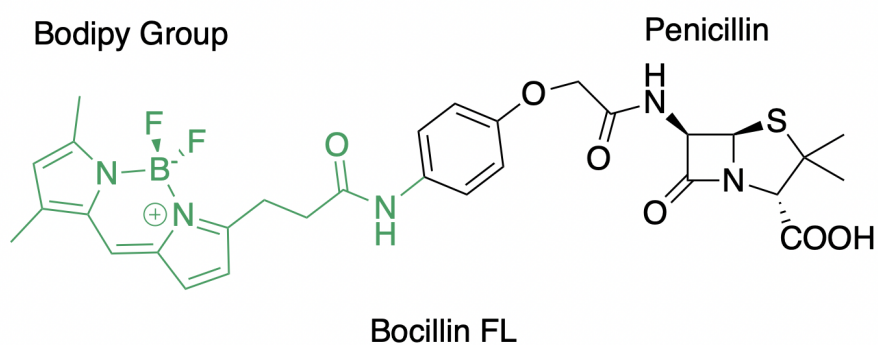


Figure 3.2 Structure of bocillin. The bodipy group is highlighted in green.

3.1.3 iAmp

Kotagiri *et al.*¹³⁵ reported the discovery of a fluorescent degradation product of ampicillin that retains the β -lactam ring of ampicillin. They named the compound iAmp. Kotagiri *et al.*¹³⁵ reported the mass of iAmp as 348.8 Da and the mass of ampicillin to be 349.4¹³⁵. They proposed that an oxidative deamination reaction at the primary amine leads to the creation of this fluorescent derivative (**Figure 3.3**).

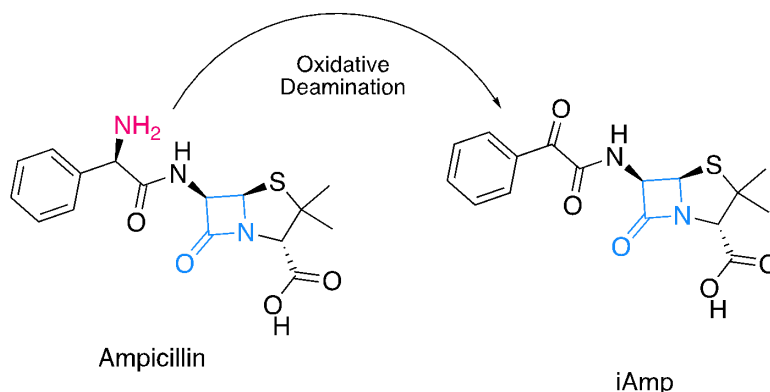


Figure 3.3 The proposed structure of iAmp. The formation of iAmp as described by Kotagiri *et al.*¹³⁵, in blue is the β -lactam ring, and in pink is the primary amine of ampicillin. iAmp is thought to lose its primary amine - gaining a ketone group. The β -lactam ring inhibits PBPs covalently binding to the highly conserved active site serine residue. Therefore, maintaining the integrity of the β -lactam ring is key for functionality comparisons.

If iAmp maintains the β -lactam ring, it may be a good substitute for monitoring ampicillin permeability fluorescently. and, for example, using the microfluidic accumulation method used to study quinolones, which are innately fluorescent^{68,79}, or more broadly by looking at fluorescence polarisation. There is an established bocillin fluorescence polarisation assay that is used to screen for inhibitors of PBPs¹⁶⁵. When bocillin is excited with polarised light it has rapid rotation due to its small size, leading to depolarised light emission. In contrast, when bocillin is bound to PBPs it has a much slower rotation (because of the large size of the complex), leading to polarised light being emitted. As β -lactams have a higher affinity for PBPs than bocillin so they displace the binding of bocillin, so depolarised light is emitted. Ideally iAmp could be used in place of bocillin, and this assay could be used as a proxy for permeability. The research presented in this chapter aimed to validate the reported structure iAmp, the method used to produce it, and evaluate its use in the study of the permeability of β -lactams.

3.1.4 Research aims

- Synthesise and purify the previously reported fluorescent derivative of ampicillin
- Establish if this fluorescent product of ampicillin degradation can be used to quantify the permeability of β -lactams.
- Investigate the effect of outer membrane contributions from porins and lipid layers on the permeability of β -lactams.

3.2 Results and Discussion

3.2.1 Synthesis of iAmp

Kotagiri *et al.*¹³⁵ report the formation of a fluorescent derivative of ampicillin upon heating to 99°C. To replicate this, 10 mg/mL of ampicillin was heated in a thermal cycler (Agilent SureCycler 8800) at 99°C for two hours. The heat degradation of ampicillin led to an observable yellow colour change in white light. Under ultraviolet (UV), blue light was emitted from the heat-treated ampicillin, iAmp (**Figure 3.4**). The fluorescent excitation and absorption spectrum were as described by Kotagiri *et al.*¹³⁵ (**Figure 3.5**). The heat-degraded ampicillin demonstrates the same fluorescence observed by Kotagiri *et al.*¹³⁵.



Figure 3.4 Eppendorfs containing ampicillin on the left, and iAmp on the right, visualised in gel imager. Ampicillin and iAmp are visualised on a UV-gel doc. The iAmp sample appears saturated and in this image black, whereas the ampicillin sample does not emit any light at this wavelength so appears transparent.

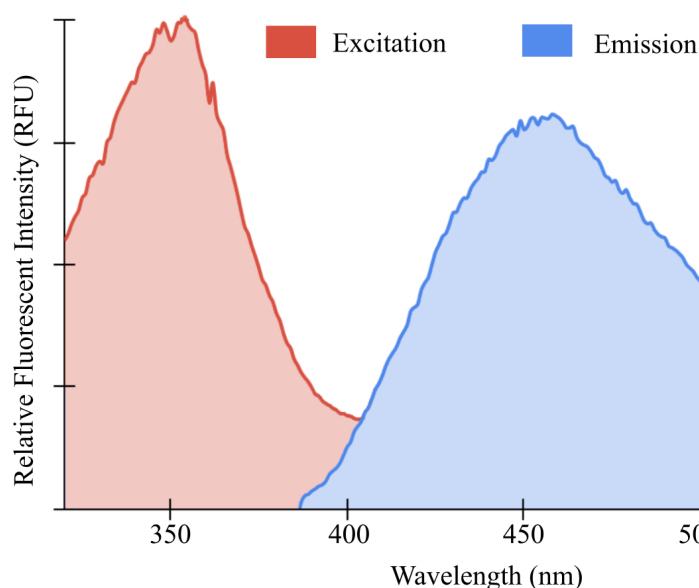


Figure 3.5 Fluorescence spectrum of the compound purified by High-performance liquid chromatography (HPLC). The observed fluorescence is inline with that which was reported by Kotagiri et al.¹³⁵. The excitation is highest at ~350 nm. Emission (at an excitation wavelength of 350 nm) has a broader peak around 440-460 nm.

3.2.2 Purification of iAmp

To investigate the structure and activity of the fluorescent derivatives of ampicillin, the heat-degraded sample (presumed to be iAmp) was purified by preparatory reverse phase C18 HPLC. In this instance, the stationary phase (the column) was hydrophobic, meaning compounds that are less polar and have a lower hydrophobicity will have a greater affinity to the column. In this purification, the mobile phase was a methanol gradient (5-95% methanol). As a result, more hydrophilic and polar compounds will have a greater affinity for the methanol and elute faster. The HPLC separation was analysed/observed using fluorescence detection (Ex./Em. 350/450 nm) and absorbance detection (at 218, 254 and 280 nm).

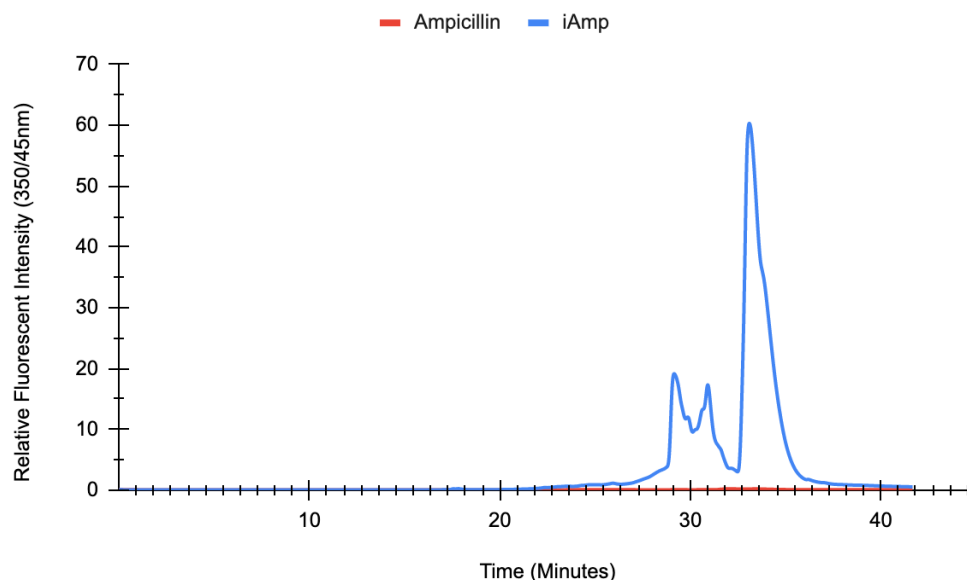


Figure 3.6 HPLC elution profile of ampicillin and iAmp from a reverse phase C18 column monitored using fluorescence detection Ex. 350nm/Em. 450 nm. 10 μ L of 10 mg/mL sample was injected. Flow rate 3 mL/min, gradient 5-95% water:methanol. Samples were monitored fluorescently at Ex./Em. 350/450 nm, and via absorbance at 218 and 254 nm, which are shown below.

When monitoring the separation using the fluorescent detector at Ex. 350 nm, Em. 450 nm, ampicillin peaks are not visible, as the ampicillin is not fluorescent (**Figure 3.6**). There are fluorescent peaks at the end of the run, 28 to 32 minutes, and 33 to 36 minutes (**Figure 3.6**). The late timing of these peaks suggests that the fluorescent compounds have hydrophobic properties. The peaks were analysed by MS, the predominant species in the latest peak had a m/z of 349, the major species in the other peaks were not consistent with the expected mass of iAmp.

HPLC fractions were collected at different wavelengths; this allowed the separation of ampicillin from iAmp and allowed the observation of different chemical characteristics. Double bonds or other conjugated systems have strong absorbances at 218 nm. Ampicillin demonstrated two characteristic absorbance peaks at 218 nm between 20 and 30 minutes (**Figure 3.7**), resolvable by HPLC. The iAmp chromatogram recorded at this wavelength was more complex, with 218 nm absorbing species eluting within the same time frame as the two prominent peaks in the ampicillin sample (**Figure 3.7**). These peaks could be consistent with the incomplete conversion of ampicillin to iAmp.

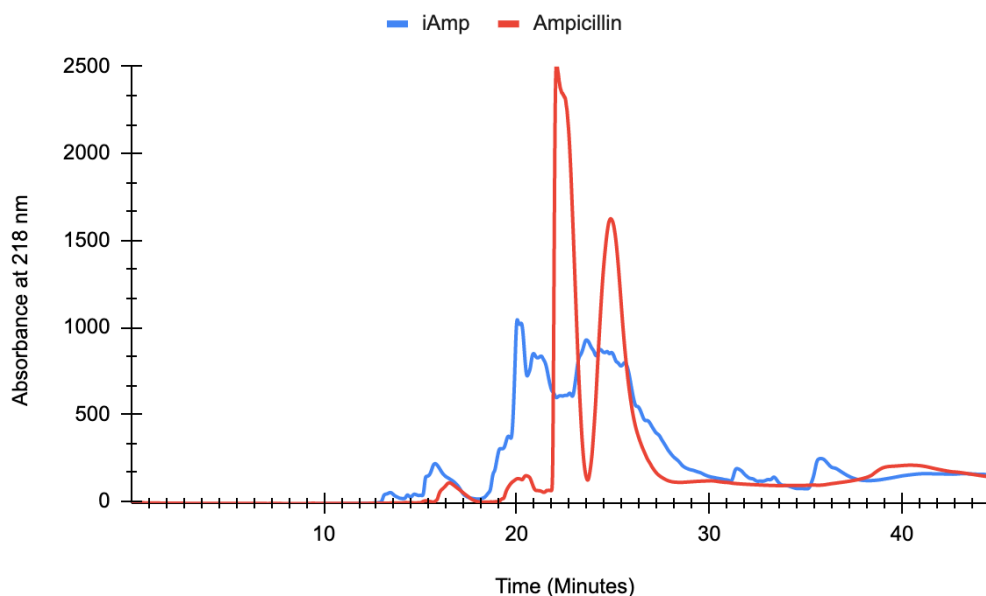


Figure 3.7 HPLC trace of iAmp and ampicillin purified by reverse phase C18 monitored at 218 nm. Sample loading as above.

Monitoring reverse phase HPLC at 254 nm of iAmp allowed detection of more highly conjugated chemical groups such as delocalised aromatic species. Again, after 30 minutes, large peaks eluted, suggesting that the compound eluting at this time was structurally distinct from and more hydrophobic than ampicillin (*Figure 3.8*).

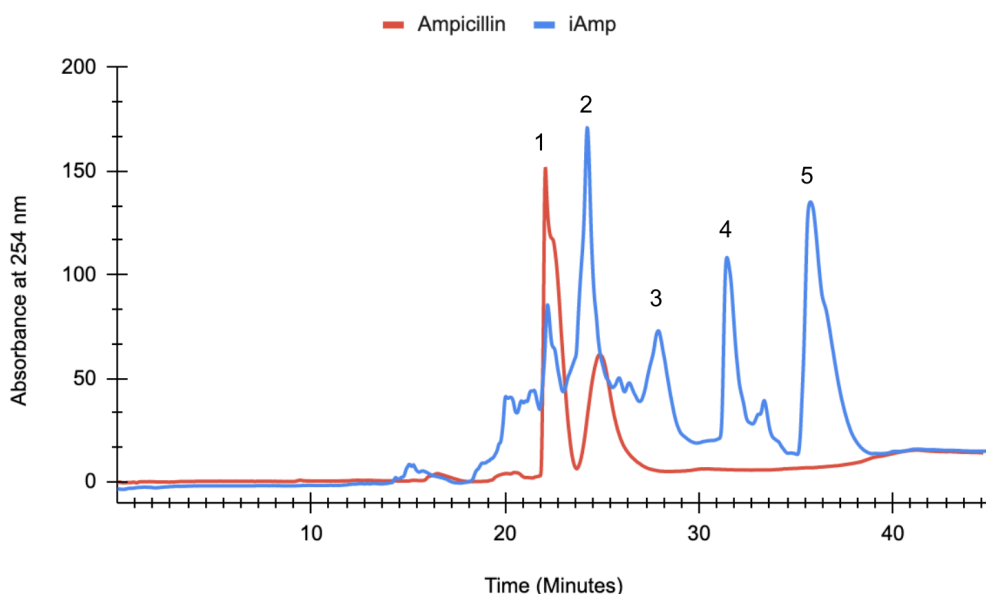


Figure 3.8 HPLC trace of iAmp and ampicillin purified by reverse phase C18 monitored at 254 nm. Sample loading as above.

The compound(s) derived from the synthesis of iAmp from ampicillin that were observed by HPLC appeared to have different physicochemical properties than ampicillin. The

fluorescence emission spectrum of iAmp following excitation at 350 nm differed from ampicillin but in addition to this, the difference in absorption at 254 nm suggests that the fluorescent compound possessed different characteristics (**Figure 3.8**). The fluorescent compound appeared to be more highly conjugated and less polar. Notably, the latter characteristic suggested a potential to change how the compound enters the cell relative to its ampicillin progenitor. Ampicillin passes through porins which are channels for hydrophilic compounds^{55,166}, iAmp seems to be much more hydrophobic. This evidence suggested that perhaps iAmp would not pass through the porins.

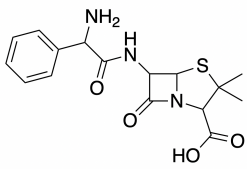
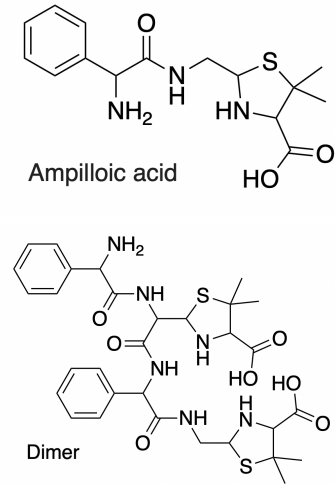
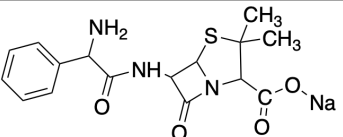
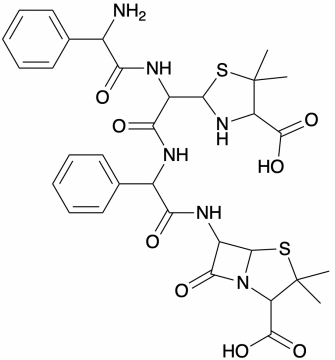
Samples were collected from both runs according to their peaks at 254 nm and their fluorescence at 350/450 nm. Fourteen samples from 20 to 38 minutes were taken to cover the five main peaks (**Figure 3.8**). These samples were taken for mass spectrometric analysis to look at the predominant species in each peak. Primarily to identify the fluorescent peaks for iAmp and other fluorescent derivatives, as there have been many reports of fluorescent ampicillin derivatives. Fluorescent derivatives of ampicillin were first described by Jusko *et al.* in 1971¹⁶⁷.

Furthermore, Barbhaiya and Turner in 1976¹⁶⁸, noted that fluorescent derivatives are formed from a range of β -lactams¹⁶⁸. Jusko *et al.*¹⁶⁷ proposed the fluorescent compound to be diketopiperazine. Following this, Uno *et al.*¹⁶⁹ identified and characterised many fluorescent derivatives of ampicillin, including one with a singly charged mass:charge ratio (m/z) of 349.1: 2-hydroxy-3-phenyl-6-[aminophenylacetyl]amino]ethylpyrazine(2-hydroxy-3-phenyl-6-pennillomethyl pyrazine)¹⁶⁹. This observation was further confirmed by Baertschi *et al.*¹⁷⁰, who identified and classified fluorescent compounds using HPLC, NMR, and MS.

3.2.3 Electrospray ionisation mass spectrometry of key HPLC Peaks

The HPLC system's automated peak detection and selection function led to collections spanning peaks 1 to 5 when monitored at 254 nm (**Figure 3.8**). Positive-mode electrospray time-of-flight MS (**Figure 3.9**) indicated that the predominant species in peaks 1 and 2 were ampicillin, sodiated ampicillin, and oligomers of ampicillin (**Table 3.1 for a list of masses for species**) (**Figure 3.9**). Ampicillin forms oligomeric species due to a

nucleophilic attack on the β -lactam ring; it can form either closed-cycle or open-cycle oligomers^{171,172}. A closed-cycle dimer is an oligomer of ampicillin with one species of ampicillin that contains a β -lactam ring, whereas, in an open-cycle dimer, both rings are hydrolysed. Closed-cycle dimers and trimers were observed¹⁷² (**Figure 3.9, 3.10**). The observed masses in **Table 3.1** were tentatively assigned based on literature from¹⁷²⁻¹⁷⁵.

	Predicted monoisotopic mass [M+H] ⁺	MS ² (m/z) [M+H] ⁺	Structure
1	350.1174 (Ampicillin)	333 [Amp-NH ₃] ⁺ 160 192	
2	324.1382 (Ampilloic acid) 673.2545 (Ampilloic acid dimer)	307	 Ampilloic acid Dimer
3	372.1001 (Sodiated ampicillin)		
4	699.2272 (Closed-cycle dimer - 1 open ring)	540.1, 381.1	

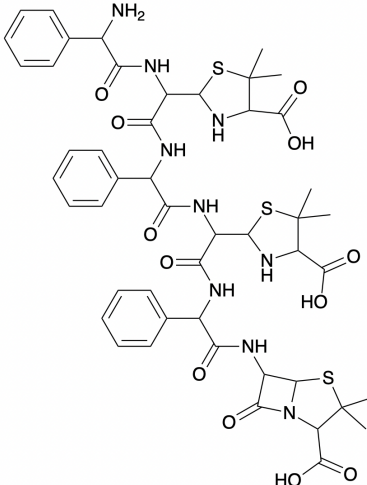
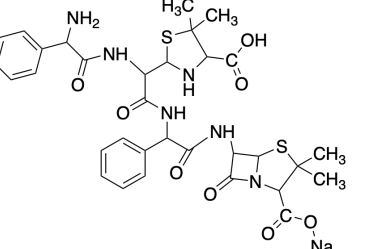
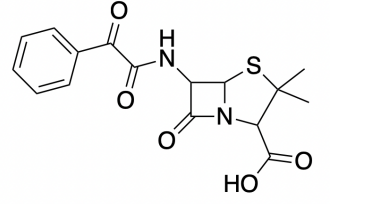
5	1048.3368 (Closed-cycle trimer, 2 open rings)	524.8 [M+H] ²⁺ 889.29 loss of thiazolidine ring	 <p>The image shows three chemical structures. The top structure is a closed-cycle trimer of ampicillin, consisting of three ampicillin units linked together in a cyclic arrangement. The middle structure is a dimer of ampicillin, with two units linked together. The bottom structure is a single ampicillin molecule, showing the characteristic fused beta-lactam and thiazolidine rings, a phenylacetamido group, and a side chain with a methyl group and a carboxylic acid group.</p>
6	721.2091 (sodiated closed cycle dimer, 1 open ring)		 <p>The image shows two chemical structures. The top structure is a sodiated closed-cycle dimer of ampicillin, where two ampicillin units are linked together in a cyclic arrangement, and the sodium ion is coordinated to the carboxylate group of one of the units. The bottom structure is a single ampicillin molecule, similar to the one in the previous row, but with a sodium ion coordinated to the carboxylate group.</p>
iAmp	Mass originally reported as 348.8. However, predicted [M+H] ⁺ is 349.0858 (ChemDraw)		 <p>The image shows the chemical structure of ampicillin, a beta-lactam antibiotic. It consists of a fused beta-lactam and thiazolidine ring system, a phenylacetamido group, and a side chain with a methyl group and a carboxylic acid group.</p>

Table 3.1 Predicted monoisotopic masses of ampicillin and related structures. Masses and structures of some of the peaks identified based on¹⁷²⁻¹⁷⁵

Below are time-of-flight (TOF) TOF-MS ESI spectra from the peaks collected by HPLC for the ampicillin samples. The first spectrum (**Figure 3.9A**) is from the earliest predominant peak (peak 1). ESI MS demonstrated that ampicillin (**compound 1 in Table 3.1**) was the most predominant species observed (**Figure 3.9A**). The spectrum also contained peaks of the isotope peaks 351.1225 and 352.1185 (**Figure 3.9 annotated with blue lettering and (i)**). An ion with a mass of 333 (m/z) was present, which is thought to be ampicillin after an ammonia loss $[M - NH_3]^+$. There was also evidence of methanolysed ampicillin¹⁷³ at 382 (m/z) (**Figure 3.9A 1_{mh}**); this species presence could be due to the methanol. There was also some evidence of oligomerised ampicillin, with an ion of 699 (m/z) evident (**compound 4 in Table 3.1**), the predicted mass of closed-cycle ampicillin dimer (**Figure 3.9A peak 4**). There were also peaks at 700 and 701 due to isotopic distribution of the molecule (**Figure 3.9A peak 4**). The peak at 1048 is indicative of a closed-cycle trimer (**compound 5 in Table 3.1**), supported by peaks at 889 and 524 (**Figure 3.9A peaks 5,**

5₍₂₎). The peak at 889 could be a fragment ion of the trimer without its thiazolidine ring, and the peak at 524 is believed to be the doubly charged trimer species (**Figure 3.9A**).

The second key peak in the ampicillin sample (**Figure 3.8 HPLC peak 2**) analysed by MS also demonstrated the presence of ampicillin (**compound 1 Table 3.1**). There was a sodiated ampicillin species 372 (**compound 3 Table 3.1**). There was evidence of closed-cycle dimers 699 (*m/z*) (**compound 4 in Table 3.1**) and its isotopic variants (**4i Figure 3.9B**). There was also evidence of closed-cycle ampicillin trimers 1048 (*m/z*) (**compound 5 in Table 3.1**) and its doubly charged species at 524 (*m/z*) (**compound 5 in Table 3.1**); also present was an ion at 889, thought to be a closed cycle dimer with a loss of thiazolidine (**Table 3.1**). The second HPLC peak appears to have more evidence of oligomerised ampicillin and other degradation products.

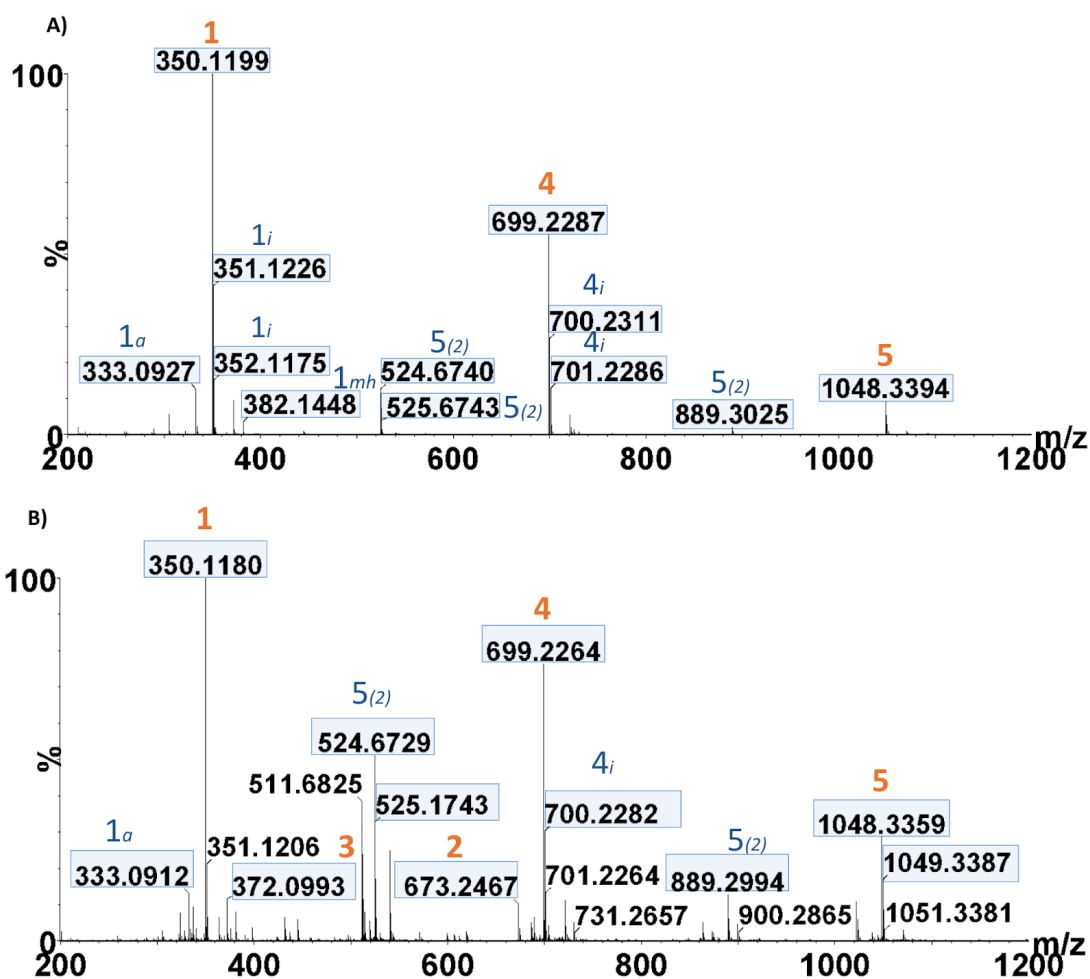


Figure 3.9A-B ESI TOF MS+ of predominant ampicillin peaks from predominant two peaks of ampicillin as detected in HPLC separation. Key peaks labelled with an orange number which corresponds to a structure in Table 3.1. The blue numbers represent either an isotopic distribution of the main peak (n_i) or a daughter ion (2), or in the case of compound 1 (Table 3.1, ampicillin) [ampicillin - NH_3] (a) and a methanolysed species (mh). The first spectrum (A.) has a predominant peak at 350.11 (m/z) as well as smaller peaks for other polymers of ampicillin. Whereas the second peak (bottom spectrum) has more oligomerised ampicillin demonstrated by peaks at 699.2 (m/z), 1048.3 (m/z), and their daughter ions.

The early peaks (1 and 2) from HPLC (**Figure 3.8**) of the iAmp sample are presumed to contain ampicillin and oligomers of ampicillin (**Figure 3.10A**). MS analysis of the first peak in the iAmp sample showed a predominant ion at 350.11 (m/z) - presumed ampicillin (**compound 1 Table 3.1**) (**Figure 3.10A**). The second most predominant ion was 699.22 (m/z), presumed to be a closed-cycle dimer of ampicillin (**compound 4 Table 3.1**)(**Figure 3.10**). There was also evidence of closed-cycle dimers, with signals at 1048.33, 899.3, and 524.67 (m/z) (**compound 5 in Table 3.1**) (**Figure 3.10A**). There was evidence of sodiated species of ampicillin, demonstrated by a signal at 372.1 (m/z) (**compound 3, Table 3.1**). In conclusion, the first peak of iAmp to come from the HPLC contains unconverted ampicillin, and the patterns demonstrated, in a large part, are similar to that of the ampicillin sample (**Figure 3.10A**).

Peak two of the iAmp HPLC run (**Figure 3.8**) demonstrated more products of degradation and oligomerisation than that of the second peak of ampicillin or that of iAmp's first peak (**Figure 3.10B**). The predominant ion in this peak appeared to be the closed-cycle dimer of ampicillin (**compound 4 Table 3.1**) with a (m/z) of 699.22 (**Figure 3.10B**). The sodiated version of the dimer was evidenced by an ion at 721 (m/z) (**compound 6 Table 3.1**) (**Figure 3.10B**). There was evidence of other oligomers and ampicilloic acids (**compound 2 in Table 3.1**) (**Figure 3.10B**). There was still evidence of ampicillin in the second peak of the iAmp sample, as an ion on 350.11 (m/z) was seen in **Figure 3.10** (**compound 1 in Table 3.1**). The second peak in the iAmp sample had more evidence of dimers and trimers.

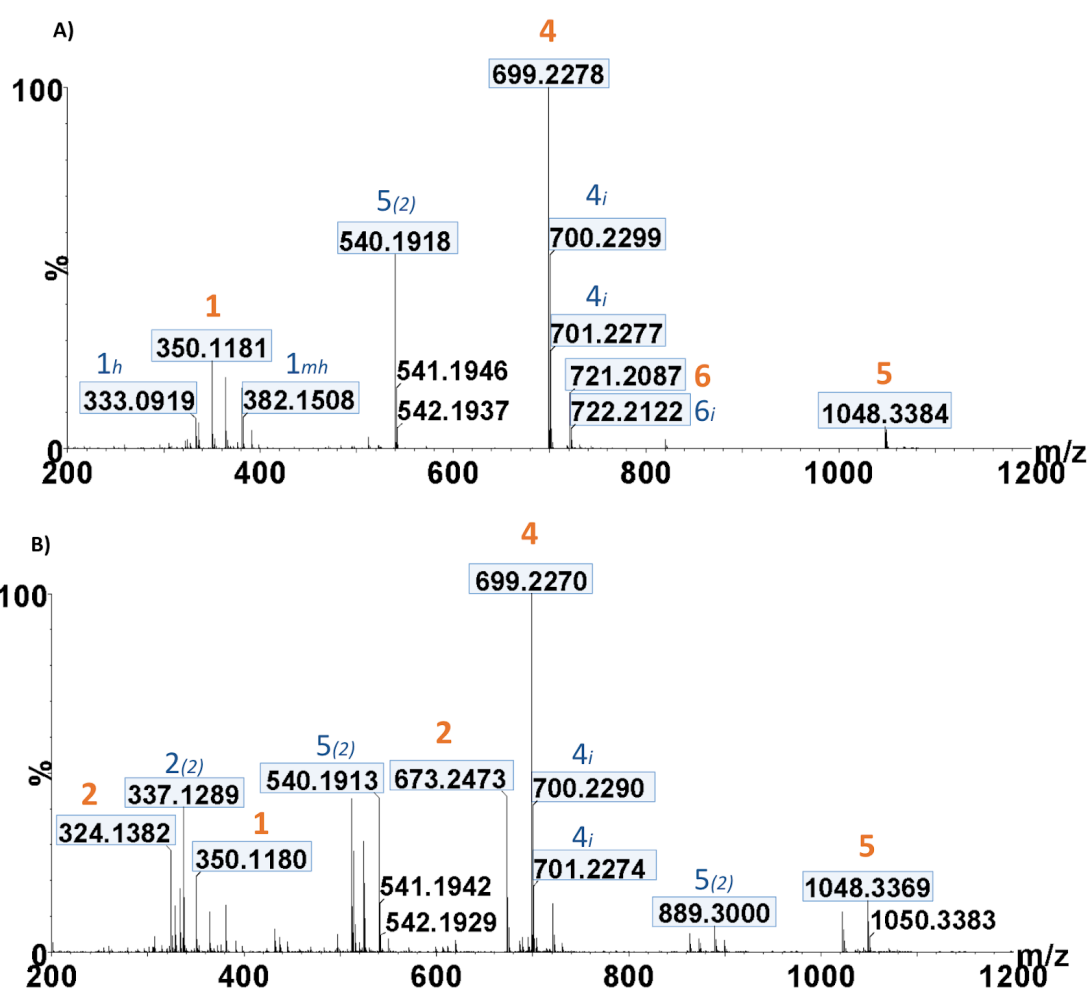


Figure 3.10A-B ESI TOF MS of iAmp fractions from peaks collected at the first (A) and second (B) peak of HPLC purification, observed at 254 nm.

As previously mentioned, iAmp elutes later in the run (**Figure 3.6, 3.7 and 3.8**); MS analysed these later peaks. The major fluorescent species were seen between 29 to 33 minutes and between 33 to 36 minutes when fluorescent detectors were used (**Figure**

3.6). The fifth peak had a predominant observed species (m/z) of 349.1675 (**Figure 3.11 below**). Peak number 4 had a predominant ion at 524.2000, thus peak 5 was assumed to contain iAmp.

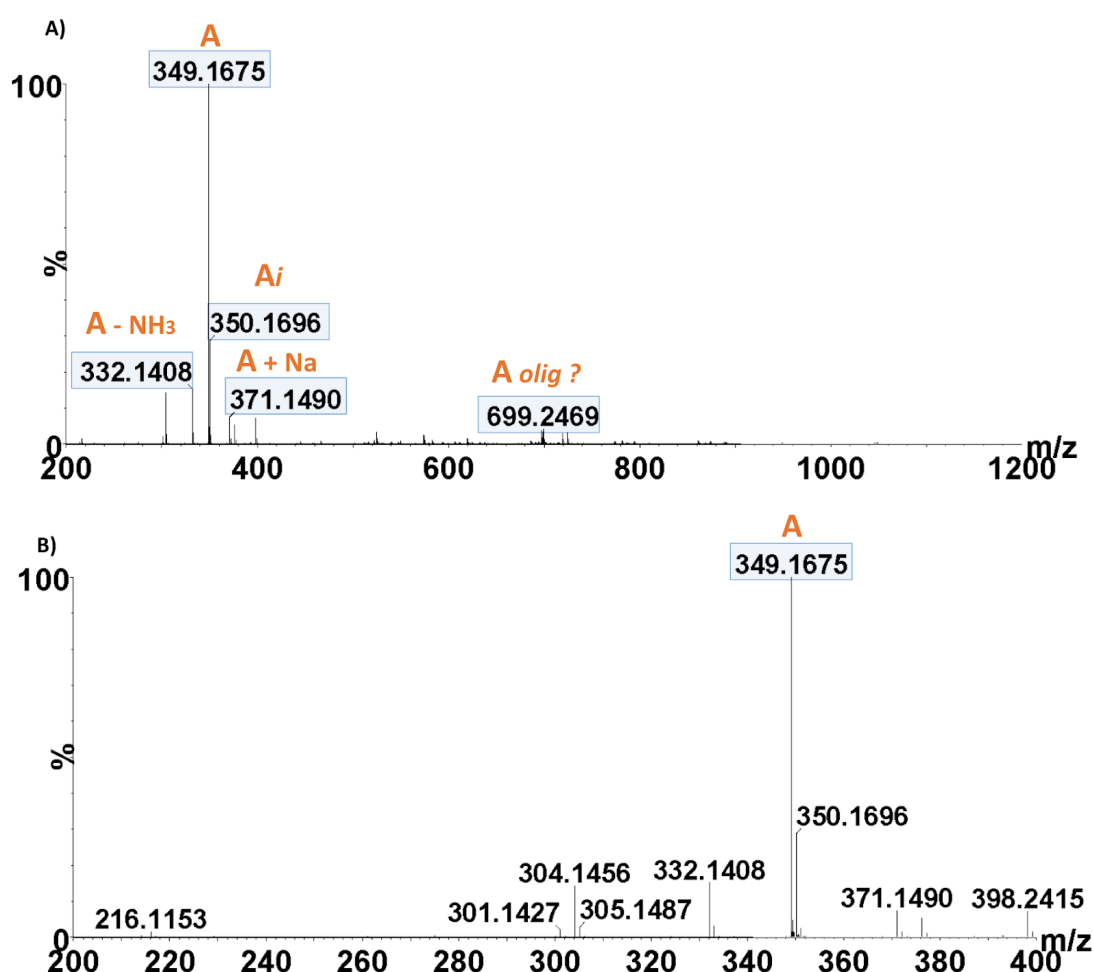


Figure 3.11 The positive mode ESI MS spectra of major fluorescent peak from HPLC separation of iAmp. Top spectrum (A) range of 200 - 1200 (m/z), and bottom spectrum (B) shows a range of 200-400 (m/z). The predominant species has a (m/z) of 349.1675 was seen - annotated with "A" in orange. There was also a peak at 350.1696 which could be an isotope of the major species (A). There was also a peak + 22 from the major species, thus presumed to be a sodiated version of the compound. There was a peak -17, potentially the molecule after an ammonia loss. There was also a peak at 699, speculatively an oligomeric version of the major species.

No species in any samples were observed with a mass of 348.8, as reported by Kotagiri *et al.*¹³⁵. The predicted mass of the compound (calculated from the molecular formula of iAmp) should be 348.0780 and 349.0858 when protonated. Here, the primary fluorescent compound had an observed protonated mass of 349.1675 (m/z) (**Figure 3.11**); this differs by 0.367 (349.1675-348.8) relative to the values reported by Kotagiri *et al.*¹³⁵. This difference could be due to different calibrations or sensitivities of the mass spectrometer.

However, the observed mass (349) (**Figure 3.11**) is aligned with other literature that suggests there is a fluorescent derivative with a mass of 349 [M+H]⁺, that demonstrates the same fluorescent properties¹⁶⁹. The sample collected in peak 5 was presumed to be iAmp.

3.2.4 Real-time analysis of iAmp synthesis

To visualise the synthesis of iAmp in real-time, quantitative Polymerase Chain Reaction (RT-qPCR) equipment (Agilent Technologies Stratagene Mx3005P) was used. The equipment can rapidly increase and sustain extreme temperatures whilst monitoring a broad fluorescence spectrum. The fluorescence was monitored at Ex. 350 nm and Em. 450 nm using water as a control. This study examined four conditions; previously synthesised iAmp (red), ampicillin degraded to iAmp (green), and amoxicillin (blue) (**Figure 3.12**). Amoxicillin has a similar structure to ampicillin but with the addition of a hydroxyl group on the phenyl ring.

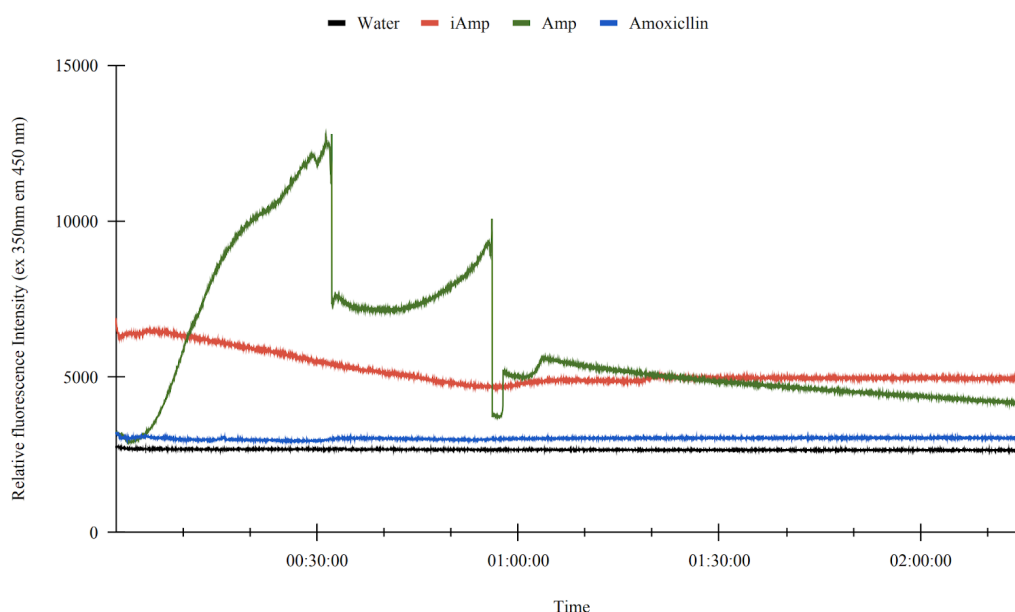


Figure 3.12 The synthesis of iAmp was followed using RT-qPCR equipment monitored at Ex./Em. 350/450 nm. In black is the water control, in blue is amoxicillin, in red is already synthesised and purified iAmp, and green is the ampicillin sample that becomes iAmp with intense heat treatment.

The Kotagiri *et al.*¹³⁵ paper states they synthesised iAmp by treating ampicillin at 99 °C for 2 hours; however, the fluorescent derivative was created within the first 30 minutes of incubation (**Figure 3.12**). In this investigation, after 30 minutes, the detector was flooded, and evaporation occurred, leading to a drop in fluorescence. This experiment

demonstrated that the original iAmp is very heat stable, maintaining fluorescence despite the high temperatures. Amoxicillin has a similar structure to ampicillin, except for the hydroxyl group on the phenyl group, which inhibits the fluorescent derivative from forming. Amoxicillin fluorescence has been observed in other research, but when salicylaldehyde was added, as opposed to intense heating¹⁷⁶. This RT-qPCR heating technique was repeated several times; however, the prolonged incubation at such high temperatures led to steam production, which unglued the adhesive film over the PCR tray leading to evaporation. Attempts were made to mitigate this using mineral oil and a heated lid, unfortunately, without success.

To visualise the iAmp creation over time, a manual approach to the above techniques was attempted. The samples were incubated at 99 °C for two hours, aliquots were taken every 10 minutes, and the end-point fluorescence was measured (excitation at 350 nm and emission at 450 nm). During this incubation, a yellow colour change was observed, and it accompanied the production of the fluorescent signal (**Figure 3.13**).

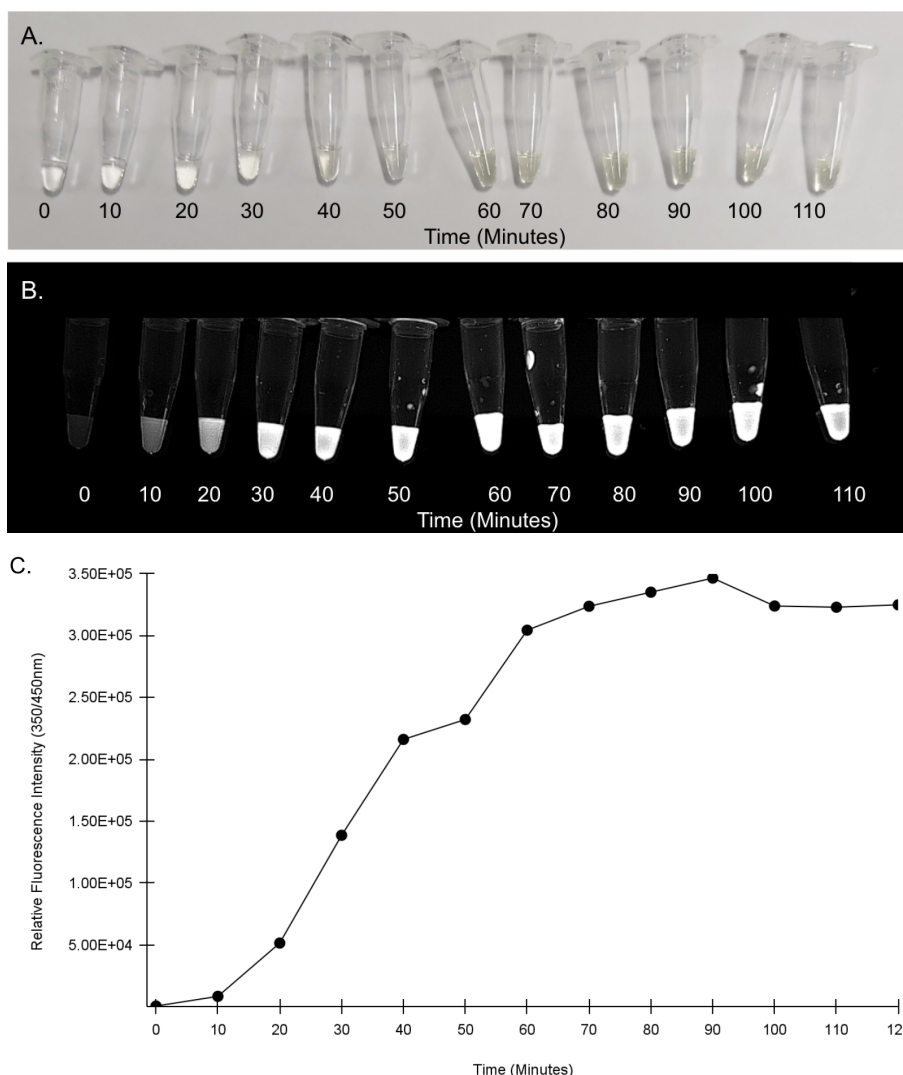


Figure 3.13A-C Time point investigation into the production of iAmp. 3.13A: a photograph of all the time points of the sample, note the yellow colour change. Below **3.13B** shows the same samples as above, but imaged on the U.V imager GelDoc. In **3.13C** Relative fluorescence intensity of iAmp samples over time.

Time-point investigation of iAmp synthesis confirmed that fluorescent derivatives were formed within the first 30 minutes (**Figure 3.13**). In this experiment, the production of iAmp increased over time when no longer hampered by evaporation effects impacting the RT-PCR experiment (**Figure 3.12**). This experiment showed that 30 to 90 minutes might be the optimum incubation period for this process. Further work could be done using HPLC, MS, and NMR to interrogate the predominant species at each time point. Furthermore, the reported iAmp could be externally synthesised, allowing us to probe the activity of the compound.

3.3 - Investigating the activity of iAmp

3.3.1 Minimal Inhibitory Concentration of iAmp

The catalytic activity of ampicillin is derived from the β -lactam ring; it is from here that PBPs are acylated and inactivated. As Kotagiri *et al.*¹³⁵ report the β -lactam ring to be intact, the compound should retain its ability to bind PBPS. The original paper¹³⁵ suggests that the changes to the amine may preclude its ability to enter cells and bind PBPs. To test this, the minimal inhibitory concentration (MICs) of iAmp and ampicillin were performed with K12 *E. coli* and with a strain of *E. coli* with a modified FhuA channel, rendering it permanently open^{33,102}. With ampicillin treatment K12 *E. coli* had an MIC of 4 $\mu\text{g}/\text{mL}$, in comparison to less than 0.5 $\mu\text{g}/\text{mL}$ in the FhuA *E. coli*. With iAmp treatment K12 *E. coli* had an MIC of 64 $\mu\text{g}/\text{mL}$ iAmp, and *E. coli* with modified FhuA channels had an MIC of iAmp of 16 $\mu\text{g}/\text{mL}$. The decrease in the MICs in iAmp between the K12 and the FhuA strain could be indicative of a permeability problem with respect to penetration of iAmp across the outer membrane of *E. coli*. However, *E. coli* that are 'open' will be more susceptible to any changes, *i.e.* osmotic changes, and thus can be considered fragile. The potential impermeability of the compound may negate its use as a permeability marker. The inactivity of iAmp needs to be better understood to ascertain if it is a helpful compound.

3.3.2 iAmp activity investigated using bocillin binding gels

Structure and activity assays of antibiotics and their targets frequently use fluorescent antibiotics^{134,163}. Bocillin is a bodipy-labelled penicillin that binds PBPs if the active site is available¹³⁴. This binding is seen as a band on an SDS-PAGE gel. Non-fluorescently labelled β -lactams, like ampicillin, have a much higher affinity for PBPs than bocillin¹³⁴. As a result, if PBPs are preincubated with β -lactams, bocillin cannot bind, so the fluorescent band on an SDS-PAGE gel is lost¹³⁴. As iAmp reportedly maintains its β -lactam ring, it should prevent the binding of bocillin. However, iAmp did not inhibit bocillin from binding; as a result, there is an observable band for PBP1b-bocillin (**Figure 3.14a**). In contrast, the ampicillin sample inhibited bocillin from binding, so a band was not observed. This result suggests that iAmp does not bind to the active site serine or that if it does, it is either not covalent or of a very low affinity. The β -lactam ring may not be intact in iAmp, most likely due to the intense heating that may have broken bonds.

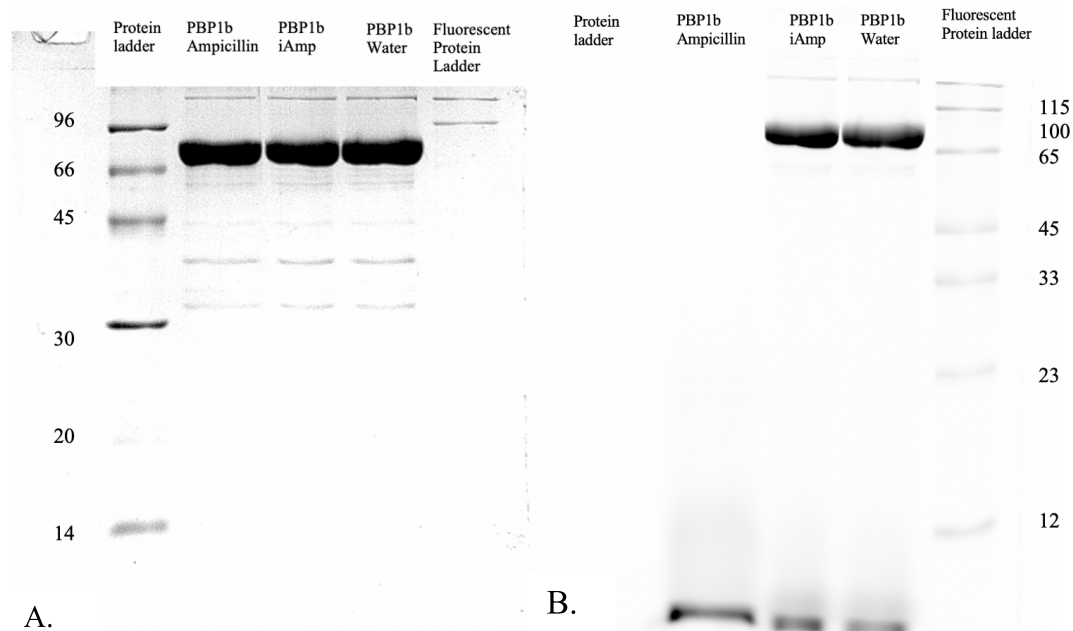


Figure 3.14A-B SDS-PAGE gels used to assess iAmp binding to PBP1b. On the right is: (A) coomassie stained protein gel, containing a protein ladder on the right (markers of 96, 66, 45, 30, 20, 14 kDa). Followed by PBP1b *E. coli* incubated with ampicillin, then a sample incubated with iAmp, followed by a sample incubated with water. Figure 3.14B is a bocillin gel of PBP1b *E. coli* incubated with either water; 200 μ M iAmp; or 200 μ M Ampicillin. Gel was visualised at Ex. 488 nm and Em. 530 nm. Strong fluorescent bands are visualised in water and iAmp conditions at 90 kDa, which corresponds to the size of *E. coli* PBP1b.

3.3.3 D-ala release in DacB carboxypeptidation with iAmp and ampicillin

To further understand if iAmp has lost its β -lactam functionality, a PBP4 (DacB) activity (carboxypeptidation) assay was used. The bocillin gel above (**Figure 3.14**) is indicative of a loss of PBP binding but not necessarily activity. DacB (PBP4) is a low molecular weight PBP that can catalyse carboxypeptidation of UDP MurNAc (DAP) pentapeptide. As this carboxypeptidation occurs, D-ala is released. D-ala release is followed through the combined activity of D-amino acid oxidase and horseradish peroxidase leading to resorufin production from amplex red, which was detected as an increase in absorbance at 555 nm^{136,177,178}. This activity would be curtailed by ampicillin (which is recognised by the PBP as a D-alanyl-D-alanine mimetic) which reacts covalently with the active site serine of the PBP, bringing about a cessation of its activity¹⁷⁸. As it has been previously shown that ampicillin has no impact on D-amino acid oxidase or horseradish peroxide (Catherwood *et al.*¹³⁶ 2020), this constituted an assay to quantify any inhibitory potency possessed by iAmp for PBP activity.

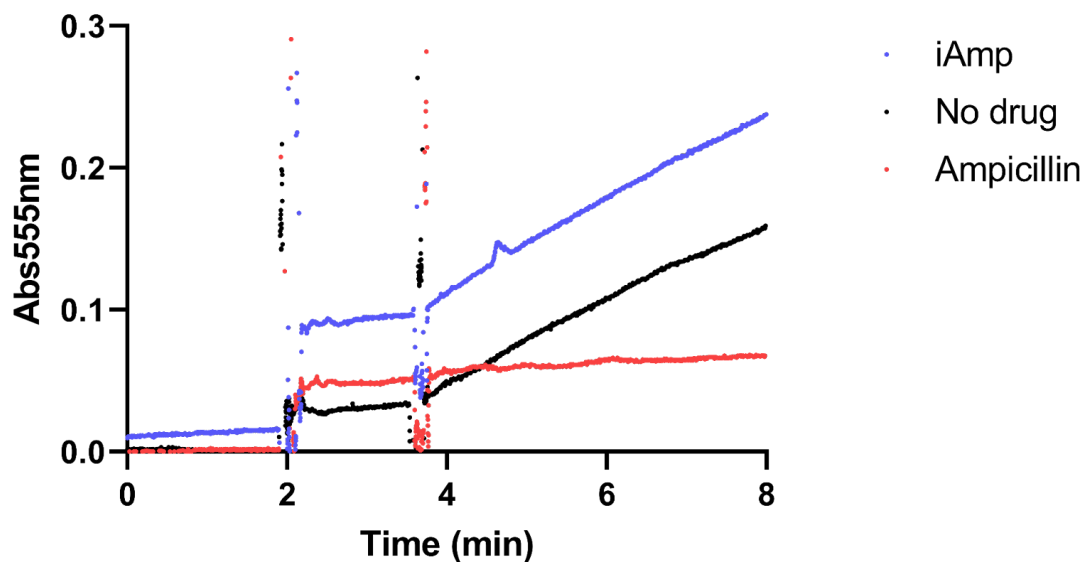


Figure 3.15 D-ala release by PBP4 D-D carboxypeptidase activity in the presence and absence of Ampicillin and iAmp. Red - in the presence of 20 μ M ampicillin. Blue - 20 μ M iAmp. Black - complete D-Ala release assay with no additional components. Ampicillin or iAmp were added at two minutes, followed by initiation of the reaction by the addition of UDP-MurNac pentapeptide at around 4 minutes (large spikes indicate the addition of compound at 2 minutes and UDP-MurNac at 4 minutes). Comparison in progress curves for blue (iAmp) and black (no inhibitor) would indicate that iAmp had no impact on the reaction and unlike ampicillin (red) did not inhibit PBP4 (DacB).

Figure 3.15 shows that, unlike ampicillin, iAmp failed to inhibit DacB activity as monitored by D-ala released on carboxypeptidation. Ampicillin caused complete inhibition of PBPs (**Figure 3.15**), leading to no product being produced, and as such, an almost flat line was observed. iAmp failed to inhibit the D-ala release assay (**Figure 3.15**), meaning it probably could not bind to the PBP4 protein. This observation could be a result of a hydrolysed β -lactam ring, consistent with the synthesis of this compound requiring intense heat or loss of other interactions with the ampicillin amino group lost on oxidative deamination of ampicillin during the formation of iAmp.

3.3.4 Imaging iAmp in *E. coli*

An important use of fluorescent antibiotics is in imaging studies; for example, bocillin can be used to stain *E. coli* cells (**Figure 3.16B**). If iAmp maintained its β -lactam ring or some semblance of a penicillin structure, it should function similarly to bocillin. To investigate

this, *E. coli* were treated with either bocillin, water or iAmp. Cells were then imaged using fluorescence microscopy (**Figure 3.16**).

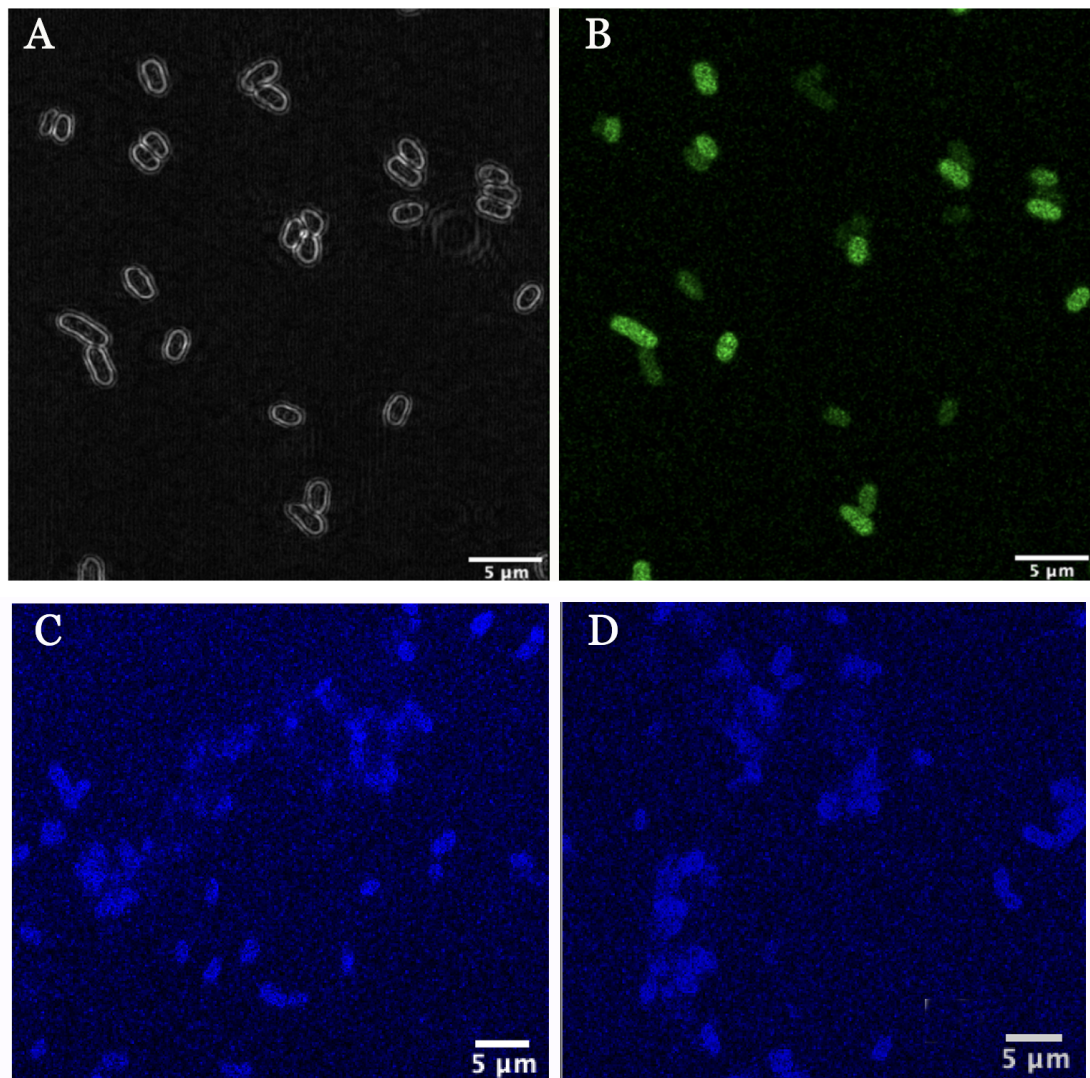


Figure 3.16A-D Staining *E. coli* with iAmp, and bocillin. 3.16A *E. coli* cells that have been treated with bocillin without fluorescence excitation, **B** is the same cross-section as **3.16A** but with fluorescence visualisation. **3.16C** is a view of *E. coli* incubated with water and excited at 350 nm. **Figure 3.16D** is *E. coli* treated with iAmp excited at 350 nm.

Bocillin can effectively stain the cells, making them visible under fluorescence microscopy (**Figures 3.16A-B**). In **Figure 3.16C** *E. coli* cells were treated with water, and in **Figure 3.16D** they have been treated with iAmp. At this wavelength there was some weak intrinsic fluorescence; the gain had to be set very high to enable any visualisation. There seemed to be no observable difference between treating with water versus treating with iAmp. Therefore, iAmp is not effective at staining cells. One possible explanation might be that intrinsic fluorescence was sufficient to obscure iAmp fluorescence.

3.4 Structural Analysis of iAmp

3.4.1.1 ESI MS of iAmp

The molecular weight of iAmp, according to Kotagiri *et al.*¹³⁵, was 348.8; however, in the mass spectrum provided by the authors in the supplemental information of the paper, the peak is annotated at 348.8 in positive mode ionisation. This suggests that the observed mass is $[M+H]^+$ and that the actual monoisotopic mass is 347.7922 Da. This chapter has calculated the predicted monoisotopic mass of iAmp (based on the mechanism they suggest) to be 348.0780 Da and 349.0858 $[M+H]^+$. Demonstrating that the authors observed masses differ from the predicted. These masses also differ from the mass observed in this chapter. The predominant fluorescent compound (synthesised and purified in this chapter) had a mass of 349.1675 $[M+H]$ (m/z) on TOF-ESI MS in positive ionisation mode (**Figure 3.17**).

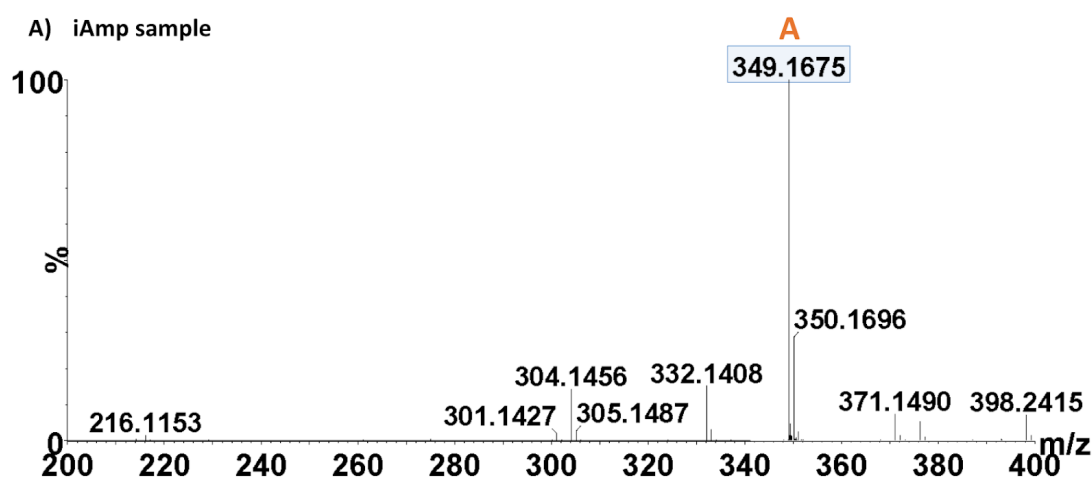


Figure 3.17 This is the observed mass spectrum for iAmp. The predominant fluorescent sample from heat degradation of ampicillin, isolated from peak 5.

In the original iAmp paper, Kotagiri *et al.* had appended the positive ion mass spectrum of ampicillin, with a peak at 349.7 reported as the molecular weight and had annotated the spectrum with 350 as $[M+H]^+$ (m/z). The monoisotopic mass of ampicillin is 349.1096 Da, and the positively charged species will be observed at 350.11742. As with the iAmp mass spectra, the masses reported by Kotagiri *et al.* are inaccurate. This could suggest that Kotagiri *et al.*¹³⁵ failed to correctly calibrate their mass spectrometer, meaning that their synthesised compound may be miss-reported in size. The mass difference between the

predicted size of ampicillin and that previously reported in¹³⁵ is 1 (m/z). Kotagiri *et al.*'s ESI ion trap MS of the compound is complex, and the software seems to have not picked up on the peak they have suggested as iAmp, as it is not a predominant peak and sits within the range that could be deemed as 'noise', whereas in Figure 3.17 iAmp is the predominant peak (**Figure 3.17**).

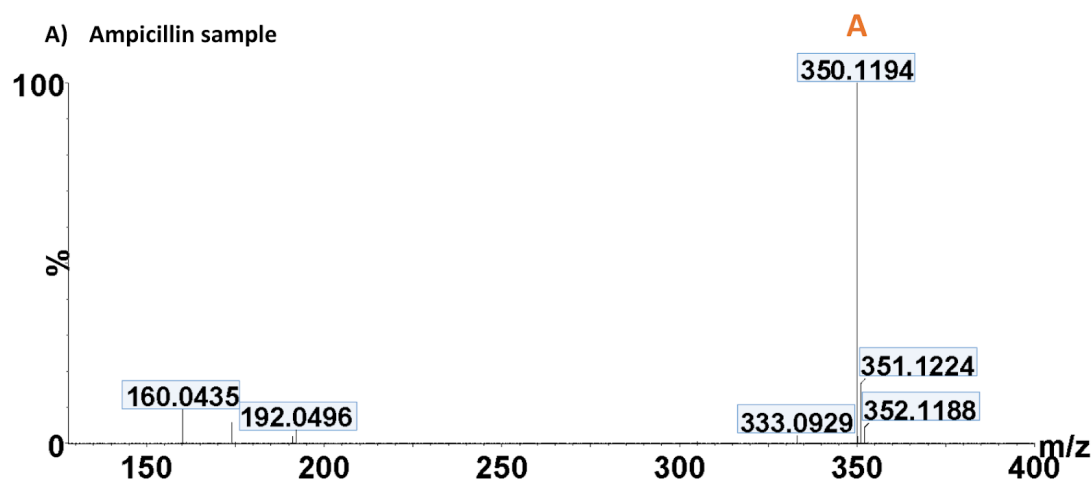


Figure 3.18 Ampicillin sample ESI+ TOF MS . The predominant species (A) has a (m/z) of 350.1194. There are also other isotopes of ampicillin present (351.1224, and 352.1188). There is also an ampicillin ion that has lost an ammonia ion (333.0929).

Thus far, this chapter has demonstrated that iAmp forgoes β -lactam functionality and that chemical properties differ from those reported by Kotagiri *et al.*¹³⁵. Therefore, we believe that the structure they published is incorrect. This chapter reports an alternative structure and proposes a synthesis pathway for this new structure. From here onwards, the structure reported by Kotagiri *et al.*¹³⁵ will be known as KiAmp, and the structure reported in this chapter will be termed Compound 8.

3.4.1.2 NMR Structural Analysis

¹H NMR and ¹³C NMR was performed and the structure of the compound was solved by R. Reinbold, (Chemistry Department, Oxford (personal communication)). The formula of Compound 8 was revealed to be C₂₀H₂₀N₄O₂, with an expected m/z of consistent with the MS spectrum collected. The ¹³C NMR and ESI MS performed by R. Reinbold is available in **Appendix 9.1** and **9.2**.

3.4.2 Solved structure of Compound 8

The proposed fluorescent compound from this investigation is 2-amino-*N*-(2-(6-oxo-5-phenyl-1,6-dihydropyrazin-2-yl)ethyl)-2-phenylacetamide (**Figure 3.19**). The chemical formula is C₂₀H₂₀N₄O₂, which has an exact mass of 348.16.

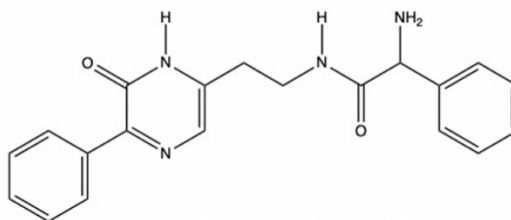


Figure 3.19 The chemical structure of Compound 8. Derived from NMR analysis, MS, and supporting literature.

A pathway accounting for the production of Compound 8 from ampicillin is presented below. This mechanism was proposed by J. Deering (personal communication, University of Warwick)(**Figure 3.21**). The new Compound 8 structure is more conjugated than that of KiAmp. This is consistent with the observation that Compound 8 absorbed more light at 254 nm than ampicillin (in later peaks). Notably, Compound 8 is missing the β -lactam ring, consistent with the observation that it can no longer inhibit PBPs. The β -lactam ring is a highly strained chemical group, and therefore one could presume that under the intense conditions (heating to 99°C), it would break. This is supported by studies that demonstrate that the hydrolysis of the β -lactam ring is observed at ambient temperatures¹⁷⁹. The breaking of the β -lactam ring is often seen as an observation in MS¹⁷², and it is a fragmentation pattern used to identify β -lactams^{172,174}. Therefore, the lysis of the β -lactam ring on the formation of Compound 8 reported here is likely. Full structural analysis of each HPLC peak and their resultant compounds is warranted to fully confirm the absence of KiAmp.

3.4.3 Proposed Mechanism of Compound 8 synthesis

Below is the proposed mechanism for the synthesis of Compound 8 from heat degraded ampicillin (**Figure 3.20**). It is based on similar mechanisms that have been proposed for other degradative pathways¹⁷⁴.

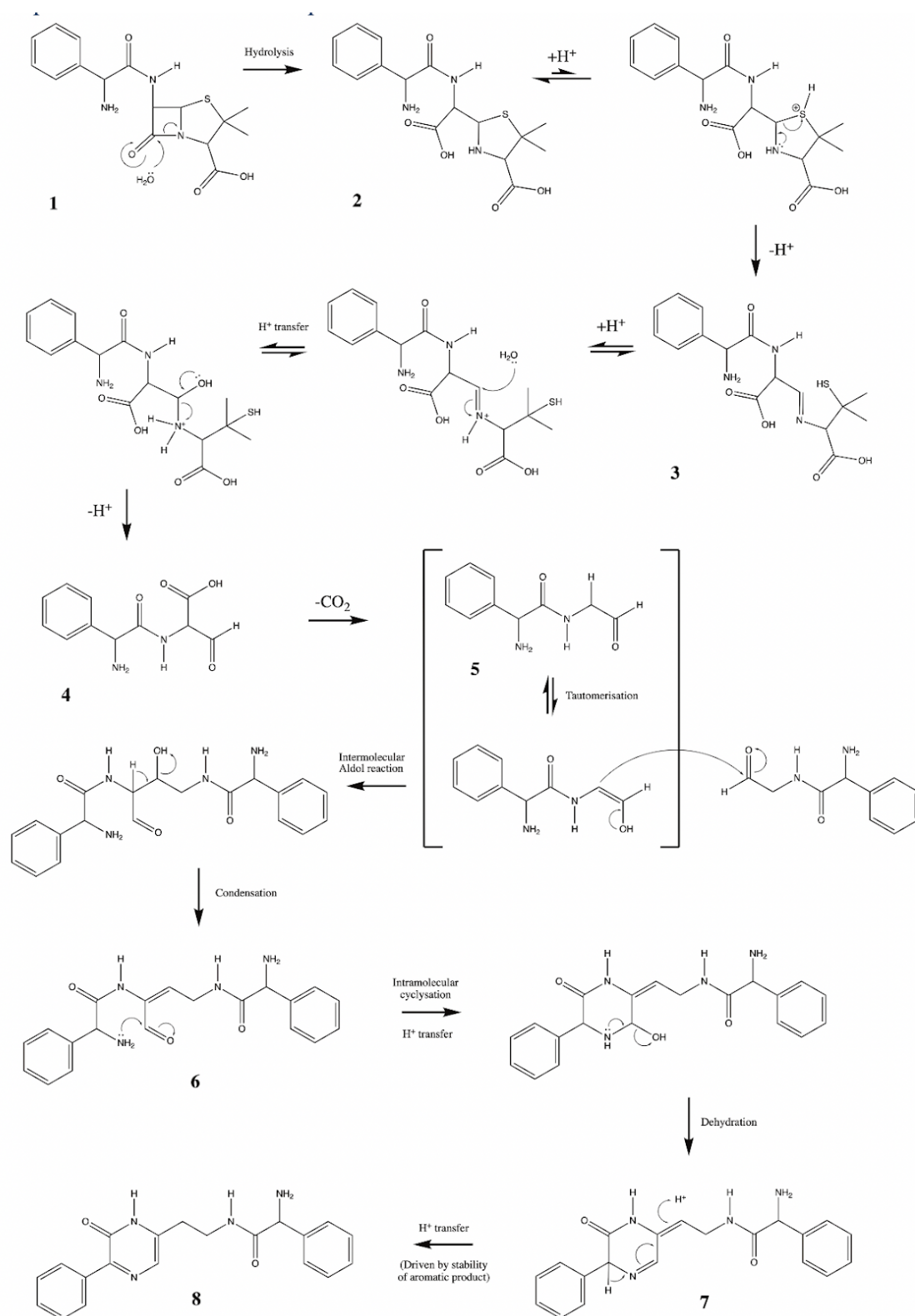


Figure 3.20 Proposed mechanism for the formation of Compound 8 (8) from ampicillin (1). First, the β -lactam ring is hydrolysed to give 2. Schiff base formation drives the opening of the thiazolidine ring to give 3 which is hydrolysed to give the aldehyde 4. The carboxylic acid then undergoes decarboxylation giving 5 and CO_2 . Two molecules of 5 then react via Aldol condensation to give 6. An intramolecular cyclisation followed by Schiff base formation and dehydration gives 7 which undergoes proton transfer giving the highly conjugated compound 8 (8).

Importantly, this proposed mechanism (**Figure 3.20**) has the hydrolysis of the β -lactam ring as the initial step in the production of compound 8. Biological assays in this chapter demonstrated that the β -lactam ring was likely hydrolysed. To further establish and support this proposed mechanism, ampicillin could be treated with β -lactamases to hydrolyse the ring and see if compound 8 is formed when heat-treated. Alternatively, ampicillin could have been dissolved in an alternative, anhydrous solvent to investigate the acid water hydrolysis of the ring. Finally, this proposed mechanism implies that the production of compound 8 requires 2 mols of ampicillin; this could be quantified by HPLC.

Richter *et al.*⁸⁹ identified some key rules that predict compound accumulation in Gram-negative cells (eNTRy-way). Richter *et al.*⁸⁹ proposed that primary amines are indispensable in facilitating compound accumulation. The authors also suggested that rigid compounds (they define rigidity as less than 5 rotatable bonds) are better at accumulating within bacteria. They also stated globularity is another key determinant in this process, categorising globularity of less than 0.25 as ideal, meaning low globularity is important for permeability.

Attribute	KiAmp	Compound 8	Ampicillin
Number of amide bonds	2	1	2
Number of aromatic rings	1	3	1
Number of rotatable bonds	4	6	4
Hydrogen bond acceptors	5	4	5
Hydrogen bond donors	2	3	3
Number of primary amines	0	1	1
Globularity	0.045	0.023	0.102
Plane of Best Fit (PBF)	0.738	0.63	1.091
LogP	0.5010	1.7956	0.3181
Topological polar surface area (TPSA)	103.78	100.87	112.73

Table 3.2 Permeability properties of KiAmp, Compound 8 and ampicillin. Comparing the permeability properties of KiAmp, Compound 8, and ampicillin. RDKit library was used to calculate the values in this table, apart from globularity, this was calculated using the Richter *et al.* eNTRy-way calculator¹⁸⁰. PBF is used to characterise the 3D structure of the model, it is based on work by Firth, Brown and Blagg 2012¹⁸¹.

Ampicillin satisfies all of the eNTRY-way rules; it has low globularity, is rigid (4 rotatable bonds) and has a primary amine (**Table 3.2**). Ampicillin has a low LogP, suggesting it is hydrophilic, supporting its passage through porins. KiAmp also has low globularity and is rigid. KiAmp and ampicillin have the same number of rotatable bonds (**Figure 3.21** of rotatable bonds). However, KiAmp is predicted to lose its primary amine in the oxidative deamination that forms it. Richter *et al.*⁸⁹ stated that the non-sterically encumbered amine is the most important determinant of accumulation⁸⁹. The lack of the primary amine caused Kotagiri *et al.*¹³⁵ to believe that the compound is an ineffective antimicrobial¹³⁵. Compound 8 has the lowest globularity score of the three compounds and has a primary amine. It is, however, more flexible. Compound 8 has a much higher LogP value implying it is more lipophilic, which was demonstrated by HPLC (**Figure 3.6, 3.7, 3.8**).

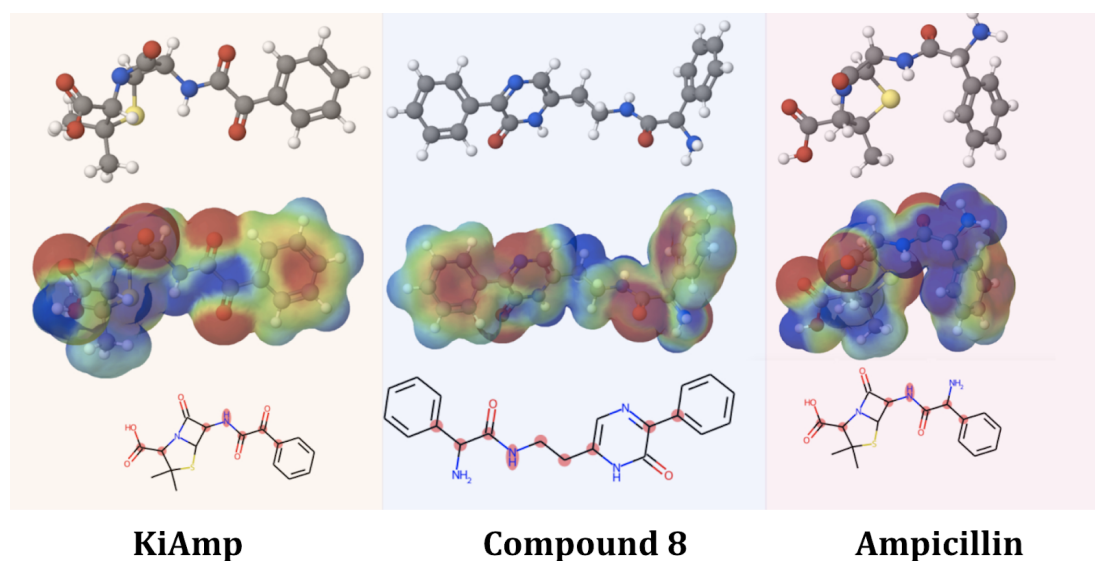


Figure 3.21 Comparing the structures of; KiAmp, Compound 8, ampicillin. Across the top row are 3D stick and ball interpretations of each molecule. The second row has an overlaid space filling model. The bottom row has the rotatable bonds of the molecule highlighted in red circles.

It is important to note that the Richter *et al.*⁸⁹ study was on non- β -lactam compounds⁸⁹. β -lactam compounds were used as a non-accumulating control because they bind covalently to penicillin-binding proteins, meaning they did not accumulate the same way as other compounds⁸⁹. The chemical predictors of this study might differ from what is seen *in vivo* as β -lactams do accumulate in Gram-negative bacteria and are one of the most commonly prescribed antibiotics for Gram-negative infections.

3.5.1 Modelling the binding of Compound 8 and ampicillin in porins

Interactions of β -lactams such as ampicillin with porins have been investigated by techniques such as X-ray crystallography and molecular dynamics studies^{57,61,182}. It is believed that fundamental interactions occur during passage through the porin, where β -lactams make contact with different residues in the constriction zone of the channel. Many of the interactions come from the third position carboxylic group of β -lactams. The compound proposed by Kotagiri *et al.*¹³⁵ would cause some changes to the interactions, the ketone group could modify hydrogen bonding interactions along the channel, but because they suggest that the β lactam ring is still intact and that the carboxylic acid is still present, this would perhaps conserve some of the interactions. Some important residues in the pore – concerning ampicillin, in particular, are four basic residues: R42, R82, R132 and K16¹⁶². Compound 8 lacks this carboxylic acid and therefore may not interact with arginines or lysines in the same way. However, this would require further investigation by X-ray crystallography, molecular dynamics or docking studies, all of which is not necessary given the resolved structure of Compound 8.

3.6 Future work

Previously fluorescent antibiotics have been used to quantify intracellular permeability, usually using fluorescent imaging to observe the process⁸⁶ and flow cytometry to quantify accumulation¹⁸³. Stone *et al.*⁸⁶ made fluorescent fluoroquinolone probes and investigated their accumulation via fluorescence microscopy. They noted that antimicrobial efficacy was lower for their fluorescent compounds and that controlling efflux was critical in observing accumulation. As mentioned previously, auto-fluorescent antibiotic accumulation has been quantified via microfluidic approaches⁸³. If Compound 8 had retained antimicrobial activity and permeability properties more akin to ampicillin, these techniques could have been applied. However, the microfluidic techniques are specialised and would require access to the equipment and specialist training. Dumont *et al.*⁹⁷ outline a method of measuring the permeability of fleroxacin and ciprofloxacin using spectrophotometric techniques combined with mass spectrometry⁹⁷.

If Compound 8 was as reported, an approach similar to Dumont *et al.*⁹⁷ could have been established, which would have allowed an investigation into permeability, efflux, and β -lactamases. The relationship between permeability and efflux is synergistic^{33,184} thus an important consideration. As efflux should be controlled, accumulation of Compound 8 in bacteria with efflux pumps knocked out would have been studied. Alternatively, an efflux pump like phenylalanine-arginine- β -naphthylamide could have been added to block efflux.

Again, if Compound 8 functioned as a β -lactam, then further characterisation of permeability could have been achieved by using strains that have altered permeability profiles. Zgurskaya *et al.*¹⁰² FhuA strains or strains with porin mutations, or strains with porins knock out, all provide an insight into permeability. Compound 8 could have been added to each of these strains and accumulation quantified by spectrophotometry or MS to build a better picture of β -lactam permeability. To assess lipid-mediated uptake of β -lactams strains with compromised lipid layers could also have been used. The lipid layer is adaptable to the environment, and its fluidity changes in response to stressors¹⁸⁵. It has been shown that *E. coli* grown at low temperatures become susceptible to vancomycin¹⁸⁵. These conditions could have allowed the exploration of the impact of these critical factors on β -lactam permeability using Compound 8 as a fluorescent analogue of ampicillin.

However, the contributions of porins, lipid changes, efflux, and β -lactamases on β -lactams could still be explored by other methods. Rosselet *et al.*⁷⁰ established a method to measure permeability based on starch-iodine reaction monitoring β -lactamase activity⁷⁰. Similarly, Ross *et al.*¹⁸⁶ established a β -lactamase assay using the hydrolysis of cephalosporins using UV detection¹⁸⁶. Hydrolysis of cephalosporins such as ceftazidime has been a popular method for measuring permeability^{84,187}. This technique could be used to investigate the initial aims of Compound 8 work.

Furthermore, label-free methods such as MS would be beneficial in monitoring β -lactam permeability. The hydrolysis of β -lactams is used to identify β -lactamase genes in clinical isolates. The use of MS as a label-free method to measure permeability will be discussed in the following chapter.

3.7 Conclusions

Compound 8 has lost the β -lactam ring, which is essential for interactions with PBPs. The structure of Compound 8 was misreported. Therefore, it is not a good indicator of permeability. It is also unaffected by β -lactamases, which are essential to the *in vivo* environment. Modelling of Compound 8 through OmpF could determine if Compound 8 can be used to measure the ingress of this small molecule through OmpF. Molecular dynamics or other modelling techniques would probably demonstrate the lipophilic qualities of Compound 8 compared to ampicillin. Ultimately, an alternative to this compound should be used to achieve the initial aims of this study.

Chapter 4. Monitoring β -Lactam acylation using mass spectrometry techniques

4.1 Background

4.1.1 Penicillin-Binding Proteins

Penicillin-binding proteins (PBPs) are essential for producing and maintaining peptidoglycan, which is a major constituent in the cell wall. There are three classes (A, B, C) that group PBPs by their function¹⁸⁸. Class A PBPs contain bifunctional N-terminal glycosyltransferase and C-terminal D,D-transpeptidase domains¹⁸⁸. Class B PBPs are monofunctional D,D-transpeptidases and Class C PBPs include; monofunctional carboxypeptidases, endopeptidases and L,D-transpeptidases¹⁸⁸. PBPs are conserved across bacterial species. Three specific motifs involving the transpeptidation/carboxypeptidation are well conserved: SxxK, SxN, and KTG¹⁸⁹. The serine of the SxxK motif provides the active site nucleophile that is central to catalytic activity¹⁸⁹. The other two motifs are near to the active site and are thought to potentially play a role in supporting the substrate binding and activation¹⁹⁰.

PBPs polymerise lipid II, which results in a supportive mesh around the cell, providing structural integrity, shape and protection of the cytoplasmic membrane from physical (for example, osmotic) stresses¹⁹¹. β -lactams target PBPs¹⁹². The β -lactam ring mimics the transpeptidation substrate, which prevents the natural substrate from binding, as the β -lactam ring is subject to nucleophilic attack by the catalytic serine in the active site of PBPs^{190,192}. This leads to the opening of the β -lactam ring and the formation of a covalent β -lactam adduct (**Figure 4.1**)¹⁹³. The covalent interaction between the β -lactam ring and the serine in the SxxK of the PBP is specific¹⁸⁹. The inability to catalyse transpeptidation compromises peptidoglycan synthesis, which leads to a weaker cell wall and eventually lysis¹⁹².

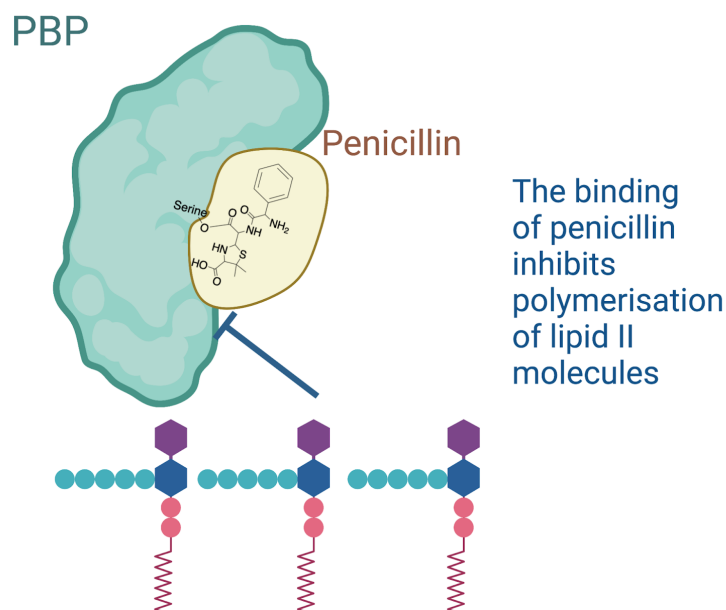


Figure 4.1 Penicillin covalently acylating PBP inhibiting the polymerisation of lipid II. Antibiotic adducts inhibit the polymerisation of lipid II, preventing the creation and preservation of peptidoglycan, weakening the cell wall. **Figure 4.1** made using Biorender.

The covalent interaction between PBPs and β -lactams is stable enough to be monitored using mass spectrometry (MS)¹⁹⁴. Mass spectrometry is an analytical technique that uses ionised atoms to establish the mass-to-charge ratio of a compound and possibly its structure and quantity¹⁹⁵ (**Figure 4.2 mass spectrometry workflow**).

4.1.2 Mass Spectrometry

MS works by ionising molecules and analysing the resultant masses. Molecules can be ionised by methods such as electrospray ionisation or MALDI (**Figure 4.2**)¹⁹⁵. The ionisation can cause a compound to fragment in a characteristic way¹⁹⁵. For example, when looking at β -lactams, the breaking of the β -lactam ring is a characteristic fragment^{196,197}. The charged compounds and their accompanying fragments are separated *via* their mass to charge ratios¹⁹⁸. This process takes place in a vacuum to minimise collision-induced fragmentation¹⁹⁵. The final step of MS involves measuring the current of the separated ions; this is displayed as a graph (mass spectrum), where m/z is along the X-axis, and relative abundance is along the y axis¹⁹⁸.

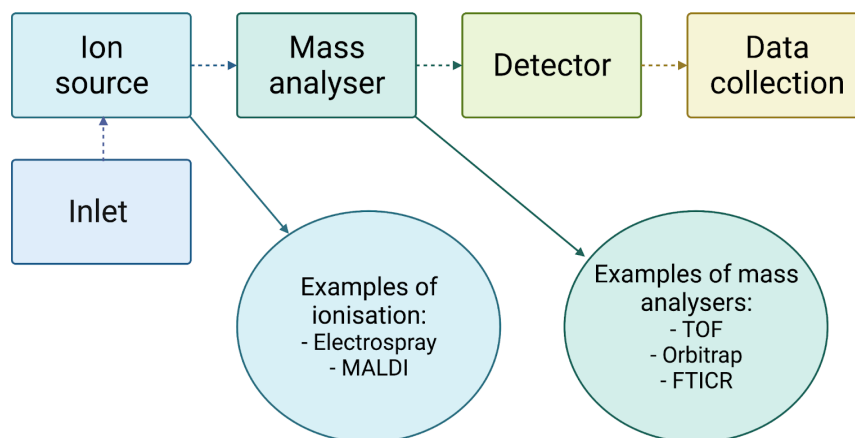


Figure 4.2 The workflow of mass spectrometry. The inlet transfers the sample to ionisation source, for example via needle. The ion source is responsible for ionising samples into gas-phase. The mass analyser is the means of separating the ions. The detector measures the current of ions and amplifies this signal. The data collection step is the computation front-end of the system showing the raw data, which can subsequently be manipulated and interpreted via correct software. Figure adapted from Chabliiss Fundamentals of Contemporary Mass Spectrometry¹⁹⁵.

4.1.3 Ionisation

Ionisation can be performed in negative mode, *i.e.*, on molecules that have gained a negative charge through, or positive mode on molecules that have gained a positive charge¹⁹⁹. Choosing negative or positive mode ionisation depends on the molecule being analysed. For example, β -lactams have amine groups suitable for positive ionisation, but they also contain a carboxylic acid that can be analysed negatively.

As alluded to in **Figure 4.2** there are many methods of ionisation suited for different molecules^{195,199}. The following ionisation sources were used in this work: electrospray ionisation (ESI) and matrix-assisted laser desorption ionisation (MALDI). ESI is a form of liquid phase atmospheric pressure ionisation^{195,199}. ESI requires volatile samples and samples of a smaller size. In contrast, MALDI (a solid phase ionisation technique) can be used to look at large non-volatile compounds^{195,199}.

The mass analysers in the following work were; time of flight (TOF), Fourier Transform Ion Cyclotron Resonance (FT-ICR), and an Orbitrap. TOF analysers work by separating ions based on their velocity through the tube analyser¹⁹⁵. Smaller ions travel to the

detector faster than larger ones. This type of analyser resolves information very quickly¹⁹⁵. MALDI is often coupled with a TOF analyser¹⁹⁹.

4.1.5 FT-ICR MS

In FT-ICR MS mass analysers, the ions enter a cell - also known as a Penning trap, where there are two detector plates, two excitation plates, and trapping plates²⁰⁰. The cell has a strong magnetic field^{195,201}. The ions rotate due to the magnetic force¹⁹⁵. Radiofrequency pulses generated by excitation plates amplify this rotation¹⁹⁵. When the charge is removed, the ions rotate at their cyclotron frequencies, detected by the detector plates¹⁹⁵. The observed rotation of the ions is characteristic of their m/z ratio. A convoluted frequency spectrum is generated, which *via* Fourier transform is then converted into a mass spectrum¹⁹⁵.

4.1.6 Orbitrap Mass Analyser

Like FT-ICR MS, the Orbitrap analyser uses fourier transform to deconvolute oscillation data to a mass spectrum²⁰². It forgoes the use of ICR and instead uses a thin capillary wire around a cylindrical electrode¹⁹⁵. The ions are trapped in oscillation around the electrode, and again like in FT-ICR, a radiofrequency pulse is applied¹⁹⁵. Both orbitrap analysers and FT-ICR analysers are frequently used in proteomics^{201,203}.

4.1.7 MS permeability studies

Mass spectrometry has been used to look at the permeability of antibiotics for years^{77,89,92,204}. Most mass spectrometry permeability studies focus on small compound accumulation⁸⁹. This method is unsuitable for some compounds, such as β -lactams, as they covalently bind to PBPs. When the small compound binds to a large molecule, there is a significant mass shift. The mass spectrometer may not be calibrated to this scale, and it would require optimisation, reducing the throughput of the process. To monitor covalent modifications of proteins resulting from compound permeability, proteomic techniques could be used as they regularly measure the mass of large molecules.

4.1.8 Proteomics

Proteomics is the large-scale study of the proteins in an organism by mass-spectrometry²⁰⁵. There are two approaches: top-down and bottom-up proteomics²⁰⁶. The most widely used form of proteomics is bottom-up²⁰⁷, whereby the proteins are extracted from an organism and then tryptically digested into peptides²⁰⁶. These peptide fragments are sequenced to identify proteins²⁰⁸. Top-down proteomics separates and analyses intact proteins²⁰⁹. In bottom-up proteomics the digestion of peptides increases the sample complexity, the increase of complexity coupled with the low abundance of the modified peptide makes isolating modified peptides (e.g. a phosphorylated peptide) harder²⁰⁶. This is because these factors increase the error rate in fitting the observed masses to the calculated masses. This makes identifying modified peptides in bottom-up proteomics more difficult, especially with labile modifications²⁰⁶. As β -lactams are covalently bound to the active site serine, the adduct should be relatively stable, meaning there should be a high abundance of the modified peptide (peptide containing the active-site serine with a covalently bound β -lactam), meaning bottom-up proteomics should be applicable.

This work proposes that bottom-up proteomic peptide mapping techniques could study the covalent interaction between β -lactams and PBPs. Bottom-up proteomics techniques could quantify the binding of PBPs to β -lactams in conjunction with isotopically heavy labelled β -lactams. This study begins using purified proteins and then progresses to complex cellular extracts. This chapter Aims to use resistant strains or modified strains to quantify *in vivo* permeability of β -lactams eventually.

4.1.9 Research aims

- To use MS to accurately quantify β -lactam binding to PBPs. Providing a label-free method.
- Applications of which could include, using β -lactam-PBP3 binding as a permeability readout, or to identify the affinity of PBPs for various β -lactams.

4.2 Results

4.2.1 Protein purifications

Ideally, this work would look at the naturally abundant proteins in *E. coli* and monitor the active site of PBPs and their acylation by β -lactams. To begin with, however, this work started with purified proteins. A selection of the proteins mentioned were kind gifts from other researchers in the group (*E. coli* PBP1a, PBP1b and *P. aeruginosa* and *Burkholderia pseudomallei* (*B. pseudomallei*) PBP3). The others (*E. coli* PBP3 and PBP4) were purified by IMAC chromatography, with further purification steps taken when required. **Figure 4.3** shows the purified *E. coli* PBPs used in this work.

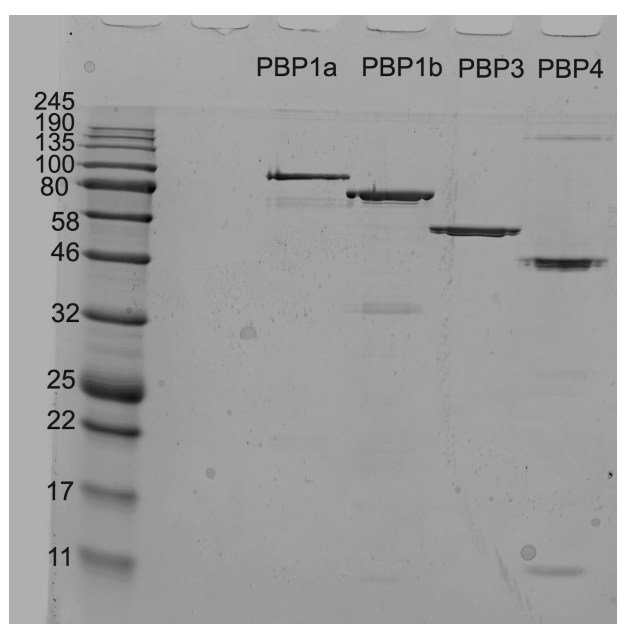


Figure 4.3- SDS-PAGE of *E. coli* PBPs used in this chapter. On the left hand side is a protein standard ladder (NEB) with marker sizes of; 245, 190, 135, 100, 80, 58, 46, 32, 25, 22, 17 and 11kDa. From left to right the lanes are; PBP1a (93.6 kDa), PBP1b (83.7 kDa), PBP3 (59 kDa), and PBP4 (52 kDa). 8 μ g of protein was loaded into each lane of this 12% SDS PAGE protein gel. PBP1a and 1b were kind gifts from Dr A. O'Reilly and Dr K. Smart.

4.2.2 Proteomic Results

The bottom-up proteomics workflow used is outlined below:

1. Proteins were purified.
2. Proteins were treated with β -lactams.

3. The resulting samples were prepared for bottom-up proteomics by either in-gel digestion or in-solution digestion following the filter-aided sample preparation (FASP) protocol.
 - a. Trypsin was added to digest proteins.
 - b. Peptides were collected.
4. Samples were analysed by LC-MS/MS using an Orbitrap mass spectrometer.
5. The raw data was searched using MaxQuant, and Scaffold visualised and processed data.

The PBPs were treated with trypsin; trypsin cleaves at the carboxy terminus of arginine and lysines²¹⁰. Consequently, the peptide fragments from the tryptic cleavage of purified proteins were predictable. To calculate the peptides, ExPasy PeptideMass²¹¹ was used. To calculate the m/z of these species and their fragments, proteomics tools by SPC were utilised²¹². MaxQuant¹³⁸ searched and assigned the raw data, and Scaffold¹³⁹ visualised the results.

4.2.3 *E. coli* PBP4 acylated with ampicillin

The first PBP tested was DacB PBP4 from *E. coli*, a Class C PBP function as D,D-carboxypeptidase²¹³. PBP4 was treated with ampicillin and then prepared according to the in-gel digest method (**Appendix 9.3A**). Below (**Table 4.1**) are the potential parent ions of the active site of PBP4. The m/z is calculated for a range of charged states and with ampicillin (monoisotopic mass of ampicillin is 349.1096).

	Main active site SXXK
Peptide sequence	VGASAP AidYHSQQMALPASTQK Oxidised methionine +16
$[M+H]^+$	2371.1711
$[M+H]^+$ with ampicillin	2720.2807
$[M+2H]^{2+}$ with ampicillin	1360.6440
$[M+3H]^{3+}$	907.4318
Drug with ampicillin	

Table 4.1 Predicted peptide mass of active site serine of *E. coli* PBP4, with and without ampicillin treatment. MaxQuant software also calculated the masses for each peptide with other modifications such as methylations and oxidations. It also calculates the masses of misscleaved peptides.

In the PBP4 ampicillin sample, only one peptide containing the ampicillin adduct was observed. This peptide was tentatively identified in its triply charged state, and it was observed only once (**Figure 4.4**). The predicted m/z of this peptide was 907.4318, and a unique peak of 907.10 m/z was observed. These results demonstrated that the technique could be used to identify antibiotic acyl adducts in PBP4 *E. coli*. However, due to the infrequent observations of the acylated peptide, modifications need to be made to the technique.

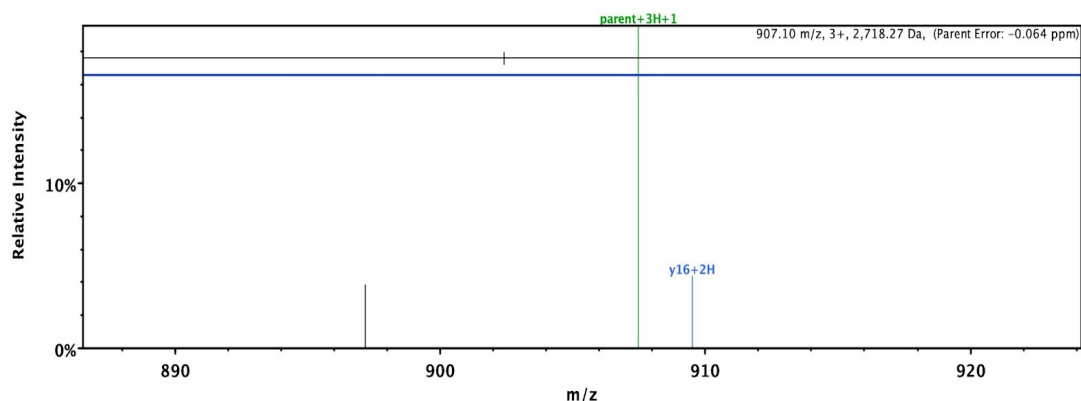


Figure 4.4 *E. coli* PBP4 treated with ampicillin. The mass spectrum of PBP4 treated with ampicillin demonstrates an ampicillin adduct at the first active-site serine (SxxK). The intensity is low, and there is some error observed.

4.2.4 *E. coli* PBP1a acylated with Ampicillin

To investigate whether this technique applies to a range of PBPs, full-length *E. coli* PBP1a was used. PBP1a has both transglycosylase and transpeptidase activity. PBP1a was treated with ampicillin, before being prepared using the in-gel digest method (**Appendix 9.3B**). However, no active site peptides containing ampicillin peptides were observed. Further research into Class A PBPs is needed to assess why this did not work, perhaps in the form of biochemical assays to monitor acylation rates¹⁶⁵. Therefore, tailoring the antibiotic to the PBP may be necessary; a range of Class A PBPs should be screened against a few different classes of β -lactams.

4.2.5 Method Development

Due to the limited success of bottom-up proteomic techniques in mapping acylated active site serines, amendments to the protocol were then investigated. Method development aimed to increase the peptide coverage and preserve the β -lactam PBP bond. PBPs and β -lactams form a covalent ester bond, which in theory should be

resistant to sample preparation. However, when monitoring the acylation of the active site serine of PBPs, some instability of the adduct was observed.

There are two approaches to preparing peptides for proteomics; in-gel preparation and in-solution preparation (peptides prepared in-solution are further processed by FASP protocol). Both techniques prepare the samples using overnight incubations with trypsin at 37°C. Bocillin gels were used to assess the stability of the β -lactam PBP bond over a range of conditions. PBPs were incubated with ampicillin for 15 minutes at ambient temperatures and then exposed to a range of temperatures for overnight incubation. The temperatures investigated are as follows; -20°C, 4°C, 21°C, 37°C, 42°C. The following day (16 hours), the samples were incubated with bocillin, as per the protocol and visualised on a bocillin and SDS-PAGE Gel¹³⁴.

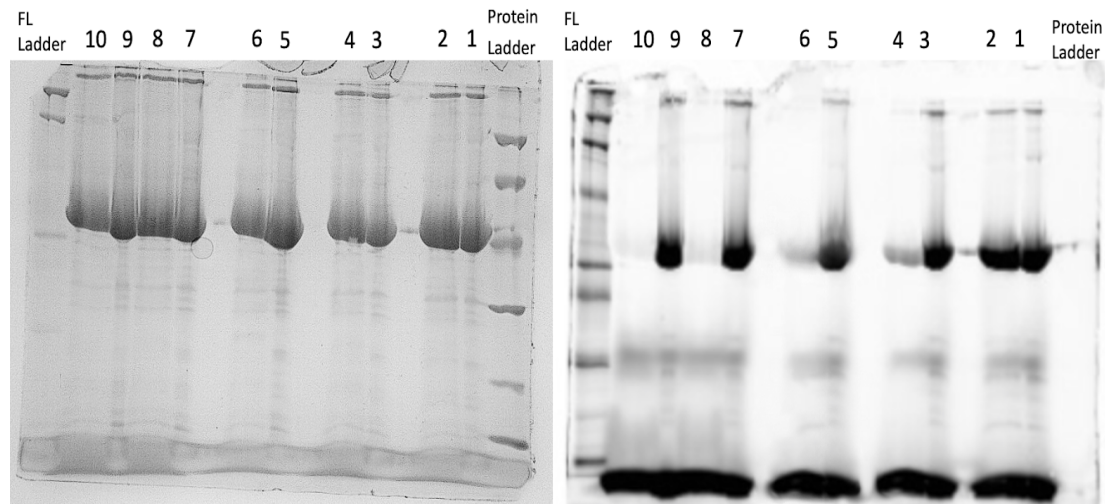


Figure 4.5 The Effect of heat degradation on PBP-antibiotic adduct stability. Protein gel on left, Bocillin gel on right; of temperature investigation. On the left of both gels is the fluorescent protein ladder (Invitrogen) which is only (clearly) visible in the bocillin gel. Fluorescent ladder (FL) has band sizes of; 158, 98, 63, 40, 32, 21, and 11 kDa. The prestained protein ladder (on the right of both images) has ladder sizes of 96, 66, 45, 30, 20, and 14 kDa. All lanes contain 16 μ g of protein. In lane 10 is the PBP4 and ampicillin incubated at -20°C then incubated with bocillin, and the control for this reaction is in lane 9 which is PBP4 incubated at -20°C then treated with Bocillin. In lane 8 is the 4°C condition, and its control in lane 7. In lane 6 is 22°C with its control in lane 5. In lane 4 is a 37°C incubation, lane 3 is the control. In lane 2 is the 42°C incubation, and the control is in lane 1.

The covalent bond was intact for the -20°C and 4°C incubation step, evidenced by the lack of band present, as the bocillin was unable to bind to the PBP (**Figure 4.5**). For the room temperature incubations, there seems to be some binding of bocillin, indicating that some of the covalent bonds were broken. Clear bocillin bands were present in the

37°C and 42°C conditions. These results suggest there are no intact PBP4-ampicillin bonds and that, therefore, prolonged incubations at high temperatures are unsuitable. As a result, samples were incubated at ambient temperatures as opposed to 37°C. To further probe the stability of the ester bond between PBP4 and ampicillin, the protein's melting temperature could have been investigated using RT-qPCR equipment.

4.2.6 PBP3s

PBP3 proteins are Class B high molecular weight PBPs; involved in cellular division¹⁹⁰. The following work used truncated mutants of PBP3 proteins lacking the transmembrane domain. Soluble PBP3s were used, as membrane proteins require detergent solubilisation, and detergents can interfere with mass spectrometry; as a result, removal of the detergent is complicated²⁰⁸.

4.2.7 Nitrocefin and PBP3s

Nitrocefin is a chromogenic cephalosporin β -lactam; it is not a clinically used antibiotic but an essential tool for investigating β -lactamase and PBP activity²¹⁴. Nitrocefin has an interesting acylation mechanism involving a second serine (Newman 2021²¹⁵). There are two proposed active sites for PBPs. First, the primary active-site serine is in the conserved SXXK sequence, and the second active-site serine implicated in nitrocefin acylation is in the conserved SXN sequence²¹⁵. The monoisotopic mass of nitrocefin is 516.04096 Da. Therefore, the aim was to identify a mass of approximately this size localised to both serines.

4.2.8 *E. coli* PBP3 acylated with nitrocefin

His-tagged *E. coli* PBP3 was purified by IMAC; it was then treated with nitrocefin for 15 minutes before being prepared by in-gel preparation (**Appendix 9.3C**). Below (**Table 4.2**) are the predicted sizes of the peptide fragment parent ion.

	Main active site SxxK	Second active site SxN
Peptide sequence	TITDVFEPGSTVKPMVMTA LQR Oxidised methionine +16	SSNVGVSK
[M+H] ⁺	2520.3200	777.41015
[M+H] ⁺ Nitrocefin	3036.3610	1293.4511
[M+2H] ²⁺ Nitrocefin	1518.6841	647.2292
[M+3H] ³⁺ Nitrocefin	1012.7919	431.8219

Table 4.2 Predicted masses of *E. coli* PBP3 active-site serines with nitrocefin adducts

The first active-site serine (**SxxK**) in *E. coli* PBP3 was acylated by nitrocefin. This acylated peptide was observed in its triply charged species. The acylated *E. coli* **SxxK** peptide had a predicted m/z of 1012.7919, and this exact mass was observed (**Table 4.2**); this species had a relative intensity of 14,943,000 (**Figure 4.6**). The active site serine in this peptide is at position 10; it would be identified in b_{10} of the peptide or y_{14} upwards. The actual peptide containing the active site serine was not identified, but other fragments of the peptide were identified.

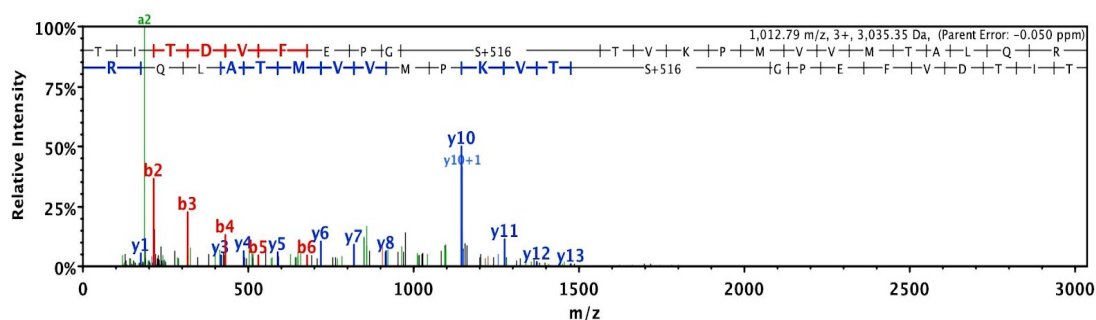


Figure 4.6 The active-site serine of *E. coli* PBP3 treated with nitrocefin. MS spectra demonstrates nitrocefin adduct on the first active-site serine (SxxK).

The spectrum (**Figure 4.6**) shows many of the associated fragment ions of the parental peptide. However, the fragment ions containing the active-site serine are not observed. The largest y ion is y_{13} ; it contains the amino acids from the C terminus up to the serine; it (y_{13}) has a monoisotopic mass of 1473.82805 Da, it has a very low intensity. The y fragment ion containing the active site serine has a mass of 2076.90103 Da, significantly higher than other observed peaks. In this preliminary investigation, there is some evidence of acylation of active site serine.

4.2.9 *Burkholderia pseudomallei* PBP3 acylated with nitrocefin

B. pseudomallei is the causative agent of melioidosis infections which have high mortality rates²¹⁶. *B. pseudomallei* is also known to have high levels of intrinsic resistance²¹⁶. Therefore, developing drugs that can treat *B. pseudomallei* is essential for antibiotic drug discovery. The truncated *B. pseudomallei* PBP3 was a gift from Dr D. Bellini. The predicted peptides of *B. pseudomallei* active sites, acylated with nitrocefin, are outlined in **Table 4.3**.

	Main active site SxxK	Second active site SxN
Peptide sequence	VLTDVFEPGSIMKPFTVSLALDLHR Oxidised methionine +16	(K)SSNIGATK(I)
[M+H] ⁺	2784.4780	776.3920
[M+H] ⁺ with nitrocefin	3300.5189	1,292.4330
[M+2H] ²⁺ with nitrocefin	1650.2594	647.2243
[M+3H] ³⁺ with nitrocefin	1101.1808	431.8211

Table 4.3 The predicted masses of *B. pseudomallei* PBP3 active site peptides acylated by nitrocefin.

No acylation of the first active-site serine (SxxK) in *B. pseudomallei* PBP3 was observed. However, nitrocefin's acylation of the second active-site serine (SxN) was observed at the predicted mass:charge ratios (**Table 4.3, Figure 4.7A**). The parent ion of doubly charged species of SxN-nitrocefin was the most abundant species in the spectrum with an m/z of 647.23; it also had a low error score. Evidence of SxN-nitrocefin was supported by broad coverage of y and b ions that contained the serine of the fragmented peptide with the nitrocefin adduct. In addition, there was also a signal from free nitrocefin 517 [M+H]⁺, further corroborating nitrocefin's interactions with the second active-site serine (SxN) in *B. pseudomallei* PBP3 (**Figure 4.7B**).

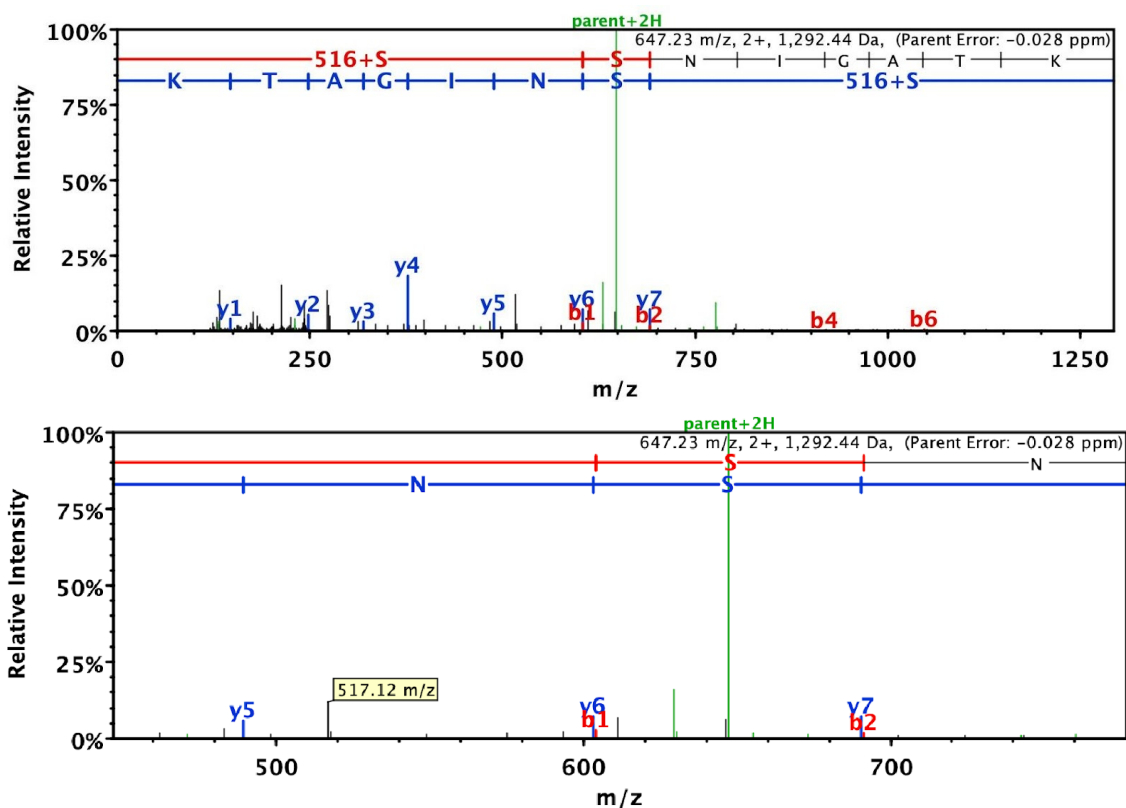


Figure 4.7 *B. pseudomallei* PBP3 treated with nitrocefim. Acylation of the second-active site serine (SxN) was observed in A. In A the full spectrum of the identified peptide is shown, with good coverage from fragment ions. In B a peak at 517 is observed, most likely due to dissociated nitrocefim.

4.2.10 *Pseudomonas aeruginosa* PBP3 acylated with a range of β -lactams

Pseudomonas species are particularly impenetrable to antibiotics and are also clinically significant, particularly for patients with cystic fibrosis, due to their rising antibiotic resistance exhibited by this pathogen²¹⁷. *P. aeruginosa* PBP3 was purified and gifted by Dr H. Newman. Below (**Table 4.4**) are the calculated masses of the active-site peptide fragments containing the conserved serines, with adducts of nitrocefim, aztreonam, and ceftazidime.

	Main active site SXXX	Second site SXN
Peptide sequence	AMIDVFEPGSTVKPFSMSAA LASGR	SSNVGISK
[M+H] ⁺ <i>m/z</i>	2569.2788	790.4257
[M with nitrocefin +H] ⁺ <i>m/z</i>	3085.3199	1306.4657
[M with nitrocefin +2H] ²⁺ <i>m/z</i>	1543.1636	654.2401
[M with nitrocefin +3H] ³⁺ <i>m/z</i>	1029.1115	435.4886
[M with aztreonam +H] ⁺ <i>m/z</i>	3004.3308	1226.4777
[M with aztreonam +2H] ²⁺ <i>m/z</i>	1502.6690	613.7425
[M with aztreonam +3H] ³⁺ <i>m/z</i>	1002.1151	409.4974
[M with ceftazidime +H] ⁺ <i>m/z</i>	3115.3780	1337.5249
[M with ceftazidime +2H] ²⁺ <i>m/z</i>	1558.1926	669.2661
[M with ceftazidime +3H] ³⁺ <i>m/z</i>	1039.1309	446.5132

Table 4.4 The predicted *m/z* ratios of *P. aeruginosa* PBP3 peptide fragments with the antibiotics; nitrocefin, aztreonam, and ceftazidime. The monoisotopic masses of nitrocefin, aztreonam and ceftazidime (M) are 516.0410, 435.0519 and 546.0991, respectively.

4.2.11 *Pseudomonas aeruginosa* PBP3 acylated with Nitrocefin

In *Pseudomonas* PBP3, acylated peptides containing the second active-site serine SxN were observed (**Figure 4.8**). The peptide containing this serine was observed in a doubly charged state at the predicted *m/z* of 654.24 (**Table 4.4**). The peptide has many supporting *y* and *b* ions, suggesting good evidence of nitrocefin localisation. This SxN fragment also contained a peak at 517.17, suggesting some unbound nitrocefin (**Figure 4.8**). It may have dissociated due to the collision energy. The preparation steps involve several ultrafiltration steps with 10 kDa membranes, so molecules with a mass less than that should have been removed.

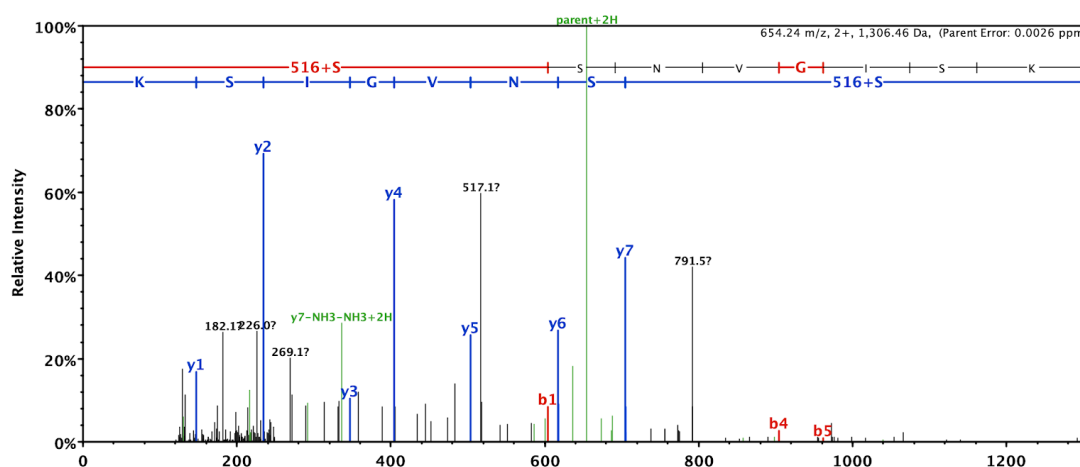


Figure 4.8 MS spectrum of *P. aeruginosa* PBP3 acylated by nitrocefin, acylation at second-active site serine SXN. There is also a peak at 517.17, the mass of unbound nitrocefin.

A non-specific acylated serine was identified in the *P. aeruginosa* PB3 nitrocefin sample. The non-specific serine was identified in this peptide: WKPSDIVDVYPGTLQIGR (**Figure 4.9**). The unexpected acylation of this serine in this peptide suggests issues in this method. A crystal structure of PBP3 in a complex with ceftriaxone is shown below (**Figure 4.10**). It demonstrates the proximity of ceftriaxone to SxN and SxxK, whereas the third serine is some distance from the active site (**Figure 4.10**).

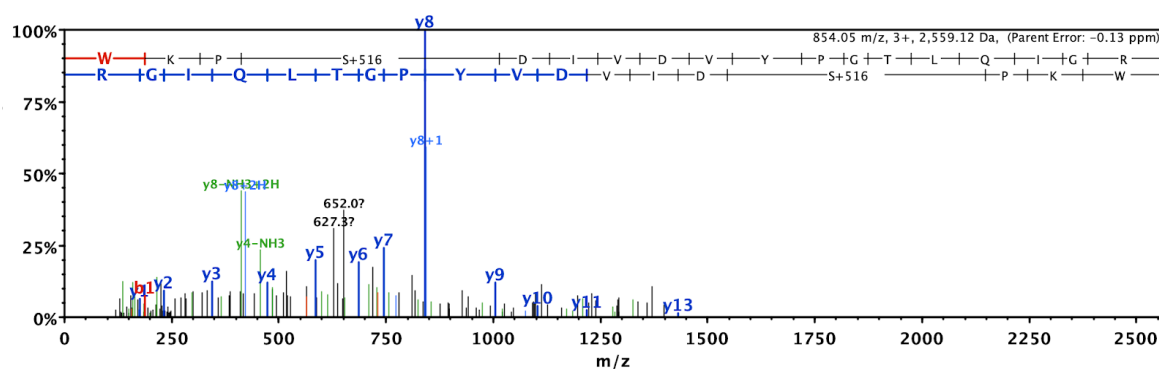


Figure 4.9 MS spectrum of *P. aeruginosa* PBP3 treated with nitrocefin, acylation observed on an unexpected serine. The y ions here cover up to y13 which is up until the serine.

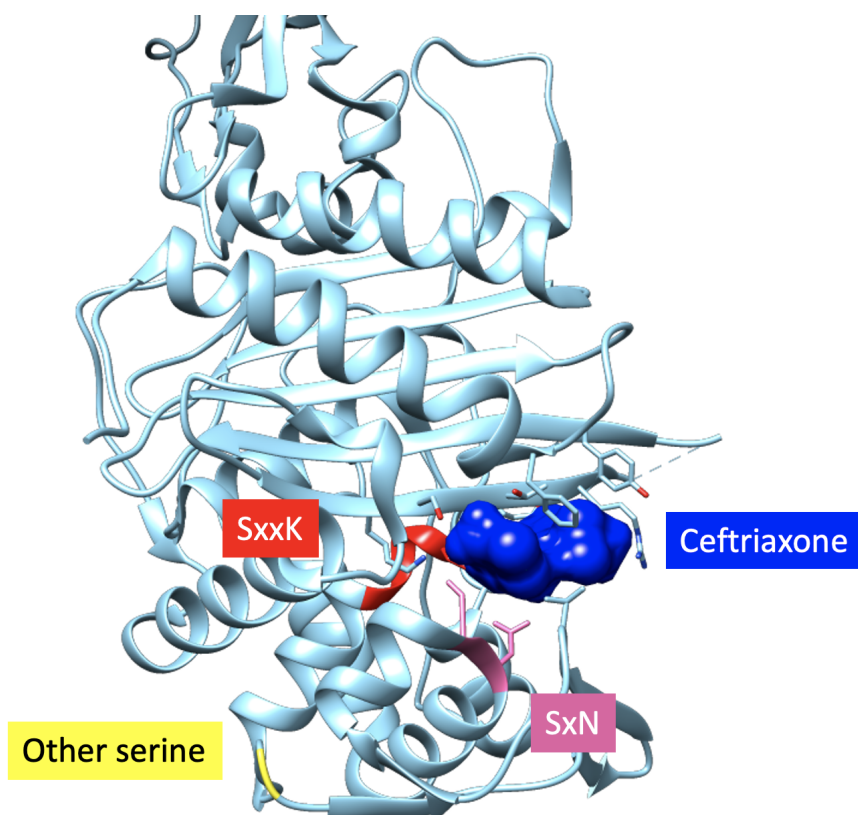


Figure 4.10 *P. aeruginosa* PBP3 in complex with ceftriaxone, with SxxK and SxN sites highlighted. Created on Chimera PDB 3OCN²¹⁸ of PAO1 PBP3 in complex with ceftriaxone. Ceftriaxone in royal blue colour space-fill representation is shown. The red portion of the alpha helix is the active site serine SxxK, it is in contact with the drug. The pink portion is the SxN second conserved serine, thought to be involved in orientating or binding the drug. The yellow region is the mis-acylated serine on a linker region, far from the active site.

4.2.12 *P. aeruginosa* PBP3 acylated with aztreonam

PBP3s from *E. coli*, *B. pseudomallei* and *P. aeruginosa* were treated with aztreonam and ceftazidime before being prepared for proteomics using the in-gel digestion method (**Appendix 9.3**). Unfortunately, aztreonam and ceftazidime adducts were only observed in *P. aeruginosa* PBP3.

Aztreonam acyl-adducts were found in the first active-site serine (**SxxK**) of PBP3 *P. aeruginosa*. The aztreonam acyl-adduct was found in the peptide: AAMRNRAMIDVFEPG**STVK**PFMSAALASGR (**Figure 4.11**). In this peptide, all three methionines were oxidised; this is due to the in-gel protocol. The peptide was observed in its triply charged state, at the predicted m/z of 1002.12 (**Table 4.4**).

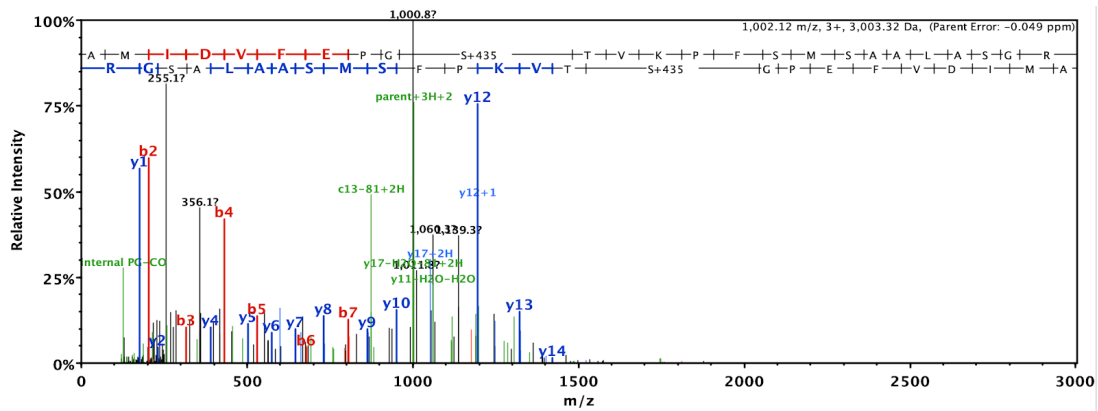


Figure 4.11 *P. aeruginosa* PBP3 treated with aztreonam. An aztreonam adduct was localised to the first active-site serine SxxK.

4.2.13 Ceftazidime PBP3 adducts *P. aeruginosa*

Ceftazidime adducts were located on the SxxK active site serine of *P. aeruginosa* PBP3 and were observable with high intensity. Many of the predicted fragments ions (*b* and *y*) were located. The calculated mass of the peptide was 3130 *m/z*, and it was found in its doubly charged form (**Figure 4.12**). Ceftazidime has a monoisotopic mass of 546.0099 Da; ceftazidime contains a 3'-methyl pyridine ring that acts as a leaving group due to the electron shift upon acylation, causing the loss of the group. The mass of ceftazidime bound to the serine should mean the acyl adduct should have a mass of 452. An adduct of this size was not identified in any peptides.

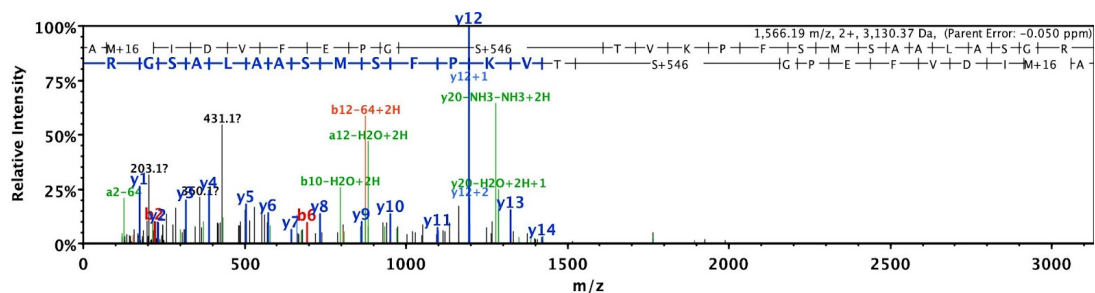


Figure 4.12 *P. aeruginosa* PBP3 treated with ceftazidime. Ceftazidime has acylated the first active-site serine (SxxK).

4.3 Acylation of *E. coli* PBPs by ampicillin

4.3.1 *E. coli* PBPs mixture treated with ampicillin

To investigate how the bottom-up proteomic technique fares when the sample complexity increases, four PBPs from *E. coli* were used; PBP1a, PBP1b, PBP3, and PBP4.

The purified proteins were mixed, treated with ampicillin and aztreonam. These proteomic samples were prepared using the in-gel method (**Appendix 9.3D**). There was no evidence of acylation with aztreonam treatments in any of the samples.

4.3.2 *E. coli* PBP1b with ampicillin

The predicted *E. coli* PBP1b peptide mass to charge ratio, when cut with trypsin, is 1946.0957 m/z +H. When treated with ampicillin a mass of +349 should be localised to the serine. The peptide sequence in PBP1b containing the active-site serine (SxxK) is; **SIGSLAKPATYLTALSQPK**. The ampicillin adduct should be on this peptide's second serine (SLAK). However, the software localised the acylation to the N-terminal serine (SIG) (**Figure 4.13**). This could be a mislocalisation by the software; this demonstrates the potential drawbacks of this technique.

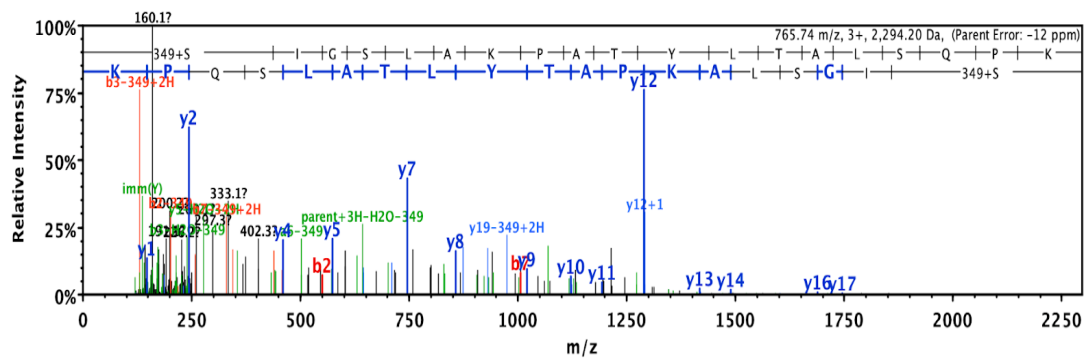


Figure 4.13 *E. coli* PBP1b treated with ampicillin. An ampicillin acylated serine is observed however it is mislocalised.

In the *E. coli* PBP1b-ampicillin sample, there was another unexpected observation of ampicillin bound to the **SxN** active-site serine (**Figure 4.14**). This was unexpected as the second active-site serine is thought to be involved in nitrocefin acylation predominantly and not inclusive of other β -lactams.

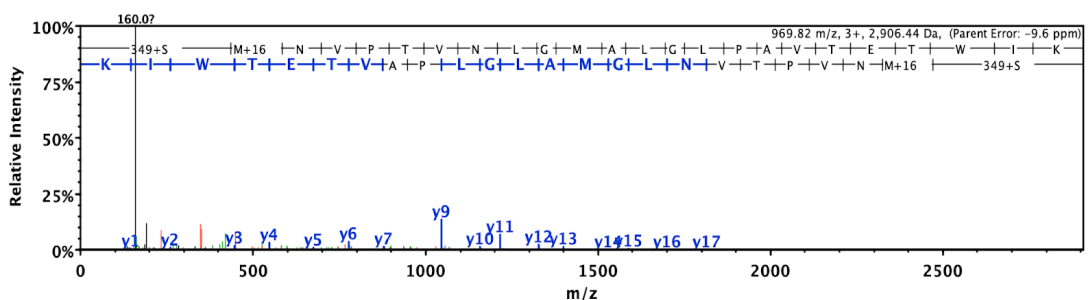


Figure 4.14 *E. coli* PBP1b treated with ampicillin, a serine with an ampicillin adduct is located in the second conserved serine of PBPs at SxN.

4.3.3 *E. coli* PBP1a treated with ampicillin

By treating *E. coli* PBP1a with ampicillin, an acylated first active-site serine (SxxK) should be observed. The peptide in PBP1a containing the first active-site serine is QVGSNIKPF~~LY~~TAAMDK. This peptide has a predicted singly charged m/z ratio of 1882.9731, which, when acylated with a molecule of ampicillin, rises to 2232.0827 m/z . This peptide was not identified using the usual Scaffold parameters, whereby the threshold for results is 95% confidence. However, when this threshold was lowered to 80%, peptides with an acylated SxxK active site serine were observed (**Figure 4.15**). As expected, there are fewer b and y ions to support the evidence of this peptide, including a lack of b and y ions that contain the active site serine.

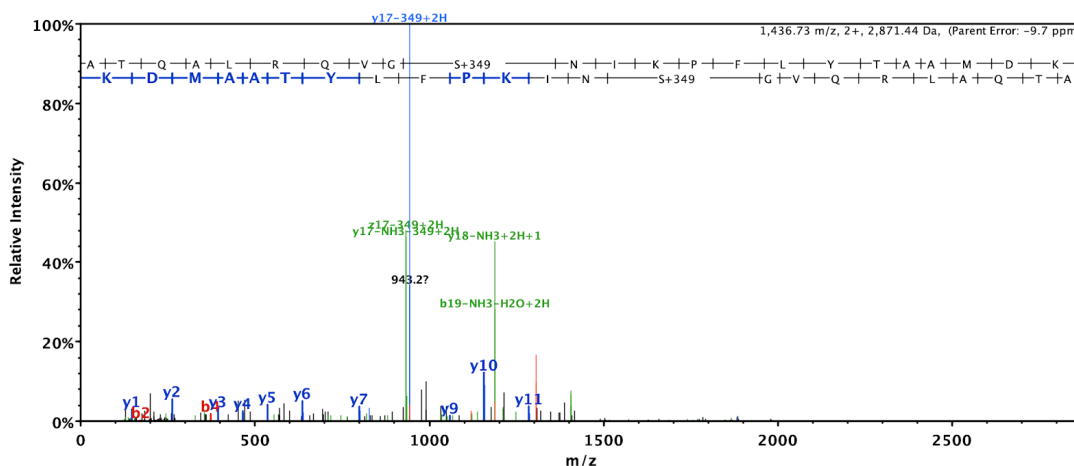


Figure 4.15 *E. coli* PBP1b treated with ampicillin, an ampicillin adduct is located on the SxxK active-site serine.

4.3.4 *E. coli* PBP4 and ampicillin

In earlier experiments using PBP4 with ampicillin, there was evidence of ampicillin acylating the active site serine (SxxK) (**Figure 4.6**). However, in this mixture of PBPs, the adduct was not located. The adduct observed earlier was of lower certainty, and the lack of an adduct in this experiment suggests the technique is unreliable. It also may not have been observed as one of the other PBPs may have sequestered the ampicillin.

4.3.5 *E. coli* PBP3 ampicillin

PBP3 from *E. coli* also did not show any ampicillin adducts. It has previously reliably shown nitrocefin adducts across multiple species. However, the *E. coli* PBP3 was also unable to demonstrate adducts with aztreonam or ceftazidime.

4.3.2 Discussion of bottom-up approach

As complexity increased in this experiment, accuracy and reliability were lost. When a mixture of proteins was treated, the acetylated adduct was lost on half of the proteins. Ampicillin adducts were seen in *E. coli* PBP1a and PBP1b, but not PBP3, PBP4. The lack of acyl-adducts in all proteins could be because either not enough antibiotic was added or because the sample is too complex for the software to interpret, and with increased searches of modifications, the error rate also increases. Alternatively, potentially one PBP had a stronger interaction with ampicillin and sequestered it, preventing other PBPs from binding to it. Ampicillin should have had the most prominent reaction to PBP4²¹⁹, so, surprisingly, it did not show any adducts. This demonstrates the need to focus on one PBP to capture the signal for permeability, preferably a soluble PBP such as 4 or 5, as it is quicker to prepare an extract from these proteins. However, PBP3 seemed to be a good target for most species tested. The original goal for this work was to work with crude lysates (not purified proteins) to investigate β -lactam binding on each PBP as a measure of permeability.

This proteomic approach provides a sensitive means of identifying antibiotic-acylated PBP peptides of interest. The control samples for each protein never demonstrated active sites acylated by β -lactams. However, occasionally non-specific acylation sites were observed. Either due to over-saturation with β -lactam causing spontaneous acylation or mislocalisation via software. These discrepancies raise issues about the efficacy of this technique. Reproducibility issues further this frequently just one modified peptide or a few modified peptides were observed, demonstrating it is not a reliable method.

As specificity was lost when complexity increased, the experiment needed to be simplified or higher specification equipment employed. Firstly, to reduce complexity,

the focus was switched to one PBP, *E. coli* PBP3. To resolve complexity, FT-ICR MS was used. FT-ICR MS has advanced resolving power and is used in top-down proteomics.

4.4 Top-Down Proteomics MS

Due to the limited reproducibility achieved by bottom-up proteomic peptide mapping for PBPs and β -lactams, a top-down approach using FT-ICR MS was attempted. FT-ICR MS can achieve up to femtomole accuracy; it is instrumental in resolving post-translational modifications^{195,201}. FT-ICR MS can employ electron capture dissociation and collision-induced energy to extract structural information^{195,208}.

For the FTI-CR MS work, *E. coli* PBP3 was treated with ceftriaxone. Ceftriaxone also has a leaving group, meaning when it binds PBPs and the β -lactam ring 3' leaving group (3-mercapto-2-methyl-1,2-dihydro-1,2,4-triazine-5,6-dione) is displaced concurrent with acylation of the PBP serine by the drug (**Figure 4.16**). The acylation of the active site serine causes a mass shift from 554 to 395. The truncated version of the protein used in this work had a predicted molecular weight 59847 Da. The table of the predicted masses is shown below (**Table 4.5**). The FT-ICR MS work was carried out and analysed by Dr Y. Lam (University of Warwick).

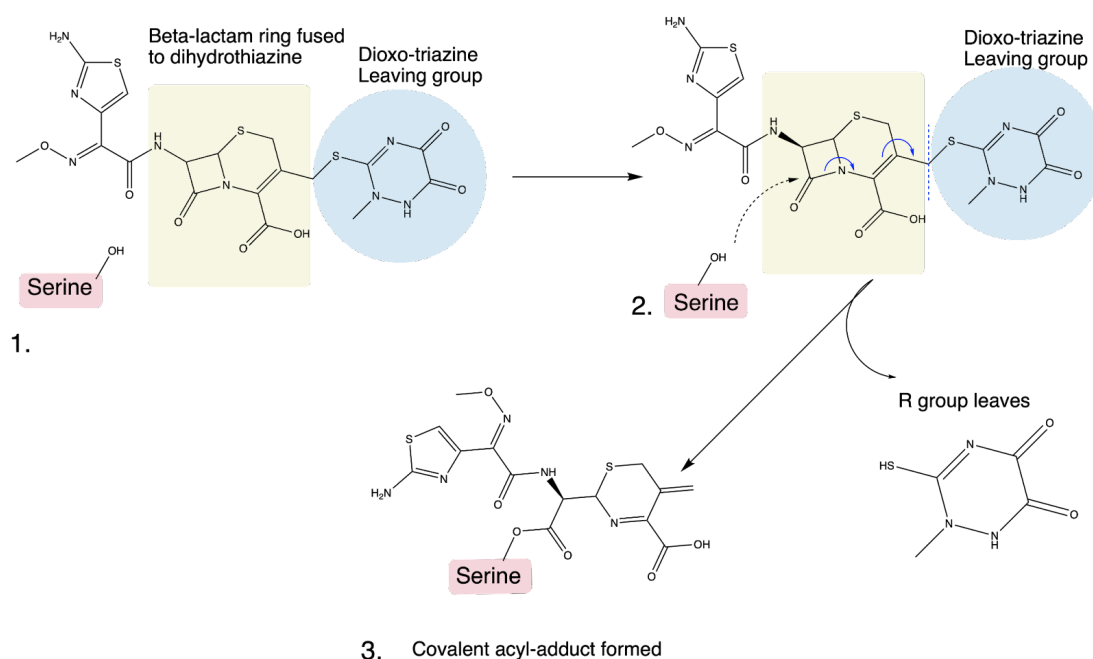


Figure 4.16 Mechanism of ceftriaxone acylating the active site serine of PBPs. (1.) The hydroxyl group of the active site serine is brought into proximity to ceftriaxone. (2.) Nucleophilic reaction of the β -lactam ring with the serine hydroxyl group leads to the opening of the β lactam ring. The rearrangement of electrons expels the R group – Dioxo-triazine group. (3.) A covalent bond is formed between the ceftriaxone derivative and the active site serine. (4.) Deacylation reaction. Figure is adapted from Edoo et al 2018²²⁰.

	Predicted molecular weight
Predicted molecular mass of <i>E. coli</i> PBP3	59847.6263
Molecular weight of ceftriaxone	554.5710
Predicted molecular weight of ceftriaxone without the leaving group	395.4080
Predicted molecular weight of <i>E. coli</i> PBP3 with ceftriaxone bound [M with ceftriaxone +H] ⁺	60403.2263
Predicted monoisotopic mass of <i>E. coli</i> PBP3 with ceftriaxone bound and leaving group gone	60243.32

Table 4.5 Predicted masses of *E. coli* PBP3 treated with ceftriaxone.

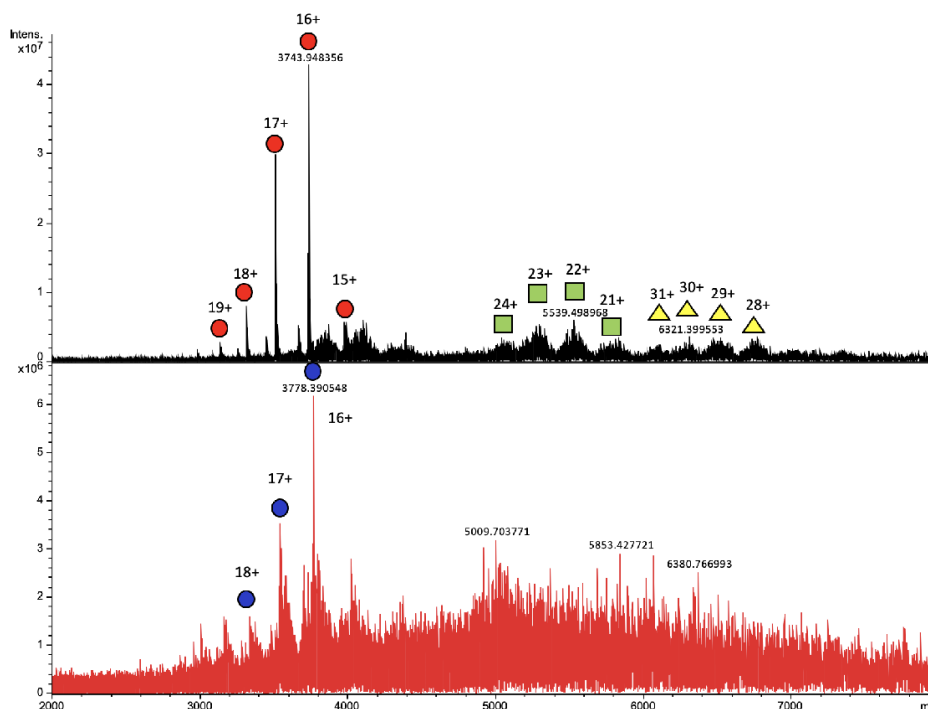


Figure 4.17A-B The FT-ICR MS spectrum of *E. coli* PBP3. The top spectrum is the untreated *E. coli* PBP3 (no antibiotic). And the bottom spectrum (red) is *E. coli* PBP3 with ceftriaxone treatment. The red and blue circles over peaks annotate the observed charged states of the protein (19+ to 15+).

The most abundant peak has an m/z of 3743.948356 ($[M+16H^+]/16$), giving an average protein molecular mass of $\sim 59,887.1737$. In **Figure 4.17A**, dimers and trimers are seen and denoted by the squares (estimated Mr 121,846.977) and triangle (estimated Mr 189,641.9866), respectively. The bottom spectrum shows the protein with ceftriaxone treatment; it has an estimated molecular mass of $\sim 60,438.24677$ (**Figure 4.17**). The observed masses of the ceftriaxone treated and untreated proteins have larger masses than expected (**Table 4.5**). The predicted molecular mass of *E. coli* PBP3 is 59847.6263, and the observed was 59,887.1737; this is a mass difference of 39.5474 Da, this could potentially be the mass of a potassium adduct (**Table 4.5**). The predicted mass of *E. coli* ceftriaxone and PBP3 is 60242.6263, as the leaving group should not be present (**Table 4.5**). The observed molecular mass of the ceftriaxone treated PBP3 was 60,438.24677. The observed mass was 195.62047, greater than expected.

The difference between the observed PBP3 and ceftriaxone treated PBP3 was 551, smaller than the expected mass of ceftriaxone in its complete form and larger than the expected mass in its acylated form. The mass differences observed between the ceftriaxone treated and untreated PBP3 and the differences between observed and expected are not consistent with the loss or mutation of an amino acid. The construct

for the expression of the PBP3 was sequenced and showed no mutations (**Appendix 9.4**). Furthermore this experiment was undertaken using the same *E. coli* PBP3 stocks used in earlier proteomic experiments for which there was broad spectrum sequence coverage. The mass shift with ceftriaxone should have been 395 Da, as the leaving group should have left following the interactions with the active-site serine (**Table 4.5**).

The monoisotopic mass of ceftriaxone sodium salt is 576.028; when protonated, this rises to 577.0358; if the compound were hydrolysed, a mass increase of +18 would be observed, and if the compound were decarboxylated, a mass decrease of -44 would be expected. These changes would result in a predicted mass of 551.035. This mass is consistent with the observed mass difference between the treated and untreated PBP3. This hydrolysed and slightly degraded version of ceftriaxone could interact with the PBP3 in a non-covalent way. Using the observed mass difference of 551.035 and accounting for a potassium adduct with a mass of 39, the new predicted mass of the ceftriaxone treated PBP3 rises to 60,437.66 which is 0.5855 less than the observed.

FT-ICR MS can trap charge and reanalyse molecules using fragmentation¹⁹⁵. Fragmentation techniques were used to analyse these proteins further; collision-induced dissociation (CID) and electron capture dissociation (ECD).

4.4.1 CID of PBP3-ceftriaxone acyl enzyme

For CID of *E. coli* PBP3, the samples were treated with 22eV (**Figure 4.18**). Before being trapped, the parent ion had a molecular mass of = 59,887.1737; after trapping with 22 eV CID, the molecular mass dropped to 59,810.04666, a difference of 77.12704 (**Figure 4.18**). The ceftriaxone treated PBP3 lost a mass of 245, going from 60,438.24677 to 60,193.03962 (**Figure 4.18**).

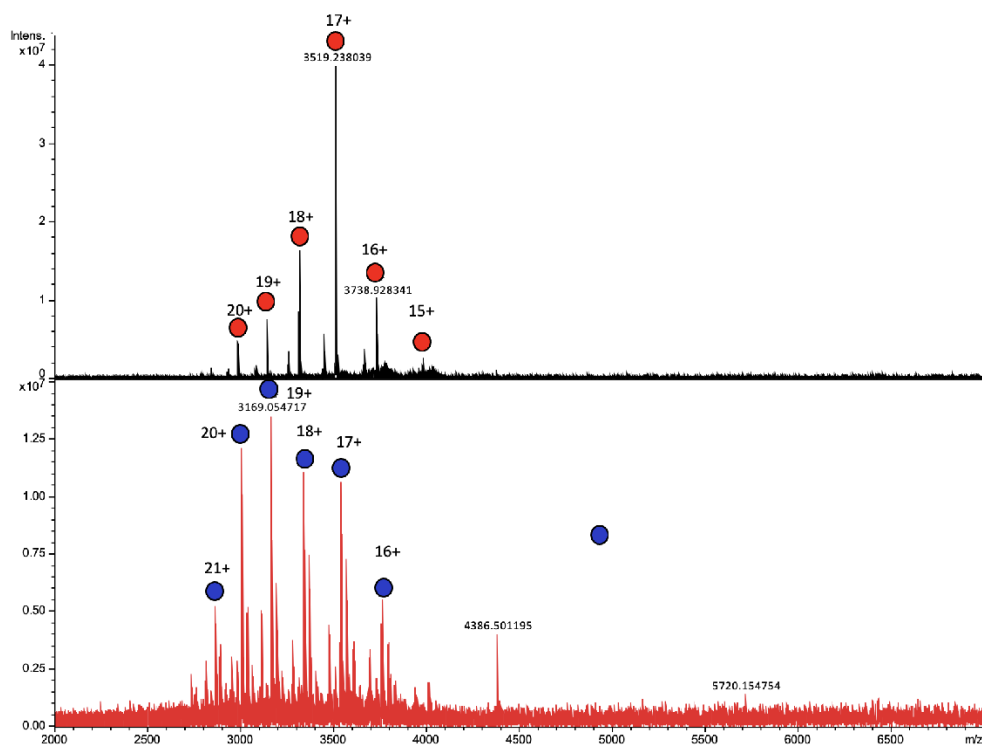


Figure 4.18 The FT-ICR MS CID spectrum of *E. coli* PBP3. The top spectrum (in black) is the untreated PBP3 control, and the bottom spectrum (in red) is *E. coli* PBP3 treated with ceftriaxone. The circles indicate the charge states of the protein.

The difference between the parent ion of *E. coli* PBP3 treated with ceftriaxone and the resulting fragment ion is 245 Da (**Figure 4.18**). 245 is not the size of the adduct. The untreated parental ion decreased to 77 Da (**Figure 4.18**). The expected mass loss was not observed, perhaps due to the stability of the covalent bond between the protein and the β -lactam.

4.4.2 CID ECD of PBP3-ceftriaxone acyl enzyme

Electron capture dissociation is used by FT-ICR MS as a fragmentation technique¹⁹⁵. It is particularly useful in protein analysis, as it fragments across the peptide backbone, whilst maintaining labile modifications¹⁹⁵. As previous work has demonstrated some lability in the β -lactam-PBP3 ester bond, ECD was an attractive option. Below is the observed CID ECD spectrum of PBP3 treated with ceftriaxone, trapped, and exposed to CID ECD at 20eV(**Figure 4.19**).

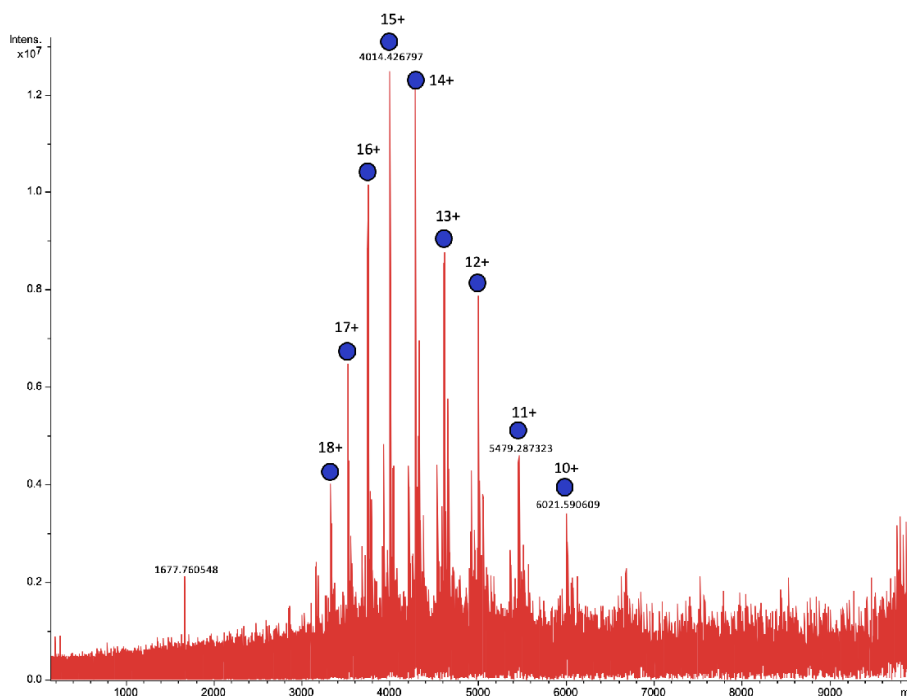


Figure 4.19 FT-ICR MS spectra of ECD CID PBP3-ceftriaxone sample. Again, blue circles indicate charge states of protein.

The observed mass of the protein after this treatment was 60,201.40196 Da. A 236.84481 Da change from the parental ion (60,438.24677)(**Figure 4.19**). This is not the mass expected of bound ceftriaxone (395 Da). This result was similar to the CID spectrum. It suggests that the adduct is stable under CID and ECD CID. Using CFM-ID 4.0²²¹, potential fragments of ceftriaxone were generated to search for any potential fragments of 236 or 77 that fit a plausible explanation of the results. Unfortunately, none of the fragments explained the mass difference observed in this work. As a result, this technique is not beneficial to the goals of this project.

4.5 Using MALDI-TOF MS to monitor the acylation of PBPs

Due to equipment availability and increased throughput, a new MS technique was picked. MALDI has been extensively used for complex biological samples due to its inherent ability to resolve complex mixtures^{199,222}. MALDI-TOF had a lower resolution than other mass spectra used in this work, so mass is calculated from the average m/z of triplicates. MALDI MS tolerates salts much better than in FT-ICR MS or Orbitrap MS^{199,222}. As a result, the proteins were not desalted before analysis.

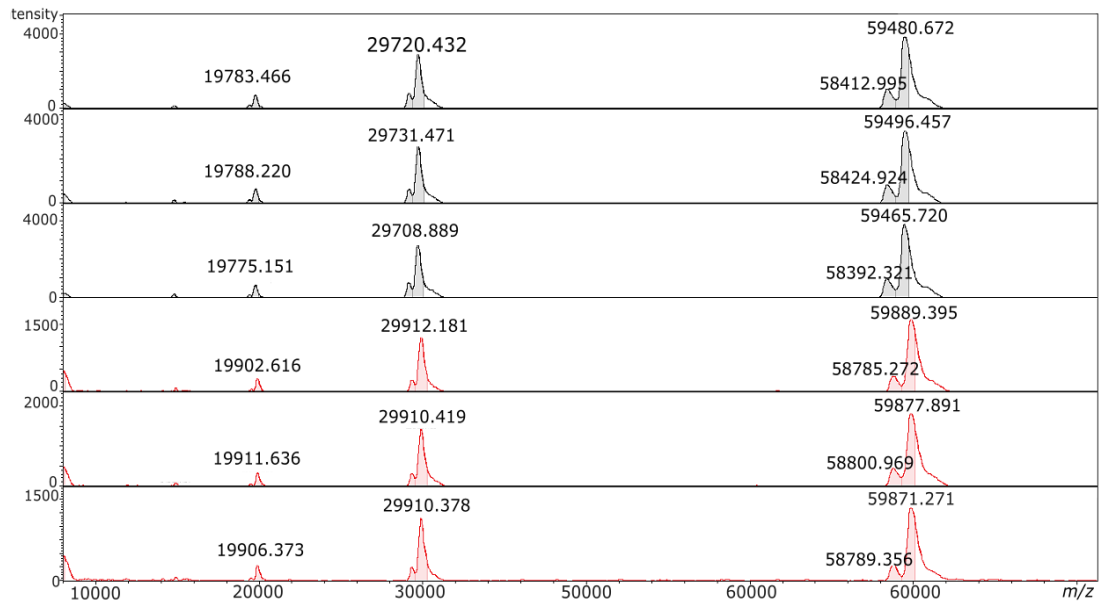


Figure 4.20 MALDI spectrum of *E. coli* PBP3 treated with ceftriaxone. The top three spectra in black are *E. coli* PBP3 control samples, and the bottom three spectra are with *E. coli* PBP3 with ceftriaxone treatment. The average m/z of the control samples was taken from the predominant peaks. The average mass of the untreated protein is 59480.95, and the average of the treated protein is 59879.52. The mass difference between the two averages is 398.56.

The predicted masses of *E. coli* PBP3 were as before (**Table 4.5**). The observed mass of *E. coli* PBP3 was 59480.95, 366.67 Da less than predicted (**Figure 4.20**). The mass of PBP3 ceftazidime acyl enzyme was 59879.52, this is 363.8 Da less than predicted (**Table 4.5**).

The difference between the observed masses of the two samples was 398 Da (**Figure 4.20**). This mass is close to the predicted mass of the ceftazidime PBP3 acyl-enzyme. The MALDI mass spectra have two peaks for the proteins and a broad shoulder on all the primary peaks (**Figure 4.20**). The presence of two peaks and a broad shoulder makes the deconvolution of the spectra more complex and increases the likelihood of inaccurate masses. This may explain why the mass difference was out by 3 Da. There are also doubly and triply charged species of the protein present at peaks around 26 kDa and 19 kDa.

4.5.1 MALDI-TOF of other PBP4

The interaction of PBP4 with ceftriaxone could not be observed. Some literature suggests the ionisation of ceftriaxone is more difficult than other cephalosporins due to

'weak molecular polarity'²²³. There is also literature that suggests that ceftriaxone is 'highly protein bound'¹⁹⁶. Both of these factors suggest that another cephalosporin may have been more appropriate. Nitrocefin seemed to provide more reliable adducts, so it should have been used, although the acylation method is reportedly different from other cephalosporins²¹⁵.

Furthermore, the sample preparation conditions for the MALDI may have been suboptimal. In general, when PBPs were mixed with ceftriaxone the proteins appeared to precipitate. This observation was echoed by other researchers who observed precipitation of PBP2a when treated with β -lactams²²⁴. Despite attempting to optimise this with different buffers and concentrations, it was still observed, and could have impeded this experiment. It is believed that the acylation of PBPs leads to a conformational change that may encourage precipitation²²⁴.

4.6 Discussion and Future prospects

Each protein and each peptide extracted from a protein (when we look bottom-up) upon acylation ionises to a greater or lesser extent dependent on many factors. This is true for different β -lactams. Meaning despite how much optimisation could be done for each protein, the favourable outcome could suppress the ions of another sample²⁰⁸.

According to Fontana *et al.* (2000)²¹⁹, relative affinities of *E. coli* PBPs for various β -lactams varies (**Table 4.6**)²¹⁹. Fontana *et al.* reported PBP1 has marked affinity to cephalosporins, PBP2 has an affinity to mecillinam and imipenem, PBP3 has an affinity for some cephalosporins, piperacillin and aztreonam, and PBP4 has affinity for benzylpenicillin, ampicillin and imipenem. Idealistically, this would have been proved in this work.

This work has focused on narrowing the complexity of resolving mass spectra of PBPs when treated with β -lactams. The Orbitrap and FT-ICR MS are substantially more resource-heavy without yielding better results. The goal of using these high-resolution mass spectrometers was general accuracy and accurate quantification. ¹³C isotopically labelled β -lactams would have been used for quantification after establishing a calibration curve²²⁵. The method would have been based on this paper's²²⁵ calibration curves, and enzyme assays and MICs would have supported the results. However, the

reproducibility and the lack of resources prohibited achieving this. Therefore, high throughput, low-resolution MALDI-TOF was tested.

A more straightforward MALDI method to quantify β -lactams would be to analyse the shift between the intact and the hydrolysed version of β -lactams²²⁶. Some authors looking at β -lactam resistance/ β -lactams, in general, quantify the proportion of the hydrolysed β -lactam^{226,227}. For example, when ampicillin is hydrolysed, a mass shift of +18 is observed²²⁶. This method was not used as it would not be sensitive enough for quantification due to the proclivity of the β -lactam ring to break. Furthermore, the focus of this work was to use β -lactam binding to PBPs to quantify permeability.

The MALDI allowed us to visualise the shift between untreated and treated PBPs, however, the results are essentially approximations. The next steps of this work could have been; developing a calibration curve with PBP3 and ceftriaxone on MALDI. This would have consisted of measuring the area under the curve of the peaks. This would have been confirmed by a higher sensitivity MS like FT-ICR-MS or Orbitrap using a heavy isotope labelled ceftriaxone²²⁵. The binding of PBP3 and ceftriaxone measured by MS would have been confirmed by concurrent enzyme assays. To move on to a more complex system, soluble His-PBP3 is over-expressed in *E. coli* cells, and then the binding is quantified by MS methods. This is built upon porin knockout strains, that are made chemically competent to express the protein.

4.7 Conclusion

The results from this chapter were not compelling enough to pursue this work further. The time and expertise needed to optimise an MS protocol for each β -lactam and PBP were unachievable. Furthermore, there are alternative methods to investigating β -lactam permeability, such as the micro-iodometric assay. The subsequent chapters focus on developing general assays and tools to investigate permeability.

Chapter 5. Creating synthetic cells to observe antibiotic permeability

5.1 Background

5.1.1 Liposomes

Historically, methods of measuring permeability have included techniques such as liposome swelling assays⁶⁶. Liposomes are unilamellar vesicles formed of a phospholipid bilayer, and they can be made in a range of sizes (**Figure 5.1**)²²⁸. Small unilamellar vesicles (SUVs) are 20 to 100 nm in diameter, large vesicles (LUVs) are 100 to 1000 nm, and giant unilamellar vesicles (GUVs) are 1 to 200 μm (**Figure 5.1**)²²⁸. Liposomes are heavily used in biological assays due to liposomes' adaptability; the internal environment can hold targets or reporters for assays, the phospholipid composition can be altered, and membrane proteins can be integrated. Liposome assays have been used to monitor radioactive, fluorometric and spectroscopic reaction²²⁹. As a result of the number of controllable variables, the application of liposomal assays is vast²²⁹.

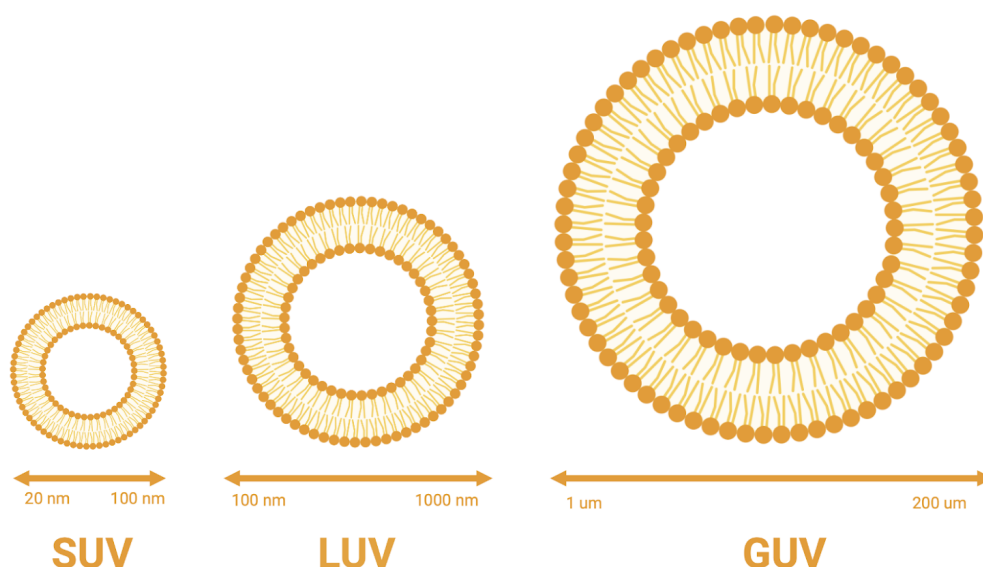


Figure 5.1 Size range of liposomes²²⁸. They are formed of a phospholipid bilayer, and are unilamellar (in this work). Membrane proteins can be integrated into the membrane of phospholipids during rehydration and with the use of BioBeads. The internal volume of liposomes can hold the components of the reporter.

Liposomes are a good model for the study of antibiotic permeability as both movement of compounds through the lipid layer and porins can be monitored²³⁰ and have been used extensively in this manner. Monitoring permeability through liposomes has usually required either a fluorescent reporter or fluorescent antibiotics. The necessity of fluorescence limits the broader application of the techniques to drug discovery.

Both small and large vesicles can be formed using rehydration and freeze-thaw cycles, followed by extrusion^{228,231}. Historically GUVs were made using a 'gentle hydration' technique or, more recently, GUVs have been made using the inverted emulsion transfer technique (used in this work) or electro-swelling formation^{228,231}. However, microfluidics is the most current and precise method to form homologous GUVs^{67,79,232}.

Liposomes have long been used as cell models. However, liposome models are reductive. The aim of this work is that by combining liposome technology with synthetic biology techniques, a more complex model of *E. coli* can be built.

5.1.2 Synthetic cells

Synthetic biology aims to study and imitate biological phenomena and utilise them to expand the contemporary edge of science²³³. Synthetic cells are a goal of biological research. There are two critical approaches to this goal: a top-down approach and the second being a bottom-up approach²³³. Top-down approaches aim to knock out all but the essential genes in organisms to rebuild and programme them with specific functions^{233,234}. The bottom-up approach is centred around building up an artificial cell system from individual components¹⁴². The following chapter outlines the use of the later approach to building a system to measure permeability and aid in drug discovery.

Bottom-up synthetic cells are formed of four key components; a liposome, a cell-free transcription/translation system, energy sources, and the target DNA. Cell-free expression systems have been used to create proteins and nucleic acid²³⁵⁻²³⁷. There are two main categories of cell-free protein synthesis systems (CFPSs); either a 'PURE' expression system²³⁸ or a cell-free lysate system. PURE based expression systems are composed of purified proteins^{238,239}. In comparison, lysate based methods use concentrated cell-lysate to synthesise proteins (**Figure 5.2**)²⁴⁰⁻²⁴². The lysate system can be made using different species' machinery. The main species used to make lysate based CFPS are; *Saccharomyces cerevisiae*, *E. coli*, rabbit reticulocyte lysate, Chinese

hamster ovary, and wheat germ²⁴¹. Each system comes with its own advantages and disadvantages, and are useful in different applications. The system used in this chapter is an *E. coli* lysate system based on the work of Noireaux *et al.*^{129,142,143} (**Figure 5.2**). Noireaux *et al.* have done extensive work developing their system, and have explored the efficacy in an interdisciplinary manner, looking at proteomics²⁴³, modelling²⁴⁴, and microfluidics²⁴⁵.

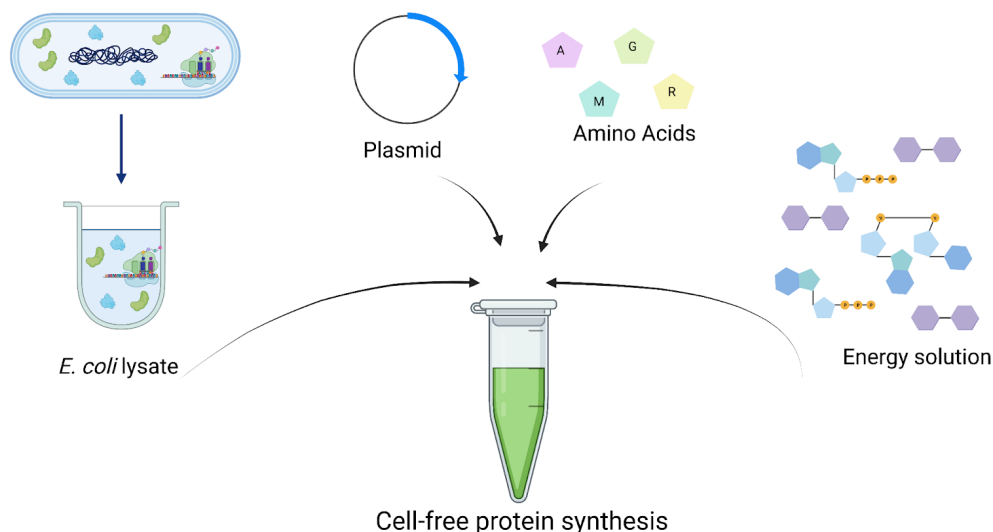


Figure 5.2 Components of CFPS. The *E. coli* lysate should contain the key transcription and translation machinery like RNA polymerase, ribosomes, elongation factors²⁴⁶. This is supplemented with an energy solution which contains ATP but also 3-Phosphoglyceric acid to regenerate ATP. The energy solution also contains nucleotides and tRNA as building blocks for transcription and translation. This is supplemented with amino acids. The plasmid of interest is supplied, and this results in protein synthesis.

Synthetic cells allow for much greater experimental control of variables, and more manipulation, allowing more modifications of the platform for this or other applications. Synthetic cells also allow us to create complex and dynamic systems compared to proteoliposomes whilst still maintaining control. Using synthetic cells as a platform allows us to investigate the effect of different lipid layers, be it artificial or native. It could also allow for the integration of membrane proteins^{247,248}. Synthetic cell systems allow for the introduction of multiple plasmids, which means multiple proteins can be expressed. In this work, synthetic cells will be created to mimic *E. coli*, in order to develop a new permeability assay.

5.1.3 Outline of work

The inverted emulsion transfer technique was used to create the GUVs that encapsulated the CFPS reaction (**Figure 5.3** for a visual representation of the technique)^{228,249,250}. Two versions of the inverted emulsion transfer technique were trialled, one using an Eppendorf tube²⁴⁹ and the second using a 96-well plate^{146,147}. There were three key components, outer solution, interface, and emulsion. The outer solution was an aqueous solution that, in this application, was used as a 'feeding solution' for liposomes¹⁴¹. Noireaux *et al.* defined 'feeding solutions' as the aqueous solution used to encapsulate and maintain protein synthesis¹⁴¹. There are three feedings solutions, F1, F2, and F3¹⁴¹. F1 consists of phosphate-buffered saline and 200-400 mM glucose¹⁴¹. F2 is Tris-buffered magnesium glutamate, potassium glutamate, PEG, and maltodextrin to feed liposomes and maintain osmotic balance¹⁴¹. F3 is the most complex and expensive feeding solution formed of all of the components of the CFPS system, without the lysate, meaning the energy solution, the ATP, the magnesium and potassium glutamate, PEG and maltodextrin¹⁴¹.

Forming liposomes in the inverted emulsion technique involves layering phospholipids on top of a feeding solution (also known as the outer solution), creating an interphase layer. The interface is a thin layer of phospholipids dissolved in oil; this is layered on the outer solution and left for at least thirty minutes to form an interface^{146,147,249}. The emulsion is composed of phospholipids and an inner solution; the inner solution (in this application) is formed from a CFPS reaction. The emulsion is formed either by vortexing (Eppendorf)^{145,249} or by mechanical agitation (96-well plate)^{146,147}, creating inverted micelles of inner solution surrounded in phospholipid. This emulsion is added on top of the interphase layer. The encapsulation happens when the solution is centrifuged, creating GUVs in the outer solution^{145,146,249}. In the Eppendorf approach, GUVs need to be extracted; this requires puncturing the tube with a needle and collecting a droplet of GUVs. In contrast, the GUVs formed in the 96-well microplate are not extracted for visualisation.

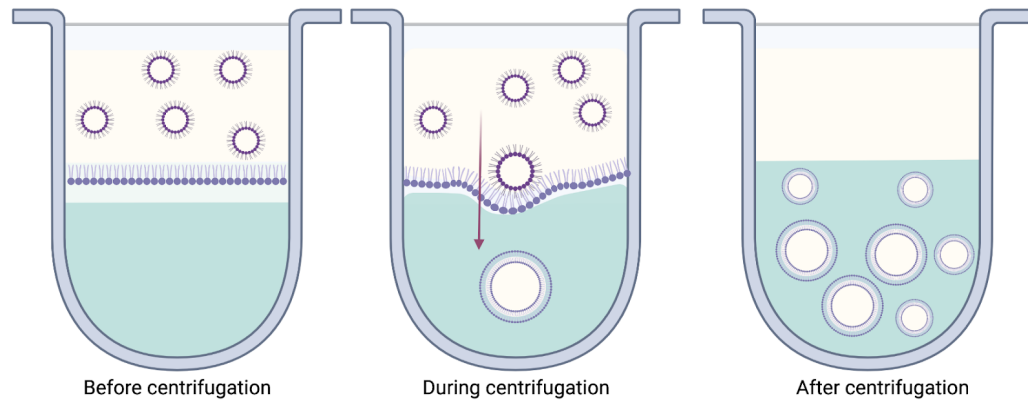


Figure 5.3 The inverted emulsion transfer technique used to form GUVs. In the bottom of the wells is the outer solution which in this case was either F1, or F2¹⁴¹. Followed by a layer of lipid in oil forming the interface. The lipid at the interface layer will form the outer leaflet of the bilayer. After this interface has formed, the inner solution is mixed with lipid and oil to form inverted emulsions which are layered over the interface.

Using 96-well plates to create synthetic cells will result in synthetic cells being available in a high-throughput format. By having the CFPS produce a fluorescent protein, any drugs that inhibit transcription or translation should reduce the amount of fluorescent reporter created. The aim is to gradually increase the complexity of the GUV until it is an accurate model of the *E. coli* outer membrane. Therefore, the platform is not only set up to screen new drugs but also, it can monitor permeability across the outer membrane. Thus, creating an activity screening platform that can be used in tandem with a permeability screen.

5.1.4 Research aims

- Create an in-house CFPS system
- Encapsulate this CFPS system in GUVs
- Scale up these synthetic cells to a 96-well format for antimicrobial activity screens

5.2 Results

5.2.1 Creating cell free protein synthesis system

The first step of work was to establish a CFPS system. Working from the literature of Noireaux *et al.*^{129,140,141,251} and Levine *et al.* 2020¹²⁴, an in-house *E. coli* cell-free expression system was implemented. *E. coli* Rosetta (DE3) were grown in cell-free auto-induction media¹²⁴, harvested and lysed as reported by Levine *et al.*¹²⁴. This lysate was used in conjunction with amino acids (biotech rabbit) made up to 1.5 mM, with an energy source consisting of an ATP regeneration system. Reactions were made up to a volume of 20 μ L, whereby the production of a green fluorescent protein (deGFP) was monitored in the PHERAstar plate reader in 384-well (Greiner) plates at Ex./Em. 488/510 nm, a temperature of 30°C. deGFP construct from Noireaux *et al.*¹⁴³ was chosen due to its supporting literature and evidence that it is more translatable than other forms of GFP. The deGFP construct was expressed from a p70a plasmid that uses endogenous *E. coli* RNA polymerase and sigma factor 70^{140,143}.

Many components of the energy solution require optimisation; DTT, ATP, energy sources, and magnesium glutamate concentrations. The concentration and preparation of the DNA can influence protein synthesis. DNA was prepared using a MAXIprep (Qiagen) kit to ensure the removal of any contaminating RNase.

5.2.2 Optimising CFPS system

Setting up a CFPS was an iterative process. As the reaction had many constituents, there were many potential points of failure; it is also not the most robust process. The system's activity was impacted by a range of factors such as DNA concentration²⁵² and preparation and magnesium concentrations. Characterisation of each lysate preparation is necessary for an effective CFPS system²⁵¹. To assess the system's efficacy, a collaborator kindly gifted their CFPS components for comparison (S. Liyanagedera, University of Warwick). The CFPS was given the same deGFP (15 nM) plasmids incubated at 30°C in parallel, and the relative fluorescent intensity of deGFP values was read after 2 hours (**Figure 5.4**).

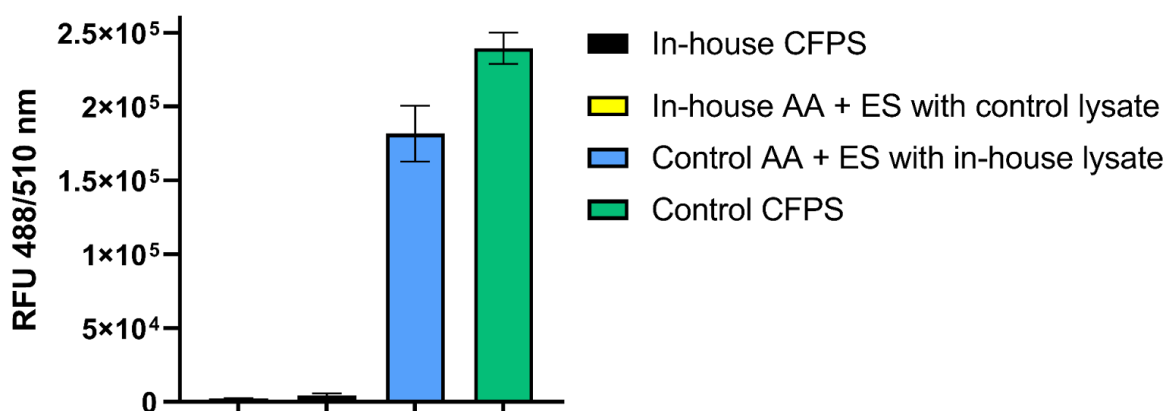


Figure 5.4 Troubleshooting CFPS. In-house set-up was compared to collaborators (S. Liyanagedera) set-up. 15nM deGFP plasmid was maintained across reactions. Amino acids have been abbreviated to AA and the energy solution was abbreviated to ES. On the left of the bar chart is the in-house CFPS and next to this is the in-house amino acids and energy solution used with control lysate. In the third bar (blue) is the control amino acids and energy solution used in conjunction with the in-house lysate. On the right hand side (green) is the control CFPS reaction.

The first test on the lysates functionality appeared to be 25% less active than the control reaction (**Figure 5.4**). However, the main issue appeared to be with the rest of the CFPS reaction components. As a result, different concentrations of DTT, amino acids, magnesium-glutamate, DNA and ATP were tested. Additional ATP (additional 1mM) significantly boosted the reaction. Supplementing the other key components in the energy solution, such as folinic acid, cAMP, and NAD, also increased deGFP productions. By adjusting the energy solution, deGFP was produced to the same levels as when using the control energy solution.

5.2.3 Purifying deGFP

The deGFP used in the CFPS reaction was purified using traditional *E. coli* cell-based methods, to compare to the CFPS reaction. The deGFP was expressed and purified by nickel IMAC affinity chromatography. The protein concentration was quantified using the nanodrop at 488 nm, extinction coefficient 55,000 (cm⁻¹ M⁻¹). A dilution series of these values was used to create a calibration curve to convert relative fluorescence units (RFU) to concentration (**Figure 5.5**).

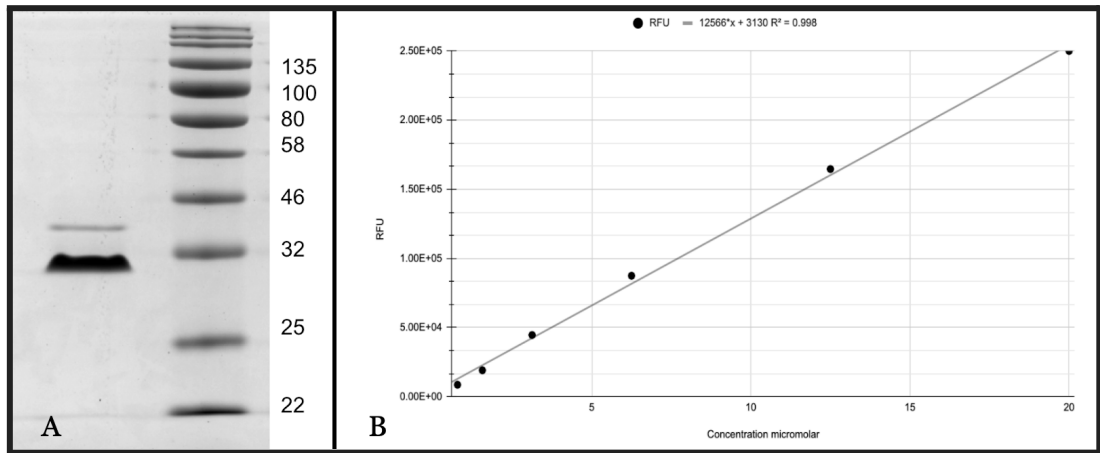


Figure 5.5AB Purified deGFP and the accompanying concentration curve created. A Left; SDS-PAGE gel of purified deGFP. On the left of the gel is the purified protein, and on the right is the protein ladder (NEB prestained protein ladder). **Figure 5.5B**; right is the calibration curve constructed of purified protein. 1 mg/mL deGFP (29. kDa) is 34.5 μ M.

The purification of deGFP and the calibration curve allows for the estimation of deGFP production (**Figure 5.5B**). deGFP production by the CFPS was monitored over three hours at 30°C, where fluorescent readings were taken every three minutes (Ex./Em. 488 nm/510 nm). The monitoring of the fluorescence in this way allows us to understand the kinetics of the process, which would be necessary for the analysis of the impact of antibiotics on encapsulated CFPS.

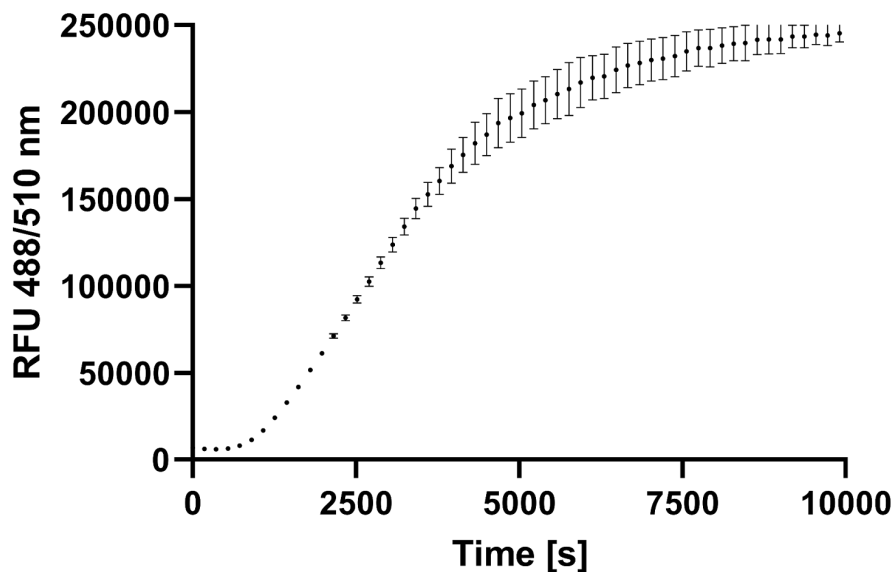


Figure 5.6 The CFPS deGFP production over time. deGFP production was monitored using fluorescence intensity readings over time on the PHERAstar plate reader (BMG) Ex./Em. 488/510 nm

Over three hours the reaction produced 20 μ M of deGFP (**Figure 5.6**). 10 μ L of one of the reactions was mixed with a protein loading buffer and analysed by SDS-PAGE alongside the control (**Figure 5.7**). The SDS-PAGE gel offered little information about GFP synthesis as the CFPS reaction was rich in proteins, and the abundance of deGFP was comparatively low. However, the gel was imaged fluorescently prior to stains to allow for clear visualisation of GFP.

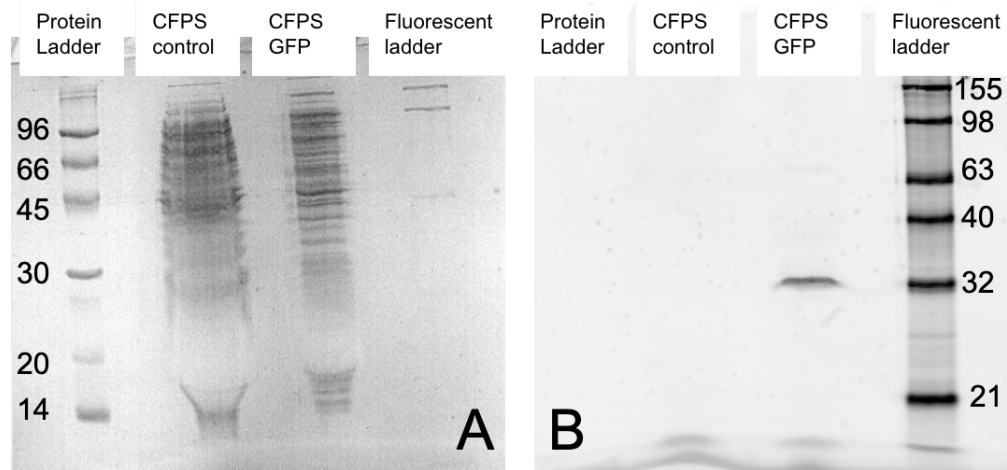


Figure 5.7 Visualising deGFP produced from CFPS reactions. **5.7A:** Coomassie stained SDS-PAGE Gel of CFPS reaction. The prestained protein ladder had band sizes of 96, 66, 45, 30, 20 and 14 kDa. The fluorescent ladder had band sizes of 155, 98, 63, 40, 32, 21 kDa. **5.7B** Fluorescent image of the SDS-PAGE gel. **5.7A** order; protein ladder, CFPS no DNA, CFPS deGFP plasmid, fluorescent ladder. **5.7B;** LtoR, Protein ladder not seen, then no bands in control lane, the CFPS lane, then fluorescent ladder, deGFP has a mass of 29 kDa.

Noireaux's paper demonstrates that deGFP production continues for up to 12 hours, whereas this gel (**Figure 5.7**) was visualised after a two-hour incubation^{140,143}. Noireaux *et al.* state that deGFP production in their transcription/ translation (TXTL) system began after 1 hour of incubation, and there was linear production of deGFP for 8 hours^{244,253}. Perhaps a longer incubation period is needed to produce a clear band on SDS-PAGE gels. When the protein gel was imaged fluorescently, deGFP was clearly visualised.

5.2.4 Inhibiting CFPS with antibiotics

Researchers using CFPS often employ antibiotics that inhibit protein synthesis, such as rifampicin or gentamicin, as negative controls. To further confirm the establishment of

a CFPS, the response of deGFP synthesis to 50 μM gentamicin (a protein synthesis inhibitor) was analysed.

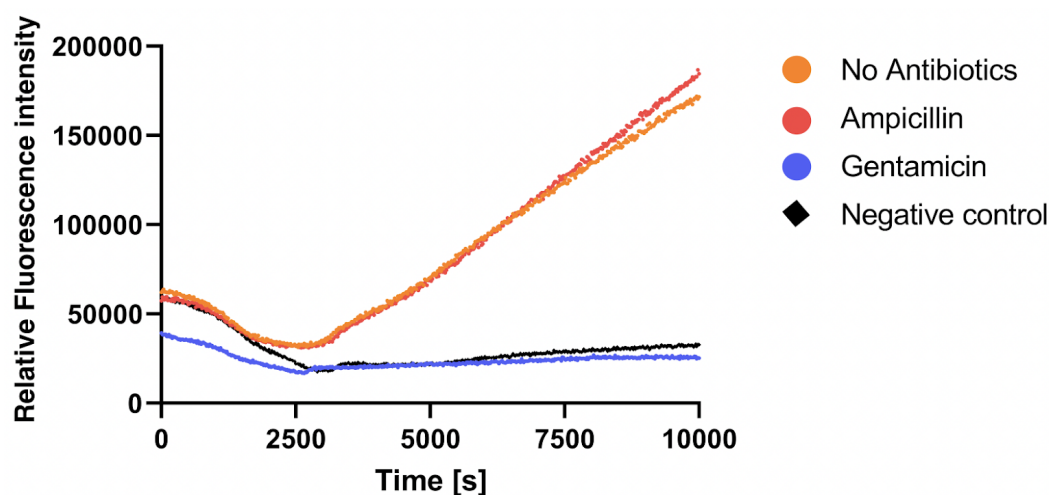


Figure 5.8. The deGFP produced by CFPS reaction in the presence of different antibiotics. The CFPS was set up with either: no antibiotics - regular set-up (orange), ampicillin (red), gentamicin (blue), no antibiotics and no DNA - negative control (black). The reaction was monitored in the PHERAstar for three hours at 30°C.

The antibiotic study showed that gentamicin inhibited deGFP production, demonstrating fluorescence rates in line with the negative control (no DNA) (**Figure 5.8**). Ampicillin and the positive control produced fluorescent deGFP (**Figure 5.8**). Ampicillin should not interfere with deGFP production in general as it targets cell wall synthesis and because it is the selection marker on the plasmid. In conclusion, the CFPS system is working as expected, and it can make deGFP (**Figure 5.8**).

5.3 Creating Synthetic Cells

5.3.1 Creating Synthetic Cells using Eppendorf technique

As the CFPS platform was producing protein, this reaction was encapsulated in POPC lipid to create 'synthetic cells'. To create synthetic cells, a bottom-up approach was trialled. Here the cell-free extract was encapsulated in POPC lipid using a technique adapted from Matsuura *et al.*¹⁴⁵ and Noireaux *et al.*¹⁴¹. This involved encapsulating cell-free extract in lipids surrounded by an energy solution, using the emulsion transfer method. This technique was done in small volumes in an Eppendorf tube. After the synthetic cells were created, they were incubated at 30°C and visualised. The protocol yielded some GUVs that synthesised deGFP.

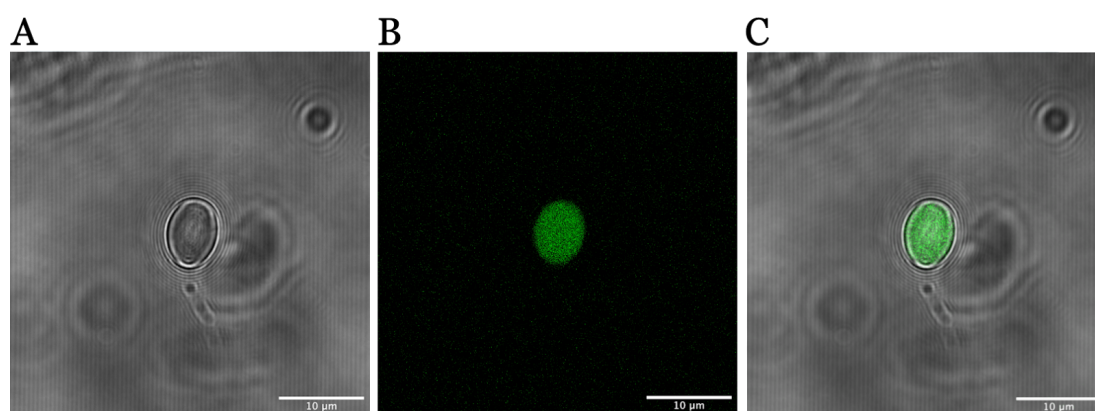


Figure 5.9A-C A synthetic cell synthesised using Eppendorf technique. 5.9A GUV under brightfield microscopy. **5.9B** (same) GUV visualised using fluorescence microscopy Ex./Em.: 488nm /510 nm. In **5.9C** both images overlaid using Imagej Fiji. Images were taken at 100x magnification under oil immersion on ZEISS 880 microscope argon laser. This work arose from a collaboration with S. Liyanagedera (University of Warwick), methods were developed collaboratively and reagents were shared.

This technique created the synthetic cells; however, they were small (for GUVs - which can be up to 200 μm), with sizes around 6 to 8 μm (**Figure 5.9**). The yield using this method was repeatedly low; on average, only 10 GUVs per Eppendorf preparation were observed. Although this technique created GUVs that synthesised deGFP, it was labour intensive, slow, and had poor yields. Therefore, this method would not be applicable to drug screening methods. As a result, a new method of creating synthetic cells was investigated to lead to a higher yield and provide a more amenable format for drug discovery platforms. It was also anticipated that adding a membrane stain would also improve visualisation.

5.3.2 Creating GUVs using the microplate technique

A new technique described by Litschel *et al.*¹⁴⁶ was used to create synthetic cells using the 96-well plate method (**Figure 5.3**). A lipophilic membrane dye (Dil iodide) was added to aid in visualising GUVs. To begin with, the microplate technique was used to make POPC GUVs that encapsulated purified deGFP as opposed to the cell-free reaction. Purified deGFP was encapsulated before CFPS because the CFPS reaction is more challenging to encapsulate than purified protein¹⁴¹.

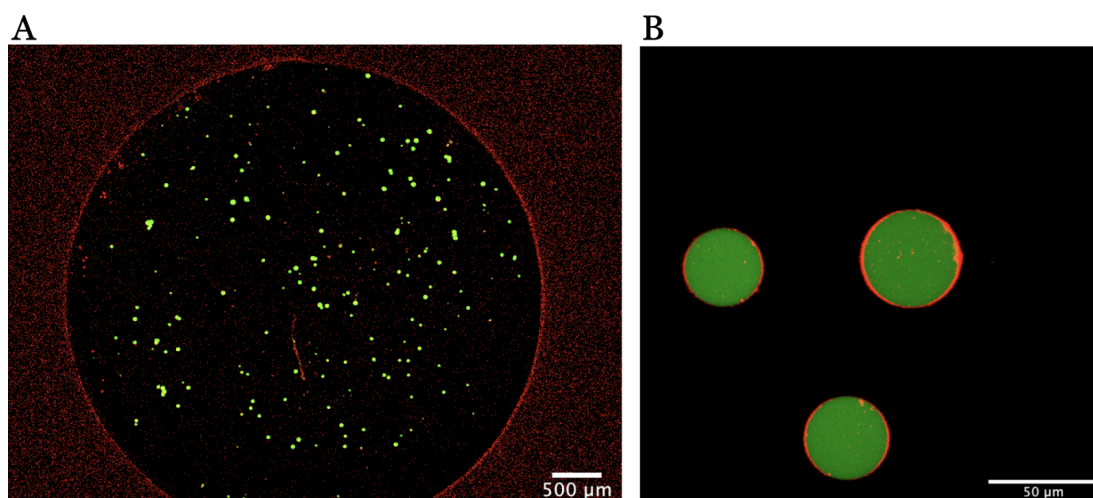


Figure 5.10A-B GUVs created by the microplate method that encapsulates deGFP. **5.10A** is an image of the microplate containing GUVs taken at 5x magnification; the image was created from a series of tile scans of the well. In **5.10B** is an image of the vesicles at 40x magnification. The deGFP and Dil fluorescence were observed on different channels, which were merged. Images were taken on the ZEIS-LSM 880, both images had the brightness and contrast autocorrected on FIJI ImageJ.

From a tile scan of the wells, there are more than ten vesicles visualised (**Figure 5.10**). The number of deGFP-containing vesicles could be quantified adequately by either particle counting programs²⁵⁴ or flow cytometry^{249,255}. An enhanced magnification of the vesicles shows that the deGFP has been encapsulated and that the use of a membrane dye (DIL) has enhanced the GUV images (**Figure 5.10**). The microplate technique has allowed the creation of more vesicles, and these vesicles are in a format that is more permissible for drug screening (*i.e.* 96-well plate for assays). As a result, this technique was trialled for encapsulation of cell-free extract.

5.3.3 Creating synthetic cells using the microplate technique

The 96-well microplate method to generate GUVs worked well with purified deGFP. As a result, this technique encapsulated the CFPS reaction to form synthetic cells. The cell-free reaction was encapsulated and incubated at 30°C for two hours before visualisation.

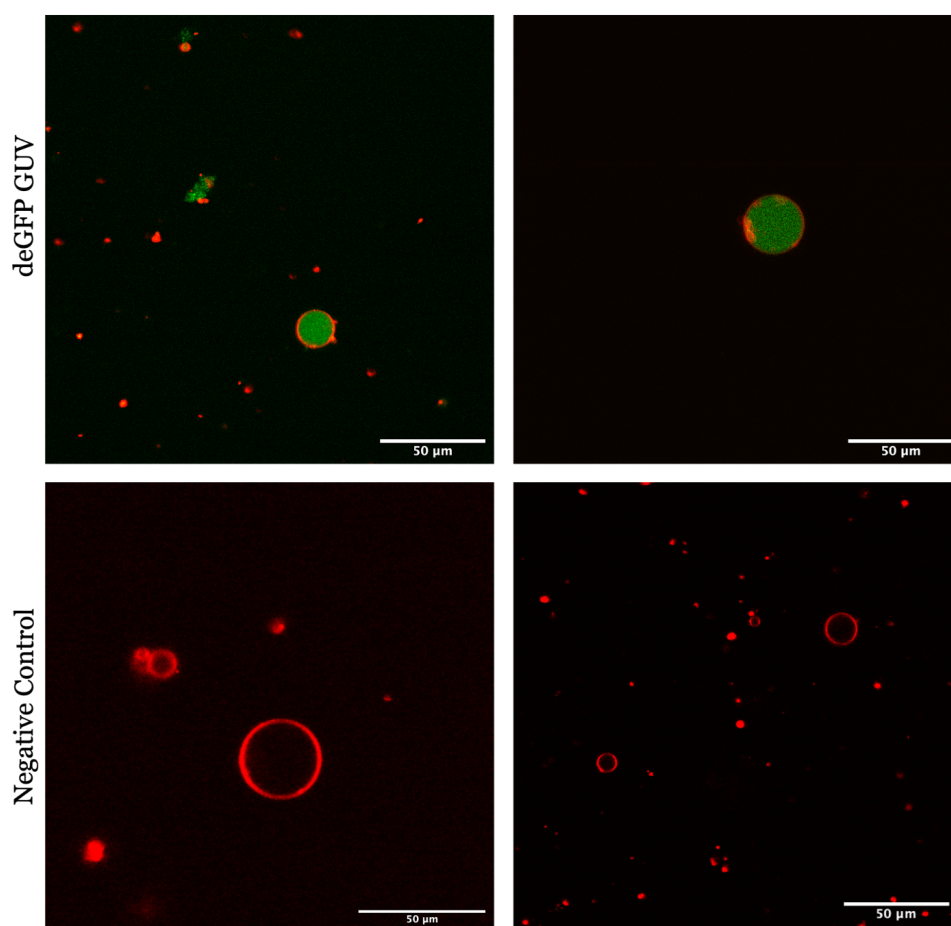


Figure 5.11 Synthetic cells formed using 96-well microplate method. 5.11 top-row: Z-stack of GUVs making deGFP, 40x magnification. 5.11 bottom-row GUVs that had not encapsulated the plasmid (control), 40x magnification, and as a result did not make deGFP. Again, images were taken on the ZEISS-LSM 880, and the figure was produced with FIJI ImageJ. This work arose from a collaboration with S. Liyanagedera (University of Warwick), methods were developed collaboratively and reagents were shared.

Using the microplate method, GUVs were created that produce deGFP (**Figure 5.11**). However, the intensity of the deGFP is much lower than that of purified deGFP. In addition, there are some empty vesicles and fewer vesicles in general, suggesting that perhaps the encapsulation efficiency of the CFPS is lower, so it has lost some of the essential transcription-translation machinery. Another quantitative technique should have been used to differentiate the functioning synthetic cells from the control well. Ultimately, this technique did work, synthetic cells were created, and they were in a 96-well plate format. The following steps for this work were to optimise the process.

5.3.4 Monitoring the synthetic cells creating deGFP over time

A time-course study of deGFP expression was undertaken to monitor the production of the protein over time. The CFPS reaction was encapsulated by POPC lipid forming GUVs as before, but this time the incubation period took place within the incubation chamber of the microscope. Over two hours, pictures of the vesicles were taken every three minutes to visualise deGFP production (*Figure 5.12*).

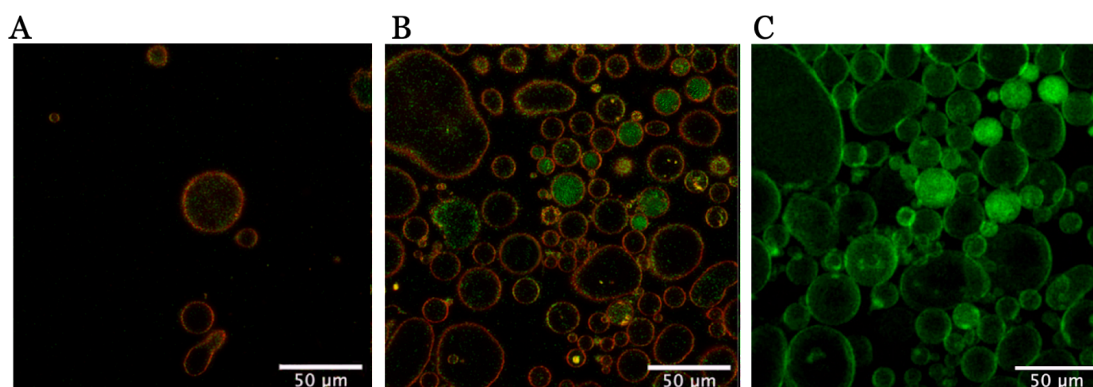


Figure 5.12A-C Synthetic cells producing deGFP over two hours. In **5.12A** a single image (single z plane) at time point zero, and in **5.12B** an image (single z plane) taken after two hours of incubation. **5.12C** is a Z-stack of the vesicles at the end of the incubation just the deGFP channel is shown. Images were captured on the ZEISS-LSM 880, and the figure was produced with FIJI ImageJ.

The time-course experiment captured some deGFP production over time; however, the amount of deGFP produced seemed low (*Figure 5.12*). There were also many empty cells (*Figure 5.12*). Interestingly cells seemed to coalesce, and budding was observed. The cause of this could be due to phase separation of the lipid²⁵⁶. It could also be due to osmolarity imbalances²⁵⁷. The blebbing and coalescing of vesicles could be impacting the production of deGFP and could impact future experiments. This experiment showed that deGFP production can be monitored over time, however it reinforced the complexities of the goal of generating synthetic cells. The main pitfalls of low numbers and inconsistent deGFP production are still present.

5.3.5 Using antibiotics to inhibit synthetic cells producing deGFP

Gentamicin was added to the outer solution so that after GUV formation, the inhibitor surrounds them. The gentamicin should permeate across the lipid layer and inhibit the

production of deGFP. To better mimic *E. coli*, *E. coli* polar lipid extract (Avanti) was used to form the vesicles.

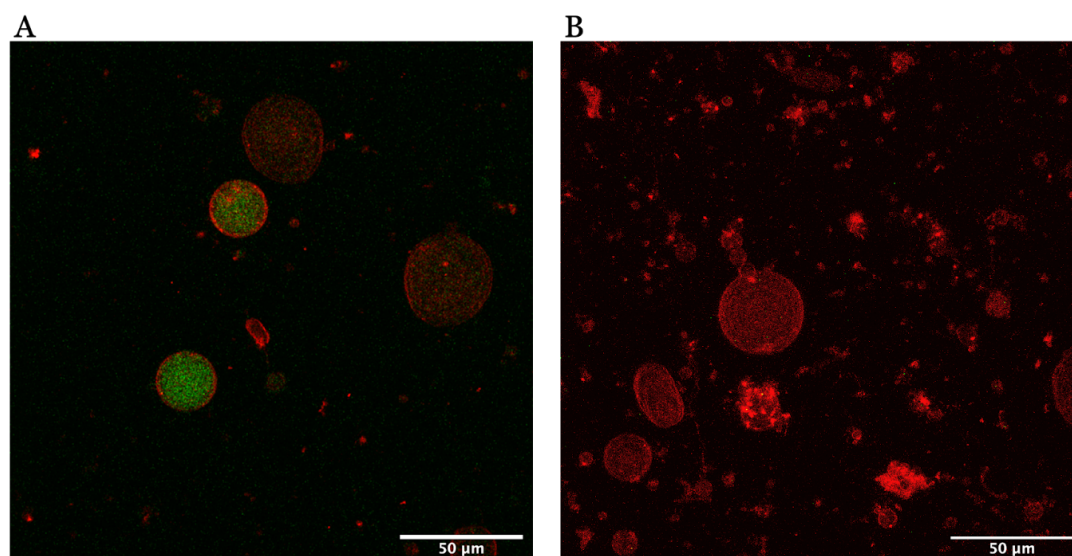


Figure 5.13A-B Gentamicin inhibiting deGFP produced by synthetic cells 5.13A on the left are uninhibited GUVs, they have produced deGFP. **5.13B** On the right are GUVs that are surrounded in 50 μM gentamicin.

The microplate method was used to encapsulate CFPS and create synthetic cells made of *E. coli* phospholipid (**Figure 5.13**). The *E. coli* polar lipid extract is not an exact mimetic for the outer membrane. However, it is a complex of natural phospholipids and provides a good starting point for increasing complexity. The gentamicin appeared to suppress deGFP fluorescence, as predicted, but the low vesicle production rate suggested there was room for improvement. Furthermore, the inhibition of deGFP in synthetic cells by gentamicin should have been further investigated to ensure the inhibition was coming from the gentamicin and not empty vesicles.

The synthetic cells in 5.13 were made with an F2 feeding solution as outlined by Noireaux *et al.*¹⁴¹. Perhaps the lack of an energy solution prevented effective CFPS. Noireaux *et al.* suggested that F3 as an outer solution will lead to increased deGFP production but decreased vesicle numbers¹⁴¹. Whereas F1, the simplest of the outer solutions, would lead to more vesicles but decreased protein production¹⁴¹. As the F2 feeding solution was a compromise between the two, it was used here¹⁴¹. The ionic strength of the buffer greatly influences GUV yield²⁵⁶. Buffers with high ionic strength lead to lower yields of GUVs²⁵⁶. The protocol was optimised since both the vesicle numbers and the deGFP production were consistently low.

5.3.6 Optimisation of microplate method

As the production of GUVs was low in general, the microplate method was optimised following Moga *et al.*¹⁴⁷. Firstly, the volume of the outer solution was decreased from 100 to 50 μL , and the interface volume (400 μM lipid in oil) was halved to 20 μL . The interface was incubated for the same period (30 minutes). The emulsion was made of 5 μL inner solution in 250 μL of lipid in oil, which was mechanically agitated (as before). This time 50 μL of the emulsion was added over the interface as opposed to 40 μL . As in previous experiments, the protocol was completed in an ATMOS bag filled with nitrogen to ensure water was not absorbed by the mineral oil^{146,147}. As always, particular care was taken in preparing and drying the lipids to prevent oxidation. These amendments were made with an inner solution either consisting of 20 μM deGFP or BSA, or CFPS reaction.

The lipid composition of bacteria plays a role in permeability^{54,55}. Bacteria adapt their lipid composition to their environment (temperature and pH), which can increase or decrease membrane permeability⁵⁴. Antibiotic exposure can cause bacteria to change the composition of their phospholipids to become antibiotic-resistant²⁵⁸. As a result, it is essential to build representative models of bacterial membranes. A feature of the emulsion transfer approach is that an asymmetrical bilayer can be formed by having two different lipid compositions, one at the interface and one mixed with the inner solution²⁵⁹. The *E. coli* outer membrane is asymmetric; the inner layer contains roughly 70-80% PE, 10-20% PG and 2-10% CL. The outer leaflet of the outer membrane is mainly composed of lipopolysaccharide (LPS). *E. coli* polar phospholipid was used for the inner leaflet and was mixed with the inner solution. For the outer leaflet, KDO2-Lipid A was used; this is the lipid A and core sugars of LPS.

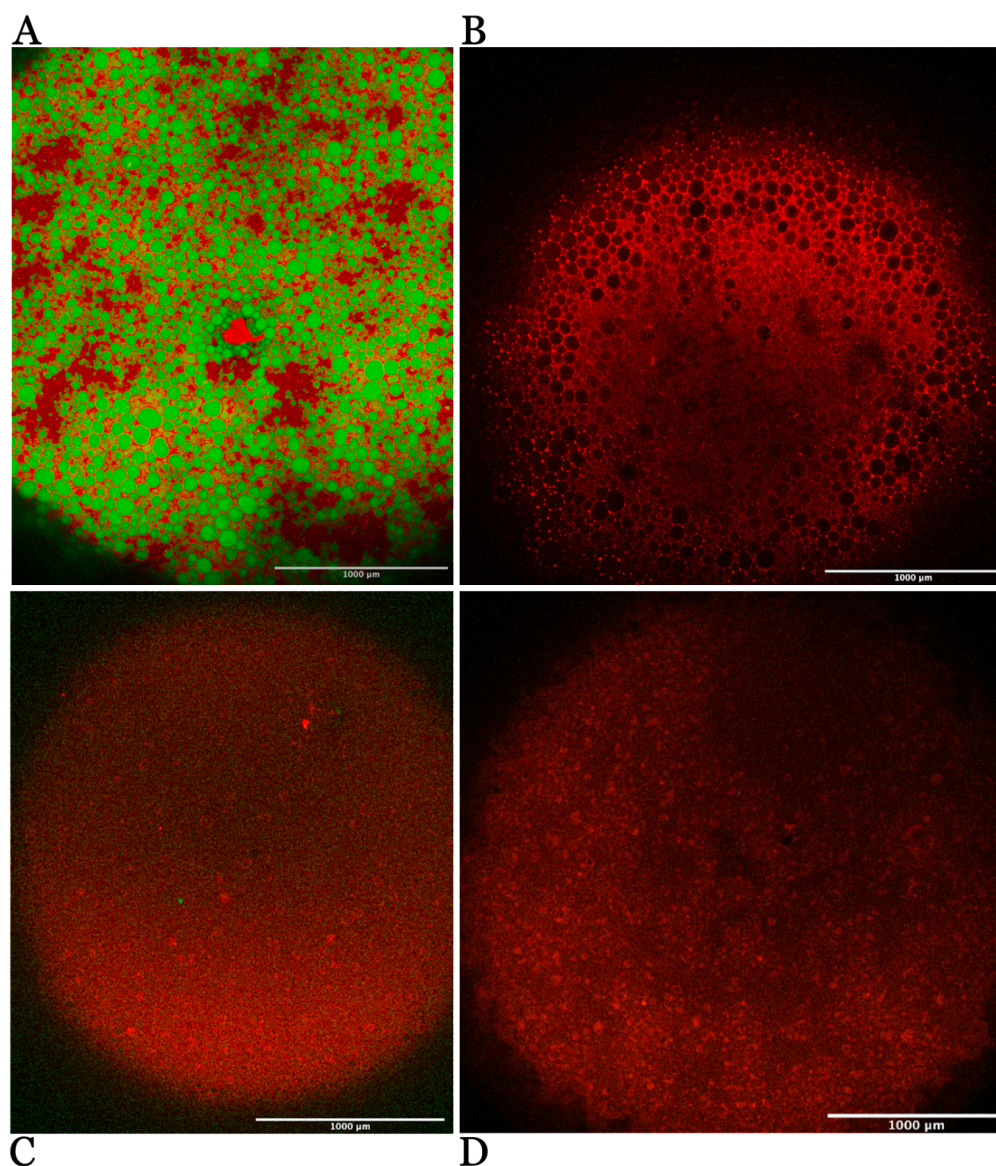


Figure 5.14A-D Tile scans from the optimised microplate method. Top left: GUVs with pure deGFP, top right: pure BSA GUVs. Bottom left is CFPS with deGFP plasmid. Bottom right is CFPS with no plasmid.

The optimised protocol resulted in a significant increase in GUV creation in deGFP, BSA and CFPSs GUVs (**Figure 5.14**). The tile scans (5x) of the wells show a difference between the GUVs containing pure protein and CFPS (**Figure 5.14**). The GUVs containing purified proteins were larger and distinguished vesicles, whereas the CFPS GUVs were smaller in size. The difference between the deGFP and BSA in terms of fluorescence was clear. This means the fluorescence difference will be significant enough to differentiate on a plate reader, increasing this technique's ability to drug discovery. Unfortunately, the CFPS reaction failed in the vesicles and failed when in a plate reader. The CFPS has many fragile components with a limited life span. By the time this experiment was undertaken, the reaction components were ~two years older

than from the initial preparation. Due to the limited time and financial constraints, more lysate was not purified, and the components were not repurchased. Instead, the focus of this work shifted towards creating GUVs with asymmetric lipid bilayers.

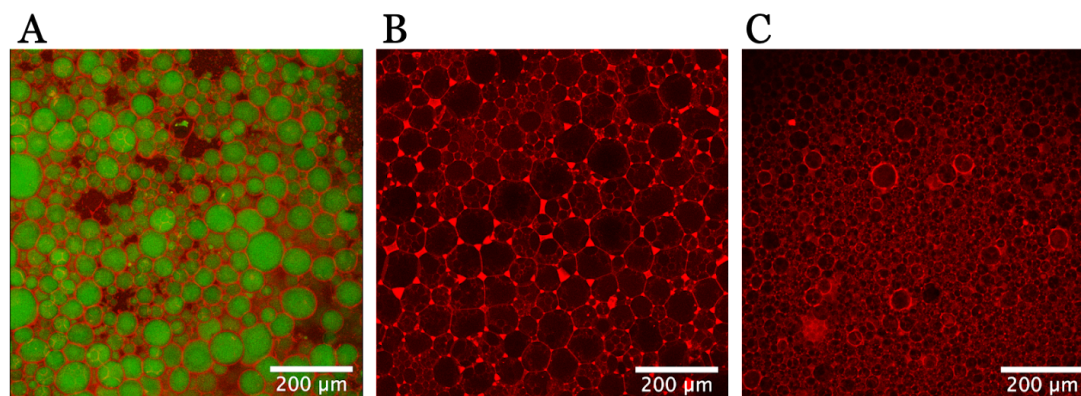


Figure 5.15A-C The vesicles formed from the optimised protocol at 10x magnification. 5.15A (left) GUVs containing purified deGFP, 5.15B (middle) GUVs containing BSA, 5.15C (right) GUVs containing CFPS reaction

Upon a greater magnification, the degree of variability in the vesicles of CFPS versus pure protein is evident (**Figure 5.15**). There also appears to be some variation between the BSA and the deGFP GUVs. Ultimately, the optimisation of the protocol (guided by¹⁴⁷ paper) has resulted in a much greater number of vesicles.

The optimised technique resulted in many liposomes. Most importantly, these liposomes contained an asymmetrical bilayer that mimicked *E. coli*. The outer leaflet was composed of KDO2-Lipid A, and the inner leaflet contained the *E. coli* polar lipid extract. The creation of an asymmetric bilayer is an improvement on POPC lipid. However, the ratio used is only a snapshot of phospholipid compositions that *E. coli* could have.

Furthermore, the diversity of phospholipid composition in the outer membrane between Gram-negative bacteria species is huge and has not been addressed by this work⁵². In addition, the lack of membrane proteins will also affect the interactions in the membrane. In the future work, it is hoped that membrane proteins can be expressed, therefore better modelling the outer membrane of *E. coli*. The asymmetry of this bilayer may not have been maintained throughout this experiment. In bacteria, there are flip/floppases that maintain lipid asymmetry²⁵⁹. To better understand the asymmetry this should be explored with fluorescent lipids.

5.3.7 Creating GUVs with asymmetric bilayers

Fluorescently labelled phospholipids NBD-PG and marina blue-PE (1%) were added to the inner and outer leaflet. The aim was to form biologically relevant GUVs and use the fluorescent groups to visualise the interface of the asymmetric bilayer.

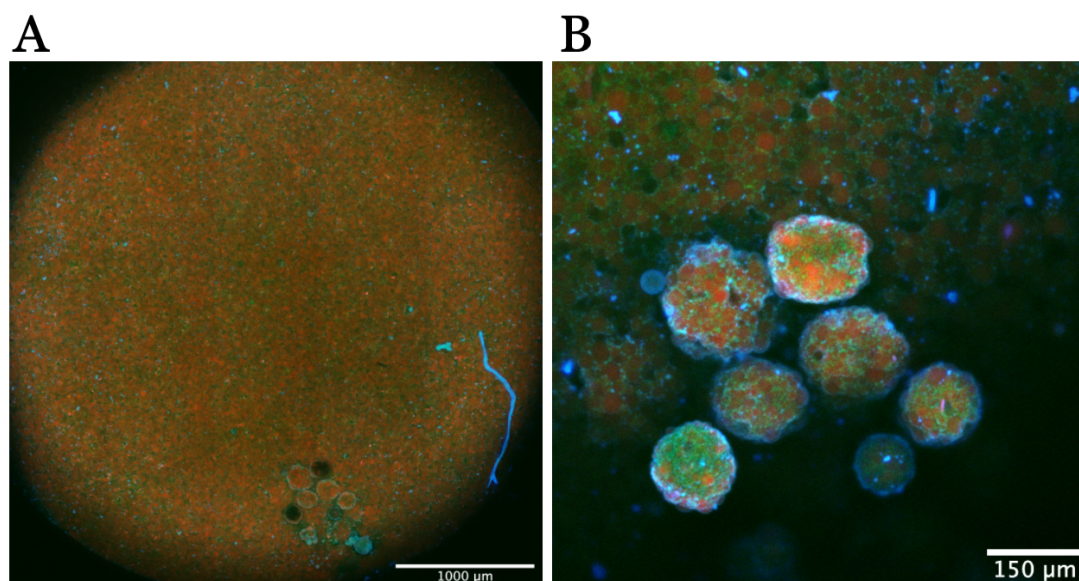


Figure 5.16A-B GUVs that have fluorescent lipids integrated into the asymmetric bilayer. **5.16A** (left) is a 5x magnification tile scan of GUVs containing mPlum and encapsulated in a complex phospholipid mix. **5.16B** (right) is a 10x magnification image on the lipid aggregation seen on the bottom right of **5.16A**.

As NBD-PG was used, the fluorescent wavelengths would interfere with deGFP, so mPlum was encapsulated instead (kind gift from M. Sambrook). The GUVs formed were smaller, and more lipid aggregation was seen (**Figure 5.16**). Giant aggregated and distorted GUVs formed²⁶⁰ (**Figure 5.16**), whereby the phase separation of lipids occurs due to charge and osmosis, leading to remodelling of the membrane. These misshapen vesicles could be due to the charge of the lipid; the system could be optimised to adapt to combat this with pH and salt – since the technique can withstand changes to these variables, however, CFPS will not be able to withstand these adaptations¹⁴⁷. It has been suggested that more lipid aggregation occurs due to centrifugation at higher speeds; by reducing the g force, lipid aggregation could be reduced.

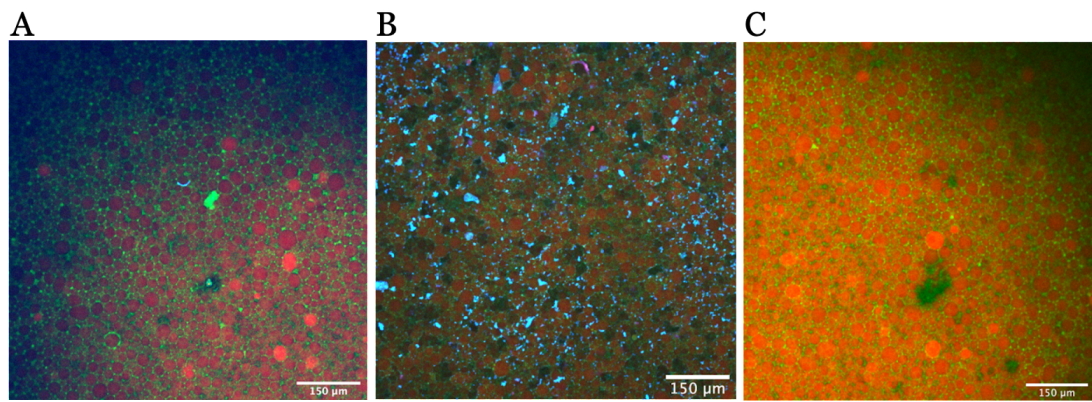


Figure 5.17A-C A closer look at GUVs that have fluorescent lipids integrated into the asymmetric bilayer. 5.17A (left) A 10x magnification demonstrating only one vesicle containing marina blue PE. **5.17B** (middle) GUVs contain marina blue and marina blue aggregation. In **5.17C** (right) no marina blue PE appears in the membrane.

GUVs containing *E. coli* phospholipids in a realistic composition can be formed (**Figure 5.17**). Using fluorescent phospholipids to visualise GUVs resulted in aggregation. To use these phospholipids, more investigations into aggregation and blebbing of the GUVs are warranted. The CFPS reaction contains magnesium ions which may interact with the anionic phospholipids PG and CL in the membrane, causing fusion²⁶¹. The fluorescent lipids may have different transition temperatures; this may have caused phase separation leading to aggregation.

5.3.8 GUV calcein leakage

As the 96-well microplate, the inverted emulsion method can form GUVs that have an outer membrane mimetic; this platform was used to look at antimicrobial peptides (AMP). The calcein quenching assay has frequently been used to study AMPs and leakage from vesicles in general. Calcein is fluorescent at Ex./Em. 495/515 nm is self-quenched at concentrations greater than 70 mM²⁶². Some researchers have performed experiments where the calcein is self quenched inside vesicles, and then as it permeates out (via membrane disruption or pores), it fluoresces outside the vesicle²⁶². Others have used lower concentrations of calcein (1 mM) and observed the loss of fluorescence as calcein leaks from the GUVs^{148,149}. The latter approach was trialled in this work.

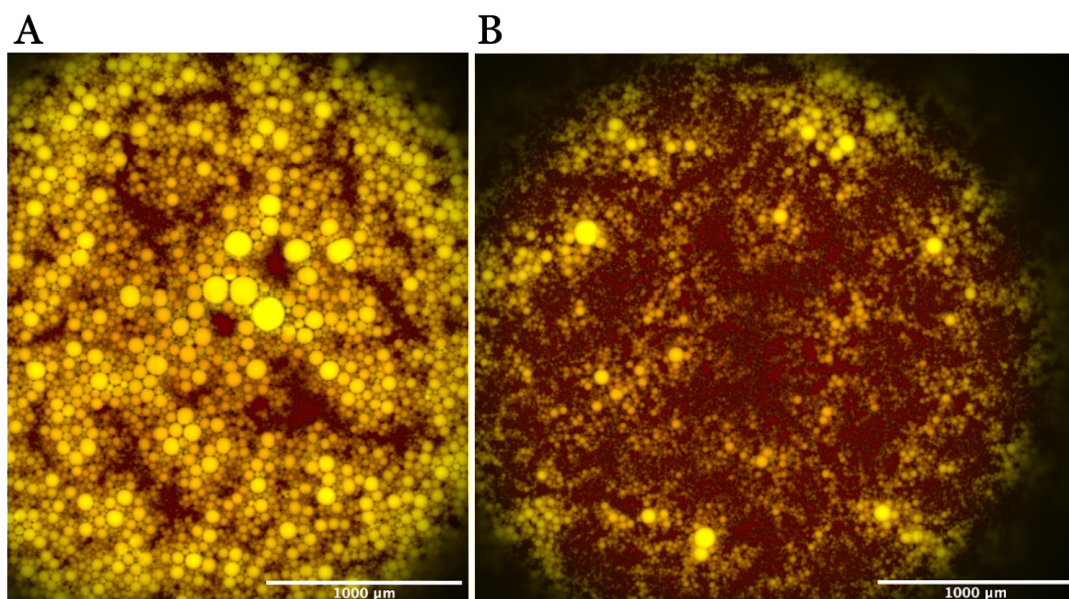


Figure 5.18 Calcein leakage assay performed on GUVs created using the 96-well microplate method. Both **5.18A** and **B** are composed of a tile scan created from a 6 image Z-stack of each tile stitched. The yellow in this work is from calcein fluorescence (recoloured to yellow for figure). Dil iodide membrane stain was used as before. **5.18A** (left): GUVs formed from *E. coli* phospholipid on the inner leaflet, and KDO2-lipid A on the outer leaflet of the bilayer, encapsulating 1 mM calcein. **5.18B** (right) On the right: GUVs of the same composition, also encapsulating 1 mM calcein, but the outer solution (surrounding the GUVs) contained 50 μM polymyxin.

The *E. coli* outer membrane GUVs demonstrated good encapsulation of calcein, suggesting this technique can be used with a range of reporters (**Figure 5.18**). The vesicles on the right (**Figure 5.18**) contained 50 μM polymyxin B, a cationic antimicrobial peptide that disrupts the outer membrane of Gram-negative bacteria. The concentration of polymyxin was chosen to be consistent with previous work (**Figure 5.13**). In this experiment, there seemed to be significantly fewer GUVs. The results suggest that polymyxin has disrupted the GUVs resulting in the lack of liposomes. This initial work suggests that this technique could be used to screen AMPs for activity. The lipid composition could be altered to mimic other bacteria or the inner membrane to expand this work. This technique could be applied to CFPS, and the synthetic cells could produce the AMP, and if it were effective, there would be fewer vesicles and calcein fluorescence.

5.4 Discussion and Future Prospects

The work in this chapter has met some of the outlined aims of this work. A bottom-up approach to forming synthetic cells has been established in a 96-well plate. These

synthetic cells can transcribe and translate genes to synthesise proteins, and protein-synthesis directed antibiotics inhibit this process. The technique, in general, can be optimised and adapted to analyse the impact of variation of several cellular parameters. For instance, the lipid layer was adjusted to make an outer membrane mimetic, or other proteins can be synthesised. The adjustable nature of this technique means there are many potential uses.

However, the generation of synthetic cells needs optimisation. Currently, the encapsulation of the CFPS reaction is lower than that of purified proteins. The poor encapsulation of CFPS is most likely due to osmolarity imbalances; therefore, using an osmometer to balance the inner and outer solution properly would improve encapsulation(31). And therefore, the reliability of this technique.

In general, the variation of vesicles is considerable, and this is where microfluidic techniques are advantageous as they can control the creation of GUVs. In addition to this, microfluidic devices have been able to move away from the use of mineral oil due to the design of the chips and can now use other carriers. The 96-well plate technique still uses oil, which could have biological implications, resulting in artefacts such as oil in the membrane^{263,264}.

A microfluidic platform was created that encapsulated fluorescent dye in vesicles, which was released and quenched when exposed to antimicrobial peptides (AMPs)²³². A similar approach could be taken by using the synthetic cells created by the microplate method. The synthetic cells could either encapsulate AMPs or the DNA sequence for the AMPs, which, when expressed, would cause the disruption of the vesicles and then leakage and a subsequent change in fluorescence. This adaptation would allow the investigation of potential AMPs. However, the microplate method of generating GUVs cannot control the size of GUVs or the ability to monitor individual vesicles, whereas microfluidic methods can create homogenous vesicles and monitor individual GUVs^{265,266}.

If more time and resources were available, more work on quantifying the permeability of antibiotics into synthetic cells would have been done. The original plan for quantification was to use software that analyses microscopy images using machine learning algorithms, such as “Vesicle analyser pro” software published by Hjartarson 2019²⁶⁷ or Lee 2022²⁵⁴. Flow cytometry would also have been a good option. Flow

cytometry would have allowed the number of cells and fluorescence to be quantified. Flow cytometry would be able to distinguish and count the number of empty vesicles^{249,255}. Flow cytometry could allow us to better understand to what degree protein synthesis occurs and its inhibition.

Initially, there was an aim of incorporating outer membrane proteins. The initial idea had intended to use multiple plasmids. However, there were many obstructions to this. Firstly, the insertion of OmpF into membranes is still widely discussed. Due to its stability, some believe it could self-insert into the vesicle membrane²⁶⁸. However, others believe many proteins, such as the BAM complex, are involved in its insertion²⁶⁹. The expression of OmpF within cells is further complicated by the capacity of *E. coli* based systems to fold membrane proteins²⁷⁰. OmpF in its natural state has a membrane tag²⁶⁹, which is cleaved. However, the CFPS system cannot cleave the tag. The tag could have been omitted in cloning, but this may have inhibited membrane insertion via BAM²⁷¹. Moreover, the inclusion of BAM brings about more questions; some of the complex is reportedly present in CFPS lysate; however, to what degree is uncertain. In general, *E. coli* lysate contains many chaperones that enable protein folding of soluble proteins²⁷⁰. However, large membrane proteins are more difficult to express; for example, YiaD an outer membrane protein failed to be expressed in CFPS but was expressed in C43 cells using traditional recombinant protein expression purification techniques²⁷².

There has been some success in membrane protein purification by expressing the protein in the CFPS and then diluting the purified protein into liposomes²⁷³; this technique has allowed for the refolding of tetrameric channels into liposomes²⁷³. To apply this technique to OmpF optimisation surrounding OmpF purification, refolding and dilution into membranes would be required. The use of a purified recombinant protein was planned, and recombinant OmpF was expressed and purified. However, as OmpF is a membrane protein, a detergent would be necessitated, further complicating the technique. Unfortunately, as time was limited, this was not started.

The reporter used in this work was deGFP, and deGFP is a stable protein, especially in this cell-free system which should lack some proteases like OmpT. Therefore, using a deGFP with a degradation tag such as SSrA²⁵³ would have been helpful at the quantification stage, as it would indicate real-time expression. The creators of the p70a-deGFP plasmid have done extensive work looking at the system's kinetics²⁴⁴ and

deGFP production, and calculating degradation²⁵³. Based on their extensive body of work, mathematical modelling of protein synthesis inhibition would be possible.

Alternatively, plasmids encoding the reporters of assays such as the luciferase assay could be expressed. If the luciferase enzyme is expressed, it should be able to convert D-luciferin to oxyluciferin, which is luminescent²⁷⁴. Dihydrofolate reductase is another reporter that could have been expressed. Dihydrofolate reductase catalyses the reduction of dihydrofolate into tetrahydrofolate using the oxidation of NADPH, resulting in increasing levels of NADP⁺²⁷⁵. The increase in NADP⁺ levels decreases the signal at 340 nm, meaning the enzyme's activity can be followed spectrophotometrically²⁷⁵. Therefore, compounds that reduce the expression of dihydrofolate reductase or inhibit the enzyme's activity - like trimethoprim can be screened for in a spectrophotometric assay. Ultimately, this reiterates the critical benefits of this system; the target can be adapted based on the drug screen.

5.5 Conclusion

Although there is a great deal more optimisation necessary, the potential applications of synthetic cells in drug discovery are enormous. The lipid layer can be altered to mimic the compositions of different membranes. Hopefully, protein channels can be introduced. Perhaps even peptidoglycan could be made by the cell if lipid II was generated by either the encapsulated pathways responsible for its synthesis or if lipid II itself was incorporated into GUV construction and CFPS encapsulation. Many potential avenues of bacterial cell walls could be explored. More optimisation into building a robust CFPS system would allow this technique to be used as a screening method. The use of flow cytometry would give us a better indication of quantification, which is crucial for permeation. Nevertheless, if this technique was optimised, it could provide a tandem permeability activity screen which is desperately needed to save money and streamline the antimicrobial drug discovery pathway. Unlike the previous two chapters, this chapter is a broad spectrum approach to permeability. The aim of this work was to create a permeability chapter that could be applied to all antibiotics - not just β -lactams.

Chapter 6. Machine learning to predict antibiotic permeability

6.1 Background

6.1.1 Machine learning in medicinal chemistry

Modern drug discovery is hugely expensive, meaning new drugs that do not demonstrate long-term profitability are not developed²⁷⁶. The lack of investment has led to the antimicrobial drug pipelines “drying up”. Novel approaches to drug discovery are necessary to reinvigorate the drug development pipeline. Artificial intelligence (AI) infiltrates all scientific exploration and exploitation, innovating how we approach many problems such as speech recognition, recommendation engines and self-driving cars. Furthermore, machine learning is more recently and frequently being used to aid in drug discovery. The surge in using machine learning algorithms in this application is revolutionising medicinal chemistry (MedChem)²⁷⁷. Previous attempts at integrating AI, in the form of deep neural networks to the field of MedChem, had failed^{278,279}. Nevertheless, with the advent of more complex algorithms, large open-source databases, and advanced technology^{277,279,280}, AI offers the hope of reinvigorating MedChem²⁷⁹.

Machine learning is a form of artificial intelligence that uses large datasets and mathematical algorithms to tackle previously unsolved problems^{281,282}. It encompasses a broad range of algorithms; it also includes the field of deep learning. Machine learning is being employed in many areas of drug development, including antibiotic research²⁸² (**Figure 6.1**). Recently a deep learning technique was utilised to discover a novel antibiotic called Halicin by Stokes *et al.* 2020¹¹¹. Many other groups use machine learning algorithms to identify and find novel antimicrobials^{283,284}. In particular, many researchers are looking at antimicrobial peptides²⁸⁵. Groups use machine learning models to predict antimicrobial resistance genes from whole-genome sequencing and metagenomics^{286,287} alongside work to predict the mechanism of action of

antimicrobials²⁸⁸. Moreover, most relevantly, groups have used various forms of machine learning to look at the permeability and accumulation of antibiotics^{89,103,104}.

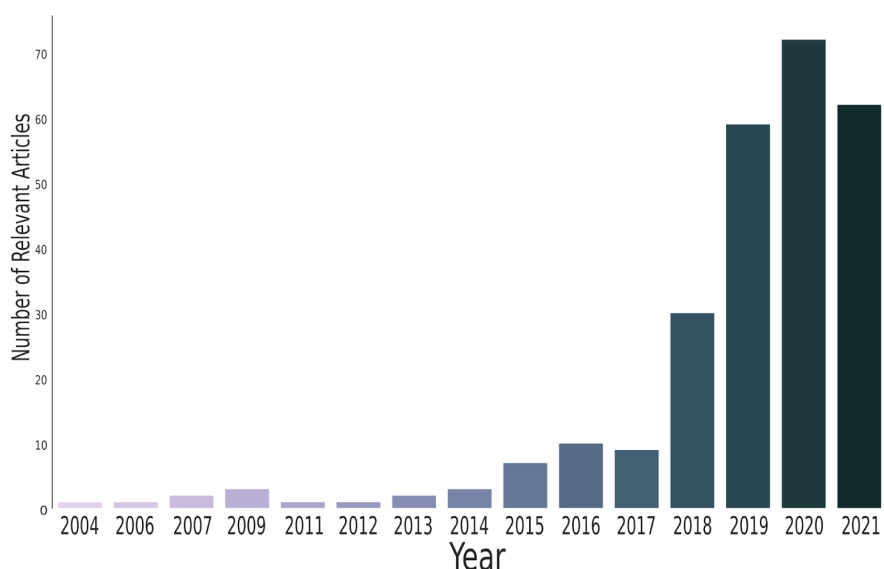


Figure 6.1 The boom of machine learning in relation to antibiotics. PubMed search “machine learning antibiotic” returned 262 articles on the 10th of August 2021, demonstrating a recent boom in the field.

6.1.2 Key concepts in machine learning

In this chapter, machine learning techniques aid in antimicrobial drug discovery and the elucidation of permeability patterns. To better describe the specifics of machine learning, some definitions will be briefly described in **Table 6.1**, and these terms are contextualised in **Figure 6.2**. Following that is an overview of crucial concepts in machine learning. In-depth descriptions of the algorithms, particularly the equations, have been omitted.

There are two main learning techniques **in machine learning: supervised and unsupervised learning**²⁸¹.

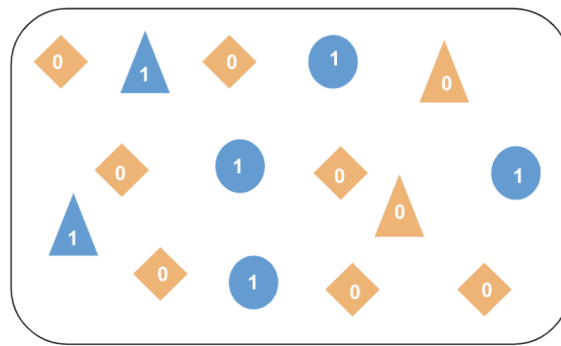
Supervised learning techniques tend to be used to fit predictive/classification models. This technique requires labelled data as the input; techniques in this category model the interconnections between the features of the dataset and the output target, *i.e.* the class²⁸¹. **Unsupervised learning** is used to identify patterns in data. It does not require labels as these techniques are not used to classify data, but to cluster it. Unsupervised learning discerns the concealed patterns of a dataset without labels. Unsupervised learning is mainly involved in pattern recognition²⁸¹.

Within these different learning techniques, there are hybrid problem solvers such as *semi-supervised learning*, a combination of supervised and unsupervised learning, using unsupervised methods to fill in for missing labels to complete supervised learning algorithms.

Term	Definition	Example
Model	The output of an algorithm trained on a dataset. It contains both the information on processing and the data itself	
Target	The final output or 'answer' for models to learn (in supervised models). In unsupervised models the target could be a label.	In drug discovery the final output could be 'Active' or 'Inactive' in a classification problem. When predicting solubility – a regression problem, the target is a continuous value.
Feature	These are the variables from which the model learns, and uses them to make predictions. These are also known as 'attributes'. These can be considered the independent variables.	In drug discovery the features of the dataset of drugs might include; hydrophobicity, charge, and other physicochemical descriptors.
Algorithm	The process used to solve the question at hand. Used to 'learn' and 'fit' data.	Examples of algorithms include; Linear Regression, Decision Trees, K-nearest Neighbours.
Hyperparameter optimisation	The process of testing a range of parameters on the efficacy of model using a validation dataset	Parameters can include number of neural layers, learning rates, and momentum.

Table 6.1 A Glossary of key terms. These key terms are further elaborated upon below in Figure 6.2

The Dataset



It contains different **features** such as:

- Three different shapes ○ △ ▢
- Two colours ■ ■

The data here is **labelled** - meaning it is a **supervised learning** problem. It is labelled either 0 or 1. This is a **binary classification**

The goal of creating a **model** based on this data would be to classify unknown data to the **target** either 0 or 1.



1 Splitting Dataset

The dataset is split into a training dataset, a test dataset and validate dataset

2 Fitting an Algorithm

An appropriate **algorithm** is selected. An **algorithm** is used to fit the data. This is a mathematical model used to associate the **features** to the **labels**

3 Model is Created

A **model** is the outcome of training an **algorithm** on a dataset. A **model** can be used to make decisions on unknown data - assigning **targets**.

4 Hyper Optimisation

Efficacy of the **model** is optimised using the validate dataset

Figure 6.2 Contextualising the glossary terms used in ML. At the top is a rectangle of shapes - this is the dataset. This dataset contains different features used to train a binary classification. In the bottom half of the figure is a general workflow used in machine learning.

6.1.3 Algorithms

Different machine learning algorithms have different uses based on input, output, and function. Briefly, some of the key algorithms used in this work are discussed below:

1. Decision trees have been around since the 1960s; they are sequential algorithms, akin to flowcharts, composed of nodes, branches, and leaves^{289,290}. The model splits and tests data on various attributes at nodes, and the threshold at each node determines the outcome and whether the attribute tested will lead to further branching and tests or return a classification.

2. Random forest (RF) models were first described by Breiman in 2001²⁹¹. They are similar in structure to decision trees, but are an ensemble approach, so instead of having one decision tree it is a collection of decision trees²⁹¹ (**Figure 6.3A**). They are an improvement upon decision trees as they offer more variability, creating models that tend to learn better on new data²⁹². The initial input data is then randomly selected to make 'bootstrap' datasets²⁹². A decision tree is built using a random selection of the features of a dataset (usually, the number of features selected is derived from the square root of the total number of features in the dataset)²⁹². The bootstrapped data is fed into the decision tree; this process is repeated for each tree in the forest²⁹². In this chapter, the random forest was built from 100 decision trees. The predictions used in the final classification come from the majority results of all the trees in the forest²⁹². The accuracy of the prediction is assessed by monitoring if the decision tree can correctly predict the data left out of the bootstrap dataset (the 'out-of-bag' data)²⁹².
3. Artificial Neural networks were created to mimic the way the brain works, with a series of neurons that feed into each other. There are a few different types of neural networks, the main one being the multilayer perceptron (MLP)^{293,294}. In the MLP, the input value is entered into an input layer of neurons (the number of neurons corresponds to the number of features)²⁹². The input value is transformed through a series of weights and biases²⁹². After the input has been multiplied by a weight value and has a bias value added, it is fed into a hidden layer with an activation energy²⁹². The activation energy transforms the value based on the activation function²⁹². In this work, the activation functions used were rectified linear unit (ReLU) and the hyperbolic tangent activation function (tanh). The sum of the outputs from the activation functions of all the neurons in hidden layers is calculated, and another bias function is applied before the value is fit and the output is generated²⁹². The weights and biases are iteratively adjusted (backpropagation) to best fit the data by measuring and reducing loss in a process called gradient descent (**Figure 6.3B**)²⁹².

In this chapter, two of the above mentioned algorithms (RF and MLP) and a deep learning algorithm (discussed later) were used.

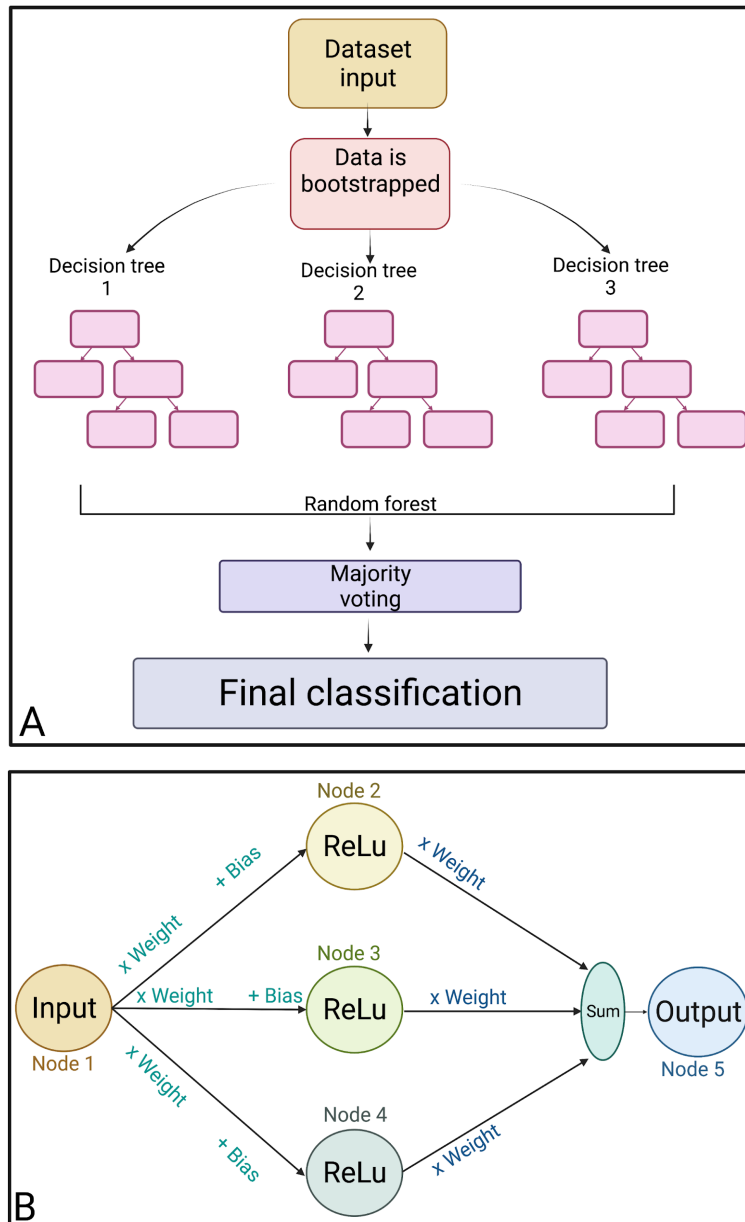


Figure 6.3A-B An overview of two of the machine learning algorithms used in this chapter. (A) - Random forest model, In this example there are 3 decision trees calculating the probability of each class, the majority of these decision trees output is taken as the final classification. Figure **6.3B** demonstrates the architecture of a simple neural network. This neural network has 1 input neuron fed into a hidden layer that contains 3 neurons. Each neuron here has a Rectified Linear unit (ReLU) activation function. The sum of the outputs of the three neurons in the hidden layer feedforward to the output. The input value is transformed by a series of weight and bias transformations throughout the learning process; these weights and biases are altered throughout the process to better fit the data using gradient descent function.

6.1.4 Deep learning

Deep learning is a form of neural networks that uses larger datasets, and larger numbers of neural layers in order to decipher and solve problems. Deep learning can encompass the use of the MLP, but also more complicated models such as recurrent neural networks (RNN), convolutional neural networks (CNN), Generative adversarial networks (GAN), and graph convolutional networks (GCN)²⁸¹. The use of Deep Learning is having a renaissance due to increased datasets, data sharing, and increased computational power, allowing better models to be created and used²⁷⁸.

6.1.4.1 Graph Convolutional Networks (GCN)

Graph convolutional networks use the architecture of convolutional neural networks, meaning that instead of the traditional layers of an MLP, there are layers of convolution and pooling to reach classification at a gathering layer^{152,295}. GCNs have a significant and emerging role in drug discovery as they are well suited to embed molecular information^{279,282}. In a molecule, the atoms will become the 'nodes' of the graph, and the bonds will be the 'edges'^{152,280,296}. 'Feature matrices' are generated based on the chemical structure of the compound²⁸⁰. These 'feature matrices' are based on molecules' atomic features and spacing. The information on spacing is generated by convolutional layers that aggregate information on neighbouring atoms²⁹⁵. The feature matrices generated can be huge, and thus a pooling layer is applied to pull out the most prominent features²⁸⁰. Through aggregating and pooling information, the data dimensionality is reduced to critical points to inform classification²⁸².

6.1.5 Data engineering

Before a model can learn and make predictions on a dataset, a data engineering step has to occur. Data engineering can involve two main categories of work, data preprocessing and feature engineering. Data preprocessing can involve steps such as: reformatting the dataset into a comprehensible format, removing null data, and removing unnecessary data (de-noising). Dimensionality reduction algorithms like

Principle Component Analysis (PCA) can be used to investigate the dataset's variability and reduce some features²⁹⁷.

6.1.5.1 Feature engineering

Feature engineering is the process of translating the input into a readable format; for example, if the model's input is a picture, then the picture is converted into a numerical matrix, where each pixel has a value²⁷⁹. Feature engineering can also be broadly categorised into two forms; user picked features and learnt features (learnt via deep learning algorithms)²⁷⁹.

To embed chemical structures with some molecular representations for machine learning, two essential libraries were used; RDKit¹⁵⁴ and DeepChem¹¹¹. RDKit can embed meaning from general chemical inputs like SMILES codes, translating them to a "mol" object, from which features can be interpreted¹⁵⁴. There are many functions in the library that allow **user-defined features** to be pulled-out, such as: the number of different chemical groups, the number of rings and the quality of the fit to Lipinski's rule of 5²⁹⁸. These physicochemical descriptors can be used as the input features to train a model, this method was used by both Richter *et al.*⁸⁹ and Zgurskaya *et al.*¹⁰³, who created models to establish rules surrounding permeability.

Although **user-defined features** were used in this work, the primary way molecular representations were derived was by using **learnt features** in three ways: Extended Connectivity Fingerprints¹⁵¹ (**ECFPs**), **Mol2Vec**¹⁵⁰, and **Graph representations**^{152,279,296}.

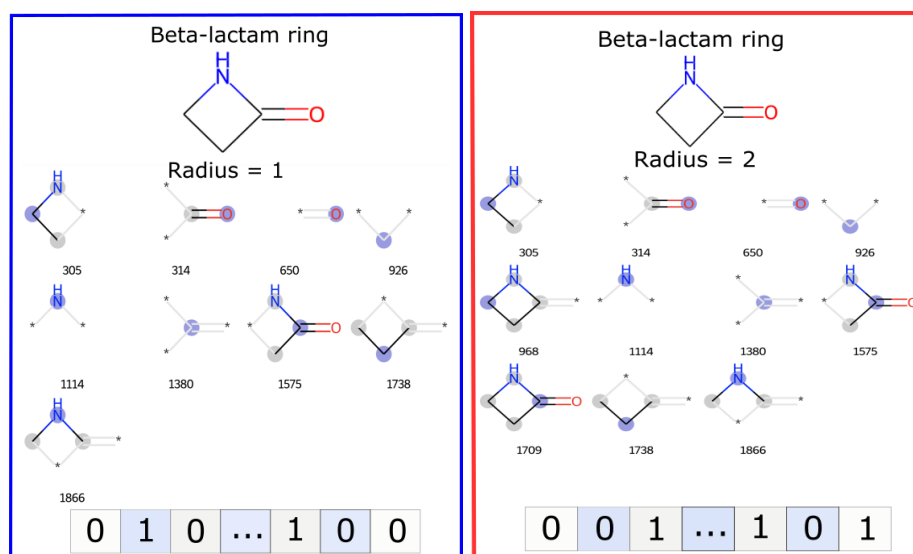


Figure 6.4 a β -lactam ring featurised using ECFP. In the blue box using a radius of one, and in the red box with a radius of 2. These substructures are given a calculated value, which is then embedded into a bit vector (at the bottom). The bit vector length used in this work was 1024 and the radius used was 3. By increasing the radius around each atom it increases the computational permutations of the substructures, meaning there is more embedded information of the compound.

The RDKit program can also generate ‘fingerprints’ for molecules. ECFP molecular representations are known as “bag-of-fragments” as the key substructures of a molecule are encoded based on a defined radius, generating fragments of a molecule^{151,279}. To summarise the generation of **ECFPs**, firstly, the user defines the radius (from an atom) used to define substructures, then each substructure is identified and given an identifier (numerical value)^{151,279}. These identifiers are based on a number of properties such as valency, hydrogen bonding, atomic number, atomic mass, atomic charge, and whether the atom is aromatic¹⁵¹. These identifiers are iterated (updated over the search) based on other substructures²⁷⁹. From the list of identifiers generated, the repeated values are removed¹⁵¹. The list of identifiers is then converted into a bit vector (a string of 1s and 0s), and the length of this vector is usually 1024 or 2048 bits¹⁵¹. **ECFPs** are frequently used in MedChem, particularly in similarity searches.

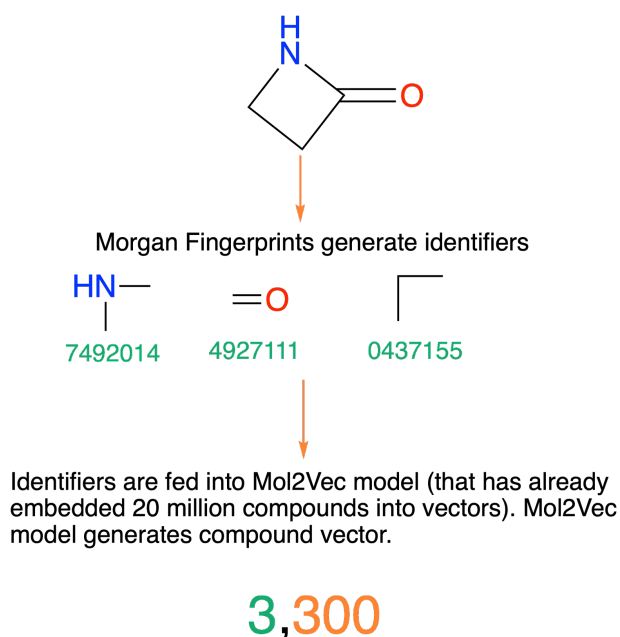


Figure 6.5 Mol2Vec featurisation of the β -lactam ring. In this example, the β -lactam is converted into a series of substructures using Morgan Fingerprints with a radius of 1. In this example three ‘chemical words’ are generated and each has an identifier (green number). These identifiers are fed into a pre-trained Mol2Vec Model, based on the identifiers the model returns a compound vector (3,300), in this case it is a molecular sentence made of 3 “chemical words”, and these words have been embedded in a 300 dimension vector.

Another molecular representation used in this work was the **Mol2Vec** algorithm (**Figure 6.5**). **Mol2Vec** is a model-based natural language processing (NLP) technique based on the Word2Vec algorithm¹⁵⁰. **Mol2Vec** algorithm was trained on the Morgan Fingerprints¹⁵⁵, as these are very similar to ECFPs (ECFPs are derived from Morgan Fingerprints)¹⁵¹. The substructures generated from the Morgan fingerprints are known (in this application) as ‘chemical words’, and the compounds are known as ‘chemical sentences’¹⁵⁰. These ‘chemical sentences’ are trained on the **Mol2Vec** algorithm, which groups ‘chemical words’ together based on the distance between the embeddings and returns a 300 dimension vector, embedding the chemical structure of a compound¹⁵⁰. The vector is the molecular representation of the whole molecule. The **Mol2Vec** ‘corpus’ (model) has been trained on twenty million compounds and has learnt the similarities and differences between each substructure based on the distances between the ‘chemical sentences’¹⁵⁰. The user has the **Mol2Vec** model create ‘chemical sentences’ for the compounds in their dataset and uses these representations to train their model¹⁵⁰. The advantage of this embedding over traditional fingerprints is that the model has learnt more about the similarities and differences of substructures so that it may have more information embedded.

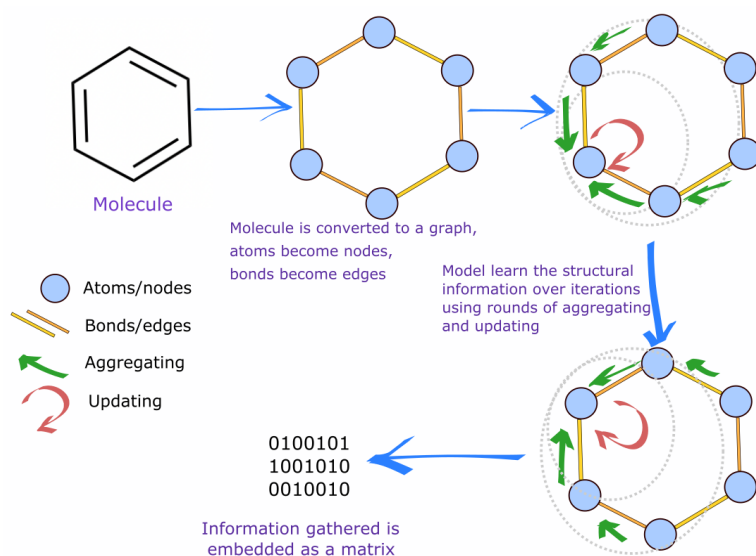


Figure 6.6 Graph featurisation of a benzene ring. The atoms become nodes, and the bonds become edges. As the model trains, it learns about the structure of the molecule by aggregating information about the neighbouring atoms, eventually resulting in a final matrix.

Finally **Graph** representations were also used to generate molecular representations from SMILES of compounds in the dataset. Graph representations were generated for a **Graph** Convolutional Neural Network, by the program DeepChem¹⁵³, based on Duvenald *et al.* 2015¹⁵². Graphs in GCNs are composed of edges and nodes^{152,279,280}. Molecules can be translated into graphs by using the atoms as nodes and bonds as edges (**Figure 6.6**)²⁸⁰. These graph representations are learnt during model training^{152,279,280}. In the GCN model, the nodes and edges of the molecule are then embedded in two ways, first as a 'feature matrix' and an 'adjacency matrix'²⁸⁰. The features identified are based on 75 features involving atom type, number of bonds, number of valencies, charge, hybridization and aromaticity²⁹⁹. The adjacency matrix encodes the spatial information learnt from the graph. In both of these matrices, the details are usually "one-hot-encoded" - if the feature is present, a value of 1 is awarded, and if it is absent, a value of 0 is given. The two matrices containing information about the molecule and its spacing are combined.

6.1.6 Evaluating models

One of the first stages of any model learning is the splitting of the data (**Figure 6.2**), occurring after null data is removed but before feature engineering and any data

normalisation. Datasets are split into training, test and validate datasets. In the learning process, the model is fitted on the training dataset and evaluated on the test dataset³⁰⁰. The test set scores help assess how generalisable a model will be on a new dataset³⁰⁰. However, to optimise a model (hyperparameter optimisation), a validation set is also needed³⁰⁰. Validation scores are more indicative of the model's generalisability³⁰⁰. If the training dataset is small, a method of cross-validation known as k-folds can be used, where data is divided into k non-overlapping sets, trained on $(k-1)$ sets, evaluated on the remaining set, and then repeated for each of the k sets³⁰⁰. K -fold cross-validation generally reduces bias³⁰¹.

6.1.6.1 Metrics

To evaluate the function of machine learning models there are different approaches suited for different algorithms, due to output differences²⁸¹. Key evaluation metrics for classification problems include accuracy, F-scores, recall, confusion matrices, and receiver operator curve – area under curve (ROC-AUC).

- Accuracy is the percentage of correct predictions out of all predictions, this is a useful metric when assessing a dataset with balanced targets³⁰².
- Precision measures the number of times the model correctly predicts a positive outcome³⁰². It reports how many positive results are actually positive results. It is calculated by dividing the true positive rate by the sum of false positives and true positives.
- Recall calculates the number of positive predictions that were correctly assigned³⁰². It is also known as the true positive rate, it is calculated by dividing the number of true positives, by the sum of the true positives and false negatives.
- The F1-score is the harmonic mean of precision and recall scores³⁰².
- ROC-AUC demonstrates the sensitivity of the binary classification model. Plots the rate of true positives against the rate of false positives³⁰². This metric can be plotted and visualised as ROC-AUC curves. (**Figure 6.7b**)
- Confusion matrices summarise the predicted results of a classification problem (**Figure 6.7a**). They allow easy, class-by-class visualisation of predictions.

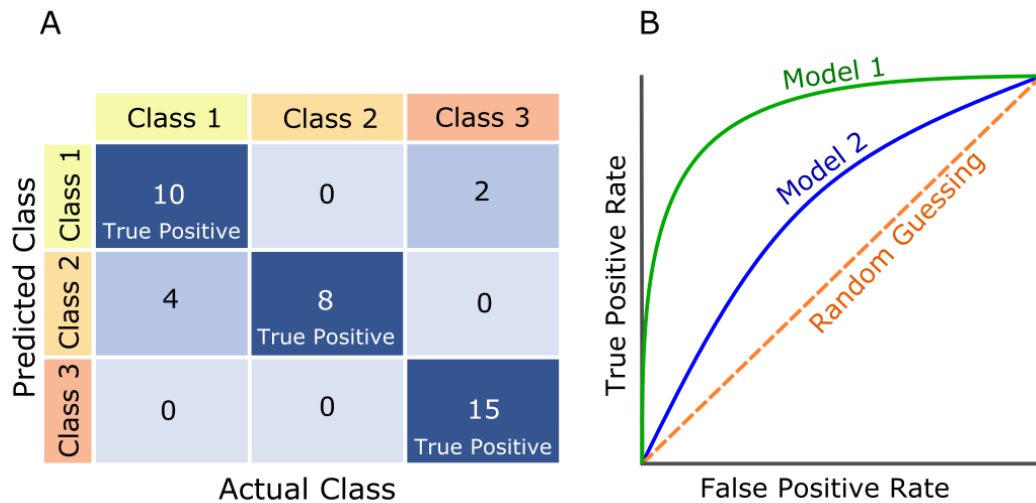


Figure 6.7 Overview of Confusion matrices and ROC-AUC curves. In 6.7A; an example of a multiclass confusion matrix. In the confusion matrix, the true positive – accurate labels are along the diagonal (darkest blue) and the other misclassified classes are in other boxes In 6.7B; is an ROC-AUC curve that is comparing two models. Model 1 is performing better than model 2 but both are better than randomly assigning classes.

6.1.6.2 Bias – Variance Trade off

Machine learning algorithms aim to achieve ‘generalisability’, meaning a model scores highly on both the training data, and on new unseen data^{297,300}. A generalisable model needs to find a balance between *bias* and *variance*^{300,303,304} (Figure 6.8).

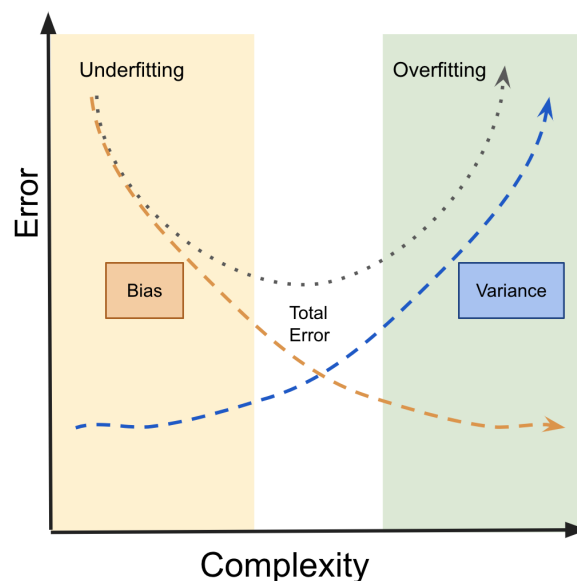


Figure 6.8 Building a generalisable mode. Figure 6.8 is adapted from Kocak *et al.* 2021²⁹⁷. To build a generalisable model, a balance between bias and complexity is sought. This figure illustrates the tradeoff between the two. The ideal balance between error and complexity is in the white box.

Bias is how many incorrect predictions are made; models with high **bias** will frequently misclassify predictions³⁰⁴. As a result, the accuracy for the training and test datasets (unseen data) is low^{284,297}. In contrast, **variance** is the amount of complexity or noise in a class³⁰⁴. Models can become overly biased and underfit the data, meaning the pattern detected to fit the data is oversimplified and imposes an overly simplistic and inaccurate trend (**Figure 6.9**)³⁰³⁻³⁰⁵. High variance can occur when not enough features or data, in general, is provided for the model. As a result, the model has to solve a problem with insufficient information to discern the whole picture. The degree of **bias** in a model is visualised nicely by a learning curve (**Figure 6.10**)²⁹⁷.

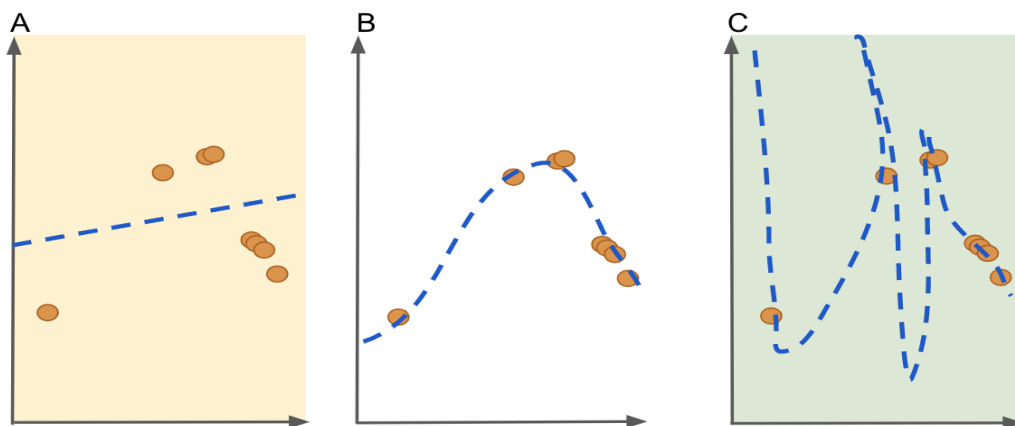


Figure 6.9A-C Model Underfitting vs overfitting. Figure 6.9 is adapted from reference³⁰⁰. In 6.9A the model is underfitting, the trend line (blue) used to make predictions does not account for the nuances of the dataset. In 6.9B the model has established a general trend and should make accurate predictions on new data. In 6.9C The model is overfitting, each datapoint is explicitly fit, meaning it has not learned a general trend and therefore will struggle with new data.

6.1.6.3 Overfitting

Conversely, the opposite, overfitting can also occur when a model has a high degree of variance (**Figure 6.9**)³⁰⁵. Here the model does not fit the general trend of the data; in fact, it fits the explicit points of a dataset and therefore performs well on the training dataset but performs poorly on new data (**Figure 6.9**)³⁰³. A generalisable model has an adequate balance of the two (**Figure 6.8**)³⁰³

In order to assess whether a model is underfitting or overfitting, two techniques can be used; **validation curves** and **learning curves**³⁰⁰. **Validation curves** allow us to

visualise the difference between training and validation scores; it allows us to identify whether a model is overfitting (**Figure 6.10**)²²⁵. **Learning curves** allow us to visualise if a model is receiving adequate data so as not to become overly biased (**Figure 6.10**)³⁰⁰.

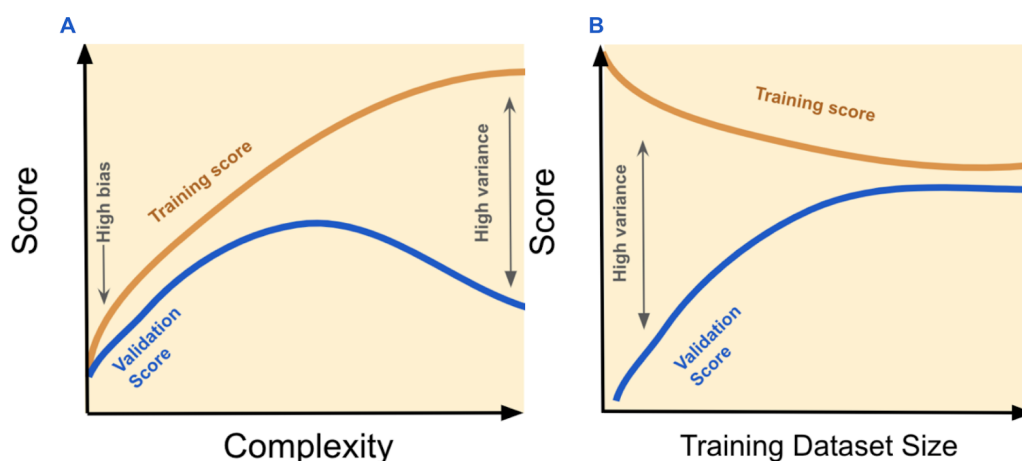


Figure 6.10A-B: Validation curves and learning curves. Adapted from reference³⁰⁰ In **6.10A** is an illustration of a validation curve, and **6.10B** (right) is an illustration of a learning curve. **6.10A** Validation curve- As complexity increases, bias is reduced and scores increase to a certain degree. As a problem becomes too complex and there is a high degree of variance, the validation scores drop-off whilst training scores remain high (as the model is overfit). **6.10B** Learning curve - For large datasets, the validation scores and training scores should converge. If the dataset is too small there is a high degree of variance, so the model can not identify patterns. However once the two scores align in a learning curve, other variables besides dataset size must be investigated.

Several methods prevent overfitting, such as validation call-backs³⁰⁰, where the model is trained over many 'epochs' of a dataset and assessed after each epoch using loss measures (e.g. accuracy). If the loss increases after a certain point, the model will cease training to prevent overfitting. Dropout layers can also be implemented in neural networks; here, the model can drop a percentage of the neurons after each layer.

6.1.7 Research aims

- Create a dataset to predict permeability
- Build a model that predicts permeability
- Predict permeability of new compounds

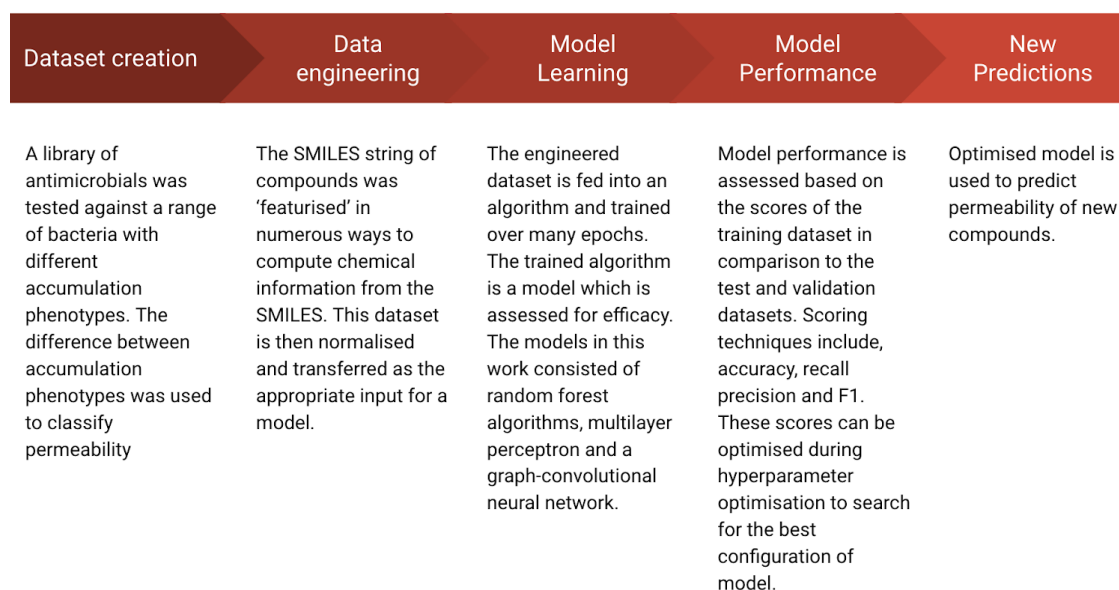


Figure 6.11 An overview of the general workflow of this chapter. A dataset was created using MIC data of compounds. The physicochemical features of compounds were derived from both user based features and learnt molecular representations. These features were used in a range of models to learn which future compounds may encounter permeability issues.

6.2 Results and discussions

6.2.1 Using MICs as a permeability indicator

Machine learning uses existing data to predict an unknown event. In the current context, permeability data is needed to predict permeability issues. Minimal inhibitory concentrations (MICs) were calculated for a range of antibiotics against a collection of bacterial strains and species. Importantly, different strains of *E. coli* that demonstrate different characteristics were used, primarily strains with mutations to porins. The strains used:

- *E. coli* Δ *OmpF* (keio collection 8925)¹²⁵
- *E. coli* Δ *OmpC* (keio collection 9781)¹²⁵
- *E. coli* Δ *OmpA* (keio collection 8942)¹²⁵
- *E. coli* Δ *OmpFCAB* (strain #102270)¹²¹
- *E. coli* Δ *OmpFCA*(strain #102268)¹²¹
- *E. coli* supplemented with sub-inhibitory amounts of polymyxin
- *S. aureus* ATCC 29213
- *P. aeruginosa* PAO1

These different strains were used to test the MICs of almost 70 antibiotics, spanning the most widely used classes in treating bacterial infections (**full MIC results available in**

Appendix 9.5). Some non-antimicrobial compounds were included to diversify the chemistry in the dataset and provide more variation for the model. These decoy compounds were included as machine learning classification models improve predictions when inactive compounds are included^{299,306}. It was anticipated that the differences in MICs between certain strains coupled with the molecular representations of the drug would inform future predictions of permeability.



Figure 6.12 MIC results of nine bacteria against a broad spectrum of bacteria. A heatmap of MICs, compounds tested along the y axis, and against strains along the x-axis. The lighter colours (beiges and pinks), are higher values. Darkest colours (black) represent lowest MIC values. There are some significant differences between OmpF knockout strains and the BL21 *E. coli*. *E. coli* permeabilised with polymyxin behaves similarly to the *Staphylococcus aureus* control. The strain with the highest MICs is *P. aeruginosa*. The total porin knockout (*E. coli* ΔOmpFCAB (strain #102270)¹²¹) also has higher MICs.

Porins are the main entry point for many antibiotics; hence knocking them out should affect MIC results. There seemed to be similar compounds affected by the loss of porins. In particular, β -lactams (cefoperazone, ceftiofur, cephalexin, aztreonam, cefotetan), and non- β -lactams (fleroxacin, minocycline, nitrofurantoin) were affected (**Figure 6.12**). The *ompF* knockout strain is less susceptible to antibiotic treatment than $\Delta ompC$ and $\Delta ompA$ strains (**Figure 6.12**). There were two strains (ΔACF and $\Delta ACFB$ *E. coli*) with all porins knocked out, with one of these strains also lacking *lamB*. These two strains shared the same pattern of intrinsic resistance to the drugs above but had even higher MICs than the single knockouts (**Figure 6.12**). This work reiterated that porins are the key pathway for many antibiotics, and changes to the porin or their expression could result in resistance to some compounds. It also demonstrated that certain compounds favour different pathways. For example, the phosphomycin MIC in the $\Delta ompA$ strain was higher than the OmpF and OmpC knockouts (**Figure 6.12**).

MICs were also tested in *Pseudomonas aeruginosa* strain PAO1 and *Staphylococcus aureus* ATCC29213. *P. aeruginosa* have a more impermeable membrane than *E. coli*, and instead of having porins that are homotrimers, they have porins that are single channels such as OprD³⁰⁷⁻³⁰⁹. As a result, *Pseudomonas aeruginosa* are more selective in their sensitivity to antibiotics, and infections are harder to treat^{308,309,309}. Therefore, using PAO1 should give us an indication of which drugs have more challenges to permeability.

PAO1 demonstrated a marked increase in resistance to some cephalosporins such as cephalothin, cefotetan, and cefixime compared to BL21 *E. coli* and the porin knockout derivatives (**Figure 6.12**). Conversely, *S. aureus* was used as an outer membrane absent control, with this strain demonstrating similarities to the BL21 *E. coli* treated with polymyxin (**Figure 6.12**). The *Staph* strain was more sensitive (in general) when compared to the Gram-negative strains (**Figure 6.12**).

An additional condition was BL21 *E. coli* treated with sub-inhibitory concentrations of polymyxin B (0.0625 $\mu\text{g}/\text{mL}$). Polymyxin B is a cationic peptide which acts as a membrane permeabilising agent. Therefore, its use could negate the effect of the outer membrane as a limit of antibiotics. When the polymyxin treated *E. coli* was exposed to vancomycin, the MIC dropped to 2 $\mu\text{g}/\text{mL}$ (**Figure 6.12**). This low MIC demonstrated that the outer membrane is no longer a limiting factor for permeability as vancomycin is a large compound (1447 Da) that cannot cross the Gram-negative outer membrane.

The permeability control used in this experiment was not necessarily ideal, as the potential synergistic effects of polymyxin B have not been controlled. Other controls that could have been used for this and have been used in the other work include; spheroplasts and Zgurskaya's $\Delta fhuA$ strains^{33,310}. Spheroplasts are formed from *E. coli* treated with a combination of lysozyme and osmotic gradients that cause *E. coli* to shed their outer membrane³¹¹. Spheroplasts were not used here as β -lactams were included in this study, and the process of forming spheroplasts could have damaged the targets, impacting the results. In addition to this, spheroplasts are fragile so cell death may be attributed to fragility as opposed to antibiotics. Zgurskaya's strains have a modified iron channel (FhuA) that is rendered permanently open^{33,310}. Using these strains in this work resulted in an excessive variation between repeats and thus excluded.

6.2.2 Feature engineering and extraction of compounds in permeability dataset

Relative changes in MIC values between *E. coli* and permeabilised *E. coli* were calculated. The calculated relative changes informed the creation of 4 permeability outputs, shown in **Figure 6.13**. In Class 0, the relative change between the two conditions was 0, meaning there was no decrease in MICs in permeabilised *E. coli*; this class included drugs like aztreonam and cefoxitin. In Class 1, there was a 50% reduction in MICs in permeabilised *E. coli*; for example, the doxycycline MIC decreased from 1 to 0.5 $\mu\text{g}/\text{mL}$. Class 2 contained compounds with a 75% reduction in MICs in permeabilised *E. coli*; for example, the MIC of azlocillin decreased from 8 to 2 $\mu\text{g}/\text{mL}$. Class 3 contained compounds that had a more than 75% reduction in MICs, compounds in this category included rifampicin and clarithromycin. For compounds in Class 3, the outer membrane is a significant barrier.

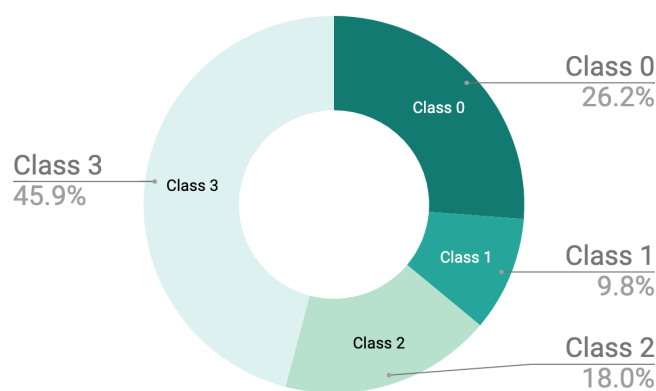


Figure 6.13 Distribution of classes in the MIC dataset. Classes were determined based on relative change. Relative change was calculated using the difference between wildtype (BL21) MICs and permeable MICs (BL21 supplemented with polymyxin).

With the initial MIC dataset, RDKit was used to calculate a number of chemical features from the SMILES of the compounds. The features input to help the model distinguish chemical features were as followed:

- Number of rotatable bonds
- Molecular weight
- Number of carboxylic acids
- Number of amines
- Number of Hydrogen bond acceptors
- Number of Hydrogen bond donors
- LogP
- Globularity
- Number of heterocycles
- Number of pyridines

These features were chosen based loosely on Richter *et al.*⁸⁹ rules for predicting permeability.

The dataset for predicting permeability consisted of the above features calculated by RDKit, the compounds, and the classes for predictions (based on the relative change between *E. coli* and hyperpermeable *E. coli*). A small neural network based on this dataset was built. The neural network was a simple MLP consisting of an input layer of 10 neurons, feeding into two fully connected hidden layers that led to 4 potential outputs.

6.2.3 Results of permeability dataset

The initial results of this model were poor, with no predictions being made for classes other than Class three (*results Table 6.3*). Class three are compounds hindered by permeability and have a higher efficacy when the outer membrane has been removed. This model is biased to this Class; as a result, showing the model new compounds will result in the model predicting the compound as Class 3.

Metric	Score
Weighted F1 score	33%
Accuracy	50%
Precision weighted average	12%
Recall weighted average	50%

Table 6.2 Metric scores from the MLP MIC model predicting permeability. Poor scores across all metrics. The weighted F1 is calculated for each class then a weight is multiplied to each class depending on the number of true labels of each class, then all the F1s for each class are added together.

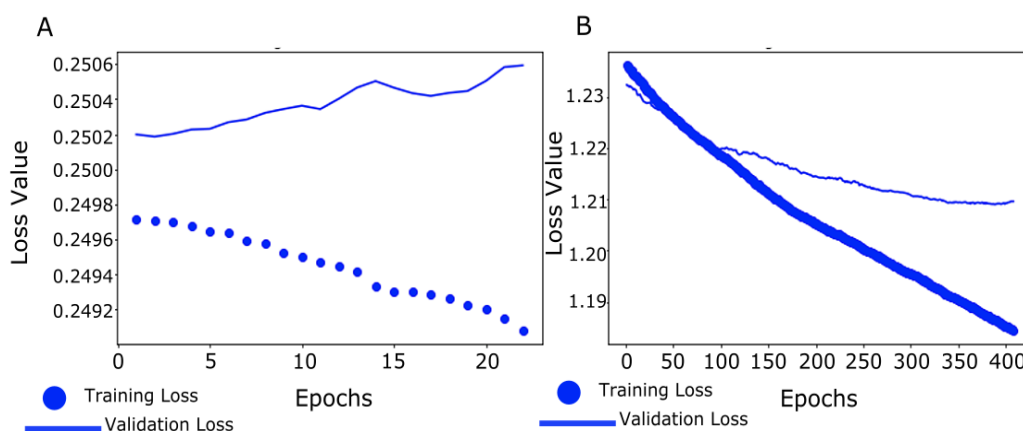


Figure 6.14A-B Training curves of the permeability prediction made by the MLP model of permeability. Training curves over 20 epochs (A) and 400 epochs (B). In A there is no convergence of loss, but in (B) there is some, but the training loss rapidly drops suggesting overfitting, in addition the validation loss is high, suggesting high bias.

This early model failed to identify any relevant patterns from the dataset. Hyperparameter optimisation was carried out on the model to improve metrics. In hyperparameter optimisation, changes to several qualities were investigated, such as learning rates, optimisers, number of neurons, and number of hidden layers. Despite the hyperparameter optimisation, no significant improvements were made. The training and validation losses were plotted (*Figure 6.14*). The training curves indicate that there is a high degree of bias in the model and that when the model is over trained (validation callbacks removed) the model overfits. This early attempt demonstrates the

initial issues of implementing deep learning in drug discovery, the model is trying to resolve a complex issue with a small and varied dataset. The results using user-picked features have not been able to solve the complex issue of predicting permeability. This is likely due to the complex problem and the small sample size, which has created a high degree of bias. To improve this many more compounds are necessary, ideally, tens of thousands of compounds.

6.3 Creating a new dataset from SPARK database

6.3.1 Data preprocessing of new SPARK derived database

Ultimately, this model does not have enough data to identify patterns that lead to permeability issues; more data was sought to solve this issue. The PEW trust partnered with the technology of the Collaborative Drug Discovery Inc. (CDD) and created the Community for Open Antimicrobial Drug Discovery Shared Platform for Antibiotic Research and Knowledge (CO-ADD SPARK), a centralised data sharing resource providing access to a wealth of antimicrobial research³¹². A new database of 750 compounds was constructed by utilising the CO-ADD SPARK database. Compounds with MICs for wild type *E. coli* (ATCC 25922) and hyperpermeable *E. coli* were used. A large amount of data engineering was necessary to create a new dataset from the CO-ADD SPARK database. This feature engineering involved removing compounds from the released SPARK data without relevant MICs or repeated data, any that did not fit the patterns of standard MICs (**Figure 6.15**). The released SPARK data contained many MICs for other bacteria like *Pseudomonas aeruginosa* and *Staphylococcus aureus*. It also contained data on efflux.

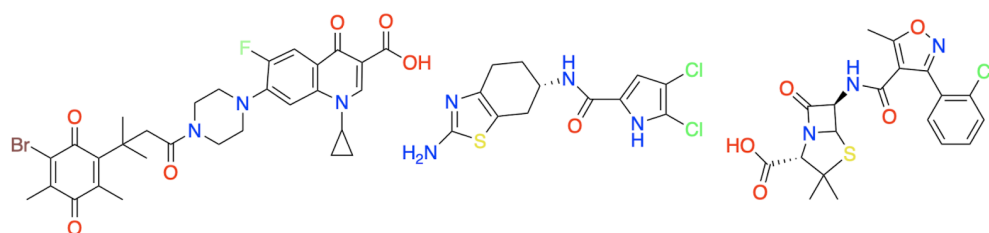


Figure 6.15 Feature engineering of the SPARK derived dataset. Three compounds that were removed during the data engineering stage as their MICs reportedly changed by less than 50%. Therefore not a whole dilution step as per CLSI guidelines¹²⁸.

The compounds were classified into three groups (**Figure 6.16**), where Class 0 contained compounds that had no relative change; these are compounds that do not have a permeability barrier. Class 2 contains compounds with 50% relative change in MICs; they exhibit potential permeability issues. Class 3 contains more than 50% relative change compounds, meaning the compounds most affected by permeability. The dataset was imbalanced in terms of classes, with the number of compounds in Class 3 being much larger. Class 0 had 146 compounds, Class 1 contained 98 compounds, and Class 2 contained 505 compounds.

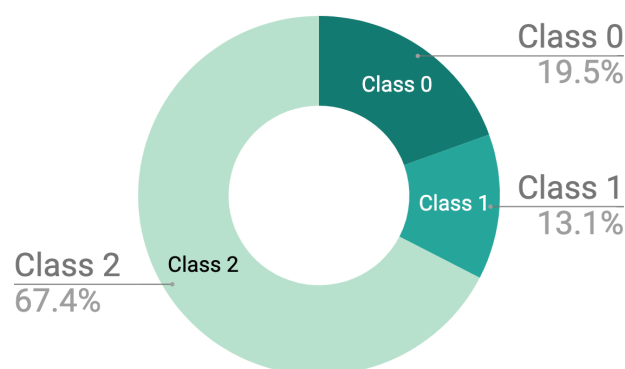


Figure 6.16 Class distribution in the new SPARK derived dataset. In this GCN model the number of outputs was three - as opposed to earlier models. The classes, however, are still imbalanced.

6.3.2 Featurising the SPARK derived dataset using Mol2Vec

The initial permeability investigation that used a small MIC dataset did not yield accurate permeability predictions (**Table 6.3**). The new dataset used to predict permeability contains 750 compounds. Unlike the previous investigation that used *user-defined* features, this study used a different type of molecular embedding created using Mol2Vec (**Figure 6.5**)¹⁵⁰. The result of compounds featurised with Mol2Vec was a 300 dimension vector for each compound¹⁵⁰. The Mol2Vec featurisation resulted in more input features, and therefore more input neurons were needed in the MLP. The data was split into training and test datasets, and the test dataset was further split into five cross-validations.

6.3.3 Results of new dataset

The MLP model prediction permeability using Mol2Vec molecular representations had accuracy scores of 70.8% (average over 5 *k*-folds) on the training dataset and 64.4% (average over 5 *k*-folds) on the test dataset. There was a reasonable difference between the test dataset and the training dataset in terms of accuracy; this could mean the model is not overfitting. However, most compounds were assigned as false negatives due to the class imbalance. As a result, a balanced F1 score (across classes) was much lower, <40%. This model struggled to fit compounds in Class 1. More work could have been done in investigating the permeability classes and distributions. In addition, different algorithms, like a random forest model, could have been utilised to improve scores.

6.4 Using learnt molecular representations

6.4.1 Featurising the new SPARK derived dataset

More complex algorithms, namely, graph convolutional neural networks, were implemented to improve upon this work, as they have had success in MedChem applications^{280,313}. GCN models learn molecular representations of molecules during training and are therefore believed to be better at distinguishing smaller changes. To achieve this, the DeepChem¹⁵³ python package library was used, which contains an inbuilt graph convolutional neural network that 'featurises' and transforms the data in a self-contained way (2.6.5).

6.4.2 Results of GCN model predicting permeability from SPARK derived dataset

The GCN model was trained on the SPARK derived dataset, composed of MIC data of 750 compounds. This was split 80:10:10 into training, test and validation datasets. The training dataset had 600 compounds with three outputs (determined by relative

change in MICs). This multiclass GCN model returned scores of 75% for the training dataset and 65% for the test set (**Figure 6.16**). Hyperparameter optimisation of the model was carried out; this returned the following best parameters; learning rate 0.0001, dropout rates of 0.2, and a dense layer size of 128. With optimisation, the training dataset ROC-AUC was improved to 95% (**Table 6.4**), and the test dataset ROC-AUC was improved to 72%. These scores indicate a considerable improvement in the prediction efficacy. However, the difference between training and test datasets suggests the model is overfitting.

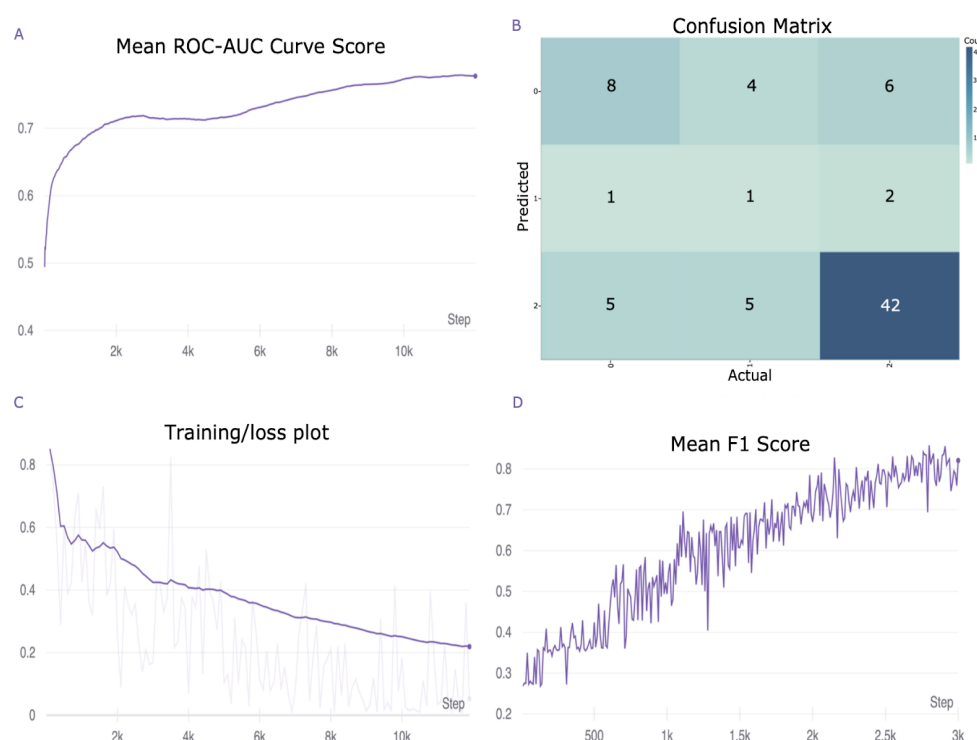


Figure 6.17A-D Results of the multiclass GCN model predicting permeability (A) a plot of the mean ROC-AUC curve scores over iterations of training. (B) the confusion matrix generated for the predictions of the multiclass GCN model. (C) the training over loss plot over iterations of training. (D) the mean F1 score of the multiclass GCN over time. The figures in 6.17A-D were generated using the WandB program.

	Class 0	Class 1	Class 2	Overall
Recall	57%	10%	84%	50%
Precision	44%	25%	84%	51%
Accuracy				70%
ROC-AUC				Training = 95% Test = 72%

Table 6.3 Metric results from the multiclass GCN model predicting permeability barriers. The precision and recall scores were calculated from the confusion matrices.

As mentioned, accuracy scores used previously have returned high scores in prior work. ROC-AUC scores are better at accounting for false positives and true positives³⁰². Ultimately, for a classification problem, a confusion matrix is best at allowing us to visualise the class predictions (**Figure 6.17B**). The model excels in predicting Class 2; these compounds are affected by permeability (**Figure 6.17B and Table 6.3**). It is less accurate at predicting Class 0 than the least accurate at predicting Class 1 (**Figure 6.17B and Table 6.3**). The metrics for each class are also calculated (**Table 6.3**). They demonstrate that Class 1 predictions have poor recall scores at 10% (**Table 6.3**) and poor precision at 25%. These scores are both worse than randomly assigning compounds to this Class. The low precision and recall scores demonstrate a failure to assign compounds to Class 1 accurately.

At this point, the model's output was debated; although previous work had used multiple outputs¹⁰⁴, perhaps binary classification would give better results. Binary classification may be an overgeneralization of permeability barriers. However, Class 1 contains molecules that have halved in MICs, in the scale of MICs, this is not a necessarily significant change. MICs are done as a serial dilution, with the range typically being 256, 128, 64, 32, 16, 8, 4, 2, 1, 0.5... $\mu\text{g/mL}$. A change of 1 to 0.5 could be due to error, and in drug testing, it would not usually be considered that significant.

6.5 Comparing graph convolutional model to a random forest model

6.5.1 Creating random forest model to predict permeability barriers

Currently, there is debate in the literature as to whether GCNs are that advantageous in contrast to 'simpler' methods²⁹⁶. GCN can perform excellently; however, they require a great deal of fine-tuning to reach this point; there are many factors in the model that could be optimised. In contrast, random forest classifiers that are learning from ECFPs can outperform GCN models^{279,296}. Random forest models have been used by other researchers investigating the rules around permeability^{89,103}. A random forest classifier was built using the same library (DeepChem) as before, but this time a SKlearn

(library) algorithm was used through Deepchem (as a wrapper). In this instance, three outputs were used for comparison to GCN, later two outputs will be compared.

For the random forest model, the SMILES of compounds were transformed into ECFP vectors. These vectors map out the atomic spacing of the molecule¹⁹. A radius of 3 was used to generate identifiers of the atoms, and a bit length of 1024 was used.

6.5.2 Random forest prediction results on SPARK derived dataset

The RF model created could not predict multiple outputs simultaneously, as a result each class was predicted sequentially. The random forest model has good ROC-AUC scores across all classes (**Figure 6.18**). However it demonstrated low precision (16.7%) and recall (10% scores) (**Table 6.4**) as the GCN model did (**Figure 6.18 and Table 6.3**).

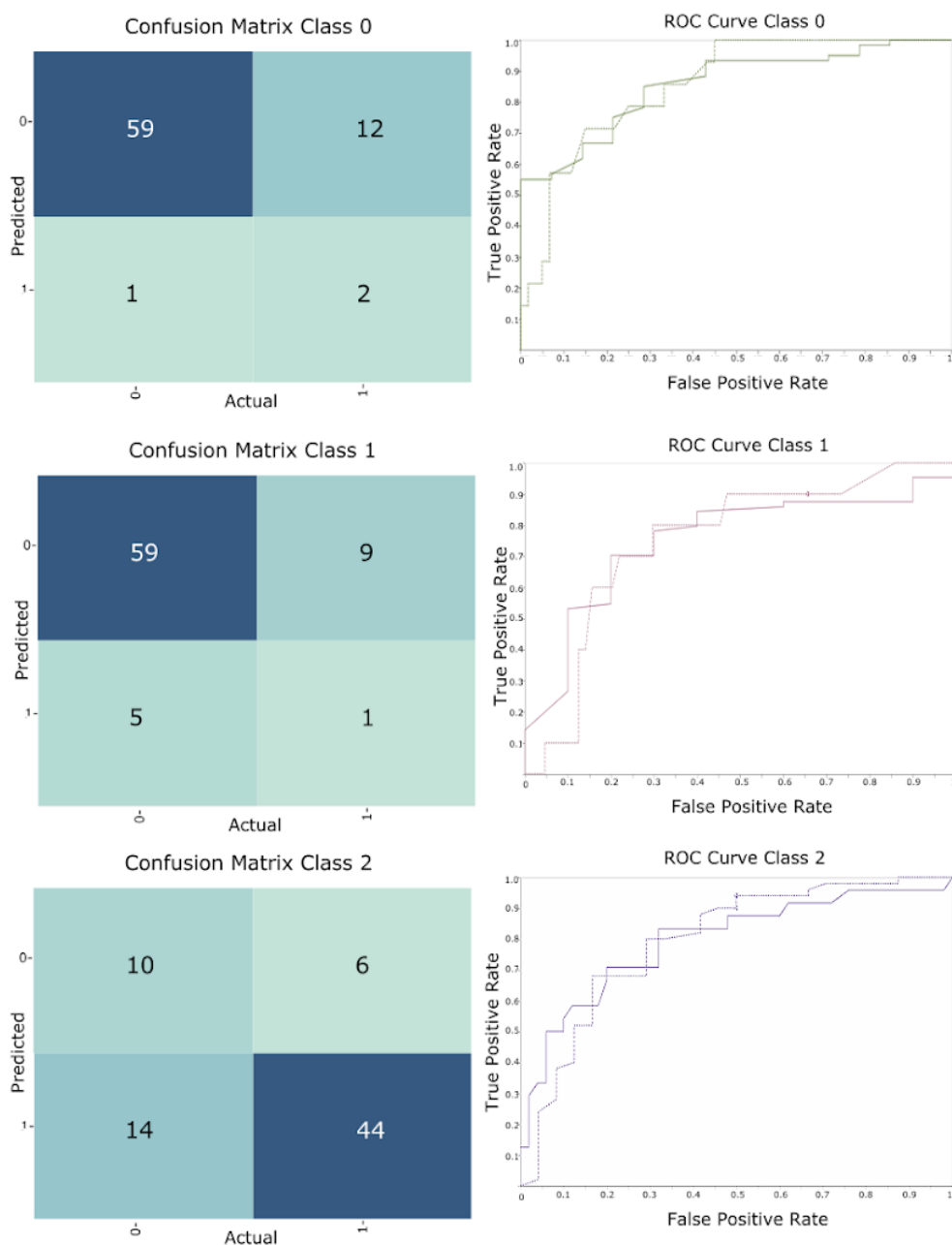


Figure 6.18 Confusion matrices and ROC curve scores for each class output of RF model. First row is Class 0 - compounds without permeability issues. In the middle row is Class 1. The bottom row is Class 2, these are compounds affected by permeability. Table 6.4 contains the metrics for this model.

	Class 0	Class 1	Class 2	Overall
Recall	14%	10%	88%	37.3%
Precision	66.7%	16.7%	75.8%	53%
Mean ROC-AUC	Training - 99% Test - 78% Valid - 73%	Training - 99% Test - 69% Valid - 72%	Training - 99% Test - 80.1% Valid - 86.4%	Training - 99% Test - 75.7% Valid - 77%

Table 6.4 Metrics calculated for each output of the random forest model generating predictions for three permeability outputs.

The random forest model had high mean ROC-AUC scores and nice ROC-AUC curves, particularly for Class 2. However, the recall and precision scores were poor for Classes 0 and 1 (**Table 6.4**). These low precision and recall scores indicate that samples that should be classified as Class 0 or 1 are misclassified. Class 1 has a low precision rate, suggesting that many of the compounds classified as Class 1 are false positives. This failure could be attributed to the smaller size of Class 1. The results from the training datasets are much higher than on the test datasets - particularly for class 1, indicating the model is overfitting.

6.5.3 Approaching permeability predictions from a binary perspective

A recurring theme of the models above is that the 'middle' classes perform much worse, so this multiclass problem was converted into a binary classification. Class 1 and Class 0 were merged. This reclassification meant that compounds in Class 0 contained compounds with no relative change and compounds with up to 50% decrease in MICs. Class 0 are compounds not impeded by a permeability barrier. In contrast, Class 1 (now) contains compounds that experience more than 50% reduction in MICs when permeabilised. Class 1 are compounds affected by a permeability barrier.

The first binary model tested was the GCN model; after hyperparameter optimisation, good ROC-AUC scores were found (**Table 6.5**). The binary GCN model returned ROC-AUC scores of 94% on the training dataset, 82% on the test dataset, and 82% on the validation training set. This model is performing much better than previously, but it is still overfitting (based on the difference between training and test results) (**Table 6.4 for metrics of multi-output GCN**). The precision and recall scores have significantly improved for both outcomes and are better than random assignment. (**Table 6.6 for results of the binary model**). The confusion matrix (**Figure 6.19**) demonstrates the amount of correctly classified compounds.

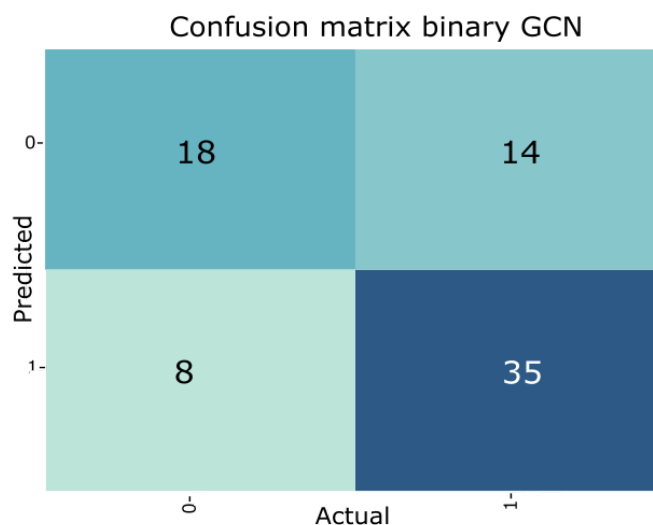


Figure 6.19 Confusion matrix of results from binary GCN model predicting permeability issues on the SPARK derived dataset. These are the classification results from the binary GCNs predictions on the the test dataset, which is composed of 75 randomly assigned compounds. The dataset is unbalanced, with more Class 1 compounds.

	Class 0	Class 1	Overall
Recall	69%	71%	70%
Precision	56%	81%	68.5%
Mean ROC-AUC			Train: 94% Test: 82% Valid: 82%

Table 6.5 Metric scores of binary GCN model of predicting permeability issues of the SPARK derived dataset.

The random forest model for predicting permeability also performed much better when a binary output was used (**Figure 6.20**). The random forest model returned scores of 99% on the training dataset, 86% on the test dataset, and 80% on the valid dataset (**Table 6.6**). The results are comparable to the GCN binary model (**Figure 6.19**). The cohesion in results across both models (when using a binary output) suggests that the use of a more complex model (GCN) may not be advantageous. Removing the other classes means this model now lacks the distinctions in different permeability classes; that other researchers have demonstrated¹⁰⁴. However, the models both meet the aims of the work - they can predict if permeability issues will hinder a compound. The dataset size could be increased to improve accuracy, or the root of the dataset (MICs in this case) could be changed to a different measure of accumulation. For example, the amount of intracellular accumulation could be quantified by MS techniques. Using

different data to inform permeability predictions could reduce the emphasis on compounds with effective antimicrobial activity, leaving the predictions more open to different permeability properties.

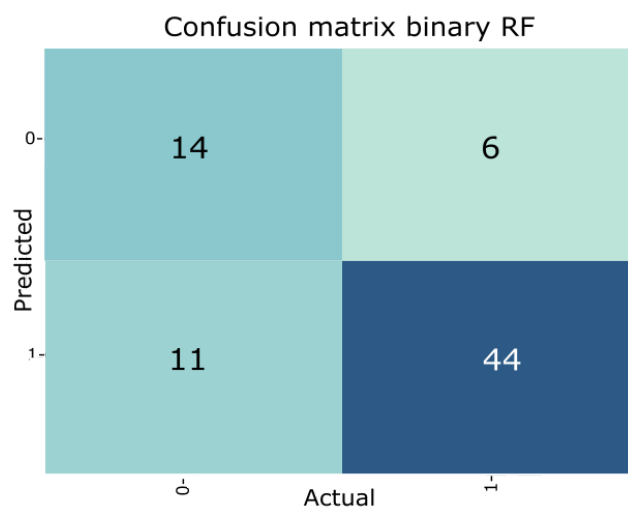


Figure 6.20 Confusion matrix of the results from the binary random forest model used to predict permeability barriers on the SPARK derived dataset. These are the classification results from the binary RFs predictions on the the test dataset, which is the same 75 randomly assigned compounds as was tested against the binary GCN.

	Class 0	Class 1	Overall
Recall	56%	88%	72%
Precision	70%	80%	75%
Mean ROC-AUC			Train: 99% Test: 86% Valid: 80%

Table 6.6 The metric results of the binary random forest model predicting permeability barriers of compounds.

6.6 Predictions and atomic contributions

Nine compounds were randomly removed from the dataset when this dataset was established. This small selection of compounds was used to test the efficacy of the binary RF and binary GCN model. This was a preliminary investigation into the efficacy of the models on unseen compounds. The nine compounds were translated into convolutional molecules for the GCN model and ECFPs with a radius of 3 for the RF model.

	Class 0	Class 1
Binary GCN recall	60%	75%
Binary GCN precision	75%	60%
Binary RF recall	40%	100%
Binary RF precision	100%	57%

Table 6.7 The results of the predictions of binary GCN and binary RF model on nine unseen compounds.

Both the binary versions of the models perform well. The binary GCN model was able to correctly predict 7 out of 9 compounds and maintained a high balance of precision and recall (**Table 6.7**). The compound failed to correctly predict compound #5 as having permeability issues and failed to predict compound #9 as not having permeability issues (**Figure 6.21**). Meaning the binary GCN misclassified compounds in both classes. The RF model predicted 6 out 9 compounds correctly. The model was more biased to predicting permeability issues for compounds and predicted compounds #8, #9 and #3 as having permeability issues (**Figure 6.21**). Both models incorrectly assigned #9 as being impermeable, suggesting the models could have learnt similar rules around impermeability. These nine compounds have come from the same compiled database as the other compounds, and as antibiotics make up one of the most chemically diverse drug groups, the models may perform poorly on compounds from outside of this dataset. Moreover, nine compounds are a small percentage of this dataset and a negligible fraction in drug discovery screens that often begin with thousands of compounds. These models ideally need to be tested on a new database to assess the permeability of new compounds, to establish the efficacy of both models. By testing the models on new compounds it would help establish which model is better.

The atomic contributions of the predictions made by the binary GCN model were investigated. Code adapted by DeepChem¹⁵³ based on the work of Riniker and Landrum 2013³¹⁴ was used to investigate these atomic contributions³¹⁻³³. It assesses the atomic contribution of each atom of a molecule by comparing the model's efficacy with and without each atom³¹⁴. The code returns the contribution of the atoms used to make predictions of classes³¹⁴. In **Figure 6.21**, the atomic mappings of each compound are presented alongside the class assigned by GCN and RF and the correct class. The red, in this case, means the atoms would contribute to permeability issues (**Figure 6.21**), whereas the blue indicates that the atom does not cause permeability issues (**Figure 6.21**).

As mentioned, on this limited dataset, the binary GCN model predicted correctly 7/9 times. In the first column on the left (**Figure 6.21**), the three compounds are quite similar in their structure, the top two differing by one nitrogen. These small changes are typical in drug discovery screens and can significantly affect activity. In this limited application, the change did not alter the permeability of the compound, but the binary GCN model did detect these changes (based on the returned atomic predictions). The alteration of a lead compound to produce many potential hits is common in drug discovery but it poses an issue for the implementation of machine learning methods in drug discovery. Usually in machine learning applications the data points are independent of each other, independent data that is distinct is easier to separate and therefore classify into groups²⁷⁹.

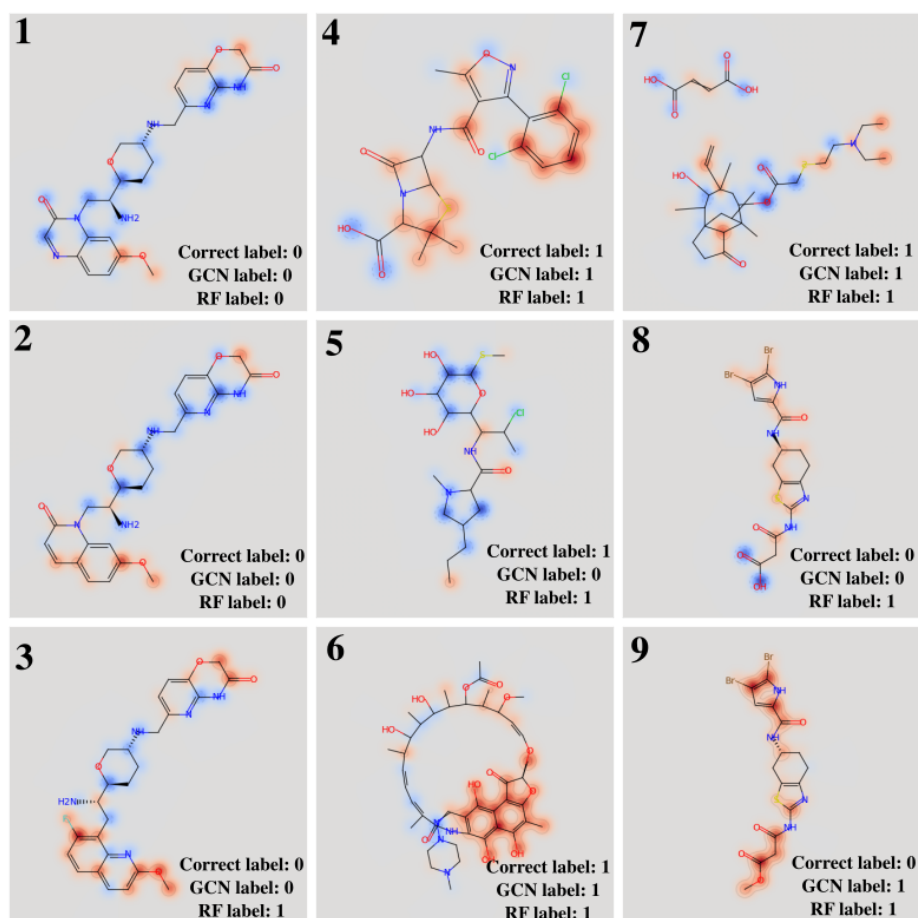


Figure 6.21 Nine unseen compounds used to test the efficacy of binary models. The correct and predicted labels are shown. The red indicates which atoms may cause permeability issues. This figure demonstrates the atomic contributions that resulted in the classification from the binary GCN model.

Both models were then used to predict whether two compounds from Tomassi *et al.*¹⁰¹ had permeability issues. These compounds were designed as derivatives of avibactam but manipulated to achieve better permeability in *Pseudomonas aeruginosa* using their permeability assay, crystallography, molecular dynamics and MICs. The initial compound (compound 2) had changes made to R1 to optimise translocation through porins, whilst functionality (**Figure 6.22**) was maintained with the carboxylic acid, leading to compound EX0462. The resultant molecule (EX0462) achieved target specificity and permeated into a range of Gram-negative organisms¹⁰¹.

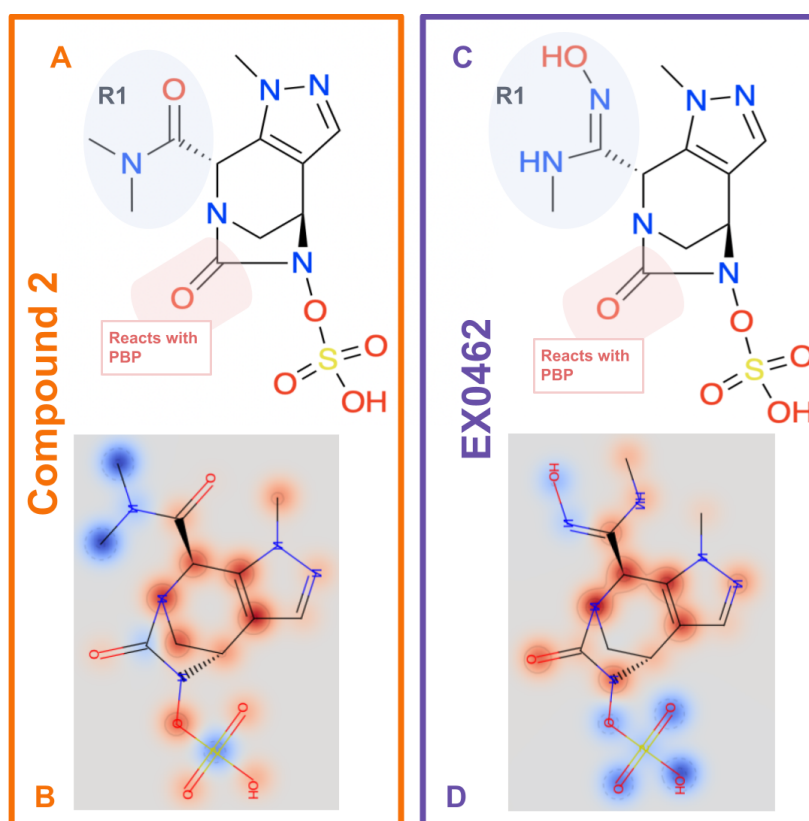


Figure 6.22A-D Two compounds used to probe the accuracy of permeability predictions. 6.22A the structure of compound 2, with R1 and the active site labelled. (B) the returned atomic contributions resulting in the predictions of this class. On the right in the purple box is EX0462, the compounds modified to increase permeability in *Pseudomonas* by altering R1 whilst maintaining activity at the active site. (C) the labelled structure of EX0462, and (D) the atomic contributions on the classification of this structure.

Tomassi *et al.*¹⁰¹ report that the MIC of compound 2 in the PAO1 strain is 4 $\mu\text{g}/\text{mL}$ and 32 $\mu\text{g}/\text{mL}$ in clinical isolates¹⁰¹. In *E. coli* ATCC 25922, the MIC of compound 2 was 2 $\mu\text{g}/\text{mL}$ and 1-8 $\mu\text{g}/\text{mL}$ in clinical isolates. EX0462 had a MIC of 0.5 $\mu\text{g}/\text{mL}$ in both reference strains; 25922 (*E. coli*) and PAO1 (*P. aeruginosa*). This MIC increased to 2 $\mu\text{g}/\text{mL}$ for clinical isolates of *P. aeruginosa* and 1 $\mu\text{g}/\text{mL}$ for *E. coli* clinical isolates. When

these compounds (compound 2 and ETX0462) were tested in both models (binary RF and GCN), they predicted that both compounds would not have permeability issues.

The atomic contributions of permeability predictions were visualised (**Figure 6.22**). The changes at R1 of ETX0462 appeared to have a negative contribution to permeability. This is in contrast to the design of the molecule. There were some similarities in atomic contributions between the two molecules, in particular the core ring of the molecules was predicted to be a permeability barrier to both compounds. EX0462 should be more permeable than compound 2 because of the alterations at R1 (this is a distinction the binary GCN model was not able to make)¹⁰¹. However, both compounds have reportedly low MICs in *E. coli*. Compound 2 was deemed as having permeability issues in the Tomassi *et al.*¹⁰¹ paper for two key reasons. Firstly, when modelling the interactions of the compound with the predominant porin (OprD) in *Pseudomonas aeruginosa*, and secondly from MIC values. OprD is different from OmpF; it has more selectivity of its substrates; it is generally narrower and is a monomer, not a homotrimer. This work used *E. coli* as the model organism instead of *Pseudomonas aeruginosa*, which may be why the compounds were both predicted as not having permeability issues.

6.7 Discussion and Future Prospects

The final binary models (GCN and RF) were built through many iterations of algorithms and datasets, resulting in accurate and useful predictions of new molecules. The atomic contributions of the binary GCN model indicated it identified small changes to compounds (**Figure 6.21**). The binary RF and GCN both demonstrated high ROC-AUC scores, however they demonstrated differences in predictions on new compounds. The binary RF appeared to be more biased in predicting Class 1 so compounds that have a permeability issue. To further this work, three approaches could be taken, firstly, to explore more algorithms to comprehensively compare different algorithms' success in predicting permeability and use this to better weigh in on the current literature in this field.

Secondly, the models could have been improved with a more complex dataset and may have elucidated more information on permeability, especially without reliance on MICs.

Diverging from MICs would have reduced the selectivity for antimicrobial activity and narrowed the scope to permeability. The SPARK database would contain more data for each compound available in an ideal scenario. Ideally, the results would be generated from a range of assays and modelling studies, including some of the following examples: TOMAS^{100,101} permeability assay, OMV swelling assay¹⁰⁵, MICs, molecular dynamic studies, and docking studies of the compounds in porins. Many more permeability assays have been used, but they have been created to quantify specific chemicals such as fluoroquinolone fluorescence⁷⁹, β -lactam iodometric assay⁷⁰, and specific fluorescent bacterial probes³¹⁸. By using permeability results from a range of assays, a permeability score could be generated for each compound. This system might create a better picture of permeability, thus making it easier for a model to predict permeability issues. The SPARK database is a good starting point for this sort of complex library, owing to the platform's open-source nature.

Thirdly, the models could have been improved by using more data. Predicting permeability is a complex problem that needs extensive data to inform classification. AI helps solve complex issues but requires extensive datasets; for example, a computer vision problem may involve many hours of videos, and from each frame of the video, each pixel could be used to inform the model. This is a much greater wealth of information than used in this study²⁷⁹. However, with more data comes more noise, so a greater effort to denoise the data should be made if using the random forest model that uses ECFP molecular representations³¹⁹. Noise reduction could be achieved using dimensionality reduction techniques such as principal component analysis.

The final binary models from this work should have been applied to a more extensive database such as the ZINC15 database or the drug repurposing hub. The predictions generated from these databases should then have been purchased and tested for permeability issues. The compounds with and without permeability barriers should then have had their physicochemical properties probed to discover any underlying permeability rules. The results of this larger screen could then have been used to inform antimicrobial drug design.

The aim of having a model that predicts permeability issues is to apply it to compounds before drug screening to assess which molecules may have issues. The binary RF and GCN models were used to predict when compounds may have permeability issues. This is an essential consideration for drug discovery. However, compound accumulation is

not just dependent on permeability but also on active efflux. The SPARK database does contain efflux data, so a model could be built based on efflux or accumulation by combining data on permeability and efflux. Zgurskaya *et al.*³³ have done some work on predicting permeability and have included efflux as a consideration.

Furthermore, this work focused mainly on *E. coli* as opposed to other more impermeable bacteria. Different compounds may have been restricted by building a different model based on the permeability of other bacteria such as *Pseudomonas*. *Pseudomonas aeruginosa* in particular has been used frequently in accumulation studies due to its impermeable outer membrane³⁰⁷⁻³⁰⁹. Perhaps a myriad of MICs from a range of bacteria could also be used to build a more comprehensive model of permeability.

Conversely, precise models could be built based on individual species' porin or lipid permeability. This was an initial aim of the early iterations of this work. The role of different porins in different compound permeability was going to be used to predict permeability (using the knockout strains). However, as more data was needed to predict permeability accurately and the SPARK database was utilised, the investigation into porins was lost. In future applications of this work, perhaps the compounds in the final database could have been purchased, and a permeability database around porins could have been built. These compounds could have had MIC with the knockout strains to build a specific role of each porin in permeability. They also could have been screened with a library of resistant mutants whose permeability phenotype was elucidated by genomic and proteomic studies. This would result in a precise database that may be able to illuminate more of the rules guiding model predictions.

AI offers many exciting avenues for the field of MedChem. However, AI comes with drawbacks and errors, particularly when looking at molecules. In most other fields where AI is used, there is an assumption that each data point is independent of another (commonly known as independent and identically distributed data (IID))²⁷⁹. This is not the case in MedChem, as many small changes made to a parent compound can result in chemically similar compounds that may have vastly different functions, resulting in compounds not being independent of each other²⁷⁹, which can impact the distinguishability of classes. However, such an issue could be tackled by good molecular representations of compounds that can sensitively comprehend and distinguish small changes to molecules²⁷⁹. This is where learnt representations, such as the ones used in

the binary GCN model are useful. Alternatively, the data can be split by using a scaffold splitter, separating the dataset based on different chemical structures, this differentiation based on chemistry may aid in classification.

Additionally, interpretability can become an issue, especially when using deep-learning representations to 'featurise' compounds instead of *user-defined feature* engineering (the GCN uses deep learning representations)^{279,320}. Indeed, understanding what is going on 'under the hood' is a key issue. There has been a push towards explainable artificial intelligence approaches³²⁰. Explainable AI aims to address critical questions such as why the model predicted the way it did, why the prediction is correct, what rules have been learnt from the model, and how certain the prediction is³²⁰. In this chapter, some interpretations were gleaned from the model by looking at atomic contributions. Much more work should have been done to understand the model better and whether its predictions could be extrapolated³²⁰.

The DeepChem¹⁵³ library was essential for this work, as it made the field and the techniques approachable and provided many avenues to explore (such as atomic contributions). DeepChem¹⁵³, in particular, has allowed biologists and chemists an accessible avenue into the field of machine learning, allowing interdisciplinary development, which has proven essential to the discovery of new antibiotics.

6.8 Conclusions

The future of novel antimicrobial discovery is an interdisciplinary endeavour. It is a global health issue requiring attention from all angles. AI could streamline the expensive task of drug discovery, enabling far more research to be done by start-ups and academic institutions. AI could generate new compounds, predict which compounds could be 'hit', and predict which compounds are experiencing permeability issues. It could also be used more broadly to tackle themes of AMR. This chapter has demonstrated some success in predicting permeability issues in potential antimicrobials.

7. Conclusions and future work

This thesis approached the theme of Gram-negative permeability in an interdisciplinary manner, aiming to develop tools to measure the accumulation of β -lactams and create a broader high-throughput method of measuring accumulation. Each of these chapters will be briefly summarised before contextualising the results in relation to the current state of antimicrobial drug development.

A paper in 2016 by Kotagiri *et al.*¹³⁵ reported that the heat degradation of ampicillin creates a fluorescent compound that maintains its β -lactam ring (Compound 8)¹³⁵. The paper states that heating ampicillin to high temperatures results in the oxidative deamination of ampicillin's primary amine, giving the β -lactam derivative fluorescent properties¹³⁵. The compound reported in this paper is a fluorescent β -lactam created without modification by a fluorophore. Therefore, it could have been used as intrinsically fluorescent antibiotics, like fluoroquinolones are, in permeability studies^{68,79}. Given the potential of Compound 8, the compound was therefore chosen as a potential candidate to study β -lactam permeability in this thesis. Chapter three investigated the structure and function of Compound 8, and, ultimately, found that the structure reported was inaccurate. The new structure described in Chapter three does not have a β -lactam ring and will not be a good indicator of β -lactam permeability due to its differing chemical properties. Further experiments could have further evidenced the results in this chapter, but ultimately, they were not necessary. There are other studies of β -lactam permeability using micro-iodometric assay⁷⁰, radiolabelled β -lactams⁵⁵ and MS¹⁰⁹. The use of MS to quantify permeability is explored in Chapter four.

Chapter 4 aimed to develop a label-free tool to measure β -lactam permeability through the covalent acylation of PBPs. This chapter used three mass spectrometers; Orbitrap, FT-ICR MS, and MALDI. The chapter began by looking at the binding of; ampicillin, nitrocefin, aztreonam and ceftazidime to PBPs from various species, using a bottom-up proteomic technique. The acylation of the active-site serine of PBPs by β -lactams was observed. However, results varied depending on PBP, β -lactam and bacteria, creating variability and unreliability in results. A top-down proteomic technique was attempted with FT-ICR MS to look specifically at ceftriaxone and *E. coli* PBP3. The results from this experiment resulted in masses that were inconsistent with calculations. Subsequently a more straightforward approach was taken using MALDI-TOF; this was more successful

than FT-ICR-MS. Ultimately, this chapter was not progressed due to the limited results and the availability of other assays to measure β -lactam permeability. Ultimately, it was clear from this chapter that MS is not an appropriate method of measuring β -lactam permeability, in this case. The direction of this thesis then shifted to looking at general permeability assays and predicting permeability.

In Chapter five, the work shifted from attempting to monitor the permeability of just β -lactam antibiotics, to establishing a method to measure permeability across many classes. Liposome based permeability assays have been a mainstay of the field for many years. This chapter sought to build on this foundation to create a better model of *E. coli* for permeability assays and establish a universal permeability screen. To do this, techniques from synthetic biology were applied to build synthetic cells in 96-well microplates. This chapter laid the groundwork for using this assay for drug screening uses. Preliminary data from this chapter suggest that synthetic cells could be a good model for permeability studies, the assay is highly adaptable and broadly applicable. Unfortunately, further development fell outside of the time frame of this project. But the prospects of this study are huge, from monitoring porins to altering lipid composition to mimic the membrane of a variety of bacteria, adapting the target of the screen, and creating a tandem activity-permeability screen, essential for drug discovery purposes.

Chapter six sought to establish whether machine learning techniques could be used to predict permeability. By having tools to predict permeability into *E. coli*, the process of drug development could be streamlined. The results from this chapter demonstrated the efficacy of two models (binary RF and GCN) in predicting whether a compound will face permeability barriers. The complex nature of the permeability barrier means fitting a model to this dataset is challenging, there is certainly room for improvement. Future work from this chapter includes; applying the model to large datasets to see if it can predict permeability from more diverse screens, training a model on different bacteria or different permeability characteristics, and training a model on accumulation, including efflux data. More work developing the model and applying this model are necessary. However, this work could be used prior to drug development screens to save time and money which is desperately necessary.

This thesis has approached Gram-negative permeability from an interdisciplinary standpoint. Techniques from microbiology, biochemistry, chemistry and computer science have been used. The breadth of the project was vast, perhaps preventing the

explanation of the rules around permeability. Several tools enabled this interdisciplinary effort in the name of ‘open-source science’; in the form of shared protocols (<https://bio-protocol.org/>), databases (SPARK³¹²) and tutorials for machine learning (GoogleColab notebooks, DeepChem³²¹ and <https://www.kaggle.com/>), that has enabled this project. The expansion of the SPARK database or the establishment of a similar database could generate the wealth of information needed to predict permeability issues for compounds resulting in the creation of effective antimicrobials. Both target and phenotypic screens would benefit from assays, models, and rules informing drug design.

Open-source science could be used to fill the knowledge gap left by big pharma exiting antimicrobial drug discovery. Open-source science allows small pharmaceutical companies and academic institutions to reinvigorate the drug discovery pipeline. Previous open-source science projects identified antimalarial compounds; one of the winners of this project used the DeepChem library³²². A new open-source project has been started to search for novel antimicrobial compounds that inhibit the Mur Ligases (<https://github.com/opensourceantibiotics/murligase>). The Mur ligase project asks contributors to use generative adversarial models to create compounds predicted to inhibit Mur ligases. Organisers of the project are willing to synthesise the hits from the screens and assay the hits for activity. The nature of this project enables experts from around the world to use their knowledge to find new antibiotics. However, this project is still affected by the same economic constraints that have burdened drug discovery. The compounds discovered in this project are patent-free, meaning the drug will be off-label immediately, limiting any financial returns. This means pharmaceutical companies are unlikely to fund the expensive clinical trial process for a drug that will not make them money.

Ultimately, antibiotic development will continue to be impeded due to economic reasons. Until the threat of AMR is appropriately acknowledged, the discovery of antibiotics will continue to suffer. The process of developing antimicrobials needs to be de-risked to entice businesses into antimicrobial drug discovery. In the meantime, small teams will continue to attempt to bridge the gap. Having tools such as permeability assays will allow for a streamlined approach to developing antimicrobials. Better still, having rules or models that can guide the rational drug design of compounds will remove a great deal of screening. The preliminary data in the synthetic cell chapter suggests it could be used as a platform for permeability screening. It could

have broad applicability with some adaptations like standardised CFPS reactions and a microfluidic platform. The results from the machine learning chapter suggest that with more data, a model that predicts permeability barriers to compounds could be established. With more expertise, the rules surrounding permeability barriers could be better explained. To conclude, this thesis has explored both specific and general permeability assays from an interdisciplinary standpoint, with some chapters demonstrating promising applications with further development.

8. Bibliography

1. Ventola, C. L. The Antibiotic Resistance Crisis. *P T* **40**, 277–283 (2015).
2. Adedeji, W. A. THE TREASURE CALLED ANTIBIOTICS. *Ann Ib Postgrad Med* **14**, 56–57 (2016).
3. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae. *Br J Exp Pathol* **10**, 226–236 (1929).
4. Chain, E. *et al.* PENICILLIN AS A CHEMOTHERAPEUTIC AGENT. *The Lancet* **236**, 226–228 (1940).
5. Henry, R. J. THE MODE OF ACTION OF SULFONAMIDES *. *Bacteriol Rev* **7**, 175–262 (1943).
6. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.*†. *Proceedings of the Society for Experimental Biology and Medicine* **55**, 66–69 (1944).
7. Ehrlich, J., Bartz, Q. R., Smith, R. M., Joslyn, D. A. & Burkholder, P. R. Chloromycetin, a New Antibiotic From a Soil Actinomycete. *Science* **106**, 417 (1947).
8. KOYAMA, Y. A new antibiotic ‘colistin’ produced by spore-forming soil bacteria. *J. Antibiot.* **3**, 457–458 (1950).
9. Duggar, B. M. Aureomycin; a product of the continuing search for new antibiotics. *Ann N Y Acad Sci* **51**, 177–181 (1948).
10. Farzam, K., Nessel, T. A. & Quick, J. *Erythromycin. StatPearls [Internet]* (StatPearls Publishing, 2021).
11. Griffith, R. S. Introduction to Vancomycin. *Reviews of Infectious Diseases* **3**, S200–S204 (1981).
12. Sensi, P. History of the Development of Rifampin. *Reviews of Infectious Diseases* **5**, S402–S406 (1983).
13. Leshner, G. Y., Froelich, E. J., Gruett, M. D., Bailey, J. H. & Brundage, R. P. 1,8-NAPHTHYRIDINE DERIVATIVES. A NEW CLASS OF CHEMOTHERAPEUTIC AGENTS. *J Med Pharm Chem* **91**, 1063–1065 (1962).
14. Eisenstein, B. I., Oleson, F. B., Jr & Baltz, R. H. Daptomycin: From the Mountain to the Clinic, with Essential Help from Francis Tally, MD. *Clinical Infectious Diseases* **50**, S10–S15 (2010).
15. Slee, A. M. *et al.* Oxazolidinones, a new class of synthetic antibacterial agents: in vitro and in vivo activities of DuP 105 and DuP 721. *Antimicrob Agents Chemother* **31**, 1791–1797 (1987).
16. *2020 Antibacterial agents in clinical and preclinical development: an overview and analysis.* 76 (2021).
17. SOARES, G. M. S. *et al.* Mechanisms of action of systemic antibiotics used in periodontal treatment and mechanisms of bacterial resistance to these drugs. *J Appl Oral Sci* **20**, 295–304 (2012).
18. Lebar, M. D. *et al.* Forming Cross-Linked Peptidoglycan from Synthetic Gram-Negative Lipid II. *J. Am. Chem. Soc.* **135**, 4632–4635 (2013).
19. Vollmer, W., Blanot, D. & De Pedro, M. A. Peptidoglycan structure and architecture.

- FEMS Microbiology Reviews* **32**, 149–167 (2008).
20. GRAM, C. Ueber die isolirte Färbung der Schizomyceten in Schnitt- und Trockenpräparaten. *Fortschritte der Medicin* **2**, 185–189 (1884).
 21. Silhavy, T. J. Classic Spotlight: Gram-Negative Bacteria Have Two Membranes. *J Bacteriol* **198**, 201 (2015).
 22. Silhavy, T. J., Kahne, D. & Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb Perspect Biol* **2**, a000414 (2010).
 23. Krause, K. M., Serio, A. W., Kane, T. R. & Connolly, L. E. Aminoglycosides: An Overview. *Cold Spring Harb Perspect Med* **6**, a027029 (2016).
 24. McCoy, L. S., Xie, Y. & Tor, Y. Antibiotics that target protein synthesis. *WIREs RNA* **2**, 209–232 (2011).
 25. Hooper, D. C. & Jacoby, G. A. Topoisomerase Inhibitors: Fluoroquinolone Mechanisms of Action and Resistance. *Cold Spring Harb Perspect Med* **6**, a025320 (2016).
 26. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol* **33**, 300–305 (2017).
 27. O’Neill, J. *Tackling drug-resistant infections globally: final report and recommendations*. (2016).
 28. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* **399**, 629–655 (2022).
 29. Rosenblatt-Farrell, N. The Landscape of Antibiotic Resistance. *Environ Health Perspect* **117**, A244–A250 (2009).
 30. Reygaert, W. C. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol* **4**, 482–501 (2018).
 31. Corona, F. & Martinez, J. L. Phenotypic Resistance to Antibiotics. *Antibiotics (Basel)* **2**, 237–255 (2013).
 32. Harms, A., Maisonneuve, E. & Gerdes, K. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science* **354**, aaf4268 (2016).
 33. Krishnamoorthy, G. *et al.* Synergy between Active Efflux and Outer Membrane Diffusion Defines Rules of Antibiotic Permeation into Gram-Negative Bacteria. *mBio* **8**, (2017).
 34. Aslam, B. *et al.* Antibiotic resistance: a rundown of a global crisis. *Infect Drug Resist* **11**, 1645–1658 (2018).
 35. Projan, S. J. Why is big Pharma getting out of antibacterial drug discovery? *Current Opinion in Microbiology* **6**, 427–430 (2003).
 36. Plackett, B. Why big pharma has abandoned antibiotics. *Nature* **586**, S50–S52 (2020).
 37. Årdal, C. *et al.* Antibiotic development — economic, regulatory and societal challenges. *Nat Rev Microbiol* **18**, 267–274 (2020).
 38. Dheman, N. *et al.* An Analysis of Antibacterial Drug Development Trends in the United States, 1980–2019. *Clinical Infectious Diseases* **73**, e4444–e4450 (2021).
 39. Mullard, A. Achaogen bankruptcy highlights antibacterial development woes. *Nat Rev Drug Discov* **18**, 411 (2019).
 40. Cama, J. *et al.* To Push or To Pull? In a Post-COVID World, Supporting and Incentivizing Antimicrobial Drug Development Must Become a Governmental Priority. *ACS Infect. Dis.* **7**, 2029–2042 (2021).

41. Rybenkov, V. V. *et al.* The Whole Is Bigger than the Sum of Its Parts: Drug Transport in the Context of Two Membranes with Active Efflux. *Chem. Rev.* (2021) doi:10.1021/acs.chemrev.0c01137.
42. Prajapati, J. D., Kleinekathöfer, U. & Winterhalter, M. How to Enter a Bacterium: Bacterial Porins and the Permeation of Antibiotics. *Chem. Rev.* **121**, 5158–5192 (2021).
43. Tommasi, R., Iyer, R. & Miller, A. A. Antibacterial Drug Discovery: Some Assembly Required. *ACS Infect. Dis.* **4**, 686–695 (2018).
44. Looper, R. E., Boger, D. L. & Silver, L. L. Small Molecular Weapons against Multi-Drug Resistance. *Acc. Chem. Res.* **54**, 2785–2787 (2021).
45. Singh, S. B., Young, K. & Silver, L. L. What is an “ideal” antibiotic? Discovery challenges and path forward. *Biochemical Pharmacology* **133**, 63–73 (2017).
46. Kim, G.-L., Seon, S.-H. & Rhee, D.-K. Pneumonia and Streptococcus pneumoniae vaccine. *Arch Pharm Res* **40**, 885–893 (2017).
47. Ghosh, C., Sarkar, P., Issa, R. & Haldar, J. Alternatives to Conventional Antibiotics in the Era of Antimicrobial Resistance. *Trends in Microbiology* **27**, 323–338 (2019).
48. Furfaro, L. L., Payne, M. S. & Chang, B. J. Bacteriophage Therapy: Clinical Trials and Regulatory Hurdles. *Frontiers in Cellular and Infection Microbiology* **8**, (2018).
49. Sulakvelidze, A., Alavidze, Z. & Morris, J. G. Bacteriophage Therapy. *Antimicrob Agents Chemother* **45**, 649–659 (2001).
50. Dimitriu, T. *et al.* Bacteriostatic antibiotics promote CRISPR-Cas adaptive immunity by enabling increased spacer acquisition. *Cell Host & Microbe* **30**, 31-40.e5 (2022).
51. Silver, L. L. A Gestalt approach to Gram-negative entry. *Bioorganic & Medicinal Chemistry* **24**, 6379–6389 (2016).
52. Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol Rev* **40**, 133–159 (2016).
53. Kojima, S. & Nikaido, H. Permeation rates of penicillins indicate that Escherichia coli porins function principally as nonspecific channels. *PNAS* **110**, E2629–E2634 (2013).
54. Nikaido, H. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
55. Delcour, A. H. Outer Membrane Permeability and Antibiotic Resistance. *Biochim Biophys Acta* **1794**, 808–816 (2009).
56. Cowan, S. W. *et al.* Crystal structures explain functional properties of two E. coli porins. *Nature* **358**, 727–733 (1992).
57. Danelon, C., Nestorovich, E. M., Winterhalter, M., Ceccarelli, M. & Bezrukov, S. M. Interaction of Zwitterionic Penicillins with the OmpF Channel Facilitates Their Translocation. *Biophys J* **90**, 1617–1627 (2006).
58. Simonet, V., Malléa, M. & Pagès, J.-M. Substitutions in the Eyelet Region Disrupt Cefepime Diffusion through the Escherichia coli OmpF Channel. *Antimicrobial Agents and Chemotherapy* **44**, 311–315 (2000).
59. Ziervogel, B. K. & Roux, B. The Binding of Antibiotics in OmpF Porin. *Structure* **21**, 76–87 (2013).
60. Bajaj, H. *et al.* Bacterial Outer Membrane Porins as Electrostatic Nanosieves: Exploring Transport Rules of Small Polar Molecules. *ACS Nano* **11**, 5465–5473 (2017).

61. Bajaj, H. *et al.* Molecular Basis of Filtering Carbapenems by Porins from β -Lactam-resistant Clinical Strains of *Escherichia coli*. *J Biol Chem* **291**, 2837–2847 (2016).
62. Ude, J. *et al.* Outer membrane permeability: Antimicrobials and diverse nutrients bypass porins in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences* **118**, e2107644118 (2021).
63. Janssen, A. B. & Schaik, W. van. Harder, better, faster, stronger: Colistin resistance mechanisms in *Escherichia coli*. *PLOS Genetics* **17**, e1009262 (2021).
64. Henry, R. *et al.* Colistin-Resistant, Lipopolysaccharide-Deficient *Acinetobacter baumannii* Responds to Lipopolysaccharide Loss through Increased Expression of Genes Involved in the Synthesis and Transport of Lipoproteins, Phospholipids, and Poly- β -1,6-N-Acetylglucosamine. *Antimicrob Agents Chemother* **56**, 59–69 (2012).
65. Lou, H. *et al.* Altered Antibiotic Transport in OmpC Mutants Isolated from a Series of Clinical Strains of Multi-Drug Resistant *E. coli*. *PLoS One* **6**, (2011).
66. Nikaido, H. & Rosenberg, E. Y. Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins. *J Bacteriol* **153**, 241–252 (1983).
67. Kuhn, P., Eyer, K., Allner, S., Lombardi, D. & Dittrich, P. S. A Microfluidic Vesicle Screening Platform: Monitoring the Lipid Membrane Permeability of Tetracyclines. *Anal. Chem.* **83**, 8877–8885 (2011).
68. Cama, J. *et al.* Quantification of Fluoroquinolone Uptake through the Outer Membrane Channel OmpF of *Escherichia coli*. *J. Am. Chem. Soc.* **137**, 13836–13843 (2015).
69. Six, D. A., Krucker, T. & Leeds, J. A. Advances and challenges in bacterial compound accumulation assays for drug discovery. *Curr Opin Chem Biol* **44**, 9–15 (2018).
70. Zimmermann, W. & Rosselet, A. Function of the outer membrane of *Escherichia coli* as a permeability barrier to beta-lactam antibiotics. *Antimicrob Agents Chemother* **12**, 368–372 (1977).
71. Novick, R. P. Micro-iodometric assay for penicillinase. *Biochem J* **83**, 236–240 (1962).
72. Perret, C. J. Iodometric assay of penicillinase. *Nature* **174**, 1012–1013 (1954).
73. Williams, K. J. & Piddock, L. J. Accumulation of rifampicin by *Escherichia coli* and *Staphylococcus aureus*. *J. Antimicrob. Chemother.* **42**, 597–603 (1998).
74. Capobianco, J. O. & Goldman, R. C. Macrolide transport in *Escherichia coli* strains having normal and altered OmpC and/or OmpF porins. *International Journal of Antimicrobial Agents* **4**, 183–189 (1994).
75. Li, X. Z., Livermore, D. M. & Nikaido, H. Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: resistance to tetracycline, chloramphenicol, and norfloxacin. *Antimicrobial Agents and Chemotherapy* **38**, 1732–1741 (1994).
76. Thanassi, D. G., Suh, G. S. & Nikaido, H. Role of outer membrane barrier in efflux-mediated tetracycline resistance of *Escherichia coli*. *Journal of Bacteriology* **177**, 998–1007 (1995).
77. Cai, H., Rose, K., Liang, L.-H., Dunham, S. & Stover, C. Development of a liquid chromatography/mass spectrometry-based drug accumulation assay in *Pseudomonas aeruginosa*. *Analytical Biochemistry* **385**, 321–325 (2009).
78. Piddock, L. J. V., Jin, Y.-F., Ricci, V. & Asuquo, A. E. Quinolone accumulation by

- Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Escherichia coli*. *Journal of Antimicrobial Chemotherapy* **43**, 61–70 (1999).
79. Cama, J. *et al.* Direct Optofluidic Measurement of the Lipid Permeability of Fluoroquinolones. *Scientific Reports* **6**, srep32824 (2016).
 80. Leive, L., Telesetsky, S., Coleman, W. G. & Carr, D. Tetracyclines of various hydrophobicities as a probe for permeability of *Escherichia coli* outer membranes. *Antimicrobial Agents and Chemotherapy* **25**, 539–544 (1984).
 81. Chopra, I. Transport of tetracyclines into *Escherichia coli* requires a carboxamide group at the C2 position of the molecule. *Journal of Antimicrobial Chemotherapy* **18**, 661–666 (1986).
 82. Vergalli, J. *et al.* Spectrofluorimetric quantification of antibiotic drug concentration in bacterial cells for the characterization of translocation across bacterial membranes. *Nat Protoc* **13**, 1348–1361 (2018).
 83. Cama, J. *et al.* Single-cell microfluidics facilitates the rapid quantification of antibiotic accumulation in Gram-negative bacteria. *Lab on a Chip* **20**, 2765–2775 (2020).
 84. Allam, A. *et al.* Microspectrofluorimetry to dissect the permeation of ceftazidime in Gram-negative bacteria. *Sci Rep* **7**, 986 (2017).
 85. Łapińska, U. *et al.* Fast bacterial growth reduces antibiotic accumulation and efficacy. *eLife* **11**, e74062 (2022).
 86. Stone, M. R. *et al.* Fluoroquinolone-derived fluorescent probes for studies of bacterial penetration and efflux. *MedChemComm* **10**, 901–906 (2019).
 87. Stone, M. R. L. *et al.* Fluorescent macrolide probes – synthesis and use in evaluation of bacterial resistance. *RSC Chem. Biol.* **1**, 395–404 (2020).
 88. Pagès, J.-M. *et al.* New Peptide-based antimicrobials for tackling drug resistance in bacteria: single-cell fluorescence imaging. *ACS Med Chem Lett* **4**, 556–559 (2013).
 89. Richter, M. F. *et al.* Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **545**, 299–304 (2017).
 90. Zhou, Y. *et al.* Thinking Outside the “Bug”: A Unique Assay To Measure Intracellular Drug Penetration in Gram-Negative Bacteria. *Anal. Chem.* **87**, 3579–3584 (2015).
 91. Davis, T. D., Gerry, C. J. & Tan, D. S. General Platform for Systematic Quantitative Evaluation of Small-Molecule Permeability in Bacteria. *ACS Chem Biol* **9**, 2535–2544 (2014).
 92. Iyer, R. *et al.* Evaluating LC-MS/MS To Measure Accumulation of Compounds within Bacteria. *ACS Infect Dis* **4**, 1336–1345 (2018).
 93. Bhat, J., Narayan, A., Venkatraman, J. & Chatterji, M. LC–MS based assay to measure intracellular compound levels in *Mycobacterium smegmatis*: Linking compound levels to cellular potency. *Journal of Microbiological Methods* **94**, 152–158 (2013).
 94. Schumacher, A. *et al.* Intracellular accumulation of linezolid in *Escherichia coli*, *Citrobacter freundii* and *Enterobacter aerogenes*: role of enhanced efflux pump activity and inactivation. *Journal of Antimicrobial Chemotherapy* **59**, 1261–1264 (2007).
 95. Prochnow, H. *et al.* Subcellular Quantification of Uptake in Gram-Negative Bacteria. *Anal. Chem.* **91**, 1863–1872 (2019).
 96. Widya, M. *et al.* Development and Optimization of a Higher-Throughput Bacterial Compound Accumulation Assay. *ACS Infectious Diseases* (2019)

doi:10.1021/acsinfecdis.8b00299.

97. Dumont, E. *et al.* Antibiotics and efflux: combined spectrofluorimetry and mass spectrometry to evaluate the involvement of concentration and efflux activity in antibiotic intracellular accumulation. *Journal of Antimicrobial Chemotherapy* **74**, 58–65 (2019).
98. Rath, C. M. *et al.* Optimization of CoaD Inhibitors against Gram-Negative Organisms through Targeted Metabolomics. *ACS Infect. Dis.* **4**, 391–402 (2018).
99. Tian, H., Six, D. A., Krucker, T., Leeds, J. A. & Winograd, N. Subcellular Chemical Imaging of Antibiotics in Single Bacteria Using C60-Secondary Ion Mass Spectrometry. *Anal Chem* **89**, 5050–5057 (2017).
100. Iyer, R. *et al.* Whole-Cell-Based Assay To Evaluate Structure Permeation Relationships for Carbapenem Passage through the *Pseudomonas aeruginosa* Porin OprD. *ACS Infect Dis* **3**, 310–319 (2017).
101. Durand-Reville, T. F. *et al.* Rational design of a new antibiotic class for drug-resistant infections. *Nature* **597**, 698–702 (2021).
102. Krishnamoorthy, G., Wolloscheck, D., Weeks, J. W., Croft, C. & Rybenkov, V. V. Breaking the Permeability Barrier of *Escherichia coli* by Controlled Hyperporination of the Outer Membrane. *Antimicrob. Agents Chemother.* **60**, 7372–7381 (2016).
103. Cooper, C. J. *et al.* Molecular Properties That Define the Activities of Antibiotics in *Escherichia coli* and *Pseudomonas aeruginosa*. *ACS Infect Dis* **4**, 1223–1234 (2018).
104. Mansbach, R. A. *et al.* Machine Learning Algorithm Identifies an Antibiotic Vocabulary for Permeating Gram-Negative Bacteria. *Journal of Chemical Information and Modeling* (2020) doi:10.1021/acs.jcim.0c00352.
105. Ferreira, R. J. & Kasson, P. M. Antibiotic Uptake Across Gram-Negative Outer Membranes: Better Predictions Towards Better Antibiotics. *ACS Infect Dis* **5**, 2096–2104 (2019).
106. Tran, Q.-T., Williams, S., Farid, R., Erdemli, G. & Pearlstein, R. The translocation kinetics of antibiotics through porin OmpC: insights from structure-based solvation mapping using WaterMap. *Proteins* **81**, 291–299 (2013).
107. Ceccarelli, M., Vargiu, A. V. & Ruggerone, P. A kinetic Monte Carlo approach to investigate antibiotic translocation through bacterial porins. *J. Phys.: Condens. Matter* **24**, 104012 (2012).
108. Kumar, A., Hajjar, E., Ruggerone, P. & Ceccarelli, M. Structural and dynamical properties of the porins OmpF and OmpC: insights from molecular simulations. *J. Phys.: Condens. Matter* **22**, 454125 (2010).
109. Richter, R. *et al.* An Outer Membrane Vesicle-Based Permeation Assay (OMPA) for Assessing Bacterial Bioavailability. *Advanced Healthcare Materials* **11**, 2101180 (2022).
110. Durand-Réville, T. F. *et al.* ETX2514 is a broad-spectrum β -lactamase inhibitor for the treatment of drug-resistant Gram-negative bacteria including *Acinetobacter baumannii*. *Nat Microbiol* **2**, 1–10 (2017).
111. Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688–702.e13 (2020).
112. Kulp, A. & Kuehn, M. J. Biological Functions and Biogenesis of Secreted Bacterial Outer Membrane Vesicles. *Annu. Rev. Microbiol.* **64**, 163–184 (2010).

113. Jan, A. T. Outer Membrane Vesicles (OMVs) of Gram-negative Bacteria: A Perspective Update. *Front Microbiol* **8**, (2017).
114. Kim, S. W. *et al.* Outer membrane vesicles from β -lactam-resistant Escherichia coli enable the survival of β -lactam-susceptible E. coli in the presence of β -lactam antibiotics. *Scientific Reports* **8**, 5402 (2018).
115. Kim, S. W. *et al.* The Importance of Porins and β -Lactamase in Outer Membrane Vesicles on the Hydrolysis of β -Lactam Antibiotics. *Int J Mol Sci* **21**, (2020).
116. O'Shea, R. & Moser, H. E. Physicochemical Properties of Antibacterial Compounds: Implications for Drug Discovery. *J. Med. Chem.* **51**, 2871–2878 (2008).
117. Brown, D. G., May-Dracka, T. L., Gagnon, M. M. & Tommasi, R. Trends and Exceptions of Physical Properties on Antibacterial Activity for Gram-Positive and Gram-Negative Pathogens. *J. Med. Chem.* **57**, 10144–10161 (2014).
118. Bertani, G. STUDIES ON LYSOGENESIS I.: The Mode of Phage Liberation by Lysogenic Escherichia coli1. *Journal of Bacteriology* **62**, 293 (1951).
119. Grossman, T. H., Kawasaki, E. S., Punreddy, S. R. & Osburne, M. S. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* **209**, 95–103 (1998).
120. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology* **166**, 557–580 (1983).
121. Ina Meuskens, M. M. A New Strain Collection for Improved Expression of Outer Membrane Proteins. *Frontiers in Cellular and Infection Microbiology* **7**, (2017).
122. Meuskens, I., Michalik, M., Chauhan, N., Linke, D. & Leo, J. C. A New Strain Collection for Improved Expression of Outer Membrane Proteins. *Frontiers in Cellular and Infection Microbiology* **7**, (2017).
123. A Protein-Free Medium for Primary Isolation of the Gonococcus and Meningococcus - J. Howard Mueller, Jane Hinton, 1941. <https://journals.sagepub.com/doi/abs/10.3181/00379727-48-13311>.
124. Levine, M. Z. *et al.* Activation of Energy Metabolism through Growth Media Reformulation Enables a 24-Hour Workflow for Cell-Free Expression. *ACS Synth. Biol.* **9**, 2765–2774 (2020).
125. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2**, 2006.0008 (2006).
126. Hanahan, D., Jessee, J. & Bloom, F. R. [4] Plasmid transformation of Escherichia coli and other bacteria. in *Methods in Enzymology* vol. 204 63–113 (Academic Press, 1991).
127. Green, R. & Rogers, E. J. Chemical Transformation of E. coli. *Methods in enzymology* **529**, 329 (2013).
128. Weinstein, M. P. & Patel, J. B. *National Committee for Clinical Laboratory Standards. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically—11th Edition: Approved Standard M07-A11.* (Committee for Clinical Laboratory Standards, 2018).
129. Shin, J. & Noireaux, V. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **1**, 29–41 (2012).
130. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254 (1976).
131. Smith, P. K. *et al.* Measurement of protein using bicinchoninic acid. *Analytical*

- Biochemistry* **150**, 76–85 (1985).
132. Sambrook, J. *Molecular Cloning: A Laboratory Manual. volume 3.* (Cold Spring Harbor Laboratory Press, 2001).
 133. Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685 (1970).
 134. Zhao, G., Meier, T. I., Kahl, S. D., Gee, K. R. & Blaszcak, L. C. BOCILLIN FL, a Sensitive and Commercially Available Reagent for Detection of Penicillin-Binding Proteins. *Antimicrob Agents Chemother* **43**, 1124–1128 (1999).
 135. Kotagiri, N., Sakon, J., Han, H., Zharov, V. P. & Kim, J.-W. Fluorescent ampicillin analogues as multifunctional disguising agents against opsonization. *Nanoscale* **8**, 12658–12667 (2016).
 136. Catherwood, A. C. *et al.* Substrate and stereochemical control of peptidoglycan crosslinking by transpeptidation by *Escherichia coli* PBP1B. *J. Am. Chem. Soc.* (2020) doi:10.1021/jacs.9b08822.
 137. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).
 138. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367–1372 (2008).
 139. Searle, B. C. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **10**, 1265–1269 (2010).
 140. Garamella, J., Marshall, R., Rustad, M. & Noireaux, V. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synth. Biol.* **5**, 344–355 (2016).
 141. Garamella, J., Garenne, D. & Noireaux, V. Chapter Nine - TXTL-based approach to synthetic cells. in *Methods in Enzymology* (eds. Schmidt-Dannert, C. & Quin, M. B.) vol. 617 217–239 (Academic Press, 2019).
 142. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences* **101**, 17669–17674 (2004).
 143. Shin, J. & Noireaux, V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J Biol Eng* **4**, 8 (2010).
 144. Sun, Z. Z. *et al.* Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J Vis Exp* e50762 (2013) doi:10.3791/50762.
 145. Soga, H. *et al.* In Vitro Membrane Protein Synthesis Inside Cell-Sized Vesicles Reveals the Dependence of Membrane Protein Integration on Vesicle Volume. *ACS Synth. Biol.* **3**, 372–379 (2014).
 146. Litschel, T. *et al.* Freeze-thaw cycles induce content exchange between cell-sized lipid vesicles. *New J. Phys.* **20**, 055008 (2018).
 147. Moga, A., Yandrapalli, N., Dimova, R. & Robinson, T. Optimization of the Inverted Emulsion Method for High-Yield Production of Biomimetic Giant Unilamellar Vesicles. *ChemBioChem* **20**, 2674–2682 (2019).
 148. Tamba, Y. & Yamazaki, M. Single Giant Unilamellar Vesicle Method Reveals Effect of Antimicrobial Peptide Magainin 2 on Membrane Permeability. *Biochemistry* **44**, 15823–15833 (2005).
 149. Moniruzzaman, Md., Alam, J. Md., Dohra, H. & Yamazaki, M. Antimicrobial Peptide Lactoferricin B-Induced Rapid Leakage of Internal Contents from Single Giant

- Unilamellar Vesicles. *Biochemistry* **54**, 5802–5814 (2015).
150. Jaeger, S., Fulle, S. & Turk, S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J. Chem. Inf. Model.* **58**, 27–35 (2018).
 151. Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010).
 152. Duvenaud, D. *et al.* Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv:1509.09292 [cs, stat]* (2015).
 153. Ramsundar, B., Eastman, P., Walters, P. & Pande, V. *Deep Learning for the Life Sciences: Applying Deep Learning to Genomics, Microscopy, Drug Discovery, and More.* (O'Reilly Media, 2019).
 154. Landrum, G. Rdkit: Open-source cheminformatics software. (2016).
 155. Morgan, H. L. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *J. Chem. Doc.* **5**, 107–113 (1965).
 156. Dolk, F. C. K., Pouwels, K. B., Smith, D. R. M., Robotham, J. V. & Smieszek, T. Antibiotics in primary care in England: which antibiotics are prescribed and for which conditions? *Journal of Antimicrobial Chemotherapy* **73**, (2018).
 157. Puiggali-Jou, A. *et al.* Properties of Omp2a-Based Supported Lipid Bilayers: Comparison with Polymeric Bioinspired Membranes. *ACS Omega* **3**, 9003–9019 (2018).
 158. Jaffe, A., Chabbert, Y. A. & Semonin, O. Role of porin proteins OmpF and OmpC in the permeation of beta-lactams. *Antimicrob. Agents Chemother.* **22**, 942–948 (1982).
 159. Cowan, S. W. *et al.* The structure of OmpF porin in a tetragonal crystal form. *Structure* **3**, 1041–1050 (1995).
 160. Chevalier, J., Pagès, J. M. & Malléa, M. In vivo modification of porin activity conferring antibiotic resistance to *Enterobacter aerogenes*. *Biochem Biophys Res Commun* **266**, 248–251 (1999).
 161. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612 (2004).
 162. Yamashita, E., Zhalnina, M. V., Zakharov, S. D., Sharma, O. & Cramer, W. A. Crystal structures of the OmpF porin: function in a colicin translocon. *EMBO J* **27**, 2171–2180 (2008).
 163. Stone, M. R. L., Butler, M. S., Phetsang, W., Cooper, M. A. & Blaskovich, M. A. T. Fluorescent Antibiotics: New Research Tools to Fight Antibiotic Resistance. *Trends in Biotechnology* **36**, 523–536 (2018).
 164. Miao, L., Liu, W., Qiao, Q., Li, X. & Xu, Z. Fluorescent antibiotics for real-time tracking of pathogenic bacteria. *Journal of Pharmaceutical Analysis* **10**, 444–451 (2020).
 165. Shapiro, A. B., Gu, R.-F., Gao, N., Livchak, S. & Thresher, J. Continuous fluorescence anisotropy-based assay of BOCILLIN FL penicillin reaction with penicillin binding protein 3. *Anal Biochem* **439**, 37–43 (2013).
 166. Ghai, I. & Ghai, S. Understanding antibiotic resistance via outer membrane permeability. *Infect Drug Resist* **11**, 523–530 (2018).
 167. Jusko, W. J. Fluorometric analysis of ampicillin in biological fluids. *Journal of Pharmaceutical Sciences* **60**, 728–732 (1971).
 168. Barbhaiya, R. H. & Turner, P. Fluorimetric determination of ampicillin and

- cephalexin [proceedings]. *Br J Pharmacol* **58**, 473P (1976).
169. Uno, T., Masada, M., Kuroda, Y. & Nakagawa, N. Isolation and structural investigation of the fluorescent degradation products of ampicillin. *Chemical and Pharmaceutical Bulletin* **Volume 29**, 1344–1354 (1981).
 170. Baertschi, S. W. *et al.* Isolation and Structure Elucidation of the Major Degradation Products of Cefaclor Formed Under Aqueous Acidic Conditions. *Journal of Pharmaceutical Sciences* **86**, 526–539 (1997).
 171. Roets, E. *et al.* Isolation and structure elucidation of ampicillin and amoxicillin oligomers. *Journal of Chromatography A* **303**, 117–129 (1984).
 172. Zhang, L. *et al.* Identification of the related substances in ampicillin capsule by rapid resolution liquid chromatography coupled with electrospray ionization tandem mass spectrometry. *J Anal Methods Chem* **2014**, 397492 (2014).
 173. Frański, R., Czerniel, J., Kowalska, M. & Frańska, M. Electrospray ionization collision-induced dissociation tandem mass spectrometry of amoxicillin and ampicillin and their degradation products. *Rapid Communications in Mass Spectrometry* **28**, 713–722 (2014).
 174. Li, T. *et al.* Structural elucidation of stress degradation products of ampicillin sodium by liquid chromatography/hybrid triple quadrupole linear ion trap mass spectrometry and liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **28**, 1929–1936 (2014).
 175. Wu, Q., Zhang, X., Du, J. & Hu, C. Discussion on the dimerization reaction of penicillin antibiotics. *Journal of Pharmaceutical Analysis* (2021) doi:10.1016/j.jpha.2021.06.005.
 176. Robinson-Fuentes, V. A., Jefferies, T. M. & Branch, S. K. Degradation Pathways of Ampicillin in Alkaline Solutions. *Journal of Pharmacy and Pharmacology* **49**, 843–851 (1997).
 177. Gutheil, W. G., Stefanova, M. E. & Nicholas, R. A. Fluorescent Coupled Enzyme Assays for d-Alanine: Application to Penicillin-Binding Protein and Vancomycin Activity Assays. *Analytical Biochemistry* **287**, 196–202 (2000).
 178. Clarke, T. B. *et al.* Mutational Analysis of the Substrate Specificity of Escherichia coli Penicillin Binding Protein 4. *Biochemistry* **48**, 2675–2683 (2009).
 179. Mitchell, S. M., Ullman, J. L., Teel, A. L. & Watts, R. J. pH and temperature effects on the hydrolysis of three β -lactam antibiotics: ampicillin, cefalotin and cefoxitin. *Sci Total Environ* **466–467**, 547–555 (2014).
 180. Entryway. <http://www.entry-way.org/>.
 181. Firth, N. C., Brown, N. & Blagg, J. Plane of Best Fit: A Novel Method to Characterize the Three-Dimensionality of Molecules. *J. Chem. Inf. Model.* **52**, 2516–2525 (2012).
 182. Aunkham, A., Schulte, A., Winterhalter, M. & Suginta, W. Porin Involvement in Cephalosporin and Carbapenem Resistance of Burkholderia pseudomallei. *PLoS One* **9**, (2014).
 183. Stone, M. R. L., Phetsang, W., Cooper, M. A. & Blaskovich, M. A. T. Visualization of Bacterial Resistance using Fluorescent Antibiotic Probes. *JoVE (Journal of Visualized Experiments)* e60743 (2020) doi:10.3791/60743.
 184. Ferrer-Espada, R. *et al.* A permeability-increasing drug synergizes with bacterial efflux pump inhibitors and restores susceptibility to antibiotics in multi-drug resistant Pseudomonas aeruginosa strains. *Scientific Reports* **9**, 3452 (2019).

185. Stokes, J. M. *et al.* Cold Stress Makes Escherichia coli Susceptible to Glycopeptide Antibiotics by Altering Outer Membrane Integrity. *Cell Chemical Biology* **23**, 267–277 (2016).
186. Ross, G. W. *et al.* Comparison of assay techniques for β -lactamase activity. *Analytical Biochemistry* **54**, 9–16 (1973).
187. Matsumura, N., Minami, S., Watanabe, Y., Iyobe, S. & Mitsuhashi, S. Role of Permeability in the Activities of β -Lactams against Gram-Negative Bacteria Which Produce a Group 3 β -Lactamase. *Antimicrob Agents Chemother* **43**, 2084–2086 (1999).
188. Zapun, A., Contreras-Martel, C. & Vernet, T. Penicillin-binding proteins and β -lactam resistance. *FEMS Microbiol Rev* **32**, 361–385 (2008).
189. Ghuysen, J.-M. Bacterial Active-Site Serine Penicillin-Interactive Proteins and Domains: Mechanism, Structure, and Evolution. *Reviews of Infectious Diseases* **10**, 726–732 (1988).
190. Sauvage, E. & Terrak, M. Glycosyltransferases and Transpeptidases/Penicillin-Binding Proteins: Valuable Targets for New Antibacterials. *Antibiotics (Basel)* **5**, (2016).
191. Macheboeuf, P., Contreras-Martel, C., Job, V., Dideberg, O. & Dessen, A. Penicillin Binding Proteins: key players in bacterial cell cycle and drug resistance processes. *FEMS Microbiol Rev* **30**, 673–691 (2006).
192. Tipper, D. J. & Strominger, J. L. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A* **54**, 1133–1141 (1965).
193. Beadle, B. M., Nicholas, R. A. & Shoichet, B. K. Interaction energies between β -lactam antibiotics and E. coli penicillin-binding protein 5 by reversible thermal denaturation. *Protein Sci* **10**, 1254–1259 (2001).
194. Lu, W.-P., Kincaid, E., Sun, Y. & Bauer, M. D. Kinetics of β -Lactam Interactions with Penicillin-susceptible and -resistant Penicillin-binding Protein 2x Proteins from Streptococcus pneumoniae: INVOLVEMENT OF ACYLATION AND DEACYLATION IN β -LACTAM RESISTANCE *. *Journal of Biological Chemistry* **276**, 31494–31501 (2001).
195. Dass, C. *Fundamentals of Contemporary Mass Spectrometry*. (John Wiley & Sons, 2007).
196. Ongas, M. *et al.* Liquid chromatography–tandem mass spectrometry for the simultaneous quantitation of ceftriaxone, metronidazole and hydroxymetronidazole in plasma from seriously ill, severely malnourished children. *Wellcome Open Research* **2**, (2017).
197. Wang, H., Huang, H., Cao, J., Chui, D. & Xiao, S. Mass Spectral Profile for Rapid Differentiating Beta-Lactams from Their Ring-Opened Impurities. *BioMed Research International* **2015**, e697958 (2015).
198. Ho, C. *et al.* Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Rev* **24**, 3–12 (2003).
199. Awad, H., Khamis, M. M. & El-Aneed, A. Mass Spectrometry, Review of the Basics: Ionization. *Applied Spectroscopy Reviews* **50**, 158–175 (2015).
200. N. Nikolaev, E., N. Vladimirov, G., Jertz, R. & Baykut, G. From Supercomputer Modeling to Highest Mass Resolution in FT-ICR. *Mass Spectrom (Tokyo)* **2**, S0010 (2013).

201. Kovalev, S. V. & Lebedev, A. T. Chapter 14 - Identification of biologically active peptides by means of Fourier transform mass spectrometry. in *Fundamentals and Applications of Fourier Transform Mass Spectrometry* (eds. Kanawati, B. & Schmitt-Kopplin, P.) 425–468 (Elsevier, 2019). doi:10.1016/B978-0-12-814013-0.00014-4.
202. Lange, O., Damoc, E., Wiegghaus, A. & Makarov, A. Enhanced Fourier transform for Orbitrap mass spectrometry. *International Journal of Mass Spectrometry* **369**, 16–22 (2014).
203. Zubarev, R. A. & Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **85**, 5288–5296 (2013).
204. Schumacher, A. *et al.* Intracellular accumulation of linezolid in *Escherichia coli*, *Citrobacter freundii* and *Enterobacter aerogenes*: role of enhanced efflux pump activity and inactivation. *J Antimicrob Chemother* **59**, 1261–1264 (2007).
205. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837–846 (2000).
206. Gregorich, Z. R., Chang, Y.-H. & Ge, Y. Proteomics in heart failure: top-down or bottom-up? *Pflugers Arch* **466**, 1199–1209 (2014).
207. Bludau, I. *et al.* Systematic detection of functional proteoform groups from bottom-up proteomic datasets. *Nat Commun* **12**, 3810 (2021).
208. Dupree, E. J. *et al.* A Critical Review of Bottom-Up Proteomics: The Good, the Bad, and the Future of This Field. *Proteomes* **8**, (2020).
209. Catherman, A. D., Skinner, O. S. & Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochem Biophys Res Commun* **445**, 683–693 (2014).
210. Zhang, Y. *et al.* Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical reviews* **113**, 2343 (2013).
211. PeptideMass. https://web.expasy.org/peptide_mass/.
212. SPC Proteomics Tools. <http://tools.proteomecenter.org/software.php>.
213. Sauvage, E., Kerff, F., Terrak, M., Ayala, J. A. & Charlier, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews* **32**, 234–258 (2008).
214. Sutton, L. D., Biedenbach, D. J., Yen, A. & Jones, R. N. Development, characterization, and initial evaluations of S1 a new chromogenic cephalosporin for β -lactamase detection. *Diagnostic Microbiology and Infectious Disease* **21**, 1–8 (1995).
215. Newman, H. Biophysical and biochemical studies of penicillin binding proteins and novel PBP3 inhibitors. (University of Warwick, 2021).
216. Wiersinga, W. J., van der Poll, T., White, N. J., Day, N. P. & Peacock, S. J. Melioidosis: insights into the pathogenicity of *Burkholderia pseudomallei*. *Nat Rev Microbiol* **4**, 272–282 (2006).
217. Stefani, S. *et al.* Relevance of multidrug-resistant *Pseudomonas aeruginosa* infections in cystic fibrosis. *Int J Med Microbiol* **307**, 353–362 (2017).
218. Sainsbury, S. *et al.* Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. *J Mol Biol* **405**, 173–184 (2011).
219. Fontana, R., Cornaglia, G., Ligozzi, M. & Mazzariol, A. The final goal: penicillin-binding proteins and the target of cephalosporins. *Clinical Microbiology and Infection* **6**, 34–40 (2000).

220. Edoó, Z., Arthur, M. & Hugonnet, J.-E. Reversible inactivation of a peptidoglycan transpeptidase by a β -lactam antibiotic mediated by β -lactam-ring recyclization in the enzyme active site. *Sci Rep* **7**, 9136 (2017).
221. Wang, F. *et al.* CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem.* **93**, 11692–11700 (2021).
222. El-Aneed, A., Cohen, A. & Banoub, J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* **44**, 210–230 (2009).
223. Hong, Y.-H. *et al.* A high-accuracy screening method of 44 cephalosporins in meat using liquid chromatography quadrupole-orbitrap hybrid mass spectrometry. *Anal. Methods* **9**, 6534–6548 (2017).
224. Roychoudhury, S., Kaiser, R. E., Brems, D. N. & Yeh, W. K. Specific interaction between beta-lactams and soluble penicillin-binding protein 2a from methicillin-resistant *Staphylococcus aureus*: development of a chromogenic assay. *Antimicrobial Agents and Chemotherapy* **40**, 2075 (1996).
225. Decosterd, L. A. *et al.* Validation and clinical application of a multiplex high performance liquid chromatography – tandem mass spectrometry assay for the monitoring of plasma concentrations of 12 antibiotics in patients with severe bacterial infections. *Journal of Chromatography B* **1157**, 122160 (2020).
226. Sparbier, K., Schubert, S., Weller, U., Boogen, C. & Kostrzewa, M. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry-Based Functional Assay for Rapid Detection of Resistance against β -Lactam Antibiotics. *Journal of Clinical Microbiology* **50**, 927–937 (2012).
227. Kostrzewa, M., Sparbier, K., Maier, T. & Schubert, S. MALDI-TOF MS: an upcoming tool for rapid detection of antibiotic resistance in microorganisms. *PROTEOMICS – Clinical Applications* **7**, 767–778 (2013).
228. Rideau, E., Dimova, R., Schwillle, P., Wurm, F. R. & Landfester, K. Liposomes and polymersomes: a comparative review towards cell mimicking. *Chem. Soc. Rev.* **47**, 8572–8610 (2018).
229. Sforzi, J., Palagi, L. & Aime, S. Liposome-Based Bioassays. *Biology* **9**, (2020).
230. Nasr, G., Greige-Gerges, H., Elaissari, A. & Khreich, N. Liposomal membrane permeability assessment by fluorescence techniques: Main permeabilizing agents, applications and challenges. *International Journal of Pharmaceutics* **580**, 119198 (2020).
231. Akbarzadeh, A. *et al.* Liposome: classification, preparation, and applications. *Nanoscale Res Lett* **8**, 102 (2013).
232. Nahas, K. A. *et al.* A microfluidic platform for the characterisation of membrane active antimicrobials. *Lab on a Chip* **19**, 837–844 (2019).
233. Elani, D. Y. Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology. *Angewandte Chemie (International Ed. in English)* **60**, 5602 (2021).
234. Pelletier, J. F. *et al.* Genetic requirements for cell division in a genomically minimal cell. *Cell* **184**, 2430–2440.e16 (2021).
235. Chakrabarti, A. C., Breaker, R. R., Joyce, G. F. & Deamer, D. W. Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *J Mol Evol* **39**, 555–559 (1994).
236. Oberholzer, T., Albrizio, M. & Luisi, P. L. Polymerase chain reaction in liposomes.

- Chemistry & Biology* **2**, 677–682 (1995).
237. Stano, P., Carrara, P., Kuruma, Y., Souza, T. P. de & Luisi, P. L. Compartmentalized reactions as a case of soft-matter biotechnology: synthesis of proteins and nucleic acids inside lipid vesicles. *J. Mater. Chem.* **21**, 18887–18902 (2011).
 238. Mavelli, F., Marangoni, R. & Stano, P. A Simple Protein Synthesis Model for the PURE System Operation. *Bull. Math. Biol.* **77**, 1185–1212 (2015).
 239. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc* **10**, 1328–1344 (2015).
 240. Gregorio, N. E., Levine, M. Z. & Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods and Protocols* **2**, 24 (2019).
 241. Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, D. S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *Chembiochem* **16**, 2420 (2015).
 242. Perez, J. G., Stark, J. C. & Jewett, M. C. Cell-Free Synthetic Biology: Engineering Beyond the Cell. *Cold Spring Harb Perspect Biol* **8**, a023853 (2016).
 243. Garenne, D., Beisel, C. L. & Noireaux, V. Characterization of the all-E. coli transcription-translation system myTXTL by mass spectrometry. *Rapid Communications in Mass Spectrometry* **33**, 1036–1048 (2019).
 244. Marshall, R. & Noireaux, V. Quantitative modeling of transcription and translation of an all- E. coli cell-free system. *Scientific Reports* **9**, 11980 (2019).
 245. Izri, Z., Garenne, D., Noireaux, V. & Maeda, Y. T. Gene Expression in on-Chip Membrane-Bound Artificial Cells. *ACS Synth. Biol.* **8**, 1705–1712 (2019).
 246. Hu, X.-P., Dourado, H., Schubert, P. & Lercher, M. J. The protein translation machinery is expressed for maximal efficiency in Escherichia coli. *Nat Commun* **11**, 5260 (2020).
 247. Sharma, B., Moghimianavval, H., Hwang, S.-W. & Liu, A. P. Synthetic Cell as a Platform for Understanding Membrane-Membrane Interactions. *Membranes* **11**, 912 (2021).
 248. Garenne, D. & Noireaux, V. Membrane functions genetically programmed in synthetic cells: A barrier to conquer. *Current Opinion in Systems Biology* **24**, 9–17 (2020).
 249. Nishimura, K. *et al.* Cell-Free Protein Synthesis inside Giant Unilamellar Vesicles Analyzed by Flow Cytometry. *Langmuir* **28**, 8426–8432 (2012).
 250. Tsuji, G., Sunami, T. & Ichihashi, N. Production of giant unilamellar vesicles by the water-in-oil emulsion-transfer method without high internal concentrations of sugars. *Journal of Bioscience and Bioengineering* **126**, 540–545 (2018).
 251. Sun, Z. Z. *et al.* Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *JoVE (Journal of Visualized Experiments)* e50762 (2013) doi:10.3791/50762.
 252. Donnelly, D. P. *et al.* Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat Methods* **16**, 587–594 (2019).
 253. Shin, J. & Noireaux, V. Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. *J Biol Eng* **4**, 9 (2010).
 254. Lee, I.-H., Passaro, S., Ozturk, S., Ureña, J. & Wang, W. Intelligent fluorescence image analysis of giant unilamellar vesicles using convolutional neural network. *BMC Bioinformatics* **23**, 48 (2022).

255. Nishimura, K. *et al.* Population Analysis of Structural Properties of Giant Liposomes by Flow Cytometry. *Langmuir* **25**, 10439–10443 (2009).
256. Morales-Pennington, N. F. *et al.* GUV preparation and imaging: Minimizing artifacts. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1798**, 1324–1332 (2010).
257. Dreher, Y., Jahnke, K., Bobkova, E., Spatz, J. P. & Göpfrich, K. Division and Regrowth of Phase-Separated Giant Unilamellar Vesicles**. *Angewandte Chemie International Edition* **60**, 10661–10669 (2021).
258. May, K. L. & Grabowicz, M. The bacterial outer membrane is an evolving antibiotic barrier. *Proc Natl Acad Sci USA* **115**, 8852–8854 (2018).
259. Paulowski, L. *et al.* The Beauty of Asymmetric Membranes: Reconstitution of the Outer Membrane of Gram-Negative Bacteria. *Frontiers in Cell and Developmental Biology* **8**, (2020).
260. Tresset, G. The multiple faces of self-assembled lipidic systems. *PMC Biophys* **2**, 3 (2009).
261. Düzgüneş, N., Wilschut, J., Fraley, R. & Papahadjopoulos, D. Studies on the mechanism of membrane fusion. Role of head-group composition in calcium- and magnesium-induced fusion of mixed phospholipid vesicles. *Biochim Biophys Acta* **642**, 182–195 (1981).
262. Patel, H., Tscheka, C. & Heerklotz, H. Characterizing vesicle leakage by fluorescence lifetime measurements. *Soft Matter* **5**, 2849–2851 (2009).
263. Deshpande, S., Caspi, Y., Meijering, A. E. C. & Dekker, C. Octanol-assisted liposome assembly on chip. *Nat Commun* **7**, 10447 (2016).
264. Yandrapalli, N., Petit, J., Bäumchen, O. & Robinson, T. Surfactant-free production of biomimetic giant unilamellar vesicles using PDMS-based microfluidics. *Commun Chem* **4**, 1–10 (2021).
265. Al Nahas, K. & Keyser, U. F. Standardizing characterization of membrane active peptides with microfluidics. *Biomicrofluidics* **15**, 041301 (2021).
266. Wimley, W. C. Describing the Mechanism of Antimicrobial Peptide Action with the Interfacial Activity Model. *ACS Chem. Biol.* **5**, 905–917 (2010).
267. Hjartarson, H. Vesicle Analyser Pro. (2019) doi:10.5281/zenodo.2592469.
268. Phale, P. S. *et al.* Stability of trimeric OmpF porin: the contributions of the latching loop L2. *Biochemistry* **37**, 15663–15670 (1998).
269. Schiffrin, B., Brockwell, D. J. & Radford, S. E. Outer membrane protein folding from an energy landscape perspective. *BMC Biology* **15**, 123 (2017).
270. Garenne, D. *et al.* Cell-free gene expression. *Nat Rev Methods Primers* **1**, 1–18 (2021).
271. Struyvé, M., Moons, M. & Tommassen, J. Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *Journal of Molecular Biology* **218**, 141–148 (1991).
272. Savage, D. F., Anderson, C. L., Robles-Colmenares, Y., Newby, Z. E. & Stroud, R. M. Cell-free complements in vivo expression of the E. coli membrane proteome. *Protein Science* **16**, 966–976 (2007).
273. Focke, P. J. *et al.* Combining In vitro Folding with Cell-Free Protein Synthesis for Membrane Protein Expression. *Biochemistry* **55**, 4212 (2016).
274. Smale, S. T. Luciferase assay. *Cold Spring Harb Protoc* **2010**, pdb.prot5421 (2010).
275. Widemann, B. C., Balis, F. M. & Adamson, P. C. Dihydrofolate Reductase Enzyme

- Inhibition Assay for Plasma Methotrexate Determination Using a 96-Well Microplate Reader. *Clinical Chemistry* **45**, 223–228 (1999).
276. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **323**, 844–853 (2020).
 277. Bajorath, J. State-of-the-art of artificial intelligence in medicinal chemistry. *Future Science OA* **7**, FSO702 (2021).
 278. Baskin, I. I., Winkler, D. & Tetko, I. V. A renaissance of neural networks in drug discovery. *Expert Opinion on Drug Discovery* **11**, 785–795 (2016).
 279. Chuang, K. V., Gunsalus, L. M. & Keiser, M. J. Learning Molecular Representations for Medicinal Chemistry. *J. Med. Chem.* **63**, 8705–8722 (2020).
 280. Alves, L. A. *et al.* Graph Neural Networks as a Potential Tool in Improving Virtual Screening Programs. *Frontiers in Chemistry* **9**, (2022).
 281. Vamathevan, J. *et al.* Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov* **18**, 463–477 (2019).
 282. Carracedo-Reboredo, P. *et al.* A review on machine learning approaches and trends in drug discovery. *Computational and Structural Biotechnology Journal* **19**, 4538–4558 (2021).
 283. Ivanenkov, Y. A. *et al.* Identification of Novel Antibacterials Using Machine Learning Techniques. *Frontiers in Pharmacology* **10**, 913 (2019).
 284. Zoffmann, S. *et al.* Machine learning-powered antibiotics phenotypic drug discovery. *Sci Rep* **9**, (2019).
 285. Yoshida, M. *et al.* Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides. *Chem* **4**, 533–543 (2018).
 286. Arango-Argoty, G. *et al.* DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).
 287. Liu, Z. *et al.* Evaluation of Machine Learning Models for Predicting Antimicrobial Resistance of *Actinobacillus pleuropneumoniae* From Whole Genome Sequences. *Frontiers in Microbiology* **11**, 48 (2020).
 288. Espinoza, J. L. *et al.* Predicting antimicrobial mechanism-of-action from transcriptomes: A generalizable explainable artificial intelligence approach. *PLOS Computational Biology* **17**, e1008857 (2021).
 289. Kotsiantis, S. B. Decision trees: a recent overview. *Artif Intell Rev* **39**, 261–283 (2013).
 290. Song, Y. & Lu, Y. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **27**, 130 (2015).
 291. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
 292. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. (Springer Science & Business Media, 2009).
 293. Gardner, M. W. & Dorling, S. R. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* **32**, 2627–2636 (1998).
 294. Ramchoun, H., Ghanou, Y., Ettaouil, M. & Idrissi, M. A. J. Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence* **4**, (2016).

295. Chen, J., Si, Y.-W., Un, C.-W. & Siu, S. W. I. Chemical toxicity prediction based on semi-supervised learning and graph convolutional neural network. *Journal of Cheminformatics* **13**, 93 (2021).
296. Jiang, D. *et al.* Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. *Journal of Cheminformatics* **13**, 12 (2021).
297. Kocak, B., Kus, E. A. & Kilickesmez, O. How to read and review papers on machine learning and artificial intelligence in radiology: a survival guide to key methodological concepts. *Eur Radiol* **31**, 1819–1830 (2021).
298. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods* **44**, 235–249 (2000).
299. Sakai, M. *et al.* Prediction of pharmacological activities from chemical structures with graph convolutional neural networks. *Sci Rep* **11**, 525 (2021).
300. VanderPlas, J. *Python Data Science Handbook: Essential Tools for Working with Data*. (O'Reilly Media, Inc., 2016).
301. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R*. (Springer New York, 2014).
302. Zheng, A. *Evaluating Machine Learning Models*. (O'Reilly Media, Inc., 2015).
303. Geman, S., Bienenstock, E. & Doursat, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation* **4**, 1–58 (1992).
304. Doroudi, S. The Bias-Variance Tradeoff: How Data Science Can Inform Educational Debates. *AERA Open* **6**, 2332858420977208 (2020).
305. Mehta, P. *et al.* A high-bias, low-variance introduction to Machine Learning for physicists. *Physics reports* **810**, 1 (2019).
306. Mervin, L. H. *et al.* Target prediction utilising negative bioactivity data covering large chemical space. *J Cheminform* **7**, 51 (2015).
307. Lister, P. D., Wolter, D. J. & Hanson, N. D. Antibacterial-Resistant *Pseudomonas aeruginosa*: Clinical Impact and Complex Regulation of Chromosomally Encoded Resistance Mechanisms. *Clinical Microbiology Reviews* **22**, 582–610 (2009).
308. Slama, T. G. Gram-negative antibiotic resistance: there is a price to pay. *Crit Care* **12**, S4 (2008).
309. Lambert, P. A. Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. *J R Soc Med* **95**, 22–26 (2002).
310. Zgurskaya, H. I., López, C. A. & Gnanakaran, S. Permeability Barrier of Gram-Negative Cell Envelopes and Approaches To Bypass It. *ACS Infect. Dis.* **1**, 512–522 (2015).
311. Sun, Y., Sun, T.-L. & Huang, H. W. Physical Properties of *Escherichia coli* Spheroplast Membranes. *Biophysical Journal* **107**, 2082 (2014).
312. Thomas, J., Navre, M., Rubio, A. & Coukell, A. Shared Platform for Antibiotic Research and Knowledge: A Collaborative Tool to SPARK Antibiotic Discovery. *ACS Infect. Dis.* **4**, 1536–1539 (2018).
313. Gaudet, T. *et al.* Utilizing graph machine learning within drug discovery and development. *Briefings in Bioinformatics* **22**, bbab159 (2021).
314. Riniker, S. & Landrum, G. A. Similarity maps - a visualization strategy for molecular fingerprints and machine-learning methods. *Journal of Cheminformatics* **5**, 43 (2013).

315. Polishchuk, P. *et al.* Structural and Physico-Chemical Interpretation (SPCI) of QSAR Models and Its Comparison with Matched Molecular Pair Analysis. *J. Chem. Inf. Model.* **56**, 1455–1469 (2016).
316. Matveieva, M. & Polishchuk, P. Benchmarks for interpretation of QSAR models. *Journal of Cheminformatics* **13**, 41 (2021).
317. Matveieva, M., Cronin, M. & Polishchuk, P. Interpretation of QSAR Models: Mining Structural Patterns Taking into Account Molecular Context. *Molecular Informatics* **38**, (2018).
318. Phetsang, W. *et al.* Fluorescent Trimethoprim Conjugate Probes To Assess Drug Accumulation in Wild Type and Mutant Escherichia coli. *ACS Infect Dis* **2**, 688–701 (2016).
319. S. Karlov, D., Sosnin, S., V. Tetko, I. & V. Fedorov, M. Chemical space exploration guided by deep neural networks. *RSC Advances* **9**, 5151–5157 (2019).
320. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell* **2**, 573–584 (2020).
321. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning*. (MIT Press, 2016).
322. Tse, E. G. *et al.* An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials. *J. Med. Chem.* **64**, 16450–16463 (2021).

9. Appendix

9.1 Compound 8 NMR

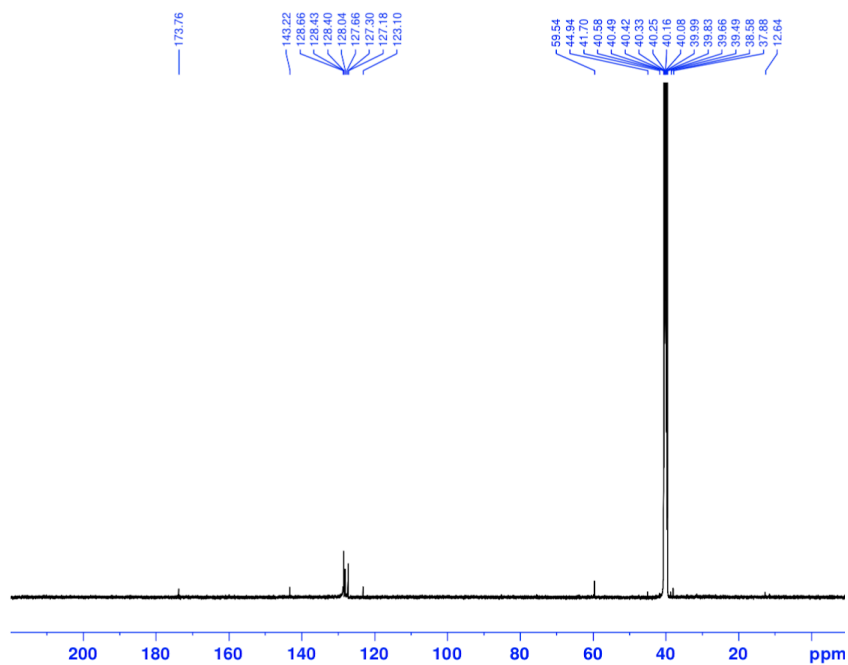


Figure 9.1 NMR analysis of Compound 8. 1.2 mg of Compound 8 in deuterated DMSO was analysed by C-13 NMR (AVC500) by R. Reinbold (University of Oxford)

9.2 Compound 8 MS

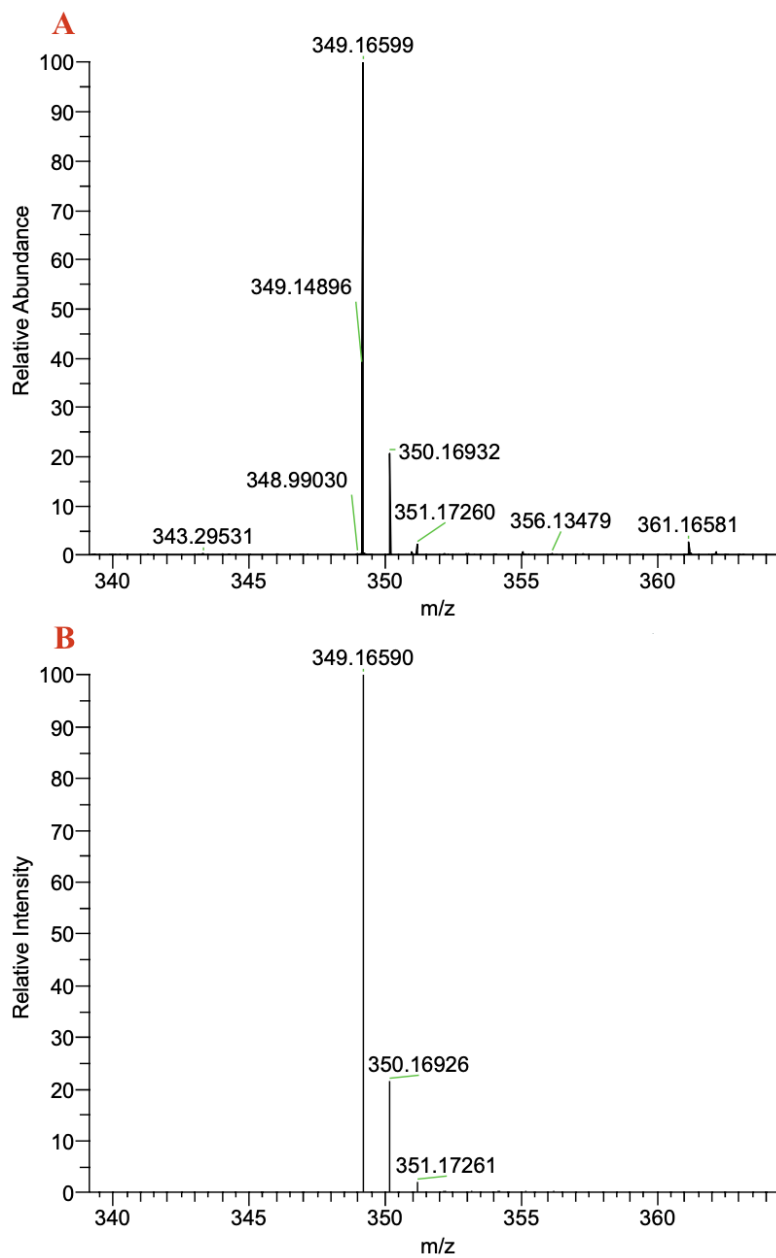
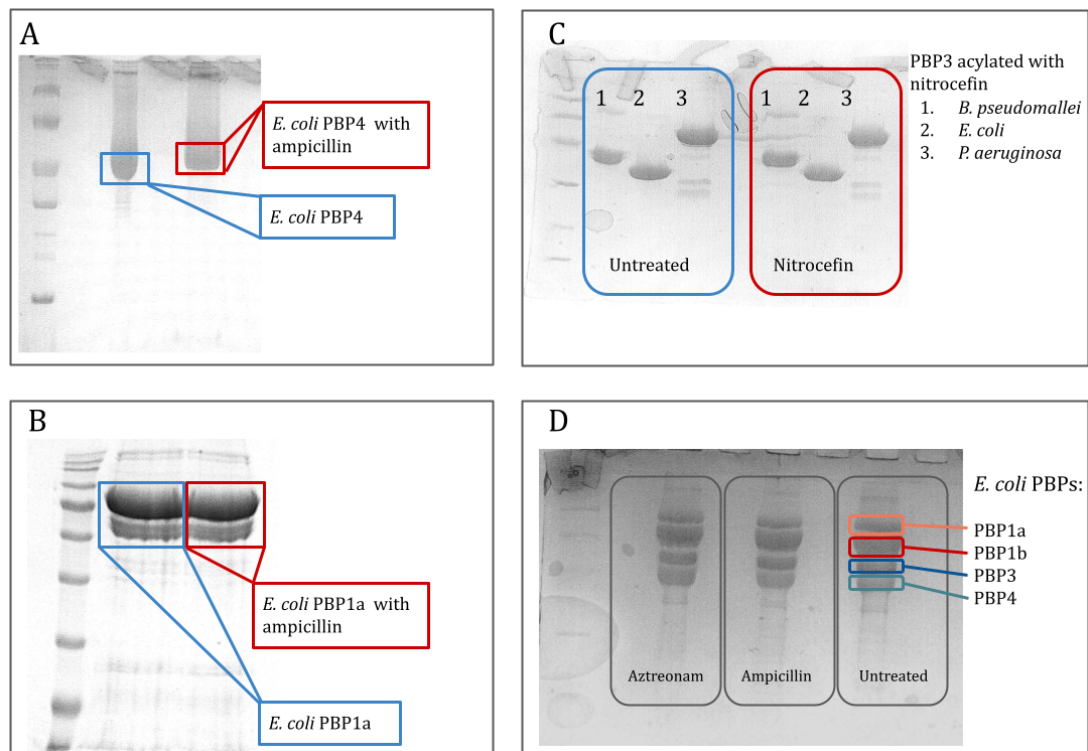


Figure 9.2 ESI+ MS spectra of Compound 8 - theoretical and measured. MS spectra from collaborators R. Reinbold. In spectra (A) is the measure MS of Compound 8, in (B) is the theoretical MS of Compound 8.

9.3 SDS-PAGE gels used in in-gel digest proteomic preparations



9.4 Sequenced *E. coli* PBP3

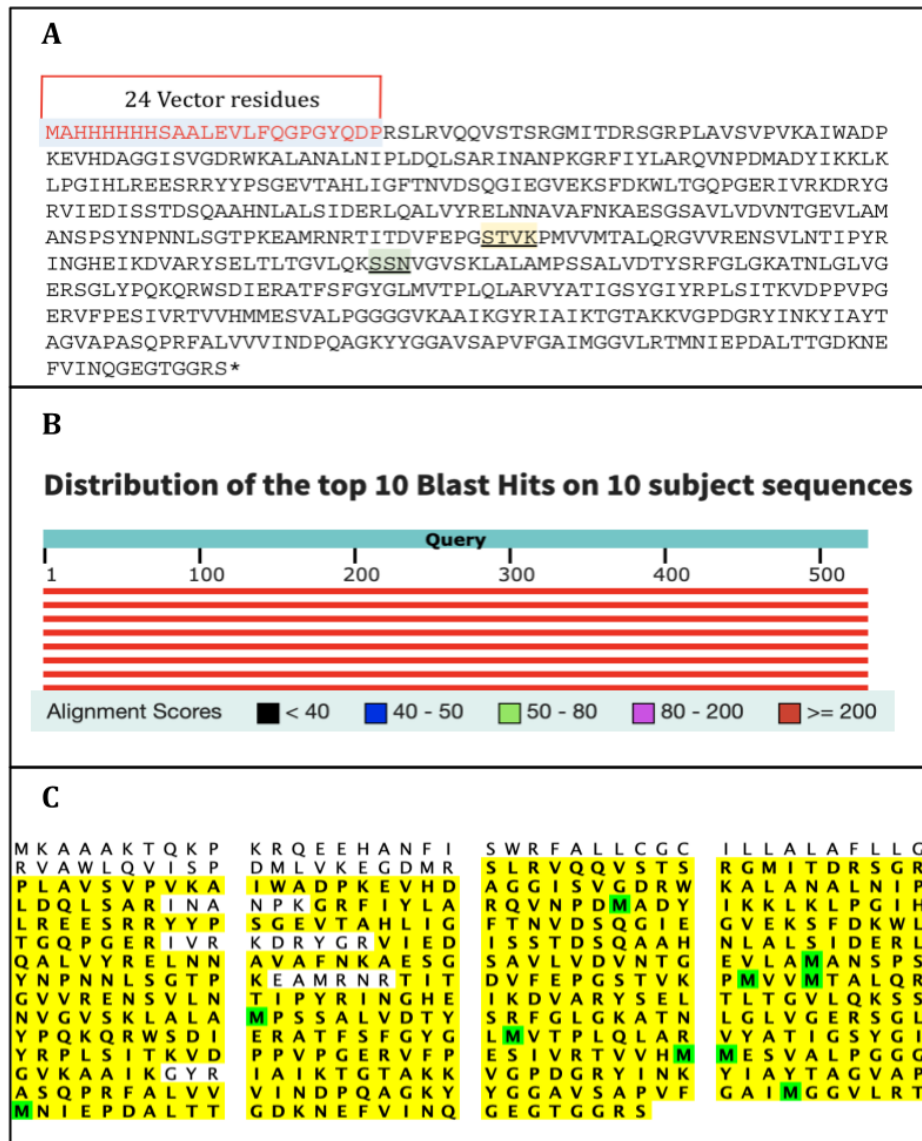


Figure 9.4A-C Sequencing alignment of *E. coli* PBP3. (A) The translated protein sequence of the PBP3 pet47b vector used in this work. The 24 vector residues including the His-tag are highlighted in red lettering. The first active-site serine in PBP3 is STVK which is underlined and highlighted in yellow. The second active-site serine is SSN which is also underlined and highlighted in blue. (B) The blast alignment of the translated sequencing results. The sequence demonstrated 100% alignment to *E. coli* K12 FTSI (PBP3). (C.) Proteomic sequencing of *E. coli* PBP3, highlighted in yellow, is the peptide coverage. The construct used in this work was missing the transmembrane domain - residues 60-588.

9.5 MIC results

Drug	BL21	Δ OmpF	Δ OmpC	Δ OmpA	BL21+ Polymyxin	Δ ACFB	<i>S. aureus</i>	<i>P. aeruginosa</i>	Δ ACF
Flucloxacillin	128	128	128	128	64	128	0.125	128	32
Vancomycin	128	128	128	128	2	128	2	128	128
Metronidazole	128	128	128	128	128	128	16	128	128
Fusidic acid	128	128	128	128	2	128	0.25	128	64
Linezolid	64	128	128	128	0.5	64	2	128	128
d-cycloserine	128	128	64	64	32	128	128	128	64
Novobiocin	128	128	128	128	4	64	0.03125	128	128
6-APA	128	128	128	128	128	128	128	128	128
Tedizolid	128	128	128	128	0.5	64	2	128	128
Methicillin	128	128	128	128	128	128	1	128	128
Penicillin G	8	64	32	32	2	32	1	128	32
Clarithromycin	128	128	128	128	0.5	128	0.125	128	128
Cefsulodin	32	64	64	64	0.25	128	4	8	128
Quinidine	16	32	64	32	8	32	8	128	128

Subbactam	128	128	128	128	128	8	64	16	128	128
Pentamidine	128	128	128	128	128	8	128	16	128	128
β -lactamase	128	128	128	128	128	128	128	128	128	128
Azlocillin	8	16	8	16	16	2	32	4	16	16
Spectinomycin	16	16	16	16	16	16	32	8	128	32
Cefaclor	4	16	8	8	8	1	64	1	128	64
Ampicillin	2	8	8	8	8	2	8	0.5	128	16
Carbenicillin	4	8	8	8	8	0.5	8	8	128	64
Amoxicillin	4	16	16	8	8	0.5	32	0.5	128	32
Chloramphenicol	2	8	16	4	4	0.5	2	4	32	2
Nitrofurantoin	1	8	8	8	8	0.25	8	32	128	1
Cefoxitin	2	32	16	8	8	2	64	4	128	128
Amikacin	2	8	2	8	8	2	8	4	16	4
Phosphomycin	16	32	16	64	64	8	32	0.5	64	16
Rifampicin	16	16	16	16	16	0.5	4	0.0625	64	8
Cephalexin	0.5	16	8	8	8	1	8	2	64	8
Minocycline	0.25	2	2	2	2	0.0625	2	0.0625	32	4
Polymyxin B	0.5	0.25	0.25	0.125	0.125	32	1	32	4	0.5

Piperacillin	1	4	4	4	4	0.25	8	4	16	8
Doxycycline	1	2	2	2	0.5	0.125	8	0.125	32	32
Neomycin	1	4	4	2	0.125	1	2	1	32	4
Colistin	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	1	32	8	0.25
Gentamicin	0.5	1	0.5	2	1	0.5	4	0.5	8	4
Cephalothin	8	32	32	8	1	0.5	32	0.5	32	64
Oxolinic acid	1	2	2	2	0.25	0.25	2	8	64	0.125
Aztreonam	0.015625	0.25	0.25	0.125	0.015625	0.015625	0.5	0	8	1
Levofloxacin	0.0078125	0.0625	0.03125	0.0625	0.0078125	0.0078125	0.015625	0.25	1	0.015625
Cefepime	0.015625	0.125	0.0625	0.0625	0.03125	0.03125	0.25	2	4	1
Cefoperazone	0.0078125	0.5	0.125	0.0625	0.0078125	0.0078125	0.25	4	8	2
Trimethoprim	0.5	2	2	2	0.03125	0.03125	4	64	64	0.25
Cefotetan	0.125	2	1	0.25	0.0625	0.0625	8	16	128	16
Fleroxacin	0.015625	0.25	0.125	0.125	0.015625	0.015625	0.125	0.5	4	0.0625
Ciprofloxacin	0.0078125	0.03125	0.03125	0.03125	0.0078125	0.0078125	0.0078125	0.25	0.25	0.0078125
Triclosan	0.0625	0.25	0.5	0.5	0.0078125	0.0078125	0.126	0.5	64	0.25
Sparfloxacin	0.0078125	0.015625	0.03125	0.03125	0.0078125	0.0078125	0.0078125	0.125	2	0.0078125
Ceftazidime	0.03125	0.25	0.25	0.25	0.0078125	0.0078125	0.0625	16	4	0.5

Cefotaxime	0.015625	0.125	0.125	0.125	0.125	0.0078125	0.5	1	32	0.125
Ceftriaxone	0.03125	0.25	0.125	0.125	0.125	0.0078125	0.5	2	32	0.5
Cefixime	0.25	1	0.5	0.5	0.5	0.015625	1	16	128	1
Azithromycin	128	128	128	128	128	0.5	128	0.5	128	128
Clindamycin	128	128	128	128	128	0.5	128	0.0625	128	128
Erythromycin	128	128	128	128	128	0.5	128	0.5	128	128
Dalbavancin	128	128	128	128	128	0.125	128	0.03125	128	128
Daptomycin	128	128	128	128	128	0.125	128	0.25	128	128
Dirithromycin	128	128	128	128	128	2	32	4	128	32
Nafcillin	128	128	128	128	128	0.5	128	0.25	128	128
Oritavancin	128	128	128	128	128	0.5	128	0.125	128	128
Teicoplanin	128	128	128	128	128	2	128	0.25	128	128
Telavancin	128	128	128	128	128	2	128	0.125	128	128
Telithromycin	128	128	128	128	128	2	128	0.125	128	128

Figure 9.5 The MIC results used to create the dataset for the MLP model. The MICs for this experiment were done in triplicate. The starting concentration was 128 µg/mL and the lowest concentration was 0.0078125 µg/mL

9.6 Review: Quantifying Antibiotic Permeability



Breaching the Barrier: Quantifying Antibiotic Permeability across Gram-negative Bacterial Membranes

Jehangir Cama^{1,4}, Abby Mae Henney² and Mathias Winterhalter³

¹ - Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom

² - Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom

³ - Department of Life Sciences and Chemistry, Jacobs University Bremen, Campus Ring 1, D-28759 Bremen, Germany

⁴ - Living Systems Institute, University of Exeter, Stocker Road, Exeter EX4 4QD, United Kingdom

Correspondence to Jehangir Cama and Mathias Winterhalter: j.cama@exeter.ac.uk, m.winterhalter@jacobs-university.de
<https://doi.org/10.1016/j.jmb.2019.03.031>

Edited by C.G. Dowson

Abstract

The double-membrane cell envelope of Gram-negative bacteria is a sophisticated barrier that facilitates the uptake of nutrients and protects the organism from toxic compounds. An antibiotic molecule must find its way through the negatively charged lipopolysaccharide layer on the outer surface, pass through either a porin or the hydrophobic layer of the outer membrane, then traverse the hydrophilic peptidoglycan layer only to find another hydrophobic lipid bilayer before it finally enters the cytoplasm, where it typically finds its target. This complex uptake pathway with very different physico-chemical properties is one reason that Gram-negative are intrinsically protected against multiple classes of antibiotic-like molecules, and is likely the main reason that *in vitro* target-based screening programs have failed to deliver novel antibiotics for these organisms. Due to the lack of general methods available for quantifying the flux of drugs into the cell, little is known about permeation rates, transport pathways and accumulation at the target sites for particular molecules. Here we summarize the current tools available for measuring antibiotic uptake across the different compartments of Gram-negative bacteria.

© 2019 Elsevier Ltd. All rights reserved.

Antimicrobial failure and resistance are now widely recognized across the scientific community as a major threat to the modern medical system [1,2]. A combination of significant drug discovery challenges [3], poor economic incentivisation [1,4], and frankly disastrous misuse in both medicine [5] and particularly in agriculture [6] has led us to the present scenario in which multi-, extreme-, and now pan-drug-resistant pathogens are leading to patient hospitalizations and mortalities across the globe [7,8]. There is a dire need to refresh the antibiotic pipeline with new drugs, with new modalities of action, in order to address the current lack of treatment options for these infections.

To tackle the worldwide problem of antibiotic resistance, the European Union currently supports a platform called New Drugs for Bad Bugs (ND4BB, www.ND4BB.eu), which is a broad series of public-private partnerships under the umbrella of the Innovative Medicines Initiative (see IMI, www.imi.europa.eu). A particular bottleneck in the development of new

antibiotics is their poor permeability in Gram-negative bacteria, which therefore require high drug doses which in turn lead to toxic side effects [1,9–11]. The IMI sub-project *Translocation* in particular has focused on the permeability challenge, with a report on their results due to be published soon [9]. Along the same lines, the US-based PEW foundation also recognizes low permeability as a major bottleneck, and has a stated goal to “understand and overcome barriers to drug penetration and efflux avoidance for Gram-negative bacteria” (The PEW Charitable Trusts: The PEW Roadmap [12]).

The low drug permeability of Gram-negative arises due to their complex double-membrane cellular envelope [10,13–15], depicted schematically in Fig. 1 [16]. To reach their (typically) cytoplasmic targets, small-molecule antibiotics need to first traverse the outer membrane, which is an asymmetric bilayer containing a lipopolysaccharide polymer network on its outer surface. This polymer mesh is a formidable

permeability barrier to large molecules irrespective of whether they are hydrophobic or hydrophilic. On the other hand, smaller solutes like nutrients or antibiotics depend on their ability to translocate through transmembrane β -barrel proteins, known as *porins* [13,17,18], to penetrate the hydrophobic outer membrane. It is also well established that the water-filled porins are typically selective for *hydrophilic* molecules. If the target is cytoplasmic, the therapeutic molecule next has to traverse the cytoplasmic membrane, a phospholipid bilayer that favors the transport of small *hydrophobic* molecules via solubility-diffusion mechanisms [19,20]. It is this dual-entry barrier that underpins the ability of Gram-negatives to survive antibiotics and drug-like molecules that might otherwise be active against Gram-positive bacteria. In addition, Gram-negatives also have in-built mechanisms for expelling toxic compounds from their interiors—these so-called *efflux pumps* actively recognize and expel antibiotics and other toxins from the cell [21–24]. The net result of this is that pharmaceutical *in vitro* target-based screening programs routinely struggled to deliver an agent active against Gram-negatives, simply because the candidates failed to accumulate in the vicinity of their targets at lethal concentrations. There is thus a very limited subset of molecules capable of breaching Gram-negative permeability barriers and, with the

inexorable spreading of resistant strains, we are running out of treatment options at an alarming rate.

Despite the obvious importance of this problem and decades of work, the permeability issues associated with Gram-negative cell envelopes have remained challenging to overcome; we still lack a fundamental understanding of the rules governing molecular transport across these membranes. Early work using radioactive tracer molecules revealed selective uptake across the so-called “outer cell wall” in Gram-negative bacteria and pointed toward the potential role of porins in mediating transport [25,26]. To reduce the number of parameters, porins were isolated and reconstituted into artificial planar lipid membranes [27,28]. Conductance measurements suggested single-channel pore sizes of about a nm, a value surprisingly close to that revealed by high-resolution x-ray structures a few years later [29]. In a different approach, outer-membrane porins were reconstituted into multilamellar liposomes and kinetic information about transport through the porins was obtained from the osmotic swelling induced by successful solute penetration [30]. These so-called liposome swelling assays became an important tool for understanding the uptake of nutrients or antibiotics through porins. Initial experiments with this

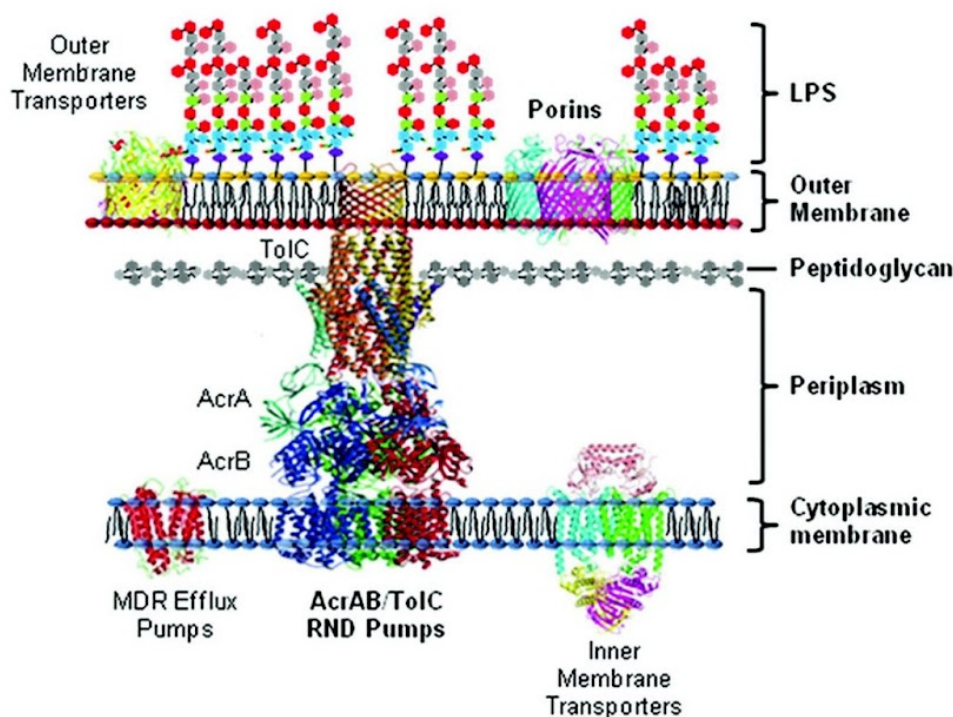


Fig. 1. Schematic of the Gram-negative cell envelope. Adapted with permission from Manchester *et al.* [16]. Copyright (2012) American Chemical Society.

approach, which involved measuring the transport of variously sized sugars, established the size exclusion cutoff of the OmpF porin to be about 550 Da [31,32]. Ions, amino acids, and small sugars may use general diffusion porins to enter the periplasmic space, whereas larger sugars and other molecules need to use dedicated pathways for outer-membrane transport [18]. These early studies established the molecular sieving properties of porins and provided an explanation for the high diffusion rates of these compounds through the outer membrane [33]. Comparison of the diffusion rates of solutes of various sizes gave remarkably reliable values of the channel size with respect to the corresponding crystallographic structure. However, swelling occurs in response to the movement of all the solutes, including components of the buffer, and extreme care is needed when this method is used to study the diffusion of charged solutes [31]. A comparison of some previously available techniques for studying antibiotic permeability was also given by Mortimer and Piddock [34], including the use of the natural autofluorescence of quinolones to study their accumulation in a range of bacterial species.

A number of experimental approaches have thus been utilized to quantify antibiotic permeability in Gram-negative, spanning both precise *in vitro* techniques for studying individual porins or efflux pumps, as well as whole-cell screens. However, what is still missing is a quantitative technique capable of precisely determining drug accumulation in the sub-cellular compartments where their targets are actually located. This is a crucial challenge that the field must overcome; controlled studies on individual porins and efflux pumps provide invaluable information about the precise routes taken by drugs to enter/exit cells, but the extraordinary variety of pathways and proteins in different species make it experimentally unfeasible to study each and every pathway of interest in reconstituted systems. Furthermore, there is mounting evidence to show that the expression of these proteins is highly dependent on the nutrient microenvironment around the cells [35,36]. Inherent variability in gene expression, especially in efflux pump activity, within an isogenic colony leads to variation in drug uptake and hence in drug tolerance [24,37]. These are all naturally occurring complications that confound attempts at understanding drug uptake at the (sub)cellular level. It is absolutely critical to develop techniques where cells can be studied in their native state, in well-controlled microenvironments, rather than after repeated washing and resuspension steps in contrived nutrient conditions.

A number of recent reviews have examined the topic of antibiotic accumulation in Gram-negatives, focusing either on the latest technical developments for quantifying accumulation in cells [38], summarizing what we know about the structure of the cell envelope and the challenges associated therein [10], or

specifically focusing on efflux systems [39] and porins [40]. However, we believe that an understanding of this complex problem of antibiotic permeation into Gram-negatives will involve insights from both bottom-up biomimetic approaches as well as top-down whole-cell measurements, and therefore review here the latest techniques in both general areas. We will discuss recent whole-cell assays based on mass spectrometry (MS) and single-cell microscopy. However, we will also give an overview of the latest developments in electrophysiology measurements of individual transport channels, as well as recent liposome-based transport assays; these techniques combine biomimetic model membranes with microfluidic platforms to study antibiotic transport across lipid bilayers and porins with precise control over the drug mixing and exposure times. We will identify the advantages and drawbacks of current approaches, and suggest how we might combine expertise from different scientific backgrounds to overcome the present technical challenges in the field. Finally, we have also included a brief discussion at the end on the economic challenges facing antibiotic development, which we feel need to be discussed in the scientific, industrial, and regulatory communities in order to ensure that our academic investigations do indeed translate into real-world, clinical impact.

Whole-Cell Accumulation Assays with MS

There have been a number of techniques proposed for measuring antibiotic accumulation via MS. Some directly measure the intracellular concentration of the compound, whereas others measure the change in the extracellular concentration and use this to determine compound accumulation [41,42]. Figure 2 outlines the general workflow of both approaches. Methods measuring the intracellular concentration of the drug are more labor-intensive and can involve multiple wash stages such as a silicone oil wash [23,41,43], centrifugation washes [44], and filtration processes [38,45] to remove the extracellular compound. These wash stages potentiate the risk of cell lysis and, as such, loss of compound leading to inaccurate results [38]. The silicone oil technique involves layering bacteria that have been incubated with the drug over the oil and pelleting it. The silicone oil is then pipetted off along with the supernatant. The largest intracellular investigations of accumulation utilized silicone oil to limit extracellular binding [41,46]. This is because, despite the increased manual workload, this technique leads to less lysis and a cleaner spectrum due to reduced non-specific binding [46]. The concentration of compounds used in the assays is important; limits of detection restrict the use of highly potent antibacterials which, due to the loss of cellular integrity, leak out of the cells and thus invalidate the assay [23,38].

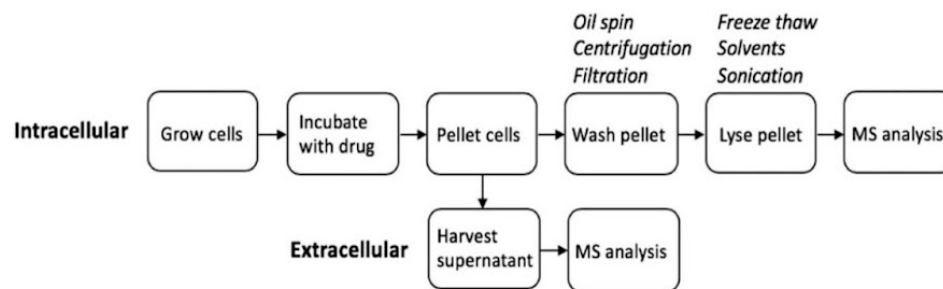


Fig. 2. General workflow of MS assays. An outline of the general workflow utilized in MS based intracellular and extracellular drug concentration measurements.

Richter *et al.* [41] have progressed intracellular assays to screen up to 180 diverse compounds. They sought to discover and define guidelines for compound permeability. It was revealed that in *Escherichia coli*, molecules which are rigid and flat and possess a primary amine group preferentially permeate [41]. This study investigated intracellular concentrations through LC–MS/MS in *E. coli* MG1655 [41]. Richter *et al.* tested more compounds than have been previously tested, but also examined a broader collection of compounds [41]. Another accumulation study by Davis *et al.* [47], which involved an LC–MS/MS assay on 10 compounds in *E. coli*, *Bacillus subtilis*, and *Mycobacterium smegmatis*, struggled to find distinct rules governing permeability. This may be due to the differences in the cell wall and the cell membranes of these organisms, or even an environmental parameter, such as growth temperature. In addition, the limited diversity of compounds tested also restricted their ability to draw conclusions on permeability.

The potential risk of cell lysis due to various assay preparation steps suggests that deducing the change in the extracellular concentration of the drug due to cellular accumulation might be more useful for high-throughput screening (HTS) [42]. As outlined in Fig. 2, the bacteria that have been incubated with the drug are pelleted, following which the supernatant is harvested and analyzed [42]. This method requires a minimally permeating cold control to deduce accumulation [38,42]. However, cold controls must also be treated with caution—for instance, Stokes *et al.* [48] showed that cold stress makes *E. coli* susceptible to glycopeptide antibiotics via alterations in the outer membrane. This demonstrates the challenges of performing appropriate controls for accumulation assays in bacteria.

Data generated from these accumulation assays can indicate which compounds accumulate in greater concentrations, and as a result are more permeable. These data, however, do not resolve accumulation in the various subcellular compartments. There has been some progress by Spangler

et al. [49] involving the use of click-chemistry and MS to better understand subcellular accumulation. This method has unveiled some useful information about the accumulation of a range of drugs which contain an azide group [49]. However, it is not applicable to a large library screening of compounds, as it would require the conjugation of a probe to each compound [49]. An LC–MS/MS study on 132 ligA compounds by Iyer *et al.* [46] looked at accumulation and on-target activity and reiterated the fact that cytoplasmic accumulation is difficult. Another method used to determine the subcellular localization of drugs is 3D imaging time-of-flight secondary ion MS by Tian *et al.* [50]. Here a cluster ion beam was used to resolve the subcellular localization of the antibiotics. The C_{60}^+ method of ionization utilizes a label-free method to reveal the subcellular localization of ions with a spatial resolution of approximately 300 nm [50]. The equipment used in this work offers exceptional imaging and spectral analysis of compounds. In contrast to standard LC–MS, it is able to combine depth analysis and two-dimensional imaging, making a three-dimensional molecular rendering feasible [50]. This technique generates a great deal of data as every pixel in the image contains a spectrum. As a result, the data analysis is complex, as is the experimental setup itself, and it remains to be seen how easily the technique can be transferred into a commercial setting. However, as a label-free technique with sub-micron resolution, it represents an important technological advance that promises to be a valuable addition to the study of compound accumulation in bacterial compartments.

Prochnow *et al.* [51] have also recently reported a new workflow that involves subcellular fractionation via cold osmotic shock, ultrasonic treatment, and multiple centrifugation steps to quantify the subcellular accumulation of four different classes of antibiotics in WT and efflux deficient *E. coli* cells. Although the workflow involves multiple steps and hence may not be able to examine antibiotic uptake in real time in various environmental conditions, it is

nevertheless another important advance in the field which enables the measurement of a variety of drug classes in a label-free manner directly in the respective vicinities of the drug targets. As an example of the information made available, the authors were able to quantify differences in the concentrations of ciprofloxacin in the periplasm versus the cytoplasm of *E. coli*; correcting for the volume of the respective compartments revealed that the concentration of ciprofloxacin was 6-fold lower in the cytoplasm than in the periplasm. Such quantitative information will be crucial for the rational design of novel drug candidates in the next stages of antimicrobial drug discovery.

Despite its many advantages, it is important to note that MS generates a spectrum that reports the mass to charge ratio of the observed molecule and the intensity of ions at this peak, which is not directly quantitative. The determination of concentrations by ion intensities can be complicated by contaminants and detergents, and as such, direct quantification of mass spectra is challenging. The current techniques also involve a great deal of manual labor, which is incompatible with HTS. Preparing the samples for intracellular accumulation includes not only several cycles of centrifugation or silicone oil washing, but also manual lysis methods such as freeze–thaw cycles or sonication. To use MS for HTS, techniques such as acoustic MS have been proposed. This involves using an acoustic dispenser coupled to a mass spectrometer [52,53]. A sound wave hits the sample causing the formation of a droplet, which is fired into the ionization beam; this technique can analyze three samples a second [52,53]. Furthermore, the acoustic mass spectrometer is set up in a high throughput 384-well plate format, meaning that the amount of compound required is greatly reduced [52,53]. The acoustic coupling bypasses LC–MS loading, which increases the speed of analysis, but could mean that complex samples are not separated as efficiently [52].

A single LC–MS run of one compound generates a vast amount of information, but to appropriately deduce accumulation, a range of controls are necessary; these include measuring signals from the antibiotic with no bacteria, the bacteria with no antibiotic, a non-permeating control (low temperature, although subject to the caveat about cold controls mentioned above), and a high accumulating control (efflux deficient) [38]. The sheer amount of data generated by these methods suggests the need for machine learning techniques to analyze the data and predict permeability rules. The work by Richter *et al.* [41] has been amalgamated into a website (<http://www.entry-way.org/>) where the ability of a proposed molecule to traverse the outer membrane based on predictive rules can be estimated. Thus, MS data are now being utilized for “real-world” output, to show how accumulation data can be rationalized into drug design.

Single-cell Level Microscopy

A number of recent techniques, currently published or under development, involve studying antibiotic accumulation at the single-cell level. By combining single-cell level uptake information with recently developed mathematical models [54], there is potential for a breakthrough in the field that would enable the prediction of sub-cellular drug localization using model parameters extracted from the experimental data.

Deep UV autofluorescence microscopy and microfluorimetry has been successfully implemented by the Pagès group and collaborators to study the accumulation of antibiotics in single bacterial cells [55–57]. These techniques typically involve suspending bacterial cells between two quartz coverslips and investigating them with a tunable deep UV light source (a synchrotron) [58]. Bacteria are prepared by culturing them to exponential phase, pelleting them via centrifugation followed by resuspension in a NaPi-MgCl₂ buffer; the cells are then mixed with the compounds of interest just before microscopy [58]. Using this technique, the group was able to study the accumulation of fluoroquinolones label-free in *Enterobacteriaceae*, including studies on cells overexpressing the AcrAB efflux system [59]. However, the deep UV fluorescence measurements require a number of controls for separating cellular autofluorescence and crosstalk from the drug signals [58]. Furthermore, deep UV imaging requires quartz optics, including quartz objectives, which are not easily available. A number of controls for cell viability must also be performed, since deep UV excitation is detrimental for the cells.

The same group has also worked with fluorescent derivatives of ceftazidime to study intracellular accumulation at the single-cell level using microspectrofluorimetry and epifluorimetry [60]. The derivatives accumulated in the periplasm once the outer membrane was permeabilized, and periplasmic accumulation was found to correlate with antibiotic activity [60]. However, the authors also noted that modifications of the compounds with fluorophores changed the antibiotic activity of the drug in *E. coli* strains, which needs to be accounted for when interpreting permeation rates for fluorescent derivatives. The Cooper and Blaskovich groups (Queensland) have also pioneered the development of a range of fluorescent derivatives of antibiotics, which are being applied to single-cell level fluorescence assays, including using super-resolution structured illumination microscopy [61,62]. They have recently reviewed the latest developments in the field, showcasing the various ways that fluorescent antibiotic derivatives are being used to investigate mechanisms of resistance [63].

However, techniques based on single-cell fluorescence still, in general, lack the ability to resolve accumulation in subcellular compartments. As mentioned previously in relation to MS approaches,

centrifugation and washing steps are less than ideal, and yet without them, it is challenging from single-cell microscopy to distinguish drug molecules bound to the outer membrane from drug molecules that actually permeate into the cell. What is also missing is the development of accumulation assays in which the microenvironment and the drug exposure time are well controlled—this is where nanofluidic or microfluidic techniques could provide solutions. Indeed, a nanofluidics assay, combining bacterial trapping in microchannels with controlled drug exposure, was recently reported by the Lee group; they quantified the translocation of clindamycin through wild-type and an efflux deficient *E. coli* strain, although they used an LCMS assay to quantify the clindamycin flux per cell rather than a fluorescence readout [64]. However, although the calculation is performed “per cell”, the actual measurement accumulates perfusion data across a population of approximately 200 cells. Furthermore, the technique measures the flux of drug molecules across the cell rather than cellular accumulation itself, and it is unclear whether any subcellular concentration information can be inferred from this technique. The technique also relies on having an excellent seal between the cells and the trap channels, which may limit its applicability—different cell types and even isogenic cells in different metabolic states will have different widths and shapes, which would affect the quality of the seals and lead to leakage of the drug molecules between the loading and collection channels. Nevertheless, combining innovative microfabrication and single-cell trapping approaches [37] with microfluorimetry or fluorescence has the potential to enable measurements of drug accumulation in well-defined microenvironments. This is important when considering that different growth conditions lead to changes in porin expression [35]; these systems could be used to analyze the role of different media conditions in influencing drug accumulation kinetics. These assays are currently under development, and combining microfluidics with the latest microscopy techniques has immense potential for solving long-standing problems in the field.

Raman microscopy and microspectroscopy have also been used to study the uptake and interactions of β -lactamase inhibitors in groups of freeze-dried *E. coli* cells [65,66] in a label-free manner. The Carey group built upon their expertise in detecting different inhibitor–enzyme intermediates with Raman microscopy [67] to study the penetration of clavulanic acid and tazobactam into *E. coli* cells; their inhibitory effects on β -lactamases were also characterized [65]. Although Raman spectroscopy has in the past been technically challenging when used to study bacteria due to low signals and noisy backgrounds, the use of freeze-drying and subsequent difference spectroscopy has now enabled these semi-quantitative measurements in groups of cells [66].

Alternative techniques to measure uptake in cells involve using enzymatic activity inside the bacteria as a readout. For example, the influx of cephalosporins can be measured by coupling their uptake to their hydrolysis by periplasmic β -lactamases [68–70]. However, to connect whole-cell experiments with single-channel recordings, it is important to be able to quantify the amount of functional protein in the outer membrane. Recently the team of D. Bumann at Biozentrum Basel was able to quantify the number of porins in the outer membrane of *Acinetobacter baumannii* in infected mouse and rat lung tissue [71]. Combining new techniques for single-cell level uptake measurements with developments in proteomics would be extremely beneficial for the field, although at present there remain significant technical challenges that still need to be overcome before this can be made a reality.

Developments in Electrophysiology

Molecular details of antibiotic transport through specific porin pathways can be obtained from single-channel characterization experiments in “black lipid membranes,” which are described in detail in Gutschmann *et al.* [72]. Briefly, purified detergent-solubilized channel proteins or proteoliposomes are added to one side of the membrane and spontaneously insert into the bilayer over time. The sequential insertions of open channels in the membrane lead to discrete current jumps due to ion movement through the open channels. The conductance (i.e., the amount of current per unit voltage) of a channel can be obtained by measuring the size of these current jumps. For a number of applications, the reconstitution of proteins into giant unilamellar vesicles (GUVs) and subsequent patch-clamp measurements are advantageous [73]. For this technique, a commercial tool (Port-a-Patch, Nanion Technologies) may be used, which facilitates the experimental handling as the liposome containing preparation is simply pipetted on top of a hole in glass and aspirated to form a giga-seal [74,75]. After insertion, the channel activity can be studied under various conditions (e.g., in the presence of DNA, polymers, or small molecules). Investigations by the Benz and Rosenbusch groups in the 1970s and 1980s established some of the hallmark properties of the general diffusion porins, such as the relatively large pore size, low ionic selectivity, and high opening probability [75–77].

The main advantage of black lipid membrane measurements is their ease of access and the low consumption of protein. Although this is a single-molecule technique, the information originates from the modulation of the ion current with time; the extracted information is an average over the entire path along the channel. More molecular details about how the molecules move along the channel and their main point of interaction can be obtained in

combination with all atom molecular dynamics (MD). Indeed, with the numerous high-resolution structures upcoming forthwith, MD simulation is a perfect complementary technique to single-channel electrophysiology. For example, some years ago, the ion pathways in OmpF were investigated in more detail, and these studies revealed that the paths taken by anions and cations are divergent at the constriction region; cations are drawn close to the negative

charges of the L3 loop, and anions flow near the positively charged cluster of the opposite barrel wall (see also Fig. 3a) [78,81,82]. This type of work emphasizes the notion that the permeating ions interact with the wall of the channel and that ion movement does not follow simple diffusion. Moreover, measuring the temperature-dependent OmpF channel conductance and normalizing over the temperature-dependent bulk ion conductances

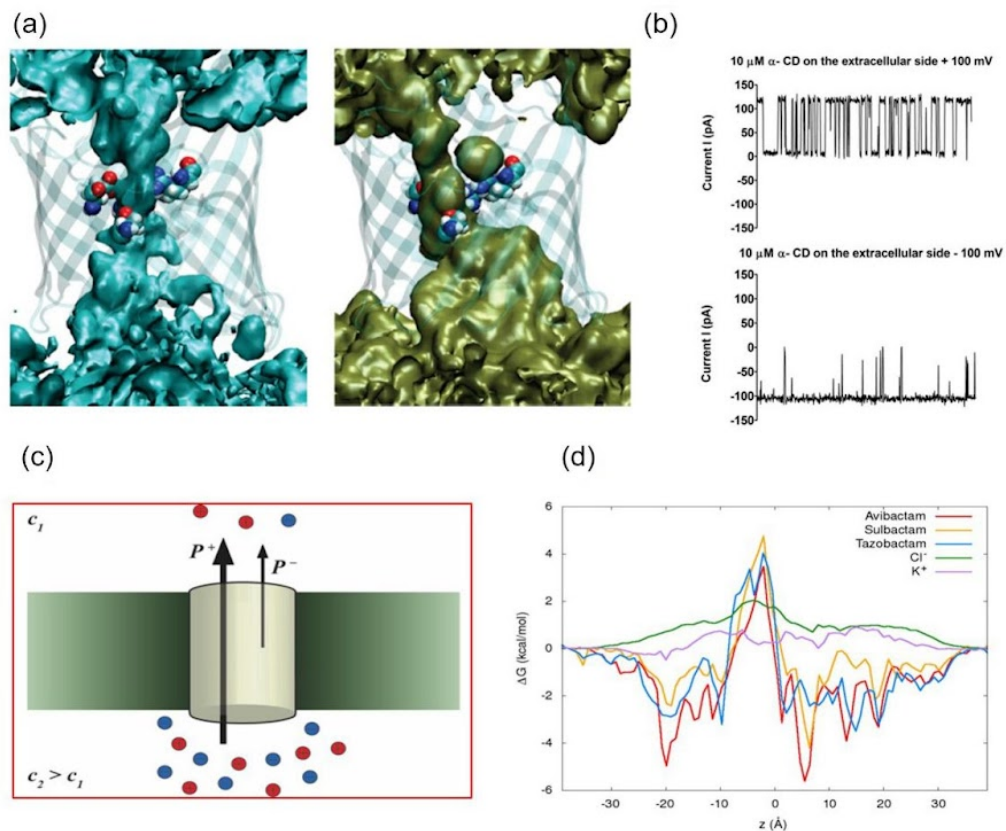


Fig. 3. (a). Iso-density surfaces of the Cl⁻ (left) and K⁺ ion densities (right) in an OmpF channel averaged over the full trajectory. The important residues in the constriction zone are shown as well. For details, please see Pezeshki *et al.* [78]. Reprinted from Pezeshki *et al.* [78]. Copyright (2009), with permission from Elsevier. (b). Effect of electroosmotic flow. Typical ion-current recordings at positive voltage +100 mV (top) and negative voltage -100 mV (bottom) in the presence of 10 μM α-CD on the extracellular side of CymA in 1 M KCl. Here, the electroosmotic flow drags the neutral α-CD into the channel and briefly blocks the ion current through the channel. Reversing the voltage reverses the flow direction and only a few α-CD molecules reach the channel, with substantially shorter residence times. Variation of the applied external transmembrane voltage will modulate the residence times and allows one to determine whether or not translocation was successful [79]. Reprinted from Bhamidimarri *et al.* [79]. Copyright (2016), with permission from Elsevier. (c). For charged molecules, a concentration gradient can be applied driving cations and anions through the channel. A difference in the electrophoretic mobility causes a diffusion potential and allows one to identify translocation, and to distinguish translocation events from binding. Combining this with single-channel recordings allows one to estimate the flux. (d). All atom modeling performed in the group of M. Ceccarelli shows the energy barrier for translocation. OmpF as a cation selective channel has a low and broad barrier for K⁺ and a higher one for Cl⁻. In contrast, beta-lactam inhibitors like avibactam, sulbactam or tazobactam have sharp energy barriers [80]. Reprinted in part with permission from Ghai *et al.* [80]. Copyright (2017) American Chemical Society.

revealed a clear deviation from the bulk. An Arrhenius plot allows for the quantification of these ion interactions [78,82]. MD modeling revealed two distinct ion pathways when the ion concentration is below 150 mM KCl, whereas above this concentration, ions fill almost the entire pore volume [78,82].

In a similar manner, substrate permeation is facilitated by an affinity to specific channels. The addition of malto-oligosaccharides to LamB (Maltoporin) from *E. coli* drives the sugar to enter the binding site and block the channel entirely for ions. The occupancy of the binding site with sugar was obtained by measuring the reduction in channel conductance [83]. The permeabilities obtained for malto-oligosaccharides using ion current noise agreed well with liposome swelling results, except in the case of sucrose [30,83]. Inspection of the structure revealed that sucrose has a binding site but is not able to permeate [84]. It is also possible to extend the noise analysis technique to antibiotic permeation. However, the analysis requires a sufficiently strong binding of the drug to the channel, and the antibiotic in the binding site must sufficiently block the ion current [85]. A further point to note is the inherent instability of antibiotics [86]. For example, in the case of ampicillin, a short exposure to basic pH to enhance the solubility can degrade large amounts of the compound.

Binding can be distinguished from transport by applying an external force on charged molecules. As such, an electric field can drag or repel charged permeating molecules, which is then translated into a modification of the residence time in the channel. Thus, measuring the average translocation time for different external electric fields should provide a clear picture [87]. In the case of uncharged molecules, the molecular movement could be modulated by electro-osmotic flow [79]. Most of the porin channels have an excess of one charge species in the constriction zone; thus, an externally applied electric field would cause a net unidirectional incoming flow; molecules diffusing with the flow would have shorter residence times (see Fig. 3b). In a series of investigations, the specific permeation of α -cyclodextrin across CymA of *Klebsiella oxytoca* was quantified; the individual steps have also been simulated [79]. A more recent suggestion for enhancing the signal for *translocation* is to engineer a barrier for exit within the pore [88]. By comparing the residence times of molecules in the porin in the presence and absence of the barrier on the periplasmic side of the channel, one can identify true permeation events and identify molecules that reach the exit. Starting from the well-studied porin OmpF, a single point mutation at position 181 OmpF^{E181C} was introduced and crosslinked with either sodium (2-sulfonatoethyl) methanethiosulfonate (MTSES) or glutathione. The modification of OmpF^{E181C} by MTSES builds a barrier which is

sufficient to block the pathway of norfloxacin through the porin. The modulation of the interaction dwell time allows one to determine the successful permeation of norfloxacin across wild-type OmpF. This approach might enable the discrimination of blockage events from translocation events for a wide range of substrates while working in the μ M range. As the data analysis is straightforward, parallelization of the experiments might be possible. A potential application of this technique could include screening for molecular structures to improve the permeability of antibiotics.

A different approach involves the use of an unbalanced charge accumulation [80,86]. Creating a concentration gradient between both sides of the channel induces a concentration driven flux (see Fig. 3c). In most cases, one of the ions diffuses slower than the counter ion; the difference in flux creates a *diffusion potential*. Measuring the diffusion potential as a function of the concentration gradient reveals relative fluxes. In combination with a single-channel conductance measurement, a more quantitative analysis is possible. Note that the cation's permeation strongly depends on that of the anion and vice versa, and thus the permeability is not a simple constant but varies with concentration and the counterions. In the case of charged antibiotic molecules, their channel permeability can be estimated by the application of a concentration gradient and the measurement of the potential built up by an unequal permeation of the cation *versus* the anion. This technique is similar to a selectivity measurement, and a numerical solution based on the Goldman–Hodgkin–Katz ion current equation can be used. The permeation of three β -lactamase inhibitors (avibactam, subactam and tazobactam) through OmpF and OmpC orthologues from four enterobacterial species was recently characterized using this approach. The information from electrophysiology can be combined with existing high-resolution structures and modeling to obtain the energy barrier. For example, in the case of OmpF, the energy barrier for small ions is low and broad, whereas for the beta-lactam inhibitors, the barrier is substantially higher and narrower (see Fig. 3d). Surprisingly, there is also a shallow affinity site just before the entry at the extracellular side [80]. Well-designed single-channel conductance measurements, in combination with MD simulations, thus continue to be a powerful tool for investigating transport through individual components of the bacterial cell envelope at the molecular level.

Biomimetic Liposome Assays

Liposomes or lipid vesicles are versatile tools for studying membrane processes [89]. We have already discussed how multilamellar (multiple lipid bilayers) liposomes were used in swelling assays to

study solute transport through porins. On the other hand, unilamellar vesicles containing a single lipid bilayer can be used to mimic the lipid bilayer of cell membranes, with precise control of the lipid content and composition of the membrane. Small unilamellar vesicles (SUVs) typically have diameters between 50 and 200 nm, large unilamellar vesicles (LUVs) range from 200 to 1000 nm in diameter, and GUVs typically have diameters ranging from 1 to 100 μm . SUVs and LUVs can be formed relatively easily and quickly using lipid extrusion techniques [90], and are then often subjected to bulk assays such as spectrofluorimetry where transport or binding processes are studied using changes in the fluorescence or absorbance of the bulk solution as a whole over time [91]. This is typically correlated with the leakage of a dye or fluorophore encapsulated within the vesicles or, alternatively, the transport of a solute across the SUV bilayers, which then reacts with an encapsulated target to change the fluorescence of the solution [92]. The latter principle was combined with TIRF microscopy and microfluidics by the Dittrich group to study tetracycline transport across SUV membranes, where the SUVs were immobilized on a glass substrate via an avidin–biotin bond [93]. The process involved encapsulating an Eu^{3+} salt inside the vesicles, which formed a fluorescent complex with tetracycline as it diffused across the SUV bilayers [93]. They used the technique to investigate the influence of membrane composition on permeation rates and showed that although drug lipophilicity as predicted by traditional octanol–water partition coefficients correlates well with membrane permeability, it is not the only significant parameter [93].

Similar results were observed with a related approach, developed in the Keyser laboratory at the University of Cambridge, where GUVs were used as the model membranes. The advantage of using GUVs is that membrane transport can be investigated at the single vesicle level using standard microscopy approaches [94]. However, traditional techniques using GUVs are limited in throughput—the GUVs are typically suspended in a droplet on a coverslip, and when fluorescent compounds are added, their uptake is measured by studying the change in fluorescence inside the GUVs. This is limited in throughput to the number of vesicles in the field of view of the microscope [95]. The other problem with such experiments is that the initial time of drug arrival and its concentration in the vicinity of the GUVs are both ill defined.

These technical problems were overcome using microfluidics. The GUVs were formed off chip using electroformation [96] and were then introduced into one of the inlets of a simple T-junction microfluidic device, whose channels were approximately 40 μm in width and height [97,98]. The drug of interest was added using the second inlet of the T-junction, and the flows were controlled by applying suction at the

outlet using a syringe pump [97]. The design enables the equal mixing of the vesicle fluid stream with the antibiotic at the T junction, and thus the time of vesicle exposure to the drug and the drug concentration surrounding the vesicle are both well defined [97]. The use of a flow system enabled measurements on hundreds of vesicles rather than the tens of GUVs normally studied in such assays [94,95]. Importantly, the technique utilized the autofluorescence of quinolone antibiotics in the ultraviolet ($\lambda_{\text{ex}} = 340 \text{ nm}$) to track their uptake in GUVs label-free; this simplifies the analysis since there are no secondary binding interactions that need to be accounted for when calculating the accumulation of the drug in the GUVs. Furthermore, unlike the industry standard PAMPA (parallel artificial membrane permeability assay) technique, there are no significant complications from unstirred water layers [98], which may affect the interpretation of compound permeability [99].

Using this technique, the group also confirmed that lipid composition plays a critical role in antibiotic permeability [100]. For example, at pH 7, norfloxacin permeability was found to be an order of magnitude higher for GUVs made of DOPC lipids as compared to DPhPC—these lipids have the same phosphatidylcholine (PC) head-groups but differ in the composition of their hydrocarbon tails, with the branched methyl groups in the DPhPC tail believed to be responsible for its lower permeability to small molecules [100]. The group investigated norfloxacin transport across GUV membranes formed from a range of different lipid mixtures, and further reiterated the need to quantify drug transport across the specific lipid bilayer composition of interest [100].

Such investigations will also be particularly important for comparing the permeation of drugs across Gram-negative membranes and the cytoplasmic membrane of Gram-negatives. There is some debate in the field about the relevance of the cytoplasmic membrane of Gram-negatives as a barrier, since a wide range of drugs are able to accumulate in the cytoplasm of, for instance, *E. coli* cells with compromised outer membranes [101]. However, as mentioned in the introduction, the difficulty with Gram-negative cytoplasmic entry arises from the apparently orthogonal sieving properties of the outer *versus* the cytoplasmic membrane, as has been extensively discussed previously [10,101]. It is also worth noting that Gram-negative membranes and the cytoplasmic membranes of Gram-negatives differ significantly in their lipid compositions [102]. This is likely to lead to quantifiable differences in drug transport rates across these membranes, which should be a future avenue for liposome-based investigations.

Vesicles are thus good models for studying passive transport across the cytoplasmic membrane of bacteria, where solubility and diffusion across the lipid bilayer are important. However, transport across

the outer membrane is generally governed by the porins. We have reviewed electrophysiology approaches to study antibiotic binding and transport through individual porins in a separate section. However, although these measurements have proved invaluable in understanding the molecular interactions between different drugs and channels, they are limited by the fact that actual fluxes are still challenging to measure.

To solve this conundrum, the authors (J.C. and M. W.) collaborated to reconstitute OmpF into GUV membranes, creating proteoliposomes which could then be studied in the microfluidic transport assay described above (Fig. 4) [103]. This provides a controlled, bottom-up approach to studying antibiotic transport across biomimetic bacterial membranes—norfloxacin transport was first quantified across pure GUV lipid membranes, and then the same experiments were performed with the OmpF

containing proteoliposomes. The experiments directly measured the enhancement in flux due to the porins in the membrane, paving the way for quantifying antibiotic permeability through transport proteins of interest using a direct optical readout [103].

A difficulty with GUV-based approaches has been that traditional vesicle formation protocols have a number of limitations. Electroformation, for example, involves applying alternating electric fields to bud lipid films off a conducting surface—this makes the use of physiological salt conditions challenging [104]. The populations also contain vesicles of varying diameters, and there are often difficulties associated with batch-to-batch variability, especially in the presence of salts. Furthermore, there are various reconstitution techniques for incorporating proteins in GUVs, but they each have certain drawbacks and there is no one standard technique

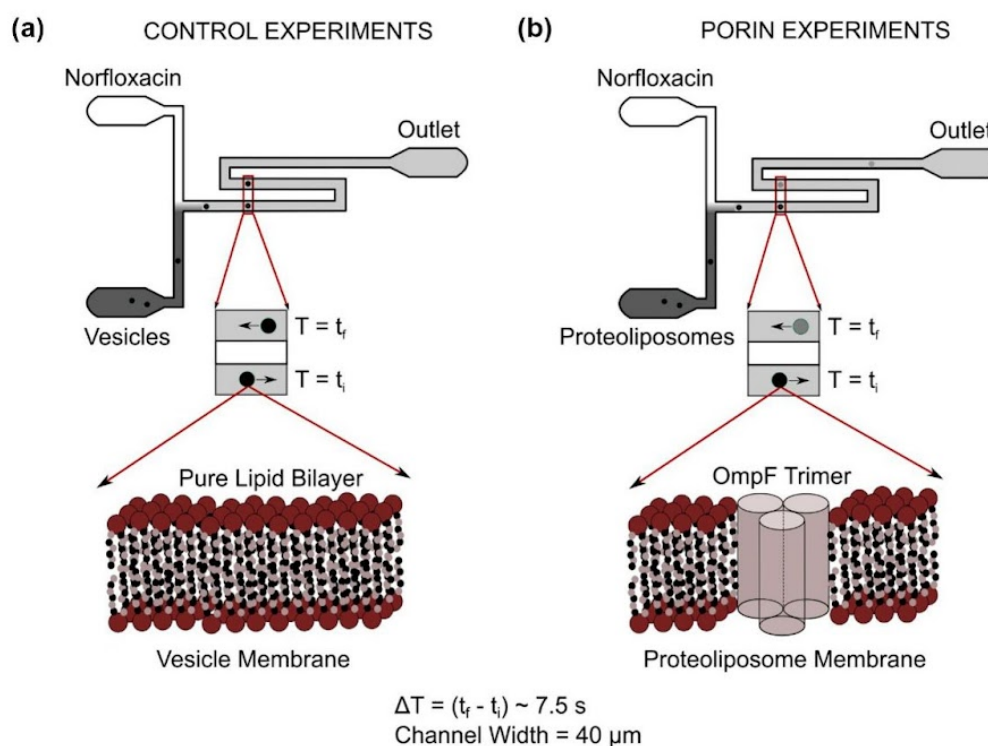


Fig. 4. Schematic of proteoliposome-based microfluidic assays for studying antibiotic transport across biomimetic membranes. Liposomes and OmpF embedded proteoliposome model membranes are mixed with an autofluorescent antibiotic (norfloxacin, $\lambda_{\text{ex}} = 340 \text{ nm}$) in a T-junction microfluidic device. As the control liposomes (a) or proteoliposomes (b) move through the bath of the drug, they accumulate the drug at different rates. Video fluorescence microscopy at multiple time points is used to determine the permeability coefficient of the drug crossing the model membranes. The difference in the permeabilities of the liposomes and proteoliposomes to the drug is used to quantify antibiotic flux through the porins in the proteoliposome membranes. The number of porins per proteoliposome is quantified separately using electrophysiology. Figure reprinted with permission from Cama *et al.* [103]. Copyright (2015) American Chemical Society.

that works better than the rest [105]. In order to overcome these challenges, a number of groups have now developed microfluidic GUV formation protocols [106]. One promising technique is octanol-assisted liposome assembly (OLA), developed by the Dekker group in Delft [107]. This involves a microfluidic jetting technique where lipids are dissolved in octanol as the organic phase—post GUV formation, the octanol forms a pocket which then spontaneously buds off as a droplet. The octanol can be separated [108] and then the pure GUVs analyzed downstream of formation. We have now successfully integrated the OLA platform with an antibiotic transport assay downstream of GUV formation, thus providing a complete lab-on-chip platform to create and test vesicles in physiological salt conditions [109].

These new microfluidic liposome formation platforms will also enable the investigation of an important question with regard to the transport of charged molecules across cytoplasmic membranes. As mentioned previously, the challenge with Gram-negative cytoplasmic uptake is that charged, hydrophilic molecules preferentially permeate the outer membrane, but then face a hydrophobic cytoplasmic lipid membrane across which their transport rates are much slower than that of neutral uncharged species; this has been quantified directly for drugs in different charge states [98]. It has been proposed that the transport of charged molecules across the cytoplasmic membrane may be mediated by the proton motive force and electrochemical gradients, leading to energy-dependent but carrier-independent mechanisms of transport [10]. In the past, membrane vesicles from species such as *Lactococcus lactis* and *Leuconostoc mesenteroides* have been coupled with membrane potentials to examine the transport of amino acids, citrate and malate [110,111], but detailed investigations on antibiotic uptake in such conditions have not yet been realized. Using the OLA formation platform, we have recently demonstrated that it is possible to trap thousands of vesicles encapsulating a pH sensitive dye in physical traps [112]; the vesicles are then subjected to antibiotic treatment over a time period of hours. This platform could in the future be utilized to study the role of electrochemical and pH gradients on drug transport by utilizing the high encapsulation efficiency of the OLA platform. One could potentially set up precise electrochemical and pH gradients between the interior and the exterior of these synthetic vesicles and then flush charged antibiotics into the system, monitoring the uptake of the drugs via a fluorescence-based readout similar to that described above. Such investigations could be used to quantify the role of electrochemical potentials on the uptake of charged antibiotics across model cytoplasmic membranes.

Vesicles are therefore promising model systems for the study of passive antibiotic transport across specific

lipid and porin pathways in bacteria. However, they have also been used to study active efflux via the reconstitution of bacterial efflux systems in proteoliposomes [113]. The LmrP multidrug transporter from *L. lactis* was successfully incorporated in a proteoliposome system by the van Veen group in Cambridge, where it was used to study calcium transport [113]. Complexes of OprM and MexA, two proteins from the MexA–MexB–OprM multidrug efflux pump of *Pseudomonas aeruginosa*, have also been reconstituted in SUVs and LUVs for the cryo-electron tomography of the protein complexes, suggesting that entire MDR efflux pump complexes can be successfully reconstituted [114]. The Picard group combined proteoliposomes containing the MexAB and the OprM parts of the complex to develop an assay that measured energy-dependent substrate transport in a system designed to mimic the Gram-negative double membrane [115]. The authors proposed that their model system be used to screen for inhibitors of the efflux pump, which would be valuable assets for restoring the activity of older antibiotics that would otherwise be subjected to efflux from the cell [115].

Protein reconstitution in SUVs or in multilamellar vesicles has traditionally been far easier than reconstitution in GUVs. Therefore, techniques to form GUV sized proteoliposomes from SUV proteoliposomes are of great interest, enabling the screening of multiple different bacterial transport proteins using the modern microfluidics approaches described above. The Spatz group has recently published a technique [116] that uses SUVs encapsulated in block-copolymer surfactant-stabilised water-in-oil droplets; by tuning the charge at the inner interface of the droplet, SUVs can be made to adsorb at the interface forming droplet-stabilised (ds) GUVs. They demonstrated a simple step to release these GUVs from the surfactant and oil, resulting in a high yield of GUV formation [116]. This paves the way for converting SUV-proteoliposomes into GUV-proteoliposomes which could then be investigated as described above, with high-resolution microscopy/microfluidics approaches.

Both OLA and the dsGUV techniques are relatively recent developments in the field and hold great promise for the construction of truly biomimetic models of bacterial membranes. The next technical leap would involve the *controlled* formation of double bilayers to better mimic Gram-negative cell envelopes. If this were to be achieved, it is not inconceivable to imagine the reconstitution of entire transporter complexes such as MDR efflux pumps, spanning the “inner” bilayer, the “periplasmic” inter-membrane space, and the “outer” bilayer, in these model systems. By subjecting them to the antibiotic transport measurements described above, a wealth of information could be untapped regarding the fluxes, activation barriers and selectivity of these crucial antibiotic transport pathways in bacteria.

A Note on the Economics of Antibiotic Research and Development

We end this review with a slight diversion from our topic of studying transport across bacterial membranes. As professional scientists, it is sometimes easy to lose track of the bigger picture when working on the technical minutiae of a problem. Why are we interested in understanding the details of antibiotic transport? Obviously, the major aim is to exploit our knowledge and techniques to develop more effective antibiotics and to circumvent existing drug resistance and tolerance mechanisms. However, to translate this work into antibiotic development and ultimately, to the benefit of actual patients, we require a drug development ecosystem whereby promising academic discoveries are implemented in a timely and effective manner in biotechnology and pharmaceutical companies.

However, as the field knows very well, the antibiotic pipeline has dried up—no new broad-spectrum antibiotic classes have been brought to the market in the past five decades [117]. This is not only a scientific but also an economic problem, as has been recognized by numerous studies on the issue [1,2,118,119]. In a volume-based payment system, revenues are driven by sales, and sales are driven by prescriptions; but with antibiotics, the more prescriptions there are, the faster resistance develops, thereby lowering the intrinsic value of the drug [120]. New antibiotics typically need to be preserved for last-resort cases in which all other options have been exhausted, which also translates to lower sales. Even extremely promising drug candidates such as Novartis' new monobactam derivative LYS228 are currently being pulled from the pipeline, with the company seeking to outlicense the therapeutic [121]. Unless the economic challenges are resolved, irrespective of the scientific advances made in understanding the basic biology and pharmacokinetics, real-world impact in clinics will not be observed.

In his seminal review of the field, Lord O'Neill stressed the importance of tackling this antibiotic development and supply problem [1]. Both he and others [120,121] have suggested giving market entry rewards of the order of \$1 billion to the developers of new antibiotics, to both develop and actually manufacture stocks of new antibiotics, which would be made available in clinical settings as appropriate. It is pertinent here to stress the importance of manufacturing and *maintaining* stock supplies, even if the drugs are used sparingly; it is worrying to note that economic considerations are leading to supply failures of commonly prescribed antibiotics in today's market, further exacerbating problems in infectious disease control [122]. Market entry rewards need participation from various governments, to tackle what is very much a global problem—pathogens do not recognize passport controls. A global antibiotics reward fund,

sponsored by the world's major economies, would be able to address this issue. Along these lines, it is heartening to note that the UK government has, in January 2019, announced a 20-year vision and a 5-year national action plan on antimicrobial resistance, which includes proposals to address the antibiotic market failure crisis and incentivise antibiotic development [123].

Besides governments, it is also worth considering the massive public and charitable engagement that other fields of medical research have generated, and contrast that with the case of antibiotic research. In 2017–18, Cancer Research UK raised £634 million [124]. The British Heart Foundation raised £159 million in 2016–17 [125]. In contrast, Antibiotic Research UK, the first charity in the World focusing specifically on antimicrobial resistance, had an income of £235,524 in 2016–17 [126], *3 orders of magnitude lower* than the others. These figures show just how much scope there is to increase funding for this crucial area of research through public engagement. Public awareness and charity involvement would boost basic research into antibiotics, which would help refresh the development pipeline and potentially lower entry barriers for pharmaceutical companies looking to enter or re-enter the field. We urge the scientific community, policy makers, heads of charitable organizations and the general public across the globe to consider these issues seriously and begin planning ways to massively increase funding to this vital sector, both in the areas of basic research such as the transport studies discussed here and in the translational programs involved in getting new drugs from the bench into the clinic.

Concluding Remarks

Despite the challenging economic environment for antibiotic R&D in general, research into drug permeation in bacteria has developed considerably over the past decade. These new techniques will pave the way for identifying and characterising the physicochemical properties of permeating compounds, which should help guide medicinal chemists in their quest for discovering novel antimicrobials with Gram-negative activity. The various methodological innovations involving MS, fluorimetry and microfluidic assays also offer new opportunities to better understand the relationships between genetic control of the transport pathways and bacterial adaptation to diverse environmental stresses. Real-time measurements of the intrabacterial accumulation of specific molecules, inducers, inhibitors and antimicrobial drugs represent important breakthroughs, which have been made possible by collaborations extending across traditional scientific boundaries. Recent developments in whole-cell assays, using MS and single-cell microscopy, have been vital to the field and show great promise in

providing crucial information for the rational design of new antibiotics. Complementary research programs with continuous feedback between the whole cell and model system based transport assays will be crucial for generating a holistic understanding of transport in these systems. Such programs will require molecular biologists, microbiologists, geneticists, medicinal chemists, biophysicists, mathematical modelers, bio-engineers and image analysis experts to succeed. Although the problem is extremely complex, the recent developments reviewed here suggest that, with the right scientific networks and funding modalities, a more fundamental understanding of drug permeation in Gram-negatives is accessible. This will be invaluable in the search for new drugs to tackle the next generation of resistant pathogens.

Acknowledgments

J.C. acknowledges support from the BBSRC and an ERC Consolidator grant awarded to Prof. Ulrich F. Keyser (Designerpores No. 647144). J.C. also acknowledges support from a Wellcome Trust Institutional Strategic Support Award to the University of Exeter (204909/Z/16/Z). A.M.H. acknowledges funding from Warwick University and Cross Council AMR research funding through grant no. MR/N002679/1. A. M.H. also thanks the Medicines Discovery Catapult for allowing access to their equipment, and Prof. Christopher Dowson and Dr. Adrian Lloyd for guidance and support. M.W. was part of the TRANSLOCATION consortium, which received support from the Innovative Medicines Initiatives Joint Undertaking under Grant Agreement No. 115525, resources that are composed of financial contributions from the European Union's seventh framework program (FP7/2007–2013) and in kind contributions from EFPIA companies.

Received 1 November 2018;

Received in revised form 29 January 2019;

Accepted 28 March 2019

Available online 05 April 2019

Keywords:

porins;
outer-membrane proteins;
envelope permeability;
efflux;
antibiotic resistance

Abbreviations used:

MS, mass spectrometry; HTS, high-throughput screening; MD, molecular dynamics; SUV, small unilamellar vesicle; LUV, large unilamellar vesicle; GUV, giant unilamellar vesicle; OLA, octanol-assisted liposome assembly.

References

- [1] J. O'Neill, Tackling Drug-Resistant Infections Globally: Final Report and Recommendations, 2016.
- [2] C. Årdal, et al., DRIVE-AB Report: Revitalizing the Antibiotic Pipeline, 2018.
- [3] D.J. Payne, M.N. Gwynn, D.J. Holmes, D.L. Pompliano, Drugs for bad bugs: confronting the challenges of antibacterial discovery, *Nat. Rev. Drug Discov.* 6 (2007) 29–40.
- [4] L.J.V. Piddock, The crisis of no new antibiotics—what is the way forward? *Lancet Infect. Dis.* 12 (2012) 249–253.
- [5] P. Kardas, S. Devine, A. Golembesky, C. Roberts, A systematic review and meta-analysis of misuse of antibiotic therapies in the community, *Int. J. Antimicrob. Agents* 26 (2005) 106–113.
- [6] G.G. Khachatourians, Agricultural use of antibiotics and the evolution and transfer of antibiotic-resistant bacteria, *CMAJ* 159 (1998) 1129–1136.
- [7] R. Ofori-Asenso, “When the bug cannot be killed”—the rising challenge of antimicrobial resistance, *Medicines* 4 (2017) 40.
- [8] L. Chen, R. Todd, J. Kiehbauch, M. Walters, A. Kallen, Pan-resistant New Delhi metallo-beta-lactamase-producing *Klebsiella pneumoniae*—Washoe County, Nevada, 2016, *Morb. Mortal. Wkly Rep.* 66 (2017) 33.
- [9] R.A. Stavenger, M. Winterhalter, TRANSLOCATION project: how to get good drugs into bad bugs, *Sci. Transl. Med.* 6 (2014) 6–8.
- [10] L.L. Silver, A gestalt approach to gram-negative entry, *Bioorg. Med. Chem.* 24 (2016) 6379–6389.
- [11] T. Kostyanov, et al., The innovative medicines Initiative's new drugs for bad bugs programme: European public-private partnerships for the development of new strategies to tackle antibiotic resistance, *J. Antimicrob. Chemother.* 71 (2016) 290–295.
- [12] K. Talkington, C. Shore, K. Pooja, The PEW Charitable Trusts: A Scientific Roadmap for Antibiotic Discovery, 2016.
- [13] J.-M. Pagès, C.E. James, M. Winterhalter, The porin and the permeating antibiotic: a selective diffusion barrier in gram-negative bacteria, *Nat. Rev. Microbiol.* 6 (2008) 893–903.
- [14] H.I. Zgurskaya, C.A. Lopez, S. Gnanakaran, Permeability barrier of gram-negative cell envelopes and approaches to bypass it, *ACS Infect. Dis.* 1 (2015) 512–522.
- [15] A.H. Delcour, Outer membrane permeability and antibiotic resistance, *Biochim. Biophys. Acta* 1794 (2009) 808–816.
- [16] J.I. Manchester, E.T. Buurman, G.S. Bisacchi, R.E. McLaughlin, Molecular determinants of AcrB-mediated bacterial efflux implications for drug discovery, *J. Med. Chem.* 55 (2012) 2532–2537.
- [17] H. Nikaido, Porins and specific diffusion channels in bacterial outer membranes, *J. Biol. Chem.* 269 (1994) 3905–3908.
- [18] H. Nikaido, Molecular basis of bacterial outer membrane permeability revisited, *Microbiol. Mol. Biol. Rev.* 67 (2003) 593–656.
- [19] A. Missner, P. Pohl, 110 years of the Meyer–Overton rule: predicting membrane permeability of gases and other small compounds, *ChemPhysChem* 10 (2009) 1405–1414.
- [20] Q. Al-Awqati, One hundred years of membrane permeability: does Overton still rule? *Nat. Cell Biol.* 1 (1999) E201–E202.
- [21] X.-Z. Li, H. Nikaido, Efflux-mediated drug resistance in bacteria, *Drugs* 64 (2004) 159–204.

- [22] D. Du, et al., Structure of the AcrAB–TolC multidrug efflux pump, *Nature* 509 (2014) 512–515.
- [23] X.-Z. Li, D.M. Livermore, H. Nikaido, Role of efflux pump(s) in intrinsic resistance of *Pseudomonas aeruginosa*: resistance to tetracycline, chloramphenicol, and norfloxacin, *Antimicrob. Agents Chemother.* 38 (1994) 1732–1741.
- [24] Y. Pu, et al., Enhanced efflux activity facilitates drug tolerance in dormant bacterial cells, *Mol. Cell* 62 (2016) 284–294.
- [25] T. Nakae, H. Nikaido, Outer membrane as a diffusion barrier in *Salmonella typhimurium*, *J. Biol. Chem.* 250 (1975) 7359–7365.
- [26] T. Nakae, Identification of the outer membrane protein of *E. coli* that produces transmembrane channels in reconstituted vesicle membranes, *Biochem. Biophys. Res. Commun.* 71 (1976) 877–884.
- [27] R. Benz, K. Janko, W. Boos, P. Luger, Formation of large, ion-permeable membrane channels by the matrix protein (porin) of *Escherichia coli*, *BBA* 511 (1978) 305–319.
- [28] H. Schindler, J.P. Rosenbusch, Matrix protein from *Escherichia coli* outer membranes forms voltage-controlled channels in lipid bilayers, *Proc. Natl. Acad. Sci. U. S. A.* 75 (1978) 3751–3755.
- [29] S.W. Cowan, et al., Crystal structures explain functional properties of two *E. coli* porins, *Nature* 358 (1992) 727–733.
- [30] M. Luckey, H. Nikaido, Specificity of diffusion channels produced by lambda phage receptor protein of *Escherichia coli*, *Proc. Natl. Acad. Sci. U. S. A.* 77 (1980) 167–171.
- [31] H. Nikaido, E.Y. Rosenberg, Porin channels in *Escherichia coli*: studies with liposomes reconstituted from purified proteins, *J. Bacteriol.* 153 (1983) 241–252.
- [32] H. Nikaido, E.Y. Rosenberg, Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of *Escherichia coli*, *J. Gen. Physiol.* 77 (1981) 121–135.
- [33] H. Nikaido, Porins and specific channels of bacterial outer membranes, *Mol. Microbiol.* 6 (1992) 435–442.
- [34] P.G. Mortimer, L.J. Piddock, A comparison of methods used for measuring the accumulation of quinolones by Enterobacteriaceae, *Pseudomonas aeruginosa* and *Staphylococcus aureus*, *J. Antimicrob. Chemother.* 28 (1991) 639–653.
- [35] A. Smith, et al., The culture environment influences both gene regulation and phenotypic heterogeneity in *Escherichia coli*, *Front. Microbiol.* 9 (2018) 1739.
- [36] X. Liu, T. Ferenci, Regulation of porin-mediated outer membrane permeability by nutrient limitation in *Escherichia coli*, *J. Bacteriol.* 180 (15) (1998) 3917–3922.
- [37] R.A. Bamford, et al., Investigating the physiology of viable but non-culturable bacteria by microfluidics and time-lapse microscopy, *BMC Biol.* 15 (2017) 121.
- [38] D.A. Six, T. Krucker, J.A. Leeds, Advances and challenges in bacterial compound accumulation assays for drug discovery, *Curr. Opin. Chem. Biol.* 44 (2018) 9–15.
- [39] D. Du, et al., Multidrug efflux pumps: structure, function and regulation, *Nat. Rev. Microbiol.* 16 (2018) 523–539.
- [40] S. Chevalier, et al., Structure, function and regulation of *Pseudomonas aeruginosa* porins, *FEMS Microbiol. Rev.* 41 (5) (2017) 698–722.
- [41] M.F. Richter, et al., Predictive compound accumulation rules yield a broad-spectrum antibiotic, *Nature* 545 (2017) 299–304.
- [42] Y. Zhou, et al., Thinking outside the “bug”: a unique assay to measure intracellular drug penetration in gram-negative bacteria, *Anal. Chem.* 87 (2015) 3579–3584.
- [43] A. Schumacher, et al., Intracellular accumulation of linezolid in *Escherichia coli*, *Citrobacter freundii* and *Enterobacter aerogenes*: role of enhanced efflux pump activity and inactivation, *J. Antimicrob. Chemother.* 59 (2007) 1261–1264.
- [44] H. Cai, K. Rose, L.-H. Liang, S. Dunham, C. Stover, Development of a liquid chromatography/mass spectrometry-based drug accumulation assay in *Pseudomonas aeruginosa*, *Anal. Biochem.* 385 (2009) 321–325.
- [45] A.R. Brown, et al., A mass spectrometry-based assay for improved quantitative measurements of efflux pump inhibition, *PLoS One* 10 (2015), e0124814.
- [46] R. Iyer, et al., Evaluating LC-MS/MS to measure accumulation of compounds within bacteria, *ACS Infect. Dis.* 4 (2018) 1336–1345.
- [47] T.D. Davis, C.J. Gerry, D.S. Tan, General platform for systematic quantitative evaluation of small-molecule permeability in bacteria, *ACS Chem. Biol.* 9 (2014) 2535–2544.
- [48] J.M. Stokes, et al., Cold stress makes *Escherichia coli* susceptible to glycopeptide antibiotics by altering outer membrane integrity, *Cell Chem. Biol.* 23 (2016) 267–277.
- [49] B. Spangler, et al., Molecular probes for the determination of sub-cellular compound exposure profiles in gram-negative bacteria, *ACS Infect. Dis.* 4 (2018) 1355–1367.
- [50] H. Tian, D.A. Six, T. Krucker, J.A. Leeds, N. Winograd, Subcellular chemical imaging of antibiotics in single bacteria using C60-secondary ion mass spectrometry, *Anal. Chem.* 89 (2017) 5050–5057.
- [51] H. Prochnow, et al., Subcellular quantification of uptake in gram-negative bacteria, *Anal. Chem.* 91 (3) (2019) 1863–1872 (Article ASAP).
- [52] H. Zhang, Acoustic dispensing-mass spectrometry: the next high throughput bioanalytical platform for early drug discovery, *Bioanalysis* 9 (2017) 1619–1621.
- [53] I. Sinclair, et al., Novel acoustic loading of a mass spectrometer: toward next-generation high-throughput MS screening, *J. Lab. Autom.* 21 (2016) 19–26.
- [54] D.A. Westfall, et al., Bifurcation kinetics of drug uptake by gram-negative bacteria, *PLoS One* 12 (2017), e0184671.
- [55] S. Kašćáková, L. Maigre, J. Chevalier, M. Réfrégiers, J.M. Pagès, Antibiotic transport in resistant bacteria: synchrotron UV fluorescence microscopy to determine antibiotic accumulation with single cell resolution, *PLoS One* 7 (2012), e38624.
- [56] J. Vergalli, et al., Spectrofluorimetric quantification of antibiotic drug concentration in bacterial cells for the characterization of translocation across bacterial membranes, *Nat. Protoc.* 13 (2018) 1348–1361.
- [57] J. Vergalli, et al., Fluoroquinolone structure and translocation flux across bacterial membrane, *Sci. Rep.* 7 (2017) 9821.
- [58] B. Cinquin, et al., Microspectrometric insights on the uptake of antibiotics at the single bacterial cell level, *Sci. Rep.* 5 (2015), 17968.
- [59] M. Masi, et al., Fluorescence enlightens RND pump activity and the intrabacterial concentration of antibiotics, *Res. Microbiol.* 169 (2018) 432–441.
- [60] A. Allam, et al., Microspectrofluorimetry to dissect the permeation of ceftazidime in gram-negative bacteria, *Sci. Rep.* 7 (2017) 986.
- [61] W. Phetsang, et al., An azido-oxazolidinone antibiotic for live bacterial cell imaging and generation of antibiotic variants, *Bioorg. Med. Chem.* 22 (2014) 4490–4498.
- [62] W. Phetsang, et al., Fluorescent trimethoprim conjugate probes to assess drug accumulation in wild type and mutant *Escherichia coli*, *ACS Infect. Dis.* 2 (2016) 688–701.
- [63] M.R.L. Stone, M.S. Butler, W. Phetsang, M.A. Cooper, M.A. T. Blaskovich, Fluorescent Antibiotics: new research tools to fight antibiotic resistance, *Trends Biotechnol.* 36 (2018) 523–536.

- [64] S. Hong, et al., Assessing antibiotic permeability of gram-negative bacteria via nanofluidics, *ACS Nano* 11 (2017) 6959–6967.
- [65] H. Heidari Torkabadi, et al., Following drug uptake and reactions inside *Escherichia coli* cells by Raman microspectroscopy, *Biochemistry* 53 (2014) 4113–4121.
- [66] P.R. Carey, H. Heidari-Torkabadi, New techniques in antibiotic discovery and resistance: Raman spectroscopy, *Ann. N. Y. Acad. Sci.* 1354 (2015) 67–81.
- [67] H. Heidari Torkabadi, et al., Raman spectra of interchanging β -lactamase inhibitor intermediates on the millisecond time scale, *J. Am. Chem. Soc.* 135 (2013) 2895–2898.
- [68] H. Nikaïdo, E.Y. Rosenberg, J. Foulds, Porin channels in *Escherichia coli*: studies with beta-lactams in intact cells, *J. Bacteriol.* 153 (1983) 232–240.
- [69] S. Kojima, H. Nikaïdo, Permeation rates of penicillins indicate that *Escherichia coli* porins function principally as nonspecific channels, *Proc. Natl. Acad. Sci.* 110 (2013) E2629–E2634.
- [70] S. Kojima, H. Nikaïdo, High salt concentrations increase permeability through OmpC channels of *Escherichia coli*, *J. Biol. Chem.* 289 (2014) 26464–26473.
- [71] S.P. Bhamidimarri, et al., A multidisciplinary approach towards identification of antibiotic scaffolds for *Acinetobacter baumannii*, *Structure* 27 (2) (2019) 268–280.
- [72] T. Gutschmann, T. Heimburg, U.F. Keyser, K.R. Mahendran, M. Winterhalter, Protein reconstitution into freestanding planar lipid membranes for electrophysiological characterization, *Nat. Protoc.* 10 (2015) 188–198.
- [73] A. Baslé, R. Iyer, A.H. Delcour, Subconductance states in OmpF gating, *Biochim. Biophys. Acta* 1664 (2004) 100–107.
- [74] K.R. Mahendran, M. Kreir, H. Weingart, N. Fertig, M. Winterhalter, Permeation of antibiotics through *Escherichia coli* OmpF and OmpC porins: screening for influx on a single-molecule level, *J. Biomol. Screen.* 15 (2010) 302–307.
- [75] N. Saint, et al., Structural and functional characterization of OmpF Porin mutants selected for larger pore size, *J. Biol. Chem.* 271 (1996) 20676–20680.
- [76] R. Benz, K. Janko, P. Läuger, Ionic selectivity of pores formed by the matrix protein (porin) of *Escherichia coli*, *BBA* 551 (1979) 238–247.
- [77] R. Benz, A. Schmid, R.E.W. Hancock, Ion selectivity of gram-negative bacterial porins, *J. Bacteriol.* 162 (1985) 722–727.
- [78] S. Pezeshki, C. Chimerel, A.N. Bessonov, M. Winterhalter, U. Kleinekathöfer, Understanding ion conductance on a molecular level: an all-atom modeling of the bacterial porin OmpF, *Biophys. J.* 97 (2009) 1898–1906.
- [79] S.P. Bhamidimarri, J.D. Prajapati, B. van den Berg, M. Winterhalter, U. Kleinekathöfer, Role of electroosmosis in the permeation of neutral molecules: CymA and cyclodextrin as an example, *Biophys. J.* 110 (2016) 600–611.
- [80] I. Ghai, et al., General method to determine the flux of charged molecules through nanopores applied to β -lactamase inhibitors and OmpF, *J. Phys. Chem. Lett.* 8 (2017) 1295–1301.
- [81] W. Im, B. Roux, Ions and counterions in a biological channel: a molecular dynamics simulation of OmpF porin from *Escherichia coli* in an explicit membrane with 1 M KCl aqueous salt solution, *J. Mol. Biol.* 319 (2002) 1177–1197.
- [82] C. Chimerel, L. Movableu, S. Pezeshki, M. Winterhalter, U. Kleinekathöfer, Transport at the nanoscale: temperature dependence of ion conductance, *Eur. Biophys. J.* 38 (2008) 121–125.
- [83] C. Andersen, M. Jordy, R. Benz, Evaluation of the rate constants of sugar transport through maltoporin (LamB) of *Escherichia coli* from the sugar-induced current noise, *J. Gen. Physiol.* 105 (1995) 385–401.
- [84] R. Dutzler, Y.-F. Wang, P.J. Rizkallah, J.P. Rosenbusch, T. Schirmer, Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway, *Structure* 4 (1996) 127–134.
- [85] K.R. Mahendran, et al., Molecular basis of Enrofloxacin translocation through OmpF, an outer membrane channel of *Escherichia coli*—when binding does not imply translocation, *J. Phys. Chem. B* 114 (2010) 5170–5179.
- [86] I. Ghai, et al., Ampicillin permeation across OmpF, the major outer-membrane channel in *Escherichia coli*, *J. Biol. Chem.* 293 (2018) 7030–7037.
- [87] P.R. Singh, M. Ceccarelli, M. Lovelle, M. Winterhalter, K.R. Mahendran, Antibiotic permeation across the OmpF channel: modulation of the affinity site in the presence of magnesium, *J. Phys. Chem. B* 116 (2012) 4433–4438.
- [88] J. Wang, J.A. Bafna, S.P. Bhamidimarri, M. Winterhalter, Permeation of small molecules through membrane channels: chemical modification for quantification of transport via OmpF, *Angew. Chem.* 131 (2019) 4788–4792.
- [89] E. Rideau, R. Dimova, P. Schwill, F.R. Wurm, K. Landfester, Liposomes and polymersomes: a comparative review towards cell mimicking, *Chem. Soc. Rev.* 47 (2018) 8572–8610.
- [90] H. Jousma, et al., Characterization of liposomes. The influence of extrusion of multilamellar vesicles through polycarbonate membranes on particle size, particle size distribution and number of bilayers, *Int. J. Pharm.* 35 (1987) 263–274.
- [91] B. Christiaens, et al., Tryptophan fluorescence study of the interaction of penetratin peptides with model membranes, *Eur. J. Biochem.* 269 (2002) 2918–2926.
- [92] G. Ghale, et al., Chemosensing ensembles for monitoring biomembrane transport in real time, *Angew. Chem. Int. Ed.* 53 (2014) 2762–2765.
- [93] P. Kuhn, K. Eyer, S. Allner, D. Lombardi, P.S. Dittrich, A microfluidic vesicle screening platform: monitoring the lipid membrane permeability of tetracyclines, *Anal. Chem.* 83 (2011) 8877–8885.
- [94] S. Li, P. Hu, N. Malmstadt, Confocal imaging to quantify passive transport across biomimetic lipid membranes, *Anal. Chem.* 82 (2010) 7766–7771.
- [95] S. Piñero-Fernandez, C. Chimerel, U.F. Keyser, D.K. Summers, Indole transport across *Escherichia coli* membranes, *J. Bacteriol.* 193 (2011) 1793–1798.
- [96] M.I. Angelova, D.S. Dimitrov, Liposome electroformation, *Faraday Discuss. Chem. Soc.* 81 (1986) 303–311.
- [97] J. Cama, C. Chimerel, S. Pagliara, A. Javer, U.F. Keyser, A label-free microfluidic assay to quantitatively study antibiotic diffusion through lipid membranes, *Lab Chip* 14 (2014) 2303–2308.
- [98] J. Cama, et al., Direct optofluidic measurement of the lipid permeability of fluoroquinolones, *Sci. Rep.* 6 (2016), 32824.
- [99] A. Avdeef, et al., PAMPA—critical factors for better predictions of absorption, *J. Pharm. Sci.* 96 (2007) 2893–2909.
- [100] S. Purushothaman, J. Cama, U.F. Keyser, Dependence of norfloxacin diffusion across bilayers on lipid composition, *Soft Matter* 12 (2016) 2135–2144.
- [101] M.F. Richter, P.J. Hergenrother, The challenge of converting Gram-positive-only compounds into broad-spectrum antibiotics, *Ann. N. Y. Acad. Sci.* 1435 (2019) 18–38.

- [102] R.M. Epanand, R.F. Epanand, Lipid domains in bacterial membranes and the action of antimicrobial agents, *Biochim. Biophys. Acta* 1788 (2009) 289–294.
- [103] J. Cama, et al., Quantification of fluoroquinolone uptake through the outer membrane channel OmpF of *Escherichia coli*, *J. Am. Chem. Soc.* 137 (2015) 13836–13843.
- [104] H. Stein, S. Spindler, N. Bonakdar, C. Wang, V. Sandoghdar, Production of isolated giant unilamellar vesicles under high salt concentrations, *Front. Physiol.* 8 (2017) 63.
- [105] I.L. Jørgensen, G.C. Kemmer, T.G. Pomorski, Membrane protein reconstitution into giant unilamellar vesicles: a review on current techniques, *Eur. Biophys. J.* 46 (2017) 103–119.
- [106] K. Kamiya, S. Takeuchi, Giant liposome formation toward the synthesis of well-defined artificial cells, *J. Mater. Chem. B* 5 (2017) 5911–5923.
- [107] S. Deshpande, Y. Caspi, A.E.C. Meijering, C. Dekker, Octanol-assisted liposome assembly on chip, *Nat. Commun.* 7 (2016), 10447.
- [108] S. Deshpande, A. Birnie, C. Dekker, On-chip density-based purification of liposomes, *Biomicrofluidics* 11 (2017), 034106.
- [109] M. Schaich, et al., An integrated microfluidic platform for quantifying drug permeation across biomimetic vesicle membranes, *Mol. Pharmaceutics* (2019) <https://doi.org/10.1021/acs.molpharmaceut.9b00086> (in press).
- [110] A.J.M. Driessen, C. Van Leeuwen, W.N. Konings, Transport of basic amino acids by membrane vesicles of *Lactococcus lactis*, *J. Bacteriol.* 171 (1989) 1453–1458.
- [111] C. Marty-Teyssset, J.S. Lolkema, P. Schmitt, C. Divies, W.N. Konings, Membrane potential-generating transport of citrate and malate catalyzed by CitP of *Leuconostoc mesenteroides*, *J. Biol. Chem.* 270 (1995) 25370–25376.
- [112] K. Al Nahas, et al., A microfluidic platform for the characterisation of membrane active antimicrobials, *Lab Chip* 19 (2019) 837–844.
- [113] T.A. Schaedler, Z. Tong, H.W. van Veen, The multidrug transporter LmrP protein mediates selective calcium efflux, *J. Biol. Chem.* 287 (2012) 27682–27690.
- [114] S. Tréput, et al., Structure of reconstituted bacterial membrane efflux pump by cryo-electron tomography, *Biochim. Biophys. Acta Biomembr.* 1798 (2010) 1953–1960.
- [115] A. Verchère, M. Dezi, V. Adrien, I. Broutin, M. Picard, In vitro transport activity of the fully assembled MexAB–OprM efflux pump from *Pseudomonas aeruginosa*, *Nat. Commun.* 6 (2015) 6890.
- [116] B. Haller, et al., Charge-controlled microfluidic formation of lipid-based single- and multicompartment systems, *Lab Chip* 18 (2018) 2665–2674.
- [117] K. Lewis, Platforms for antibiotic discovery, *Nat. Rev. Drug Discov.* 12 (2013) 371–387.
- [118] J.H. Rex, K. Outtersson, Antibiotic reimbursement in a model delinked from sales: a benchmark-based worldwide approach, *Lancet Infect. Dis.* 16 (2016) 500–505.
- [119] C. Árdal, J.-A. Røttingen, A. Opalska, A.J. Van Hengel, J. Larsen, Pull incentives for antibacterial drug development: an analysis by the transatlantic task force on antimicrobial resistance, *Clin. Infect. Dis.* 65 (2017) 1378–1382.
- [120] G.W. Daniel, M. Schneider, M.B. McClellan, Addressing antimicrobial resistance and stewardship: the priority antimicrobial value and entry (PAVE) award, *JAMA* 318 (2017) 1103–1104.
- [121] Editorial. Wanted: A reward for antibiotic development, *Nat. Biotechnol.* 36 (2018) 555.
- [122] S. Nurse-Findlay, et al., Shortages of benzathine penicillin for prevention of mother-to-child transmission of syphilis: an evaluation from multi-country surveys and stakeholder interviews, *PLoS Med.* 14 (2017), e1002473.
- [123] UK Government, Tackling Antimicrobial Resistance 2019–2024. The UK's Five-Year National Action Plan, 2019.
- [124] Cancer Research UK, Annual Report and Accounts, 2018.
- [125] British Heart Foundation, Annual Report 2016–17, 2017.
- [126] Antibiotic Research UK, Fighting antibiotic resistance today and tomorrow, Annual report and accounts 2016, 2017.