



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Perez Chavarria, Roberto

Title:

Understanding colistin resistance in Gram-negative bacteria.

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Understanding colistin resistance in Gram-negative bacteria.

By

Roberto Pérez Chavarría.



“A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of PhD in Cellular and Molecular Medicine in the Faculty of Cellular and Molecular Medicine.

July, 2021.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Roberto Perez Chavarria. DATE: 19/07/2021

Covid Impact Statement.

Lockdown measures taken to deal with the COVID-19 pandemic caused by SARS-CoV-2/ meant that my work was disrupted at crucial moments.

Access to laboratory facilities was completely interrupted between the months of March and June 2020. During this time no laboratory work at all was possible. Partial access to the University and laboratory facilities was restored, starting mid-June 2020, however, social-distancing limitations mean that since this date access to the laboratory is operating on a rota system, and access to essential shared equipment and facilities housed in other laboratories has been similarly restricted, reducing the amount of laboratory work possible. Relevant affected facilities include equipment for protein purification (cell disruptor, centrifuges and ultracentrifuge), biophysical characterisation (CD spectropolarimeter, SEC-MALLS equipment) and protein crystallisation (crystallisation robot).

Abstract:

The antibiotic colistin is considered a “last resort” treatment for bacterial infections caused by multi-drug resistant Gram-negative bacteria. The efficacy of this antibiotic is now threatened by the emergence of the plasmid-borne mobilised colistin resistance (*mcr*) genes. MCR-1 is a zinc-dependent membrane-bound phosphoethanolamine transferase, that modifies the Lipid A in the bacterial outer membrane; conferring resistance against colistin. This project aims to better understand this recently discovered resistance mechanism, MCR-1, and its homologues, as well as develop a platform that could serve as a basis for the discovery and development of possible inhibitors.

Structure and sequence-based analysis were performed to gather a better understanding of the position of MCR-1, and other phosphoethanolamine transferases, within the alkaline phosphatase protein superfamily to which these enzymes belong. MCR-1 was recombinantly produced and purified in detergent micelles and its integrity was analysed by Size-exclusion chromatography coupled to a multi-angle light scattering system (SEC-MALS), circular dichroism (CD) as well and mass spectrometry analysis. An activity assay was developed by employing using “catalytic domain only” constructs of MCR-1 its homologs MCR-2 in conjunction with the synthetic chromophoric substrate 4-nitrophenol-phosphoethanolamine (pNP-PEtN). Another phosphoethanolamine transferase, PmrC from *Acinetobacter baumannii*, was recombinantly produced and used, its activity was assessed, and crystallization attempts were made. Analysis of the sequence and structure revealed that MCR-1 has a partial second metal site resembling those of the multimeric members of the alkaline phosphatase superfamily. SEC-MALS, CD and melting temperature analysis showed that recombinantly produced MCR-1 is stable in detergent micelles. The developed activity assay efficiently monitors the cleavage of the phosphoethanolamine (PEtN) motif from the chromophoric pNP-PEtN and allowed the estimation of kinetic parameters. Finally, a “catalytic

domain only” construct of PmrC was recombinantly produced, using the activity assay kinetic parameters were estimated and starting conditions for protein crystallization were found.

The work presented here makes a significant contribution towards understanding phosphoethanolamine transferase-mediated colistin resistance in Gram-negative bacteria by producing and characterising relevant enzymes (MCR-1 and PmrC) and developing an activity assay that can serve as a platform for the identification and screening of inhibitors. Structural and sequence-based analysis highlighted key conserved residues involved in the catalytic action of this family of enzymes. Exploiting the findings from this project in the design of inhibitors for phosphoethanolamine transferases will help to preserve the efficacy of colistin against multi-drug resistant gram-negative bacterial infections.

Acknowledgments.

I would like to sincerely thank:

My supervisor, Prof. James Spencer, for trusting in me and allowing me to work as part of his research group and undertake my PhD. He has been nothing but a terrific supervisor, supporting me not only for the research project but during my transition moving to the United Kingdom from a different country.

To Prof. Matthew Avison, my co-supervisor, for his advice and humour during group meetings.

To Dr Philip Hinchliffe for all the help and advice provided during the elaboration of this project.

To Prof. Paul Pringle and research associate Adam Gorman, from the Chemistry department, for their help synthesizing the chromogenic compounds used in this work. To Prof. Chris Wills and Dr Luoyi Wang, from the Chemistry department, for their help and support with the LC-MS analysis done in this work.

To my fellow PhD students Emily Lythell and Zongfang Yang, for all the help and advice provided regarding the use of specialized software and the idea of sharing and stimulating conversations.

To the faculty staff members, especially John Shaw, for helping me with the technical aspects using various equipment and sorting administrative issues.

To my family for always being there for me, to my mother for lending me an ear and listening to my rantings, to my father for helping me keep a cool head and a clear sight of my goals, to my brother for keeping me grounded and to my partner for supporting me during all the crazy moments (especially writing this document).

To my best friend Eva for always being here, through sunny and rainy days, and for her incredible support during all these years.

To my friends near and far, old and new. For all their support, enthusiasm and motivation.

To CONACyT for the financial help provided to accomplish this project.

Table of Contents.

Understanding colistin resistance in Gram-negative bacteria.	1
Author’s declaration.....	i
Covid Impact Statement.	ii
Abstract:	iii
Acknowledgments.....	v
Table of Contents.	vi
List of Figures:	xii
List of Tables:	xvii
List of Abbreviations.....	xviii
Chapter Outline:.....	xx
Chapter 1 General Introduction.....	1
1.1 Antibiotics and Antibiotic Resistance.....	1
1.1.1 The emergence of multi-drug resistant Gram-Negative bacteria.....	5
1.2 The outer membrane of Gram-negative bacteria and Lipid A.....	6
1.2.1 Lipopolysaccharide and Lipid A.....	7
1.3 Cationic peptides and LPS/outer membrane targeting antibiotics.....	8
1.4 Polymyxins.....	9
1.4.1 Structure and mechanism of action.....	9
1.4.2 Last resort antibiotic.....	11
1.4.3 Polymyxin Resistance.....	12
1.5 Mobile Colistin Resistance.....	17

1.5.1	The emergence of MCR-1.	17
1.5.2	MCR-1 dissemination.	19
1.6	MCR-1 structure and characteristics.	24
1.7	Aims of the project:	28
Chapter 2 Materials and Methods.		29
2.1	Bioinformatics and structural modelling	29
2.1.1	MCR-1 and ScPmrC 3D modelling.	29
2.1.2	Multiple Sequence Analysis of known Phosphoethanolamine transferases.	29
2.1.3	MCR-1 structural classification	30
2.1.4	Protein Sequence analysis.	30
2.1.5	Structural similarity database generation.	31
2.1.6	Structural Data Analysis using Dali.	31
2.2	Experimental Methods.	32
2.2.1	Expression Plasmids and gene cloning.	32
2.2.2	Small expression trials.	36
2.2.3	Large scale expression of recombinant proteins.	40
2.2.4	Full-length MCR-1 recovery and purification.	41
2.2.5	Large scale recovery and purification of cMCR-1, cMCR-2 and ScPmrC.	42
2.2.6	Protein analysis by Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE). .	44
2.2.7	Protein concentration determination.	44
2.2.8	Protein identification by MS/MS analysis.	45
2.2.9	MCR-1 SEC-MALS Analysis.	46
2.2.10	Circular Dichroism Spectroscopy of MCR-1 and ScPmrC.	48
2.2.11	Synthesis of 2-aminoethyl (4-nitrophenyl) hydrogen phosphate (pNP-PEtN).	49
2.2.12	Buffer pH and pNP-PEtN stability.	50
2.2.13	Assessing the activity of recombinant cMCR-1, cMCR-2 and ScPmrC.	50
2.2.14	Analysis of the enzyme reaction by LC-MS.	51

2.2.15	Estimation of the steady-state kinetic parameters for cMCR-2 and ScPmrC.	53
2.2.16	Evaluation of Thioglycolic Acid as a small-molecule inhibitor of cMCR-2.....	55
2.2.17	Crystallisation trials of ScPmrC.	55
2.2.18	Crystallization optimization.	56
Chapter 3	<i>Structure based phylogenetic study of the alkaline phosphatase superfamily.</i>	
	57	
3.1	Introduction:	57
3.1.1	The Alkaline Phosphatase superfamily.	58
3.1.2	Structural and sequence phylogenetic study of enzymes.....	59
3.2	Results.....	62
3.2.1	Phosphoethanolamine transferases as members of the alkaline phosphatase superfamily.....	62
3.2.2	Structural analysis using the Dali Server.	64
3.2.3	Analysis of the Dali PDB25 data set.	66
3.2.4	Selection and DALI Analysis of a set of representative AlkP structures.	70
3.2.5	Analysis of metal ion binding sites from structure-based sequence alignment.	84
3.2.6	Sequence-based analysis of the position of MCR-1 in the alkaline phosphatase superfamily.	88
3.3	Conclusions.	91
Chapter 4	<i>Production of Recombinant MCR-1.</i>	95
4.1	Introduction.	95
4.1.1	Optimization of the production and purification process of membrane proteins.....	95
4.1.2	Biophysical Characterization of Membrane Proteins.....	97
4.2	Results.....	98
4.2.1	Small scale MCR-1 protein expression and purification.....	98
4.2.2	Small scale titratable expression using Lemo21(DE3) system.	99
4.2.3	Media optimization and protein production.	100
4.2.4	Large-scale production and purification of MCR-1.	105

4.2.5	MCR-1 and tMCR-1 identification by peptide mass fingerprint.	112
4.2.6	SEC-MALS analysis.....	115
4.2.7	Circular Dichroism and thermal stability analysis.	121
4.3	Conclusions.	123
Chapter 5	<i>Development of an in vitro biochemical assay to assess MCR-1 activity. ...</i>	127
5.1	Introduction.	127
5.1.1	Previous approaches to assay phosphoethanolamine transferase activity.	129
5.1.2	Criteria for detection of phosphoethanolamine transferase activity	130
5.2	Results and discussion.	131
5.2.1	Small scale expression and purification of the C-terminal catalytic domains of MCR-1 and MCR-2. 131	
5.2.2	Large Scale Production and purification of cMCR-1 and cMCR-2.	132
5.2.3	Development of a biochemical assay.....	135
5.2.4	Assessing the activity of recombinant MCR catalytic domains.	138
5.2.5	Estimation of the steady-state kinetic parameters for cMCR-2.....	146
5.2.6	Inhibition of MCR-2 by Thioglycolic acid.....	148
5.3	Conclusions.	151
Chapter 6	<i>Production and characterisation of the MCR homologue PmrC from</i>	
	<i>Acinetobacter baumannii.</i>	155
6.1	Introduction:	155
6.1.1	<i>Acinetobacter baumannii</i> microbiology and clinical relevance.....	155
6.1.2	Community-acquired <i>A. baumannii</i> infections.	157
6.1.3	Multi-Drug resistance of <i>A. baumannii</i>	157
6.1.4	<i>A. baumannii</i> and colistin resistance.	159
6.2	Results.....	168

6.2.1	Cloning, small scale expression and purification of the recombinant catalytic domain of <i>A. baumannii</i> PmrC.....	168
6.2.2	Large Scale ScPmrC production.....	173
6.2.3	Circular dichroism spectroscopy of ScPmrC.....	178
6.2.4	ScPmrC Activity: Cleavage of PNP-PEtN.	181
6.2.5	Estimation of the Kinetic parameters of ScPmrC.	182
6.2.6	ScPmrC Crystallisation.....	184
6.2.7	Optimisation of ScPmrC crystallization conditions.	187
6.3	Conclusions.	189
Chapter 7	General conclusions and future work.	194
References:	201
Supplementary data:	229
Appendix A.	<i>E. coli</i> MCR-1 DNA and amino acid sequence.	229
Appendix B.	<i>E. coli</i> Full-length MCR-1 Properties.	230
Appendix C.	Full-length MCR-1 protein identification by mass fingerprint analysis. 231	
Appendix D.	Full-length MCR-1 Mass fingerprint analysis of peptide modifications by tryptic and chymotryptic digestion.	232
Appendix E.	Truncated MCR-1 protein identification by mass fingerprint analysis. 235	
Appendix F.	<i>E. coli</i> cMCR-1 DNA and amino acid sequence.	236
Appendix G.	<i>E. coli</i> cMCR-1 Properties.	237
Appendix H.	<i>E. coli</i> cMCR-2 DNA and amino acid sequence.	238
Appendix I.	<i>E. coli</i> cMCR-2 Properties.	239

Appendix J.	cMCR-2 mass fingerprint analysis.....	240
Appendix K.	cMCR-2 Mass fingerprint analysis of peptide modifications.	241
Appendix L.	PDB25 data set obtained from Dali Server.....	243
Appendix M.	PDB50 data set obtained from Dali Server.....	245
Appendix N.	PDB90 data set obtained from Dali Server.....	247
Appendix O.	Structural dendrogram of the PDB90 data set.....	249
Appendix P.	Correspondence analysis showing the principal groups from the PDB90 data set.....	250
Appendix Q.	Superposition of the representative structures for comparison against MCR.	251
Appendix R.	<i>A. baumannii</i> ScPmrC DNA and Amino Acid Sequence.	252
Appendix S.	<i>A. baumannii</i> ScPmrC properties.	253
Appendix T.	Phosphoethanolamine Transferases amino acid sequences used for <i>A. baumannii</i> PmrC Multi Sequence Alignment (MSA) analysis	254
Appendix U.	ScPmrC protein identification by mass fingerprint analysis.	257
Appendix V.	ScPmrC Mass fingerprint analysis of peptide modifications.....	258
Appendix W.	Size Exclusion Chromatography S75 column void volume estimation using Blue Dextran 2000.	260
Appendix X.	Size Exclusion Chromatography S200 column void volume estimation using Blue Dextran 2000.....	261

List of Figures:

Figure 1. Timeline of the last 110 years showing key elements on the development of antibiotics/antibiotic therapy and antibiotic resistance emergence.....	3
Figure 2. Structure of Gram-Negative Bacterial cellular envelope.....	7
Figure 3: Chemical structures of polymyxins a) Polymyxin B and b) Polymyxin E, also known as colistin.....	11
Figure 4: Covalent modifications of lipid A.....	13
Figure 5: Pathways of LPS-modifying mechanisms in Gram-Negative bacteria.....	15
Figure 6: Proposed Phosphoethanolamine Transfer Reaction.....	18
Figure 7: Presence of <i>mcr-1</i> genes on Chinese <i>E. coli</i> isolates, from poultry.....	19
Figure 8: Interactive transmission network of pathogens harbouring antibiotic resistance genes.....	20
Figure 9: Global spread of <i>mcr</i> variants from humans, animals and food.....	23
Figure 10: 3D structure of the MCR-1.....	25
Figure 11: Structure comparison of the catalytic site of the MCR-1 and EptA.....	27
Figure 12: Chemical synthesis of 4-Nitrophenol-PETN.....	50
Figure 13: Structures of the catalytic domains of MCR-1 and Alkaline Phosphatase.....	58
Figure 14: Characteristic $\alpha/\beta/\alpha$ fold of the alkaline phosphatase superfamily and MCR-1.....	64
Figure 15: Structural dendrogram of the MCR-1 catalytic domain and 63 closest matches from the PDB25 dataset calculated by Dali.....	68
Figure 16: DALI Correspondence analysis of the MCR-1 catalytic domain and 63 closest matches from the PDB25 dataset.....	69

Figure 17: Z-score distribution of the 72 structures identified from a DALI search of the PDB50 data set against the MCR-1 catalytic domain (PDB 5lrn).	71
<i>Figure 18: Heat map analysis of DALI pairwise comparison of alkaline phosphatase superfamily proteins.</i>	76
Figure 19: Structure-based dendrogram of alkaline phosphatase superfamily proteins based upon DALI pairwise comparisons.	77
Figure 20: DALI Correspondence analysis of alkaline phosphatase superfamily proteins showing the principal groups belonging to the different families members of the alkaline phosphatase superfamily.	78
<i>Figure 21: Structures of representative proteins from alkaline phosphatase subfamilies.</i>	80
<i>Figure 22: Catalytic sites of the 8 representative structures of alkaline phosphatase subfamilies.</i>	81
Figure 23: DALI alignment of structurally matching regions of alkaline phosphatase superfamily proteins	87
Figure 24: Sequence-based phylogenetic tree of alkaline phosphatase superfamily proteins. ..	89
Figure 25: Alignment of phosphoethanolamine transferase catalytic domain sequences.	90
Figure 26: Recombinant MCR-1 production in E. coli Lemo (DE3).	100
Figure 27: Recombinant MCR-1 production comparison between different culture media.	102
Figure 28: Recombinant MCR-1 production comparison between different culture media and bacterial strains.	103
Figure 29: MCR-1 expression in the presence of detergents DDM and LMNG.	104
Figure 30: SDS-PAGE gel of the scale-up culture for MCR-1 production in 2x YT media.	107

Figure 31: SDS-PAGE comparing the production of MCR-1 and its truncated form in 2x YT, 32 YT, and 2xYT Supplemented media.	108
Figure 32: Bacterial biomass yield in different culture media compositions.	109
Figure 33: SDS-PAGE of the SEC purification process of MCR-1.	110
Figure 34: Chromatogram showing the elution profile by Size Exclusion Chromatography of MCR-1 (1st peak) and its truncated form tMCR-1 (2nd peak) in the presence of DDM.	111
Figure 35: Mass spectrometry analysis of MCR-1 and tMCR-1.	113
Figure 36: MCR-1 and tMCR-1 MS analysis.	115
Figure 37: SEC-MALS profile of BSA MCR-1 purified in DDM micelles and Molar mass estimation.	120
Figure 38: Circular Dichroism spectrum of MCR-1.	122
Figure 39: Location of the truncation site of MCR-1.	125
Figure 40: Reaction catalysed by MCR-1.	128
Figure 41: SDS-PAGE comparing the small scale production of a) MCR-1 and b) MCR-2 catalytic domains 2.	132
Figure 42: Large scale production, recovery and purification of cMCR-1 and cMCR-2.	134
Figure 43: Elution profile by Size Exclusion Chromatography of cMCR-1 and cMCR-2.	135
Figure 44: Possible reactions occurring in the assay.	136
Figure 45: pH Dependence of non-enzyme catalysed Hydrolysis of 1 mM pNP-PEtN.	137
Figure 46: Reaction of pNP-PEtN at different concentrations of the soluble domain cMCR-2.	139
Figure 47: LCMS analysis of pNP-PEtN, pNP and PEtN.	140

Figure 48: LCMS analysis of the hydrolysis of pNP-PEtN catalysed by cMCR-2.....	142
Figure 49:pNP standard curve.	143
Figure 50: Reaction profile of varying concentrations of pNP-PEtN with the soluble domain of MCR-2 (50 μ M).....	147
Figure 51: Estimation of the kinetic parameters for cMCR-2 activity.....	148
Figure 52: Thioglycolic acid as an MCR-1 inhibitor candidate.	149
Figure 53: Effect of Thioglycolic acid (TGA) and EDTA on activity of cMCR-2.....	150
Figure 54: Global reports of colistin resistance in <i>A. baumannii</i>	160
Figure 55: Lipid-A modification mediated by the pmrCAB operon.	162
Figure 56: Model of ScPmrC and structure of cMCR-1.....	164
Figure 57: Phylogenetic tree and similarity matrix of several phosphoethanolamine transferases from gram-negative bacteria.	167
Figure 58: a) LcPmrC and b) ScPmrC PCR products.	169
Figure 59: Attempted recovery of ScPmrC from <i>E. coli</i> BL21 (DE3).	171
Figure 60: Expression trials of recombinant ScPmrC.	172
Figure 61: Large scale production and recovery of ScPmrC.....	175
Figure 62: Size Exclusion Chromatography of ScPmrC.....	177
Figure 63: SDS-PAGE showing the elution (E) and concentration (C) steps of a 6-His-tagged ScPmrC.....	178
Figure 64: CD data from ScPmrC.....	181
Figure 65: Cleavage of pNP-PEtN by the recombinant catalytic domain of ScPmrC	182

Figure 66: Reaction of pNP-PEtN at varying concentrations with the ScPmrC recombinant catalytic domain (50 μM).....	183
Figure 67: Estimation of the kinetic parameters for ScPmrC activity by fitting to a Michaelis-Menten model.....	184
Figure 68: ScPmrC protein crystallisation trials.	186
Figure 69: ScPmrC crystals.....	188

List of Tables:

Table 1: Antibiotic classes, mechanism of action and structural characteristics.	4
Table 2: <i>E. coli</i> strains, plasmids and antibiotics used in this study.....	32
Table 2: Primers used to amplify the catalytic domain of PmrC.	34
Table 3: PCR Thermocycling conditions.	34
Table 5: List and composition of the buffers used for recovery and purification of the different recombinant proteins produced.....	39
Table 6: Growth media compositions.	41
Table 7: Biomass production for the L-rhamnose optimization trials.	100
Table 8: BSA three detectors (UV, dRI, and LS) values used for column constant K determination.	118
Table 9: MCR-1 three detectors (UV, dRI, and LS) values used for molar mass estimation.....	118
Table 10: cMCR-2 Mass fingerprint analysis and peptide modifications.....	145
Table 11: Mass fingerprinting analysis and sequence search against the Uniprot <i>E. coli</i> BL21-DE3 database and the ScPmrC sequence.....	173

List of Abbreviations.

AlkP	Alkaline Phosphatase Superfamily
AS	Arylsulfatases
BSA	Bovine Serum Albumin
CAI	Community-Acquired Infections
CAMPs	Cationic Antimicrobial Peptides
CAP	Community-Acquired Pneumonia
CD	Circular Dichroism
CDC	American Centre For Disease Control And Prevention
cMCR	Catalytic Domain Only MCR
CR-AB	Carbapenem-Resistant A. Baumannii
CRE	Carbapenem-Resistant Enterobacteria
Dab	A, Γ -Diamino Butyric Acid.
DDM	N-Dodecyl-B-D-Maltoside
DPA	Dipicolinic Acid
dRI	Differential Refractive Index
EcBcsG	E. Coli Bacterial Cellulose Synthase Subunit G
ECDC	European Centre For Disease Control And Prevention
EDTA	Ethylenediaminetetraacetic Acid
ESBLs	Extended-Spectrum B-Lactamases
ESKAPE	Enterococcus Faecium, Staphylococcus Aureus, Klebsiella Pneumoniae, Acinetobacter Baumannii, Pseudomonas Aeruginosa, And Enterobacter.
GNB	Gram-Negative Bacteria
HPLC	High-Performance Liquid Chromatography
ICU	Intensive Care Units (ICU).
IM	Inner Membrane
IPTG	Isopropyl B- D-1-Thiogalactopyranoside
KDa	Kilo Daltons.
LB	Lysogeny Broth
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LNMG	Lauryl Maltose Neopentyl Glycol
LPS	Lipopolysaccharide
LS	Light Scattering
MALDI-ToF	Matrix-Assisted Laser Desorption/Ionization-Time of Flying.
MALLS	Multiple Angle Light Scattering Detector

MCR	Mobile Colistin Resistance
MDR	Multidrug-Resistant.
MP	Membrane Proteins
MRSA	Methicillin-Resistant Staphylococcus Aureus.
MS/MS	Tandem Mass Spectrometry.
MuGSI	Program And Multi-Site Gram-Negative Surveillance Initiative
Ni ⁺ -NTA	Nickel-Charged Affinity Resin.
NMR	Nuclear Magnetic Resonance
NPP	Nucleotide Pyrophosphatase/Phosphodiesterase
OD600	Optical Density At 600 Nm
OM	Outer Membrane
PDB	Protein Data Bank
PDC	Protein-Detergent Complex
PDR	Pan-Drug Resistant
PEtN	Phosphoethanolamine
pNP	4-Nitrophenol
pNP-PEtN	4-Nitrophenol-Phosphoethanolamine
RPM	Revolutions Per Minute.
ScPmrC / LcPmrC	Short Construct / Long Construct Of A. Baumannii's PmrC
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SOC	Super Optimal Broth With Catabolite Repression
TCS	Two-Component System
TGA	Thioglycolic Acid
TLC	Thin-Layer Chromatography
T _m	Melting Temperature
tMCR-1	Truncated Mobile Colistin Resistance -1
UVAbs280	Absorbance At 280 nm
WHO	World Health Organization

Chapter Outline:

Chapter 1: Introduction.

Explains the emergence of multi drug resistant Gram negative bacteria and the problem this represents, as well as a brief review on the importance of the antibiotic colistin and the polymyxin family of antibiotics as a “last line of defence”; introduces the several resistance mechanisms against this class of antibiotic developed by Gram-negative bacteria, and finally reviews the recently discovered mobile colistin resistance mechanism MCR-1, its homologues and its wide spread across the globe.

Chapter 2: Materials and Methods.

This chapter enlists the bioinformatical tools and the methodology used for the development of this work.

Chapter 3: Structure-based phylogenetic study of the alkaline phosphatase superfamily.

This chapter combines structure and sequence-based phylogenetic analysis of MCR-1 and the alkaline phosphatase superfamily, focusing on the catalytic site architecture and its metal content.

Chapter 4: Production of Recombinant MCR-1.

This chapter presents efforts towards the optimization of recombinant production and purification, and characterisation, of full-length MCR-1.

Chapter 5: Development of an *in vitro* biochemical assay to assess MCR-1 activity.

This chapter presents the development and validation of an *in vitro* activity assay for phosphoethanolamine transferase activity by the MCR catalytic domain to provide a platform to better understand this resistance mechanism.

Chapter 6: Production and characterisation of the MCR homologue PmrC from *Acinetobacter baumannii*.

This chapter presents the recombinant production, biophysical characterisation and *in vitro* activity assay of another colistin resistance mechanism, PmrC from the bacterium *Acinetobacter baumannii*

Chapter 1 General Introduction.

1.1 Antibiotics and Antibiotic Resistance.

The discovery and clinical use of antibiotics drastically changed the way bacterial infections were treated, allowing clinicians to change from a passive approach, focused on diagnosis, to a more aggressive one; actively treating diseases [1,2]. The introduction of antibiotics transformed medicine, reducing the risk of common infectious diseases and becoming an essential aid for many medical procedures [3]; e.g. reducing the incidence of post-surgical infections by 20 fold [4]. Modern antimicrobial therapy began in the early 20th century with the discovery of Salvarsan, the first antimicrobial compound and a safe and effective therapeutic against syphilis [5]. After the discovery of penicillin in 1929 [6] and its introduction in the United Kingdom, the United States of America and other countries, what is known as the “golden age” of antibiotic discovery started (Figure 1). Nearly half of the antibiotics; penicillins, chloramphenicol, polymyxins, tetracyclines and macrolides among others; that have now become the basis for the treatment of bacterial infections were discovered between the late 1940s and 1960 [7]. These were isolated mainly from soil organisms (Table 1), thanks to the developed of a screening platform for antibiotic-producing organisms, the Waksman platform, that allowed the isolation and screening of metabolites produced by the complex soil bacteria actinomycetes [8]. These natural products served as the basis for the generation of semi-synthetic and many fully synthetic antibiotics.

Semi-synthetic antimicrobials are based on modification of existing molecular backbones, obtained by fermentation as an evolutionary outcome of selective pressures e.g. competition between actinomycetes and other bacteria [5]. The semi-synthetic production of new antimicrobials began with the modification of streptomycin into dihydrostreptomycin [5]. Another example of a family of semi-synthetic antimicrobials is the beta-lactams; the determination of the relationship of the beta-lactam ring and antimicrobial activity of penicillin

facilitated the semi-synthetic production of a range of derivatives such as semi-synthetic penicillins and semi-synthetic cephalosporins [9,10]. The chemical manipulation of antimicrobial scaffolds has also permitted the generation of fully synthetic antimicrobials such as chloramphenicol and fully synthetic beta-lactam subclasses such as carbapenems and monobactams.

The “golden age” of antibiotics, however, did not last long as, within two decades, antibiotic resistance became a noticeable problem for clinicians and scientists alike [8] and it has been present ever since.

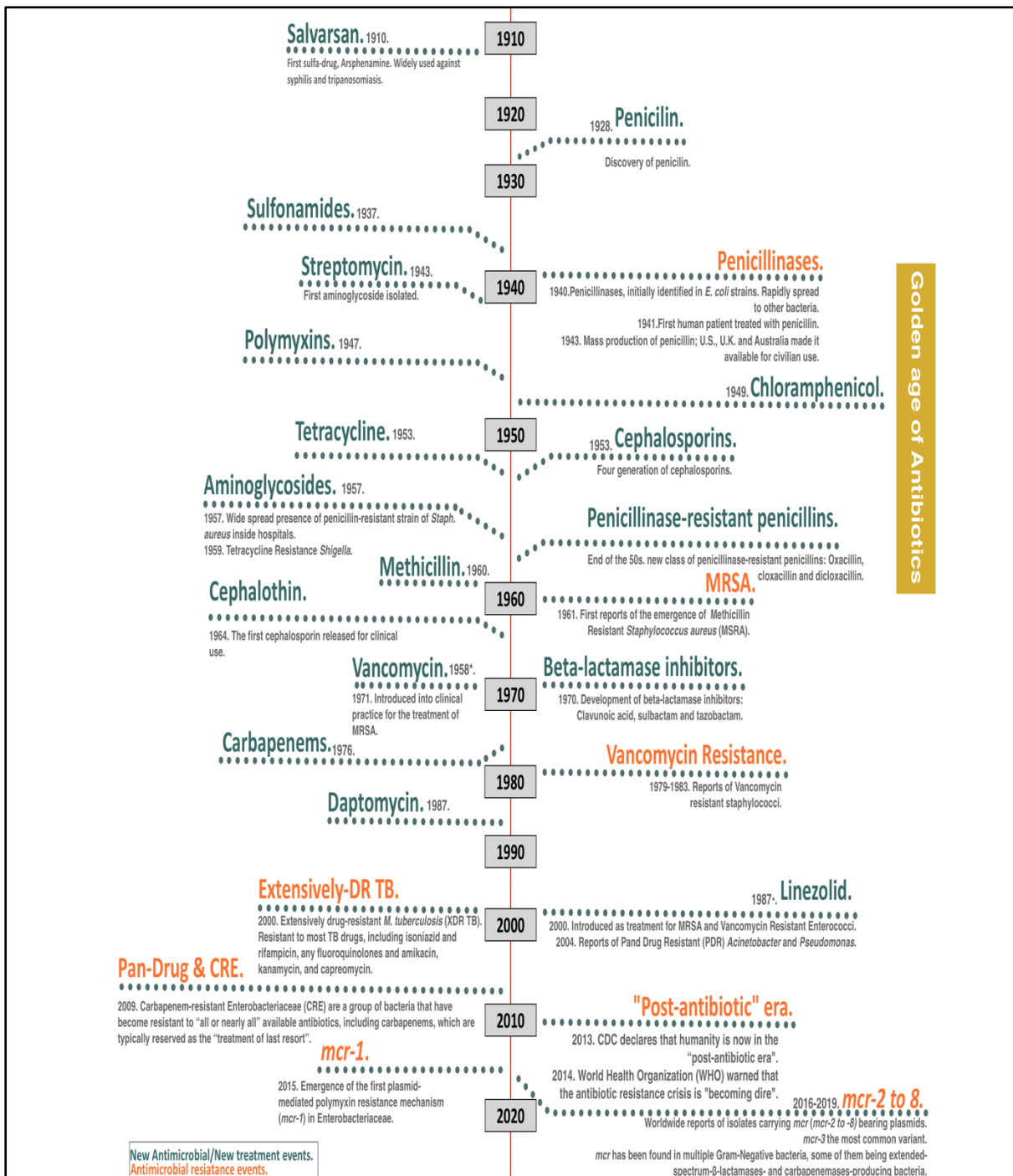
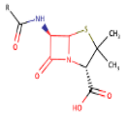
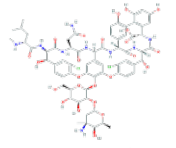
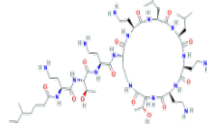
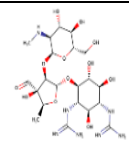
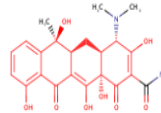
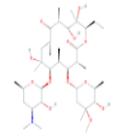
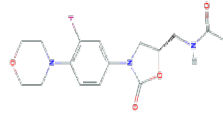
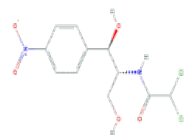
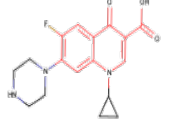
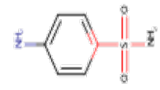


Figure 1. Timeline of the last 110 years showing key elements on the development of antibiotics/antibiotic therapy and antibiotic resistance emergence.

The first half of the timeline (1910-1960) is brimming with the discovery and development of new ways to treat infectious diseases whereas the second half (1970-2020) shows a clear increase of antimicrobial resistance events as well as a slow, almost non-existent, development of new antibiotics. Organizations like the World Health Organization have stated that we are now in a "Post-antibiotic" era [9]. *Vancomycin was developed in 1958 but introduced for clinical use later in 1971. Linezolid was developed in 1987 and was not introduced into clinical use until 2000 [1–3,6,7,13,14].

Table 1: Antibiotic classes, mechanism of action and structural characteristics.

Antibiotic Class	Basis of Mechanism of Action	Mechanism	Examples	Structural characteristic	Structure	Origin	Ref.
Beta-lactams		Inhibition of peptide cross-link formation during bacterial cell wall biosynthesis.	Penicillins (e.g. ampicillin, amoxicillin), third generation cephalosporins (e.g. ceftazidime, cefotaxime, ceftriaxone).	All compounds contain a beta-lactamic ring.		Penicillin isolated from <i>Penicillium</i> spp then later gave birth to the semi-synthetic class of penicillins. Cephalosporins identified in <i>Cephalosporium acremonium</i> (cephalosporin C) gave birth to the semi-synthetic cephalosporins class. Fully synthetic beta-lactams include carbapenems and monobactams.	
Glycopeptides	Antibiotics targeting cell wall	Inhibition of peptidoglycan synthesis by binding to D-alanyl-D-alanine portion of the peptidoglycan precursor subunit. The large size of the molecule prevents binding of D-alanyl subunit with the PBP, inhibiting cell wall synthesis.	Vancomycin (shown), teicoplanin and oritavancin.	Carbohydrate linked to a peptide chain of amino acids.		Vancomycin was discovered in the organism <i>Amycolatopsis orientalis</i> .	
Polymyxins		Interaction with the LPS in the Outer membrane of Gram-Negative bacteria, increasing bacterial membrane permeability leading to leakage of bacterial contents, however its mechanism remains largely unclear.	Colistin (shown) and Polymyxin B.	Cationic oligopeptides containing a cyclic heptapeptide with a tripeptide side chain acylated at the N terminus by a fatty acid tail.		Produced by the organism <i>Paenibacillus polymyxa</i> .	
Aminoglycosides		Inhibit the 30s ribosome subunit by targeting the A site, causing misreading and premature termination of translation of mRNA.	Streptomycin (shown) and kanamycin	All compounds contain amino sugar substructures		Streptomycin obtained from <i>Streptomyces griseus</i> , Semi-synthetic.	
Tetracyclines		Inhibit the 30s ribosome subunit by interacting with the 30S ribosomal subunit to prevent binding of t-RNA.	Tetracycline (shown), doxycycline and lymecycline	All members contain 4 adjacent cyclic hydrocarbon rings		Semi-synthetic derivatives of the natural product of <i>Streptomyces aureofaciens</i> .	
Macrolides	Inhibition of protein synthesis	Inhibitors of 50S subunit, targeting the conserved sequences of the peptidyl transferase centre of the 23S r-RNA of the 50S ribosomal subunit	Erythromycin (shown) and azithromycin.	All compounds contain a 14, 15 or 16-membered macrolide ring.		Isolated from the organism <i>Saccharopolyspora erythraea</i> .	
Oxazolidinones		Inhibition of protein synthesis by binding to 50s ribosome subunit	Linezolid	All compounds contain 2-oxazolidone in their structure		Natural product, produced by synthetic means.	
Chloramphenicol		Inhibitor of the 50s ribosome subunit, preventing binding of t-RNA to the A site of the ribosome, thus inhibiting protein synthesis.	Chloramphenicol	Distinct compound		Fully synthetic antibiotic originated from a natural product from <i>Streptomyces venezuelae</i> .	
Quinolones	Interfere with bacteria DNA replication	Disruption of DNA synthesis by interfering with type II topoisomerases DNA gyrase and topoisomerase IV during cell replication	Ciprofloxacin (shown) and levofloxacin	All members contain fused aromatic rings attached to a carboxylic acid motif		Synthetic.	
Sulphonamides	Inhibitors of folic acid metabolism.	Inhibition dihydropteroate synthase, limits bacterial synthesis of the B vitamin folate, preventing bacterial growth and multiplication.	Sulphanilamide (shown), sulfadiazine and sulfisoxazole	Sulphonamide group		Synthetic.	

[5,15-19]

1.1.1 The emergence of multi-drug resistant Gram-Negative bacteria.

Antibiotic-resistant bacterial strains have been a problem since 1940 [20], even before penicillin was introduced as a reliable way to treat bacterial infections, however, constant antibiotic overuse has accelerated the increase of drug-resistant bacteria. In response to penicillinase-producing bacteria methicillin, a penicillinase-resistant β -lactam antibiotic was developed [1]. Not so long after the introduction of methicillin into medical treatment, methicillin-resistant strains of *Staphylococcus aureus* (MRSA) were found invading hospitals (Figure 1) [21]. Vancomycin was re-introduced to treat MRSA as it was thought that vancomycin resistance was almost impossible to happen in clinical settings [22].

Third-generation cephalosporins (i.e. ceftriaxone, cefotaxime, and ceftazidime) were designed specifically to combat common β -lactamases, a defence mechanism among penicillin-resistant organisms, however, a few years after their introduction nosocomial organisms, such as *Klebsiella pneumoniae*, were found to develop mutated versions of β -lactamases, extended-spectrum β -lactamases (ESBLs), making them resistant to this class of antibiotics [23]. Resistance to third-generation cephalosporins in Enterobacteriaceae is commonly associated with ESBLs [24]. The genes responsible for encoding ESBLs, such as TEM, SHV and CTX-M; are often found in plasmids that harbour other genes that confer bacteria with resistance against other antimicrobial agents, such as aminoglycosides and sulphonamides. This means that nosocomial strains of ESBL-producing Enterobacteriaceae are, commonly, multi-drug resistant; making treatment of nosocomial infections particularly challenging.

The beginning of the 21st century marked a turning point in humanity's fight against bacterial infections. By the early 2000's one-fifth of the *K. pneumoniae*, one-third of *Enterobacter spp* and one-twentieth of the *Escherichia coli* isolates, from intensive care unit (ICU) patients, in the United States, were resistant to third-generation cephalosporins [25]. Increasing reports of ESBL-harboring bacteria, Pan-drug resistant (PDR) and Carbapenem-resistant (CRE) Enterobacterales, and extensively drug-resistant bacteria, such as *Mycobacterium*

tuberculosis (Figure 1) and the decreasing effectiveness of antibiotics have opened the way to what is now considered the “last-resort antibiotics” [26] and the “post-antibiotic” era [12,27].

1.1.1.1 Multi-Drug resistant Gram-Negative bacteria and nosocomial pathogens.

The overuse of antibiotics in modern life (hospitals, agriculture, food industry, etc.) over the past seven decades has contributed to maintaining a selective pressure allowing resistant bacteria to survive and spread [28]; just in the United Kingdom alone, in 2017, a total of 773 tonnes of antibiotics were used; out of these 773 tonnes, 491 tonnes were used in human patients and 282 for animal treatment [29]. The increase of infections caused by MDR bacteria and the emergence of Gram-negative bacteria (GNB), such as *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*, resistant to most antibiotics including carbapenems (frequently drugs of last resort for such organisms) [30], is now a worldwide health issue, turning nosocomial infections into a serious threat especially in countries where certain antibiotics are not available [31]. While nosocomial drug-resistant pathogens are the main concern for developed countries, low and middle-income countries, such as Mexico, face additional challenges dealing with MDR organisms like *Mycobacterium tuberculosis* [32], pneumococci, *Shigella flexneri*, *Vibrio cholerae* and *Salmonella* [21].

Thanks to this phenomenon, the use of “last resort” antibiotics, like carbapenems and polymyxins, has increased in the last decade [33]. However, due to the lack of progress in the discovery of new drugs against multidrug-resistant bacteria, the abuse of antibiotics and the rapid increase of carbapenem-resistant bacterial strains, the use of the polymyxin antibiotic colistin has been re-evaluated as an effective way to treat multidrug-resistant Gram-negative infections [34,35].

1.2 The outer membrane of Gram-negative bacteria and Lipid A.

Gram-negative bacteria have the distinctive characteristic of possessing a cellular envelope that consists of an inner membrane (IM), a bilayer made out of phospholipids, a mesh-like layer formed by peptidoglycan and an asymmetrical outer membrane (OM) composed of

phospholipids on the inner leaflet and lipopolysaccharides (LPS) on the outer leaflet (Figure 2). This OM serves as a filter of sorts, possessing selective properties that protect the bacteria from harmful compounds in the environment, such as antibiotics [36]. Both membranes (IM and OM) possess integral proteins embedded in them; however, the structures of these proteins varies between these two membranes. Inner membrane proteins normally consist of hydrophobic α -helices anchoring them to the membrane while OM proteins tend to consist of antiparallel β -barrel cylinders that facilitate the transport of nutrients (an example of these are porins as shown in Figure 2) [37].

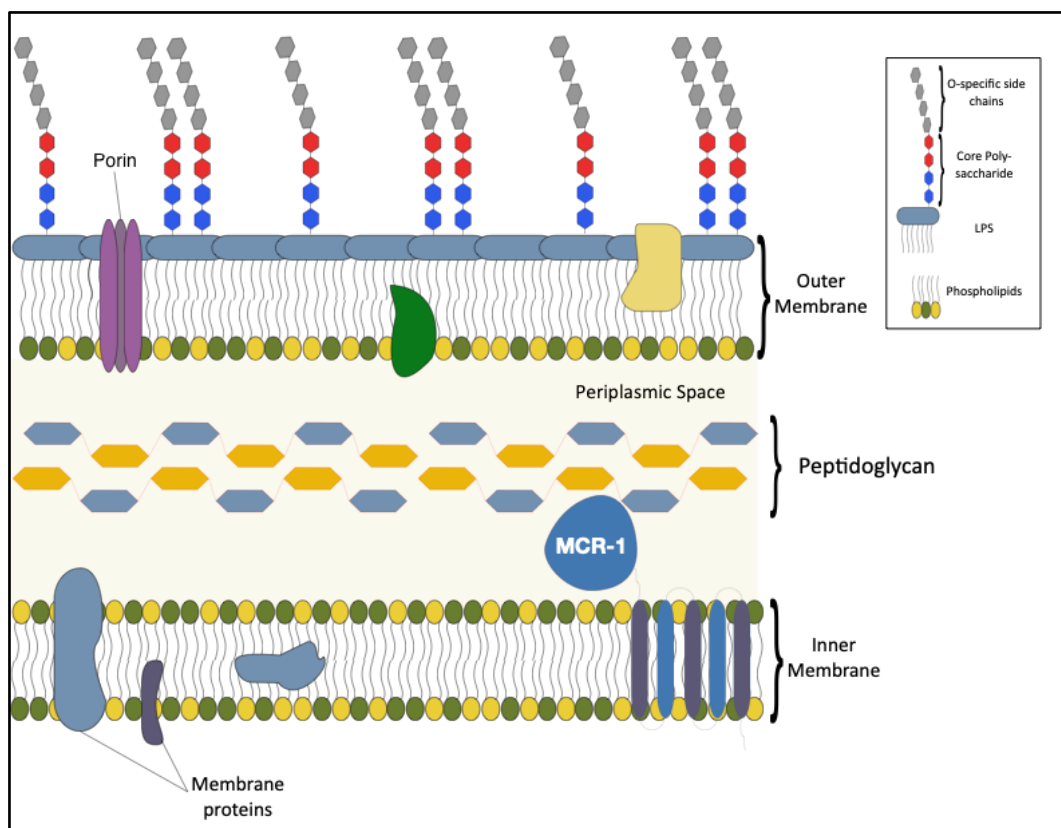


Figure 2. Structure of Gram-Negative Bacterial cellular envelope.

Schematic representation of the cell envelope in Gram-Negative bacteria.

1.2.1 Lipopolysaccharide and Lipid A.

Lipopolysaccharide (LPS) is found solely in the OM where it carries out various biological functions, such as acting as a semi-permeable barrier; given its strong amphipathic nature, making GNB innately resistant to many antimicrobial compounds, as well as playing an

important role in bacterial-host interactions and bacterial pathogenicity as LPS is the primary bacterial component interacting with the host immune system [38]; this last characteristic causes a toxic response in the host and because of this, LPS is sometimes referred as endotoxin, in reference to cell-associated toxicity observable in infections cause by gram-negative pathogens [39]. The amphipathic LPS molecule is formed by an hydrophilic polysaccharide chain and a highly conserved hydrophobic portion known as Lipid A, serving as an anchorage to the OM [40]. Lipid A consists of a β -1,6-linked d-glucosamine disaccharide carrying ester and amide-linked 3-hydroxy fatty acids at the 2, 3, 2', and 3' positions and phosphate groups at the 1' and 4' positions; differing from typical phospholipids by having six saturated fatty acid chains instead of two [36,41]. Lipid A is often glycosylated with an oligosaccharide core that can be 6 to 10 sugars long, and a long chain of repeating oligosaccharides (1 to 40 units with considerable variation between bacterial species), the O-antigenic polysaccharide [42]. The highly conserved Lipid A moiety, and its oligosaccharide core structure, is thought to be driven by its role in keeping the integrity of the semi-permeable outer membrane. These characteristics make the OM negatively charged and more hydrophobic than the IM, effectively working as a permeable barrier which only small hydrophobic drugs and membrane permeabilizers can penetrate. Bacteria can modify the basic Kdo–Lipid A structure as a response to changes in their environment; such as exposure to low pH, changes in the concentration of divalent cations such as Zn^{2+} , Fe^{3+} or as a response to enhance the outer membrane integrity for survival [43], however these modifications can signify major changes in the membrane properties, its interaction with antibiotics and bacterial virulence, given the substantial contribution of LPS to bacterial virulence [44].

1.3 Cationic peptides and LPS/outer membrane targeting antibiotics.

Cationic antimicrobial peptides (CAMPs) are a class of antimicrobials found to be part of the innate immune responses of several species including bacteria, fungi, plants and animals. Natural occurring CAMPs are usually 12 to 50 amino acids long and, despite their structural

variation, most of them share two unique traits: they are positively charged molecules, being polycationic; and can form amphipathic structures due to possessing both hydrophobic and hydrophilic domains [45,46]. These unique traits are essential for their function as antimicrobial agents; the positive charge facilitates interaction with the negative charge of membrane phospholipid head groups and the hydrophobicity allows for the insertion of the peptides into the cellular membranes, causing a physical disruption of the membranes [47] and allowing both the leakage of cellular material and nutrients and the entry into the cell of antimicrobial compounds. The main mechanism of action of CAMPs against Gram-negative bacteria is the permeabilization and disruption of the OM, driven by the electrostatic attraction between the negatively charged OM [48,49] of GNB and the positive charged CAMPs.

1.4 Polymyxins.

Polymyxins are a family of five CAMPs; polymyxin A, B, C, D and E (also known as colistin) was discovered more than six decades ago. They received their name from the organism in which they were discovered, *Bacillus polymyxa*. Their antimicrobial spectrum includes GNB such as *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *K. pneumoniae*. Out of these five polymyxins only two have been used in clinical practice: polymyxin B and colistin. Their use has been intermittent due to reports of nephrotoxicity and neurotoxicity [20,36]. However, due to the growing threat of MDR GNB infections they have been recently reintroduced as “last resort” antibiotics [50].

1.4.1 Structure and mechanism of action.

Polymyxins B and E are small antimicrobial lipopeptides formed by a cyclic polypeptide ring and a branched fatty acid tail. The fatty acid tail confers hydrophobicity on the molecule, making it amphipathic and therefore capable of making interactions in both polar and non-polar environments as water or lipid membranes [51]. The main difference between polymyxin B and colistin is a single amino acid in the peptide ring (6th residue, phenylalanine in polymyxin B is changed to leucine in colistin, see Figure 3).

Polymyxins target the outer membrane of GNB; the α,γ -diamino butyric acid (Dab) residue of polymyxins interacts with the phosphate group of lipid A in the cellular membrane, prompting the destabilization of the lipopolysaccharide, and leading to cell death by increasing the permeability of the outer membrane [51,52]. Interaction of polymyxins and LPS can be described as a two-step mechanism: the first step is carried out by the interaction of the charged regions of both compounds and in the second step the interaction is between the non-polar parts [53]. This interaction, between the LPS and polymyxins (B and E), has been described as envelope-like, separating the hydrophobic residues (Figure 3 residues 6 and 7) from the charged Dab residues (4, 5, 8 and 9) [54]. The polymyxin antimicrobial spectrum includes GNB such as *A. baumannii*, *P. aeruginosa*, *K. pneumoniae* [30] and *E. coli*; being, in the case of MDR- GNB, the only available active antibiotic [34].

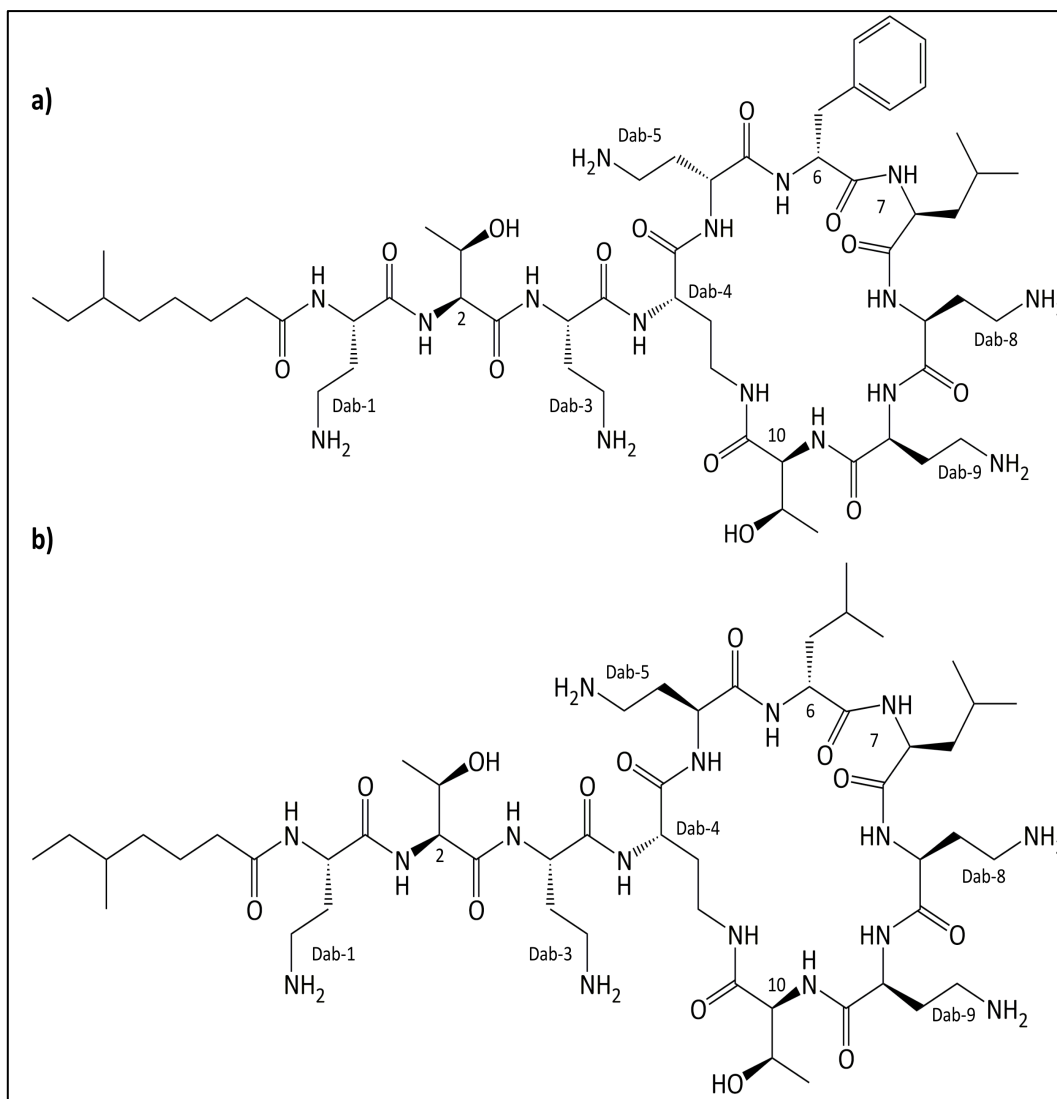


Figure 3: Chemical structures of polymyxins a) Polymyxin B and b) Polymyxin E, also known as colistin.

Polymyxins interact with Lipid A, dividing the hydrophobic residues (6 and 7) from the charged 2,4-diaminobutyric acid ((Dab) residues (4, 5, 8 and 9) [55].

1.4.2 Last resort antibiotic.

The appearance of strains of *P. aeruginosa* and *A. baumannii*, as well as *Enterobacterales*, that are resistant to β -lactams, aminoglycosides, fluoroquinolones and carbapenems, has renewed focus upon polymyxins [26,34,52,56–59]. Polymyxin use in clinical practice was discontinued in most parts of the world as they were regarded as toxic compounds [52,56],

however, this class of antibiotic has been extensively used in the agriculture and food production industry (poultry and livestock) [58,59].

Given the rise of multidrug-resistant Gram-negative bacteria (MDR GNB), its activity against MDR strains of *P. auruginosa*, *Acinetobacter* and *Klebsiella* species, and the lower levels of resistance development in bacteria compared to other antibiotic classes, polymyxins have attracted renewed interest as last-resort treatments for infections caused by MDR GNB [51,62]. The extensive use of antibiotics, including colistin, as a growth promoter in the agricultural industry in China, a country with a dense population, has led to the increase and spread of antibiotic resistance mechanisms, including the recently discovered *mcr-1* gene that confers polymyxin resistance to bacteria [60]. The surge in colistin resistance, including the emergence of a new mobile colistin resistance mechanism, is of clinical concern as it threatens the effectivity of this antibiotic. Understanding know and emerging colistin resistance mechanism is important to prolong colistin use as a last resort antibiotic.

1.4.3 Polymyxin Resistance.

Even though resistance to polymyxins and colistin has been considered rare [63], due to the low use of this class of antibiotics for treatment, polymyxin-resistant bacteria have been identified and the molecular mechanisms for this phenomenon have been characterized. The development of polymyxin resistance involves the modification of the OM in Gram-negative bacteria; more specifically the modification of the Lipid A, the main target of polymyxins, by the addition of an electropositive charge motif thus reducing the affinity between polymyxins and their target.

Resistance is generated by the enzyme-catalysed addition of positively charged residues of phosphoethanolamine (PEtN) or 4-amino-deoxy-*L*-arabinose (L-Ara4-N), or of palmitate, to the LPS in the lipid A moiety [47,52,58,64,65] (Figure 4). The addition of PEtN or L-Ara4-N to the LPS in the lipid A alters the net negative charge in the outer membrane, rendering polymyxins ineffective and are usually encoded chromosomally, except for a newly found PEtN addition mechanism (see 1.5 below).

systems is the EnvZ-OmpR system, found in *E. coli*, that regulates the expression of the porins genes *ompF* and *ompC* as a response to changes in extracellular osmolarity. EnvZ is an osmo-sensing kinase that, under a high osmolarity environment, transfers a phosphoryl group to the response receptor OmpR which in turn regulates the expressions of the porins genes *ompF* and *ompC* genes, regulating their expression [68].

Polymyxin resistance in *Klebsiella spp.* occurs due to mutations in the *phoPQ* and *pmrAB* signal transduction systems. PhoPQ is regulated by the *mgrB* gene product, therefore mutation on this gene, such as sequence insertions in the gene or relevant amino acid substitutions can lead to the inactivation of MgrB [70], thus disrupting the negative feedback it possesses over the PhoQ/PhoP signalling system, which in turn causes the modification of LPS through the upregulation of the *arnBCADTEF* operon, translating into the addition of L-Ara4-N to Lipid A through the action of ArnT (Figure 5).

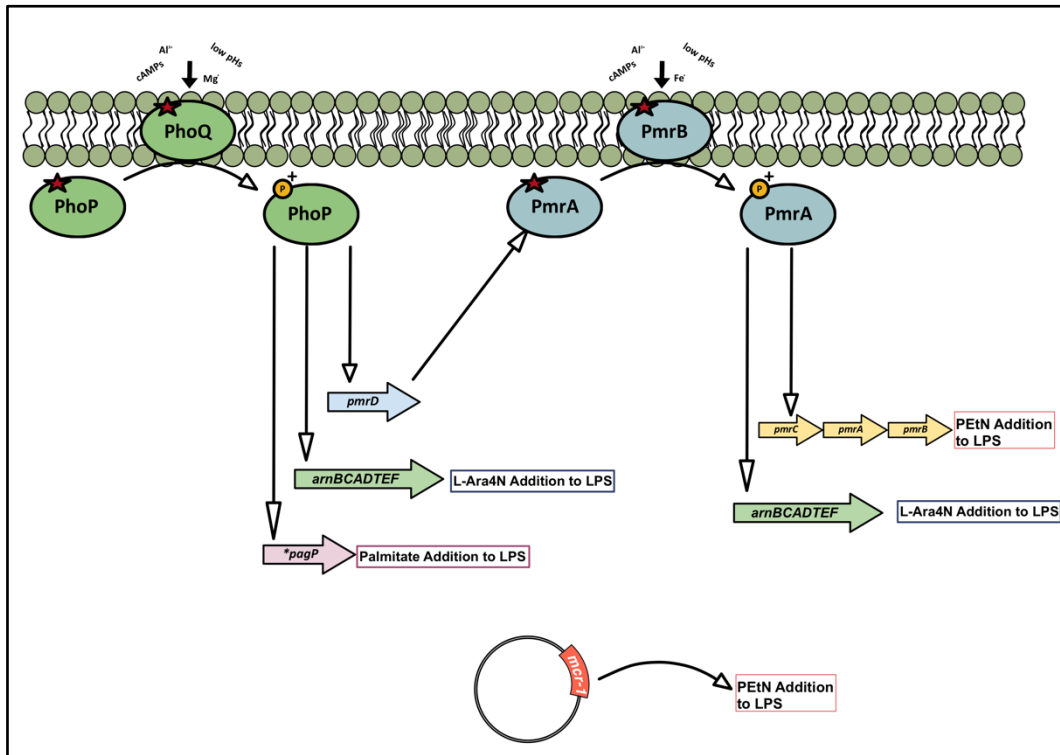


Figure 5: Pathways of LPS-modifying mechanisms in Gram-Negative bacteria.

Activation of the two-component systems (TCSs) PhoP/PhoQ or PmrAB can be induced by cationic compounds such as CAMPs, low Mg^{2+} concentrations, acidic pHs and high Fe^{3+} concentrations. The PhoP/PhoQ system (TCS) activates the transcription of the operon *arnBCADTEF* leading to the production of the L-Ara4N transferase responsible for the addition of the L-aminoarabinose (L-Ara4N) group into lipid A. This two-component system (*phoP/phoQ*) can also activate *pagP* leading to the addition of a palmitate motif to lipid A (this has only been observed in *Salmonella*). During the activation of the PmrAB system, the sensor kinase PmrB activates the regulator protein PmrA by phosphorylation, the active PmrA upregulates the *pmrCAB* leading to the overexpression of the phosphoethanolamine transferase PmrC which modifies Lipid-A by the addition of PEtN. PmrAB is also activated by PhoPQ via the *pmrD*, which in turn activates the operon *pmrCAB* leading to the addition of phosphoethanolamine to lipid A. *mcr-1* is a recently found mobile colistin resistance mechanism that mediates the addition of phosphoethanolamine to lipid A conferring polymyxin resistance to harbouring bacteria [52,67,71–75].

1.4.3.2 Polymyxin resistance generated by palmitoyltransferase activity.

Palmitoyl transferases catalyse the addition of a palmitate residue onto lipid A (Figure 4 PagP in pink). In *Salmonella typhimurium*, the PagP enzyme is under the control of the PhoP/PhoQ system, which is activated by a decrease in the extracellular Mg^{2+} concentration or the presence of cationic peptides. Homologues of this enzyme are coded on the genomes of *Yersinia pestis*, *Bordetella pertussis*, *Bordetella bronchiseptica* and various strains of *Salmonella* [58]. This reaction takes place at the surface of the outer membrane, where PagP uses phospholipids as palmitoyl donors [76].

1.4.3.3 Colistin resistance generated by the addition of phosphoethanolamine (PEtN) to Lipid A.

The addition of a phosphoethanolamine (PEtN) group to the lipid A moiety is carried out by phosphoethanolamine transferases. The addition of a PEtN group to lipid A positions 1' or 4' (Figure 4 in red), is catalysed in the periplasmic domain, using phosphatidylethanolamine as a donor. This mechanism is also regulated by enzymes controlled by the PhoP/PhoQ system that directly (or indirectly) activate the *pmrCAB* operon. The *pmrCAB* operon codes for three proteins: PmrA; a response regulator, PmrB; a sensor kinase protein and the phosphoethanolamine transferase PmrC, responsible for the addition of a phosphoethanolamine (pEtN) motif to Lipid A. High expression levels of the chromosomal PmrC results in colistin resistance in many Gram-negative organisms, such as *K. pneumoniae* and *A. baumannii* [77].

The *crrAB* (colistin resistance regulation) operon is another mechanism thought to be involved in the generation of polymyxin resistance. This operon codes for two proteins, namely the regulatory protein CrrA and the sensor protein kinase CrrB. Although little is known about the physiological role of this operon, it has been noticed that the inactivation of *crrB* promotes the overexpression of the *arnBCADTEF* and *pmrAB* operons leading to the addition of both L-Ara4N and pEtN to lipid A. mechanisms involved in polymyxin resistance [78].

All these are chromosomal mechanisms; however, a new phosphoethanolamine transferase, encoded in a plasmid, named MCR-1 [60] has been recently found. This enzyme is similar to other enzymes of the same family, such as LptA from *N. meningitidis* [79] and EptC from *Campylobacter jejuni* [73,80], but represents the first identification of transferrable polymyxin resistance.

1.5 Mobile Colistin Resistance.

1.5.1 The emergence of MCR-1.

A novel mobile colistin resistance mechanism, *mcr-1*, was discovered in late 2015 [75] This new mobile colistin resistance mechanism works similarly to previously discussed polymyxin resistance mechanisms. By catalysing the addition of a cationic phosphoethanolamine (PEtN) group to Lipid A (Figure 6) it reduces the electronegativity of Lipid A, thus diminishing the efficacy of colistin as an antimicrobial compound by preventing its binding. This is a globally concerning finding, as chromosomal colistin resistance is not generally encountered in *E. coli* and experimental data show that expression of just one gene, *mcr-1*, increases polymyxin MICs by more than 8 times [52] sufficient to confer resistance against polymyxins in *E. coli*.

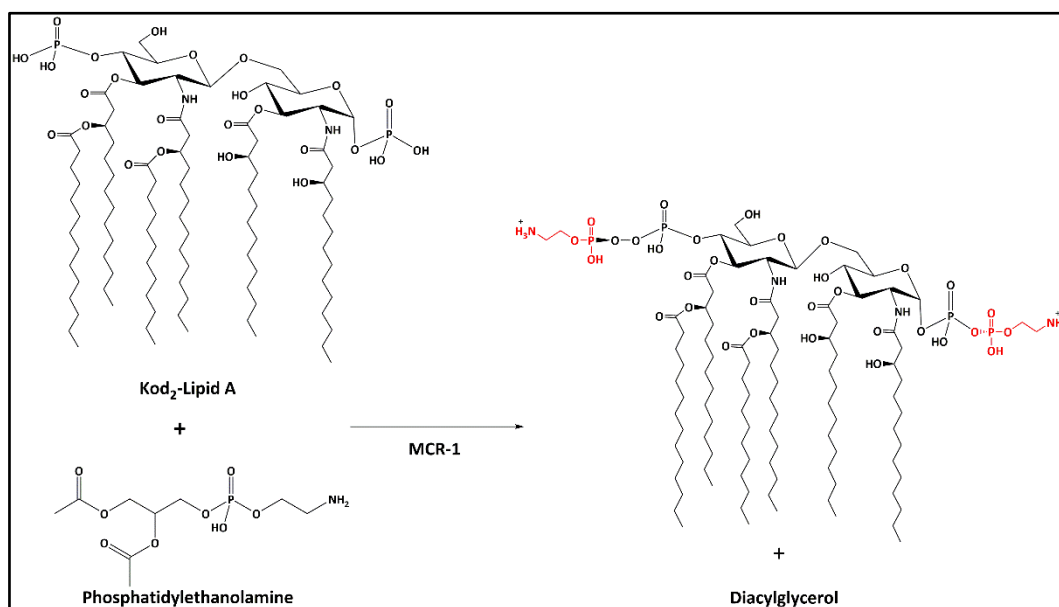


Figure 6: Proposed Phosphoethanolamine Transfer Reaction.

Transfer of a phosphoethanolamine group from a donor (Phosphatidylethanolamine) to lipid A catalysed by MCR-1. The transfer may occur to the phosphate groups at the 1' or the 4' (as shown in red) positions.

Another important aspect of this mechanism is its location on a mobile vector, a plasmid, instead of being the product of chromosomal mutations; thus giving MCR-1 the ability to be transferred between organisms [81] as well as being able to be carried by multiple plasmid backbones [82]. This is the first instance of a transferable polymyxin resistance mechanism; raising the prospect of it spreading across Gram-negative bacteria, including those already harbouring MDR genes. Bacteria form part of an intricate and interactive transmission network that facilitates their movement across different environments, and with them, clusters of resistance genes are harboured inside transferable vectors [7,83] (Figure 8).

Although only recently discovered, evidence suggests that the *mcr-1* gene has been around and silently spreading for more than a couple of decades. In China, *E. coli* isolates dated from the decade of the 1980s tested positive for *mcr-1*; at the same time that the use of colistin in food-producing animals started [84]. Colistin use in food-animal production, in 2012, was 600

times higher than its use for human treatment [85], this is consistent with an increase of *mcr-1* positive isolates found in China (Figure 7).

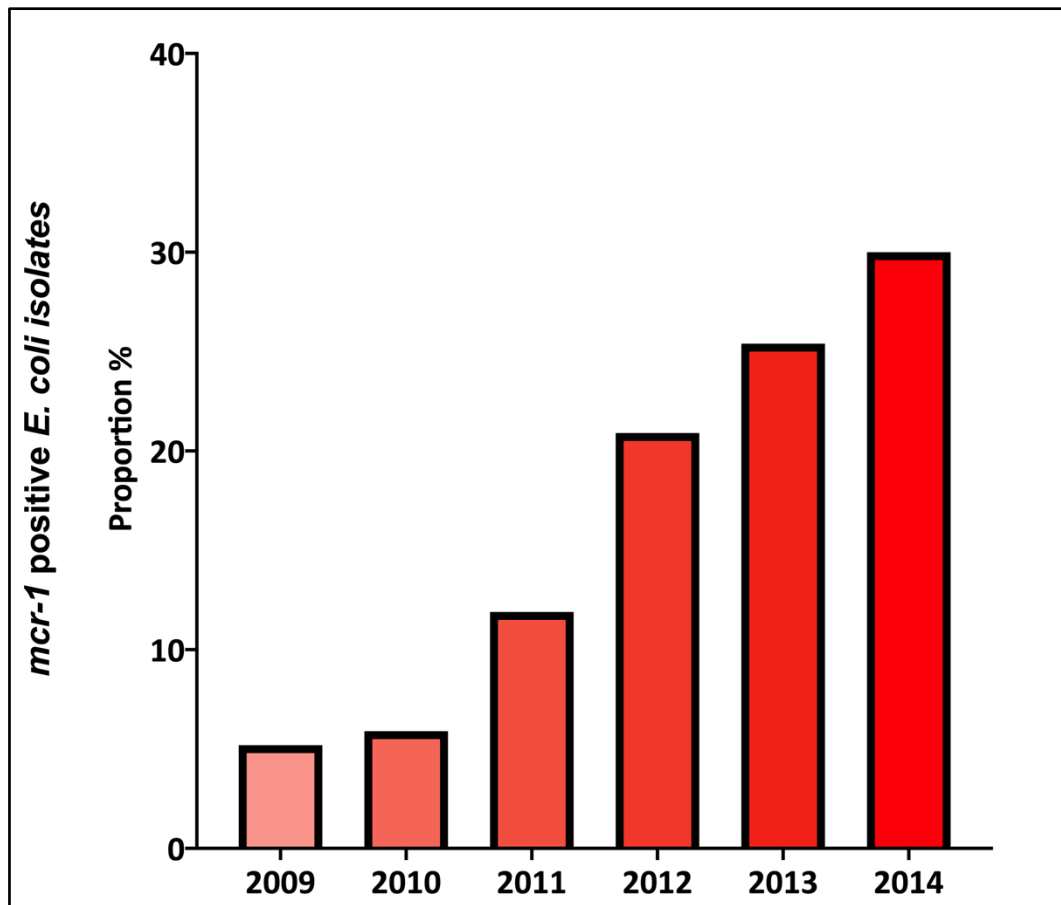


Figure 7: Presence of *mcr-1* genes on Chinese *E. coli* isolates, from poultry.

The proportion of mcr-1 positive E. coli isolates (6 out of 115 isolates from poultry) raised from 5% in 2009 to 30% (15 out of 50 isolates from poultry) in 2014. This increase in the presence of colistin resistance genes among E. coli isolates (from poultry), aligns with the increase of colistin use in food-producing animals in China before its subsequent ban in 2016 [84,86].

1.5.2 MCR-1 dissemination.

Worldwide reports show that *mcr* genes have already widespread across several countries on each continent: China, Japan, Laos, Vietnam, Malaysia, Singapore, Cambodia, Bahrain, Taiwan, Hong Kong, Thailand and South Korea (among several others) in Asia; UK, Germany, Belgium, Spain, France, Switzerland, Sweden, Lithuania and Hungary in the European region; Algeria, Egypt, Tunisia, Morocco and South Africa on the African continent, Australia in Oceania and USA and Canada in North America and Colombia, Argentina and Brazil in South

America [87–90] This widespread of *mcr-1*, across several continents, is thought to be related to its possible animal origins, mainly in cattle and pigs for human consumption; and selection pressure due to the overuse of colistin in veterinary practice [91] and its use as a growth promoter [86]. *mcr-1* and its isoforms have also been identified in bacterial isolates from seagulls and other migratory birds, this is a concerning finding since migratory animals can contribute to the global dissemination of this resistance mechanism (Figure 8) [92,93].

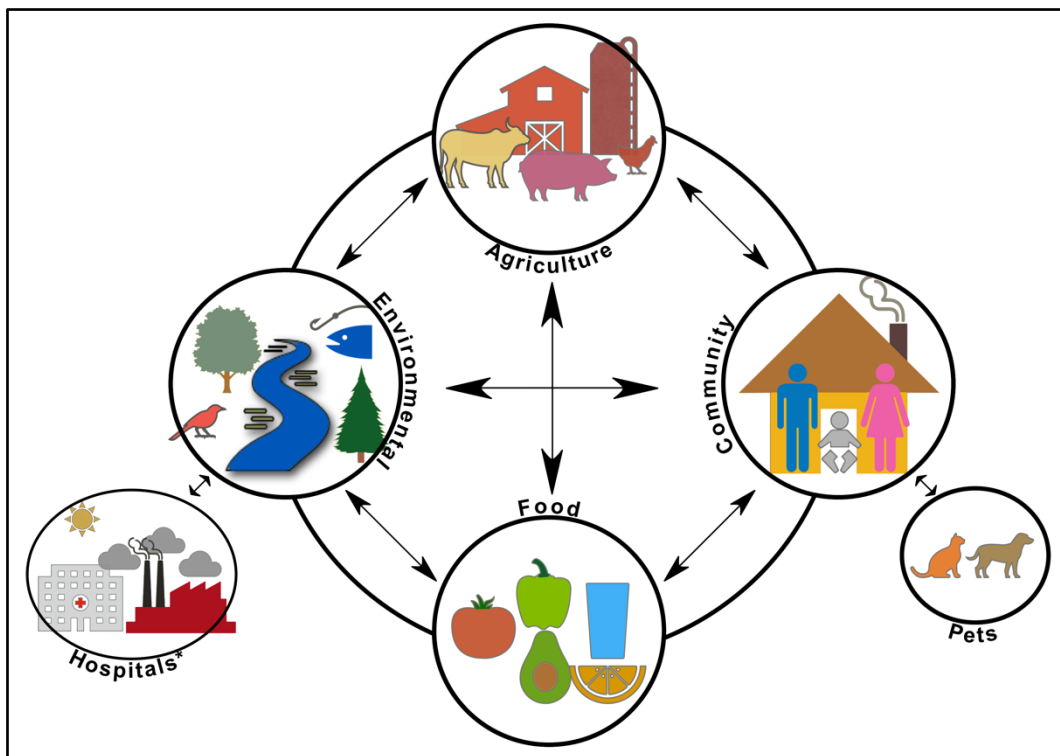


Figure 8: Interactive transmission network of pathogens harbouring antibiotic resistance genes.

Bacteria form an intricate transmission network through various paths; they can be passed between farm produce to food (“farm to table”), from human activity to the environment and indirectly through industrial activity, just to name a few. This bacteria-exchange network becomes important in the dissemination of mobile resistance mechanisms, such as *mcr-1*. *mcr-1* was isolated from farm animals and has spread now to meat, food produce, environment and human sources [83,94].

Alongside its global dissemination, this mechanism has even been reported in a diverse group of *Enterobacterales* such as *E. coli*, *Enterobacter spp.*, *Salmonella spp* [81,95], including MDR strains such as carbapenemase-producing *K. pneumoniae* [85]. MDR *K. pneumoniae* strains are already very worrisome as this is one of the main pathogens responsible for nosocomial infections, being especially difficult to treat due to the extent of antibiotic resistance

phenotypes such as the KPC-KP (*K. pneumoniae* strains producing KPC-type carbapenemases) strains [86] and the high mortality rates associated with them. *mcr-1* - carrying plasmids have been shown to contain other MDR genes providing resistance to other antibiotics like β -lactams, aminoglycosides, quinolones, sulphonamides and tetracyclines [44].

1.5.2.1 MCR-1 and horizontal gene transfer.

Due to its carriage on plasmids, *mcr-1* can be transferred among different bacteria through horizontal gene transmission. By 2020, the *mcr-1* gene has been found in 15 bacterial species: *A. baumannii*, *Citrobacter braakii*, *Citrobacter freundii*, *Cronobacter sakazakii*, *E. coli*, *Enterobacter cloacae*, *Klebsiella aerogenes*, *Kluyvera*, *P. aeruginosa*, *Providencia alcalifaciens*, *K. pneumoniae*, *Klebsiella oxytoca*, *Raoultella planticola*, *Salmonella enterica* and *Shigella sonnei* [98]. Although *mcr-1* has been found in many bacterial species, its presence in clinically relevant species, such as *E. coli* and *Klebsiella*, are the most troublesome as these bacteria are among the most common cause of nosocomial infections. A study of more than 2000 clinical isolates of *E. coli* and *K. pneumoniae* from patients suffering from bloodstream infections in China shows that *mcr-1* is more prevalent in *E. coli* than in *K. pneumoniae*; attributing colistin resistance in *K. pneumoniae* isolates to its intrinsic chromosomal mechanism, the two-component systems *phoP/phoQ* and *pmrA/pmrB* [99].

Although 15 incompatibility type (Inc) plasmids harbouring *mcr-1* have been identified; three plasmids, IncHI2, IncI2, and IncX4, are the dominant *mcr-1*-carrying plasmid types, and present high *in vitro* transfer rates [100,101]. These plasmids play a major role in the global dissemination of *mcr-1* and its emergence and prevalence in MDR gram-negative bacteria, such as Pan drug and CRE; leaving few remaining treatment options to deal with MDR-GNB infections. Even though *mcr-1* is the most common variant, having been detected in over 20 countries globally [102], several other *mcr* variants are now contributing to the increase of colistin resistance in gram-negative bacteria.

1.5.2.2 MCR variants.

In addition to the wide global spread of *mcr-1*, several related genes (ranging from MCR-1, MCR-2 to MCR-9) [103,104] have now been identified in locations across the world. *mcr* variants 1, 2, 3, 4, 5 and 8 have been found in pigs from China, *mcr-1*, 2 and 6 have been identified in pig farms across Great Britain [102]; *mcr-2* has been found in animal sources from Spain and Belgium; *mcr-3* is the most widely reported variant with reports in Thailand, Denmark, France, Brazil and Japan; *mcr-4* has been detected in *E. coli* and *Salmonella enterica* in pig isolates from Italy and Spain and *mcr-5* reported in Colombia, Japan, Spain and Germany. Multiple cases of human bacterial isolates positive for *mcr* genes (*mcr-1*, *mcr-3* and *mcr-4*) have been reported; however, most reports of *mcr* gene carriage belong to animal and food isolates [89,106–108] (Figure 9). A recent phylogenetic study classified the family of *mcr* genes, from *mcr-1* to *mcr-9*, in three main clades, one of which contained *mcr-1*, *mcr-2* and *mcr-6* coupled together, with amino acid sequence similarity above 80% among them, and the second group containing *mcr-3*, *mcr-4*, *mcr-7* and *mcr-9*, and finally *mcr-5* formed a distinct subclade [109].

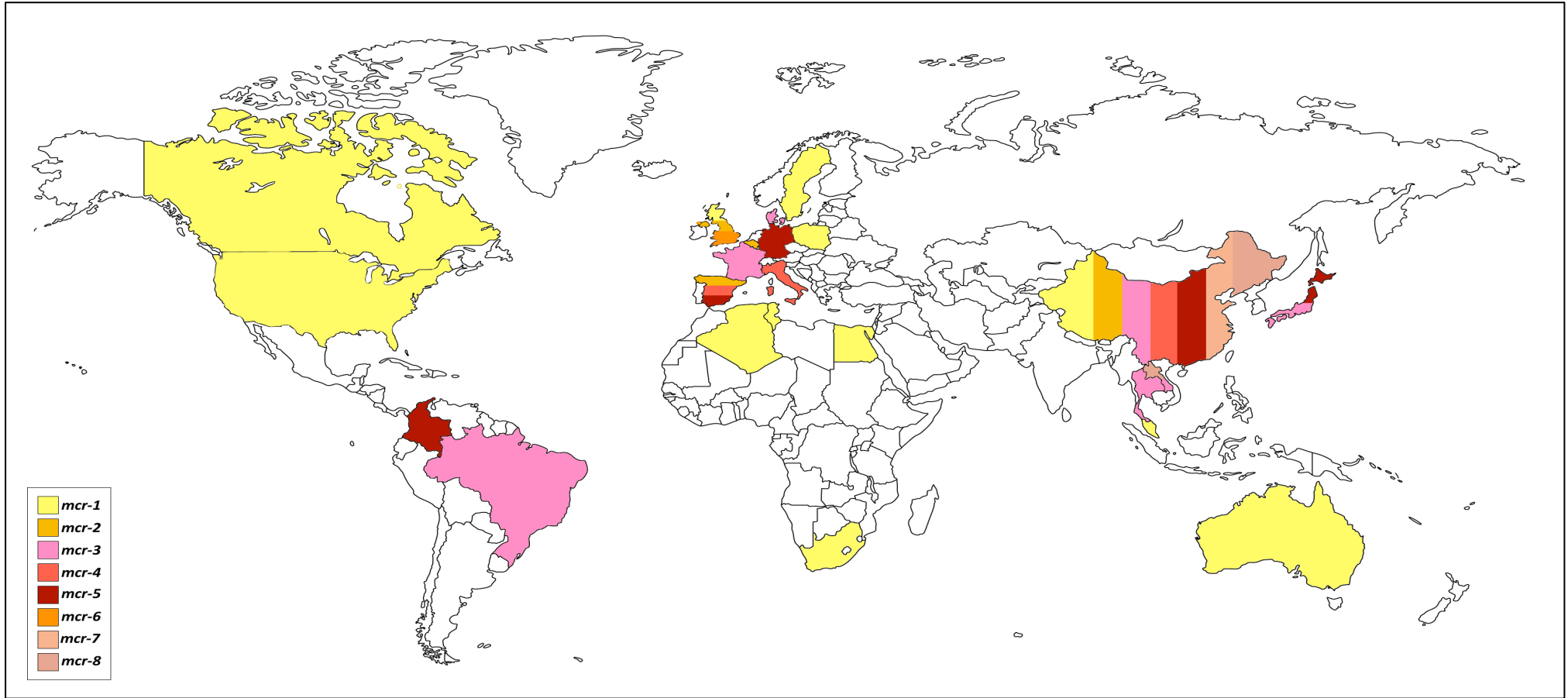


Figure 9: Global spread of mcr variants from humans, animals and food.

All mcr variants, except mcr-6, have been found in China. mcr-2,3 and 4 isolates have been reported in pig farms across the world, mcr-5 has been detected in poultry in Germany. Enterobacteriaceae strains harbouring mcr-1, mcr-3 and mcr-4 have been recovered from human isolates. Antarctica is the only continent where mcr genes have not been detected [103–107,110].

1.6 MCR-1 structure and characteristics.

The amino acid sequence of MCR-1 found in the bacterium *E. coli* exhibits resemblance to those of other characterised phosphoethanolamine transferases, such as EptA (from *Neisseria meningitides*) with 41% identity (Figure 10 d) [111]. These phosphoethanolamine transferases, members of the alkaline phosphatase superfamily (AlkP) [79], catalyse the modification of Lipid A, and show similar structures (Figure 10 c and c). The crystal structure of EptA reveals a two-domain protein with an N-terminal transmembrane portion (5 α -helices), anchoring the protein in the inner membrane, and a catalytic domain facing the periplasmic space [74,111,112], this architecture is shared by MCR-1 (as depicted in Figure 10 a and c). This placement of the catalytic domain, in the periplasmic compartment, points out that the addition of PEtN to Lipid A occurs at the membrane surface, with the transmembrane domain proposed to recruit the phospholipid phosphatidylethanolamine as a PEtN donor.

Although it is thought that MCR-1 is evolutionarily distinct from its homologues, such as EptA [103], structural analysis of the catalytic domain of MCR-1 revealed a zinc-binding pocket, in the catalytic motif, shared between MCR-1 and EptA (Figure 11) and found in other AlkP superfamily members (as described in later chapters, see 3.2.1). This zinc pocket is required for phenotypic resistance to polymyxins. Catalytic metal-binding pockets, as well as acceptor residues (such as Ser, Thr and Cys (often modified to formylglycine in sulfatases [101]) required for hydrolytic or transferase activity are highly conserved among members of the AlkP superfamily. This suggests that members of the AlkP superfamily are likely to follow a similar reaction path [114].

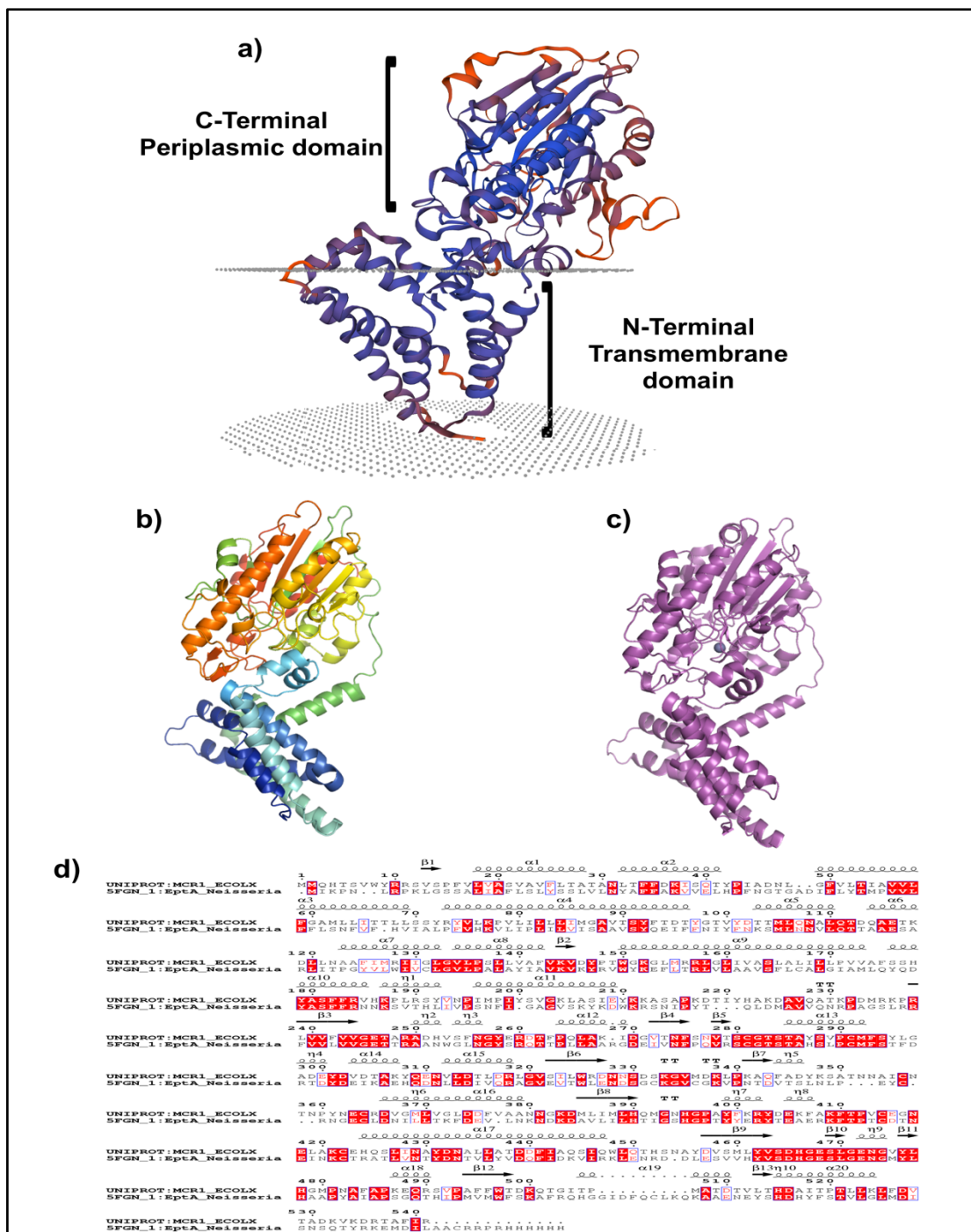


Figure 10: 3D structure of the MCR-1.

a) A 3D model of MCR-1 (obtained from Swiss-Model using the structure of EptA (PDB 5FGN) as a template) showing the N-terminal transmembrane domain and the C-terminal Periplasm facing domain. The inner membrane is represented by the grey mesh. b) 3D representation of MCR-1. c) 3D structure of *Neisseria meningitidis* EptA (PDB entry 5FGN, magenta). Zinc is represented as a single silver sphere. d) Sequence alignment of MCR-1 from *E. coli* and EptA from *N. meningitidis*; amino acids highlighted by a red box indicate strict identity, red characters indicate similarity, secondary structure analysis was based on the structure of EptA (PDB 5FGN).

Although amino acid sequence comparisons suggest that many of the active-site residues are conserved among members of the alkaline phosphatases (see below, section 3.1.1), for MCR-1 the number of zinc ions and their position in the active site varies in different experimental structures published to date (Figure 11) [64,104,105]. Members of the alkaline phosphatase superfamily are known to require the presence of one, two and in some cases up to three metal ions to carry out their catalytic activity [117–120]. The mechanism of action of phosphoethanolamine transferases has not been fully elucidated [72,77,108], with as yet few detailed mechanistic proposals for MCR, and these based on molecular dynamics and density functional theory simulations [121] rather than experimental work. This project aims to extend current knowledge of phosphoethanolamine transferases associated with colistin resistance by investigating the structural relationship between MCR-1 and the wider alkaline phosphatase superfamily, producing recombinant full-length MCR-1 from the bacterium *E. coli*, developing a biochemical assay to investigate the activity of the catalytic domain and that is suitable for use in assaying its inhibition by small molecules, and finally investigating the properties of an MCR-1 homologue, the chromosomally encoded phosphoethanolamine transferase involved in colistin resistance in the Gram-negative pathogen *Acinetobacter baumannii*, PmrC.

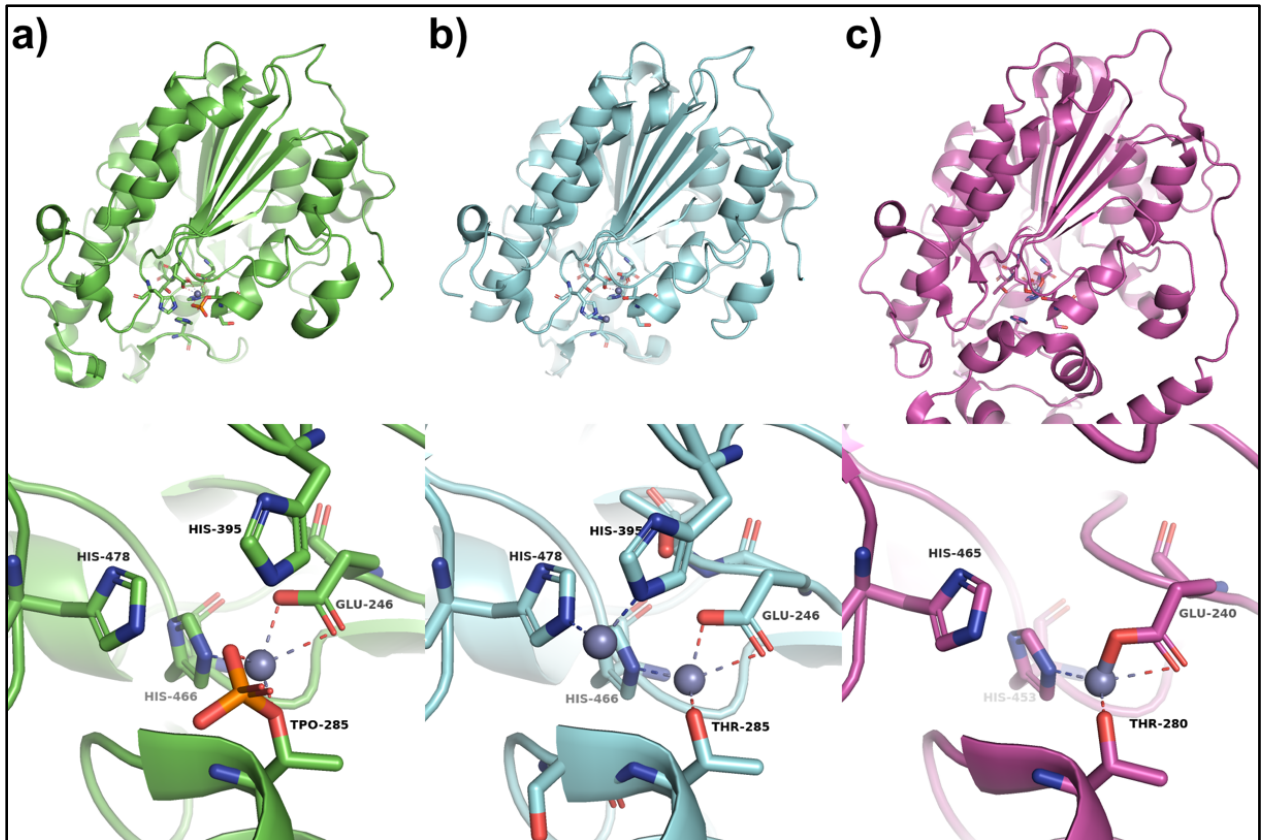


Figure 11: Structure comparison of the catalytic site of the MCR-1 and EptA.

(a) Catalytic domain of MCR-1 (*E. coli* PDB code 5LRN) containing 1 zinc in its catalytic site, (b) Catalytic domain of MCR-1 (*E. coli* PDB code 5LRM) containing 2 zinc ions in its catalytic site and (c) Catalytic domain of EptA (*N. meningitides* PDB code 5FGN) containing 1 zinc in its catalytic site. Top panels show the similarities in the fold around the active site, the active threonine (285 in MCR-1 and 280 in EptA) is phosphorylated in the MCR-1 structure containing 1 zinc ion. Active residues have been labelled according to the PDB numeration. Bottom panels show the places of the zinc ions (metallic grey spheres) and their interaction (dotted lines) [74,122].

1.7 Aims of the project:

To shed light on the recently discovered mobile colistin resistance mechanism *mcr-1* and its homologues. The objectives for each chapter are as follows:

- To understand the place of MCR-1 among the members of the alkaline phosphatase superfamily and to shed light on the architecture of the MCR-1 catalytic site and the presence of a potential secondary metal ion binding site in the catalytic pocket, by employing a structure and sequence based analysis.
- The development of a production strategy, through bacterial strain screening and growth media enhancement, to obtain the transmembrane MCR-1 in sufficient quantities for its biophysical characterization.
- The development of an enzymatic assay, through the synthetic production of a chromophoric reporter used in conjunction with a “catalytic domain only” construct of MCR-1 and its homologue MCR-2, that serves as a platform to investigate the active mechanism of MCR-1 and other phosphoethanolamine transferases and that facilitates inhibitor screening.
- To recombinantly produce the catalytic domain of another colistin resistance mechanism; the phosphoethanolamine transferase PmrC from *A. baumannii*. Looking to shed more light on its structure and the mechanism of action of this colistin resistance mechanism from an important Gram-negative pathogen.

Chapter 2 Materials and Methods.

2.1 Bioinformatics and structural modelling

2.1.1 MCR-1 and ScPmrC 3D modelling.

For the elaboration of a 3D model of the full-length MCR-1 and the catalytic domain of PmrC, here named ScPmrC, the webserver of Swiss-Model [123,124] was used. The structure of the Lipooligosaccharide phosphoethanolamine transferase A (EptA), from *Neisseria meningitidis*, was used as a template (PDB entry 5FGN [122]) for the full-length MCR-1 model. To generate a 3D model of ScPmrC, both, the catalytic domain of MCR-1 (PDB entry 5LRN [74]) and EptA (PDB entry 5FGN) were used as templates. Secondary structure analysis and visualisation were performed using the online webserver Visual-Stride (<http://webclu.bio.wzw.tum.de/stride/>) [125] and the software Pymol for MacOS (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.), respectively.

2.1.2 Multiple Sequence Analysis of known Phosphoethanolamine transferases.

Multisequence Analysis and percent similarity index of the sequences of one known phosphoethanolamine transferases of *A. baumannii* (PMRC_ABAU, accession number ATD51479.1), six from *E. coli* (EPTA_ECOLI, EPTB_ECOLI, EPTC_ECOLI, YHBX_ECOLI, MCR1_ECOLI AND OPGE_ECOLI, accession numbers UNIPROT:EPTA_ECOLI EptA, UNIPROT:EPTB_ECOLI P37661, UNIPROT:EPTC_ECOLI P0CB39, UNIPROT:YHBX_ECOLI P42640, UNIPROT:MCR1_ECOLX A0A0R6L508 and UNIPROT:OPGE_ECOLI P75785 respectively), one *Helicobacter pylori* (EPTA_HELPY, accession number UNIPROT:EPTA_HELPY O24867), one from *Haemophilus influenzae* (Y1005_HAEIN, accession number UNIPROT:Y1005_HAEIN P44974), one from *K. pneumonia* (PMRC_KPNE, accession number CDK37739.1) and *S. typhi* (EPTB_SALTY,

accession number UNIPROT:CPTA_SALTY Q7CPC0), were performed using the Clustal Omega web tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [126]. The similarity percentage indexes within the 13 phosphoethanolamine transferases were used to generate similarity matrices using GraphPad Prism version 8.2.1 for MacOS (GraphPad Software, San Diego, California USA). The sequences used for MSA analysis can be found in Appendix T

2.1.3 MCR-1 structural classification .

The PDB entry for the solved 3D structure of the catalytic domain of MCR-1, 5LRN [74], was used for the structural analysis and comparison of MCR-1 in relationship with the alkaline phosphatase superfamily. According to the CATH Protein Structure Classification database MCR-1 appear listed as a member of the Alkaline phosphatase superfamily, alongside many other phosphoethanolamine transferases, among 420 domains. An superposition of MCR-1 against these 420 domains was generated by generated by the 2DProts database available through the CAT webpage (<https://www.cathdb.info/>). The Structural Classification of Proteins, SCOP, database (<https://scop.mrc-lmb.cam.ac.uk>) entry for the Alkaline phosphatase superfamily (SCOP ID 3001353) showed nine families associated to the Alkaline phosphatase-like structural domain.

2.1.4 Protein Sequence analysis.

Sequence analysis was performed using the Clustal Omega Multiple Sequence Alignment (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [126]. 3D structure superpositions and active site architectures were performed and analysed using Pymol for MacOS (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.). Information regarding the metal content, as well as ligands, in the catalytic site was extracted from the respective PDB entries from the RCSB database (<https://www.rcsb.org>) [127,128]. Sequence and structure diversity analysis was obtained from the CATH protein family database (<http://www.cathdb.info>) [129]

and the SCOP database (<https://scop.mrc-lmb.cam.ac.uk>) entry for the Alkaline phosphatase superfamily (SCOP ID 3001353). Analysis of the conserved metal binding sites was performed using ESPRIPT (<https://espript.ibcp.fr>) [130].

2.1.5 Structural similarity database generation.

A database comparison for structurally similar neighbours, was generated using the DALI server (<http://ekhidna2.biocenter.helsinki.fi/dali/>) [131]. Dali server is an intuitive and user friendly piece of software, when compared to other available tools [132,133]. The MCR-1 structure 5LRM [74] was used as query against all structures deposited in the Protein Data Bank using the “PDB search” server function that compares the query structures against the entirety of the Protein Data Bank [127,128]. After obtaining more than 600 structural hits, the data sets generated automatically by Dali server, PDB25, PDB50 and PDB90, were used for the structural analysis using the “all-against-all” function. Only significant matches from each data set were selected for “all-against-all” as this function is restricted to 64 structures per analysis.

2.1.6 Structural Data Analysis using Dali.

“All against all” pair-wise structural comparisons were performed for all possible pairs of structural matches, with a Z-score of 2 and above, from the PDB25, PDB50 and PDB90 datasets obtained from Dali Server. Visualization of heatmap matrices and correspondence analysis (CA) was performed using GraphPad Prism version 8.2.1 for MacOS (GraphPad Software, San Diego, California USA). Structural similarity trees were performed using Dali Server.

2.2 Experimental Methods.

All the buffers and reagents were purchased from Sigma-Aldrich (Sigma Aldrich, UK) or ThermoFisher Scientific (Thermo Fisher Scientific, USA) as analytical grade reagents, unless otherwise stated.

Table 2: *E. coli* strains, plasmids and antibiotics used in this study.

Strain	Relevant Characteristics	Supplier
C43(DE3)	Expression strain carrying plasmid <i>mcr1</i> -pET24a	New England Biolabs, UK
C43(DE3)	Expression strain carrying plasmid <i>mcr1</i> -pET28b	
C41(DE3)	Expression strain carrying plasmid <i>mcr1</i> -pET24a	
C41(DE3)	Expression strain carrying plasmid <i>mcr1</i> -pET28b	
Lemo21(DE3)	Expression strain carrying plasmids <i>mcr1</i> -pET24a and pLemo	
Lemo21(DE3)	Expression strain carrying plasmids <i>mcr1</i> -pET28b and pLemo	
SoluBL21(DE3)	Expression strain carrying plasmid <i>cmcr-1</i> -pOPINF	Genlantis, USA
SoluBL21(DE3)	Expression strain carrying plasmid <i>cmcr-2</i> -pOPINF	
BL21(DE3)	Expression strain carrying plasmid <i>ScPmrC</i> -pOPINF	New England Biolabs, UK
BL21 Star (DE3)	Expression strain carrying plasmid <i>ScPmrC</i> -pOPINF	
Rosetta 2 (DE3)	Expression strain carrying plasmid <i>ScPmrC</i> -pOPINF	
ShuffleT7 Express	Expression strain carrying plasmid <i>ScPmrC</i> -pOPINF	
ArcticExpress (DE3)	Expression strain carrying plasmid <i>ScPmrC</i> -pOPINF	Agilent, UK
Plasmids	Selection antibiotic	Concentration [µg/mL]
<i>mcr1</i> -pET24a <i>mcr1</i> -pET28b	Kanamycin	50
pLemo	Chloramphenicol	30
<i>cmcr-1</i> -pOPINF <i>cmcr-2</i> -pOPINF <i>ScPmrC</i> -pOPINF	Ampicillin	50

2.2.1 Expression Plasmids and gene cloning.

Genes for *mcr-1*, *mcr-2*, *cmcr-1* and *cmcr-2* were provided subcloned into T7 expression plasmids: *mcr1*-pET24a, *mcr1*-pET28b, *cmcr-1*-pOPINF and *cmcr-2*-pOPINF, respectively, by Dr. Philip Hinchliffe. The *A. baumannii pmrC* gene sequence was synthesised as reported by Beceiro *et al* (GenBank accession number HM134921.1) [134]. Two constructs, a “short” (ScPmrC, amino acids 203-538) and a “long” (LcPmrC, amino acids 198-538) construct,

encoding the C-terminal periplasmic domain from *A. baumannii* PmrC were generated by PCR from a codon-optimized synthetic gene (Eurofins, sequence given in Appendix R), *pmrC*, using the primers (Eurofins, UK) given in Table 3 (bold segments correspond to regions complementing the *pmrC* sequence). 50 μ L PCR reactions were set up for each construct, the final primer concentration was 0.3 μ M for each primer, the PmrC template was added to a final concentration of 10 ng per μ L and the KOD Hot Start DNA Polymerase (1.0 U/ μ L) (Merck Millipore Novagen, Germany) was used at a final concentration of 0.02 U/ μ L. The PCR was carried out in a PCRmax thermocycler (PCRmax alpha, UK) with the thermocycling conditions as given in Table 4.

Table 3: Primers used to amplify the catalytic domain of PmrC.

Primer Name	Forward Primer	T _m [°C]
Forward Primer Short Construct "LcPmrC"	5' AAGTTCTGTTTCAGGGCCCG GTTATCTATGGTCAGGATGCACATC 3'	57
Forward Primer Long Construct "ScPmrC"	5' AAGTTCTGTTTCAGGGCCCG AAGAACTTGCCGCTGGTTATC 3'	58
Reverse primer	5' ATGGTCTAGAAAGCTTTA GTTACATGAGCGCAGCTG 3'	60

Table 4: PCR Thermocycling conditions.

Step	Temperature [°C]	Time [s]	Cycles
Polymerase activation	95	120	
Denaturation	95	30	
Annealing	57.0/58.0*	30	35 cycles
Extension	70	32	
Final extension	70	300	
*The annealing temperatures were set to be the lowest primer T _m			

The size of both amplified fragments was analysed and confirmed by agarose gel electrophoresis with ethidium bromide staining. A 1% agarose gel was generated by dissolving 1 g of Agarose (Sigma) in 100 mL of TAE buffer. The agarose-buffer solution was microwaved, at an intensity of 800 Watts, for 2 minutes at intervals of 30 seconds until agarose was fully dissolved. 3 µL of Ethidium bromide, to a final concentration of 0.5 µg/mL, were added to the dissolved agarose. Samples were prepared by mixing 10 µL of PCR reaction with 2 µL of 6X loading dye. 10 µL of sample were then loaded onto the 1% agarose gel and the gel electrophoresis was run for 45 minutes at 100 volts.

Both amplicons were cloned, using the InFusion method [135]: Linearized pOPINF vector was prepared by incubating 5 µg of plasmid with 50 units of both restriction enzymes, KpnI and HindIII; to a total reaction volume of 100 µL. The reaction was incubated at 37°C for approximately 4 hours. 5 µL of the purified PCR products of each construct, short and long, were mixed with 100 ng of purified linearized pOPINF vector and 2 µL of 5X In-Fusion HD

enzyme premix, distilled sterile water was added to reach a final reaction volume of 10 μ L. The reaction was incubated at 50°C for 15 mins. 2.5 μ L of the InFusion reaction mixture were added to 100 μ L of competent *E. coli* BL21(DE3) (Genlantis, USA) cells and incubated on ice for 30 minutes. The InFusion reaction-cell mixture was then heat-shocked at 42°C for 45 seconds and then incubated on ice for 5 minutes. 250 μ L of Super optimal broth with catabolite repression (SOC) were added to the transformed cells. The samples were incubated at 37°C and 180 RPM shaking for 1 hour and then plated on Lysogeny Broth (LB) agar containing ampicillin (50 μ g/mL), as the selection method, and incubated overnight. 5 colonies were recovered from each construct and grown overnight in 5 mL of LB medium supplemented with 50 μ g/mL of ampicillin. The pOPINF plasmids containing the short and long constructs (*ScPmrC*-pOPINF and *LcPmrC*- pOPINF, respectively) were recovered using a QIAprep Spin Miniprep Kit (QIAGEN N.V., Hilden, DE) and concentrated in distilled water to a final concentration of 50 μ g/mL. The final products were verified by Sanger sequencing using the service provided by Source Genomics (Source Bioscience, Nottingham, UK) using the primers listed in Table 3.

A heat-shock protocol was followed to transform the competent from *E. coli* strains with the respective plasmids, as shown in Table 2. To start the process, 50 ng of the respective plasmids were added to 100 μ L of calcium competent *E. coli* cells and incubated on ice for 30 minutes. The plasmid-cell mixtures were then heat shocked by heating them at 42°C for 45 seconds and then incubated on ice for 5 minutes. 250 μ L of Super optimal broth with catabolite repression (SOC) were added to the heat-shocked plasmid-cell mixture. The samples were incubated at 37°C with shaking at 180 RPM for 1 hour and then plated on Luria-Bertani (LB) agar containing the appropriate antibiotics as listed in Table 2.

2.2.2 Small expression trials.

After overnight incubation at 37° C, a single colony from each transformation was isolated and grown in 10 mL of LB medium with the appropriate antibiotics (Table 2) for 8 hours at 37° C and 180 RPM shaking. The overnight cultures were used to start the small-scale expression of MCR-1, cMCR-1, cMCR-2 and ScPmrC respectively. The resulting cultures were diluted using LB medium, with the with the appropriate antibiotics, to a final optical density at 600 nm (OD_{600}) of approximately 0.1 and the final volume was adjusted to 100 mL of culture. The cultures were then incubated in a shaking incubator at 37°C, 30°C for strains ArcticExpress and ShuffleT7, and 180 RPM until they reached an OD_{600} of 0.8 and MCR-1 expression was induced by adding Isopropyl β - d-1-thiogalactopyranoside (IPTG) to a final concentration of 400 μ M; additionally, cultures using *E. coli* strain Lemo21(DE3) were prepared with the following L-rhamnose concentrations: 0, 100, 250, 500, 1,000, and 2,000 μ M. After IPTG addition, the temperature was reduced to 18°C and 16°C for strains ArcticExpress and ShuffleT7 for 18 hours. Whole cell samples were recovered by centrifugation at 4000 RPMs for 10 minutes (ACL, Mod: PK 121R). Samples prior to the induction (T_0) and at harvesting time (T_f) were collected, by centrifugation at 5000 RPM and 4°C for 5 min (HERMLE Z 207 M, Labortechnik GmbH, Germany), diluted to a final OD_{600} of 0.1 and stored for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis.

2.2.2.1 Small scale purification of MCR-1.

The harvested MCR-1 producing cells recovered from small scale expression experiments, were resuspended in 30 mL lysis buffer (Table 5) with a single tablet of EDTA-free protease inhibitor (Roche) and broken down by three passages through a cell disruptor at 30 000 psi and 5°C (CF1, Cell Disrupter Constant Systems LTD, UK). Unbroken cells and insoluble debris were removed by centrifugation at 15,000 RPM and 4°C for 20 mins (Avanti J-26 XP, rotor JA-25.50, Beckman Coulter, Ireland). The supernatant from MCR-1 samples was recovered and

the biological membranes were pelleted by ultracentrifugation at 150,000 RPM and 4°C for 2 hrs (Optima L-80 XP ultracentrifuge, rotor 70-Ti, Beckman Coulter, Ireland). The pelleted membranes were re-suspended in 30 mL of membrane solubilization buffer (Table 5) and incubated in a shaker for 1 hour at 4°C. The solubilized membranes were recovered by ultracentrifugation at 150,000 RPM and 4°C for 45 minutes (Optima L-80 XP ultracentrifuge, rotor 70-Ti, Beckman Coulter, Ireland). 100 µL samples of the solubilized membranes and the supernatant were taken for SDS PAGE analysis and the rest of the solution was used for further purification of MCR-1 by Ni⁺-NTA affinity chromatography. The pellet was discarded and the supernatant stored for further experimentation.

0.4 mL of Ni⁺-NTA bead slurry (Qiagen) were washed in 20 mL of distilled water and recovered by centrifugation at 800 RPM for 5 mins (ACL, Mod: PK 121R). The water was removed and the supernatant from the high-speed ultracentrifugation was incubated with the Ni⁺ beads. Imidazole was added, to a final concentration of 5 mM, to the supernatant-Ni⁺ solution and left to incubate for 2 hours at 4°C. The supernatant-Ni⁺ solution was run down gravity-flow columns and the flow-through was collected for SDS-PAGE analysis. The Ni⁺-NTA beads were washed with 10 bed volumes of Washing step 1 buffer and then washed for the second time in 10 bed volumes of Washing step 2 buffer to remove any unbound proteins; 100 µL samples of both washing steps were recovered for SDS PAGE analysis. The bound MCR-1 was eluted in 5 bed volumes using Ni-NTA Elution buffer. 50 µL samples were collected for SDS PAGE analysis.

2.2.2.2 Small scale purification of cMCR-1, cMCR-2 and ScPmrC.

The harvested cells recovered from small scale expression experiments, were resuspended in 30 mL lysis buffer with a single tablet of EDTA-free protease inhibitor (Roche) and broken down by three passages through a cell disruptor at 30 000 psi and 5°C (CF1, Cell Disrupter Constant Systems LTD, UK). The unbroken cells and insoluble debris were removed by

ultracentrifugation at $100,000 \times g$ at 4°C for 1 h (Optima L-80 XP ultracentrifuge, rotor 70-Ti, Beckman Coulter, Ireland). The supernatants were recovered, imidazole was added to a final concentration of 10 mM and incubated for 2 hrs with 0.2 mL of Ni-NTA resin (Qiagen); Ni⁺-NTA resin was previously washed in 20 mL of distilled water and recovered by centrifuging for 10 minutes at 800 RPM and 4°C . The protein-bound resin was then washed with 5 mL of washing buffer 1 and then a second washing step using Washing step 2 buffer (according to Table 5); ScPmrC containing samples were washed a third time using Washing step 3 buffer.

Proteins were then eluted from the Ni⁺-NTA beads using elution buffer (composition in Table 5) in fractions of 500 μL . Imidazole was reduced to a final concentration of 10 mM by buffer exchange using an Amicon 10-kDa centrifugal filter. Samples were taken and analysed by SDS-PAGE. The His-tags, of both cMCR-1 and cMCR-2, were removed by adding 300 μL of recombinant 3 C protease (3 mg/mL), kindly provided by Dr Philip Hinchliffe, to 30 mL of purified protein and left to cleavage overnight. The removed His-tag, and residual un-cleaved protein, was captured on 1 mL of Ni-NTA resin slurry. All steps were performed in a cold room at 4°C . The purified proteins were subsequently injected onto a Superdex 75 column (Superdex S75 16/60, GE Healthcare, bed volume 120 mL) equilibrated overnight using 2 bed volumes of SEC purification buffer (Table 5). The sample were injected into the column from a 5 mL injection loop and the column washed with 40 mLs of buffer, after which fraction collection began. Recombinant cMCR-1 and cMCR-2 were eluted at a flow rate of 1 mL per min. 1 mL fractions were collected and analysed by SDS-PAGE and mass spectroscopy to confirm the protein identity.

Table 5: List and composition of the buffers used for recovery and purification of the different recombinant proteins produced.

Protein Recovered	Buffer name	Composition
MCR-1	Lysis Buffer	50 mM HEPES pH 7.5, 500 mM NaCl
	Membrane solubilization buffer	50 mM HEPES pH7,5, 10% glycerol, 200 mM NaCl and 1% n-Dodecyl-β-D-maltoside (DDM)
	Washing Step 1	50 mM HEPES pH 7.5, 200 mM NaCl, 0.1% DDM, 10 mM Imidazole
	Washing Step 2	50 mM HEPES pH 7.5, 200 mM NaCl, 0.1% DDM, 20 mM Imidazole
	Ni-NTA Elution	50 mM HEPES, 250 mM Imidazole, 100 mM NaCl buffer and 0.1% DDM
	SEC/SEC-MALS	50mM HEPES pH 7.5, 150 mM NaCl and 0.03% DDM
cMCR-1, cMCR-2 and ScPmrC*	Lysis Buffer	50 mM HEPES pH8, 500 mM NaCl, 2mM β-mercaptoethanol
	Washing Step 1	50 mM HEPES pH7.5, 400 mM NaCl, 10 mM Imidazole, 1mM TCEP, 0.1 mM ZnCl ₂
	Washing Step 2	50 mM HEPES pH 7.5, 400 mM NaCl, 20 mM Imidazole, 1mM TCEP, 0.1 mM ZnCl ₂
	Ni-NTA Elution	50 mM HEPES pH 7.5, 150 mM NaCl, 0.1 mM ZnCl ₂ , 250 mM imidazole and 1 mM TCEP
	SEC Purification	50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM ZnCl ₂
	Lysis Buffer	50 mM HEPES pH 7.5, 500 mM NaCl and 1 X EDTA-free protease inhibitor pellet (Roche)
	Washing Step 1	50 mM HEPES7.5, 400 mM NaCl, 10 mM Imidazole, 1mM TCEP, 0.1 mM ZnCl ₂
	Washing Step 2	50 mM HEPES pH 7.5, 400 mM NaCl, 20 mM Imidazole 20 mLs, 1mM TCEP, 0.1 mM ZnCl ₂
	Ni-NTA Elution	50 mM HEPES pH 7.5, 150 mM NaCl, 0.1 mM ZnCl ₂ , 250 mM imidazole and 1 mM TCEP
	SEC Purification	50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM ZnCl ₂ .
ScPmrC only *Washing Step 3		50 mM HEPES pH 7.5, 200 mM NaCl, 20 mM Imidazole, 1mM TCEP, 0.1 mM ZnCl ₂

2.2.3 Large scale expression of recombinant proteins.

All the transformed *E. coli* strains (Table 2) from the small-scale expression tests were kept in LB-glycerol (20 % v/v) stocks at -80°C. A starter culture was prepared by seeding 5 µL of the bacterial stock in 5 mL of LB-antibiotic (antibiotic concentrations according to Table 2) media. The culture was grown overnight at 37° C and 180 RPM. *E. coli* cultures of strain ShuffleT7, harbouring the plasmid *ScPmrC-pOPINF*, were grown at 30° C. All *E. coli* strains were cultivated in shake flasks in the different culture media. Cultivations were performed in 500 mL (working volume) shake flasks 8 shake flasks were used for each culture to a total of 4 litres of culture. After inoculation, the cultures were incubated at 37° C (30°C for strain ShuffleT7) and 180 RPM shaking until an OD₆₀₀ of approx.0.8 was reached. Recombinant protein production was induced with the addition of IPTG to a final concentration of 400 µM. The temperature was lowered to 18° C (16°C for strain ShuffleT7) and recombinant protein production lasted for 18 hours. 1 mL samples were taken and OD₆₀₀ was measured in duplicates, the samples were then diluted to a final OD₆₀₀ of 0.1 and stored for SDS-PAGE analysis. Cells were harvested by centrifugation at 5000 RPM at 4° C for 10 min (Avanti J-26, Beckman Rotor JLA-8.1000, Beckman Coulter, Ireland) and pellets frozen at -80 °C until needed.

2.2.3.1 Media composition for large-scale expression optimization.

For the optimization of full-length MCR-1 expression, different media compositions were tested to determine the best condition for both cell and recombinant protein production. The composition of the media tested can be seen in Table 6. Large-scale expression was carried out in total culture volumes of 4 litres of the selected media.

Table 6: Growth media compositions.

Media composition	Concentration	Units
1st Medium 32YT		
Yeast extract	3.2	% w/v
Peptone	0.8	% w/v
NaCl	100	mM
2nd Medium SD 32YT		
yeast extract	3.2	% w/v
Peptone	0.8	% w/v
NaCl	100	mM
Trace Metals*	1	mL
Glycerol	1.2	% w/v
Glucose	0.5	% w/v
MgSO ₄	1	mM
*Trace metals concentration FeCl ₃ 50 mM, CaCl ₂ 20 mM, ZnSO ₄ 10 mM, H ₃ BO ₃ 2 mM and HCl 40 mM.		

Trace metals and MgSO₄ solution were filtrated with a 0.22µm syringe filter and store at 4° C. The final medium used for enhanced protein production was a 2xYT medium supplemented with trace metals and MgSO₄ solution.

2.2.4 Full-length MCR-1 recovery and purification.

The harvested MCR-1 producing cells recovered from large scale expression experiments, were resuspended in 100 mL lysis buffer with four tablets of EDTA-free protease inhibitor (Roche) and broken down by three passages through a cell disruptor at 30 000 psi and 5°C (CF1, Cell Disrupter Constant Systems LTD, UK). Unbroken cells and insoluble debris were removed by centrifugation at 15,000 RPM and 4°C for 20 mins (Avanti J-26 XP, rotor JA-25.50, Beckman Coulter, Ireland). The supernatant from MCR-1 samples was recovered and the biological membranes were pelleted by ultracentrifugation at 150,000 RPM and 4°C for 2 hrs (Optima L-80 XP ultracentrifuge, rotor 70-Ti, Beckman Coulter, Ireland). The pelleted membranes were re-suspended in 30 mL of membrane solubilization buffer (Table 5) and incubated in a shaker for 1 hour at 4°C. The solubilized membranes were recovered by

ultracentrifugation at 150,000 RPM and 4°C for 45 minutes (Optima L-80 XP ultracentrifuge, rotor 70-Ti, Beckman Coulter, Ireland). 100 µL samples of the solubilized membranes and the supernatant were taken for SDS-PAGE analysis and the rest of the solution was used for further purification of MCR-1 by Ni²⁺-NTA affinity chromatography. The pellet was discarded and the supernatant stored for further experimentation. Imidazole was added to the supernatant, to a final concentration of 5 mM, to the supernatant.

The supernatant, containing MCR-1 and detergent, was filtered using a syringe filter of 0.2 µm pore size and applied to a 5 mL His-Trap chelating affinity column (Ni NTA, GE Healthcare). The column was washed twice, using first 10 bed volumes of Washing step 1 buffer and then washed for the second time in 10 bed volumes of Washing step 2 buffer to remove any unbound proteins; 100 µL samples of both washing steps were recovered for SDS-PAGE analysis. The bound MCR-1 was eluted in 5 bed volumes using Ni-NTA Elution buffer. 50 µL samples were collected for SDS-PAGE analysis. The bound MCR-1 was eluted in 5 bed volumes using Ni-NTA Elution buffer. 50 µL samples were collected for SDS PAGE analysis. Imidazole concentration was reduced, by buffer exchange, to 10 mM using an Amicon 10-kDa cut-off centrifugal filter and the protein concentrated to a final volume of 500 µL. The concentrated protein was applied to a size exclusion chromatography column (Superdex S200 10/300, GE Healthcare), the column was previously equilibrated in SEC/SEC-MALS buffer and the flow set to 0.5 mL/min. 1 mL fraction collection started after 5 mL of elution volume. Samples of the elution peaks were taken and analysed by SDS-PAGE. Fractions corresponding to the purified MCR-1 protein were pooled and concentrated using a 10 kDa molecular weight cut-off centrifugal filter (Amicon) to a final concentration of 16 mg/mL.

2.2.5 Large scale recovery and purification of cMCR-1, cMCR-2 and ScPmrC.

The harvested cells obtained from large scale expressions were resuspended in 100 mL of lysis buffer (Table 5) with four tablets of EDTA-free protease inhibitor (Roche) pellets. The

bacterial slurries were broken by three 30 000 psi passages through a cell disruptor at 5°C (CF1, Cell Disrupter Constant Systems LTD, UK). The unbroken cells were removed by ultracentrifugation at 100,000 × RPM for 1 h (Beckman Coulter Optima L-80 XP Ultracentrifuge, Beckman Rotor 70 Ti). The supernatants were recovered, imidazole was added to a final concentration of 10 mM and incubated for 2 hrs with 2 mL Ni⁺-NTA resin (Qiagen); Ni⁺-NTA resin was previously washed in 20 mL of distilled water and recovered by centrifuging for 10 minutes at 800 RPM and 4 °C. The protein-bound resin was then washed with 20 mL of washing buffer 1 and then a second washing step using Washing step 2 buffer (according to Table 5); ScPmrC containing samples were washed a third time using Washing step 3 buffer. Imidazole was reduced to a final concentration of 10 mM by exchange into imidazole free buffer using an Amicon 10-kDa centrifugal filter (20 mL volume) centrifuged at 3000 RPM. The His-tag was removed by adding 300 µL of purified recombinant 3 C protease (3 mg/mL), kindly provided by Dr Philip Hinchliffe, to 30 mL of purified cMCR-1 and cMCR-2, respectively, and left to cleave overnight. ScPmrC containing samples did not undergo 3C protease treatment. The removed His-tag, and residual un-cleaved protein, was captured on 1 mL of Ni-NTA resin slurry. All steps were performed in a cold room at 4 °C. The protein samples were concentrated and reduced to a final volume of 5 mL and subsequently injected onto a Superdex 75 column (Superdex S75 16/60, GE Healthcare, bed volume of 120 mL) equilibrated overnight using 2 bed volumes of buffer SEC Purification buffer (Table 5). The samples were injected into the column using a 5 mL injection loop and the column washed with 40 mLs of buffer, after which fraction collection began. A flow rate of 1 mL/min was set and 1 mL fractions across the eluting peaks were collected and analysed by SDS-PAGE and mass spectroscopy to confirm the presence and identity of the purified proteins. The fractions were pulled down and concentrated using a molecular weight cut off centrifugal filter (10 kDa MW cut off, 20 mL volume, Amicon) to a final concentration of 15 mg/mL for cMCR-1 and cMCR-2 and 12 mg/mL for ScPmrC.

2.2.6 Protein analysis by Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE).

Samples for protein analysis by SDS-PAGE [136], were prepared by adding an equal volume of SDS PAGE loading buffer (2% SDS, 10% glycerol, 12.5 mM Tris-Cl pH 6.8 and 0.1 % bromophenol blue; 50 μ L of β -mercaptoethanol was added to 950 μ L of this solution) and sample to a final volume of 20 μ L. For the samples recovered from bacterial cultures, 1 mL samples were taken and centrifuged for 5 minutes at 5000 RPM in an Eppendorf centrifuge (HERMLE Z 207 M, Labortechnik GmbH, Germany). The pellet was then re-suspended in 100 μ L of distilled water. 10 μ L of the resuspended pellet were mixed with 10 μ L of loading buffer and heated at 95° C for 10 minutes.

The samples were run on pre-cast 12% glycine SDS-PAGE (NuSep, USA) mini gels in running buffer (25 mM Tris-HCL, 200 mM glycine, and 0.1% SDS (w/v) [61] using a Bio-Rad electrophoretic MINI-Protean™ Tetra cell system (Bio-Rad) running at a 150 V. Gels were stained and visualized using Quick Coomassie Stain (Generon, Slough, UK).

2.2.7 Protein concentration determination.

The protein concentration was determined using the Beer-Lambert Law (Equation 1) from the absorbance at 280 nm using a Nanodrop spectrophotometer (Nanodrop Lite, Thermo Scientific) and the extinction coefficient was obtained from the ProtParam tool [137] using the ExPASy website and the protein amino acid sequence for MCR-1, cMCR-1, cMCR-2 and ScPmrC (Appendix A, Appendix F, Appendix H, Appendix R respectively).

$$A = \epsilon * b * C$$

Equation 1: Beer-Lambert law equation

Where A is absorbance measured at 280 nm, ϵ is the molar extinction coefficient of the protein at 280 nm, b is the length of the light path (in centimetres) and C is the concentration of the protein solution.

2.2.8 Protein identification by MS/MS analysis.

Prominent protein bands for MCR-1, cMCR-1, cMCR-2 and ScPmrC, located between 60-50 and 40 kDa on the SDS-PAGE were excised from the SDS-PAGE gels and analysed by mass spectrometry by Dr. Kate Heesom at the University of Bristol proteomics facility. Each gel band was subjected to in-gel enzymatic digestion, performed with trypsin or chymotrypsin, using a DigestPro automated digestion unit (Intavis Ltd.). The resulting peptides were fractionated using an Ultimate 3000 nano-LC system in line with an LTQ-Orbitrap VelosPro mass spectrometer (Thermo Scientific). Peptides in 1% (v/v) formic acid were injected onto an Acclaim PepMap C18 nano-trap column (Thermo Scientific). After washing with 0.5% (v/v) acetonitrile 0.1% (v/v) formic acid peptides were resolved on a 250 mm × 75 µm Acclaim PepMap C18 reverse-phase analytical column (Thermo Scientific) over an 80 min organic gradient (1-50% solvent B over 55 min, 50-90% B over 0.5 min, held at 90%B for 5min and then reduced to 1%B over 0.5min.) with a flow rate of 300 nl/min. Solvent A was 0.1% formic acid and Solvent B was aqueous 80% acetonitrile in 0.1% formic acid. Peptides were ionized by nano-electrospray ionization at 2.1 kV using a stainless-steel emitter with an internal diameter of 30 µm (Thermo Scientific) and a capillary temperature of 250°C. Tandem mass spectra were acquired using an LTQ- Orbitrap Velos mass spectrometer controlled by Xcalibur 2.1 software (Thermo Scientific) and operated in data-dependent acquisition mode. The Orbitrap was set to analyse the survey scans at 60,000 resolution (at m/z 400) in the mass range m/z 300 to 2000 and the top twenty multiply charged ions in each duty cycle selected

for MS/MS in the LTQ linear ion trap. Charge state filtering, where unassigned precursor ions were not selected for fragmentation, and dynamic exclusion (repeat count, 1; repeat duration, 30 s; exclusion list size, 500) were used. Fragmentation conditions in the LTQ were as follows: normalized collision energy, 40%; activation q, 0.25; activation time 10ms; and minimum ion selection intensity, 500 counts.

The raw data files were processed and quantified using Proteome Discoverer software v1.4 (Thermo Scientific) and searched against the UniProt *Escherichia coli* (strain B BL21-DE3) database (downloaded January 2020: 4172 sequences) plus the corresponding protein sequence (MCR-1, cMCR-1, cMCR-2 and ScPmrC respectively) using the SEQUEST algorithm. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.8Da. Search criteria included carbamidomethylation of cysteine as a fixed modification and oxidation of methionine as a variable modification. When required, phosphorylation of serine, threonine, and tyrosine (+79.966 Da) was also included as a variable modification. Searches were performed with full tryptic digestion and a maximum of 2 missed cleavages were allowed. The reverse database search option was enabled and all peptide data were filtered to satisfy a false discovery rate (FDR) of 1%.

2.2.9 MCR-1 SEC-MALS Analysis.

SEC-MALLS analysis of purified MCR-1 was carried out using an HP 1100 HPLC system with a Superdex 200, 10/300 column (GE Healthcare) coupled with a DAWN Heleos II (MALS) and Optilab rEX detector. The column was equilibrated with SEC/SEC-MALS buffer (Table 5) at a flow rate of 0.1 mL/min until a stable baseline was observed in the MALLS detectors. 100 μ L of the purified MCR-1, with a final concentration of 1 mg/ml in 0.03% (w/v) DDM, were applied to the column and the flow rate was set to 0.5 mL/min. The data obtained from the detectors (Abs280, light scattering, and refractive index) were analysed using the Astra 6 software. A sample of 100 μ L of Bovine serum albumin (BSA), at a concentration of 2 mg/ml, was prepared

under the same buffer conditions and use as a protein standard. Finally, 100 μL of a detergent micelle-only (1% (w/v) DDM) samples were prepared in matching MCR-1 sample buffer and run under the same conditions.

2.2.9.1.1 Molar Mass estimation using the 3 detectors method.

$$MW_{protein} = \frac{\Delta LS * \Delta UV_{280}}{K * A_{280,protein} * dRI^2}$$

Equation 2: Equation used to calculate directly the molecular mass of the protein in the protein/detergent micelle.

Where: ΔLS is the difference in the light scattering signal between the protein and the detergent samples. ΔUV is the differential in the UV signal between the protein and the detergent samples. A_{280} is the extinction coefficient in $\text{ml/mg} * \text{cm}$ of the protein. K is the column constant. dRI is the signal of the diffraction index detector for the protein sample. By rearranging Equation 2, the column constant K can be obtained as follows:

$$K = \frac{\Delta LS * \Delta UV}{MW_{protein} * A_{280,protein} * dRI^2}$$

Equation 3: Column constant K.

Using Bovine Serum albumin (BSA) as a standard to obtain the column constant K , the molecular weight ($MW_{protein}$) is replaced by BSA MW of 66,400 Da and BSA extinction coefficient (at 280 nm) for 0.66 $\text{ml/mg} * \text{cm}$. The column constant K becomes:

$$K = \frac{\Delta LS * \Delta UV}{66,400 * 0.66 * dRI^2}$$

$$K = \frac{\Delta LS * \Delta UV}{43824 * dRI^2}$$

Equation 4: Obtaining the column constant for BSA standard.

2.2.10 Circular Dichroism Spectroscopy of MCR-1 and ScPmrC.

The purified MCR-1 was buffer exchanged and diluted to a final concentration of 0.6 mg/ml, in 50mM HEPES pH 7.5, 10 mM NaCl, and 0.03% (w/v) DDM and Circular Dichroism (CD) spectra were measured between 240nm and 190 nm. Far-UV spectra were measured in a CD spectrophotometer (Jasco J-1500) at different temperature intervals, to obtain the melting temperature and determine if refolding was possible, at 5 °C, 25 °C, 35 °C, 55 °C, 75 °C, and 95°C; using a 2 mm path length quartz cuvette. Data were collected every 1 nm with a bandwidth between 190 nm and 240 nm using an integration time of 1 second per step. Melting temperature was analysed following a 222 nm wavelength. Measurements were taken at a rate of 1° C per minute for a period of nearly 4 hours, where the first 2 hours were used to reach 95° C and the other 2 to lower the temperature back to 5° C. An asymmetrical 4PL function, where X is the temperature (and the T_m is reported as the EC₅₀ value) was performed using GraphPad Prism version 8.2.1 for macOS (GraphPad Software, San Diego, California USA) to estimate the melting temperature.

For ScPmrC, the purified protein sample was buffer exchanged using an Amicon 10-kDa centrifugal filter, by centrifugation at 3000 RPM and 4 °C (ALC Multispeed Refrigerated Centrifuge, model PK 131R) and diluted to a final concentration of 10 µM in 50mM HEPES pH 7.5, 10 mM NaCl. CD spectra were obtained and data processed under the same conditions as for MCR-1. Secondary structure analysis was performed using the K2D3 server (<http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/>).

2.2.11 Synthesis of 2-aminoethyl (4-nitrophenyl) hydrogen phosphate (pNP-PEtN).

Synthesis of pNP-PEtN was carried out by Adam Gorman under the supervision of Prof. Paul Pringle (School of Chemistry, University of Bristol). 1 gramme of 4-Nitrophenyl phosphorodichloridate (Figure 12 a)) was dissolved in 15 mL of tetrahydrofuran (THF) and cooled to -15 °C. To this, a solution containing 0.63 g of *N*-*t*-butoxy carbonyl-2-amino-1-ethanol (Figure 12 b)) and pyridine (0.32 ml, 0.31 g) in THF (7 ml) was added dropwise over a period of 15 minutes. The solution was stirred at room temperature for 3 hrs. over which period a white precipitate formed. The THF was removed *in vacuo* and the resultant white oil was dissolved in 100 mL of water and basified to pH 9.0 with the addition of small portions of NaHCO₃. In air, the resultant yellow solution was washed twice in 30 mL of diethyl ether; the aqueous solution was acidified with 2 M HCl to pH 1 to give a clear solution. The crude product was extracted with ethyl acetate (3 x 100 ml) and the volatiles were removed *in vacuo* to give a cloudy oil which solidified after standing overnight. 40 mL of isopropanol (iPrOH) were added and the mixture was sonicated to form a white precipitate which was filtered and washed twice with 5 mL of cold methanol to obtain the 4-nitrophenol-PEtN (Figure 12 c)) as a white crystalline solid.

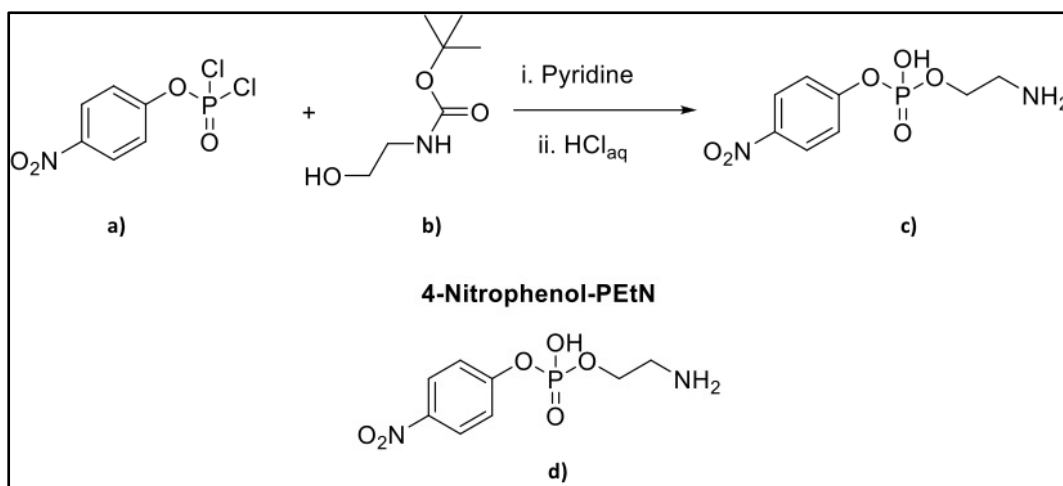


Figure 12: Chemical synthesis of 4-Nitrophenol-PEtN.

a) 4-Nitrophenyl phosphorodichloridate, b) N-t-butoxy carbonyl-2-amino-1-ethanol, c) and d) 4-Nitrophenol-PEtN. The substrate used in activity assays. Molecular weight: 262.16 g/mol. Chemical formula: $C_8H_{11}N_2O_6P$

This chromogenic substrate was used for the enzymatic assay. 2.62 mg of pNP-PEtN were dissolved in 1 mL of distilled water and heated at 37° C for approx. 5 mins to make a 10 mM stock solution. This stock was stored at -20° C.

2.2.12 Buffer pH and pNP-PEtN stability.

15 μ L of a 10 mM stock of pNP-PEtN were added to 135 μ L of Bis-Tris propane buffer (50 mM) with pH values ranging from 6.5 to 9.0 in increments of 0.5. The pH was measured using a pH meter (HI-2020 edge, HANNA Instruments, USA). NaOH (1 M stock solution) was added dropwise to adjust the pH during buffer preparation. Progress of pNP-PEtN degradation was measured at a wavelength of 405 nm for 4 hours using a ClarioStar plate reader (BMG Labtech, Aylesbury, UK).

2.2.13 Assessing the activity of recombinant cMCR-1, cMCR-2 and ScPmrC.

A concentration range of 10, 20, 50 and 100 μ M, of the purified enzymes purified, was used for the development of a phosphoethanolamine transferase activity assay. The reaction was

started by the addition of 15 μL , from a 10 mM aqueous *p*NP-PEtN stock, to 135 μL of purified enzyme (cMCR-1, cMCR-2 and ScPmrC respectively) in 50 mM Bis-tris propane pH 7.0 (Sigma Aldrich), 150 mM NaCl buffer in 96-well flat-bottom plates (Corning, clear flat bottom polystyrene 3595). The final reaction volume was 150 μL . The plates were shaken inside a CLARIOstar plate reader (BMG Labtech, UK) by orbital shaking at 100 RPM for 30 seconds before measurements started; the reaction was read at 25° C for 4 hours at 30 second intervals, with absorbance measured at 405 nm.

2.2.14 Analysis of the enzyme reaction by LC-MS.

2.2.14.1 Protein precipitation and sample preparation.

The enzymatic reaction was stopped by protein precipitation using trichloroacetic acid (TCA). 16.5 μL of 100% (v/v) TCA were added to 150 μL of the reacted solution; the sample then was incubated on ice for 10 minutes, after this incubation the sample was diluted with 500 μL of a 10% (v/v) TCA solution. This solution was left on ice for 20 minutes and then centrifuged at 13,000 RPM on a benchtop centrifuge for 30 minutes at room temperature (HERMLE Z 207 M, Labortechnik GmbH, Germany). 20 μL of supernatant was recovered (for each reaction) and analysed by LC-MS. The pellet was rinsed using 500 μL of acetone and centrifuged at 13,000 RPM and 4° C on a benchtop centrifuge for 10 minutes (HERMLE Z 207 M, Labortechnik GmbH, Germany). This step was done twice and the pellet was then dried and stored for peptide mass fingerprinting.

2.2.14.2 LC-MS analysis.

Samples from the enzyme-catalysed reactions were recovered for analysis. The release of *p*-nitrophenol from *p*NP-PEtN was analysed using liquid chromatography coupled with mass spectrometry (LC-MS) with the help of Dr Luoyi Wang at the University of Bristol Chemistry department. LC-MS data were obtained on a Waters LCMS system (Waters 2767

autosampler, Waters 515 HPLC pump, Waters 2998 Diode Array detector, Waters 2424 ELS detector and Waters Quatro Micro mass spectrometer). High-performance liquid chromatography (HPLC) grade H₂O and acetonitrile (MeCN) were mixed with 0.05% formic acid and used as a solvent system. Analytical LC-MS data were obtained using a Phenomenex Kinetex column (C₁₈, 250 x 4.60 mm, 5 μm) at a flow rate of 1 mL/min, with a gradient of 5% MeCN to 95% MeCN in 20 mins. HPLC purification was carried out using a Phenomenex Kinetex column (C₁₈, 250 x 21.20 mm, 5 μm) at a flow rate of 16 mL/min. High-resolution Mass Spectrometry (HRESIMS) data were obtained on a Bruker Daltonics micrOTOF II instrument.

To measure *p*NP formation a standard curve was created by dissolving 6.95 mg of *p*-nitrophenol (*p*NP, Sigma) in 10 mL of 50 mM HEPES pH 7.0 and 150 mM NaCl buffer to create a 5 mM stock solution. The stock solution was diluted to achieve different *p*-nitrophenol concentrations; 0.125, 0.250, 0.350, 0.5, 0.75, 1, 1.5 and 2 mM. The samples were analysed by LC-MS, as previously described. The area under the elution peak was measured using the diode array detector and plotted against *p*NP concentration in μM using GraphPad Prism version 8.2.1 for macOS (GraphPad Software, San Diego, California USA) and the equation was obtained.

$$AU = 12199012 * (pNP \text{ concentration}) + 19818$$

Equation 5: Linear regression equation for *p*NP at different concentrations in a diode array detector.

*Where AU is the area under the peak for *p*NP in the diode array detector and *p*NP concentration is given in μM.*

2.2.15 Estimation of the steady-state kinetic parameters for cMCR-2 and ScPmrC.

The activity of cMCR-2 and ScPmrC was estimated by setting up an activity assay with varying *p*NP-PEtN concentrations. The purified enzymes were diluted, to a final concentration of 50 μ M, in Bis-tris propane pH 7.0 (Sigma Aldrich), 150 mM NaCl buffer in 96-well flat-bottom plates (Corning, clear flat bottom polystyrene 3595). The reactions were started by the addition of *p*NP-PEtN in increasing concentrations from 250 μ M to a final concentration of 2 mM. The reaction final volume was 150 μ L. The plates were read at 25° C for 18 hours at 30-second intervals, with absorbance being measured at 405 nm using a ClarioStar plate reader (BMG Labtech, Aylesbury, UK). *p*NP concentration was determined from a standard curve generated by making serial dilutions from a 2 mM stock of *p*NP to a final concentration of 125 μ M *p*NP, as described below, with an extinction coefficient obtained by linear regression of the plot of absorbance against *p*NP concentration.

The initial rates of reaction were obtained from plots of absorbance against time at the various *p*NP-PEtN concentrations and plotted to determine the kinetic parameters K_M and V_{max} . The kinetic parameters were determined by both Michaelis-Menten and Lineweaver-Burk plots. The reaction rate, v , was plotted against *p*NP-PEtN concentration to generate the Michaelis-Menten plot. The parameters were estimated using Equation 6 [159].

$$v = \frac{V_{max} \times [Substrate]}{K_M + [Substrate]}$$

Equation 6: Michaelis-Menten equation.

Where v is the velocity of the reaction, V_{max} is the maximum theoretical reaction rate, when the substrate concentration is infinite, and K_M is the Michaelis constant.

Curve fitting to estimate the v_{max} , k_M and k_{cat} were carried out using GraphPad Prism version 8.2.1 for macOS (GraphPad Software, San Diego, California USA); for K_{cat} estimation E_T was set up to be 50 μ M, the total protein concentration used for these experiments. The Lineweaver-Burk plots were obtained by plotting the inverse of the reaction rate, $1/v$, against the inverse of the substrate concentration $1/[S]$ [139]. The kinetic parameters k_M and k_{max} were determined from the slope and the Y-intercept of the linear Equation 7.

$$\frac{1}{v} = \frac{K_M}{v_{max}} \frac{1}{[S]} + \frac{1}{v_{max}}$$

Equation 7: Lineweaver-Burk double reciprocal equation.

To estimate p-nitrophenol formation, a standard curve was also generated by dissolving 6.95 mg of p-nitrophenol (pNP, Sigma) in 10 mL of 50 mM HEPES pH 7.0 and 150 mM NaCl buffer to create a 5 mM stock solution. The stock solution was diluted to achieve different p-nitrophenol concentrations; 0.125, 0.250, 0.350, 0.5, 0.75, 1, 1.5 and 2 mM. 150 μ L of the each pNP concentrations were set up in a 96-well flat-bottom plate (Corning, clear flat bottom polystyrene 3595) and absorbance was measured at an optical wavelength of 405 nm using a ClarioStar plate reader (BMG Labtech, Aylesbury, UK), by technical replicates. The linear regression curve for abs vs pNP concentration (mM) was plotted using GraphPad Prism version 8.2.1 for macOS (GraphPad Software, San Diego, California USA). The resulting equation was used to estimate the formation of pNP.

$$Abs_{405} = 0.0013 * (pNP \text{ concentration}) - 0.008$$

Equation 8: Linear regression equation to estimate pNP release.

Where Abs_{405} is the absorbance reading at a wavelength of 405 nm and pNP concentration is measured in μ M.

2.2.16 Evaluation of Thioglycolic Acid as a small-molecule inhibitor of cMCR-2.

In order to test the efficacy of the developed enzymatic assay as a screening test for inhibitors a reaction using the C-terminal domain of MCR-2, cMCR-2 and the compound thioglycolic acid (TGA) as an inhibitor candidate. The reaction was set up in a 96-well flat-bottom plate (Corning, clear flat bottom polystyrene 3595). The reaction was started by the addition of 15 μ L, from a 10 mM aqueous *p*NP-PEtN stock, to 135 μ L of cMCR-2 (50 μ M) in 50 mM Bis-tris propane pH 7.0 (Sigma Aldrich), 150 mM NaCl buffer containing 1 mM of thioglycolic acid (TGA) and incubated at room temperature. A positive control followed the same conditions without the presence of TGA, reaction started by adding 15 μ L, from a 10 mM aqueous *p*NP-PEtN stock to 135 μ L of cMCR-2 (50 μ M) in 50 mM Bis-tris propane pH 7.0 (Sigma Aldrich), 150 mM NaCl buffer. A negative control was generated by soaking cMCR-2 in 100 mM EDTA overnight to remove all zinc present in the buffer; EDTA was removed by dialysis overnight, the reaction was started by adding 15 μ L, from a 10 mM aqueous *p*NP-PEtN stock, to 135 μ L of EDTA soaked cMCR-2 (50 μ M) in 50 mM Bis-tris propane pH 7.0 (Sigma Aldrich), 150 mM NaCl buffer. The final reaction volumes were 150 μ L. The plate was shaken inside a CLARIOstar plate reader (BMG Labtech, UK).by orbital shaking at 100 RPM for 30 seconds before measurements started; the reaction was read at 25° C for 16 hours at 30-second intervals, with absorbance measured at 405 nm.

2.2.17 Crystallisation trials of ScPmrC.

Crystallisation of the PmrC soluble domain, ScPmrC, was attempted by sitting-drop vapour diffusion at 4 °C in MRC 96-well crystallisation plates (Molecular Dimensions). Five sparse-matrix crystallization screens were used: Morpheus, JCSG-plus, Structure Screen 1 + 2 HT and PACT premier from Molecular Dimensions (Molecular Dimensions, Sheffield, UK) and a Top96 Crystallization Screen (Anatrace, USA); exploring a total of 480 crystallization conditions. Crystallization droplets were generated using a Mosquito LV liquid handling robot

(SPT Labtech Ltd., UK); with two mixing conditions used for the crystallisation trials: the first 400 nL of ScPmrC (11.06 mg/mL) mixed with 200 nL crystallisation reagent for a final droplet volume of 600 nL; and the second 200 nL of ScPmrC (11.06 mg/mL) mixed with 400 nL crystallisation reagent. Droplets were equilibrated against 50 μ L crystallisation reagent. The crystallization trays were sealed and incubated at 4 °C in a crystallisation hotel (RockImager, Formulatrix) and pictures of the 960 wells were taken on days 0,1,2, 3, 5, 8, 13 and 21.

2.2.18 Crystallization optimization.

Crystallization conditions for the PmrC soluble domain, ScPmrC, were optimized based on the conditions found from initial crystallization trials by the sitting drop method in 4 x 6 (24 well) plates. Citrate/ phosphate buffer (0.1 M) at 4 different pH values (3.8, 4.2, 4.6 and 5.0) was used with 6 different ethanol (precipitant) concentrations (25 %, 30 %, 35 %, 40 %, 45 % and 50 % v/v). Two droplet conditions were tested, the first one was created by mixing 1 μ L of purified ScPmrC (11.06 mg/mL) with 1 μ L of buffer + precipitant for a final droplet volume of 2 μ L; the second condition was created by mixing 2 μ L of purified ScPmrC (concentration of 11.06 mg/mL) with 1 μ L of buffer + precipitant for a final droplet volume of 3 μ L. A total of 48 conditions were tested. The crystallization trays were incubated at 4 °C, inside a cold room, and pictures of the 48 wells were taken on days 5, 7 and 14. Signs of crystals were found as early as 5 days of incubation.

Chapter 3 Structure based phylogenetic study of the alkaline phosphatase superfamily.

3.1 Introduction:

Mobile Colistin Resistance (MCR) enzymes are a group of recently discovered phosphoethanolamine transferases that confer Gram-Negative bacteria (GNB) with resistance to colistin, an antibiotic regarded as one of the last lines of defence against the growing threat that is multidrug-resistant GNB [140]. By catalysing the transfer of phosphoethanolamine (PEtN) to the Lipid A present in the bacterial outer membrane, they diminish the efficacy of colistin by up to 8 times [52]. Members of the MCR family have been identified as part of the Alkaline phosphatase (AlkP) superfamily [74,141]. A structural comparison of the crystallized catalytic domain of MCR-1 and the Shrimp Alkaline phosphatase (SAP) showed a similar $\alpha/\beta/\alpha$ fold, commonly found in alkaline phosphatases (Figure 13). The AlkP superfamily is composed of a large amount of metalloenzymes that share a highly conserved active site and overall structure but low sequence similarity [117,119] and are widely distributed across biological kingdoms.

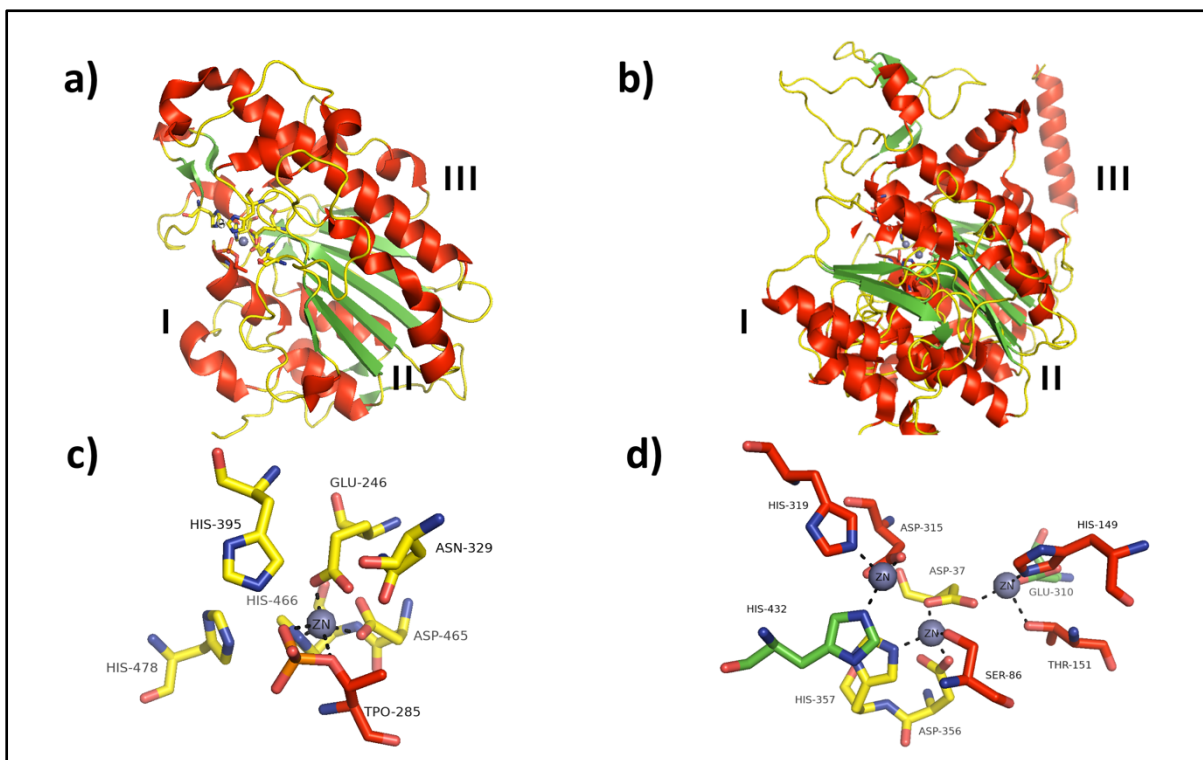


Figure 13: Structures of the catalytic domains of MCR-1 and Alkaline Phosphatase.

a) Structure of the catalytic domain of MCR-1 (PDB entry 5LRN). b) Structure of Alkaline Phosphatase (PDB entry 1K7H). The architecture of the $\alpha/\beta/\alpha$ fold can be seen in both structures: a core of beta-strands (I) is surrounded by two layers of α -helices (I and III). α -helices are shown in red, β -sheets in green and loops yellow. c) Catalytic site of MCR-1 and d) catalytic site of the alkaline phosphatase. Zinc atoms are represented as grey spheres, oxygen atoms are red, nitrogen atoms blue.

3.1.1 The Alkaline Phosphatase superfamily.

The alkaline phosphatase superfamily has an $\alpha/\beta/\alpha$ fold (a central group of up to 8 β -strands being enveloped by two layers of α -helices as seen on Figure 13 a) [142–144]. This superfamily encompasses enzymes with different activity classifications, e.g. hydrolases, isomerases, and transferases; however, family members differ in substrate preference, preferentially acting upon phospho-, sulpho- and phosphonocarbohydrate substrates [117,145]. Most members of the AlkP superfamily are metalloenzymes, however the nature of the metal ion(s) present in their catalytic site, which include Zn^{2+} , Ca^{2+} , Mg^{2+} , Mn^{2+} and Co^{2+} , as well as the number of them, vary among them [118,144,146,147]. This variation in metal

ions present in the catalytic site is accompanied by differences between members of this superfamily in the essential catalytic amino acid; the catalytic site of alkaline phosphatase (AlkP) uses serine as the nucleophile, the nucleotide pyrophosphatase/phosphodiesterase (EPP/NPP) uses threonine [148], and arylsulphatases (AS) uses a post-translationally modified cysteine (or serine), known as formylglycine [113,149]. The diversity of this superfamily, which is widespread in bacteria and eukaryotes [146], is reflected in low sequence homology; however overall structure and active site architecture are highly conserved. This highly conserved structure has helped in the identification of several previously uncharacterized proteins [117]. Extensive studies of the alkaline phosphatase superfamily have not only provided deep insight into enzyme promiscuity, the ability of an enzyme to catalyse the turnover of multiple substrates; but have also served to develop a framework for the understanding of the evolutionary development of protein structure and function and as base model of the function of catalytic metal centres in complex reaction mechanisms [114,117,143–145,148,150–153]. Performing a structural analysis of the features shared among members of this superfamily, can enable new understanding of protein evolution, mechanism and function to be obtained.

3.1.2 Structural and sequence phylogenetic study of enzymes.

Structural alignment and comparison of proteins is a useful technique that facilitates the understanding of the evolutionary origin of and evolution among protein families, even when analysing distant protein relatives, due to the fact that the three-dimensional structure of proteins is highly conserved during the evolution process. Protein folds are more susceptible to evolutionary convergence, as evidence shown that the vast majority of the know protein domains belongs to a very small number of “superfolds” [154,155] making the likelihood of fold convergence much greater than sequence convergence, even for organisms that grew in different environments [156]. The main drawback in this technique is the difficulties that arise when trying to distinguish between similar 3D protein structures that are the result of an

evolutionary process and the ones that come about due to common properties such as chemical constraints on protein folding [156].

Any set of proteins sharing a common ancestor is expected to also show similarities in structure, function and mechanism; whereas two proteins that evolved independently may not [157]. Phylogenetic analysis, based on amino acid sequence, provides relevant information regarding the evolutionary process as any set of proteins that share a common ancestry displays similar structures and functions, proportional to the similarity of their sequences [158], however, in cases where the sequence identities are below 30%, structure-based analysis has proved to be a more effective approach. A study comparing the 3D structure and sequence similarity of 108 known protein family domain suggested that protein evolution is best modelled using 3D structures when sequence similarity, among homologues, is low [159]; proposing that protein structure is better conserved than amino acid sequence [160].

Structure-based analysis can be used in order to better understand the features of enzyme catalytic mechanisms, key elements such as amino acid conservation and as a way to predict starting points for experimental procedures; at the superfamily level, sequence-based methods are useful for the study of the evolutionary relationships within large protein families [145]. Much work has been done regarding the structural analysis of the alkaline phosphatase superfamily; structural-based analyses of the alkaline phosphatase superfamily have helped in the past to obtain detailed information of the mechanism of action of proteins that are notoriously difficult to study *in vitro*, such as peptidoglycan phosphoglycerol transferase, and enzymes whose substrates tend to be complex and difficult to obtain [113,118,142,144,146,149,151,152,161–163].

Sequence-based phylogeny studies on bacterial phosphoethanolamine transferases have identified proteins likely to catalyse the addition of phosphoethanolamine to the

lipopolysaccharide lipid A in a wide range of Gram-negative bacteria [79]. Phylogenetic analysis allowed the identification and classification of two families of phosphoethanolamine transferases; the mobile colistin resistance family (MCR) and its ancestor, the Intrinsic Colistin Resistance from *Moraxella species* (ICR-Mo) [103,164–166], and the nonmobile Colistin Resistance (nMCR-1) family and its ancestor, a phosphoethanolamine transferase EptA from *Shewanella algae* [167,168]. While much work has been done to identify MCR-1 and other phosphoethanolamine transferases as members of the alkaline phosphatase superfamily [103,110,164,169–171]; none, to our knowledge, has investigated the relationship between the phosphoethanolamine transferases and the other subfamilies of the alkaline phosphatase superfamily. Moreover, despite the current information regarding the structure and conservation of the active site, the metallic ions present and their position in the catalytic site; the MCR reaction mechanism, as well as that of other PEtN transferases, has not been fully elucidated [72,108]. Comparison of PEtN transferase active sites with those of AlkP superfamily members whose catalytic mechanisms are better understood may improve understanding of the MCR reaction mechanism.

In this study we used both approaches, structure-based and sequence-based phylogenetic analysis, together with structural comparisons, to investigate the relationship between the MCR enzymes and the wider alkaline phosphatase superfamily. The aim was to understand the place of MCR-1 among the members of this superfamily of enzymes and to shed some light on the architecture of the MCR-1 catalytic site and the presence of a potential secondary metal ion binding site in the catalytic pocket, present not only in MCR-1 but other phosphoethanolamine transferases as well.

3.2 Results.

3.2.1 Phosphoethanolamine transferases as members of the alkaline phosphatase superfamily.

A starting overview was obtained by identifying the alkaline phosphatase superfamily in existing databases aiming to classify proteins based on structure. The CATH data base (<https://www.cathdb.info>) identifies domains in protein structures from the Protein Data Bank and classifies them into evolutionary superfamilies [129] using a hierarchical domain classification at four levels: Class Level (C), Architecture level (A), Topology Level (T) and Homologous Superfamily level (H); with the Class level being the most general and the Homologous Superfamily the highest level of classification. This classification is based on protein three-dimensional structures and topology [128,172,173]. The CATH domain classification locates MCR-1 as a member of the alkaline phosphatase superfamily based on the comparison of secondary structure elements, and identifies 420 structurally characterised domains as members of the alkaline phosphatase superfamily, with a total of 28 unique enzyme classifiers (ECs). Superposition of these 420 domains (Figure 14 a) identifies both the conservation of the core $\alpha/\beta/\alpha$ fold and the diversity of extensions, insertions and additions to this. Within the alkaline phosphatase superfamily (3.40.720.10) CATH identifies six distinct structural clusters each containing between one and six different functional families. The first cluster includes enzymes such as 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, arylsulphatases A and E and Phosphonoacetate hydrolase; the second cluster phosphoethanolamine transferases, including MCR, and arylsulphatases; the third cluster the ectonucleotide pyrophosphatase/phosphodiesterases; the fourth and fifth structural clusters are formed by different groups of alkaline phosphatases and related enzymes; while the sixth cluster contains phosphonate monoester hydrolases.

The SCOP (Structural Classification of Proteins, <https://scop.mrc-lmb.cam.ac.uk>) [174] database organises proteins by structural and evolutionary relationships at the level of classes, folds, superfamily and family. There is no entry for MCR-1, nor the PDB ids 5LRN or 5LRM nor MCR-1's Uniprot Id A0A0R6L508 (MCR1_ECOLX) on the database, however, upon searching for the "Alkaline phosphatase-like" term nine families were listed. SCOP divides the alkaline phosphatase superfamily (SCOP ID 3001353) into nine distinct functional families listed as: alkaline phosphatase (SCOP ID 4000782), sulphatases (SCOP ID 4000785), phosphonoacetate hydrolases (SCOP ID 4001199), DeoB-like (Phosphopentomutase-like, SCOP ID 4003235), 2,3-bisphosphoglycerate-independent phosphoglycerate mutases (SCOP ID 4003438), phosphonate monoester hydrolase-like (SCOP ID 4004457), AcpA-like (acid phosphatase-like SCOP ID 4004515), PbgA periplasmic domain-like (inner membrane YejM-like, SCOP ID 4005443) and archaeal 2,3-bisphosphoglycerate-independent phosphoglycerate mutases (SCOP ID 4007429).

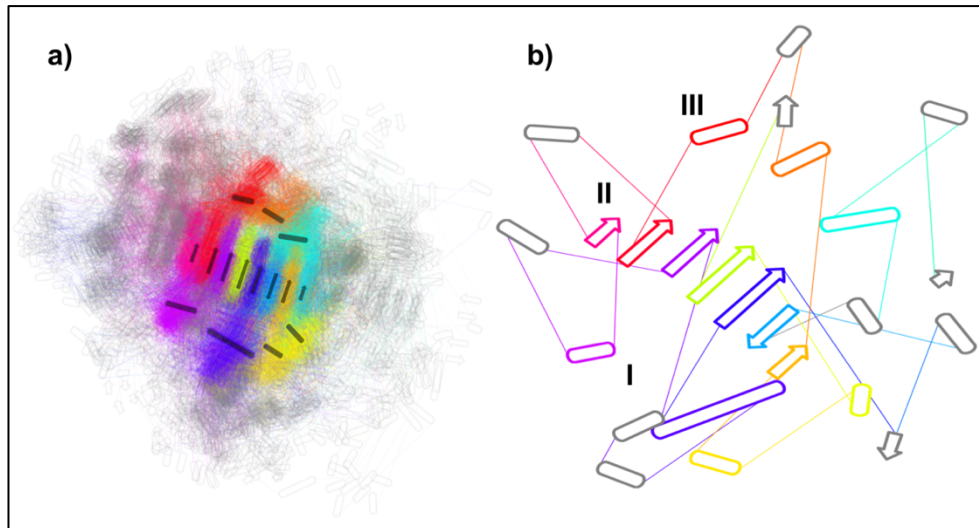


Figure 14: Characteristic $\alpha/\beta/\alpha$ fold of the alkaline phosphatase superfamily and MCR-1.

a) Superposition of 420 members of the alkaline phosphatase superfamily highlighting the $\alpha/\beta/\alpha$ fold. The superposition of the domains found in the alkaline phosphatase superfamily provides an indication of the relative distance, position and conservation of secondary structure elements characteristic of the $\alpha/\beta/\alpha$ fold. b) the $\alpha/\beta/\alpha$ of MCR-1 (PDB Entry 5LRM) is formed by 7 parallel and one antiparallel β -sheet structures (II) between two layers (I and III) of α -helices [129]. Figure generated by the 2DProts database available through the CAT webpage (<https://www.cathdb.info/>).

3.2.2 Structural analysis using the Dali Server.

Investigation of the position of MCR-1 and related enzymes within the alkaline phosphatase superfamily began with identification of relevant protein structures. A list of structures of candidate members of the alkaline phosphatase superfamily was assembled from the Protein Data Bank (PDB) using structure-based searches. The MCR-1 catalytic domain structure (PDB entry 5LRN) was used as a query to search the PDB at different levels of redundancy using the DALI server [175]. A total of 616 structurally similar hits were obtained from a search of the complete PDB, with the highest Z-score value of 60 and the lowest value of 2. The “Z-score” is a measure of the quality of the alignment between two structural superpositions provided by DALI server [175]. These structurally similar hits includes significant redundancy due to multiple chains, structure determinations of the same protein by different groups or in slightly differing conditions (e.g. multiple matches to other MCR-1 structures (i.e. 5LRN chain B, or structure 5K4P)), point mutants, and structures of very close homologues. However, the

server also searches subsets of the PDB at 3 tiers of sequence similarity; the first tier PDB25 offers a list of structures with a 25% or less sequence similarity to the query structure, with the next tiers, named PDB50 and PDB90, filtered at 50% and 90% sequence similarity respectively [175,176]. After input of the of the 5LRN structure to DALI searches of the subsets (PDB25, 50 and 90) were compared to identify a list of structural neighbours, ranked by Z-score, with which to carry out analysis of MCR-1. Further analysis required the use of an “all-against-all” analysis on DALI server; this come with the restriction of only allowing for 64 structures at the time; for this reason all similar it was needed to select sets of structures that covered all clades while excluding dissimilar hits.

The DALI Z-Score measures the quality of the structural alignment, meaning that the higher the Z-score the closer the similarity between the compared structures. A Z-score value of 2 or above can often be considered indicative of structural similarities, whereas a higher Z-score of 8 can be interpreted as a very good structural superimposition between the compared structures. Z-scores of between 8 and 20 mean that two structures are probably homologous and a Z-score above 20 means that two structures are definitely homologous, homologous proteins often share significant functional similarities [175]. Any Z-score below 2 is considered not significant [171,173]. Homologous proteins tend to be more structurally conserved and present higher Z-score values over analogous structures [133]. While members of the alkaline phosphatase superfamily tend to present low levels of sequence homology, their structures and active site architecture are highly conserved [143–145]. It is worth mentioning that these results are being highly influenced by a structural characteristic of the alkaline phosphatase superfamily, the $\alpha/\beta/\alpha$ fold [142–144] (Figure 14). It has been noticed that comparisons involving structures with this particular fold tend to have higher Z-scores than other conformations [175]. Using the measurement of Z-score as a filtering criterion allows us to easily identify key structural neighbours and shed some light on the relationships between

MCR-1 and other phosphoethanolamine transferases and the wider alkaline phosphatase superfamily.

3.2.3 Analysis of the Dali PDB25 data set.

After obtaining an exhaustive list of more than 600 structural matches against the MCR-1 catalytic domain structure 5LRN, systematic comparisons started by using the significant matches from a search against the PDB25 subset in an “all-against-all” comparison to obtain an overall view of the landscape of possible matches and identify likely thresholds for assignment to the alkaline phosphatase superfamily. The PDB25 data set is a list of nonredundant structures that has been filtered using a 25% sequence identity criterion. Filtering the data obtained from the Dali PDB search was necessary due to the limitations of the analytical methods used to analyse it, one of these being a restriction to 64 of the number of structures to analyse using the “all-against-all” pairwise comparison on the Dali server. As a result of the “all-against-all” comparison, DALI generates a dendrogram based on the pairwise similarity (Z-score) matrix and calculates a correspondence analysis (CA) which works as a multidimensional scaling method positioning proteins with the most similar structural neighbourhoods near each other. CA is a statistical technique that allows the visualization of the relationship between different groups of similar protein structures by positioning the most similar structural clusters neighbours closer to each other while the most dissimilar are far away from each other. The structure-based dendrogram of MCR-1 (PDB 5LRN) and the top 63 hits from the PDB25 subset of structures (Figure 15) showed a clear division of the structures into two clades (with separation between 2D1G and 1XJQ), cross-referencing each entry against the RCSB database confirmed that the structures in the top half of the dendrogram belong to the alkaline phosphatase superfamily; this also confirmed the low structural similarity in the low scoring matches. In the correspondence analysis (Figure 16) a clear separation can also be seen between the two groups, with all the structures with

high structural similarity to the alkaline phosphatase superfamily being close to the x axis with the rest of the structures (blue circles) being more disperse.

This division is also confirmed by the differences in their Z-scores; all of the structures identified as alkaline phosphatase superfamily members presented Z-scores in the range of 15 at the lowest and 25 at the highest value, whereas for the rest of the structures this measure drops to 7.4 at the highest Z-score of the non-alkaline phosphatase structures, with 3.8 the lowest Z-score obtained. These lower Z-scores are associated with the length of the aligned portions of the comparator structures with 5LRN, with clearly lower lengths of alignments in the case of the proteins identified as outside the alkaline phosphatase superfamily (Appendix L).

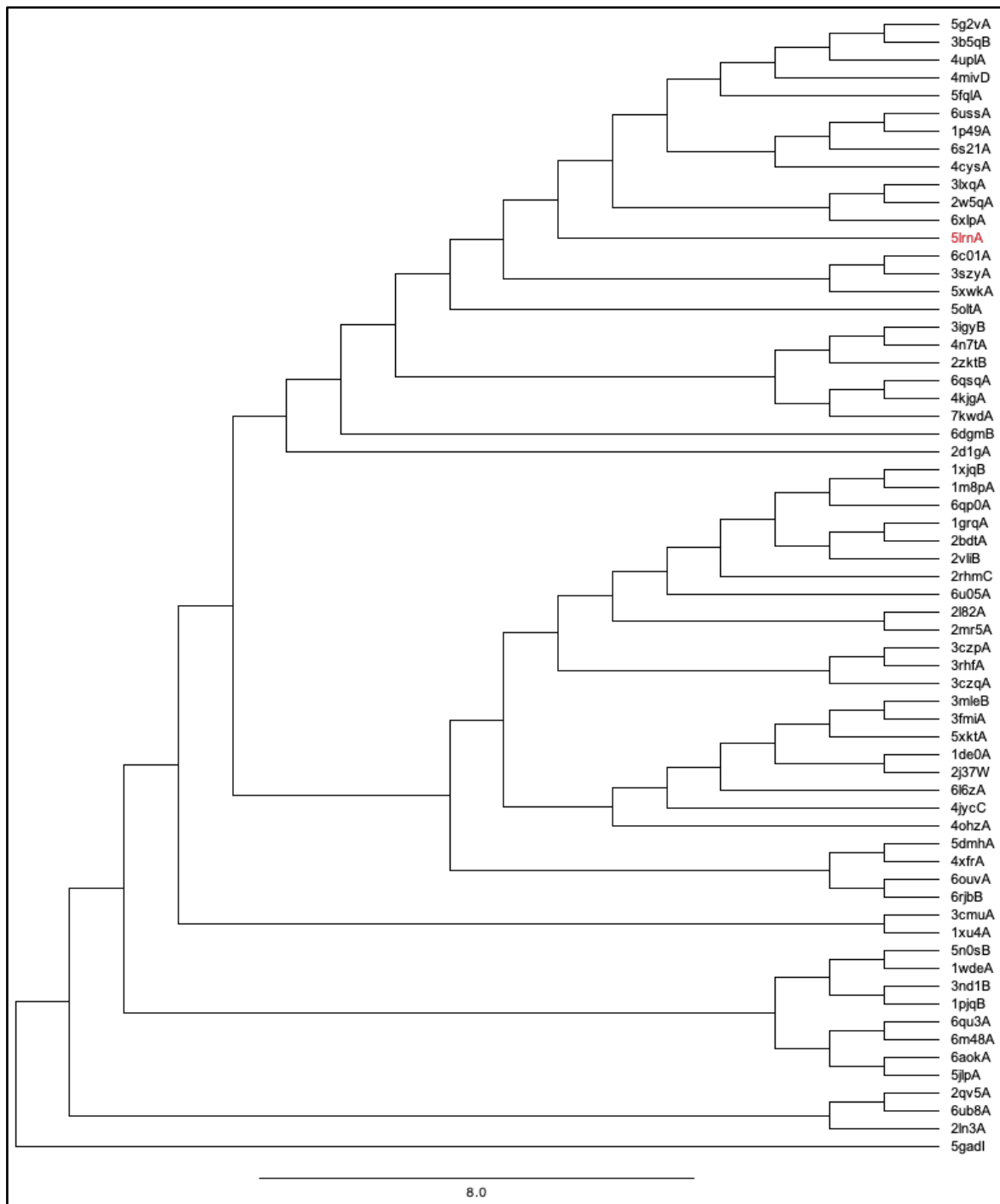


Figure 15: Structural dendrogram of the MCR-1 catalytic domain and 63 closest matches from the PDB25 dataset calculated by Dali.

Two clades can be observed. The upper clade contains members of the alkaline phosphatase superfamily as well as MCR-1 (PDB entry 5LRN labelled in red). The clade separation can be observed at structures 2D1G, at the bottom of the alkaline phosphatase clade, and 1XJQ at the top of the non-alkaline phosphatase clade.

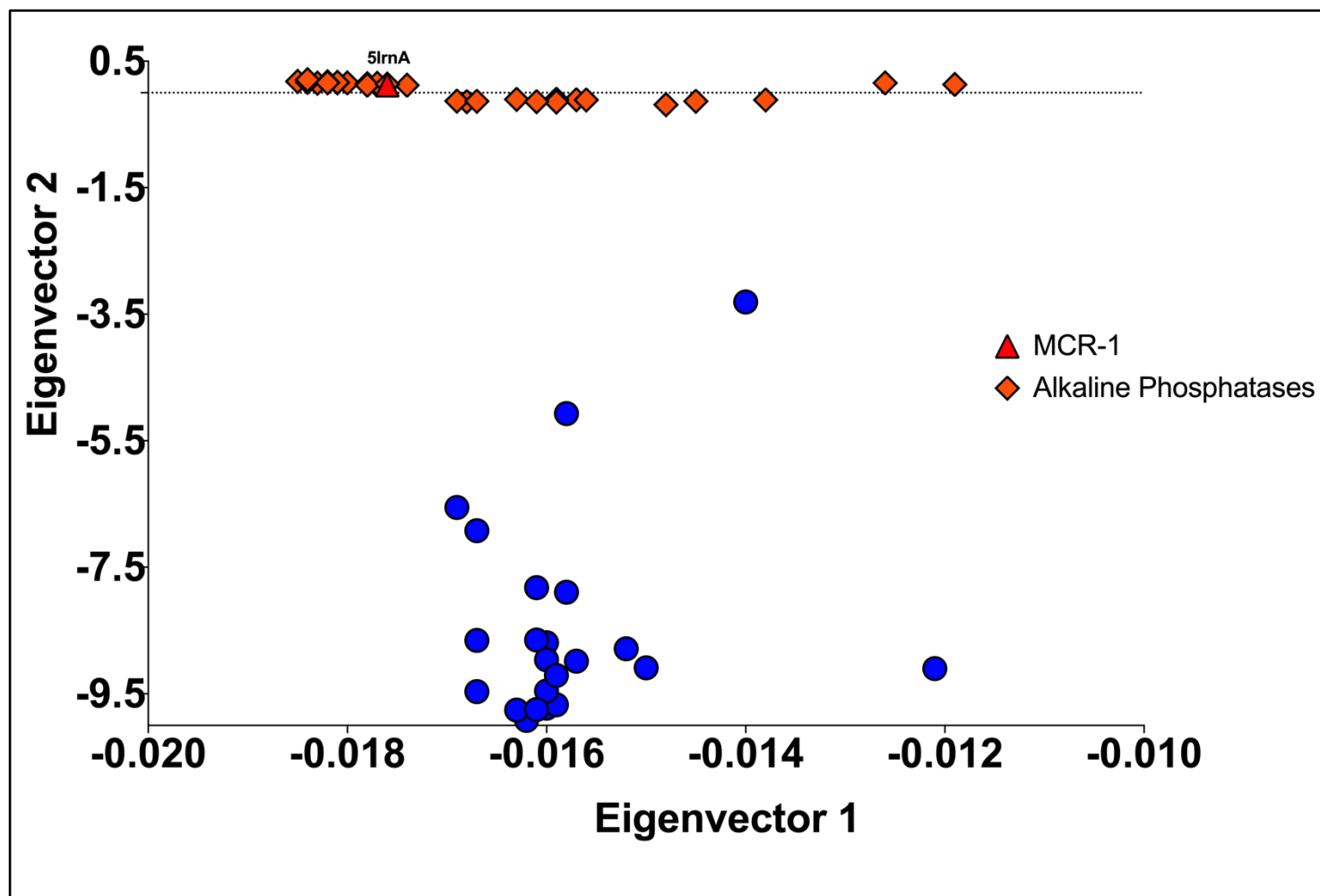


Figure 16: DALI Correspondence analysis of the MCR-1 catalytic domain and 63 closest matches from the PDB25 dataset.

Correspondence analysis (CA) obtained for the PDB25 dataset from DALI server. MCR-1 (5lrnA red triangle) can be found in the upper part of the plot, inside the group of structurally similar the alkaline phosphatases (orange diamonds). Blue circles are belong to all the other structural matches that are not identified as members of the alkaline phosphatase superfamily.

3.2.4 Selection and DALI Analysis of a set of representative AlkP structures.

The results from the analysis of the PDB25 dataset, using the “all-against-all” function on Dali, clearly showed that MCR-1 is structurally part of the alkaline phosphatase superfamily and distinct from other groups of proteins with more distant structural resemblance, however, for a deeper insight into the relationship between MCR-1 and the alkaline phosphatase superfamily; the lists of matches obtained from the PDB50 and PDB90 subsets (lists of matches at 50 % and 90 % identity, respectively) were inspected. The PDB90 dataset yielded a total of 109 structural matches; 90 of these structures presented Z-score values above 10. The DALI search of the MCR-1 catalytic domain (PDB 5lrn) against the PDB50 subset delivered a list of 72 structural matches with Z-scores above 2. The distribution of Z-scores obtained is shown in Figure 17. Superposition of 5LRN on itself gives a Z-score of 60.8, with the next four highest-scoring structures all identified as phosphoethanolamine transferases. An average Z-score of 20.1 is then maintained across structures between the 5th and 60th positions; this suggests high homology among these structures. It is important to mention that nearly 40% (28 out of the 72 structures) were found to be members of the sulfatase functional family. A clear drop in the Z-score values can be seen between the 61st and the 62nd positions; with the structure at the 61st position (PDB 3NKM) having a Z-score of 10.1 and the structure at the 62nd a Z-score of 4.2. The 61st structure, 3NKM identified as an alkaline phosphatase superfamily member during the PDB25 data set analysis (Appendix M), served as a cut-off point for structures selected for further “all-against-all” comparison. From the PDB50 dataset MCR-1 (PDB 5lrn) plus the top 61 structures up to and including 3NKM, were selected for pairwise comparison, with a further two structures, PDB entries 6PD0 and 6V8Q, added to enlarge the set to the maximum 64 structures allowed by the server. These proteins were selected due to their roles in in phosphoethanolamine transfer (PDB 6pd0 [178]) and in Lipid A binding (PDB 6V8Q [179]).

The list of the top 63 structural matches against MCR-1 from the PDB90 dataset was also inspected. The structural dendrogram obtained from an all against all pairwise comparison corroborates the division of the superfamily into two major clades, the Alkaline phosphatases and the Arylsulphatases/sulphatases (Appendix O and Appendix P) and confirms that no major subclades are omitted at the level of 50% sequence identity. However, the PDB90 list omits structures from the main alkaline phosphatase clade and so was not analysed in detail).

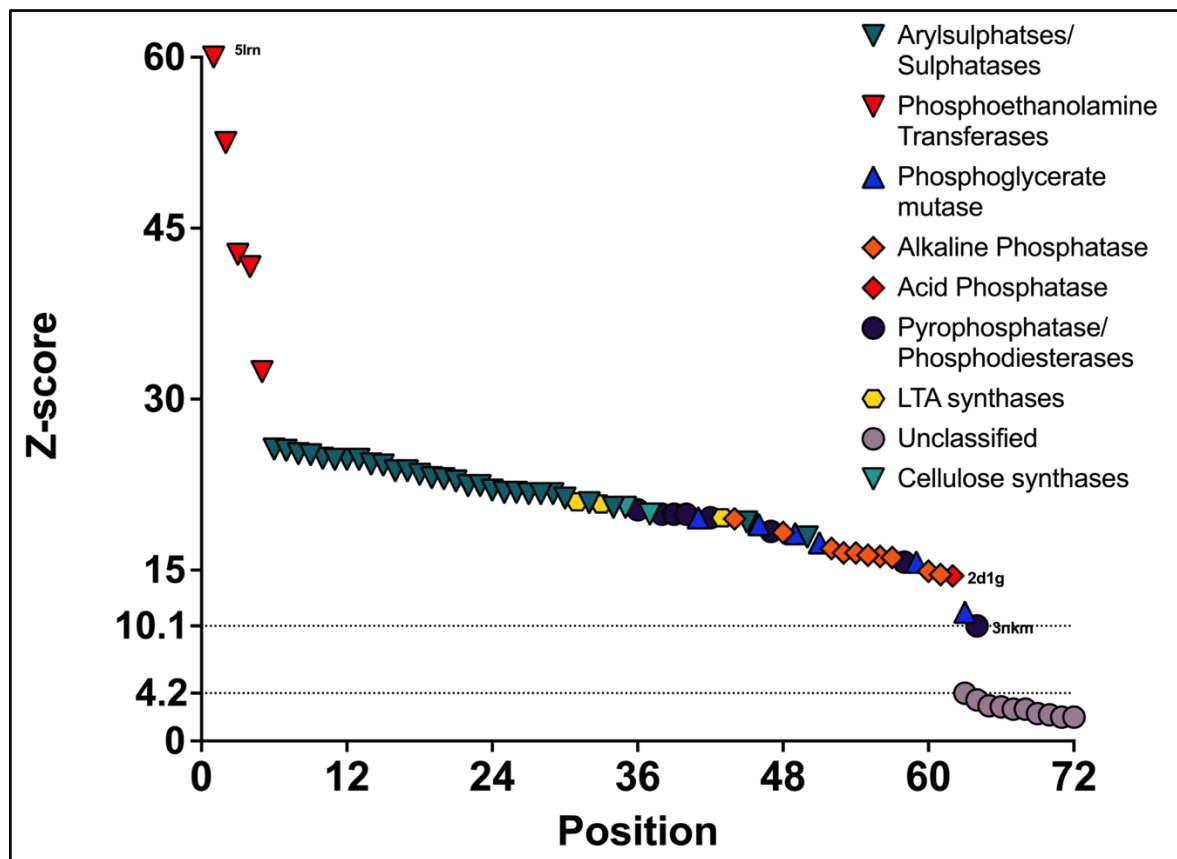


Figure 17: Z-score distribution of the 72 structures identified from a DALI search of the PDB50 data set against the MCR-1 catalytic domain (PDB 5lrn).

Figure shows Z-scores calculated by DALI for comparison of each structure with the MCR-1 catalytic domain (PDB 5LRN). Structures are colour-coded according to their functional family inside the Alkaline phosphatase superfamily. After the phosphoethanolamine transferase group (red triangles), the arylsulphatase/sulphatase group is the one with the highest structural similarity and the highest number of structures (28 out of the 72).

DALI was then used to undertake pairwise comparisons of this set of 64 proteins in an “all-against-all” analysis. A heat map derived from the set of pairwise Z-scores is shown in Figure 18, a structure-based dendrogram is shown in Figure 19 and a two-dimensional correspondence analysis calculated by DALI is shown in Figure 20. As ordered in the heat map visualisation (Figure 18) and dendrogram (Figure 19) the set of proteins can be divided into two groups typified by the alkaline phosphatases (Figure 18 group II) and the sulphatases/arylsulphatases (Figure 18 group I). These are separated on the heat map by structure 3NKM, mouse autotaxin (Enpp2) a secreted lysophospholipase D, i.e. with ectonucleotide pyrophosphatase/phosphodiesterase (ENP/NPP) activity [180]. This structure has the lowest Z-score against MCR-1 for any member of the set of enzymes, with a value of 10.1, and is clearly the most distant relative of most of the proteins in the group. The relatively low Z-scores for 3NKM can in part be due to the size of this enzyme (>805 residues) since Z-scores can be influenced by the size of the protein [175]. For this protein higher Z-scores are obtained for only a subset of other ENP/NPPases; PDB 6C01 (Z-score of 43.1 when compared against 3nkm), 6AEK (Z-score of 42.0), 5VEM (Z-score of 26.8), 2GSN (Z-score of 26.5), 5EGH (Z-score of 25.5) and 5TCD (Z-score of 24.1).

Inside these two major groups a number of other distinct clusters can be identified on the heat map. Inside the alkaline phosphatases, we have the “true alkaline phosphatase” subgroup (AlkP) and the phosphoglycerate mutases (PGM). The Nucleotide Pyrophosphatase (ENP/NPP) group seems to work as a bridge, with Z-scores indicating similarity of its members to both alkaline phosphatases (II) and the sulphatases/arylsulphatases (I). Within the sulphatase/arylsulfatase group distinct sets of proteins include the phosphoethanolamine transferases (PEtNs, the location of MCR-1 (PDB 5lrn) has been labelled in red), the lipoteichoic acid synthases (LTAs), the arylsulphatases (AS) also found in the dendrogram clade containing structures 6BJ0A to 6XLPA (Figure 19) and the phosphonate monoester hydrolases (PMH). These identified groups resemble the subfamilies found on SCOP

database; it is worth mentioning that from the nine subfamilies found in SCOP, several have a low representation with only one or two structure(s) for each: the AcpA-like phosphatases (SCOP ID 4004515, represented by structure 2D1G); the archaeal 2,3-bisphosphoglycerate-independent phosphoglycerate mutases (SCOP ID 4007429, represented by structures 3KD8 and 2ZKT); PbgA inner membrane YejM-like (SCOP ID 4005443, represented by structures 6V8Q and 6XLP); and the DeoB-like phosphopentomutase (SCOP ID 4003235, represented by structure 4N7T).

Structure 2D1G, an outlier on the dendrogram (and to a lesser degree the heat map analysis), is an acid phosphatase enzyme that, unusually, belongs to the alkaline phosphatase superfamily in which it forms a distinct branch (SCOP ID 4004515, [181]). The AcpA core domain is formed by an 8-strand beta-sheet flanked by two layers of three α -helices on either side, an architecture typical of the alkaline phosphatase superfamily's $\alpha/\beta/\alpha$ fold [141–143,181]. As mentioned before, in the SCOP database position AcpA (acid phosphatase-like SCOP ID 4004515) is identified as one of the nine subfamilies in the Alkaline phosphatase superfamily.

A clear difference between the structures in the two major clades is the number of metal ions present in the catalytic pocket. The arylsulphatase/sulphatase group, including MCR-1 and the phosphoethanolamine transferases, present mostly 1 metal ion in their catalytic pockets. The exceptions are structures 2QZU and 3LXQ, and the bacterial inner membrane protein PbgA (structures 6V8Q and 6XLP) that are not reported to contain any metal ions. Interestingly PbgA seems to contain the conserved Asp-His (D104 and H114 in 6V8Q) 178, however, it does not seem to require divalent cations for Lipid A binding 179. Meanwhile the alkaline phosphatase clade contains structures with 2 or more metal ion sites in their catalytic pockets, with the exceptions being the two proteins involved in cellulose synthesis by bacteria (structures 5OLT and 6PD0) and structure 6DGM (*Streptococcus pyogenes* GacA responsible

for addition of glycerol phosphate to cell surface carbohydrate [184]), all of which contain only a single metal ion in their respective active sites; and structure 3KD8 which does not contain any metal ions. 3KD8 is a putative cofactor-independent phosphoglycerate mutase (iPGM) from the thermophilic archaeon *Thermoplasma acidophilum* [185]. Archaeal iPGMs are a diverse group of enzymes in which 3KD8 is part of a broad grouping that also includes the enzymes from *Pyrococcus horikoshii* (2ZKT, containing two zinc ions) and *Achaeglobus fulgidus* (whose activity is reported as stimulated by Mn^{2+}). However, biochemical characterisation of this protein, and the metal dependence of its activity, is yet to be reported.

Two-dimensional correspondence analysis helped to better understand the relationship between MCR-1 and the other identified functional families. The analysis in Figure 20 groups the selected set of 64 structures of the alkaline phosphatase superfamily in the form of an inverted pyramid. At each corner we can see a major group: sulphatases at the top left, alkaline phosphatases at the top right and EPP/NPP at the bottom, forming three groups that are structurally distinct. At the top left side, near the sulphatases/arylsulphatases we can see several clusters together; the LTAs, the cellulose synthases and the PEtN transferases together clustering close to the sulphatases/arylsulphatases. These groups are not only structurally similar but all share similar biological activities, with the LTAs, the cellulose synthases and the PEtN transferases all being membrane-bound enzymes that modify bacterial saccharide substrates. Correspondence analysis positions the most similar structural neighbourhoods near to each other; this means that the clusters of structures that are closer to each other (e.g. how close PEtN transferases are to LTAs, sulphatases/arylsulphatases, and cellulose synthases in Figure 20) represent protein families related through stronger functional constraints. At intermediate distances the structural neighbours are related by shape similarity, however, this does not translate to similar biological functions (e.g. Phosphoglycerate mutases and the acid phosphatase 2DG1 in Figure 20); lastly, long distances between structural clusters (e.g. the Alkaline phosphatases and the EPP/NPP

enzymes in Figure 20) means that the structural similarity is dominated by secondary structure composition [175], in this case, the $\alpha/\beta/\alpha$ fold shared across all members of the alkaline phosphatase superfamily.

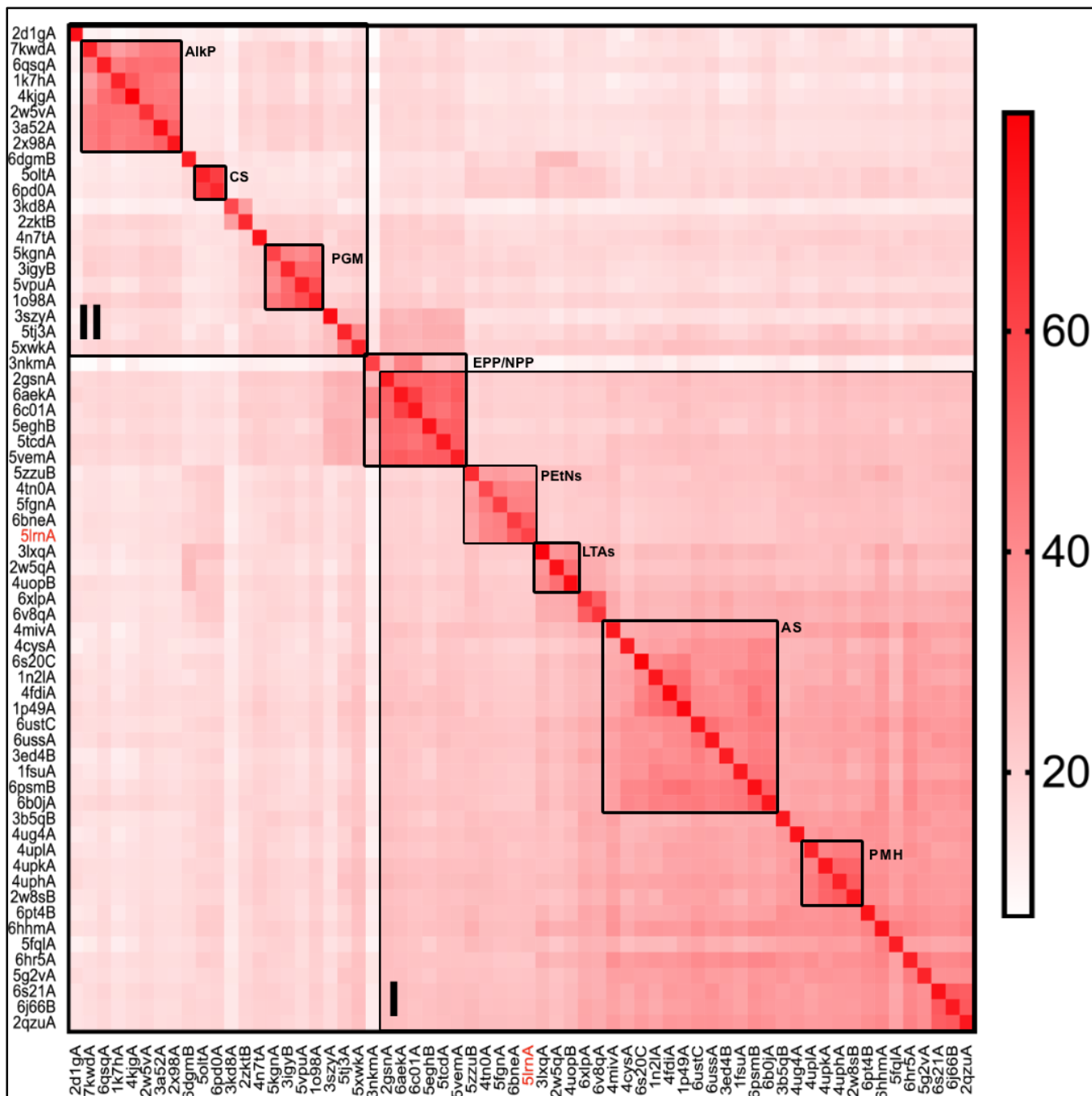


Figure 18: Heat map analysis of DALI pairwise comparison of alkaline phosphatase superfamily proteins.

The heat map was constructed using the Z-scores of the “all-against-all” analysis of the structures present in the PDB50 data set. The 6 families found through correspondence analysis have been marked by blocks. The first block (I) corresponds to the major Arylsulphatases/ Sulphatases group in which the phosphoethanolamine transferases (PEtNs, MCR-1 has been labelled in red), lipoteichoic acid synthases (LTAs), phosphonate monoester hydrolases (PMHs) and Ectonucleotide Pyrophosphatase/Phosphodiesterase (EPP/NPPs) are located. The second block (II) corresponds to the second major group, the Alkaline phosphatases which encompass the “true alkaline phosphatases” (AlkP), the Cellulose Synthases (CS) and the phosphoglycerate mutases (PGM). Positions on the map are coloured by Z-score scaled (scale at right) from 7 to 80.

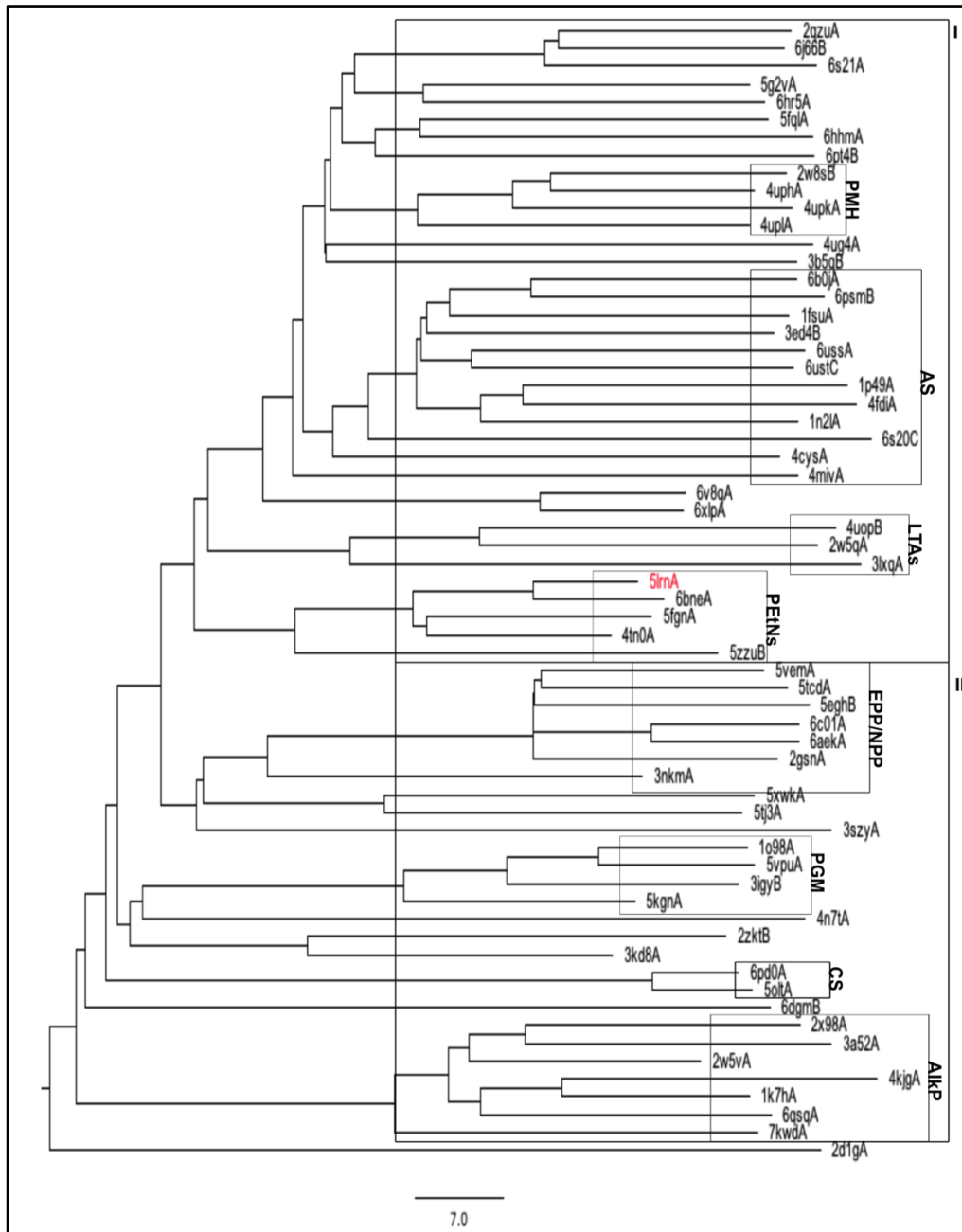


Figure 19: Structure-based dendrogram of alkaline phosphatase superfamily proteins based upon DALI pairwise comparisons.

Two main clades can be seen in a similar fashion as in the heatmap visualization; the first one (I) corresponding to the Sulphatases/Arylsulphatases containing and a second clade (II) corresponding to the Alkaline phosphatases. MCR-1 (5lrnA, labelled in red) and the phosphoethanolamine transferase (6bneA, 5fgnA, 4tn0A and 5zzuB) group lie inside the arylsulfatase/sulfatase clade. Labelled blocks are phosphonate monoester hydrolases (PMH), arylsulphatases (AS), lipoteichoic acid synthases (LTAs), phosphoethanolamine transferases (PEtNs, MCR-1 has been labelled in red), ectonucleotide pyrophosphatase/phosphodiesterase (EPP/NPPs), phosphoglycerate mutases (PGM), cellulose synthases (CS) and “true alkaline phosphatases” (AlkP).

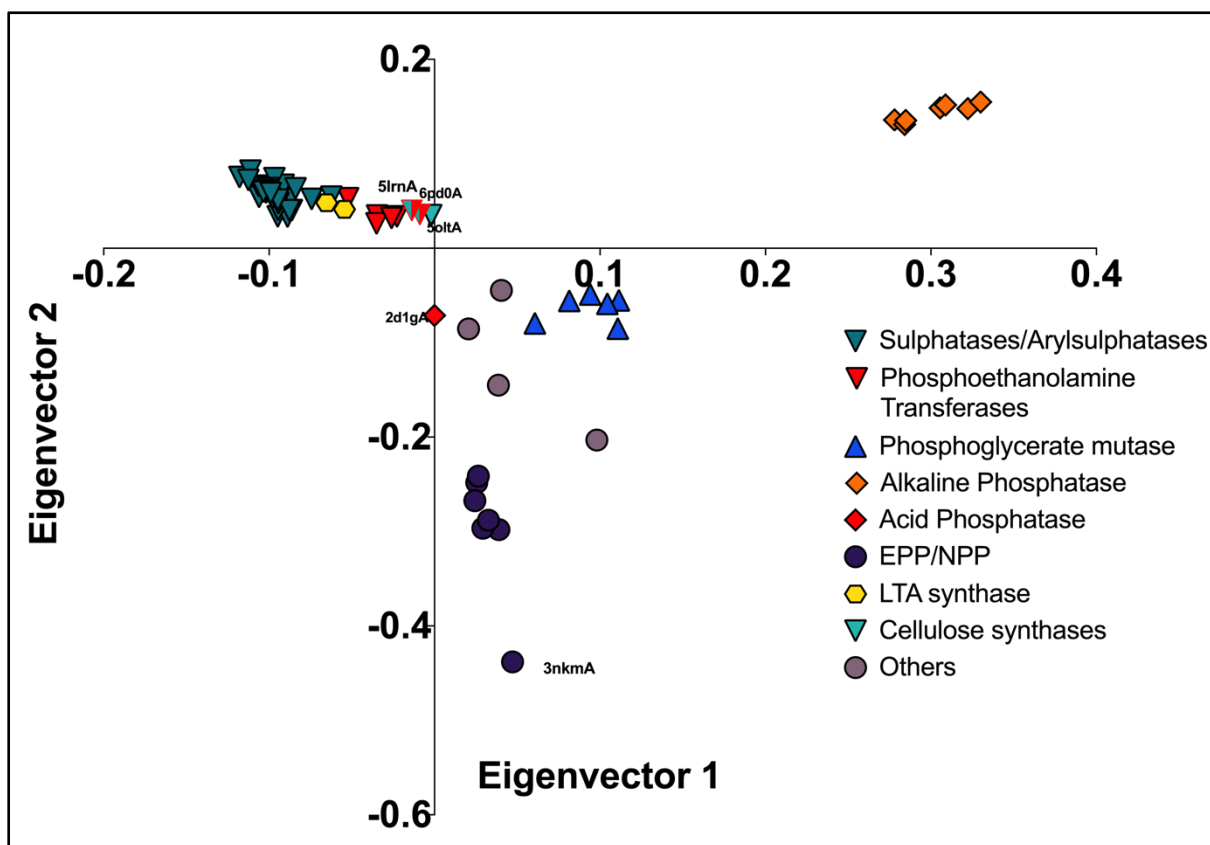


Figure 20: DALI Correspondence analysis of alkaline phosphatase superfamily proteins showing the principal groups belonging to the different families members of the alkaline phosphatase superfamily.

Principal groups are identified by colour/symbol. The phosphoethanolamine transferases can be found next to the LTA synthases and Sulphatases/Arylsulphatases. Structure 6pd0A (red and green triangle) is labelled as both a cellulase synthase and a phosphoethanolamine transferase [208]. Selected outlier proteins acid phosphatase (2dg1A) and the mouse autotaxin (3nkmA) are labelled.

A set of eight representative structures representing the main clades/groups: phosphoethanolamine transferases (PDB 5LRN), alkaline phosphatase (PDB 1K7N), cellulose synthase subunit G (BcsG, PDB 6PD0), lipoteichoic acid synthase (LtaS, PDB 2W5Q), phosphonate monoester hydrolase (PDB 2W8S), arylsulphatase (PDB 3ED4), nucleotide pyrophosphatase/phosphodiesterase (PDB 2GSN) and phosphoglycerate mutase (PDB 1O98); was selected for more detailed comparison. The respective Z-scores for the comparison of each structure with the MCR-1 catalytic domain are as follows: alkaline phosphatase Z-score of 14.6, cellulose synthase subunit G (BcsG) Z-score of 20.5,

lipoteichoic acid synthase LtaS Z-score of 19.6, phosphonate monoester hydrolase Z-score of 25.1, arylsulfatase Z-score of 22.4, nucleotide pyrophosphatase/phosphodiesterase Z-score of 19.9 and phosphoglycerate mutase Z-score of 19. As previously mentioned, a Z-score value of 20 or more between two structures means that these structures are homologous [175]; the higher scores for comparisons with the larger group of arylsulphatases/sulphatases (including phosphonate monoester hydrolase) indicates that MCR-1, and by implication other phosphoethanolamine transferases, is at the level of overall structure more similar to this group than to other enzyme classes within the alkaline phosphatase superfamily, as seen in Figure 17 and Figure 20. The most similar structures to MCR-1 (PDB 5LRN) are phosphonate monoester hydrolase, arylsulfatase and cellulose synthase subunit g (with respective Z-scores of 25.1, 22.4 and 20.5, Figure 21 c, e and f respectively). These three structures share the globular structure with $\alpha/\beta/\alpha$ topology typical of the arylsulfatase family [112,186]. The structure of the C-terminal domain of BcsG, which is demonstrated to have phosphoethanolamine transferase activity, has a high resemblance to MCR-1, and other phosphoethanolamine transferases, but the core $\alpha/\beta/\alpha$ fold is adorned by an additional large α -helical component and a small four-stranded antiparallel β -sheet at the protein C-terminus (Figure 21 c [187]). The representative structures with Z-scores lower than 20, present a more “Alkaline phosphatase”-like structure in which the $\alpha/\beta/\alpha$ fold is decorated with several large insertions (Figure 21 b, d, g and h and Appendix Q)

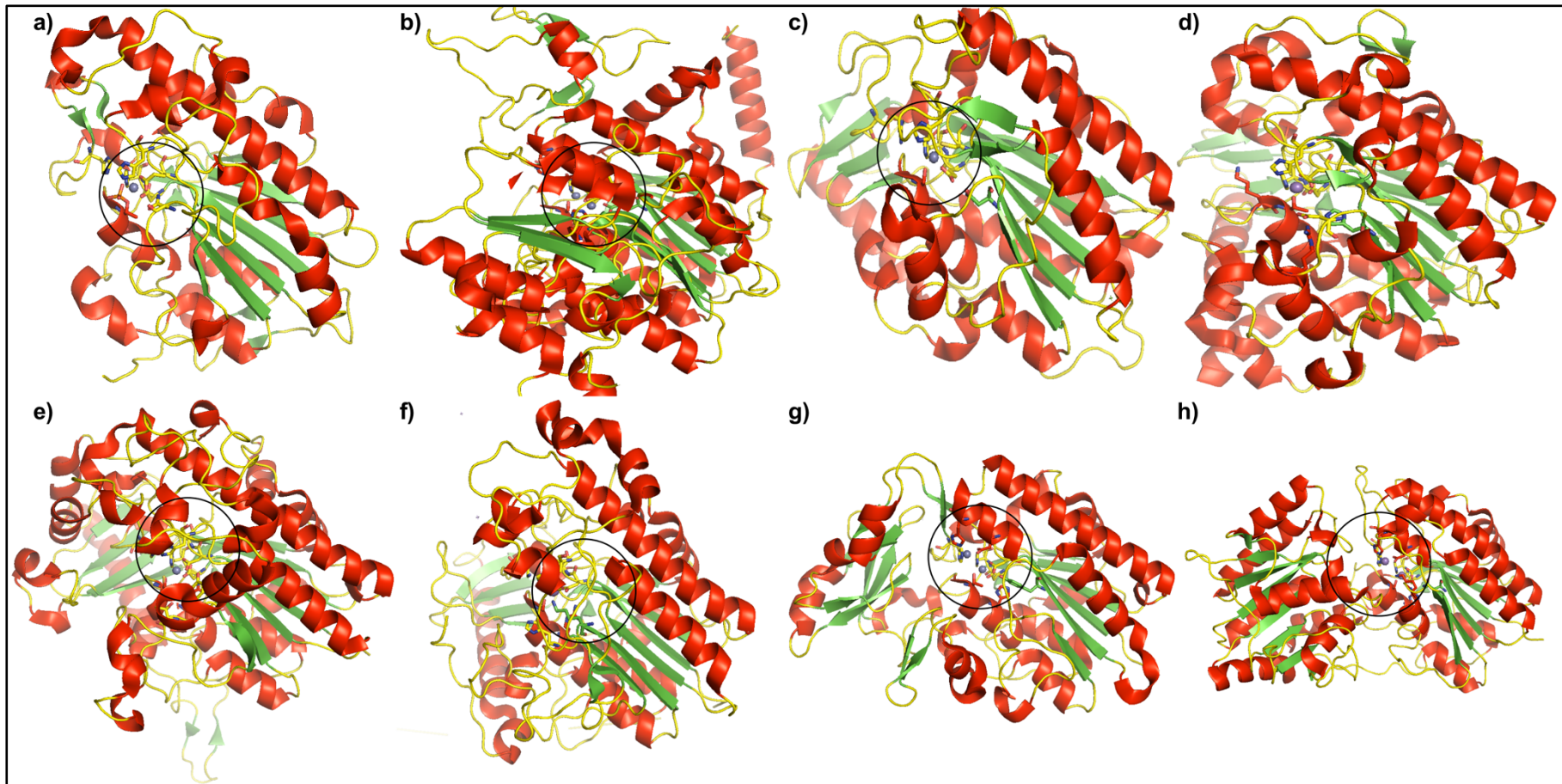


Figure 21: Structures of representative proteins from alkaline phosphatase subfamilies.

Structures of different members of the alkaline phosphatase superfamily: a) MCR-1, b) Alkaline phosphatase, c) Cellulose synthase subunit G (BcsG), d) Lipoteichoic acid synthase LtaS, e) Phosphonate Monoester Hydrolase, f) Arylsulfatase, g) Nucleotide Pyrophosphatase/Phosphodiesterase and g) Phosphoglycerate Mutase. The central $\alpha/\beta/\alpha$ fold is a highly conserved structural characteristic of the alkaline phosphatase super family. Pymol was used to generate and colour code the structures; α -helices are coloured in red, β -sheets are coloured in green and loops and coils are coloured in yellow. The catalytic pocket of each protein has been circled. PDB entries a - g in order: 5LRN, 1K7H, 6PD0, 2W5Q, 2W8S, 3ED4 2GSN and 1O98.

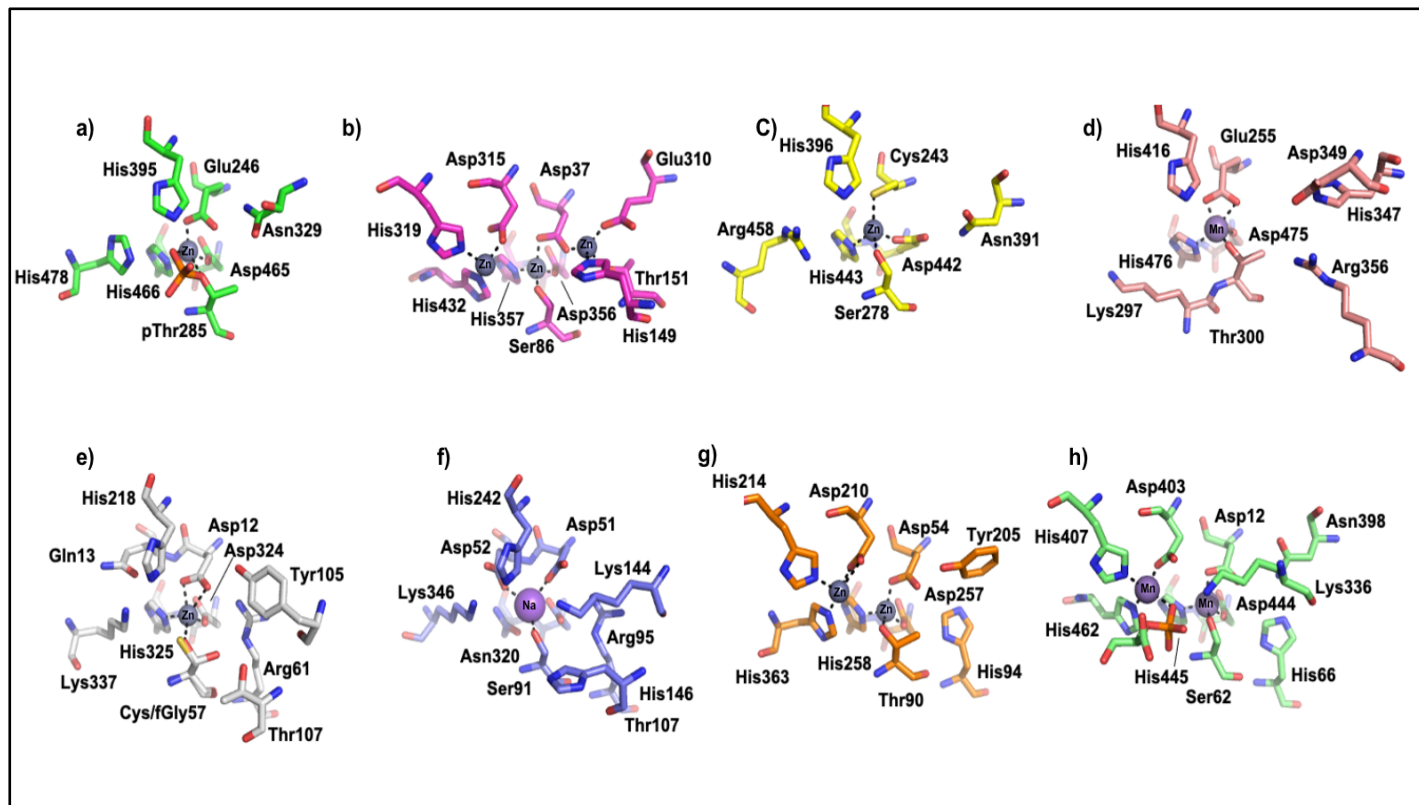


Figure 22: Catalytic sites of the 8 representative structures of alkaline phosphatase subfamilies.

Catalytic sites of a) MCR-1, b) Alkaline phosphatase, c) Cellulose synthase subunit G (BcsG), d) Lipoteichoic acid synthase LtaS, e) Phosphonate Monoester Hydrolase*, f) Arylsulfatase, g) Nucleotide Pyrophosphatase/Phosphodiesterase and h) Phosphoglycerate Mutase. The MCR-1 catalytic site is formed by six highly conserved amino acids. Glu246, Asp465, His466 and the nucleophile Thr285 interact with one zinc ion. These amino acids, can be observed in the catalytic sites of other monometallic enzymes with only the nucleophile being changed: c) Ser278, e) Cys/fGly57, f) Ser91). Equivalents of His395 and His478 can be observed in Alkaline phosphatases with multiple metal ions in their active site, however in these structures a third amino acid, aspartic acid, is present: b) Asp315, f) Asp52 and h) Asp403). The software Pymol was used for the elaboration of this figure. Zinc (Zn) is represented by metallic grey spheres, Manganese (Mn) is purple and Sodium (Na) is lilac spheres respectively; oxygen atoms are coloured red, nitrogen blue and sulphur yellow. Residue numberings are according to the PDB files. *Structure 2W8S presented a mixture of Cys and fGly according to [186].

The active site architecture and metal contents of representative members of the arylsulphatases/sulphatases group (5LRN, 6PD0, 2W5Q, 2W8S, 3ED4) and the alkaline phosphatase group (1K7H, 2GSN and 1O98) are compared in Figure 21. As mentioned above, the arylsulphatases/sulphatases generally present one metal ion in their catalytic pockets whereas the alkaline phosphatases present two or more metal ions. Of the eight structures, the sulfatase/arylsulfatase group and related structures (5LRN, 6PD0, 2W8S, and 3ED4) contain only 1 metal ion in (Zn), manganese (Mn) or sodium (Na) in their catalytic pocket. Three amino acids of this metal binding site (in MCR-1, Glu246, Asp465 and His466, Figure 22 a), which is present in all eight structures, are highly conserved across these enzymes. There are some points of variation: Glu246 is replaced by Cys432 in BscG, structure 6PD0 (Figure 22 c) and the Asp465 - His466 pair is replaced by an Asp319 - Asn320 pair in the arylsulfatase structure 3ED4 (Figure 22 f)). The fourth metal ligand is the acceptor residue for the covalent (phospho-, sulfo-, phosphono-) intermediate: threonine (phosphorylated-Thr285 in MCR-1), serine (Ser278) in 6PD0 and (Ser91) 3ED4, and formylglycine, a post-translationally modified cysteine (fGly57), in 2W8S (Figure 22 a, c, d, e and f). This formylglycine residue is present in the active site of all characterized sulphatases, both prokaryotic and eukaryotic, from the alkaline phosphatase superfamily, and is essential for catalytic activity [112,188].

Alkaline phosphatase (1K7H), EPP/NPP (2GSN) and PG mutase (1O98) all contain di-metallic centres in their catalytic sites. Where present the second zinc site is formed by a closely spaced Asp – His pair (in 1K7H Asp315 and His319, in 2GSN Asp210 and His 214, and in 1O98 Asp403 and His407); and an additional His residue (in 1K7H His432, in 2GSN His 363, and in 1O98 His462 (Figure 22 b, g and h). The distance between this metal ion pair is similar in the different structures: the two zinc atoms in the catalytic site of 2GSN are 4.26 Ångströms (Å) apart and the two manganese ions inside the 1O98 site being 4.89 Å apart. In the alkaline phosphatase structure 1K7H the distance between the equivalent zinc ions is 4.47

Å, this structure, as is the case for many alkaline phosphatase structures, contains three metal ions in its catalytic site [117–119], with a third zinc ion 4.53 Å distant from the first (Asp – Ser – Asp - His) site. The presence of three metal binding centres seems to be conserved among pure alkaline phosphatases, manifesting in both bacterial, such as ECAP, and in eucaryotic alkaline phosphatases like Shrimp alkaline phosphatase (SAP, PDB entry 1K7H) and human placental alkaline phosphatase (PLAP) [189]. Although it has been noticed that these third metal binding sites are needed to carry out catalytic activity [190,191], there is still some controversy regarding the presence metal ions with as it has been noted that both zinc and magnesium can be present in the third metal binding site of ECAP, PLAP and SAP; with however, the presence of zinc lowers the enzyme activity [189]. It has been suggested that the presence of zinc, in the third metal binding site, regulating the enzyme catalytic activity, keeping the enzyme in a conformation that has high affinity for zinc but is hindering its activity, when there are lower physiological concentration of magnesium [192].

The MCR-1 catalytic domain structure shown in Figure 22 a (PDB 5LRN) contains a single zinc ion. However, currently available crystal structures of the soluble domain of MCR-1 contain 1 (e.g. PDB entries 5LRN and 5YLF 72,189), 2 (PDB entry 5LRM 72) or up to 5 zinc ions when crystallized under high zinc conditions (PDB entry 5K4P 107). Six amino acids comprise the MCR-1 catalytic site; Glu246, Thr285, Asp465 and His466, interacting with the first zinc ion, while structures where MCR-1 is seen with 2 or more zinc ions involve His395 and His478 interacting with a second zinc ion, whose coordination is normally completed by involvement of Glu300 from a second MCR-1 chain in a dimeric structure formed by crystal packing [190].

3.2.5 Analysis of metal ion binding sites from structure-based sequence alignment.

DALI searches of the PDB also yield a set of pairwise sequence alignments for matching regions generated from structural overlays. An alignment of MCR-1 and the set of 63 structures used in the “all-against-all” structure comparison, omitting gaps and insertions from regions of the structures that do not align, is shown in Figure 23. This structure-based alignment shows the amino acids of the first metal binding site (Glu246, Asp465 and His466 in MCR-1, green highlighted residues in Figure 23) to be highly conserved across the 64 structures. Notably, His is replaced by Asn or Gln in some sulphatase structures (e.g. 3ED4, see above). The exceptions are the presence of Cys, in place of Asp or Glu, at the first position in the case of the cellulose synthases 5OLT and 6PD0 (BcsG, see above) and the absence of the Asp-His pair in 6XLP, 6V8Q and 3KD8. The structures of the PbgA proteins 6XLP and 6V8Q are phosphoethanolamine transferase-like (structurally similar to EptA) however the presence of metal ions in their catalytic sites has not been reported, implying that these are not metalloenzymes [179]. The fourth metal ligand, the acceptor residue for the covalent intermediate (threonine, (phosphorylated in the MCR-1 structure 5LRN), serine or formylglycine) can also be seen present in all structures but two, 6AEK and 3KD8. In the case of structure 6AEK, an EPP/NPP enzyme, the catalytic threonine (Thr238) residue was replaced with alanine for crystallization purposes [195]. Biochemical characterisation of the activity of 3KD8 has not been reported, but the absence both of His from the Asp – His pair and of any acceptor residue (Ser, Thr or fGly Figure 23), and the lack in the structure of suitable alternative metal binding residues, indicate that this putative enzyme is unlikely to be metal-dependent. These residues, the Asp-His pair, are important for metal binding as mutation of His-412 with Asn, in *E. coli* alkaline phosphatase, showed weak affinity for zinc binding [190].

The alignment also shows that from the second metal binding site, His395 is well, but incompletely, conserved among the 64 structures. In many alkaline phosphatases it is substituted by Asp (Asp315 in 1K7H, Figure 22), it is replaced by Asn in 6XLP and 6V8Q, is missing from 3LXQ, and replaced by Val in 5KGN and Thr in 3KD8. The PbgA enzymes 6XLP and 6V8Q are not reported to bind metal ions (see above), while 3LXQ is an uncharacterised protein from *Vibrio parahemolyticus* whose structure was determined as part of a structural genomics project and whose structure contains no metal ions. In 5KGN (cofactor-independent phosphoglycerate mutase from *C. elegans*) the second metal centre is intact with an Asp residue at the equivalent position to MCR-1 His395, but the structural alignment is complicated by two preceding proline residues. As detailed above, a metal site is unlikely to be present in structure 3KD8.

Enzymes containing two metal ions in their catalytic site; such as the alkaline phosphatases, EPP/NPP and phosphoglycerate mutase, share a particular positioning of the amino acids required for the interaction with the second metal ion; specifically an aspartic acid residue followed by a histidine positioned 4 amino acids apart in sequence, as seen on Figure 22 b) g) and h) (residues Asp315-His319 in b, Aps210-His214 in g and Asp403-His407 in h). This Asp is not present in arylsulphatases/sulphatases and is not found in phosphoethanolamine transferases. The second amino acid of this pair, His478 in the case of MCR-1 (Figure 22 a), is only present in enzymes from the alkaline phosphatase clades, including EPP/NPP and phosphoglycerate mutase, and in some phosphoethanolamine transferases including MCR-1 (Figure 25). Residues able to bind metal ions are absent from this position in enzymes from the arylsulphatase/sulphatase groups, where a basic residue (Lys or Arg) is normally found. In contrast to the findings from structural comparisons, this analysis of the active site positions the phosphoethanolamine transferases as closer to the alkaline phosphatase group of the superfamily. The distinction from other groups (sulphatases/arylsulphatases, cellulose synthases, LtaS) is supported by the existence of a partial site for a second metal ion in the

phosphoethanolamine transferases, which lack only the conserved Asp from the second metal site present in other subfamilies of the Alkaline phosphatase superfamily.



Figure 23: DALI alignment of structurally matching regions of alkaline phosphatase superfamily proteins .

Structure-based sequence alignment of the 64 structures used in DALI pairwise comparison. Individual structures were aligned to MCR-1 (PDB 5LRN) by DALI, non-matching regions are removed from the alignment. Conserved amino acids from the first metal binding site are highlighted green, the second metal site in cyan and the site of covalent intermediate formation in red. Secondary structure assignments for MCR-1 catalytic domain (pdb 5LRN) are shown at top. This figure was prepared using ESPRIPT (<https://esprict.ibcp.fr>) [126].

3.2.6 Sequence-based analysis of the position of MCR-1 in the alkaline phosphatase superfamily.

As mentioned above, structure-based phylogenetic analysis using DALI locates MCR-1 as part of the sulfatase group in the larger alkaline phosphatase superfamily. However; a multisequence alignment, performed in Clustal Omega; of the sequences of the 64 structures used in DALI-based structural comparisons, tells a different story. A phylogenetic tree based on sequence only locates MCR-1 and the group of phosphoethanolamine transferases within the alkaline phosphatase clade (between structures 5OLT and 3LXQ, Figure 24). This tree shows the phosphoethanolamine transferase cluster adjacent to the cellulose synthase (5OLT and 6PD0) and LTAs clusters ((4OUP and 2W5Q) and far away from the sulphatases/arylsulphatases, which are in the bottom half of the tree (2QZU to 6HHM). This finding is dissimilar to what was found in the phylogenetic tree derived through structural analysis (Figure 19), although both trees show a clear separation between the major alkaline phosphatase and sulphatase/arylsulphatase clades. An important thing to notice is that the sequence-based tree locates the phosphoethanolamine transferases cluster close to the cellulose synthases and the LTAs, this has also been seen in the correspondence analysis (Figure 20).

While structural analysis locates MCR-1 and other lipid A phosphoethanolamine transferases as part of the sulphatases/arylsulphatases group, purely sequence-based phylogeny positions them among the alkaline phosphatase superfamily. Our structural and sequence-based analysis suggest that MCR-1 and the lipid A PEtN transferases sit in a position between the one metal ion enzymes, such as the arylsulphatase 3ED4, and the two metal ion enzymes such alkaline phosphatase (e.g. 1K7H). given the presence and preservation of two of the three residues present in the alkaline phosphatase second metal binding site. This evidence suggests that the lipid A phosphoethanolamine transferases, such as MCR-1, are a distinct subfamily of their own within the alkaline phosphatase superfamily.

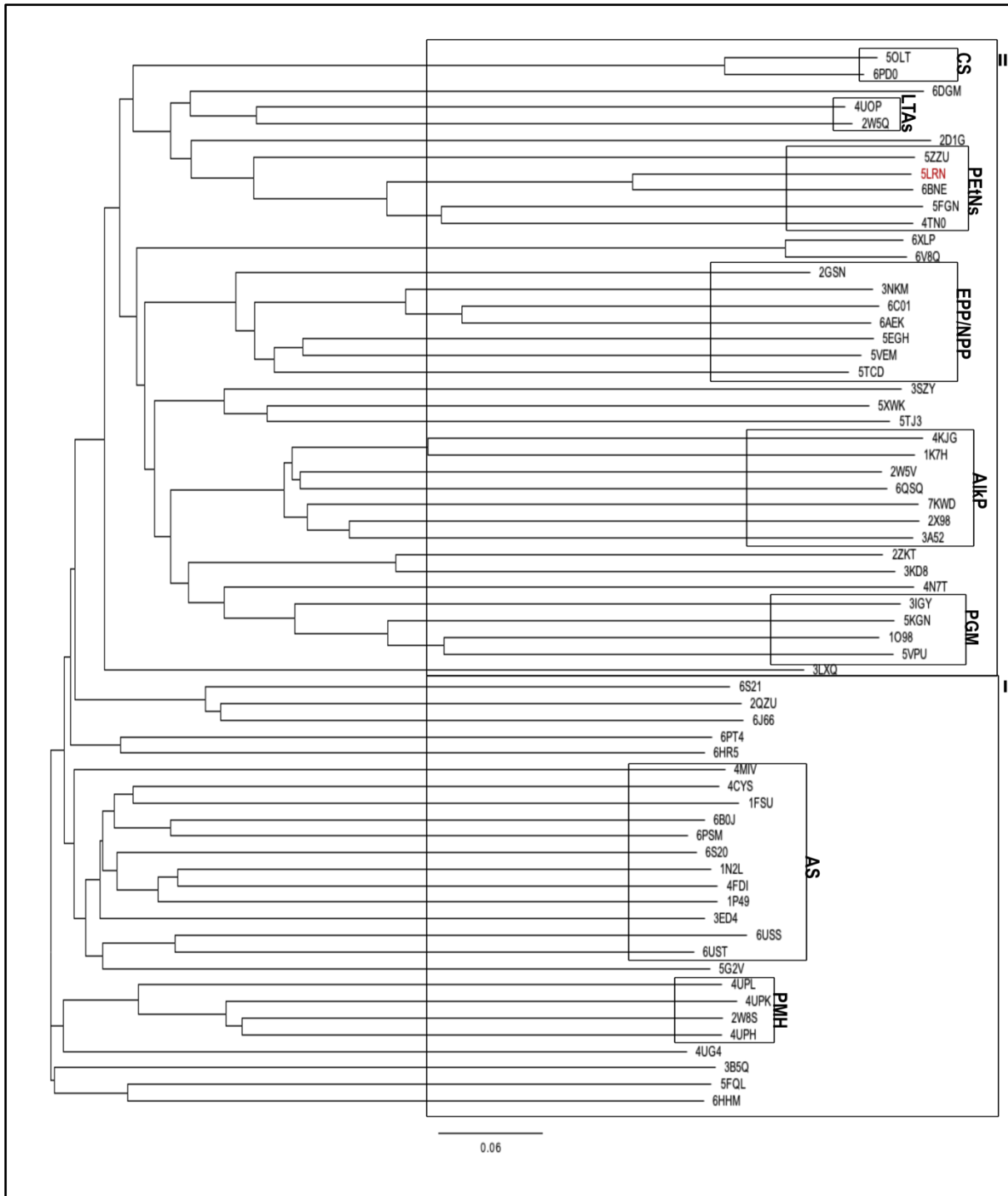


Figure 24: Sequence-based phylogenetic tree of alkaline phosphatase superfamily proteins.

Sequence-based phylogenetic tree generated in Clustal Omega for the 64 alkaline phosphatase superfamily structures used in structural comparisons. MCR-1 (PDB 5LRN) is labelled in red. Phosphoethanolamine transferases can be seen at the top of the tree adjacent to the cellulose synthases (50LT and 6PD0) and the Lipoteichoic Acid Synthases (4OUP and 2W5Q) at the upper part of the alkaline phosphate clade.

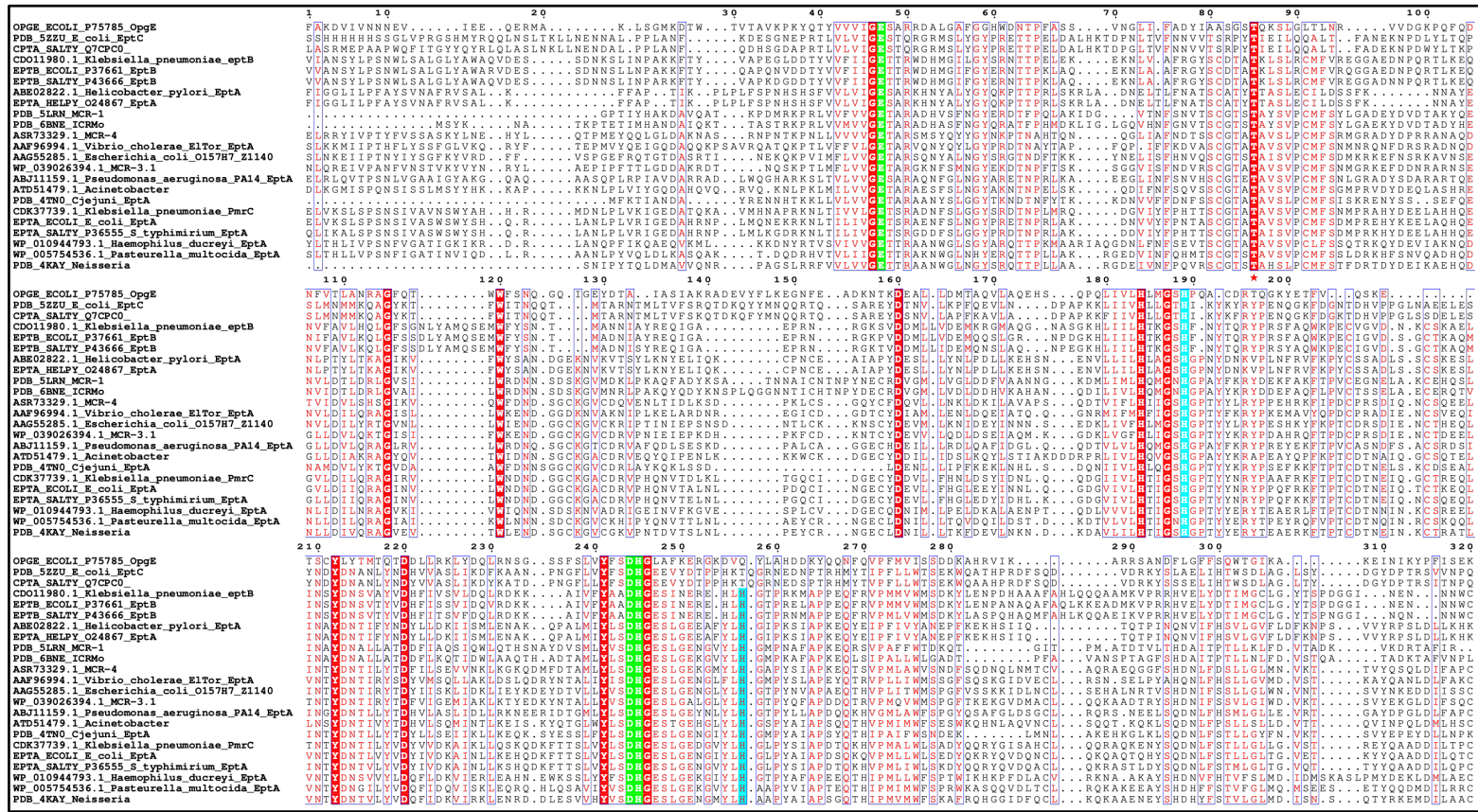


Figure 25: Alignment of phosphoethanolamine transferase catalytic domain sequences. Clustal Omega alignment of catalytic domains of Gram-negative phosphoethanolamine transferases. Conservation of the putative second metal binding site (equivalent to His395 and His478 in MCR-1) can be observed in all but EptC and related enzymes [196] where His478 is not present. Conserved amino acids from the first metal binding site are highlighted green, the second metal site in cyan and the site of covalent intermediate formation is marked with a red star. Other invariant amino acids are highlighted in red boxes. This Figure was prepared using ESPRIT (<https://espruit.ibcp.fr>) [129].

3.3 Conclusions.

Typically, sequence-based analysis is used to find possible evolutionary relationships, and identify regions of conservation and divergence, for recently discovered or uncharacterized enzymes. In our study, sequence identity between MCR-1 and other members of the alkaline phosphatase superfamily vary quite drastically, from a 67% sequence identity to MCR-1 homologue ICR(Mc) from *Moraxella catarrhalis*, to 12% sequence identity (mouse autotaxin Enpp2, structure 3NKM, Appendix N). In this scenario where sequence identity is rather low, the use of structure-based analysis is preferable, since protein 3D structures are known to be more conserved than amino acid sequences. The effectiveness of structure-based analysis depends on the number of characterized structures available, in the case of phosphoethanolamine transferases this challenge is mitigated by the extensive studies done, regarding the alkaline phosphatase superfamily, over several years [113,142,144–146].

An exhaustive structural analysis, using DALI with multiple tiers of filtering followed by pairwise comparisons of 64 non-redundant members of the alkaline phosphatase superfamily, shows that MCR-1 and other phosphoethanolamine transferases are, at the level of overall structure, more similar to the sulphatases/arylsulphatases than to the alkaline phosphatases. This structural analysis, in particular the two-dimensional correspondence analysis calculated by DALI from the set of pairwise Z-scores, positions sulphatases/arylsulphatases (and the related phosphono monoester hydrolases (PMHs)), phosphoethanolamine transferases (PEtN transferases), the lipidA binding protein PbgA, lipoteichoic acid synthase (LTaS) and the cellulose synthases (BcsG) as close structural neighbours within the alkaline phosphatase superfamily. In turn this group of enzymes is more distant from enzymes such as EPP/NPPs (that hydrolyse phospho di-, rather than mono-, esters), phosphoglycerate mutase and the alkaline phosphatases themselves. This closeness of MCR-1 to its immediate neighbours in the correspondence analysis can be associated to the structural and functional similarities shared between these enzymes. MCR (and other PEtN transferases), LtaS and BcsG are all

membrane-associated metalloenzymes with an N-terminal transmembrane domain, formed mostly by α -helices (not included in these comparisons), and a C-terminal catalytic domain that transfers a phosphorylated substrate (phosphoethanolamine in the case of MCR-1 and BcsG, glycerol phosphate in the case of LtaS) onto an acceptor [112,186,187,197,198]. While PbgA is not demonstrated to be a metalloprotein, interaction with lipid A has recently been demonstrated [179], providing some functional similarity.

The presence of one, rather than two, active site metal ions is a feature that distinguishes the large group of sulphatases/arylsulphatases from the rest of the alkaline phosphatase superfamily. Sequence and structural comparisons show that in sulphatases/arylsulphatases and their relatives the second metal site is incomplete, these proteins usually possess a His residue at the equivalent position to MCR-1 His395, but residues not expected to be involved in binding metal ions (Lys or Arg) are often found at position 478 and there is no candidate third metal ligand. By considering the architecture of the MCR-1 catalytic site, structural comparisons show this to have the necessary machinery to accommodate for the presence of 1 (having the conserved amino acids Glu246, Asp465 and His466 and Thr285 as the nucleophile) or 2 zinc ions, having two (His395 and His478) of the three ligands (lacking Asp) for the second site present in alkaline phosphatases (e.g. 1K7H Figure 22 b). There is still discussion regarding the function of this possible second zinc site in MCR-1 since structures of other lipid A phosphoethanolamine transferases, such as *N. meningitidis* EptA (PDB 5FGN), only present one active site zinc [115,120,169]. Structures of MCR-1 showing 2 or more zinc ions are usually presented with a second protein chain interacting with the second zinc ion [74,110,169,194,199]. A recent study by Lythell et al, employing molecular dynamics and density functional theory simulations, showed that in the absence of a second chain the second zinc site was unstable, and suggests MCR-1 to only require one zinc ion to catalyse the transfer of PEtN to the nucleophilic centre, Thr285 [120].

Structural and sequence comparisons across the set of 64 proteins studied in detail show that His478 is present in MCR, and the related lipid A phosphoethanolamine *transferases* *N. meningitidis* EptA (pdb 5FGN), *Moraxella catarrhalis* ICR-Mc (pdb 6BNE) and *Campylobacter jejuni* EptA (pdb 4TN0), but is absent from *E. coli* EptC (pdb 5ZZU) and other mono-metal proteins in the alkaline phosphatase superfamily, including the cellulose synthase phosphoethanolamine transferase subunits (BcsG, pdb 5OLT and 6PD0). The catalytic domain sequences of a larger set of PEtN transferase enzymes were aligned to further explore the level of conservation of His478. This analysis (Figure 25) showed that His478 is present in MCR (including the distinct MCR-3 and MCR-4 variants [200]) and EptA and EptB proteins from a range of Gram-negative bacteria, but is absent from *E. coli* EptC and OpgE and *Salmonella typhimurium* CptA.

There are several possible explanations for the presence of a conserved but partially complete second metal binding site in MCR-1. Although only one zinc seems necessary for transfer of phosphoethanolamine to the enzyme to form the covalent reaction intermediate, a second may be needed to transfer phosphoethanolamine to lipid A, as suggested by a recent computational study [201]. In this case the second site could be completed either by a contribution from the membrane-spanning domain (not considered in our comparisons), or when lipid A binds. Alternatively, the conservation of His478 may reflect a role in interaction with lipid A substrate that does not necessarily involve metal ions. The absence of His478 from *E. coli* EptC may be explained by the fact that this enzyme modifies the LPS core heptose [202], rather than lipid A (*C. jejuni* EptC (pdb 4TN0) which does possess a His478 equivalent is a multifunctional enzyme that modifies several substrates including lipid A) [203]. This, together with the lack of His478 in BcsG, which like *E. coli* EptC has an Arg at the structurally equivalent position, shows that phosphoethanolamine transferase activity does not require His at this position. In phosphonate monoester hydrolase and sulfatase enzymes an equivalent lysine (Lys337, Figure 21 e) is proposed to contribute to catalysis by leaving group stabilisation

[161], it is possible that in MCR-1 His478 does a similar job. The zinc binding Asp/ His/His (residues 315, 319, 432 in structure 1H7K Figure 22 b) seems to be a recurrent pattern in alkaline phosphatase and an archaeobacterial aminopeptidase but missing in some homologous proteins, this is due to this similar binding sites arising convergently, but are easily discarded during the divergent evolution of new functions 200. The case of MCR-1 (only containing the His/His of the trio), as well as other phosphoethanolamine transferases lacking this second zinc binding site, could be a signal that these enzymes are a result of both convergence and divergent evolution process.

In this study a structure-based phylogenetic analysis coupled to sequence analysis was used to better understand the relationship between MCR-1 and the alkaline phosphatase superfamily. The Alkaline phosphatase superfamily is known for its highly conserved structure and its $\alpha/\beta/\alpha$ fold; the structure/sequence based phylogenetic analysis in this study confirms not only that MCR-1 is a member of this superfamily but that it shares overall similarity to the sulfatase/arylsulfatase sub family, being more structurally similar to the monometallic family members of the arylsulfatase/sulphatase group, than to the multinuclear members of the alkaline phosphatase superfamily. However, at the level of sequence and catalytic site structure MCR-1 is closer to the latter group, having a more complete second metal site than other arylsulfatase/sulphatase enzymes. Our findings place MCR-1 and related phosphoethanolamine transferases of the EptA/B group as a distinct subclass within the alkaline phosphatase superfamily.

Chapter 4 Production of Recombinant MCR-1.

4.1 Introduction.

As explained in previous chapters, MCR-1 is a bacterial integral membrane protein; bound to the inner membrane of Gram-negative bacteria. MCR-1 is predicted to be composed of a transmembrane domain formed by 5 α -helices and a periplasmic-facing catalytic domain. This chapter focuses on the optimization of the production and purification of recombinant full-length MCR-1 for its biochemical characterization.

Structural and biochemical characterization studies require large amounts of stable protein, meaning that the protein needs to remain properly folded as well as remain functional. Membrane proteins (MP) are anchored into the lipidic layers of the cellular membranes and can be difficult to produce and purify, given their hydrophobic characteristics and their location inside the cells. Analysis of MP has proven to be a laborious task since their natural occurrence in the cell is frequently too low for their ready isolation [205,206]. For this reason, overproduction, using systems such as the bacterium *E. coli*, is an alternative way to obtain MP in enough quantities for their study; however, given their hydrophobic structure, overproduction tends to be deleterious for the bacterial growth, causing even death to the host, in some instances [55,207]. Therefore optimization at the different stages of the production process is required, not only to produce large quantities of material, needed for its study but to ensure that the produced protein is stable and not aggregating or degrading.

4.1.1 Optimization of the production and purification process of membrane proteins.

The Gram-Negative bacterium *E.coli* is one of the most widely known and used systems for the expression of recombinant heterologous proteins [208]. Rapid growth under lab conditions,

fast and tuneable expression systems, high production yields, low-cost production/maintenance as well as its well-understood genetics 205 and an array of available strains make it the preferred host system for recombinant protein expression. The *E. coli* BL21(DE3) strain, combined with a T7 promoter-based overexpression plasmid is one of the most used platforms for the expression of soluble recombinant proteins [210]. As mentioned before, overexpression of MPs can be deleterious for the cells, as these proteins need to be inserted into the host's membrane walls leading to stress. A way to overcome this issue is the use of an appropriate expression strain, capable of tolerating the overexpression of MPs. The Walker strains, *E. coli* C41(DE3) and C43(DE3) are mutant descendants of the *E. coli* strain BL21(DE3) capable of tolerating toxic levels of MP production [211]. Another way to reduce the problems with MP production is to have better control of the expression by tuning the transcription rate [212].

Most of the mentioned strategies focus on the enhancement of the bacteria through genetic engineering, aiming to increase productivity by improving DNA transcription, under a simple logic "more mRNA translates to more protein produced" as well as the selection of improved "superior" strains; still one of the most important factors to take into account is the chemical and physical environment to which the organism is exposed [213]. Further optimization of the culture conditions, such as temperature, growth medium components and the inducers used can enhance the yields of membrane proteins obtained.

Extraction from the membrane and purification is required for the *in vitro* study of MCR-1. Recombinant protein purification is commonly achieved by column-based chromatographic methods, such as affinity and size exclusion chromatography. Given the hydrophobic environment of MPs, the lipid bilayer, the use of detergents is vital for their purification, as they mimic the native environment of MPs allowing them to remain stable in aqueous solutions [214]. Choosing the best combination of buffer, salt concentration, detergents, and pH to

maintain the conformational stability of the protein is a key step in the optimization of the production and purification process.

4.1.2 Biophysical Characterization of Membrane Proteins.

Size Exclusion Chromatography (SEC) is an effective, and common, technique that not only allows for the purification of proteins by fractioning the samples based on size (molecular weight); but also provides some aid on their characterization. The working principle of SEC is the mobility difference between molecules based on their sizes; small molecules can interact longer with the pores of the packed bed while large molecules, such as proteins, will be “excluded”. This separation principle is the reason why SEC is a commonly used technique for protein characterization studies [215]. This however depends greatly on the protein conformation, yielding erroneous results when the protein shape is asymmetrical or then other interactions, such as dimerization occurs [216]. Since MP require the presence of detergent micelles to remain stable in an aqueous solution forming a protein-detergent complex (PDC), their molecular weight is altered. The use of a multiple angle light scattering detector (MALLS) combined with SEC is a more versatile technique for protein characterization as the different detectors used in the technique (UV, refractive index, and light scattering) provides information like the absolute mass and protein concentration independently of the PDC [217].

Analysing the structural integrity and thermal stability is a relevant setting in the characterization process of any MP. Circular dichroism (CD) is defined as the difference between the absorption of left-handed and right-handed circularly polarized light; in their native form, proteins have highly asymmetric secondary structures, α -helices and β -sheets, with distinctive CD spectra (α -helices having negative bands at 222 and 208 nm and a positive one at 190 nm and β -sheets presenting a positive band between 190 and 200 nm and a negative one between 210 and 220 nm). As they unfold these spectra change, these changes, when measured as a function of temperature, can be used to determinate the protein unfolding and

refolding thermodynamics [218], providing relevant information on the protein structure as well as its thermal stability.

The first half of this chapter will be focused on the optimization of the culture conditions, strain-plasmid combinations, and culture media enhancement on a small scale and used these to produce MCR-1 on large scale. The second half of the chapter focuses on the biophysical characterization of the purified recombinant full-length MCR-1 enzyme using techniques such as SEC-MALLS analysis, MS/MS, and Circular dichroism.

4.2 Results.

4.2.1 Small scale MCR-1 protein expression and purification.

A small-scale screening was the initial step in the optimization process for MCR-1 production and purification. The vector and expression strain are key elements for recombinant protein expression; with the right combination increasing not only the quantity but also the quality (measure as folding and aggregation state) of the protein produced. Prior to beginning this work the MCR-1 was successfully subcloned into the vectors: pET24a and pET28b by Dr. Phil Hinchcliffe at the School of cellular and molecular medicine of the University of Bristol. The bacterium *E. coli* was selected as the expression system as MCR-1 was discovered to be carried in this organism [75]. To screen for the most efficient conditions for the expression of MCR-1, three host strains were compared, the Walker strains C41(DE3) and C43(DE3) and the Lemo21(DE3) system. All of these strains are descendants of the strain BL21(DE3) 202. Strains C41(DE3) and C43(DE3) were obtained more than two decades ago as isolates, derivatives of strain BL21(DE3), that presented improved membrane protein overexpression levels [212]. The Lemo21(DE3) system is composed of the combination of strain BL21(DE3) with the plasmid pLemo. In this system, an inhibitor of the T7RNA polymerase, T7 lysosyme (T7Lys), is introduced in a pACYC vector later renamed as pLemo, under the control of the L-

rhamnose inducible promoter *rhaBAD*. The expression of recombinant proteins from T7 promoters can be controlled to an extended degree by the addition of L-rhamnose to the growth media [212,219,220] as the level of T7Lys inhibitor, and the expression of recombinant protein, is correlated to the concentration of L-rhamnose present in the culture [206,221]. This is of special value when producing membrane-associated and transmembrane proteins as overproduction can lead to toxic effects for the bacteria 208.

The pET expression system has been commonly employed for the overproduction of recombinant proteins for more than 2 decades [211]. The T7 vector pET was selected as genes cloned in this vector, controlled by the bacteriophage T7 promoter, do not cause plasmid instability and the addition of the 6-Histidine tag (His-tag) aids the purification of the recombinant protein using Ni⁺-NTA purification system [220]. Significant production of recombinant proteins can be deleterious as their overexpression is toxic for the bacteria. The combination of overexpression vectors such as the pET system with a bacterial strain, such as the Walker or Lemo strains, capable of handling toxic protein levels is a viable option for the optimization of the expression for an enzyme-like MCR-1 whose overexpression has proven to be toxic for bacterial systems [55]. Both vectors, now named *mcr1*-pET24a and *mcr1*-pET28b respectively, were transformed into *E.coli* strains C43(DE3), C41(DE3) and Lemo21.

4.2.2 Small scale titratable expression using Lemo21(DE3) system.

To assess the best L-rhamnose concentrations a small-scale expression screening was performed at varying L-rhamnose concentrations. Two L-rhamnose concentrations, 500 and 1000 µM showed the best results in terms of both, bacteria (biomass, defined as wet weight of recovered bacteria in g/L) production (as shown in Table 7) and protein recombinant MCR-1 production. Figure 26 shows the production of MCR-1 under these concentrations of L-rhamnose and 400 µM of IPTG. Something noticeable is the fact that MCR-1 runs at a lower

molecular weight, seen at 50 kDa in SDS-PAGE gels, than its predicted molecular weight of 60 kDa (predicted from its sequence Appendix A).

Table 7: Biomass production for the L-rhamnose optimization trials.

Rhamnose [μM]	Biomass [g/L]
0	9.1
100	11.93
250	13.96
500	15
1000	15
2000	13.88

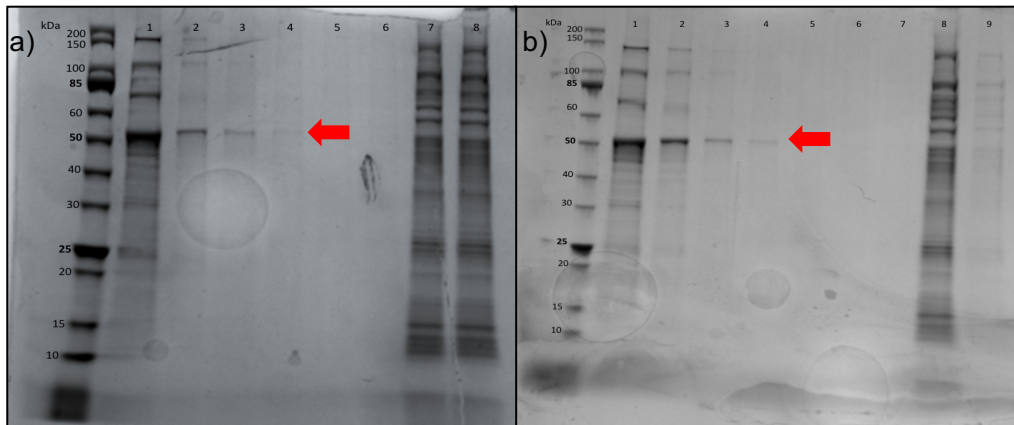


Figure 26: Recombinant MCR-1 production in E. coli Lemo (DE3).

Production of MCR-1 in E. coli strain Lemo21(DE3) harbouring mcr-1pET24a plasmid. Two Rhamnose a) 500 μM and b) 1000 μM of L-rhamnose using the pLemo system. The MCR-1 band is located between 60 and 50 kDa at the location of the red arrow. Lanes 1, 2, 3 and 4 are consecutive elution steps. Lanes 7 and 8 in a) and 8 and 9 in b) shows the cell lysate and flow-through.

4.2.3 Media optimization and protein production.

It has been reported that the use of culture media such as 2xYT and TB, which are rich in complex nutrients (e.g. yeast extract and peptone) have greater biomass yields compared with media such as Luria-Bertani (LB), however, this great production of biomass does not always

translate into an improvement for protein production and on top of this, the amount of peptone in these culture media can be detrimental for the bacterial growth. Lowering the concentration of peptone present in the culture can enhance the bacterial growth rate and sustain the bacteria during the stationary phase, however, these cultures need to be supplemented with an increased concentration of yeast extract [222] as a nitrogen source. A variation of the 2xYT media denoted 32YT was the first media alteration used to optimize biomass and protein production.

Increasing the concentration of rich nutrients in the media, such as yeast extract, can increase the final density of the culture; however, it is important to add more components to the media to maintain the viability of the culture. The addition of a low concentration of compounds such as MgSO_4 to the culture media has been shown to help maintain the viability of the culture, as well as increase the overall cell density [223]. To support high levels of protein expression a reliable carbon source is needed in addition to the amino acids present in the media. Glucose was used as a primary carbon source as it can prevent basal expression levels, by catabolite repression, as well as sustain high cell density cultures; however, its concentration must be carefully selected as it can lower the pH of high-density cultures. Glycerol was selected as a second carbon source as it functions as an energy source that does not prevent glucose depletion during the initial growth of the culture [224]. A second iteration of the 32YT medium, 32YT SD, was generated, according to Table 6, and used for the production of MCR-1. These 2 media were compared against standard 2xYT to assess the best medium configuration for MCR-1 expression.

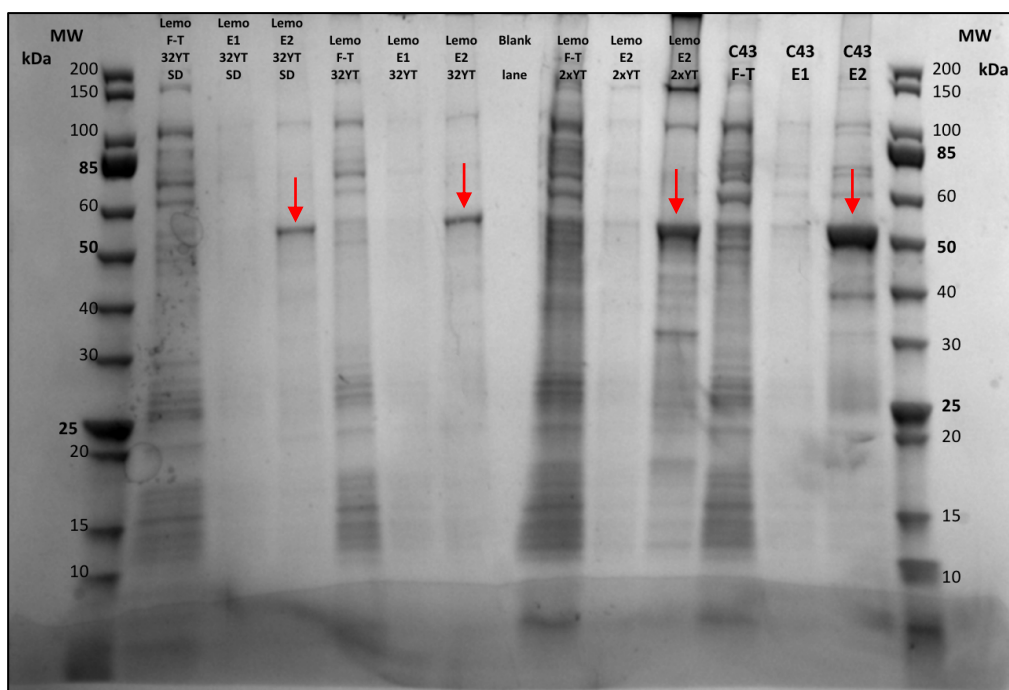


Figure 27: Recombinant MCR-1 production comparison between different culture media.

Comparison in MCR-1 production between *E. coli* strains Lemo21(DE3) in three different media configurations and the *mcr1*-pET24a vector; a yeast extract rich medium named 32YT, a glycerol-glucose supplemented version of it named 32YTSD, and a traditional 2xYT medium. Additionally, these results were compared with MCR-1 production in *E. coli* strain C43(DE3) using a 2xYT medium. The red arrows indicate the bands confirmed as MCR-1 by mass spectrometry. E1 and E2 lanes indicate the Ni⁺ purified MCR-1. L-rhamnose was added to a final concentration of 500 μ M for the Lemo cultures. Lanes marked as F-T contains cell lysate samples,

MCR-1 was extracted from solubilised bacterial membranes and purified by Ni⁺ chromatography. Analysis of these fractions using SDS-PAGE showed a band between 60 and 50 kDa, the molecular weight of MCR-1 was calculated to be 60.124 kDa from the protein amino acid sequence in the ExPASy ProtParam tool [225]. Mass spectrometry analysis of the bands confirmed the protein to be the phosphoethanolamine-lipid A transferase MCR-1 (high protein score of 809). The combination of the *E. coli* strain C43(DE3) in 2xYT medium presented a broader band when compared with the Lemo(DE3) in any of the 3 different media configurations, although the SDS-PAGE gel showed that for the Lemo strain both versions of the 32YT medium presented a clearer background, being relatively purer with minimal contaminants, compared to the production using 2xYT media (Figure 27).

Figure 27 shows strain C43(DE3), one of the Walker strains transformed, to give the best results in terms of production in 2xYT medium. Another media screening trial was carried out, this time comparing strain C43(DE3) against strain C41(DE3), using 2x YT and 32YT media. This confirmed strain C43(DE3)- *mcr-1*pET24a, in combination with medium 2x YT, to be the most suitable candidate for scaling up of the process (Figure 28).

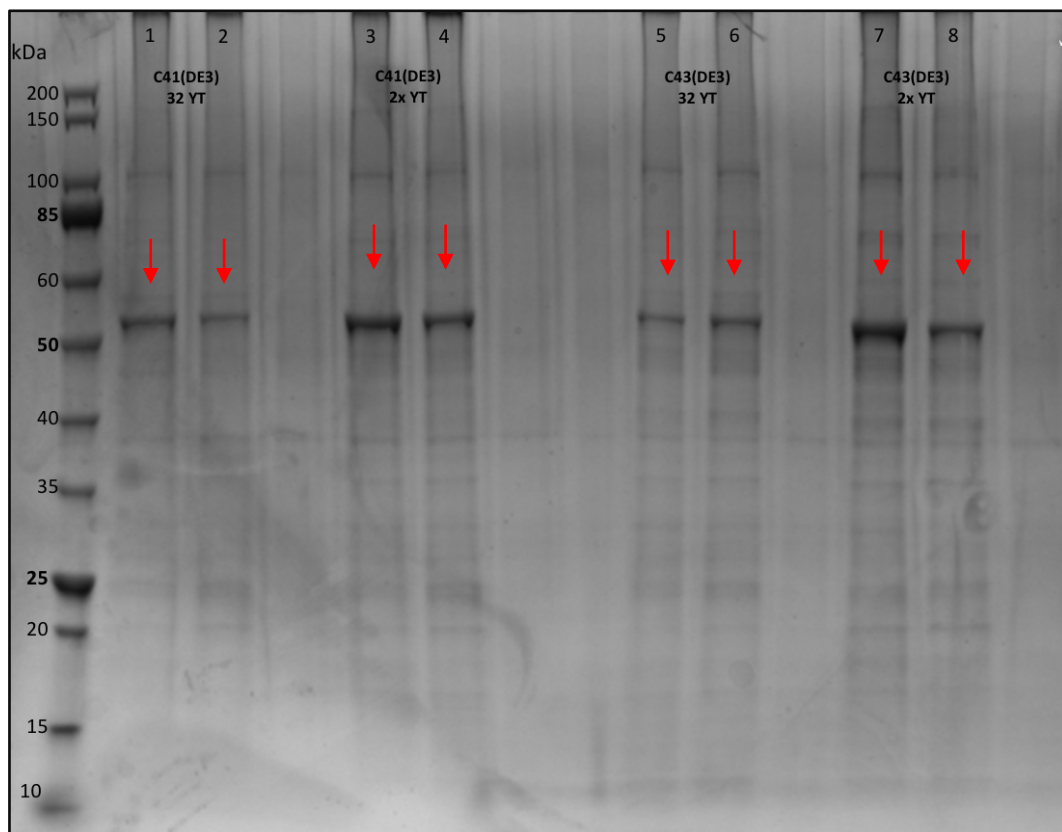


Figure 28: Recombinant MCR-1 production comparison between different culture media and bacterial strains.

Protein production comparison between E. coli strains C43 and C41, in 2xYT and 32YT media. MCR-1 can be seen located between the 50 and 60 kDa mark as indicated by the red arrows. Lanes 1 and 2 are the elution (and replicate) for the culture using C41(DE3) in 32YT media, lanes 3 and 4 are the elution (and replicate) for the culture using C41(DE3) in 2xYT media, lanes 5 and 6 are the elution (and replicate) for the culture using C43(DE3) in 32YT media lanes 7 and 8 are the elution (and replicate) for the culture using C43(DE3) in 2xYT media

To complete the optimization of the production and purification process of MCR-1 at a small scale, *E. coli* strain C43(DE3) was tested in combination with the vectors *mcr-1pET24a* and *mcr-1pET28b*, bearing C-terminal and N-terminal 6His-tags, respectively. At the same time, a detergent screening was performed comparing recovery of MCR-1 after solubilisation in the two detergents n-Dodecyl β -D-maltoside (DDM) and Lauryl Maltose Neopentyl Glycol (LMNG) Figure 29.

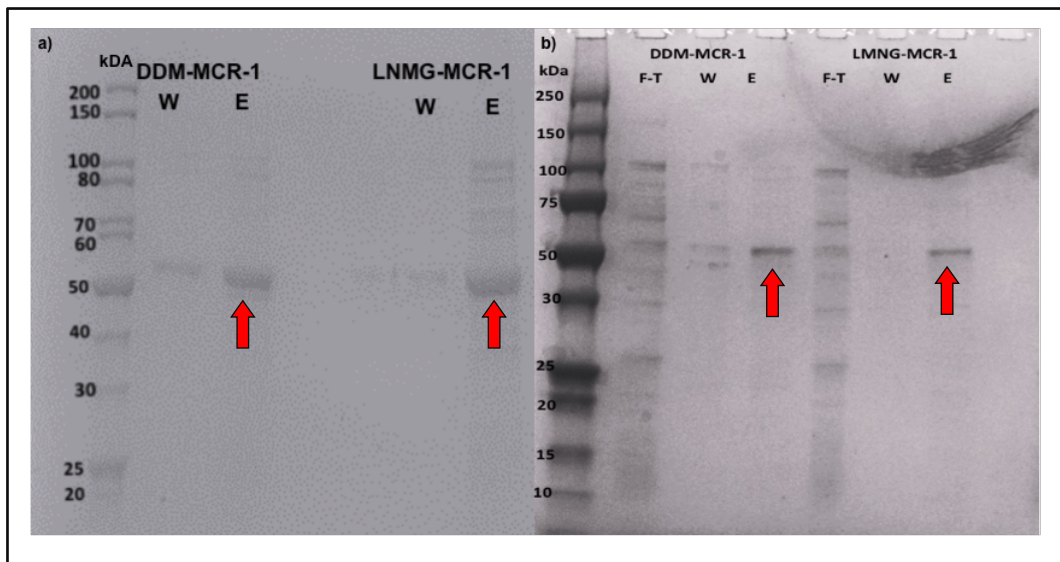


Figure 29: MCR-1 expression in the presence of detergents DDM and LMNG.

MCR-1 was expressed using a combination of strain C43(DE3) and vectors a) *mcr-1pET24a* and b) *mcr-1pET28b*. F-T refers to the flow-through, W refers to the washing step and E is the fraction corresponding to the protein Elution step. MCR-1 bands are indicated by the red arrows.

Between the plasmids *mcr-1pET24a* and *mcr-1pET28b* recovery levels of MCR-1 seemed superior with the *mcr-1pET24a* plasmid (Figure 29). Comparison of the two detergents showed better MCR-1 recovery with LMNG compared to DDM, however the latter is more cost-effective [223,226]. Overall *E. coli* strain C43(DE3) in combination with the plasmid *mcr-1pET24a* and 2xYT medium presented the best results in terms of MCR-1 expression. The *E. coli* strain C43(DE3) proved to be the best strain for the production of MCR-1. This can be attributed to the fact that this strain has a proven higher tolerance for both hosting toxic

plasmids and the production of membrane proteins, even when compared to the related strain C41(DE3) [227,228]. 2xYT is a rich medium, maintaining high levels of nutrients throughout the expression time. High-density cell cultures require large amounts of nutrients to sustain healthy bacterial growth and to achieve high levels of protein production [223,226]. The supplemented 32YT medium, named 32YTSD, is a second good option for MCR-1 expression, especially as after purification MCR-1 obtained from cultures grown on this medium had a significantly reduced number of contaminants, suggesting that not only the bacterial host tolerated the expression of MCR-1 but also that the expression of other native proteins might also be reduced. Another important factor for the overexpression of membrane proteins is the temperature of the culture. Higher temperatures during the culture and induction process can lead to higher amounts of biomass and protein; lowering the temperature during the induction phase improves solubility of recombinant protein and leads to a higher yield of membrane-integrated protein overexpression [229]. By diminishing the temperature upon IPTG induced overexpression, the organism metabolism is slowed down reducing translational stress and facilitating the insertion of membrane proteins; for the Walker strains (C41(DE3) and C43(DE3)) overexpression of recombinant proteins is optimal at temperatures nearing 20°C [230]. In these experiments the induction temperature was maintained at 18 °C throughout.

4.2.4 Large-scale production and purification of MCR-1.

Large scale expression of MCR-1 was carried out in 2L shake flasks in a shaking incubator using the best strain/plasmid/medium and detergent combination obtained from the small-scale optimization of MCR-1 production; *E. coli* strain C43(DE3), *mcr-1*pET24a plasmid, and 2xYT medium. The membranes were solubilised using 0.1% DDM and MCR-1 was purified by Ni²⁺ NTA column, the elution fractions were analysed by SDS-PAGE (Figure 30, protein bands corresponding to MCR-1 molecular weight are marked with red arrows). SDS-PAGE analysis revealed the presence of a broad band around 40 kDa in all the samples (Figure 30, bottom arrows). Mass spectrometry analysis revealed that this band is a truncated version of

MCR-1 (high protein score of 602). This truncated form seems to include the C-terminal 6 Histidine tag as it is only found during the elution steps from Ni⁺-NTA columns and not during the washing steps.

Further optimization of the production of MCR-1 for large-scale cultures was performed in 2L shake flasks, comparing growth in 2xYT with the second-best medium obtained from the small-scale optimization trials, 32xYTSD. Additionally, a third media configuration was analysed by supplementing 2xYT media with glucose, glycerol, MgSO₄, and the trace metals used for 32YT medium supplementation (2xYT SD, concentrations according to Table 6). An uninduced control was used as a reference. Compared to 2xYT, MCR-1 production from 32YT SD medium is minimal, with high levels of production of the truncated version of MCR-1, named tMCR-1. Supplementing 2xYT media with extra carbon sources and trace metals (2xYT SD) proved to be the best alternative as this not only increased the production of MCR-1 (as seen on broader bands) but also greatly reduced the production of tMCR-1 when compared to the un-supplemented version (Figure 31, lanes 7 and 15). The presence of other bands in the elution fractions indicates the need for further purification of the enzyme which was performed by size exclusion chromatography.

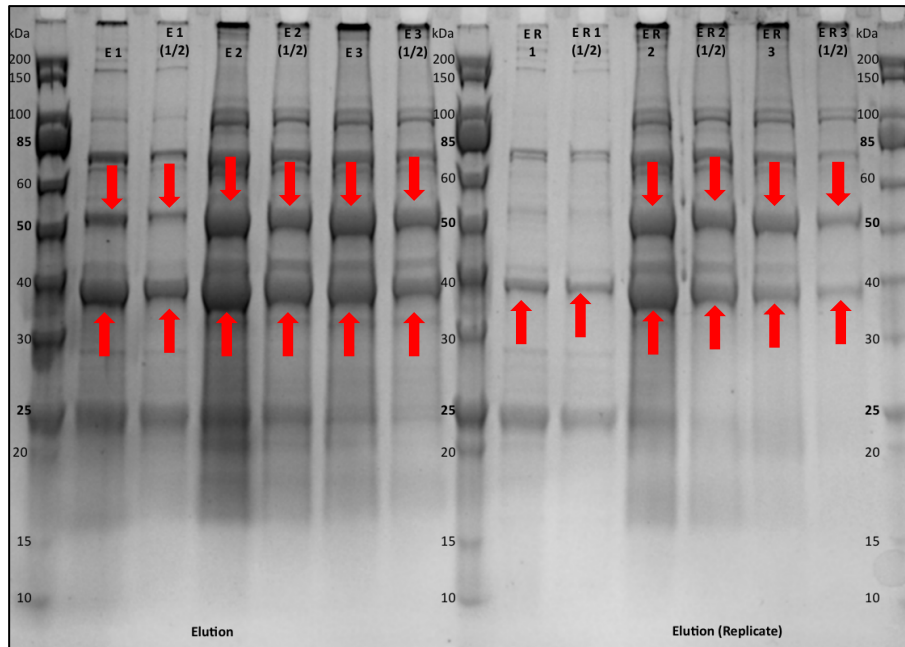


Figure 30: SDS-PAGE gel of the scale-up culture for MCR-1 production in 2x YT media.

Gel shows elution of MCR-1 from Ni²⁺-NTA column. The upper red arrow indicates full-length MCR-1, the bottom arrow the truncated form of MCR-1 (tMCR-1, approx. 40 kDa). E1, 2, and 3 corresponds to consequential elution steps. ER 1, 2 and 3 are the corresponding replicates. The lanes marked with a (1/2) are replicates with half the concentration of the sample loaded.

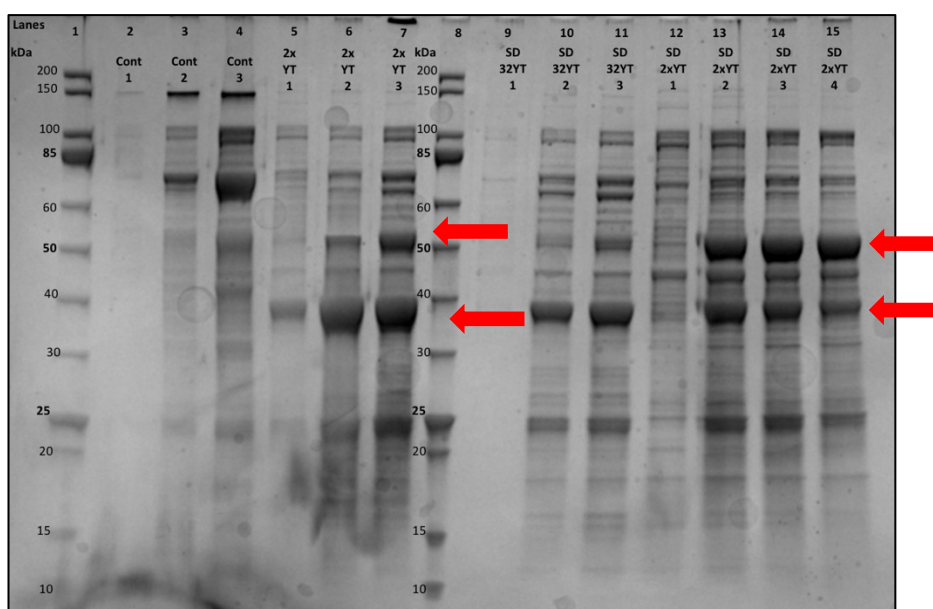


Figure 31: SDS-PAGE comparing the production of MCR-1 and its truncated form in 2x YT, 32 YT, and 2xYT Supplemented media.

The upper red arrow indicates the full-length MCR-1 (above 50 kDa) and the bottom arrow shows the truncated form of MCR-1 (approx. 40 kDa). The first lane is the Molecular weight ladder, the next 3 (2-4) lanes correspond to an uninduced culture in 2x YT labelled as control, the next 3 lanes (5-7) correspond to a 2x YT culture, lane 8 is a second molecular weight ladder, lanes 9-12 correspond to the supplemented 32YTSD medium and the last 4 lanes (12-15) correspond to the supplemented 2xYT medium (SD 2xYT).

Biomass production varied between the three different media configurations. The traditional 2xYT yielded 12.7 grams of bacteria per litre of culture, the supplemented version 2xYT SD yielded an average of 19.3 grams of bacteria per litre, and finally, 32xYTSD had an average yield of 8.9 grams of bacteria per litre of medium. The addition of the extra carbon sources as well as trace metals and $MgSO_4$, to the 2xYT medium, had a positive impact on both biomass and MCR-1 production, as well as diminished the formation of tMCR-1, and gave the most consistent results of the three media tested (Figure 32). Based on biomass yield, the relatively low formation of the truncated form of MCR-1 and the high production of MCR-1 observed by SDS-PAGE analysis, the supplemented 2xYT medium (2xYT SD) was selected for the large scale production of MCR-1. This optimized medium yielded approximately 0.1 mg/g of bacteria

per litre of culture of full-length MCR-1 and 0.06 mg/g of bacteria per litre of culture of the truncated version, tMCR-1 from 20 grams of recovered biomass after purification by SEC purification. The resulting bands can be seen in Figure 33.

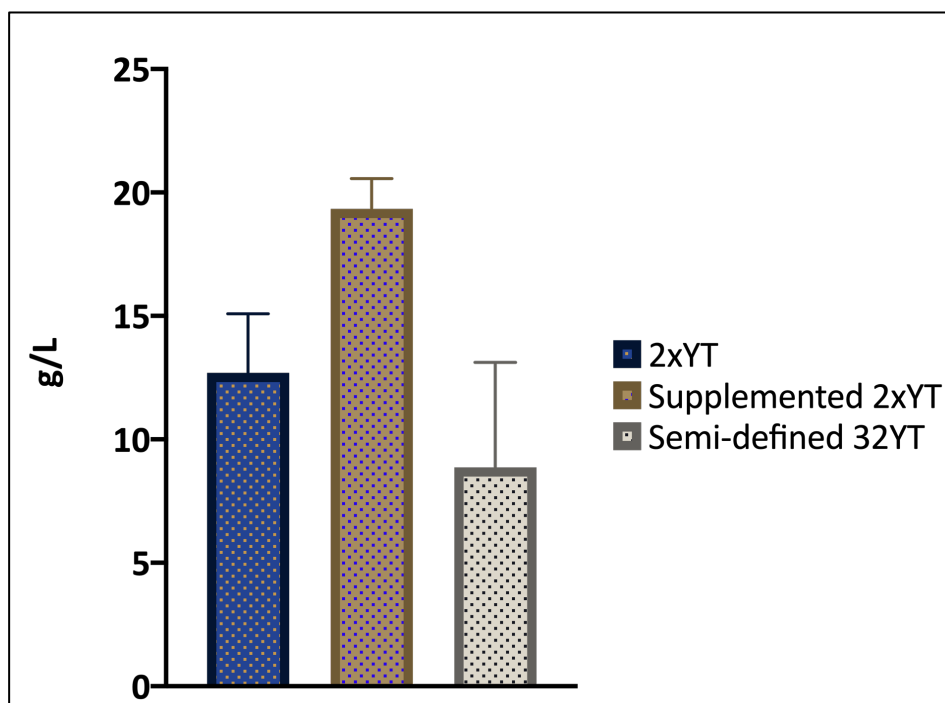


Figure 32: Bacterial biomass yield in different culture media compositions.

The supplemented 2xYT medium gave higher yields in terms of biomass production; being able to generate an approximate concentration of 20 grams of bacteria per litre of culture. Representative data from 3 biological replicates with 2 technical replicates each.

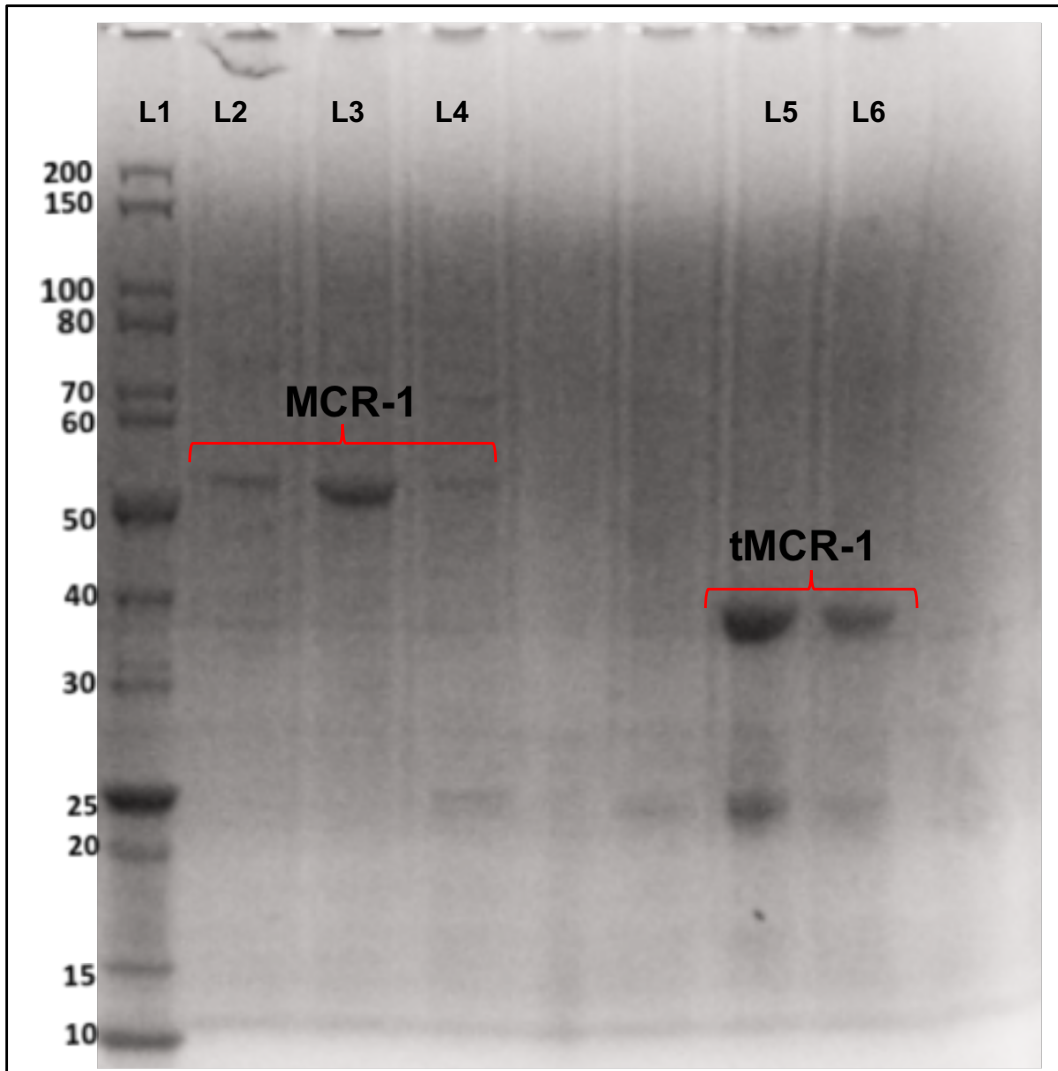


Figure 33: SDS-PAGE of the SEC purification process of MCR-1.

SDS-PAGE gel analysis of the SEC purification fractions from the large-scale expression of MCR-1 after protein production optimization. Lane L1 contains the molecular weight ladder. Lanes L2 to L4 contains the 1st eluted peak from the SEC corresponding to the full-length MCR-1 protein (mark with red brackets), Lanes L5 and L6 correspond to the 2nd eluted peak (tMCR-1 indicated with red brackets).

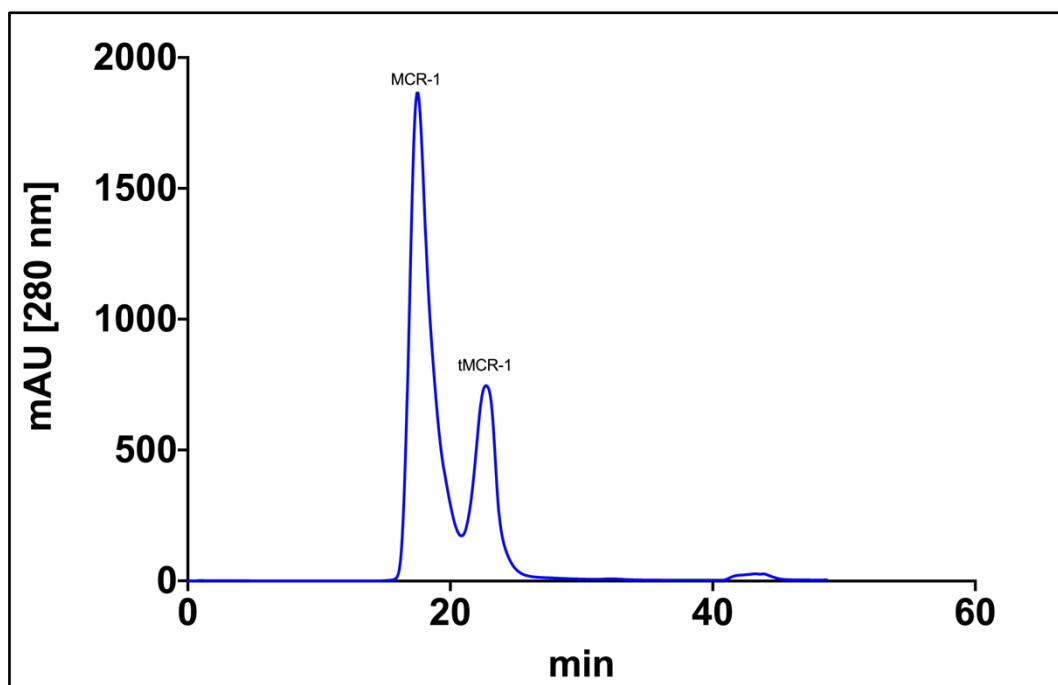


Figure 34: Chromatogram showing the elution profile by Size Exclusion Chromatography of MCR-1 (1st peak) and its truncated form tMCR-1 (2nd peak) in the presence of DDM.

SEC purification was performed on an S200 column at 0.5 mL/min. Fraction collection ($t=0$) began 10 minutes after sample injection.

Elution peaks from the Ni^{2+} -NTA column were further purified by SEC (Figure 33 and Figure 34) Both proteins, MCR-1 and tMCR-1, eluted as single sharp peaks from the SEC column. For membrane protein structure and biochemical analysis, typically, milligrams of pure protein are needed [222]. Growing kilograms of cells, to circumvent low protein production quantities, is unpractical both economically and logistically. Here we have proven that by optimizing the growth conditions, selecting the appropriate host strains as well as the nutrients present in the media, we can produce MCR-1 in the quantities needed for its further study. The optimised culture conditions, found in this study, yielded 8 mgs of purified full-length MCR-1 and 4.8 mgs of the truncated version, tMCR-1 from a 4 litter culture batch.

4.2.5 MCR-1 and tMCR-1 identification by peptide mass fingerprint.

Peptide mass fingerprinting was performed to confirm the identity of both MCR-1 and tMCR-1 (Figure 35). A first tryptic digestion was performed to digest the sample into peptides, results identified both samples as MCR-1 (Figure 35 a, c), with the full-length enzyme containing peptides corresponding to sequences near both the N- and C-termini of the protein indicating likely recovery of the intact sequence (Figure 35 a). tMCR-1 contained a single tryptic peptide apparently corresponding to amino acids 119 – 128, but no other peptides earlier than position 191 (Figure 35 c) suggesting that this is likely to be the periplasmic catalytic domain.

As many of the peptides derived from trypsin digestion of MCR-1, and particularly the one corresponding to the active site threonine, were quite large, a second peptide mass fingerprinting experiment was carried out using chymotrypsin as the digestive enzyme, aiming to obtain smaller peptides to facilitate the protein sequence identification and improve coverage of the sequence. Chymotrypsin digestion provided far better coverage of the sequence of MCR-1 by an increased number of smaller peptides (Figure 35 b), but for tMCR-1 identified no peptides earlier than those starting at position 201 (Figure 35 d). This supports the conclusion that missing region is part of the transmembrane domain of the protein.

Phosphorylation was observed on the tryptic tMCR-1 peptide containing the catalytic threonine 285 (T285, Figure 35b); consistent with other reports both on MCR-1 and other phosphoethanolamine transferases [74,95,114]. This is a large peptide (39 amino acids) was not observed in spectra from the digested full-length protein, however a smaller peptide, presenting phosphorylation on Thr285, was found during tryptic digestion of MCR-1. In the chymotrypsin digestion Thr285 is located on a smaller peptide (12 amino acids) and this was detected in digestions of both MCR-1 and tMCR-1. These data indicate that both MCR-1 and tMCR-1 are present in both phosphorylated and dephosphorylated forms.

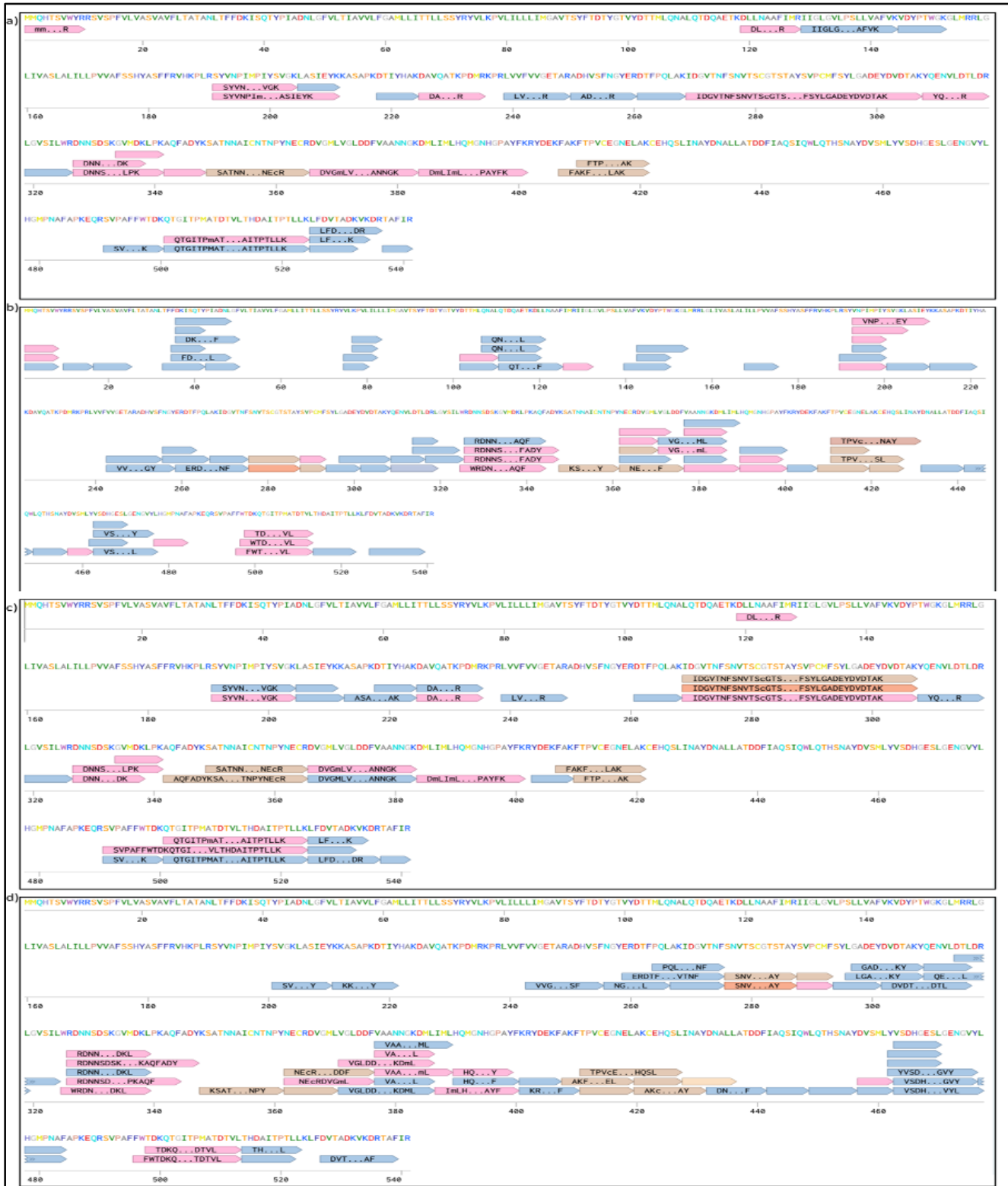


Figure 35: Mass spectrometry analysis of MCR-1 and tMCR-1.

MCR-1 and tMCR-1 identity were confirmed by mass spectrometry analysis. a) Peptide fingerprint from MCR-1 digested with trypsin, b) peptide fingerprint from MCR-1 digested with chymotrypsin; c) peptide fingerprint from tMCR-1 digested with trypsin and d) peptide fingerprint from MCR-1 digested with chymotrypsin. Peptide sequences mark in colour pink present methionine oxidation (M), brown *VOA* is for a cysteine carbamidomethyl (C) modification and orange is for amino acid phosphorylation (P).

Purified MCR-1 and tMCR-1 were subjected to further mass spectrometry analysis by electrospray, performed by Dr Karina Calvopiña Tapia at the University of Oxford according to previously published procedures [74]. Purified full-length MCR-1 contains two species (masses 61,182 Da and 61,264 Da) with a mass difference of 82 Da. From the sequence (calculated by the ExPASy ProtParam tool [136]), the predicted molecular mass of MCR-1 with a C-terminal 6His tag is 61,189 Da, but this ignores the three disulphide bonds present in the catalytic domain which when formed will reduce the mass by 6 Da to 61 183 Da. Hence the peaks in the mass spectrum correspond to intact and phosphorylated full-length MCR-1. The spectrum of the truncated form tMCR-1 contains several peaks, suggesting that multiple species are present (Figure 36 b). However there are two main peaks (38,559 Da and 38,641 Da) separated by a mass difference of 82 Da, likely corresponding to a single fragment in phosphorylated and unphosphorylated forms. The C-terminal fragment of 6His-tagged MCR-1 beginning at residue Leu 205 has a predicted mass of 38,565 Da, which after the disulphide bond is taken into consideration matches the observed peak at 38,559 Da. A gel slice containing tMCR-1 was sent for N-terminal sequencing (Alta Biosciences, Redditch, U.K.) to confirm this finding, but, consistent with the presence of multiple peaks in the mass spectrum, the results identified multiple sequences and could not identify the site of cleavage.

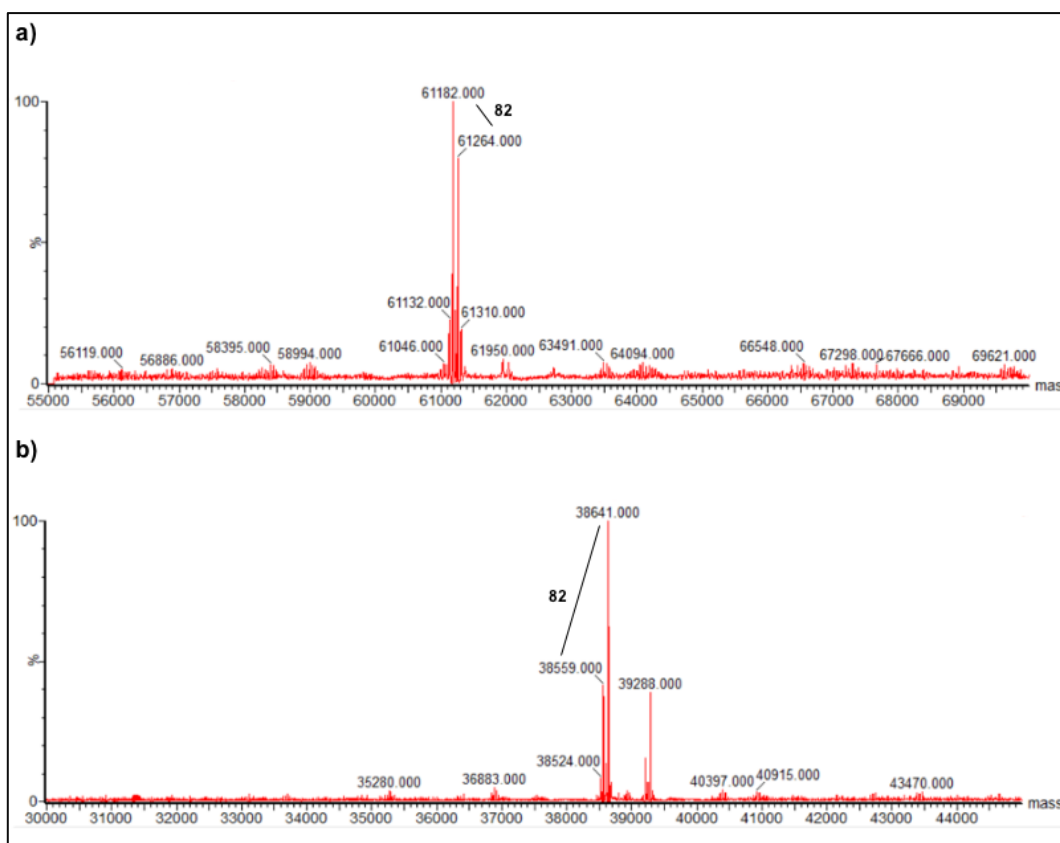


Figure 36: MCR-1 and tMCR-1 MS analysis.

a) Mass spectrum of purified full length MCR-1. The difference between the two main peaks (61182 Da and 61264 Da) is exactly 82 Da, corresponding to a mixture of the phosphorylated and unphosphorylated forms. b) Mass spectrum of purified truncated MCR-1, tMCR-1. Two main peaks can be seen (38 559 Da and 38 641 Da) with a mass difference of 82 Da, likely corresponding to a single fragment in phosphorylated and unphosphorylated forms.

4.2.6 SEC-MALS analysis

Size exclusion chromatography is one of the most common methods for protein size characterization; however, this method relies on the use of water-soluble proteins and is not suitable for the characterization of membrane proteins where the presence of a detergent is vital for protein solubilization and purification [214,231]. Size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) is a technique that allows to accurately determine a membrane protein composition, molar mass, and oligomeric state while in detergent solutions [232,233].

SEC-MALS analysis was used to analyse the behaviour, the absolute molar mass, and the conformational state of MCR-1 in its interaction with the DDM detergent micelles. The SEC-MALS technique involves an SEC column as an essential first step to allow the physical separation of possible aggregates, monomeric proteins, and detergent micelles. The SEC column is then connected to three coupled detectors; ultraviolet (UV, Abs₂₈₀), light scattering (LS), and refractive index (dRI). The UV detector is used to measure the protein concentration based on its absorbance at a wavelength of 280 nm; LS measures the light scattered by large molecules, such as detergent micelles; and lastly, a refractive index (RI) detector that measures the concentration of all the components in the sample 226. The light scattering detector is sensitive to protein aggregation, while the UV and dRI detectors are sensitive to proteins as well as other molecules such as polysaccharides, this combination allows for the technique to analyse each elution peak individually; making it suitable to characterize membrane proteins determining the MW of the protein and detergent components individually [233,234].

Calibrating the column with a protein reference helps to prevent errors during the estimation of the column constants 226. BSA was selected as the protein reference for SEC-MALS analysis of MCR-1 as its molecular weight of 66.4 kDa is well known, coincidentally similar to MCR-1 theoretical molecular weight of 61 kDa, as well as being almost monodisperse (as it can be seen in Figure 37 a) with minimal aggregate formation; the symmetric monomeric peak (around 24.6 mL elution volume) was selected and used in the SEC-MALS analysis to estimate the column constant K. Three samples were analysed by the SEC-MALS system; a BSA sample used as a reference protein, purified full-length MCR-1, and DDM micelles (Figure 37 a, b, and c respectively). The soluble BSA protein tends to form aggregates in solution, these aggregates tend to elute first as seen in the elution profile (Figure 37 a) with the main peak eluting at 24.6 mL and two small peaks eluting earlier, 20 and 22 mL respectively, representing possible aggregates. MCR-1 appears to be monodisperse as there

is only one peak showing with strong signal in all three detectors (UV, LS, and dRI) eluting at 21.7 mL (Figure 37 b). A second peak, around 25 mL, can be observed for the dRI and LS signals. This can be attributed to the presence of detergent micelles in the sample since there is no signal from the UV detector and the fact that DDM micelles are eluting around the 26 mL mark (Figure 37 c).

There are two techniques used to analyse the data obtained by SEC-MALS to estimate the protein molar mass; the *ASTRA* method, which receives its name from the commercial software from Wyatt technologies [231,235], and the *three-detector method* by Slotboom *et al.* [236]. The estimation of protein molar mass by the *three-detector method* requires the use of a detergent that does not absorb light at a wavelength of 280 nm [236]. As observable in Figure 37 c, the detergent N-Dodecyl β -D-maltoside (DDM) abides by this requirement, thus allowing us to use the *three-detector method* to estimate MCR-1 molar mass. The molar mass of MCR-1 in DDM detergent was estimated by both methods.

Analysis of MCR-1 using the Astra 6 software predicts an average molar mass of 118.74 kDa across the elution peak (Figure 37 d, marked peak). To more accurately estimate the molecular weight of the eluting MCR-1 in DDM micelles, the three-detector method, that allows for the determination of the protein molecular mass in the detergent micelle, was followed [236]. BSA was used as a reference protein to estimate the column constant K. By using the known molecular weight of BSA (66.4 kDa) and its extinction coefficient at 280 nm (0.66 ml/mg*cm) Equation 4 was obtained.

Table 8: BSA three detectors (UV, dRI, and LS) values used for column constant K determination.

	BSA	DDM	Δ Values (BSA-DDM)
Elution Volume	24.60	24.60	
UV (V)	0.05274	0.000878	0.051862
LS (V)	0.148255	0.134705	0.01355
dRI	0.0000329		

Using the values from Table 8, Equation 4 transforms into:

$$K = \frac{(0.01355) * (0.051862)}{43824 * 0000033.^2}$$

$$K = 14.81$$

Using this value for the column constant K, the molar mass for MCR-1 can be obtained by adding K to Equation 2, alongside the MCR-1 extinction coefficient (obtained from its sequence, Appendix B, using the online Expasy's Protparam tool [136]) and the signal data for the MCR-1 elution peak from Table 9 we estimated the molar mass for MCR-1 to be 122.51 kDa. The difference between this value and that obtained from the ASTRA software may be due to the ASTRA results being the average of the molar mass across the narrowest part of the elution peak for MCR-1 whereas the *three-detector method* results correspond to the highest point of the peak (21.7 mL elution volume).

Table 9: MCR-1 three detectors (UV, dRI, and LS) values used for molar mass estimation.

	MCR-1	DDM	Δ Values (MCR-1-DDM)
time (min)	21.7	21.7	
UV	0.026243	-0.00115	0.027393
LS	0.155884	0.114749	0.041135
dRI	-0.0000239		

$$MW_{MCR-1} = \frac{\Delta LS * \Delta UV_{280}}{14.81 * 1.09_{MCR-1} * dRI^2}$$

$$MW_{MCR-1} = \frac{0.041135 * 0.027393}{14.81 * 1.09_{MCR-1} * (-0.000024)^2}$$

$$MW_{MCR-1} = 122517 \text{ Da}$$

The molecular mass estimated for the purified MCR-1 in DDM (122.51 kDa) is twice the theoretical molecular weight of 61 kDa (Appendix B) suggesting that possibly an MCR-1 dimer is forming while interacting with DDM micelles.

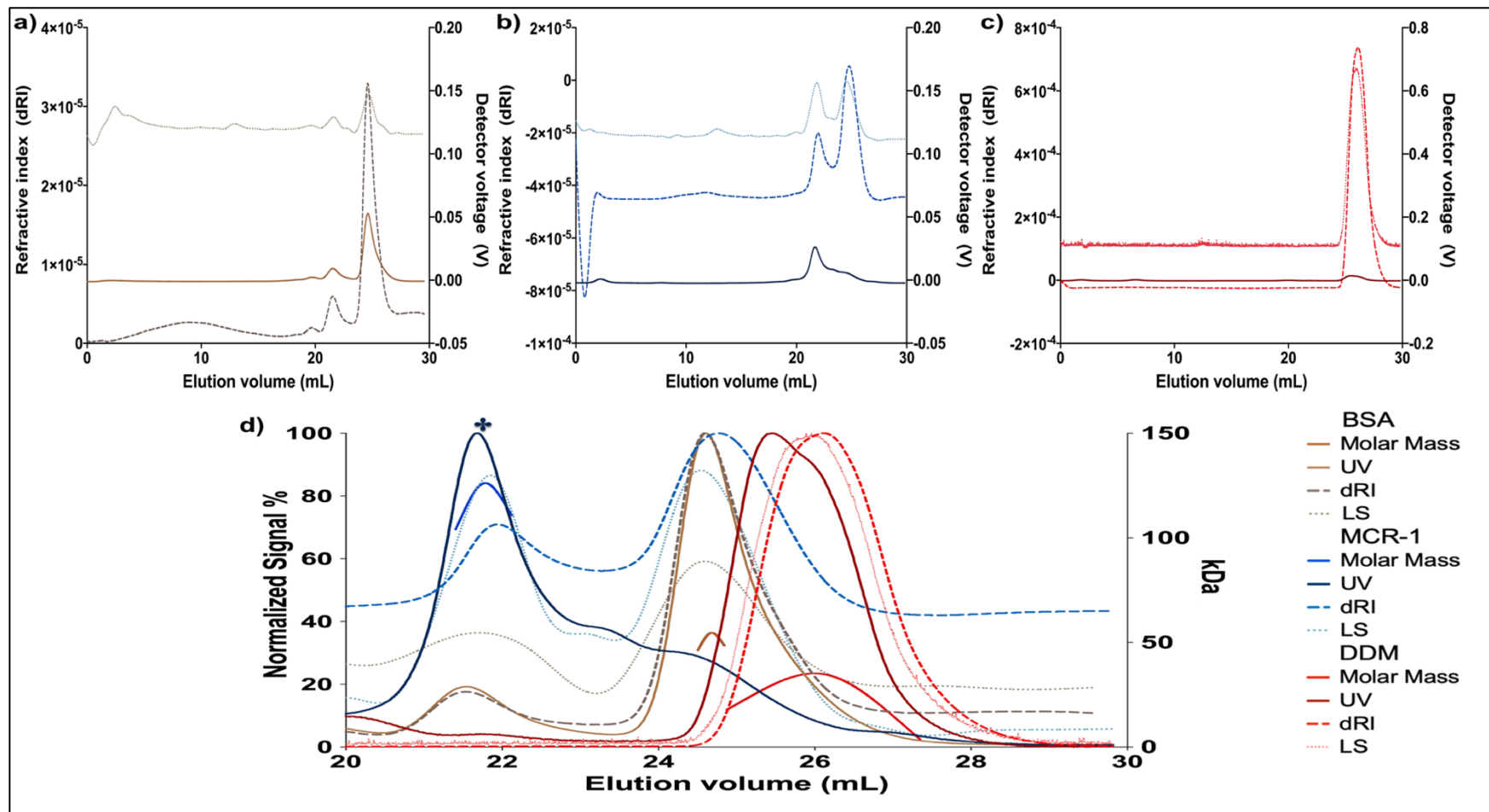


Figure 37: SEC-MALS profile of BSA MCR-1 purified in DDM micelles and Molar mass estimation.

SEC profiles of a) BSA, b) MCR-1 and c) DDM micelle monitored by differential refractive index (–dRI–) in the left Y-axis, UV Absorbance at 280 nm (—UV—) and light scattering (••LS••) in the right Y-axis. d) Comparison for the elution profiles of the three samples; the normalized signal for each detector (left) and the estimated molar mass (right axis) were plotted against the elution volume. The output signal for each detector was normalized for comparison. The marked peak was used to determine the molecular weight of MCR-1.

4.2.7 Circular Dichroism and thermal stability analysis.

Circular Dichroism (CD) spectroscopy was used to analyse the conformational properties and thermal stability of MCR-1 in DDM micelles. The spectra, taken at different temperatures, revealed a minimum at 213 nm and 219 nm at 5°C and a 215 nm and a second one at 220 nm at, both, 25°C and 35°C (Figure 38 a and b). This can be attributed to the heavily α -helix structure presence in MCR-1 as α -helical proteins tend to show negative bands around 222 nm and 208 nm and positive bands at 193 nm [237]. As temperature increases MCR-1 starts unfolding in an unreversible fashion as after lowering the temperature gradually, from 95°C to 5°C, showed no change in the CD analysis, indicating that once the protein has been denaturalized by temperature, refolding cannot be achieved. As temperature increases the stability of the protein will decrease leading to its unfolding [238]. This shows that MCR-1 is fairly stable, in the presence of DDM, at high temperatures, with the melting temperature (T_m) for MCR-1 being 60.34° C (Figure 38 c)). These results are consistent with a CD analysis performed for another phosphoethanolamine transferase sharing structural homology with MCR-1 of around 40%, EptA. CD analysis of EptA presented minima at 222 nm and 206 nm in the presence of DDM micelles [121]; however, MCR-1 proved to be more stable as the EptA T_m was measured around 57°C, approximately 4°C below the T_m of MCR-1 measured here. Secondary structure analysis of MCR-1 based upon the measured CD spectrum was attempted using the K2D3 server (<http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/>) [239]. Unfortunately the error obtained for the predicted model spectrum for MCR-1 was too high, making the predictions on secondary structure content unreliable. K2D3 uses model CD prediction based on known protein structures deposited in the Protein Data Bank [239], the fact that only a few hundred structures of transmembrane proteins, like MCR-1, in the PDB [240] makes the generation of a CD model by K2D3 unreliable.

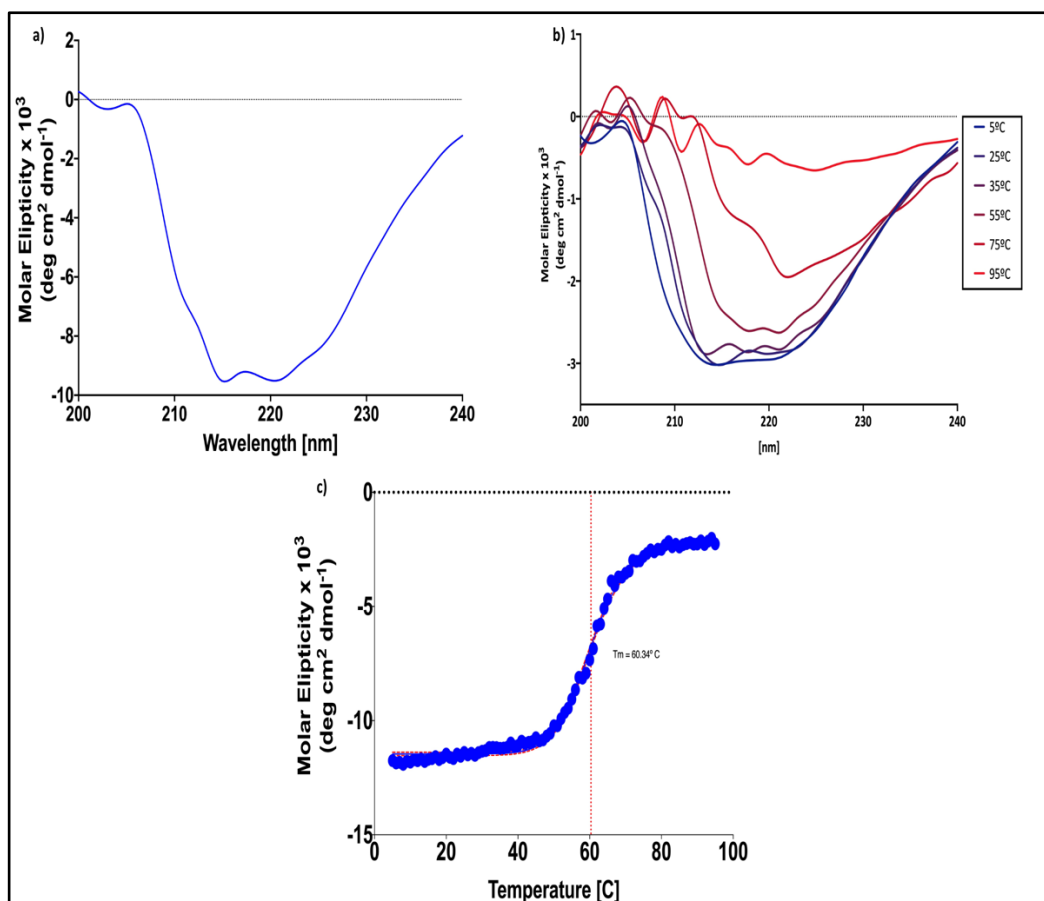


Figure 38: Circular Dichroism spectrum of MCR-1.

a) CD spectrum of MCR-1 at 25°C. b) Changes in the CD spectrum as temperature rises. C) The inflection point happens at 60.34° C. No refolding activity was observed after gradually lowering the temperature back to 5°C

These results are consistent with other biophysical characterization done in MCR-1 and EptA [241,242]; confirming that the optimization in the production and purification process done in this work yields high amounts of pure MCR-1 properly folded.

4.3 Conclusions.

This chapter describes an optimized expression protocol for the production and purification of recombinant full-length MCR-1 for structural and functional analysis. The optimization of the expression systems, as well as the culture media, resulted in a high yield of the full-length MCR-1. The growth medium composition as well as the strains selected for protein overexpression played a major role in the production of MCR-1. The *E. coli* C43(DE3) strain transformed with the *mcr1*-pET24a plasmid yielded the best results in terms of protein production, observed by SDS-PAGE (Figure 27). Although the pLemo system offers a simple route for titratable overexpression of MCR-1, it might be more suitable for high-throughput screening procedures than for large scale production given its requirements for an adjusted concentration of L-rhamnose.

An important aspect in large scale production of recombinant protein is the culture medium as it establishes the environmental conditions for bacterial growth. The use of “complex media”, such as 2x YT can help support high production levels during bacterial fermentation, however their composition, being formulated with ingredients of natural origin (e.g. yeast extract), can enhance the variation inherent to the production process. Using defined elements, can lower this undesirable variability in productivity [213,243]. By supplementing trace minerals and MgSO₄ to the complex 2xYT medium, it was not only possible to improve the yields of biomass obtained, but also allows for consistent levels of both biomass and protein production.

Although high yields of MCR-1 production were achieved, a fragmented version of the enzyme, lacking amino acids 1 – 204, and here named tMCR-1, was present during expression and purification. Despite our best efforts, where and when this cleavage is happening could not be detected. Similar fragmentation has been observed in other members of the MCR family, including MCR-3 (Dr. Nick Brown, University of Bristol School of Cellular

and Molecular Medicine, personal communication) and MCR-4 [244]. This kind of fragmentation is not unheard of and has been found on a structurally similar enzyme; *Staphylococcus aureus* lipoteichoic acid synthase, LtaS. LtaS is a membrane protein with 5 N-terminal transmembrane α -helices followed by a “sulphatase-like” extracellular C-terminal domain involved in the synthesis of polyglycerol-phosphate lipoteichoic acid (LTA) from phosphatidylglycerol. Proteomic studies have revealed that LtaS, despite being a membrane-bound enzyme, undergoes a cleavage process during bacterial growth, separating the C-terminal domain from its N-terminal anchorage; and suggesting that this enzyme might work as an extracellular enzyme [198]. The nature and causes of this cleavage of MCR-1, and the impact it may have on enzyme activity, remain unknown. Although the exact location of this cleavage was not definitively identified, a combination of electrospray mass spectrometry and peptide mass fingerprint analysis, employing trypsin and chymotrypsin for protein digestion (Figure 35), enabled a possible site to be located. This site was located at the end of the N-terminal transmembrane domain, after residues Lys204, as shown in Figure 39 and similar to the location of the cleavage site for LtaS, located at the N-terminal Ala215-Leu216-Ala217, releasing the “sulphatase-like” C-terminal domain [198]. However, the possibility of more than one cleavage site, as suggested by electrospray mass spectrometry and N-terminal sequencing of tMCR-1, is something that must be taken in consideration.

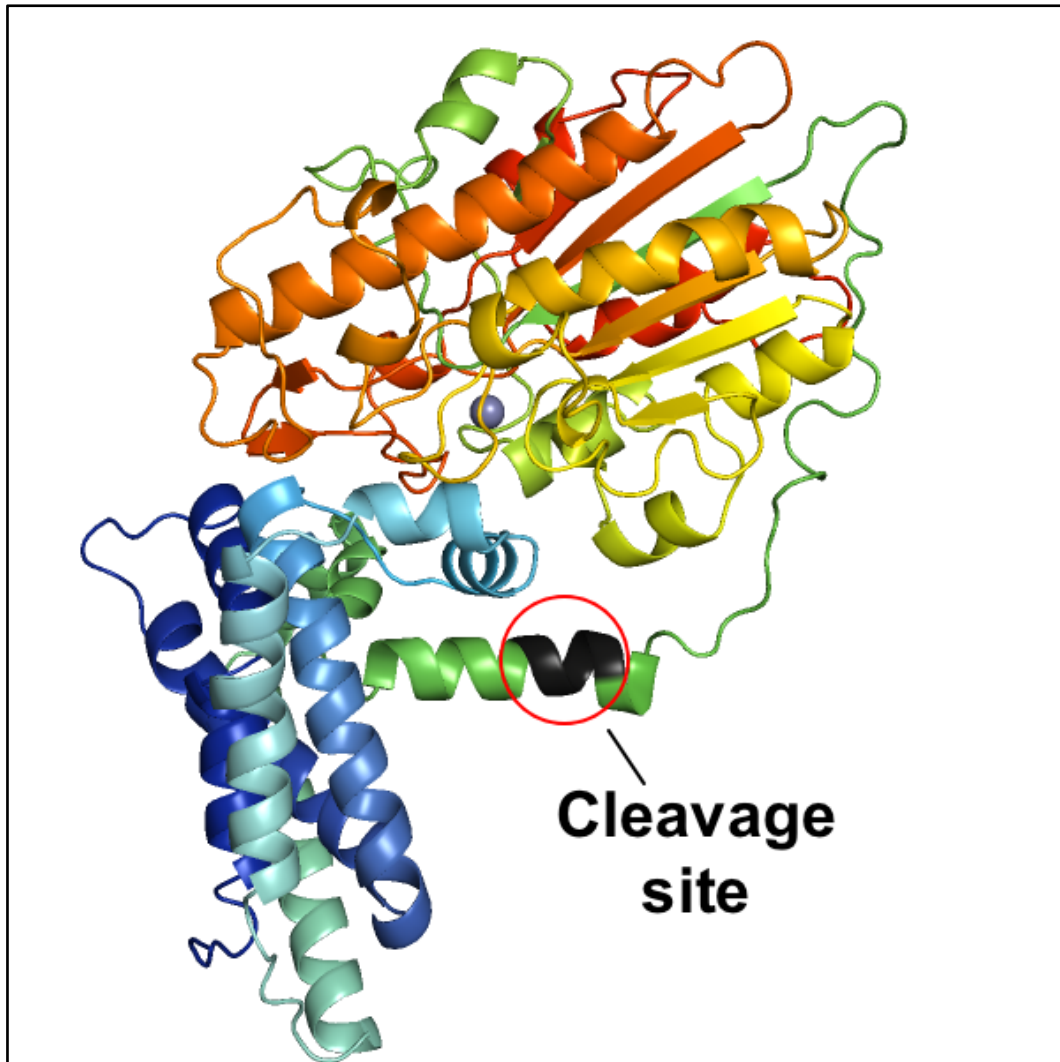


Figure 39: Location of the truncation site of MCR-1.

Mass fingerprint analysis revealed a possible truncation site before residue 205 at the end of the N-terminal membrane anchoring domain. The residues Leu205-Ala206-Ser207-Ile-208-Glu209 are black coloured.

During SEC-MALLS analysis, the molecular weight of full-length MCR-1 was estimated to be 122.51 kDa, twice the size of the molecular weight estimated from its amino acid sequence. This can be explained by considering how many MCR-1 molecules are present in a single DDM micelle, and suggests two MCR monomers, or an MCR-1 dimer, per DDM micelle. Dimeric structures of the soluble catalytic domains of phosphoethanolamine transferases have been reported in the past; for example crystal structures of the soluble domains of *Neisseria meningitidis* LptA (EptA, PDB entry 4KAY) [111,245], or of MCR-1 and MCR-2, have been

reported [74,115]; however the dimeric forms of all these proteins (LptA, MCR-1 and MCR-2) are generally considered not to be biologically relevant. To our knowledge, no potential dimerization of the full length MCR-1 has not been reported.

Despite the presence of tMCR-1, enough MCR-1 was produced and purified for its biophysical characterization. Circular dichroism spectroscopy and melting temperature analysis showed that recombinantly produced MCR-1 is stable and structured, having an estimated melting temperature of 60.34° C, very similar to the melting temperature reported of 61.14°C [246]. Even though the activity of the purified MCR-1 could not be assessed, conditions to produce and purify full-length MCR-1 were identified and biophysical characterization demonstrated the integrity of the protein in DDM micelles. Raynal et al., proposed a simple workflow to assess the quality of recombinantly produce proteins using three stages: (i) the assessment of the sample purity and integrity, (ii) sample homogeneity assessment and (iii) activity assessment [247]. MCR-1 purity and integrity was assessed through SDS-PAGE analysis and Mass Spectrometry analysis, its homogeneity was assessed via Size exclusion chromatography and it was confirmed to remain soluble and in native form, in the presence of detergent, via Circular Dichroism and SEC-MALS analysis. MCR-1's activity could not be assessed at this stage as an activity assay for the full length MCR-1 assay is still in the process of being developed. However, based on the results obtained from the aforementioned analysis, it can be concluded that MCR-1 produced is of good quality. These findings will provide the basis for future investigations of the structure and mechanism of full-length MCR-1.

Chapter 5 Development of an *in vitro* biochemical assay to assess MCR-1 activity.

5.1 Introduction.

MCR-1 confers resistance against polymyxins by catalysing the addition of a phosphoethanolamine group onto the Lipid A present in the outer membrane of Gram-negative bacteria (Figure 40) [248]. This Lipid A modification reduces the affinity between polymyxins and their target, Lipid A, rendering their antimicrobial activity against GNB ineffective, however, the exact mechanism by which MCR, and other bacterial phosphoethanolamine transferases, achieve this modification is still unknown. The development of an *in vitro* assay for phosphoethanolamine transfer is essential as not only would it allow for detection of the presence, or absence, of MCR-1; but also make it possible to assess its catalytic activity as well as aid in the identification of possible inhibitors for this antibiotic resistance mechanism. So far the activity of MCR-1, as well as other homologues such as *N. meningitidis* EptA, has been identified mostly using *in vivo* minimal inhibitory concentrations (MIC) assays such as broth microdilution (BMD), or disk diffusion (DD) [52,249,250] coupled with PCR or sequencing-based identification of gene carriage. Many of these methods incorporate the use of chelators, such as ethylenediaminetetraacetic acid (EDTA) or dipicolinic acid (DPA), as a means to detect phenotypic colistin resistance in *Enterobacterales* by exploiting the dependence of activity upon the presence of zinc in the MCR active site. However this comes with limitations as these detection mechanisms cannot distinguish between *mcr*-plasmid harbouring isolates and isolates expressing chromosomal mechanisms (such as two-component regulated systems like PhoQ/PhoP), another limitation is low diffusion levels of colistin into the agar thus producing unreliable results [251]. More recently, the development of the MALDIxin test has demonstrated that (with modern instruments) Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-ToF) mass spectrometry, as widely applied in clinical microbiology laboratories, can identify lipid A modifications from bacterial colonies,

giving hope that colistin resistance can be more reliably identified but without the ability to distinguish between plasmid-mediated and chromosomal mechanisms [252].

An *in vitro* assay is still needed to fully characterize this resistance mechanism. The development of an *in vitro* enzyme assay is of utmost importance for the understanding of the enzymatic mechanism of PEtN transferases, with a particular priority given to MCR-1 given its mobilisation and wide distribution [87,88,90,94]. Additionally, an *in vitro* activity assay would open the possibility for screening and evaluation of small molecule inhibitors of MCR-1 and other phosphoethanolamine transferases. With this in mind, the aim of this chapter is the development of a robust *in vitro* assay that would make possible investigation of the mechanism of action of MCR-1 as well as serve as the basis for a platform to screen for possible inhibitors of this resistance mechanism.

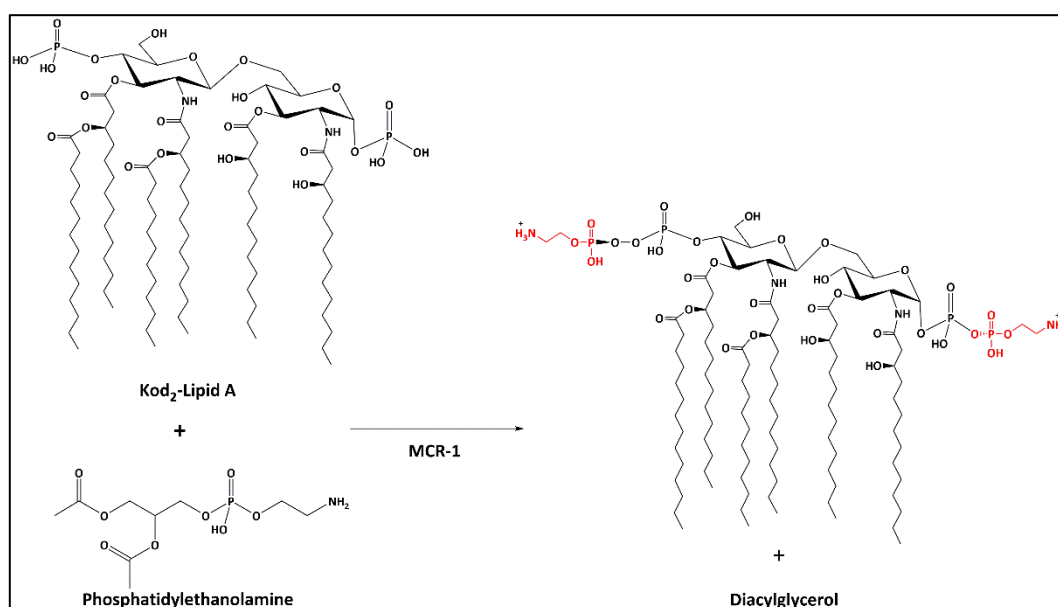


Figure 40: Reaction catalysed by MCR-1.

Transfer of a phosphoethanolamine group from a donor (Phosphatidylethanolamine) to Kdo₂-LipidA catalysed by MCR-1. The transfer may occur to the phosphate groups at the 1' or the 4' (as shown in red) positions.

5.1.1 Previous approaches to assay phosphoethanolamine transferase activity.

The enzymatic activity of other phosphoethanolamine transferases, as well as that of other colistin resistance mechanisms such as arabinose incorporation [52,67,71–75], has been identified using multiple methods *in vitro* and in bacterial cells. These broadly divide into approaches that utilise mass spectrometry or thin-layer chromatography (TLC) to detect modified lipid A in producer bacteria, and *in vitro* spectroscopic methods using chromogenic or fluorescent reporter substrates. Lipid A modification, by the incorporation into lipid A of either 4-amino-4-deoxy-L-arabinose or phosphoethanolamine, has been measured in polymyxin-resistant *Salmonella typhimurium* by coupling ^{32}P labelling with MALDI-ToF mass spectrometry and high-resolution nuclear magnetic resonance (NMR) spectroscopy. The enzymatic activity was detected by TLC analysis of the Lipid-A recovered from *S. typhimurium* grown in ^{32}P -labelled medium and subsequent mass spectrometry and NMR [65]. TLC-based assays have been used to detect modification of ^{32}P -labelled lipid A by other phosphoethanolamine transferases including EptB from *E. coli* [253], EptA from *Pseudomonas aeruginosa* [254] and EptC of *Campylobacter jejuni* [255].

Anandan *et al.* developed an assay to estimate the enzyme activity of *Neisseria meningitidis* EptA, an MCR-1 homologue, using TLC with a fluorescent-labelled substrate (1-acyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]dodecanoyl}-sn-glycerol-3-phosphoethanolamine (NBD-PEtN)) and recombinantly produced EptA in dodecyl- β -D-maltoside (DDM) micelles. The reaction was analysed by thin-layer chromatography and the products confirmed with mass spectrometry. The activity was compared with that of a “soluble-domain only” construct, lacking the transmembrane anchorage. The full-length EptA, in DDM micelles, was able to successfully remove PEtN from the substrate and transfer it to Lipid A, however, the “soluble domain-only” construct did not show any activity [118]. This same substrate, NBD-PEtN, has been used with MCR-2 and MCR-4, both homologues of MCR-1, to confirm phosphoethanolamine transferase activity [170].

In an earlier study, the activity of the recombinant soluble, periplasmic (catalytic) domain of *N. meningitidis* EptA was identified by using the chromogenic substrate, para-nitrophenol-phosphoethanolamine (pNP-PEtN). The activity of this soluble domain was identified by the release of 4-nitrophenol, yielding a yellow-coloured solution and a change in the absorbance that could be monitored at a wavelength of 405 nm [111]. Similarly, pNP-PEtN has been used to demonstrate phosphoethanolamine transferase activity of the *E. coli* cellulose synthase subunit G (BcsG) [256]. However, neither of these assays reported the time-dependence of reaction progress, important to enable the real-time study of an enzyme-catalysed reaction and ready determination of its kinetic parameters.

5.1.2 Criteria for detection of phosphoethanolamine transferase activity

The simplest enzyme assay is composed of three elements; the enzyme, the substrate and a buffer solution; and allows the visualisation of the catalysed reaction involving the transformation of the substrate into the product by monitoring the appearance, or disappearance, of a coloured compound. Identification of a substrate that is close to the enzyme's native substrate, the requirement for co-factors such as metal ions and the optimal buffer composition and conditions (pH, temperature) play a major role in the development of such an activity assay. Another important condition is that the product must differ from the substrate in the observed feature, to facilitate the observation of turnover. The use of a chromogenic compound allows for more immediate observation of the reaction as well as the use of simpler measuring techniques and equipment. Fluorometry can be used when the reaction cannot be measured photometrically and is more sensitive; however, only a handful of substrates or products emit fluorescence naturally [257].

This chapter details the development of a simple activity assay to detect phosphoethanolamine transferase activity, using a chemically synthesised chromogenic compound, 4-Nitrophenol-phosphoethanolamine (pNP-PEtN) and the soluble catalytic domain

of MCR-1 and its homologue MCR-2 [115]. The first part of this chapter will focus on the synthesis and characterisation of 4-Nitrophenol-phosphoethanolamine (carried out by collaborators in the School of Chemistry, University of Bristol) followed by the expression and purification of the recombinant catalytic domains and the development and optimization of the chromogenic enzymatic assay. Lastly, it will focus on the kinetic parameters for the isolated MCR catalytic domain, and its comparison with other *in vitro* assays used to study PEtN transferase enzymes such as EptA.

5.2 Results and discussion.

5.2.1 Small scale expression and purification of the C-terminal catalytic domains of MCR-1 and MCR-2.

To start the small-scale expression trials the plasmids *cmcr-1*-pOPINF and *cmcr-2*-pOPINF were transformed into the *E. coli* SoluBL21(DE3) strain. Cell transformations were successful for both plasmids and a small-scale expression trial was carried out. Expression and purification of both cMCR-1 and cMCR-2 was achieved in both instances, resulting in a single band around 37 kDa after SDS-PAGE analysis, consistent with the molecular weight of both cMCR-1 and cMCR-2 of 36.1 and 35.9 kDa respectively. The protein identity of both proteins was confirmed by mass fingerprint (cMCR-2 score 2085 Appendix J). There was no apparent difference in the production of both constructs.

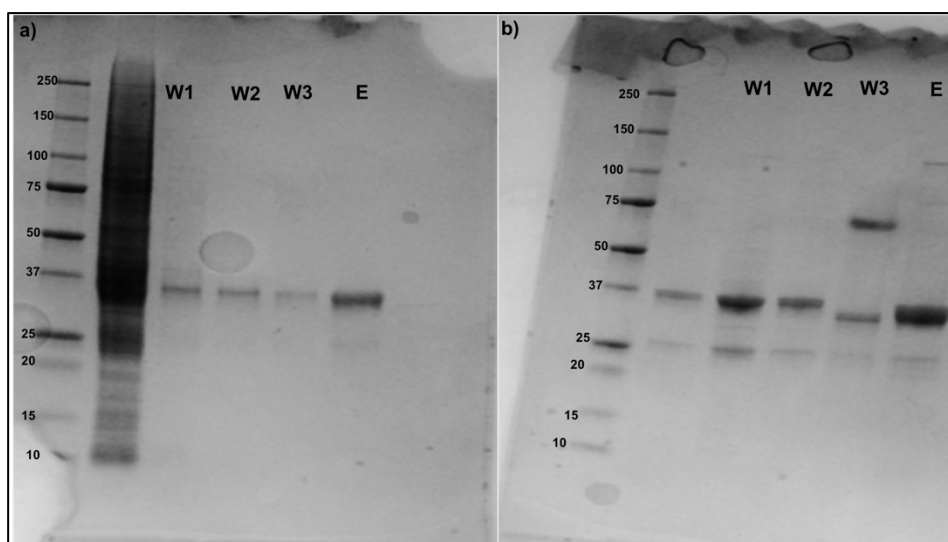


Figure 41: SDS-PAGE comparing the small scale production of a) MCR-1 and b) MCR-2 catalytic domains 2.

a) Small-scale purification of cMCR-1. cMCR-1 can be seen as a single band at the 37 kDa marker in the elution lane (E); from the sequence cMCR-1 has an estimated molecular weight of 36.1 kDa. b) Small scale purification of cMCR-2. cMCR-2 can be seen as a single band at the 37 kDa mark in the elution lane (E); from the sequence cMCR-2 has an estimated molecular weight of 35.9 kDa. Lanes W1 through W3 correspond to washing steps.

5.2.2 Large Scale Production and purification of cMCR-1 and cMCR-2.

Large scale expression of both cMCR-1 and cMCR-2 was carried in batches of 4 litres of cultured media using strain *E. coli* SoluBL21(DE3) and the optimised semi-defined medium described in previous chapters (Media optimization and protein production.). The recovered bacteria were broken down and cell lysate was centrifugated to remove cell debris. The clear cell lysate was incubated in Ni⁺ beads and soluble domains of cMCR-1 and cMCR-2 were recovered and purified, from other *E. coli* protein contaminants, by eluting using 250 mM imidazole. Fractions were analysed by SDS-PAGE. The SDS-PAGE analysis revealed a strong recovery of both domains from Ni⁺ NTA beads however further purification steps were required (Figure 42 a and b). Both cMCR-1 and cMCR-2 were treated with 3C protease to remove the 6His-Tag and then purified by SEC on a Superdex 75 column. Both proteins eluted as single well define peaks around 70 mL elution volume (Figure 43). SDS-PAGE analysis of the collected fractions, for both cMCR-1 and cMCR-2, revealed both to contain a protein of approximate mass 37 kDa, consistent with the molecular weights of cMCR-1 and cMCR-2 of

36.1 kDa and 35.88 kDa respectively (Figure 42 c and d). Protein identity was confirmed by MS/MS analysis. The fractions were recovered and concentrated to the final concentration of 13.76 mg/mL of cMCR-1 and 14.35 mg/mL of cMCR-2.

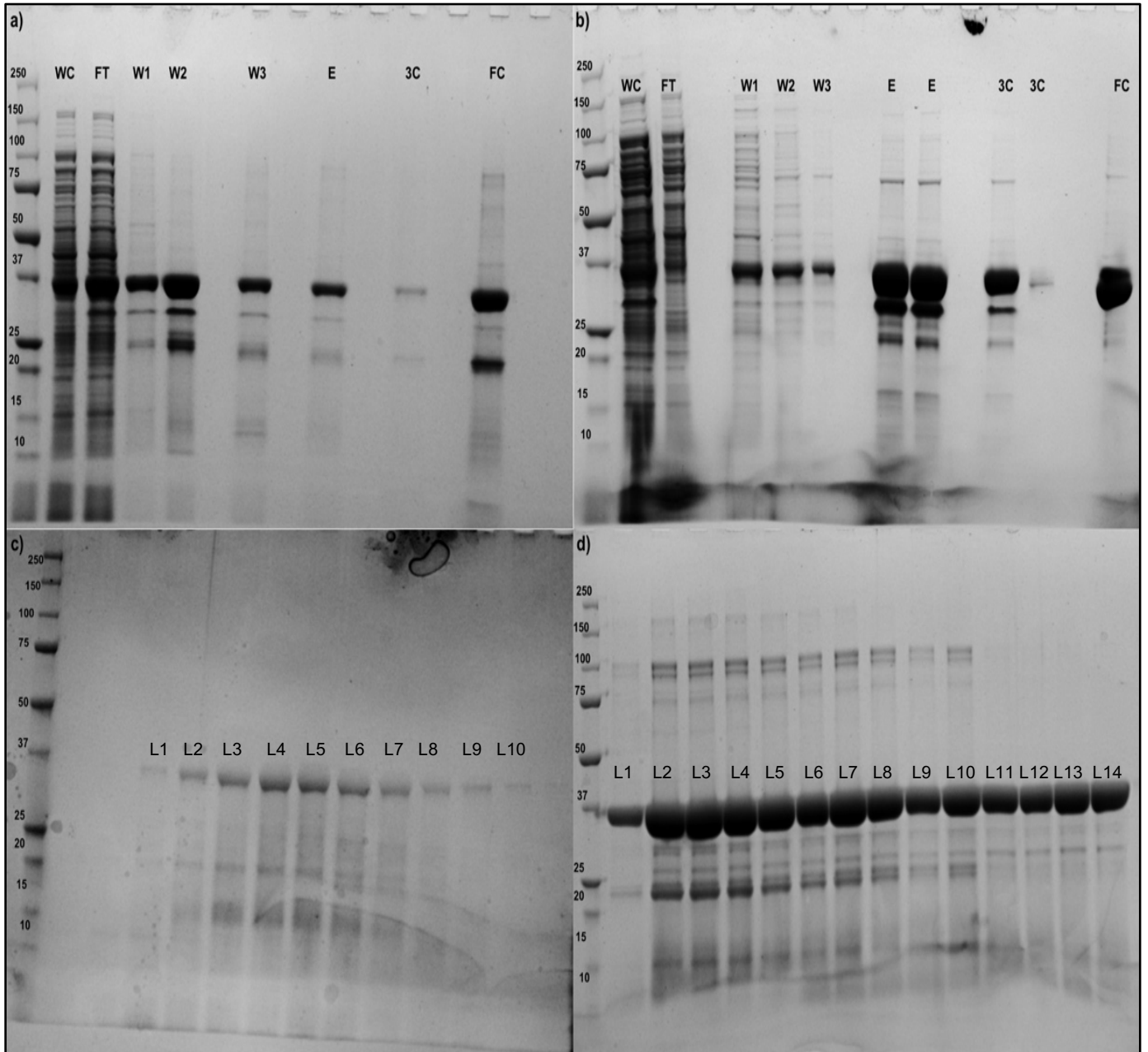


Figure 42: Large scale production, recovery and purification of cMCR-1 and cMCR-2.

SDS-PAGE gels of the large-scale production and recovery of a) cMCR-1 and b) cMCR-2 from bacteria grown in semi defined 2xYT medium. Gels a) and b) show the protein production of whole-cell samples (WC), the flow-through after Ni^{2+} NTA beads incubation (FT); the protein-bound Ni^{2+} NTA washing steps (W1, 2 and 3), the elution (E) from Ni^{2+} NTA beads, 3C protein incubation (3C) and the final concentration before further purification by SEC for both cMCR-1 and cMCR-2 respectively. Panel b) shows a replicate of the elution (second E) and the flow-through of the 3C removal step (second 3C lane). c) SDS-PAGE gel with the fractions recovered from the purification by SEC of cMCR-1 (Lanes 1 to 3 represent belong to the first tail of the peak, lanes 4 to 6 are the narrowest part of the elution peak and lanes 7 to 10 are the second tail of the peak). d) SDS-PAGE gel with the fractions recovered from the purification by SEC of cMCR-2 (Lanes 1 to 6 represent belong to the first tail of the peak, lanes 7 to 10 are the narrowest part of the elution peak and lanes 11 to 14 are the second tail of the peak). SEC profile of both, panel c) and d), can be seen in Figure 43 a) and b) respectively.

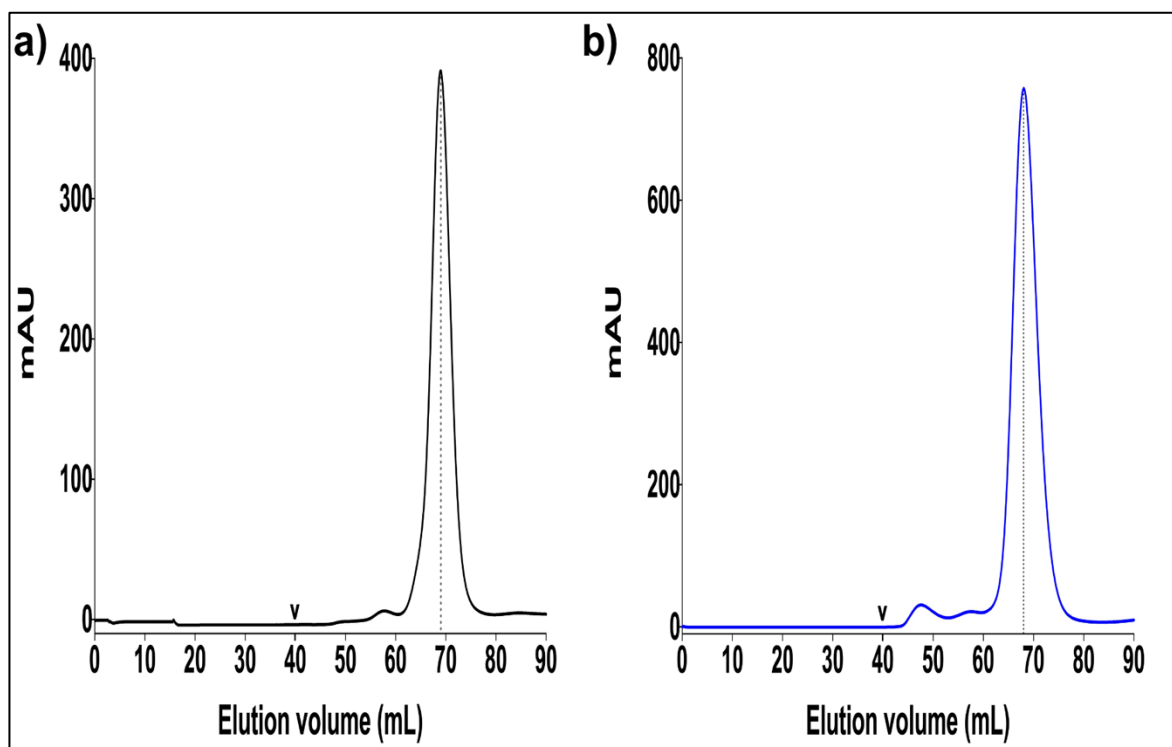


Figure 43: Elution profile by Size Exclusion Chromatography of cMCR-1 and cMCR-2.

a) Elution profile of cMCR-1 through SEC. cMCR-1 elutes at 69 mL (29 mL after fraction collection starts). b) Elution profile of cMCR-2 through SEC. cMCR-2 elutes at 68 mL (28 mL after fraction collection starts). Fraction collection starts after 40 mL, signalled by the black arrows, this is due to the void volume of the column.

5.2.3 Development of a biochemical assay.

The development of an activity assay is a first step to further understand the mechanism, investigate structure-function relationships for an given enzyme, as well as to aid in the identification of possible inhibitors. Production of recombinant cMCR-1 and cMCR-2 was achieved; these soluble domains were used as the basis for the generation of an activity assay to permit investigations of the mechanism of action of these phosphoethanolamine transferases. This assay will not only be used to determine enzyme activity but also serves to confirm the integrity of the purified enzyme, to determine the kinetic parameters and lastly provide a tool for the identification of possible inhibitors.

5.2.3.1 Establishment of pNP-PEtN as a reporter substrate and the optimization of the assay conditions.

The compound pNP-PEtN has been previously used in investigations of the activity of bacterial phosphoethanolamine transferases such as *N. meningitidis* EptA [111]. However, the published synthetic procedure produces pNP-PEtN in low yield alongside a mixture of other species (A. Gorman and P. Pringle, personal communication and [258]). An improved synthetic method developed by collaborators (Dr A. Gorman and Dr P. Pringle at the University of Bristol Chemistry Department), produced pure compound in sufficient amounts for use in assays of activity for the isolated catalytic domains of MCR enzymes.

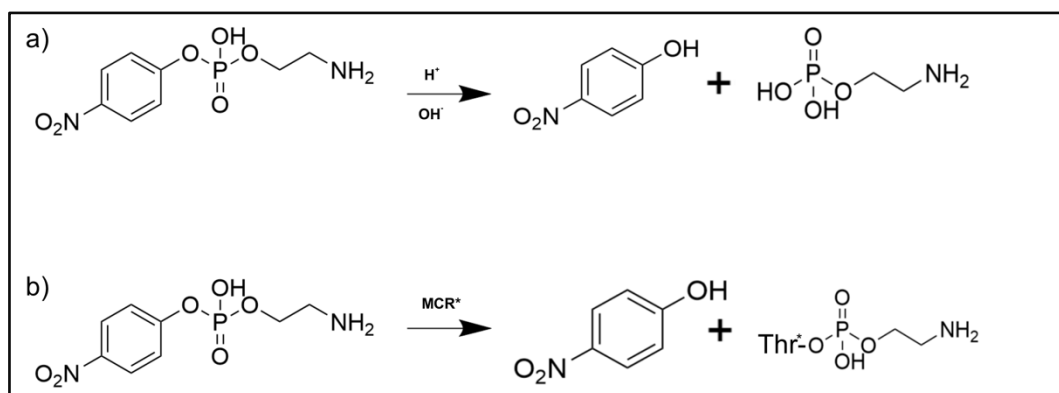


Figure 44: Possible reactions occurring in the assay.

These reactions involve a) the non-enzymatic hydrolysis of pNP-PEtN resulting in pNP and PEtN in solution and b) the formation of a covalent intermediate “PEtN-Thr*” with Threonine*, the catalytic nucleophilic amino acid in both MCR-1 and MCR-2. *Thr285 for MCR-1 and Thr283 for MCR-2.

While pNP-PEtN is a good candidate for a reporter substrate for phosphoethanolamine transferase activity, as p-nitrophenol (pNP) formation can easily be monitored with a spectrophotometer at a wavelength of 405 nm. Non-enzymatic hydrolysis of pNP-PEtN has been previously reported [111]; the fact that pNP is a good leaving group leads to the possibility of it undergoing non-enzymatic hydrolysis, depending on the pH of the environment. This non-enzymatic hydrolysis was also confirmed experimentally by Liquid Chromatography couple to Mass spectrometry (LC-MS) analysis (see below 5.2.4.1). A non-enzymatic

hydrolysis assay, varying the pH of the buffer across neutral to alkaline conditions, was done to establish the best conditions for further enzymatic activity assays. The reaction was followed for just under 4 hours, after that point, the self-hydrolysis of pNP-PEtN had remained low at pH values between 6.5 and 7.5, but increased significantly above pH 7.5 to the point where complete hydrolysis was achieved at pH 9.0 (Figure 45). These results suggest that the non-enzymatic hydrolysis of pNP-PEtN is negligible at pH values lower than 7.0, but is likely to complicate measurements of enzyme reactions at higher pH. The chromophore p-nitrophenyl phosphate (pNPP), used as a reporter substrate in assays of alkaline phosphatase activity, undergoes similar non-enzymatic hydrolysis, this also happens mostly at alkaline pHs [259]. This suggests that a buffer with a pH of 7.0 does not support significant levels of non-enzymatic hydrolysis, thus making this an appropriate pH for the activity assays.

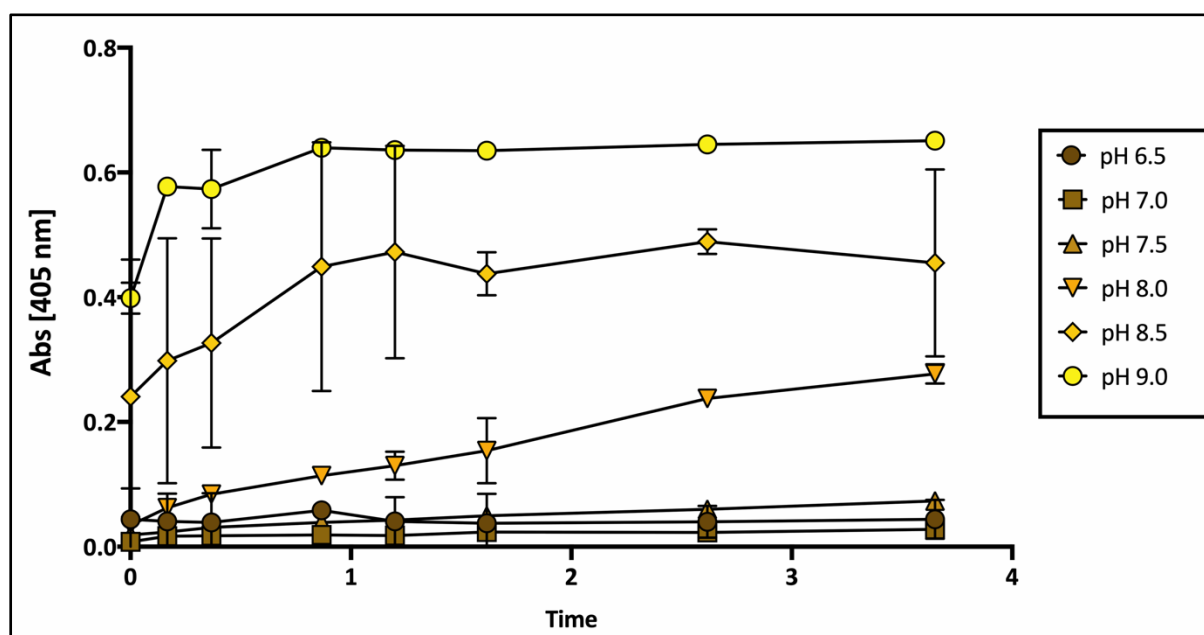


Figure 45: pH Dependence of non-enzyme catalysed Hydrolysis of 1 mM pNP-PEtN.

Increasing the pH of the buffer increases the rate of non-enzymatic hydrolysis of pNP-PEtN over time. pH 7.5 is likely the breaking point for pNP-PEtN stability. Non-enzymatic hydrolysis of pNP-PEtN was kept at a minimum at pHs 6.5 and 7.0, with a pH of 7.0 being the more stable one. This was performed with technical duplicates, standard error of the mean is presented in bars.

5.2.4 Assessing the activity of recombinant MCR catalytic domains.

The ability of the purified recombinant catalytic domain constructs cMCR-1 and cMCR-2 to cleave *p*NP-PEtN and catalyse the release of *p*NP was next assessed by incubating the purified enzymes with *p*NP-PEtN at pH 7.0. Initial experiments showed consistently higher levels of activity with cMCR-2, alongside difficulties obtaining reproducible levels of activity with cMCR-1, subsequent experiments then focused upon cMCR-2. Different concentrations of purified cMCR-2 were incubated with an excess of *p*NP-PEtN, and absorbance monitored at 405 nm, cleavage of PEtN was detected as the release of *p*NP (Figure 46). Importantly, while a clear difference between control (no enzyme) activity and *p*NP release was not found at the lowest enzyme concentration tested (10 μ M) as enzyme concentration increased up to 100 μ M, the release of *p*NP into solution accelerates. This shows that the reaction is enzyme-catalysed, i.e. that cMCR-2 is capable of accepting *p*NP-PEtN as a substrate, participating in the cleavage of the PEtN and release of *p*NP, and demonstrating that the isolated recombinant MCR-2 catalytic domain retains some catalytic activity.

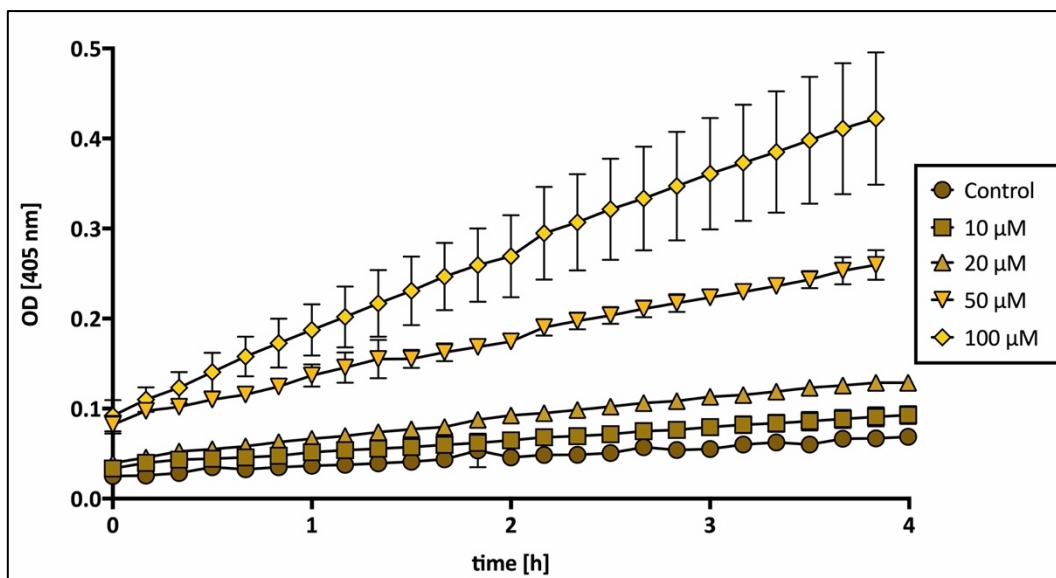


Figure 46: Reaction of pNP-PEtN at different concentrations of the soluble domain cMCR-2.

1 mM of pNP-PEtN was incubated for 4 hours at pH 7.0 with varying concentrations of cMCR-2. 1 mM of pNP-PEtN in solution with no cMCR-2 present was used as a non-enzymatic hydrolysis control. This was done with technical triplicates, standard error of the bars have been added to all plotted points.

5.2.4.1 LC-MS Analysis.

The enzyme-catalysed hydrolysis of pNP-PEtN by cMCR-2 was also confirmed by liquid chromatography coupled to mass spectrometry (LC-MS) analysis. 1 mM solutions of the pure 4-Nitrophenol-PEtN (pNP-PEtN), phosphoethanolamine (PEtN) and 4-nitrophenol (pNP) were analysed to determinate the elution times of said compounds. The chromogenic substrate pNP-PEtN elutes first at 6.7 minutes, followed by pNP, eluting around 10.5 minutes; PEtN did not show a peak and thus a standard elution time to measure PEtN was not established.

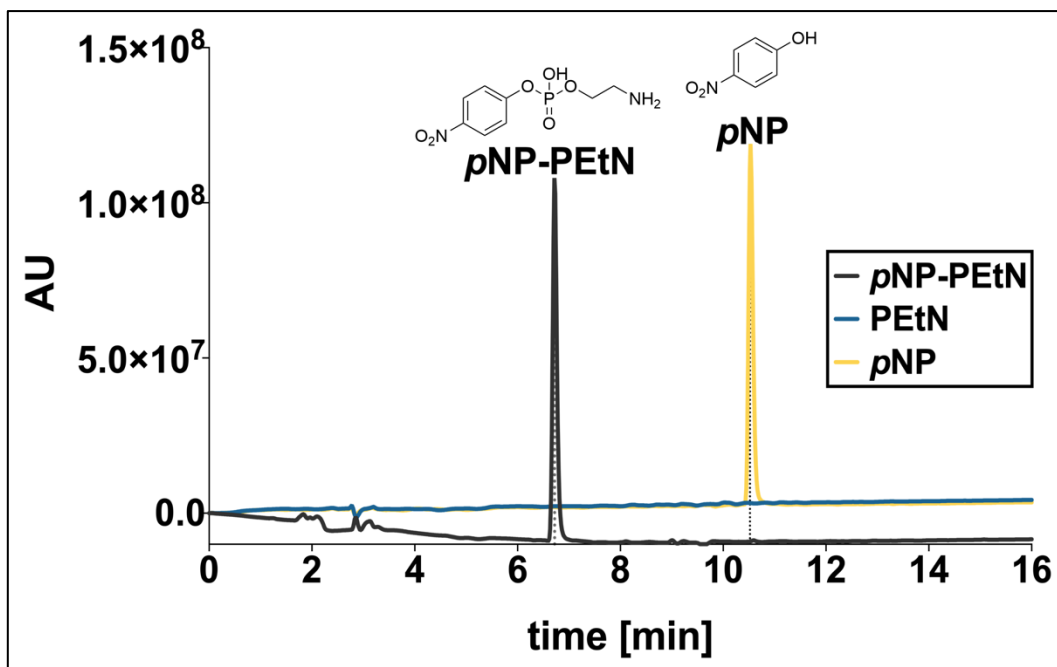


Figure 47: LCMS analysis of pNP-PEtN, pNP and PEtN.

LC-MS *pNP-PEtN*, *pNP* and *PEtN* standards. A pure sample of 1 mM *pNP-PEtN* showed an elution time of 6.71 ± 0.02 ; 1 mM *pNP* has an elution time of 10.57 ± 0.03 mins. No traces were found for *PEtN*. The units displayed are the area under the peak measured by Diode-Array Detection.

After establishing the elution times for standards an enzymatic reaction was set up using cMCR-2. After overnight incubation, the enzyme reaction was completed and stopped by protein precipitation. cMCR-2 was precipitated using trichloroacetic acid (TCA) and removed by centrifugation. The recovered pellet was analysed by peptide mass spectrometry fingerprinting to investigate the possible presence of a cMCR-2-PEtN covalent intermediate containing PEtN attached to the nucleophile cMCR-2 Thr283 (Figure 44 b). The supernatant was analysed by LC-MS to determine the release of *pNP* by enzymatic hydrolysis. A negative control was analysed to assess the degree of non-enzymatic hydrolysis that *pNP-PEtN* undergoes. LC-MS analysis of the supernatants confirmed that the C18 column could readily distinguish between intact *pNP-PEtN* (elution peak at 6.72 min Figure 47 grey line) and released *pNP* (elution peak at 10.52 min, Figure 47 yellow line) and that, in a control reaction carried out without addition of cMCR-2, non-enzymatic hydrolysis is kept at a minimum even after overnight incubation (Figure 48 b). An analysis of the area under the peak (1410138 AU units) for the non-enzyme hydrolysed *pNP*, using a calculated standard curve (Figure 49,

Equation 5), estimates a final *p*NP concentration of 139 μM , this translates to 0.02085 μmoles of *p*NP being released. This suggests that the chromophoric substrate *p*NP-PEtN is sufficiently stable in solution at room temperature throughout the assay to enable the investigation of enzyme-catalysed hydrolysis.

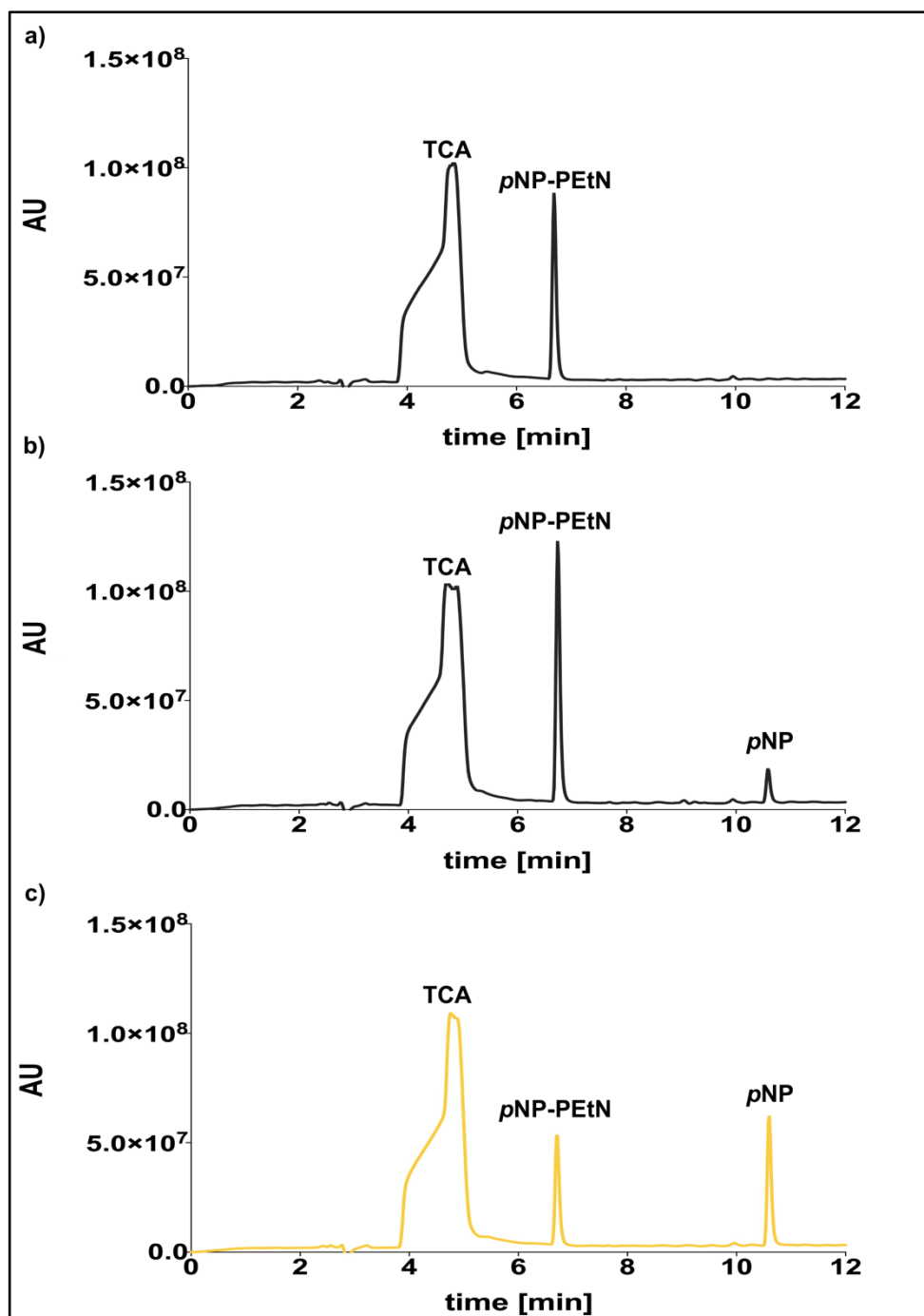


Figure 48: LCMS analysis of the hydrolysis of pNP-PEtN catalysed by cMCR-2.

a) 1 mM of pNP-PEtN at $t=0$, before enzymatic degradation. b) Negative control, the sample was taken after overnight incubation. A small peak was detected at 10.58 minutes, corresponding to the elution time of pNP. c) 1 mM of pNP-PEtN incubated overnight in the presence of cMCR-2 (50 μM). 10% TCA was used to precipitate the enzyme cMCR-2-PEtN complex. 0.082 ± 0.005 μmoles of pNP were released through enzymatic hydrolysis. All samples were treated with 10% (w/v) trichloroacetic acid (TCA) for protein precipitation, TCA elutes around 4.74 ± 0.08 mins. Experiment was done in duplicate, representative traces are shown.

A standard curve was generated to estimate the amount of pNP-PEtN being released by enzymatic hydrolysis (Figure 49). Fractionation of the reactions, after incubating pNP-PEtN (1 mM) in the presence of cMCR-2, identified the amount of pNP detected in solution as approximately 550 μ M (0.082 μ moles of pNP, area under the peak of 5489961 ± 374251.8 , Equation 5); i.e. nearly 4 times the amount of pNP released by non-enzymatic hydrolysis (Figure 48), demonstrating enhanced turnover of pNP-PEtN in the presence of cMCR-2. The amount of pNP released is also greater than the amount of enzyme present in the reaction (50 nM), indicating that under these conditions cMCR-2 is catalysing multiple turnovers of pNP-PEtN.

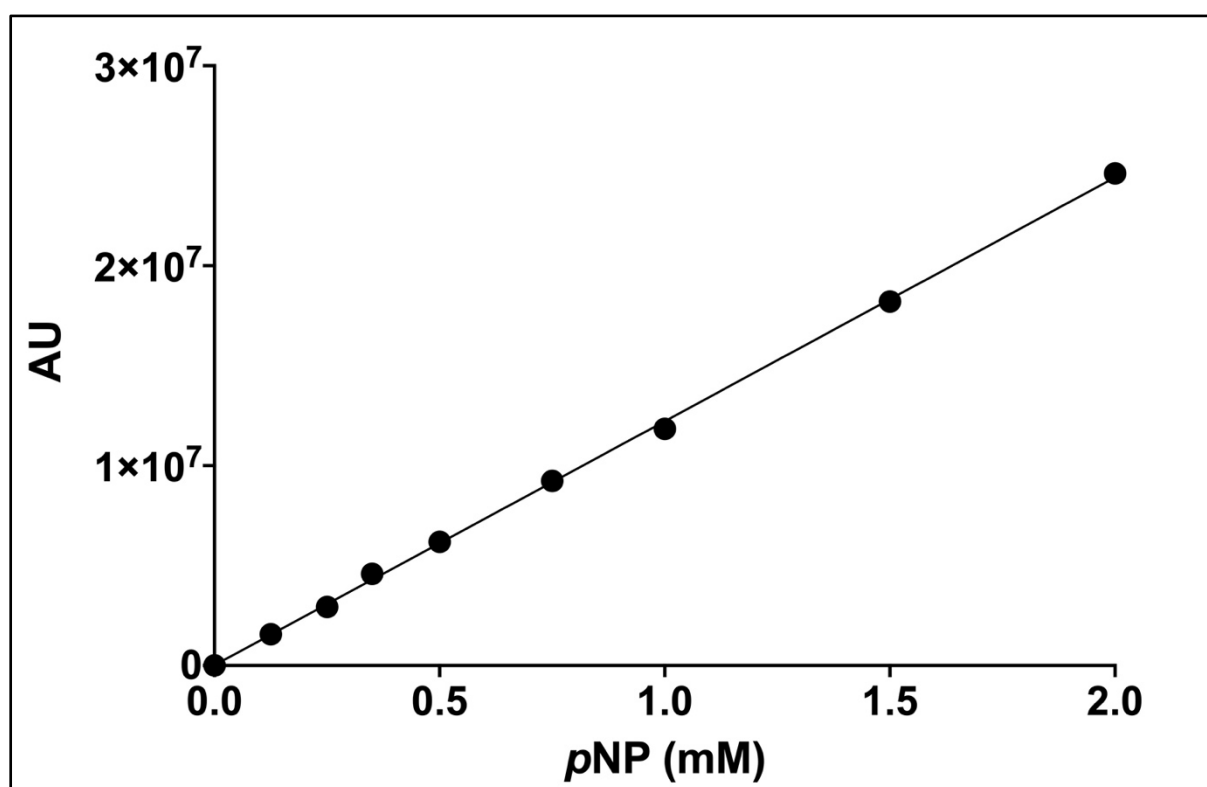


Figure 49:pNP standard curve.

Standard curve to measure the release of pNP. R2 = 0.999

Similarly, peptide mass fingerprinting of cMCR-2 after reaction with pNP-PEtN revealed that the relative abundance of peptide sequences matching phosphorylated Thr283 was twice that of the unphosphorylated version (Table 10, phosphopeptides marked in bold letters).

Importantly, however, no evidence was found for the presence of the covalent Thr²⁸³-PEtN intermediate which would be expected to accumulate if cMCR-2 was capable of catalysing PEtN addition to Thr283, but not its subsequent release. This also suggests that cMCR-2 is not only able to catalyse the release of *p*NP from the *p*NP-PEtN substrate but that PEtN is then released from the active site, enabling cMCR-2 to catalyse multiple rounds of reaction.

Table 10: cMCR-2 Mass fingerprint analysis and peptide modifications.

Protein	Sequence Coverage	Unique Peptides	Number of PSMs*
MCR-2 Catalytic domain.	98.45	45	1689
Peptide Sequence		# PSMs*	Modifications
ADHVQFNQYGR		19	
ADHVQFNQYGRETFPQLAK		11	
AVPAFFWSNNTTFKPTASDTVLTHDAITPTLLK		59	
cEHQSLINAYDNALLATDDDFIAK		85	C1(Carbamidomethyl)
DAVQTTKPSER		40	
DAVQTTKPSERKPR		1	
DAVQTTKPSERKPRLLVVFVGETAR		10	
DmLImLHQmGNHGPAAYFK		25	M2(Oxidation); M5(Oxidation); M9(Oxidation)
DmLImLHQmGNHGPAAYFK		129	M2(Oxidation); M5(Oxidation)
DMLImLHQmGNHGPAAYFK		140	M5(Oxidation)
DMLIMLHQmGNHGPAAYFK		108	
DMLImLHQmGNHGPAAYFKR		2	M5(Oxidation); M9(Oxidation)
DMLImLHQmGNHGPAAYFKR		30	M5(Oxidation)
DMLIMLHQmGNHGPAAYFKR		8	
DNNSDSKGVmDK		2	M10(Oxidation)
DNNSDSKGVmDKLPATQYFDYK		32	M10(Oxidation)
DNNSDSKGVmDKLPATQYFDYK		18	
DVGmLVGLDDYVSANNGK		8	M4(Oxidation)
DVGMLVGLDDYVSANNGK		11	
DVGMLVGLDDYVSANNGKDMLImLHQmGNHGPAAYFK		3	M23(Oxidation); M27(Oxidation)
DVGMLVGLDDYVSANNGKDMLImLHQmGNHGPAAYFK		3	M23(Oxidation)
DVGMLVGLDDYVSANNGKDMLIMLHQmGNHGPAAYFK		2	
DVGMLVGLDDYVSANNGKDMLIMLHQmGNHGPAAYFKR		2	
EQRAVPAFFWSNNTTFKPTASDTVLTHDAITPTLLK		3	
ETFPQLAK		52	
FTPVcEGNELAK		60	C5(Carbamidomethyl)
FTPVcEGNELAKcEHQSLINAYDNALLATDDDFIAK		1	C5(Carbamidomethyl); C13(Carbamidomethyl)
GVmDKLPATQYFDYK		45	M3(Oxidation)
GVMDKLPATQYFDYK		20	
GVMDKLPATQYFDYKSATNNTIcNTNPYNEcR		3	C23(Carbamidomethyl); C31(Carbamidomethyl)
KPRLVVFVGETARADHVQFNQYGR		1	
LFDVTAGK		94	
LFDVTAGKVK		7	
LFDVTAGKVKDR		6	
LGVGILWR		40	
LGVGILWRDNNSDSK		5	
LGVGILWRDNNSDSKGVMDK		1	
LPATQYFDYK		67	
LPATQYFDyKSATNNTIcNTNPYNEcR		3	Y9(Phospho); C18(Carbamidomethyl); C26(Carbamidomethyl)
LPATQYFDYKSATNNTIcNTNPYNEcR		3	C18(Carbamidomethyl); C26(Carbamidomethyl)
LVVFVGETAR		99	
LVVFVGETARADHVQFNQYGR		8	
LVVFVGETARADHVQFNQYGRETFPQLAK		1	
RYDEQFAK		47	
RYDEQFAKFTPVcEGNELAK		6	C13(Carbamidomethyl)
SATNNTIcNTNPYNEcR		82	C8(Carbamidomethyl); C16(Carbamidomethyl)
SATNNTIcNTNPYNEcRDVGmLVGLDDYVSANNGK		1	C8(Carbamidomethyl); C16(Carbamidomethyl); M21(Oxidation)
SATNNTIcNTNPYNEcRDVGMLVGLDDYVSANNGK		7	C8(Carbamidomethyl); C16(Carbamidomethyl)
SIDWLK		40	
THEANYDVAmLYVSDHGESLGENGVYLHGmPNAFAPK		12	M10(Oxidation); M30(Oxidation)
THEANYDVAMLYVSDHGESLGENGVYLHGmPNAFAPK		18	M30(Oxidation)
THEANYDVAMLYVSDHGESLGENGVYLHGmPNAFAPK		18	
THEANYDVAMLYVSDHGESLGENGVYLHGmPNAFAPKEQR		3	M30(Oxidation)
THEANYDVAMLYVSDHGESLGENGVYLHGmPNAFAPKEQR		1	
TIYHAKDAVQTTKPSER		7	
VDGLANFSQVTS ^c GT St A ^S VS ^P cMFSYLGQDDYDVDTAK		25	C13(Carbamidomethyl); T17(Phospho); C23(Carbamidomethyl)
VDGLANFSQVTS ^c GT St A ^S VS ^P cMFSYLGQDDYDVDTAK		5	C13(Carbamidomethyl); Y19(Phospho); C23(Carbamidomethyl); M24(Oxidation)
VDGLANFSQVTS ^c GT St A ^S VS ^P cMFSYLGQDDYDVDTAK		1	C13(Carbamidomethyl); C23(Carbamidomethyl); M24(Oxidation)
VDGLANFSQVTS ^c GT St A ^S VS ^P cMFSYLGQDDYDVDTAK		1	C13(Carbamidomethyl); C23(Carbamidomethyl)
YDEQFAK		57	
YDEQFAKFTPVcEGNELAK		8	C12(Carbamidomethyl)
YQENVLDTLDR		81	
YQENVLDTLDRLGVGILWRDNNSDSK		2	

*PSMs: Peptide Spectrum Matches.

5.2.5 Estimation of the steady-state kinetic parameters for cMCR-2.

Having demonstrated cMCR-2-catalysed release of *p*NP from *p*NP-PEtN, an activity assay was performed with a broad range of *p*NP-PEtN concentrations to investigate the dependence of reaction rate upon concentration and to use the resulting data to estimate the kinetic parameters for steady-state hydrolysis, the Michaelis constant, K_M , and the turnover number, k_{cat} , obtained from the maximum reaction velocity, V_{max} . To obtain the initial reaction rates of the reaction a constant enzyme concentration of 50 μ M cMCR-2 was incubated with increasing concentrations of *p*NP-PEtN at pH 7.0 for a total of 18 h (Figure 50). The rate of the reaction was measured as the slope obtained from straight-line fitting of the initial steps of the enzyme-catalysed reaction, i.e. the increase in *p*NP formation divided by the change in time. The first 4 hours of the reaction were used to obtain the reaction rates as this corresponded to the linear phase of the reaction. The amount of released *p*NP was estimated using a standard curve (Figure 51 a)) and Equation 8.

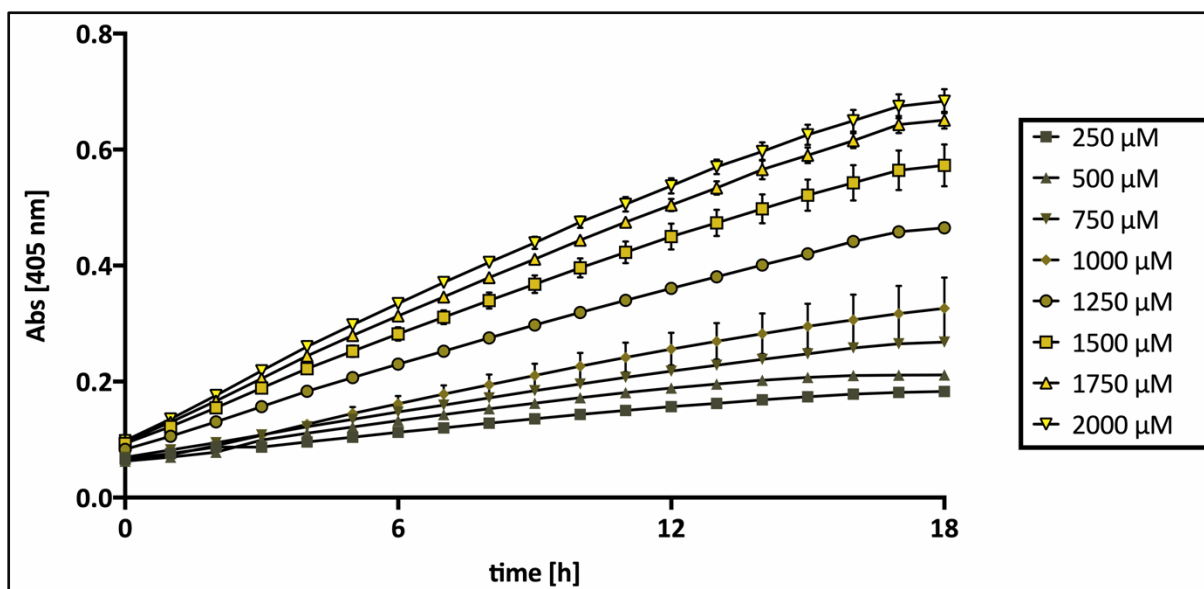


Figure 50: Reaction profile of varying concentrations of pNP-PEtN with the soluble domain of MCR-2 (50 μM).

The first 4 hours of reaction remained linear and were used to obtain the reaction rates used to estimate the kinetic parameters K_m and V_{max} . Experiment was performed in triplicate, error bars represent standard error of the mean.

The plot of the initial rate against pNP-PEtN concentration (Figure 51 b) shows a broadly hyperbolic character, indicating that the reaction is tending towards saturation at substrate concentrations above 1 mM, consistent with the expected Michaelis-Menten behaviour. The kinetic parameters were estimated from the experimental results both by the direct fitting of the data to the Michaelis-Menten equation (Equation 6 Figure 51 a) and by its double reciprocal linearization, the Lineweaver-Burk plot (Equation 7 Figure 51b)). The estimated kinetic parameters, K_M and the maximum reaction velocity V_{max} , from both graphical methods, (Figure 51), are a K_M of 0.82 mM and 1.27 mM and a V_{max} of 2213 nMmin^{-1} and 2694 nMmin^{-1} , Michaelis-Menten and Lineweaver-Burk respectively. These fits indicate a K_M value of approximately 1 mM and a k_{cat} value of 0.000738 sec^{-1} . Thus pNP-PEtN is a relatively weak substrate, with mM affinity and slow turnover, but is capable of reporting cMCR-2 activity and therefore provides a platform for further investigations of enzyme mechanism and inhibition.

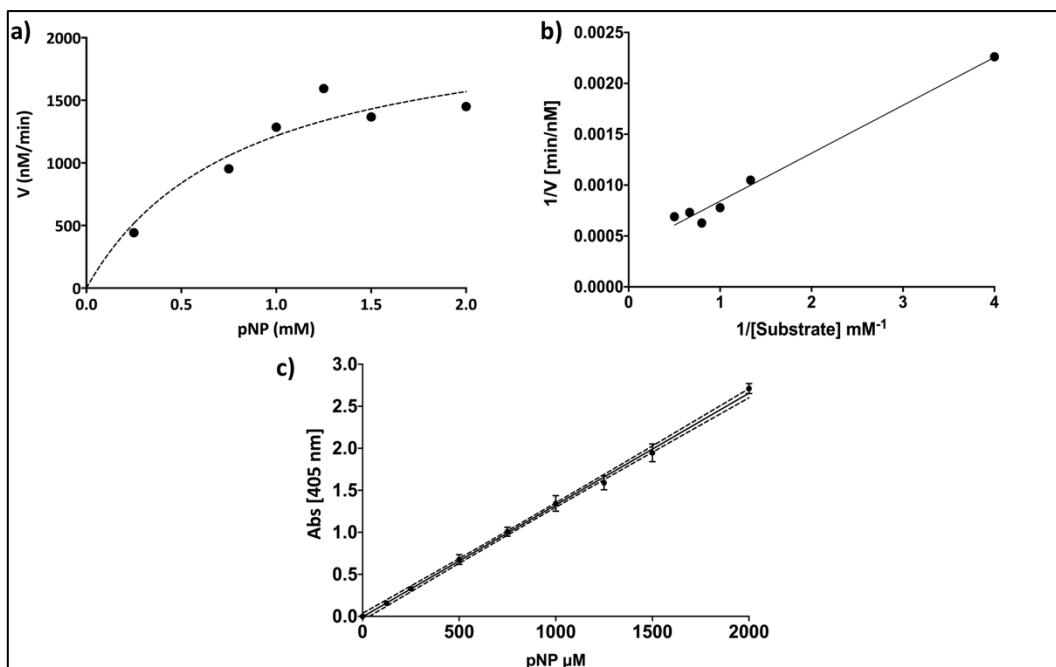


Figure 51: Estimation of the kinetic parameters for cMCR-2 activity.

a) Michaelis-Menten fitted model for the reaction as a function of the substrate, pNP-PEtN, concentration in mM. The dotted line represents the best-fitted model (R^2 of 0.07). b) Lineweaver-Burk double reciprocal plot for the reaction rate at different substrates (R^2 of 0.98). c) linear regression plot for the estimation of pNP. (R^2 of 0.99), This was done with triplicates.

5.2.6 Inhibition of MCR-2 by Thioglycolic acid.

An important use of assays for MCR activity, particularly as applied to the isolated, soluble, catalytic domain, is the assessment of inhibitor activity. As only very limited information is available regarding candidate inhibitors of MCR-1 a potential inhibitor, thioglycolic acid (TGA, also known as mercaptoacetic acid, Figure 52 a) was identified from the literature, based upon its small size, accessible zinc-ligating functional groups (thiol and carboxylate) and known affinity for other zinc metalloenzyme systems [260]. TGA has also been reported to have an inhibitory effect on alkaline phosphatases [261]. Based on this, we hypothesise that TGA has the potential to interact with the MCR zinc centre as shown in Figure 52 b. TGA inhibition of cMCR-2 was then also tested using the pNP-PEtN activity assay described above.

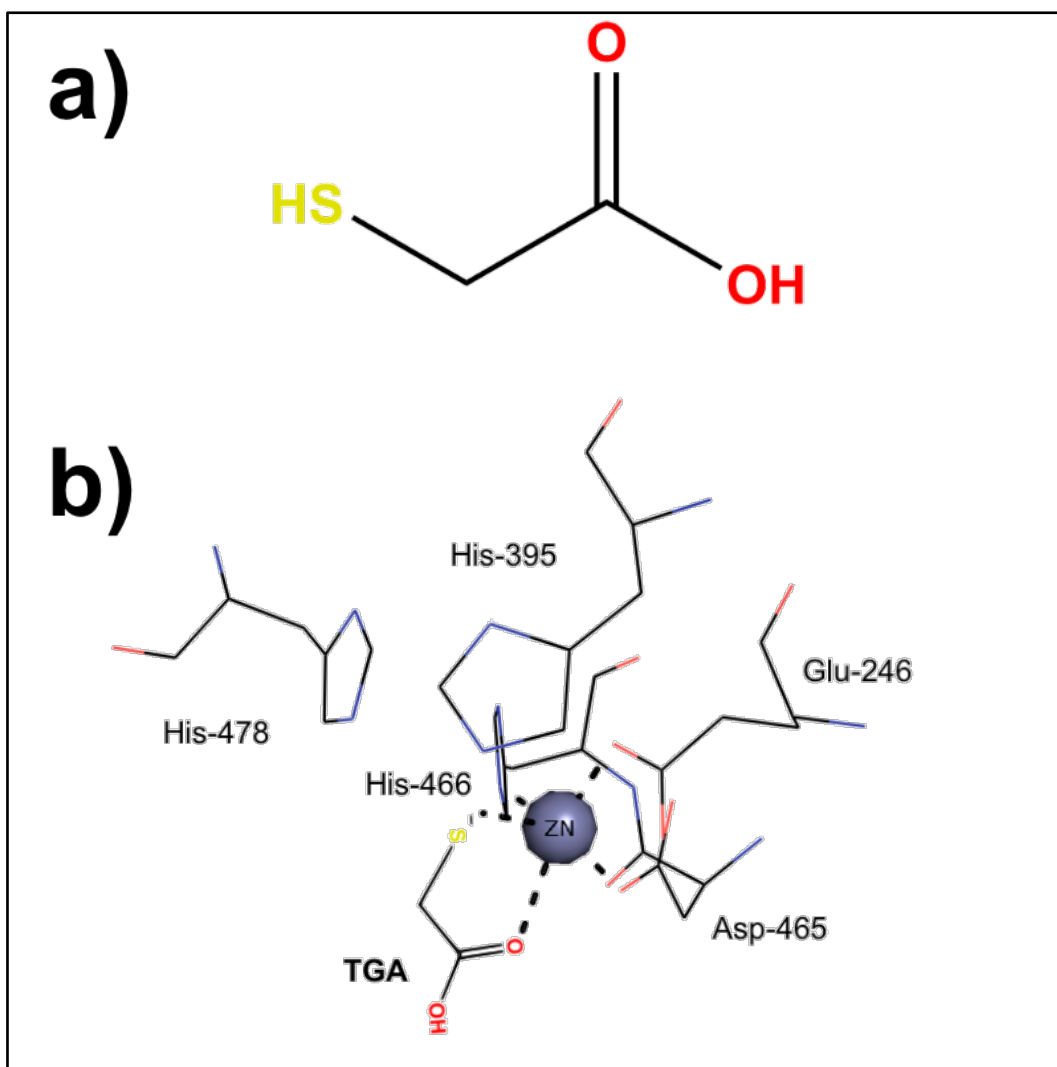


Figure 52: Thioglycolic acid as an MCR-1 inhibitor candidate.

a) Chemical structure of Thioglycolic acid (TGA). b) Possible interaction of TGA with the zinc ion present in the active site of MCR-1. TGA can interact with the zinc ion, forming chelating rings, disrupting the enzyme activity. Figure adapted from [262] using the software Pymol.

An activity assay was performed, using cMCR-2, EDTA-treated cMCR-2 and cMCR-2 in the presence of 1 mM TGA, to assess the activity of TGA as an inhibitor (Figure 53). The use of chelators, such as EDTA, in laboratory strains of *E. coli* expressing the full-length enzyme MCR-1, showed a reduction of the enzyme activity, in the form of reduction of colistin MICs [74,263] while the related alkaline phosphatases are easily inhibited by divalent cation chelating agents [264]. The presence of 1 mM TGA in the reaction seems to have a similar effect on hindering the activity of cMCR-2 compared to the EDTA-treated cMCR-2 used as a positive control. Soaking the enzyme in EDTA did not completely abolish activity, as was also

the case in a previous study of the *N. meningitidis* LptA catalytic domain [111]. With these results, we can conclude that the presence of zinc in the catalytic domain is important for the proper activity of the enzyme, as the removal of it using chelators such as EDTA harmed the enzyme activity. We also demonstrated that this activity assay works as a platform for small molecule inhibitors of MCR enzymes and other phosphoethanolamine transferases. Thioglycolic acid can act as an inhibitor for cMCR-2 as it reduced the activity of the enzyme, observed as a reduction of the final absorbance in our activity assay.

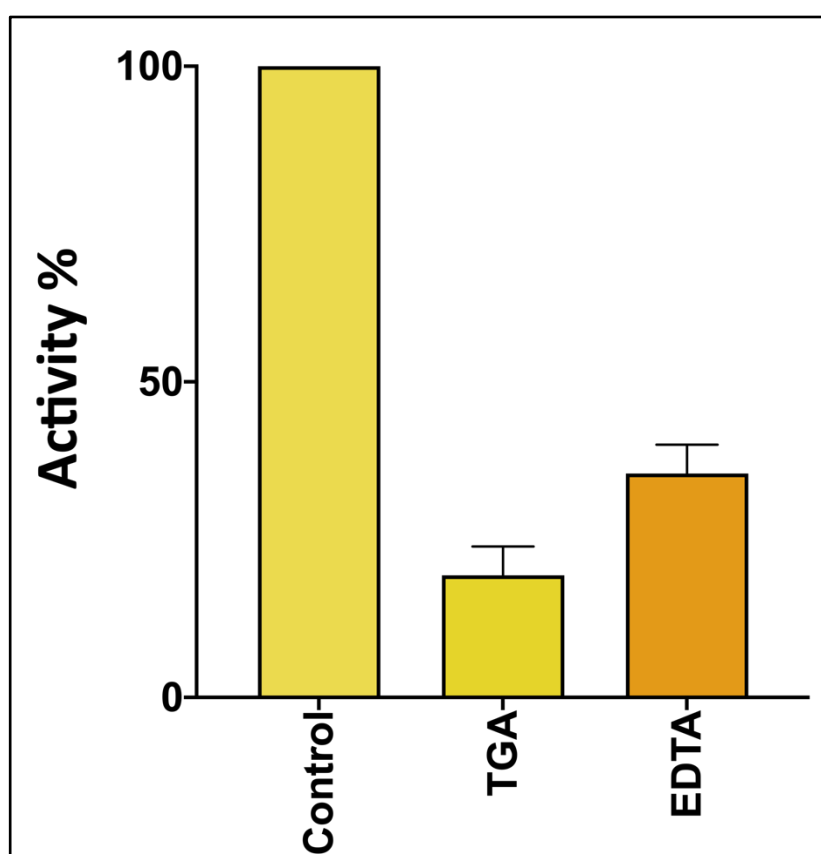


Figure 53: Effect of Thioglycolic acid (TGA) and EDTA on activity of cMCR-2.

Endpoint measurement (absorbance, 405nm, after subtraction of the no enzyme control) after 16 hours incubation of pNP-PEtN in the presence of 50 μ M cMCR-2 (Control, set to 100% activity), EDTA treated cMCR-2 (EDTA) and 50 μ M cMCR-2 plus 1 mM TGA (TGA). This experiment was performed by technical replicates.

5.3 Conclusions.

A chromogenic assay to monitor the activity of the catalytic domains of MCR-1 and MCR-2, with potential application to other phosphoethanolamine transferases, was developed to assess the functional integrity of the recombinant soluble proteins. Synthesis of the chromophore *p*NP-PEtN (para-nitrophenol-phosphoethanolamine) was successfully achieved by chemistry collaborators and it was used in combination with the soluble domains of MCR-1 and MCR- for the development of a chromogenic activity assay. The assay monitors the cleavage of the phosphoethanolamine (PEtN) motif from the chromophore *p*NP-PEtN and the release of *p*NP, however, the full catalytic reaction was not fully studied as the assay was performed in the absence of the membrane-bound natural acceptor substrate, Lipid A, thus the kinetic parameters obtained are only relevant to the first half of the reaction (transfer of phosphoethanolamine onto the acceptor threonine).

The recombinant MCR catalytic domains were used in conjunction with *p*NP-PEtN to develop a chromogenic assay and the results of this were used to estimate enzymatic parameters such as the reaction rate and the K_M . The relatively high K_M , approx. 1 mM, can be interpreted as a low affinity for the substrate *p*NP-PEtN, that is not lipidated as MCR-1 physiological substrate, Lipid A. By using *p*NP-PEtN along with the catalytic domains of MCR-1 and MCR-2 (namely cMCR-1 and cMCR-2) we were able to produce an assay that serves as a scaffold for further studies of this mechanism. The use of chromophores, such as para-nitrophenyl phosphate (*p*NPP), is common in the analysis of the enzymatic activity of alkaline phosphatases since the degradation of the substrate can be directly related to the release of the 4-Nitrophenol group (*p*NP) [259]. The release of *p*NP is easily monitored as it changes the colouration of the solution to bright yellow tone and can be easily monitored by absorbance at a wavelength of 405 nm. The chromophore *p*NP-PEtN has been previously used to assess the activity of the bacterial cellulose synthase subunit G (BcsG) from *E. coli*, an enzyme that also functions as

a phosphoethanolamine transferase. A construct of the C-terminal domain of BcsG was able to cleave the phosphoethanolamine group (PEtN) liberating free 4-nitrophenol [265]. The calculated V_{\max} figure of 2213 nMmin⁻¹ for the assay with 50mM cMCR-2 corresponds to a specific activity of 1.228 nmols/min/mg. This is similar to the value of 3.1 nmols/min/mg reported for *p*NP-PEtN hydrolysis catalysed by BcsG. By comparing these specific activities, and given that cMCR-2 was able to catalyse the removal of *p*NP from *p*NP-PEtN, we conclude that the catalytic domain of MCR-2, here named cMCR-2, possesses phosphoethanolamine transferase activity.

Wanty et al. also identified this substrate, *p*NP-PEtN, as an activity reporter for a soluble domain construct of the EptA enzyme from *Neisseria meningitidis* [111]. In this work, the reaction was measured after a 52 hours incubation, measuring the absorbance at a wavelength of 405 nm. The activity of EptA was reported as a final optical density measurement of 0.65 (measure at a wavelength of 405 nm), 3 times higher than the response obtained from a denatured enzyme control of 0.185 or by non-enzymatic hydrolysis of the substrate at (0.19). These results proved that the C-terminal soluble domain of EptA had retained its enzymatic activity and was capable of accepting *p*NP-PEtN as a substrate. It is worth mentioning that the amount of *p*NP-PEtN used in this experiment was 25-fold [111] the amount used in our study, and there is a lack of information regarding characterization of the synthetic *p*NP-PEtN substrate. As our synthetic collaborators identified significant problems with the published protocol for *p*NP-PEtN synthesis (low yields and formation of incorrect products) comparison of activity levels is difficult.

The catalytic activity of the purified cMCR-1 could not be accurately assessed despite several attempts. One reason for the lack of reproducible activity may be attributed to variability in the levels of phosphorylation of the purified enzyme. Assay of MCR activity is complicated by phosphorylation of the catalytic Thr residue (Thr285 in cMCR-1, Thr283 in cMCR-2) that

results in a mixed population of enzymes that are likely to differ in their activity towards substrates such as *p*NP-PEtN. Phosphorylation of cMCR-2 Thr283 was confirmed by mass spectrometry peptide fingerprinting, which showed a mixed distribution of peptides containing the catalytic site, consistent with the presence of phosphorylated (active cMCR-2) and non-phosphorylated (inactive cMCR-2) forms. The phosphorylation of catalytic threonine has also been seen in other phosphoethanolamine transferases such as *Campylobacter jejuni* EptC, with the residue Thr266 being the intermediate acceptor in the two-step transferase reaction [266]; or *Neisseria meningitidis* EptA 108. where a mixture of states, phosphorylated and non-phosphorylated Thr280 in the catalytic site, was described. All of these enzymes are considered part of the alkaline phosphatase superfamily; proteins of this family share a conserved ping-pong mechanism that starts with a nucleophilic attack on the central phosphorus atom, by the metal activated nucleophilic residue leading to formation of a covalent intermediate [142]. Compared to cMCR-1, the purified catalytic domain of MCR-2 (cMCR-2) was shown to provide better levels of catalytic activity along with structural stability making it the preferred candidate to test the developed activity assay.

In their study of *EcBcsG*, an *E. coli* bacterial cellulose synthase subunit G with phosphoethanolamine transferase activity, Anderson et al. presented evidence of the formation of a covalent PEtN–enzyme intermediate when using the substrate *p*NP-PEtN [265]. In that work, it was determined that the covalent PEtN–enzyme intermediate is particularly short-lived and solvent-labile and likely dissociates before tryptic digestion. The possible short life and instability of this intermediate is consistent with our observation that modification of the active Thr283 as Thr²⁸³-PEtN was not detected during mass spectrometry fingerprinting of cMCR-2 after incubation with *p*NP-PEtN.

The assay reported here uses the isolated recombinant catalytic domain of MCR-1 only. Anandan et al.[121] tested the ability of both recombinant *NmEptA* and the soluble periplasmic

domain of the same enzyme, here referred to as *cNmEptA*. The full-length *NmEptA* was found to be able to remove phosphoethanolamine from a fluorogenic lipid substrate, 1-acyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]dodecanoyl}-sn-glycerol-3-phosphoethanolamine (NBD-PEtN); however, the soluble domain *cNmEptA* was unable to catalyse the removal of phosphoethanolamine from this fluorogenic compound; leading these authors to conclude that for the enzyme to be active against lipidated substrate, the presence of the transmembrane domain is needed. However, both this, and the earlier study of Wanty et al on turnover of pNP-PEtN by *cNmEptA*, did not consider the time-dependence of the reaction, which is essential for detailed, quantitative study of an enzyme reaction and its kinetics. Our data suggest that this combination of the a soluble catalytic domain, in this case from MCR-2, with the chromogenic compound *pNP-PEtN* serves as a simple tool to measure not only enzymatic activity but to measure the effectivity of candidate inhibitors. Thioglycolic acid has been studied as an inhibitor of alkaline phosphatase enzyme activity, both as a free ligand and in a vanadyl-TGA complex, and in both instances it caused inhibition of the enzyme activity [261]. Based in our simple inhibition test using TGA as an candidate inhibitor, *cMCR-2* activity was lowered in the presence of free TGA, wen compared against the chelator EDTA, suggesting that this assay can work as a stepping stone for the testing of MCR inhibitor candidates.

The development of an activity assay is the first step to further understanding the mechanism and structure-function relationships of an enzyme, as well as to aid in the identification of possible inhibitors. Determining enzyme activity can also serve to confirm the integrity of the enzyme preparations and estimate the amount of active enzyme obtained after the production and purification process, an important consideration when (as here due to phosphorylation of the enzyme in situ during recombinant production) activity can vary between samples. It has enabled for the first time determination of kinetic parameters for the first stage of the phosphoethanolamine transfer reaction and, as demonstrated by inhibition of *cMCR-2* by TGA, can serve as a scaffold for the identification of possible inhibitors.

Chapter 6 Production and characterisation of the MCR homologue PmrC from *Acinetobacter baumannii*.

6.1 Introduction:

The Gram-negative bacterium *Acinetobacter baumannii*, a member of the ESKAPE group (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) [267], has now become one of the most prevalent causes of multidrug-resistant nosocomial infections. For more than 50 years this organism has been a leading cause of cases of nosocomial pneumonia infections [268], targeting vulnerable patients, such as those who are already critically ill. Its significance as a pathogen has increased over the last 20 years, due to its ability to acquire antimicrobial resistance genes, with the emergence of *A. baumannii* strains resistant to almost all antibiotics available turning *A. baumannii* into a major threat to current medical practices and the use of antibiotics [268].

6.1.1 *Acinetobacter baumannii* microbiology and clinical relevance.

Acinetobacter baumannii is the most clinically relevant representative of the *Acinetobacter* genus; this organism has become one of the most problematic nosocomial pathogens globally. The genus *Acinetobacter*, from the Greek *akinetos* that means “nonmotile”, was initially coined to refer to the non-motile microorganism from the *Achromobacter*, it was not until the 1970's that this genus was acknowledged and at least 15 different species were recognized to be part of it [268,269]. Members of this genus cannot be further classified into subspecies based on phenotypic characterization. This genus is comprised of aerobic, non-fermenting, Gram-negative coccobacilli, however, due to these characteristics, identification of these organisms can only reach the level of the genus as they can be easily misidentified as other bacteria with similar phenotypes. Based on DNA-DNA hybridization, used as the standard technique for the

identification of members of the *Acinetobacter* genus, 12 groups or genospecies have been identified including *A. baumannii*, *A. calcoaceticus*, *A. haemolyticus*, *A. johnsonii*, *A. junii*, and *A. Iwoffi*. Of these *A. baumannii* is the most clinically relevant [270,271].

A. baumannii is an opportunistic nosocomial pathogen, mostly infecting critically ill patients, i.e. patients that have had prolonged stays in hospital facilities, patients with a depressed immune system or patients that have gone under invasive procedures, such as surgery, mechanical respirators or catheters. It causes a range of infections from skin infections and urinary tract infections to meningitis, bacteraemia and pneumonia, being the most common infection reported in clinical settings. Measuring the mortality rates associated with *A. baumannii* infections is a difficult task given its status as an opportunistic pathogen and the critical conditions of the infected patients, however, a rough estimate for intensive care unit (ICU) patients puts its mortality rates in a range from 20% to 68% [272].

The ability of *A. baumannii* to survive for prolonged periods on dry surfaces, much longer compared to *E. coli* and other *Enterobacter* species, its ability to adhere and form biofilms as well as its propensity to acquire resistance to multiple antibiotics and disinfectants has turned *A. baumannii* into the number one priority, over other antimicrobial resistant organisms such as Carbapenem-resistant and ESBL-producing *Enterobacteriaceae*, for several international health organizations such as the World Health Organization (WHO), the European Centre for Disease Control and Prevention (ECDC) [273] and the American Centre for Disease Control and Prevention (CDC) [274]. Furthermore, the emergence of carbapenem-resistant *Acinetobacter* spp. extremely limits the therapeutic options for patients infected with this organism [275].

6.1.2 Community-acquired *A. baumannii* infections.

Outside of clinical settings, *A. baumannii* is still a feared pathogen that causes community-acquired infections (CAI), predating on those with underlying conditions such as diabetes as well as heavy smokers and alcohol drinkers [276]. Beyond nosocomial infections, community-acquired *A. baumannii* occurs primarily in countries with hot and humid environments. Although rare, community-acquired infections of *A. baumannii* have been reported in Australia and Asia. These infections were mostly documented among aboriginal communities [277]. In the tropical regions of Australia and Asia, reports of community-acquired pneumonia (CAP) due to *A. baumannii* have occurred, with the majority of cases happening during the rainy season [278]. It is worth mentioning that the *A. baumannii* strains associated with both CAI and CAP are more susceptible to antibiotic therapy (compared to the strains found in nosocomial infections). Despite being a rare incidence, the consequences of CAP can be severe, having mortality rates as high (up to 60 %) as those for nosocomial associated *A. baumannii* infections [279]. *A. baumannii* infections are also commonly associated with conflict and are commonly present in combat wounds in soldiers in Iraq and Afghanistan, with some authors referring to it as “Iraqibacter” [268,280]. This phenomenon has led to the importation of multidrug-resistant strains into other countries such as the USA, UK [281] and other European countries, where community spread of *A. baumannii* is uncommon [268].

6.1.3 Multi-Drug resistance of *A. baumannii*.

A. baumannii is an organism already resistant to many antibiotics. According to the US CDC’s Emerging Infections Program and Multi-site Gram-negative Surveillance Initiative (MuGSI), between 2012 and 2017, in the United States of America there has been an average of 9,700 cases of carbapenem-resistant *Acinetobacter* infections in hospitalized patients. 90% of the *A. baumannii* isolates recovered were resistant to fluoroquinolones, 75% to extended-spectrum beta-lactams and 66% to trimethoprim/sulfamethoxazole treatment, making this

organism resistant to nearly all antibiotics available and difficult to treat [236]. The carbapenems: imipenem, meropenem, and doripenem have been generally used to treat *A. baumannii* infections, however, given the increasing rate of carbapenem-resistant *A. baumannii* (CR-AB) cases and the fact that these are more widespread, these have now become a significant clinical challenge. Outbreaks of carbapenem-resistant *A. baumannii* (CR-AB) have been increasing over the last 15 years; with reports of CR-AB in the southeast of England [282], Germany, France, Spain, the Netherlands, and Portugal [283,284]. In the American continent, cases of CR-AB have been seen from the early 90s, with an increased rate of imported CR-AB infections due to the conflict in Iraq and recent reports indicating that in the USA alone CR-AB caused an estimate of 8,500 nosocomial infections [274] In Latin America a variety of CR-AB have been identified in countries including Brazil, Colombia, Mexico and Argentina [268,285,286].

There are only a few therapeutic options available to deal with MDR *A. baumannii* infections. These alternatives revolve around the use of a combination of antibiotics to treat MDR *A. baumannii*; almost all of these combination therapies rely on the use of colistin [280]. As described above, colistin is a cationic antimicrobial peptide (CAMP) that has been labelled as a “last line of defence” antibiotic. It possesses an antimicrobial spectrum that affects a wide range of Gram-negative bacteria including *Acinetobacter baumannii* and plays a key role in combination therapies used to treat MDR *A. baumannii* infections, as the rate of colistin resistance in MDR *A. baumannii* is relatively low (10% of studied isolates) compared to resistance rates for rifampicin or tigecycline (47% and 45% of studied isolates respectively) [287]. These data make colistin the most effective antimicrobial agent against MDR *A. baumannii*.

6.1.4 *A. baumannii* and colistin resistance.

Reports of *A. baumannii* resistance to colistin have been increasing globally (Figure 54), with colistin-resistant isolates found in more than 20 countries in the world. Although resistance against colistin is not unheard of, and in some cases is considered uncommon, colistin resistance in *A. baumannii* is of special importance as this organism is capable of presenting colistin heteroresistance. Heteroresistance is a phenomenon that occurs where a bacterial population contains a subpopulation with a substantial reduction in antibiotic susceptibility compared with the main population [288]. Resistance to colistin, as well as other polymyxins, is usually found as a modification of the Lipid-A, the main target of colistin, generated by the enzyme-catalysed addition of positively charged residues that alter its charge. These mechanisms vary among GNB, most commonly positively charged moieties such as phosphoethanolamine (PEtN) and 4-amino-4-deoxy-L-arabinose (L-Ara4N) [289]. These modifications reduce the negative charge of Lipid A and subsequently lead to the repulsion of cationic polymyxins (above 1.4.3).

In the case of *A. baumannii*, colistin resistance may occur due to a combination of a series of mutations in the genes involved in lipid A synthesis, leading to a complete loss of Lipid A [77,290], the main target of colistin, but also increasing *A. baumannii* susceptibility to other antibiotics; as well as mutation involving the operon *pmrAB* [77] leading to the expression and activation of PmrC, a phosphoethanolamine transferase that catalyses the addition of PEtN to Lipid A. Recently another study has demonstrated that efflux pumps share a major role in colistin-heteroresistance in *A. baumannii* [291]. Some strains of *A. baumannii* harbour a small subpopulation of colistin-resistant cells which, under selective pressure, can rapidly overtake the whole population, creating newly colistin-resistant populations, subsequently leading to treatment failure [292,293]. Due to these factors, and the lack of effective alternative treatments, colistin resistance in *A. baumannii* is of great concern.



Figure 54: Global reports of colistin resistance in *A. baumannii*.

Colistin resistant A. baumannii Isolates have been detected in Algeria [294], Argentina [295], Australia [293], Brazil [296], Egypt [290], France [298], Germany [299], Greece [300], India [301], Iran [302], Italy [303], Japan [304], Kuwait [305], Mexico [285], Portugal [306], Romania [307], Saudi Arabia [308], South Korea [309], Spain [310], Taiwan [311], Tunisia [312], Turkey [313], USA [314].

6.1.4.1 Colistin resistance by Loss of LPS

One of the mechanisms involved in colistin resistance in *A. baumannii* is a series of mutations leading to the complete loss of Lipid A in the outer membrane of the cells, Lipid A being the main target of colistin. These mutations occur in a series of genes, *lpxA*, *lpxC*, or *lpxD*, essential for lipid A biosynthesis. These genes are generally present in other GNB such as *E. coli*, where 9 enzymatic steps have been identified to be responsible for the biosynthesis of LipidA [315]. Sequence comparison with *A. baumannii* genes showed that the genetic arrangement for the *lpxA* and *lpxD* genes is similar to that found in *E. coli*. A screening of 12 colistin-resistant derivatives of *A. baumannii*, laboratory strain ATCC 19606 showed that all 13 isolates were LPS deficient, moreover in each isolate the first three genes responsible for Lipid-A biosynthesis, *lpxA*, *lpxC*, and *lpxD*, presented several mutations ranging from single mutations to large deletions [316], with single amino acid substitutions appearing at residues critical for the correct function of the enzymes LpxC and LpxA. The mutated residues in LpxA are located in highly conserved regions across all bacterial LpxA enzymes. These data showed that the main driver of colistin resistance in *A. baumannii* ATCC 19606 strains is due to the presence of cells with mutations in these genes, *lpxA*, *lpxC*, and *lpxD*; leading to the total loss of LPS. Despite the increased resistance to colistin that these mutations provide, the lack of LPS in the outer membrane of these strains also makes the cells susceptible to other antibiotics, even those that normally only affect Gram-positive bacteria [317]. In addition, both virulence (due to the lack of the endotoxic LPS) and cell fitness (due to the loss of outer membrane integrity) are lowered in such strains [318].

6.1.4.2 Colistin resistance by Lipid-A modification.

The second mechanism for colistin resistance in *A. baumannii* is regulated by a two-component system (TCs), the operon *pmrCAB*. This operon has a role in the modification of LPS in the outer membrane in *A. baumannii*. The *pmrCAB* operon encodes for three enzymes:

a regulator protein PmrA (also known as BasR), a cytoplasmic membrane-bound sensor kinase PmrB (BasS) and PmrC, a putative membrane phosphoethanolamine transferase that is involved in the addition of phosphoethanolamine (PEtN) to Lipid A. Environmental stimulants, such as macrophage phagosomes, a high concentration of iron (Fe^{3+}), exposure to aluminium (Al^{3+}), or low pH, leads to activation of PmrB. PmrB activates PmrA by phosphorylation, and PmrA in turns activates regulation of the *pmrCAB* operon [95] Figure 55.

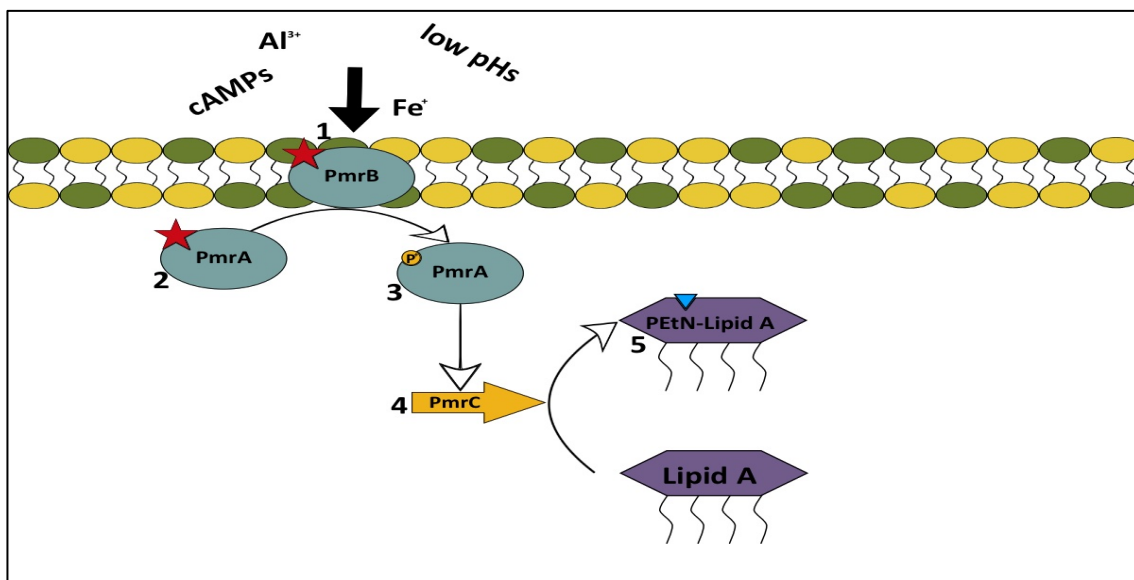


Figure 55: Lipid-A modification mediated by the *pmrCAB* operon.

As a response to the environment, the cytoplasmic membrane-bound sensor kinase PmrB (1) activates the regulator protein (2) PmrA by phosphorylation (3), the active PmrA upregulates the *pmrCAB* leading to the overexpression of the phosphoethanolamine transferase (4) PmrC which modifies (5) Lipid A by the addition of PEtN.

Genetic changes in the PmrAB regulatory system can also lead to the activation of PmrA, PmrB, and the phosphotransferase PmrC, which leads to the addition of phosphoethanolamine to Lipid A. Two distinctive genetic events are needed on the operon *pmrCAB* to generate colistin resistance in *A. baumannii*: (1) mutations in the protein PmrB and (2) the upregulated expression of both *pmrA* and *pmrB* [287]. From these mutations, the most effective in generating resistance are the ones happening on *pmrB* [318,319], the sensor

kinase. PmrB activates, by phosphorylation, PmrA, which in turns activates regulation of the *pmrCAB* operon leading to a PmrA-mediated overexpression of the phosphoethanolamine transferase PmrC. Sequence analysis of 16 polymyxin-resistant strains, derived from *A. baumannii* strain ATCC 17978 found no changes in the *pmrC* gene [320]; recent studies showed that colistin-resistant strains presented an increased transcription rate of *pmrC* with PmrC overexpression being up to 800-fold higher in colistin-resistant strains [77]. Conversely, knockout strains lacking *pmrC* were susceptible to colistin [290,320], highlighting the importance of the study of this phosphoethanolamine transferase to overcome colistin resistance.

Little is known about *A. baumannii* PmrC aside from its role in conferring colistin resistance. Sequence alignment shows that PmrC is a member of the phosphoethanolamine transferase family present in gram-negative bacteria (Figure 57). A computational analysis of PmrC from *Salmonella enterica* revealed that PmrC is a transmembrane protein formed by two main domains: (1) an N-terminal transmembrane domain most likely working as an anchor in the inner membrane and (2) a 370 amino acid periplasmic-facing catalytic domain [321]. PmrC shares a close sequence resemblance to *K. pneumoniae* EptA, another phosphoethanolamine transferase that catalysis the addition of PEtN to Lipid A; with the C-terminal periplasmic domains being 65% and the N-terminal transmembrane domains 53% identical [290]. The location of PmrC, anchored to the inner membrane [322], and the structure of its catalytic site both resemble that of MCR-1 [74], with key catalytic amino acids being relatively conserved including the catalytic Threonine (Figure 56 c and d, T29 in ScPmrC and T247 in MCR-1 catalytic domain respectively).

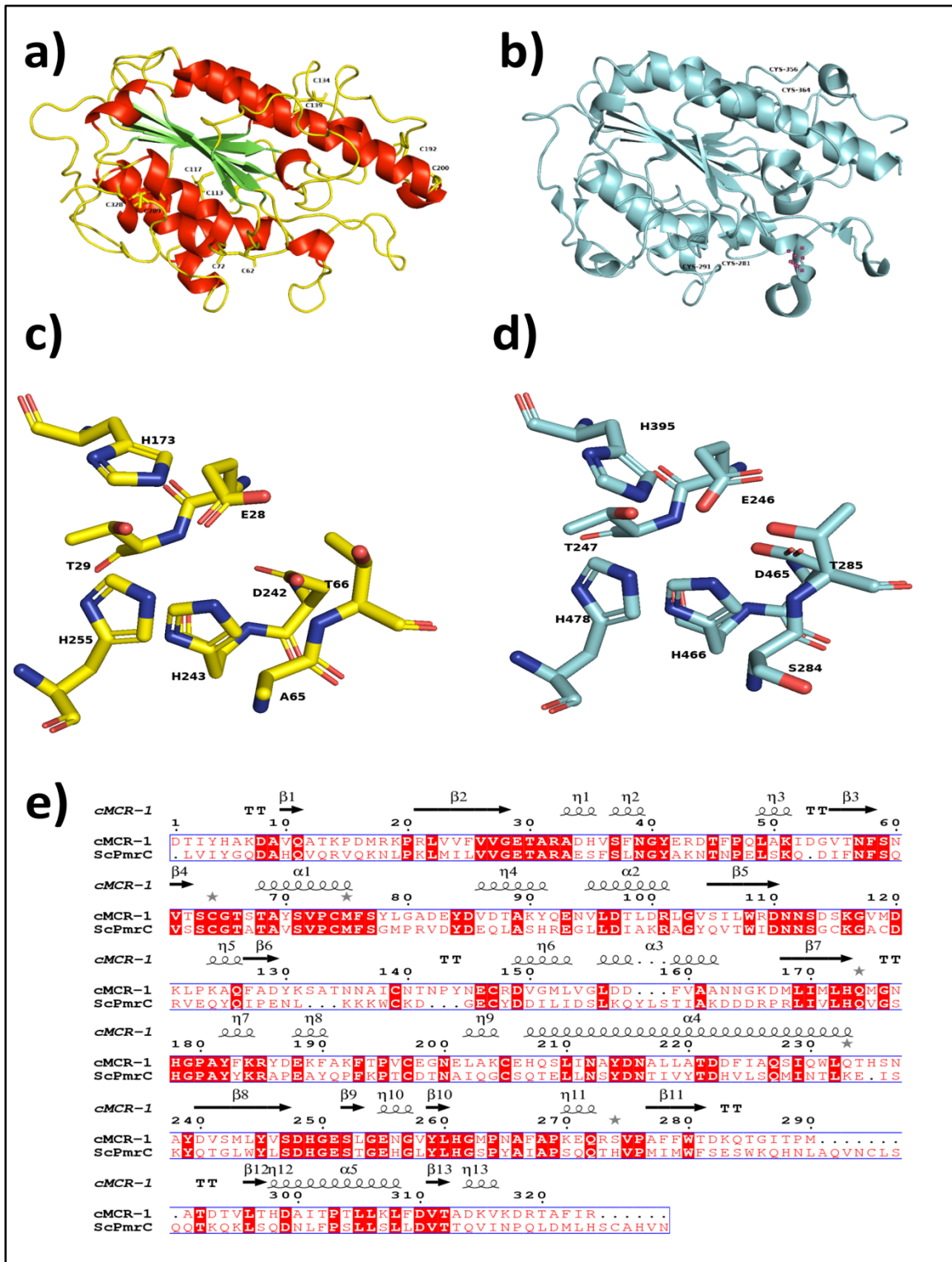


Figure 56: Model of ScPmrC and structure of cMCR-1.

A a) 3D model of ScPmrC was obtained by template modelling using the Swiss-Model server [122,123] and MCR-1 (PDB entry 5LRN) as template. Alpha-helices are coloured red, beta-sheets are green and loops and coils are presented in yellow, 10 cysteines can be seen in the surface of the protein (labelled). b) cMCR-1 structure (PDB entry 5LRM); 4 cysteines can be seen in the surface of the protein (Labelled). c) Amino acids present in the catalytic site of ScPmrC. d) Amino acids present in the catalytic site of cMCR-1 and e) Sequence alignment of ScPmrC and cMCR-1, amino acids highlighted by a red box indicates strict identity, red character indicates similarity in a group and blue box is similarity across groups

A study of 16 clinical colistin-resistant *A. baumannii* isolates found no evidence of homologs of the other regulatory and other mechanisms, such as the ArnT mechanism, involved in polymyxin resistance in other organisms (such as *K. pneumoniae*), suggesting that PmrC is an important target for the study of colistin resistance in *A. baumannii* [320]. Alongside this lack of other chromosomally encoded colistin resistance mechanisms, the plasmid-borne colistin resistance mechanism *mcr-1* and its homologues, although able to be carried in other gram-negative bacteria, is mainly found in *E. coli* and only rarely encountered in *A. baumannii* [98,322–326]. All of these data position the operon *pmrCAB* as a key element in understanding colistin resistance in *A. baumannii*, due to lipidA modification by PmrC. Little is known about *A. baumannii* PmrC however it is clearly a major contributor to colistin resistance. This chapter aims to recombinantly produce the catalytic domain of the *A. baumannii* phosphoethanolamine transferase, PmrC. This was followed by a biophysical characterization of this enzyme; looking to shed more light on its structure and its mechanism of action. Ultimately such information will aid in the identification of possible inhibitors of colistin resistance in this important Gram-negative pathogen.

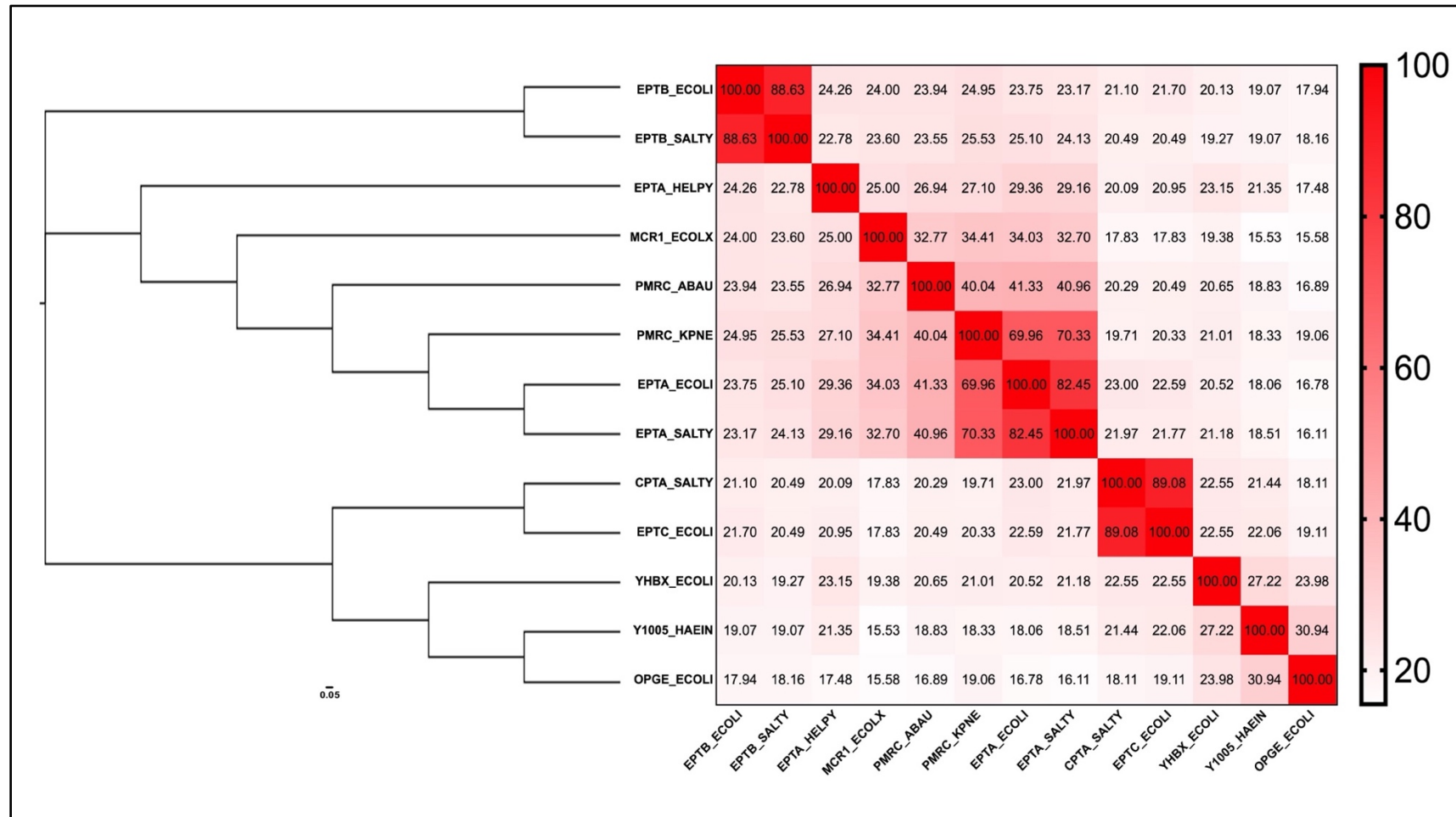


Figure 57: Phylogenetic tree and similarity matrix of several phosphoethanolamine transferases from gram-negative bacteria.

13 Phosphoethanolamine transferases from the organisms *A. baumannii* (PMRC_ABAU), *E. coli* (EPTA_ECOLI, EPTB_ECOLI, EPTC_ECOLI, YHBX_ECOLI, MCR1_ECOLI AND OPGE_ECOLI), *H. pylori* (EPTA_HELPHY), *H. influenza* (Y1005_HAEIN), *K. pneumonia* (PMRC_KPNE) and *S. typhi* (EPTB_SALTY) were analysed by multisequence alignment (accession number and sequence in Appendix T).

6.1.5 Results.

6.1.6 Cloning, small scale expression and purification of the recombinant catalytic domain of *A. baumannii* PmrC.

Two constructs of *A. baumannii* PmrC, named long construct (LcPmrC amino acids 198-538) and short construct (ScPmrC amino acids 203-538) were amplified by PCR, from a codon-optimized synthetic *pmrC* gene (Appendix R). The choice of construct length was determined by comparison of an homology model of PmrC (Figure 56) with the successfully solved 3D structure of MCR-1 catalytic domain [74]. Analysis by agarose gel electrophoresis confirmed the molecular sizes of both constructs as similar at about 1 kB. The LcPmrC construct has an expected size of 1008 bp whereas the ScPmrC construct is 993 bp (Figure 58 a) and b) respectively). Both fragments were analysed by Sanger sequencing by Eurofins (Eurofins, UK) and Source (Source BioScience, UK). After several sequencing attempts, no clear sequence was obtained for the LcPmrC construct. Further experimentation focused on subcloning the ScPmrC construct into the pOPINF T7 expression vector.

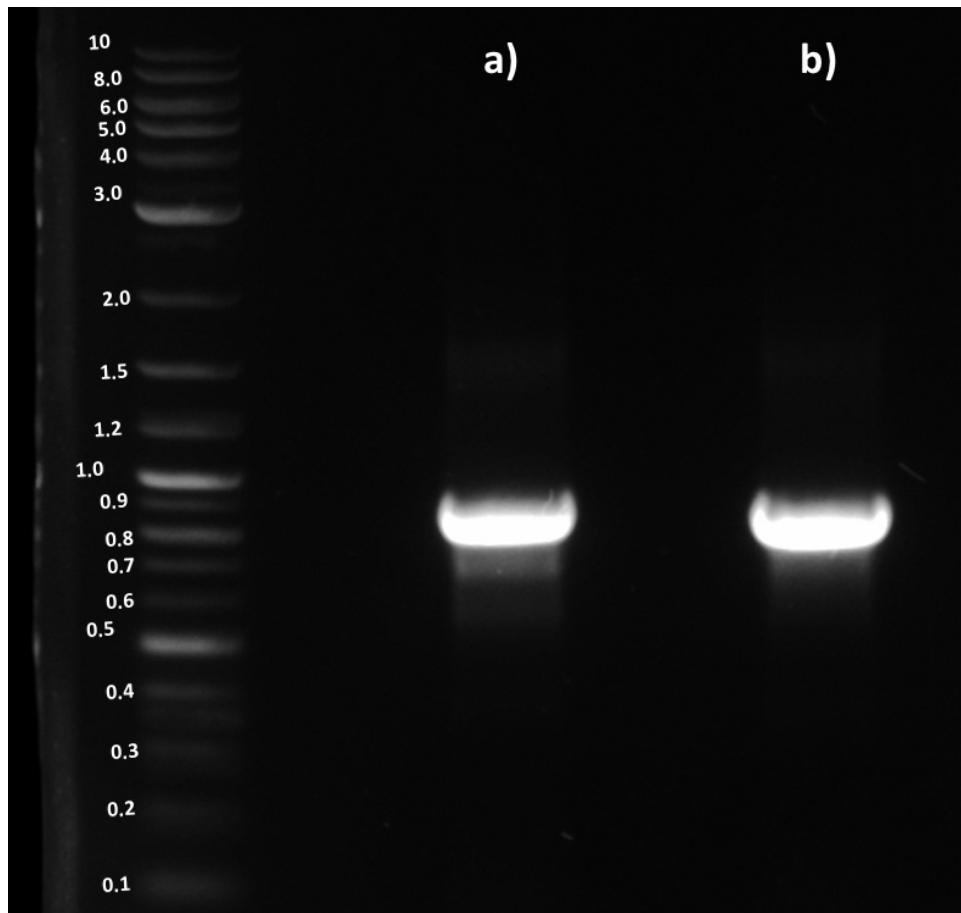


Figure 58: a) LcPmrC and b) ScPmrC PCR products.

Agarose gel showing the (a) LcPmrC and (b) ScPmrC constructs for the catalytic domain for PmrC from *A. baumannii*.

To start the small-expression trials the short construct, ScPmrC-pOPINF, was transformed into three different *E. coli* strains, BL21(DE3), BL21 Star (DE3) and Rosetta 2 (DE3). Strains BL21(DE3) and BL21(Star) were used as the strain BL21 and its derivatives are used as the standard for high levels of recombinant protein expression [210]. The strain Rosetta 2 (DE3) was also selected considering its ability to express recombinant proteins using codons rarely used by *E. coli*; rare codons, in *E. coli*, are defined as those with a use frequency lower than 1% [327], such as the arginine codon AGG that has a use frequency <0.2% in *E. coli*, whereas this same codon has a use frequency of 1% in *A. baumannii* [328].

Transformation was successful for all these strains and a small scale expression trial was carried out. SDS-PAGE of whole-cell samples showed that there was no apparent difference in ScPmrC production on induction as all the samples, including the uninduced controls, showed two prominent bands in the 30 and 40 kDa molecular weight range (Figure 60 | a, b, c and d), similar to the expected MW of ScPmrC of 37 kDa. Possible ScPmrC production was then further investigated by breaking the cells and incubating the cell lysate with Ni⁺-NTA beads. After eluting the protein from the Ni⁺-NTA beads a thick protein band was found at approximately 70 kDa molecular weight (Figure 59), inconsistent with the expected molecular mass of recombinant ScPmrC. This could represent formation of denaturation-resistant dimers during the production or recovery process, but is more likely purification of *E. coli* contaminants, possibly components of the stress response often encountered during recombinant protein production [329,330].

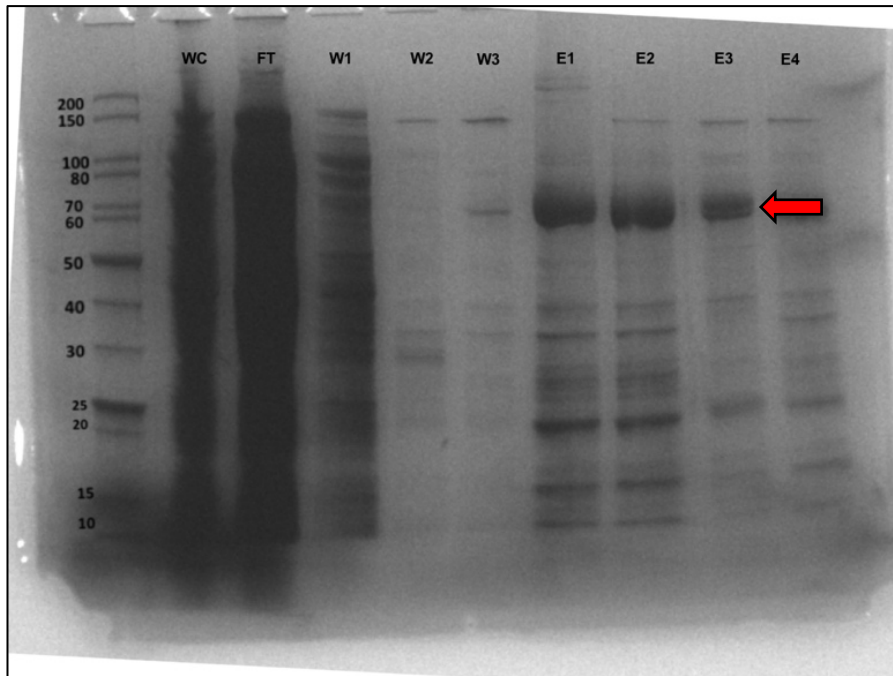


Figure 59: Attempted recovery of ScPmrC from *E. coli* BL21 (DE3).

SDS-PAGE of the small scale production and recovery of ScPmrC performed using the transformed BL21(DE3) strain. A thick band can be found at the 70 kDa mark in the lanes corresponding to the elution fractions (E1 through E4) eluting from Ni-NTA beads (indicated by the red arrow). Lane WC contains the cell lysate sample, FT contains the flow-through and lanes W1 to W3 are subsequent washing steps.

Poor protein folding and/or protein aggregation are possible explanations for the apparent low levels of ScPmrC expression and failure to recover protein in initial experiments. Aggregation can be attributed to multiple factors including incorrect disulphide bond formation, and/or incorrect protein folding [329]. Two other *E. coli* strains, ShuffleT7 Express and Artic Express were then used in efforts to overexpress ScPmrC; aiming to overcome protein aggregation. The ShuffleT7 Express strain was selected as this can promote the formation of disulphide bonds in the cellular cytoplasm by expressing a chromosomal copy of the disulphide bond isomerase DsbC correcting mis-oxidized proteins, catalysing the formation of correct disulphide bonds and correct tertiary structure formation [331,332]. The Arctic Express strain reduces the formation of inclusion bodies during recombinant protein overexpression as this expresses two exogenous chaperones, Cpn10 and Cpn60, from *Oleispira antarctica*, a

psychrophilic bacterium. These chaperones assist in the correct folding of recombinant proteins and avoid the accumulation of inclusion bodies [333,334].

SDS-PAGE of expression trials indicated that on induction strain ShuffleT7 Express produced a protein band just below 40 kDa, the MW of ScPmrC is estimated to be 37 kDa. Mass spectrometry analysis of this band (Figure 60 II red arrow) confirmed the protein to be ScPmrC, with a score of 941.31 compared to the second highest protein which scored an order of magnitude lower (98.22) and was identified as *E. coli* BL21 elongation factor G with a molecular weight of 77.5 kDa (Table 11).

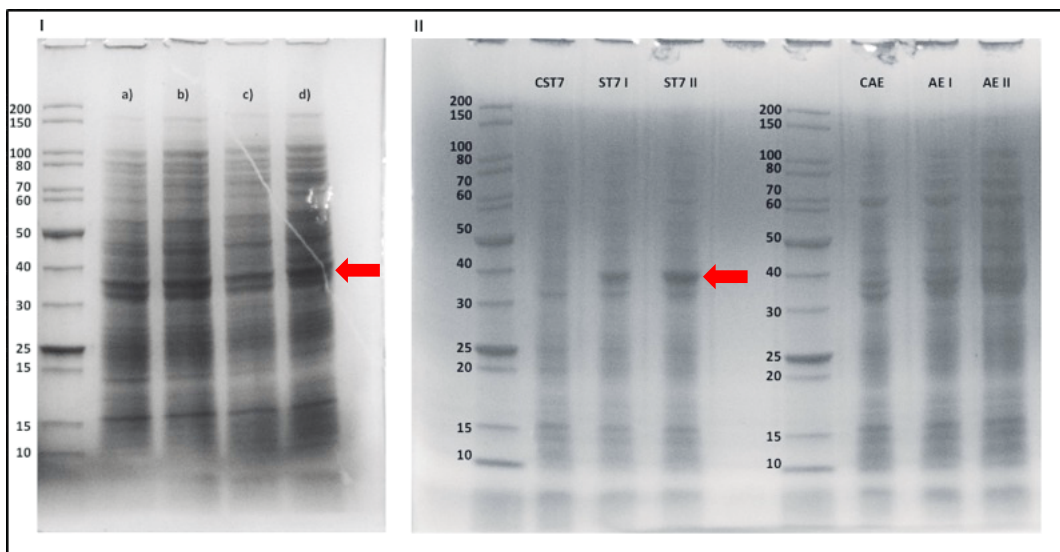


Figure 60: Expression trials of recombinant ScPmrC.

Comparison of the production of catalytic domain ScPmrC in recombinant E. coli strains. Gels show whole-cell lysates samples of : (I) a) Uninduced BL21(DE3) b) Induced BL21(DE3) c) Induced BL21 Star (DE3) and c) Induced Rosetta 2 (DE3). (II) Production of ScPmrC using E. coli strains ShuffleT7: uninduced control (CST7), induced sample and replicate (ST7 I and ST7 II respectively); and E. coli ArcticExpress (DE3) uninduced control (CAE) and induced sample and replicate (AE I and AE II respectively). Bands consistent with possible ScPmrC production are marked with red arrows.

Table 11: Mass fingerprinting analysis and sequence search against the Uniprot *E. coli* BL21-DE3 database and the ScPmrC sequence.

Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
ScPmrC	ScPmrC ScPmrC	941.31	44.58	26	332	37.3
A0A140N7C7	Elongation factor G OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=fusA PE=3 SV=1 - [A0A140N7C7_ECOBD]	98.22	45.03	24	704	77.5
A0A140ND68	Aminotransferase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2667 PE=3 SV=1 - [A0A140ND68_ECOBD]	93.73	50.00	23	396	43.5
A0A140NDI4	Phosphoserine aminotransferase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=serC PE=3 SV=1 - [A0A140NDI4_ECOBD]	87.25	52.21	17	362	39.8
A0A140NDL0	Pyruvate dehydrogenase E1 component OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_3505 PE=4 SV=1 - [A0A140NDL0_ECOBD]	78.46	24.58	20	887	99.6
A0A140NHS0	ATP synthase subunit beta OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=atpD PE=3 SV=1 - [A0A140NHS0_ECOBD]	77.13	42.83	16	460	50.3
A0A140NH65	60 kDa chaperonin OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=groL PE=3 SV=1 - [A0A140NH65_ECOBD]	75.28	37.23	20	548	57.3
A0A140N2U0	DNA-directed RNA polymerase subunit alpha OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rpoA PE=3 SV=1 - [A0A140N2U0_ECOBD]	74.17	46.20	14	329	36.5
A0A140N9F5	Outer membrane protein assembly factor BamB OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=bamB PE=3 SV=1 - [A0A140N9F5_ECOBD]	67.76	40.31	13	392	41.8

6.1.7 Large Scale ScPmrC production.

Large scale expression of ScPmrC was carried out in batches of 4 litre cultures in 2L shake flasks in a shaking incubator using *E. coli* strain ShuffleT7 Express harbouring plasmid

ScPmrC-pOPINF and the optimised semi defined medium from previous chapters. This large-scale growth yielded 6.32 grams of wet bacteria per litre of culture (on average). The recovered bacteria were broken down and cell lysate was centrifuged to remove cell debris. The clear cell lysate was incubated on Ni⁺-NTA beads and soluble *ScPmrC* was recovered and purified away from *E. coli* protein contaminants, by eluting using 250 mM imidazole. Fractions were analysed by SDS-PAGE (Figure 61, protein bands corresponding to the expected molecular weight of *ScPmrC* are marked with red arrows). SDS-PAGE analysis revealed a strong recovery of *ScPmrC* as the presence of a broad band around 40 kDa, which was confirmed to be *ScPmrC* by mass spectrometry fingerprint analysis, with a high score of 1394.96 (Appendix U). Although *ScPmrC* composes more than 70% of the eluted sample, other contaminants can be seen in the elution fraction, with two bands of molecular masses around 70 kDa and 25 kDa respectively, indicating the need for further purification. Mass fingerprint analysis showed a mixture of phosphorylation and un-phosphorylation in the active site of the enzyme (Appendix V).

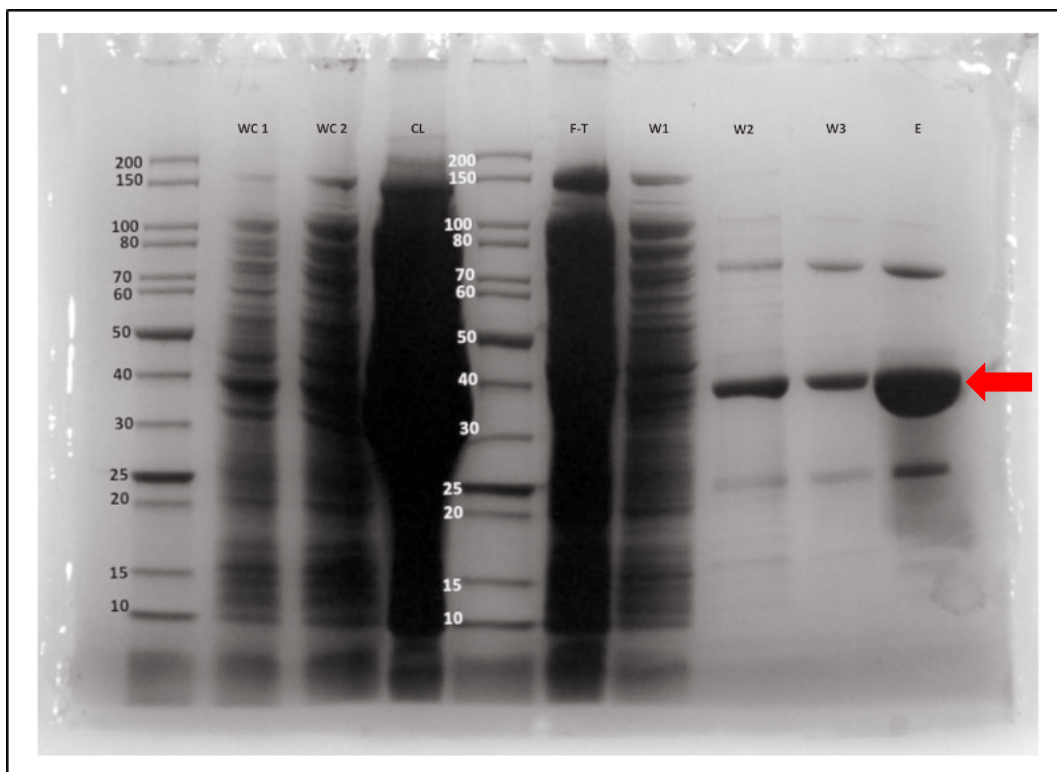


Figure 61: Large scale production and recovery of ScPmrC.

SDS-PAGE gel of the large-scale production of ScPmrC in semi defined 2xYT media. The protein production of whole-cell samples (WC1 and WC2) and cell lysate (CL) can be seen on the left half of the gel. The recovery process including the flow-through (F-T), washing steps (W1, W2 and W3 respectively) and elution (E) from Ni-NTA beads can be seen on the right half of the gel. The band corresponding to ScPmrC is marked with a red arrow located between the 30 and 40 kDa marker bands.

The elution fraction was concentrated and a buffer exchange was performed to reduce imidazole concentration. Two fractions of 3 mL volume were recovered with concentrations of 12.2 and 12.1 mg/mL, respectively, corresponding to a total of 72.8 mg of protein. This production process yielded approximately 11.5 mg of ScPmrC per litre of culture. The first 3 mL fraction was further purified using a size exclusion chromatography (SEC) column; the second 3mL fraction was treated with 3C protease to remove the 6His-Tag and then purified by SEC on a Superdex 75 column (Figure 62 c and d). The removal of the His-tag is considered a normal procedure when attempting protein crystallization for structural studies as flexible regions often hinder the crystallisation process. Both fractions presented evidence of protein aggregation at different stages of the SEC purification. The non-3C treated sample

(12.2 mg/mL) yielded a turbid solution for the recovered fractions under the peak. The chromatogram showed a high peak after 52 mLs (52 minutes of run) elution volume and a shoulder around 70 mL (70 minutes of run) of elution volume (Figure 62 a). SDS-PAGE analysis of the collected fractions, under the peak and its shoulder, revealed both to contain a protein of approximate mass 40 kDa (Figure 62 b), consistent with the calculated molecular weight of His-tagged ScPmrC of 39.4 kDa. The second sample (12.1 mg/mL) showed signs of precipitation during incubation with 3C protease. The sample was re-incubated with Ni²⁺ beads and 10 mM imidazole to remove the (6-His-tagged) 3C and the recovered, tag-less ScPmrC was then purified by SEC on the S75 column. The chromatogram showed two well-defined peaks, one eluting at 52 mLs (52 min running time), and the second one at 72 mLs (72 mins running time, Figure 62 c). Notably, the absorbance of each of the two eluted peaks was an order of magnitude lower than for the non 3C treated sample, this can be attributed to the protein precipitating during the removal of the His-tag. Samples from both peaks were taken and analysed by SDS-PAGE (Figure 62 d); both peaks presented a single band at ~40 kDa. Since aggregation was suspected, a 1 mg/mL solution of blue dextran (molecular mass of ~2,000 kDa) was used to calibrate the void volume of the SEC column. Blue dextran eluted at 48 minutes (Appendix W). This provides strong evidence that ScPmrC aggregation is happening, as in both instances the first elution peak is very close to the elution peak of blue dextran.

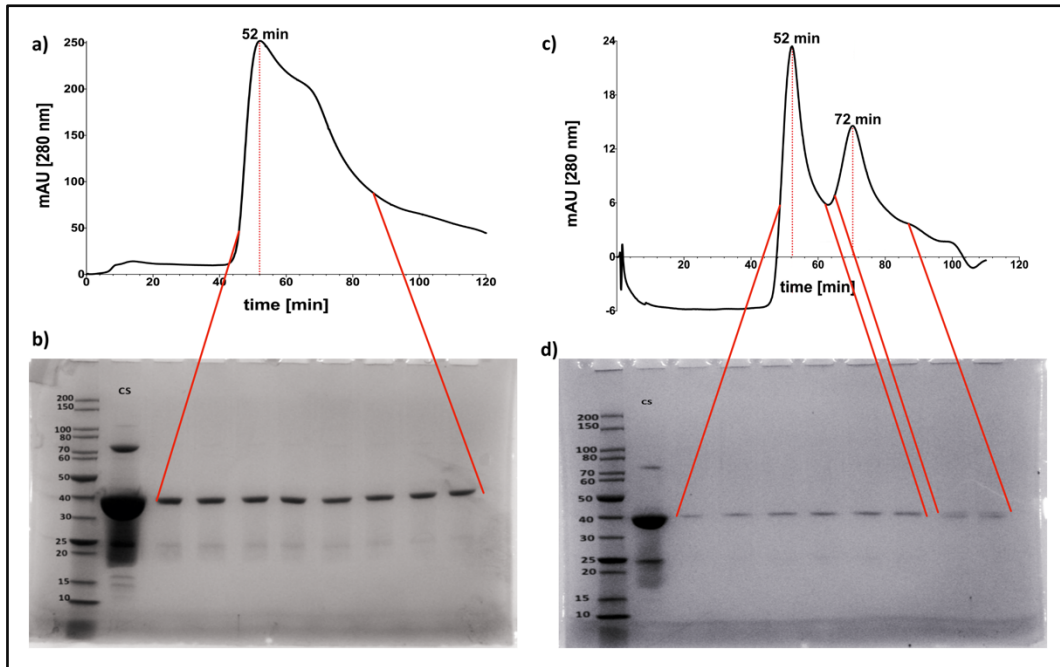


Figure 62: Size Exclusion Chromatography of ScPmrC.

a) Chromatogram showing the elution of 6-HisTagged ScPmrC from the S75 column with b) SDS-PAGE of the concentrated sample loaded into the SEC column (CS) and the recovered fractions from the elution peak (Fractions inside the red lines). c) Chromatogram showing the elution peaks and times of ScPmrC after 3C-protease treatment to remove the 6-HisTag, d) SDS-PAGE of the concentrated sample loaded into the SEC column (CS) and the recovered fraction obtained from both elution peaks (Fractions inside the red lines).

It was hypothesized that temperature was playing a major role in ScPmrC aggregation since this phenomenon was most prevalent during the SEC purification, which was the only step in the whole process that took place at room temperature; the rest of the procedures all happened at 4°C or on ice. Since removing the 6 histidine tag also proved to cause protein aggregation a second batch of 6-His-tag-ScPmrC was produced (Figure 63), 5 mL at a protein concentration of 5.9 mg/mL were obtained and used for further experimentation. This protein did not undergo further purification steps through SEC.

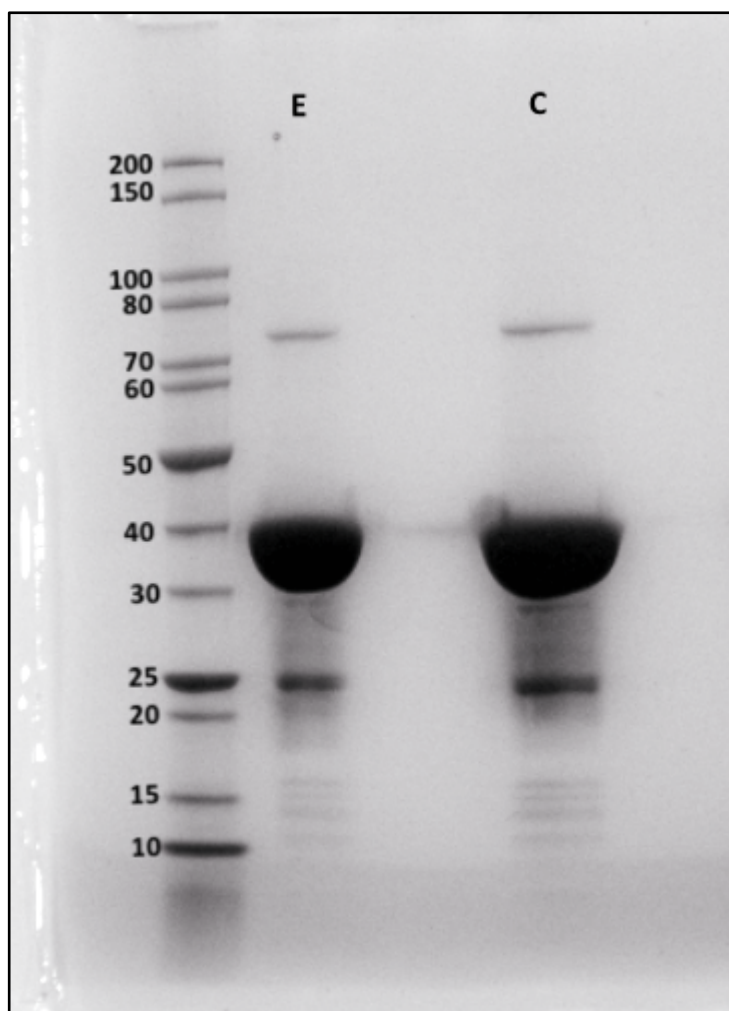


Figure 63: SDS-PAGE showing the elution (E) and concentration (C) steps of a 6-His-tagged ScPmrC.

The sample was not subjected to any further polishing step due to previously identified protein precipitation during SEC purification.

6.1.8 Circular dichroism spectroscopy of ScPmrC.

In the absence of structural information from X-ray diffraction or nuclear magnetic resonance (NMR) experiments for *A. baumannii* PmrC, circular dichroism (CD) spectroscopy was used to obtain information about the secondary structure content of the protein, confirm that the recombinant material was correctly folded and analyse the thermal stability of ScPmrC, the recombinant catalytic domain of PmrC. Analysis of CD spectra can readily provide an estimation of the fraction of the residues in the protein structure which are involved in α -helices

or β -sheets or are disordered [138]. CD was used to assess the conformational properties of the purified ScPmrC. The results confirm that ScPmrC contains secondary structure after purification, however, it is temperature sensitive and exposure to temperatures above room temperature (25°C) leads to protein unfolding and possibly aggregation.

The far -UV CD spectrum of recombinant ScPmrC is shown in Figure 64. The spectrum, from 190 to 250 nm, presents similar properties to the CD spectra of recombinant MCR-1, presented in previous chapters (Figure 38). The spectrum acquired at 25°C contains two minima of similar magnitudes at 213 and 218 nm (a) and becomes much more positive at shorter wavelengths approaching 200 nm before the detector saturates at approximately this value. (Figure 64a)) CD spectra from richly α -helical proteins present two well defined negative bands at 208 nm and 222 nm whereas proteins rich in antiparallel β -sheets have a single negative band at 218 nm and relatively positive bands between 195nm and 200 nm [218,237,238,335]. Overall the CD spectrum of recombinant purified ScPmrC is consistent with a protein with substantial α -helical content; a secondary structure content analysis was performed using the K2D3 server (<http://cbdm-01.zdv.uni-mainz.de/~andrade/k2d3/>) [239] (Figure 64 d). The K2D3 server predicts percentages of alpha helix and beta-strand content based on input CD spectra. The predictions showed an α -helical content of 40.31% and a β -sheets content of 16.03%. To corroborate this, the secondary structure content of a 3D model of ScPmrC generated using the Swiss Model server (Figure 56 a) was analysed using Pymol; this analysis showed an α -helical content of 33.23% and a β -sheet content of 12.50%, in reasonable agreement with the results of the fitting process. The discrepancy can be attributed to the fact that the comparison was with a 3D model obtained using other phosphoethanolamine transferases as a template and not with an experimental ScPmrC protein structure as well as the limitations of the technique, as readings under 200 nm wavelength become noisy and information gets lost. It is worth mentioning that K2D3 server performs the secondary structure analysis by comparing the experimentally obtained spectra against a the predicted spectra based upon structures in the PDB, and not actual experimental measurements which can lead to inaccuracies during secondary structure content analysis. Experimental measurements are also highly dependent on protein concentration; any inaccuracies in the determination of protein concentration will also affect the magnitude of your calculated signal and hence accuracy of prediction.

The thermal stability of the ScPmrC was also analysed by investigating the temperature-dependence of the CD spectrum. Spectra acquired at increasing temperatures are shown in (Figure 64 b) and a melting temperature curve showing can be found in Figure 64 c). The CD results showed clear changes to the secondary structure content of ScPmrC as the temperature increases, with spectra showing almost complete loss of signal at a temperature of 55° C consistent with thermal denaturation of the protein. [238]. An attempt at refolding the protein by lowering the temperature gradually, from 95 °C to 5 °C, was carried out however, no signs of refolding were found as ScPmrC aggregated and precipitated inside the cuvette. The fitting of the melting curve yielded a midpoint for loss of signal/unfolding temperature (T_m) of 42 °C. This is consistent with the results from size exclusion chromatography above, which indicated a propensity for the protein to aggregate at room temperature. Notably, the T_m obtained for ScPmrC is about 20 °C lower than that of full length MCR-1 (T_m of 60.34 °C) a reduction in stability which may be attributed to the lack of the transmembrane domain.

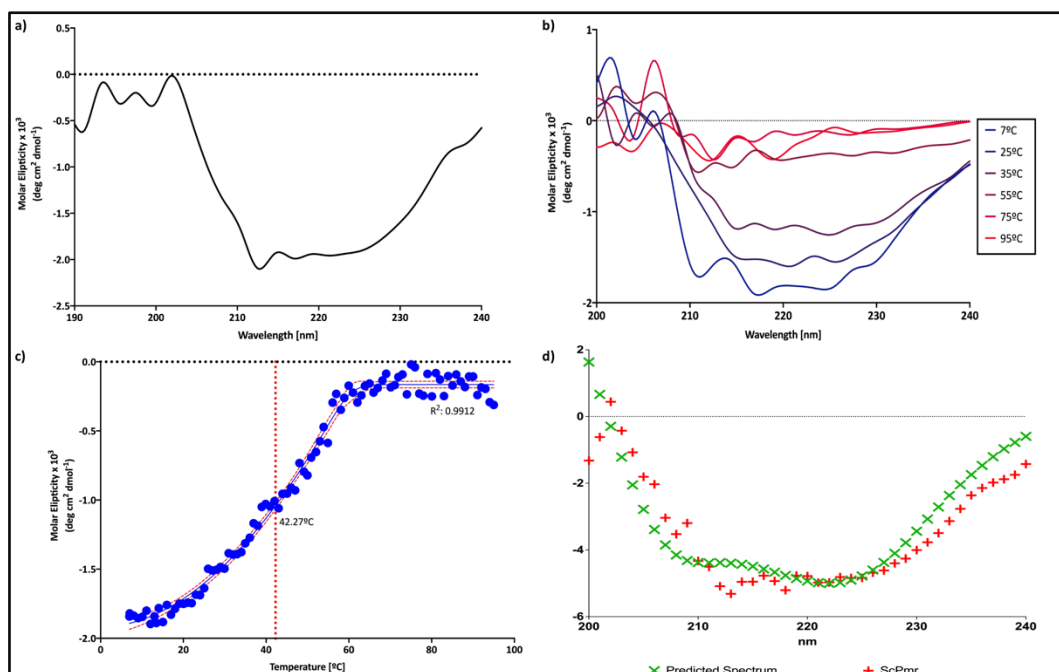


Figure 64: CD data from ScPmrC.

a) CD spectrum of the catalytic domain of PmrC, ScPmrC at room temperature. Data were obtained using 50 mM HEPES pH 7.5, 10 mM NaCl buffer with a protein concentration of $\sim 10 \mu\text{M}$. b) CD spectra of ScPmrC at varying temperatures. Changes in the spectrum can be seen at temperatures above room temperature (25°C). c) the Melting temperature curve of ScPmrC; the melting temperature as determined by CD was 42.2°C and d) Secondary structure content prediction, the estimated α -helical content was of 40.31% and β -sheets 16.03%.

6.1.9 ScPmrC Activity: Cleavage of PNP-PEtN.

After confirming that recombinant purified ScPmrC contains substantial secondary structure, an activity assay was carried out to test the ability of ScPmrC to turn over the chromogenic reporter substrate pNP-PEtN using the conditions established for cMCR-2 in previous chapters (Development of a biochemical assay.). Different concentrations of ScPmrC, ranging from 10 μM to 100 μM , were tested against pNP-PEtN at a concentration of 500 μM . The release of 4-nitrophenol (pNP) was measured at a wavelength of 405 nm (abs_{405}). No significant difference was found between the two lowest enzyme concentrations tested, 10 and 20 μM ; however, as the enzyme concentration increases, particularly above 50 μM (Figure 65), the release of pNP increases, as seen by the increase in abs_{405} . This demonstrates that ScPmrC is actively cleaving the phosphoethanolamine moiety from pNP-PEtN, indicating that

ScPmrC is able to catalyse the first stage of the phosphoethanolamine transferase reaction *in vitro*. [336]

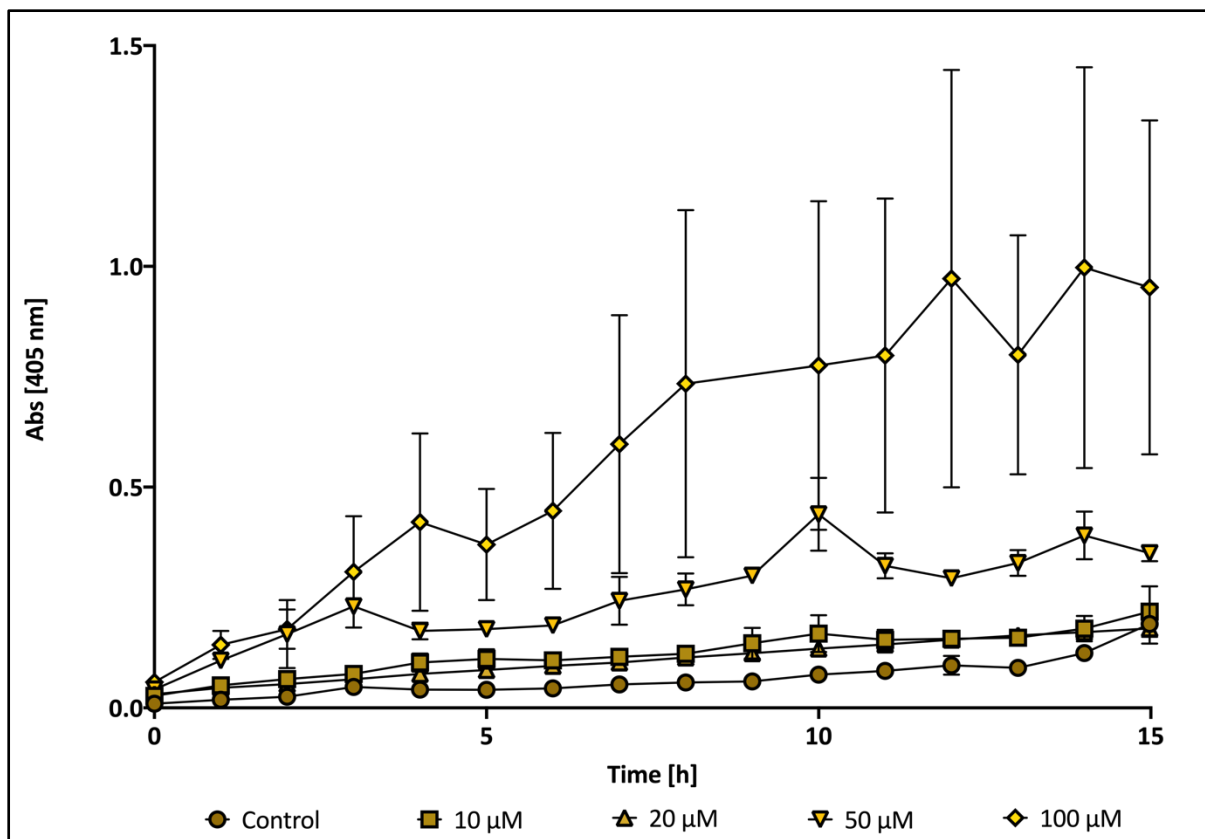


Figure 65: Cleavage of pNP-PEtN by the recombinant catalytic domain of ScPmrC

500 μM of pNP-PEtN was incubated with varying concentrations of purified recombinant ScPmrC. A no-enzyme sample was run as a control. Dashed lines are straight line fits to data. Experiment was performed in duplicate.

6.1.10 Estimation of the Kinetic parameters of ScPmrC.

An activity assay, at varying concentrations of pNP-PEtN, was performed in an effort to determine the kinetic parameters, K_m , and V_{max} , for the breakdown of pNP-PEtN by ScPmrC. To obtain the initial reaction rates of the reaction a fixed concentration of 50 μM ScPmrC was incubated with increasing concentrations of pNP-PEtN (Figure 66). The reaction initial rates were obtained, in a similar fashion as described previously (above 2.2.15), from the increase in Abs_{405} over the first 4 hours of the reaction. The amount of released pNP was estimated using a standard curve (Figure 51 c) and Equation 8.

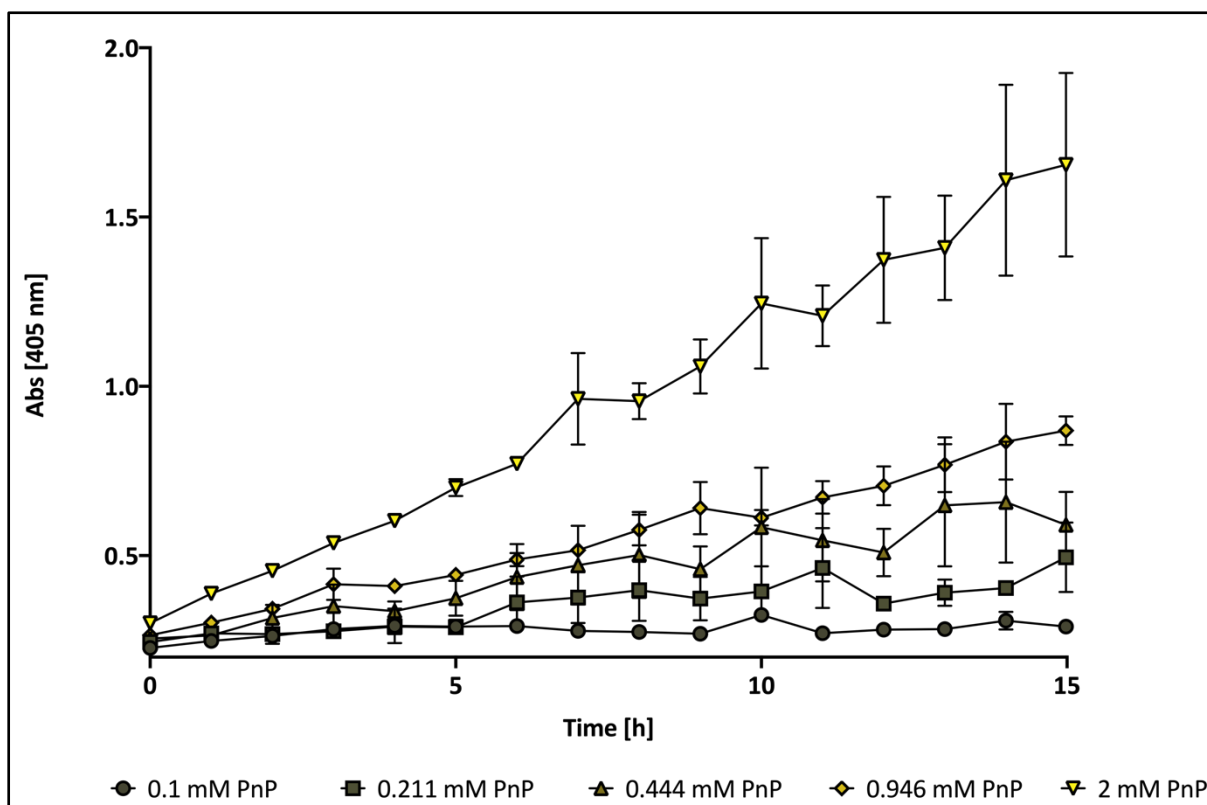


Figure 66: Reaction of pNP-PEtN at varying concentrations with the ScPmrC recombinant catalytic domain (50 μM).

ScPmrC (50 μM) was incubated with increasing concentrations of pNP-PEtN (starting at 0.1 mM to a max of 2 mM). Dashed lines are straight line fits to data. No saturation was observed during the 15 hours of incubation. The standard error of mean has been added to all data points. This was done in duplicates

The kinetic parameters were estimated from the experimental results by fitting the concentration dependence of the obtained rates to the Michaelis-Menten equation yielding estimates for the kinetic parameters K_M of 1.7 mM and the maximum reaction velocity V_{max} , of 1881 nMmin⁻¹, corresponding to a turnover number, k_{cat} , of 0.000627s⁻¹.

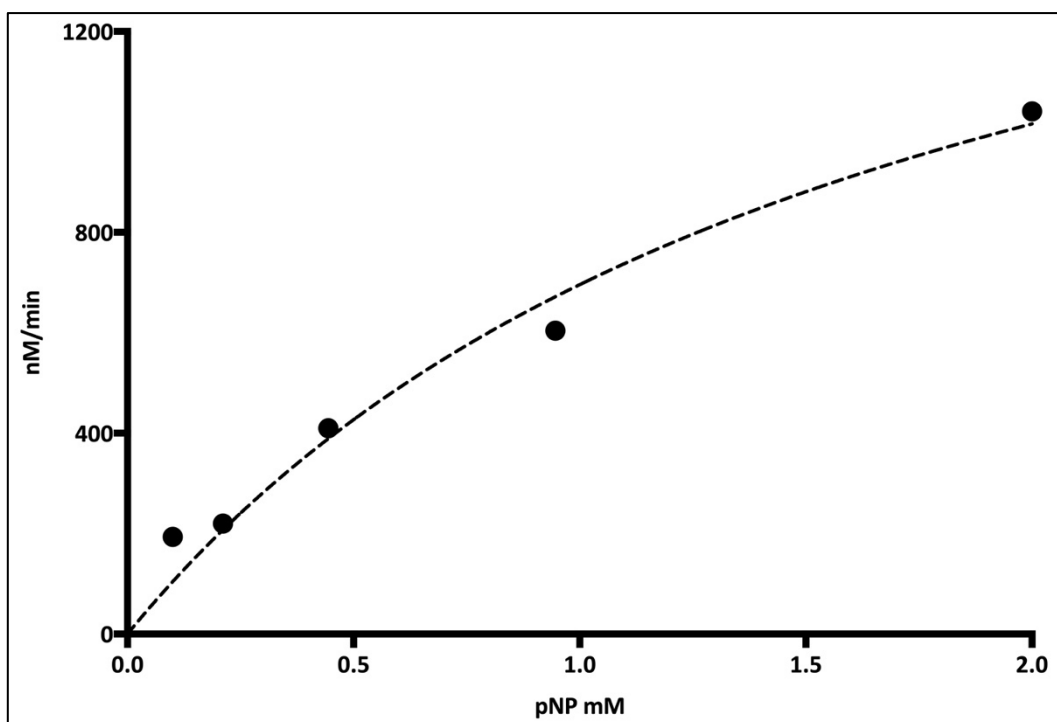


Figure 67: Estimation of the kinetic parameters for ScPmrC activity by fitting to a Michaelis-Menten model.

Michaelis-Menten fit of the reaction rate as a function of the substrate, pNP-PEtN, concentration in mM. The dotted line represents the line of best fit to Equation 6 giving kinetic parameters of ScPmrC as K_m of 1.7 mM and a V_{max} 1881 $nM \cdot min^{-1}$. $R^2 = 0.97$.

6.1.11 ScPmrC Crystallisation.

The recombinant 6-His-tagged ScPmrC maintained its structural and functional stability, as demonstrated by CD analysis, and displayed activity in *in vitro* assays with pNP-PEtN, and was then used for protein crystallization trials. The 6-His-tagged ScPmrC, purified by Ni²⁺-NTA chromatography, was subjected to five different sparse-matrix protein crystallization screens: Morpheus, JCSG-plus, Structure Screen 1 + 2 HT and PACT Premier from Molecular Dimensions (Molecular Dimensions, Sheffield, UK) and a Top96 Crystallization Screen (Anatrace, USA). The screen JCSG-plus was selected due to its design for rapid and efficient initial screening for protein crystallization conditions that can be coupled with the more systematic screen selection grid of PACT Premier [337]; the screens Structure Screen 1 + 2 HT are more traditional screens, developed in early 1990, found to be successful in the crystallization of macromolecules [338]. The Morpheus screening was selected due to its

incorporation of low molecular weight and buffer systems, to co-crystallize with proteins, observed more than 30,000 protein structures in the Protein Data Bank (PDB) [339]; this same rationale was used selecting the Top96 Crystallization Screen as it contains the most commonly occurring conditions reported in the Protein Data Bank for protein crystallization [340].

Only one out of the five screening trials, JCSG-plus, resulted in visible crystal-like material, with the other four presenting mostly heavy precipitates across most wells of the plate. Possible protein crystal hits were found in four conditions from the JCSG-plus screen trial, all of these contained citrate as salt or buffer and three of them contained PEG as the precipitant, whereas the last one contained ethanol. The three wells that contained PEG as the precipitant showed “needle” like objects after 5 days of incubation (Figure 68 a) at 4 °C whereas the last well, containing ethanol as the precipitant, presented more three dimensional objects that can be observed after 13 days of incubation and became more clear after 21 days of incubation (Figure 68 b).

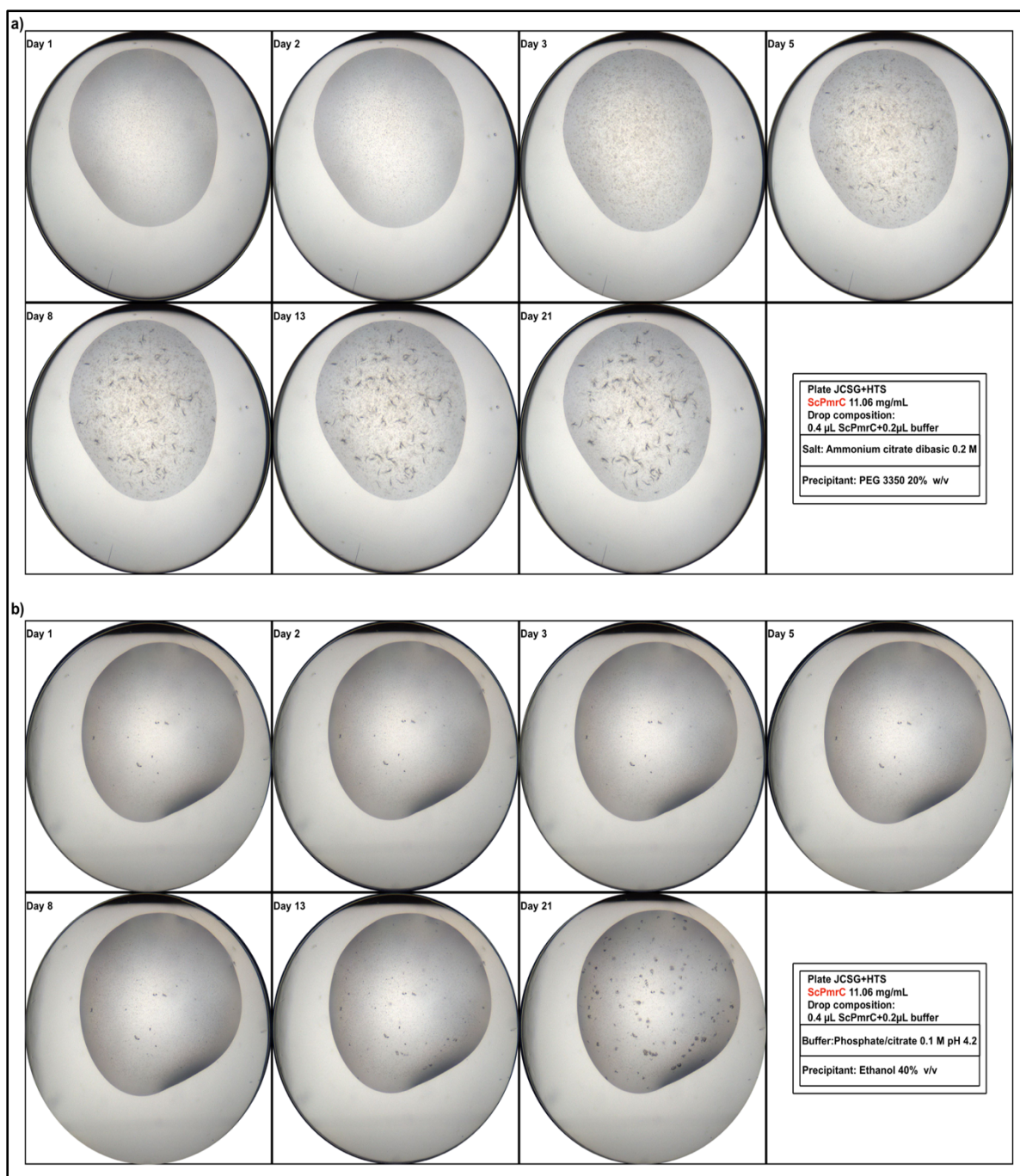


Figure 68: ScPmrC protein crystallisation trials.

Preliminary crystal screening hits for the 6-His-tag-ScPmrC enzyme obtained by vapour diffusion using the sitting drop method from a) 0.2 M Ammonium citrate dibasic and 20% (w/v) PEG 3350 and b) 0.1 M Phosphate/citrate pH 4.2 and 40% (v/v) Ethanol. The protein drop consisted of 11 mg/mL ScPmrC in 50 mM HEPES pH 7.5, 150 mM NaCl 0.1 mM ZnCl₂ and 1 mM TCEP. Crystal-like objects can be observed as early as 5 days into incubation.

6.1.12 Optimisation of ScPmrC crystallization conditions.

To optimize the production of 6-His-tagged ScPmrC crystals 2 sets of 24 well plates were used to create a screening grid using the conditions obtained from previous experiments (Figure 68 b). The combination of 0.1 M Phosphate/citrate buffer pH 4.2 and 40% (v/v) ethanol (precipitant) was selected as the best candidate for crystal formation optimization, since other conditions yielded “needle like” crystals. Two grids, consisting of 4 different pHs (3.8, 4.2, 4.6 and 5.0) with 6 different precipitant concentrations (ethanol at 25%, 30%, 35%, 40%, 45% and 50% v/v%), were generated. Experiments were carried out using the sitting drop method. For each grid a different droplet conformation was used, for the first grid a drops were set up with a protein : buffer ratio of 1:1 and the second grid had a ratio of 2:1; based on previously reported condition for the crystallization of the catalytic domain of the phosphoethanolamine transferase MCR-1 [74]. The optimization of hit conditions with varying ethanol concentrations and buffer resulted in apparent micro crystals of 6-His-tagged ScPmrC. The crystals were found after 5 days of incubation with the buffer precipitant conditions of pH 3.8 and 30% (v/v) EtOH in drops containing 1 μ L of purified 6-His-tagged ScPmrC (11 mg/mL) and 1 μ L of buffer/precipitant solution (Figure 69).

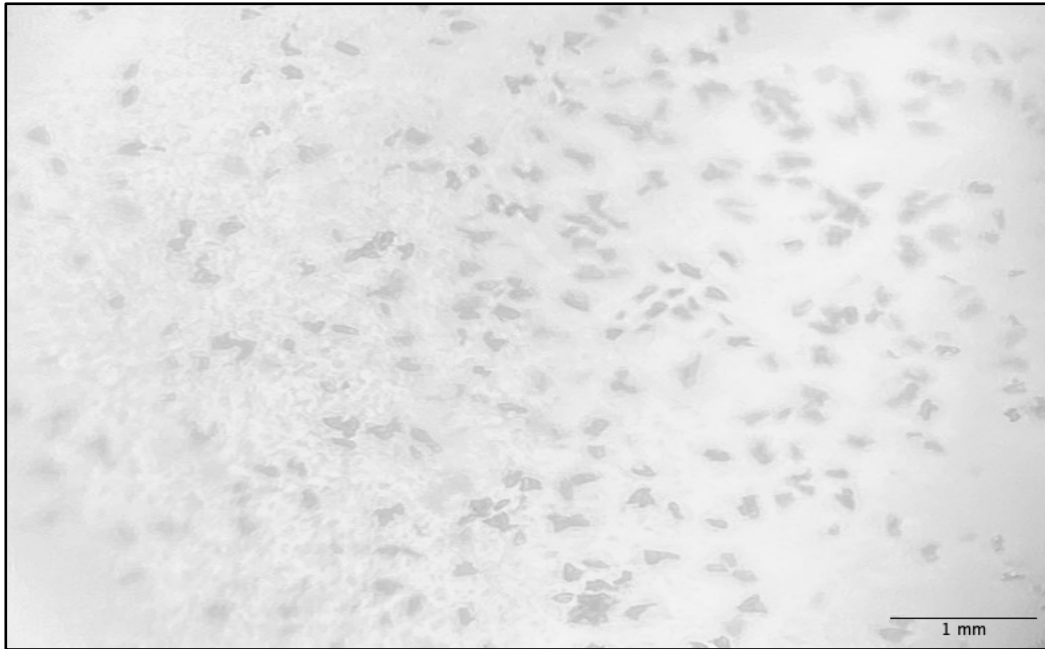


Figure 69: ScPmrC crystals.

ScPmrC crystals were found after 5 days incubation at 4° C under the 0.1 M Phosphate/citrate pH 3.8 and 30% (v/v) Ethanol (precipitant) conditions. The protein drop consisted of 1 μ L of purified ScPmrC (11 mg/mL) and 1 μ L of buffer/precipitant solution.

None of the other 47 conditions tested yielded crystals, with most of them presenting various degrees of precipitation. The crystals obtained were relatively small and not big enough to try using them for X-ray analysis, thus making it impossible to acquire a 3D structure of the protein for the time being.

6.2 Conclusions.

This chapter describes the expression, purification and biophysical characterization of ScPmrC, the soluble periplasmic C-terminal domain of PmrC from *A. baumannii*. The optimization of the expression of ScPmrC resulted in high yields of pure ScPmrC, but these could only be obtained from the Shuffle T7 Express strain, suggesting that correct disulphide bond formation is necessary for recombinant expression. During the purification process, it was found that ScPmrC is very susceptible to aggregation, especially when the temperature increases. Temperature stability analysis using CD spectroscopy, and the propensity of the purified protein to precipitate in crystallisation trials, corroborate this. Despite being prone to aggregation, recombinant 6-His-tag-ScPmrC retained its secondary structure, as evidenced by CD, and showed phosphoethanolamine transferase activity as the chromogenic assay developed in previous chapters revealed that ScPmrC has the ability to cleave PEtN from pNP-PEtN. Preliminary protein crystals of the 6-His-tagged ScPmrC were obtained using a combination of 0.1 M Phosphate/Citrate buffer and ethanol (as precipitant), with the best buffer pH/precipitant combination found to be 0.1 M Phosphate/Citrate at 3.8 pH and 30% (v/v) EtOH.

Even though there have been multiple reports of colistin resistance in *A. baumannii*, there is relatively little research on the relevant mechanisms. Colistin resistance in Gram-negative bacteria usually involves modification to lipid A, the target of colistin and other CAMPs. Currently, there are two mechanisms of modification of Lipid A that lead to colistin resistance in *A. baumannii*; the first involving the loss of LPS in the bacterial outer membrane, which removes the target of colistin but also leads to susceptibility to other antibiotics [341]. The second mechanism involves the decoration of lipid A with a positively charged motif, reducing negative charge and preventing its targeting by colistin. In Gram-negative bacteria this modification of Lipid A can be done by the addition of positively charged groups such as 4-amino-4-deoxy-L-arabinose (LAra4N) and phosphoethanolamine (PEtN) (Figure 4); however *A. baumannii* lacks the genetic equipment to perform LAra4N synthesis and Lipid A

attachment [289,342,343] thus leaving the addition of PEtN as the only modification leading to colistin resistance.

The colistin resistance mechanism causing this modification is mediated by the two-component regulatory system (TCS) PmrA and PmrB, located in the *pmrCAB* operon, which includes *pmrC*, gene that codes for the Lipid A-PEtN transferase enzyme PmrC [320,344]. Although *A. baumannii* can express additional *pmrC* homologs *eptA-1* and *eptA-2*, their roles in colistin resistance are unclear. The recently discovered plasmid borne colistin resistance *mcr-1* has been reported to be able to move through Gram-negative pathogens, however it is most common in other Gram-negative bacteria such as *E coli* [98,345] and only occasionally found in *A. baumannii*. These conditions make PmrC an important colistin resistance determinant in *A. baumannii*, but little is known about this specific colistin resistance mechanism. To our knowledge, this is the first study aiming to investigate the properties of this phosphoethanolamine transferase from *A. baumannii*. Much work has been done on other phosphoethanolamine transferases such as MCR-1 [74,98,110,120,199,201,245,246,346–350], due to its recent discovery and its ability to spread to other Gram-negative bacteria; but other such enzymes are little studied. However, the threat of carbapenem resistant *A. baumannii* strains, and the limited number of antibiotics available to deal with this threat, of which colistin is one, highlights the importance of studying PmrC as a colistin resistance mechanism and the importance of including it in future inhibitor discovery studies.

High expression levels of ScPmrC were only achievable using the *E. coli* strain ShuffleT7, a strain that contains a copy of the disulphide bond isomerase DsbC that is expressed in the cytoplasm to help with recovery of proteins from mis-oxidation [351]. This strain has been used for the successful production of several proteins containing disulphide bonds in soluble forms [332]. The sequence and 3D homology model (Figure 56) shows that ScPmrC contains 10 cysteine residues, compared to the 6 cysteines (4 close to the surface) in the periplasmic

domain of MCR-1 (PDB entry 5LRM). The model suggests that all of these cysteine residues are located in pairs close to one another in the folded protein structure, representing potential contact points for the formation of disulphide bonds. which would account for the high levels of precipitation presented in this protein. Failure to form disulphide bonds correctly during ScPmrC folding in the cytoplasm, and/or loss of disulphides near the protein surface, would account for the poor expression and high levels of precipitation presented by this protein [332,351]. By employing the ShuffleT7 strain in combination with the supplemented rich medium designed in previous chapters (above 4.2.3) the problem of low expression levels was overcome and production of ScPmrC in large quantities was achieved. This agrees with previous reports showing that disulphide bonds are important elements in the structure of phosphoethanolamine transferases, such as MCR and EptA, and indispensable for phenotypic colistin resistance [352].

Despite the aggregation issues, CD spectroscopy and activity assays confirmed that the purified protein is correctly folded in solution. When ScPmrC activity was tested using our *in vitro* assay it showed higher levels of activity than the MCR-2 soluble domain, cMCR-2, although it can be noted that after incubating ScPmrC with the chromogenic compound *p*NP-PEtN for 15 hours, the reaction maintained a linear increase and had not reached completion. The calculated V_{max} figure of 1881 nM/min for the assay with 50 mM ScPmrC corresponds to a specific activity of 1.01 nmols/min*mg protein. This is close to the value of 3.1 nmols/min*mg reported for PNP-*p*EtN breakdown catalysed by BcsG. Comparing these specific activities, and given that ScPmrC was able to catalyse the removal of *p*NP from *p*NP-PEtN, we can conclude that ScPmrC possesses phosphoethanolamine transferase activity; this is important as future studies aiming to find or development inhibitors of phosphoethanolamine transferases should include PmrC. There is still a need for further experimentation, specifically using full-length PmrC, given that studies of MCR-1 and MCR-2 using engineered strains showed that transmembrane deletion mutants lack the unique modification of Lipid A and are

not capable of conferring phenotypical colistin resistance[336], demonstrating the role that the transmembrane domain plays in the activity of the enzyme.

Melting temperature analysis showed that the catalytic domain of PmrC, here named ScPmrC, appears more unstable at higher temperatures. This can be attributed to several causes; for example the presence of extra cysteines on the surface of the catalytic domain of PmrC, as surface-located cysteines are more involved in the formation of disulphide bond formation and exchanges facilitating protein to protein disulphide linkages that promotes physical aggregation of proteins [353]. Another explanation for this is the lack of the transmembrane domain; the MCR-1 transmembrane domain is required to confer phenotypic resistance to colistin in *E. coli* [336]. Finally, this difference in thermal stability might also be attributed to the different source organisms of PmrC and MCR-1, *A. baumannii* being an environmental opportunistic pathogen [354] whereas *E. coli* is a gut commensal accustomed adapted to higher temperatures.

A 3D model of ScPmrC (Figure 56 a)) was obtained using the online Swiss-Model server [123,355] and the structures of lipooligosaccharide phosphoethanolamine transferase A from *Neisseria meningitidis* (PDB 4KAV and 5FGN, sequence identity of 43.96% and 45.66% respectively), MCR-1 (PDB 5LRN sequence identity of 41.18%) and MCR-2 (PDB 5MX9 sequence identity of 39.48%) as a template. A visualization of the secondary structure content of the protein was obtained using Stride Visual Assignment [124] and the 3D structure obtained from Swiss-Model (Figure 56 e) and an estimation of the percentage of α -helices and β -sheets was obtained from CD analysis (Figure 64 e)). The predicted model of ScPmrC is structurally similar to its homologue, cMCR-1 (PDB Entry 5LRM) as can be seen in Figure 56 a) and b), with little difference in the overall folds and similar positioning of the conserved amino acids present in the catalytic pocket (Figure 56 c and d). This suggests that, like MCR-1, PmrC requires the presence of a metal ion, most likely zinc, in its catalytic domain to carry

out its proper function, zinc binding analysis by spectroscopy or using fluorescent dyes could shed some light on this, however, obtaining a 3D structure from a protein crystal can be another effective way to confirm this. This is the first study that has achieved the production and characterization of a recombinant *PmrC* from *A. baumannii*. This is a first step for further research aiming to better study this relevant colistin resistance mechanism and the identification of possible inhibitors.

Chapter 7 General conclusions and future work.

MCR-1 and PmrC are phosphoethanolamine transferases that confer colistin resistance on Gram-negative bacteria by catalysing the addition of PEA to Lipid A, altering its charge and thus reducing binding of colistin and other cationic antimicrobial peptides. Colistin is regarded as a “last-resort” antibiotic given its efficacy against the rising number of multidrug-resistant Gram-negative bacteria pathogens; highlighting the importance of studying these colistin resistance conferring mechanisms, MCR-1 and PmrC. The aims of this project is to shed light on the recently discovered mobile colistin resistance mechanism *mcr-1* and its homologues, to obtain insights into the place of MCR-1 and other bacterial phosphoethanolamine transferases among the members of the alkaline phosphatase superfamily taking a deep interest into the structure of their catalytic site and the metal ion contents of it, to obtain full-length, membrane-bound recombinant MCR-1 for its biophysical characterization, The development of an enzyme activity assay employing the recombinant catalytic domains of MCR-1 and its homologue MCR-2, suitable for use as a platform with which to investigate the mechanism of MCR-1 and other phosphoethanolamine transferases and finally to produce the catalytic domain of the chromosomal phosphoethanolamine transferase PmrC from *A. baumannii* to investigate its activity and structure.

MCR-1 and other members of the MCR family have been identified as members of the alkaline phosphatase superfamily, sharing the structurally conserved $\alpha/\beta/\alpha$ fold; however, further than this the relationship between MCR and the alkaline phosphatase superfamily remains to be understood. This work includes a structural and sequence-based analysis to better understand the relationship between MCR-1 and other phosphoethanolamine transferases and the alkaline phosphatase superfamily, with a special interest in the conservation of the catalytic site architecture and metal ion interactions. Sequence-based analyses are often used to find

evolutionary relationships involving recently discovered or uncharacterized enzymes, such as MCR-1.

In this work, a structural and sequence-based approach was used to better understand the relationship between MCR-1 and the alkaline phosphatase superfamily and identify possible areas to focus on in future research. Our analysis allocates MCR-1, and other phosphoethanolamine transferases, as being more structurally similar to the monometallic members of the arylsulfatase/sulphatase group than to the multinuclear members of the alkaline phosphatase superfamily. However, structural comparisons and sequence-based analysis showed that, with respect to the catalytic site architecture, MCR-1 is closer to the latter group, having a more complete second metal site than the arylsulfatase/sulphatase enzymes. This conservation is present in multiple members of the phosphoethanolamine transferase enzyme family, but excludes EptC [221] and enzymes such as BcsG that transfer phosphoethanolamine onto substrates other than lipid A (Figure 25). In particular, the conservation of residues His395 and His478 in the MCR-1 active site supports the possibility of a two-metal ion mechanism for PEtN transfer to Lipid A. One way to determine the role of His478 could be to investigate the effect of replacing His478 with another residue such as arginine (as found in the active sites of enzymes such as EptC or BcsG) on colistin MICs of MCR-1-expressing bacteria. Structures of the catalytic domains of both MCR-1 and MCR-2 presenting two zinc ions in their catalytic sites have been reported [74,115]; but in these structures the second zinc ion is stabilised by interactions involving a second MCR molecule. Another possibility is that this second metal ion site might be completed by a residue present in the N-terminal transmembrane domain. A study of the residues present in the N-terminal transmembrane domain could provide better insight into this.

The current work presents an expression and purification protocol for the production of full length MCR-1 from *E. coli*. The combination of an overexpression pET plasmid, the bacterial

strain C43(DE3) and supplemented rich culture media yielded high levels of both biomass and full-length MCR-1. The major role, in this overexpression, is played by the expression strain used; *E. coli* strain C43(DE3) is a strain used for the production of proteins such as membrane proteins that are toxic and deleterious for the bacterium [356]. As has been noted, MCR-1 overexpression can be damaging for the bacteria [55], for this reason, this strain is most suitable to use as an expression vector for recombinant full-length MCR-1. Culture media plays an important role too, as it directly affects the environment of the bacteria. The use of supplementation of the carbon source as well as addition of other trace elements facilitates and supports bacterial growth at higher cellular densities. The presence of both glycerol and glucose in the medium allows the maintenance of these higher cell densities while preventing the acidification of the growth culture. During growth in high cell-density cultures, a truncated version of MCR-1 was found, named tMCR-1 in this work. This tMCR-1 is a result of MCR-1 fragmentation, losing a ~20 kDa portion from the N-terminal transmembrane domain. Despite our best efforts we could not definitively identify in which part of the expression and purification process this fragmentation was taking place, however, a possible cleavage site was identified by mass spectrometry. Protease inhibitors were used all across the recovery and purification process, aiming to prevent proteolytic degradation, but this was not fully successful in preventing degradation. Mass spectrometry analysis of purified tMCR-1 showed the presence of multiple protein species with similar molecular weight, around 38 kDa, suggesting that there might be more than one cleavage site in the enzyme, suggesting that cleavage might be caused by mechanical shearing rather than proteolytic degradation. Future work could focus on generating point mutants around the suggested cleavage site (residue Leu205), to enhance protein stability and reduce this degradation.

Even with this fragmentation happening, the production of purified recombinant MCR-1 at a high yield was achieved. Circular dichroism and thermal stability analysis confirmed that the recombinantly produced MCR-1 remains stable in DDM micelles. This can be used in future

work as a stepping stone for the crystallization and eventual elucidation of an MCR-1 crystal structure, as the structures of other phosphoethanolamine transferases, such as EptA from *N. meningitidis*, haven been solved while isolated in DDM micelles [121].

Currently, most ways of analysing the activity of phosphoethanolamine transferases such as MCR-1 have been through *in vivo* assays of colistin minimal inhibitory concentration, or discontinuous (end point) assays detecting modification of lipid A or reporter substrates. The current work aimed to develop an activity assay that would provide a more accessible method for protein activity detection as well as a tool to screen for possible inhibitors of this antimicrobial resistance mechanism. The structure of the catalytic domain of MCR-1 [74], and its homologue MCR-2[115], were available prior to the beginning of this work and were used as building blocks to develop an activity assay that serves as a platform to investigate the mechanism phosphoethanolamine transferases.

An *in vitro* activity assay was developed to monitor the activity of the catalytic domains of MCR-1 and MCR-2, as well as other phosphoethanolamine transferases (e.g. *A. baumannii* PmrC), to assess the functional integrity of the soluble domains of these recombinant enzymes and to serve as a platform for inhibitor screening. This work took advantage of development by collaborators (Dr. Adam Gorman and Prof. Paul Pringle from the School of Chemistry at the University of Bristol) of a new and efficient route to synthesis of a chromogenic substrate, 4-Nitrophenol-phosphoethanolamine (*p*NP-PEtN), to create an *in vitro* chromogenic activity assay to study MCR proteins as well as other phosphoethanolamine transferases. The assay efficiently monitors the cleavage of the phosphoethanolamine (PEtN) motif from the chromophoric *p*NP-PEtN substrate, and the release of *p*NP, and demonstrates activity of the catalytic domains of recombinant MCR-2 and *A. baumannii* PmrC. The weak activity can be associated with a number of reasons, among them the lack of a transmembranal domain as it have been proved that the N-terminal transmembranal domain is necessary to confer

phenotypic resistance in members of the MCR family of enzymes [336,357] or to catalyse the cleavage of phosphoethanolamine from the fluorescently labelled substrate NBD-PEA in the case of the phosphoethanolamine EptA [121]. Another reason is the varying phosphorylation levels presented by these enzymes. Nonetheless; this *in vitro* assay proved its function as a scaffold for the screening of small-molecule inhibitors of phosphoethanolamine transferases, with thioglycolic acid (TGA) demonstrated as a weak inhibitor of MCR-1. This assay also helped to obtain kinetic constants for the first part of the enzymatic reaction.

The full catalytic reaction was not studied as the assay was performed in the absence of the natural substrate, Lipid A, thus the kinetic parameters obtained are only relevant for the first half of the reaction, i.e. addition of phosphoethanolamine to the acceptor Thr residue. Investigation of the addition of PEtN to Lipid A requires further research. Studying full-length MCR-1, and homologues, in a more physiological environment for membrane proteins, such as liposomes, or while immobilised on e.g. assay plates, would allow for more detailed investigations into the kinetics of this family of enzymes. It has already been established that the lack of the transmembrane domain can affect the activity of the enzyme and its kinetics. This can account for the low binding affinity of the soluble domain of MCR-1, cMCR-1, for the chromogenic substrate, *p*NP-PEtN, and reflects the importance to activity of the transmembrane domain.

Finally, another clinically relevant phosphoethanolamine transferase, *A. baumannii* PmrC, was produced and characterised. Using a codon-optimized sequence, a “soluble domain only” construct of PmrC, here named ScPmrC, was recombinantly produced. A slight variation of the best conditions found in MCR-1 was used to produce this ScPmrC in large enough quantities. Due to the disulphide content, the expression strain was changed to *E. coli* strain ShuffleT7. Using this strain and the supplemented rich culture medium, enough quantities of ScPmrC were obtained to do a biophysical characterization and to obtain preliminary ScPmrC

protein crystals, though not of diffraction quality. Using the *in vitro* assay it was confirmed that this “soluble domain” only construct retained its phosphoethanolamine transferase activity as it was able to catalyse the removal of *p*NP from the chromogenic substrate with kinetic parameters similar to those obtained for MCR-2. Circular dichroism analysis demonstrated that this construct is thermolabile and prone to precipitation under relatively low temperatures, with precipitation observed even at room temperature. Even though preliminary crystals of ScPmrC were obtained and conditions for crystallization were identified, a crystal large enough to perform elucidation of the protein structure by X-ray diffraction was not obtained due to time constraints and limitations to laboratory access due to the ongoing COVID-19 pandemic. Despite the lack of a 3D structure, a model of the protein was obtained using homology modelling software with other phosphoethanolamine transferases, MCR-1 and *N. meningitidis* EptA, as templates. This model showed that ScPmrC is structurally similar to MCR-1, but with the presence of several cysteine residues and potential disulphide bonds on the protein surface, which may account for the dependence of expression on disulphide formation and the susceptibility of this protein to precipitation. Future work should focus on the production and purification of a full-length PmrC as well as obtaining a 3D structure of ScPmrC. The advances made in this work, regarding ScPmrC production and preliminary protein crystallization, can be used as a starting point for this purpose.

The antibiotic colistin is often the last defence against multi-drug resistant Gram-negative bacteria. Phosphoethanolamine transferases present a threat to the effectiveness of colistin at treating infections caused by multi-drug resistant Gram-negative bacteria, especially if these are the mobilised phosphoethanolamine transferases from the MCR family. The findings of this work serve as a platform to build upon for better understanding of the action of these enzymes, and for the development of strategies to tackle this antimicrobial resistance mechanism. The production and purification of the full length MCR-1 and several “soluble domain only” constructs as well as a simple and robust activity assay; will facilitate further

research into inhibitor discovery. The activity assay allows screening of small molecule inhibitor candidates for different phosphoethanolamine transferases, including PmrC as well as MCR-1/MCR-2, aiming to identify inhibitors active against a range of colistin-resistant species. Ultimately, agents also inhibiting other Lipid A-modifying resistance mechanisms, such as *K. pneumoniae* ArnT that catalyses the addition of 4-amino-4-deoxy-l-arabinose to Lipid A (the same substrate as phosphoethanolamine transferases) would be of the greatest clinical use. Combining these with polymyxin derivatives now in clinical development [358,359] would offer new routes to treating some of the most antibiotic-resistant infections.

References:

- [1] Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. *P & T : A Peer-Reviewed Journal for Formulary Management* (2015) 2015;40:277–83.
- [2] Zaffiri L, Gardner J, Toledo-Pereyra LH. History of antibiotics. from salvarsan to cephalosporins. *Journal of Investigative Surgery* 2012;25:67–77. <https://doi.org/10.3109/08941939.2012.664099>.
- [3] Marston HD, Dixon DM, Knisely JM, Palmore TN, Fauci AS. Antimicrobial resistance. *JAMA - Journal of the American Medical Association* 2016;316:1193–204. <https://doi.org/10.1001/jama.2016.11764>.
- [4] Wallace WC, Cinat ME, Nastanski F, Gornick WB, Wilson SE. New epidemiology for postoperative nosocomial infections. *American Surgeon* 2000;66:874–8.
- [5] Ribeiro da Cunha B, Fonseca LP, Calado CRC. *Antibiotic Discovery: Where Have We Come from, Where Do We Go?* Antibiotics (Basel, Switzerland) 2019;8. <https://doi.org/10.3390/antibiotics8020045>.
- [6] Fleming A. On the Antibacterial Action of Cultures of a *Penicillium*, with Special Reference to their Use in the Isolation of *B. influenzae*. *British Journal of Experimental Pathology* 1929;10:226.
- [7] Davies J. Where have all the antibiotics gone? *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 17, Hindawi Limited; 2006, p. 287–90. <https://doi.org/10.1155/2006/707296>.
- [8] Lewis K. Recover the lost art of drug discovery. *Nature* 2012;485:439–40. <https://doi.org/10.1038/485439a>.
- [9] Bentley R. Different roads to discovery; Prontosil (hence sulfa drugs) and penicillin (hence β -lactams). *Journal of Industrial Microbiology & Biotechnology* 2009;36:775–86. <https://doi.org/10.1007/s10295-009-0553-8>.
- [10] Deng S, Ma X, Su E, Wei D. Efficient cascade synthesis of ampicillin from penicillin G potassium salt using wild and mutant penicillin G acylase from *Alcaligenes faecalis*. *Journal of Biotechnology* 2016;219:142–8. <https://doi.org/10.1016/j.jbiotec.2015.12.034>.
- [11] Gould K. Antibiotics: From prehistory to the present day. *Journal of Antimicrobial Chemotherapy* 2016;71:572–5. <https://doi.org/10.1093/jac/dkv484>.
- [12] Reardon S. WHO warns against “post-antibiotic” era. *Nature* 2014. <https://doi.org/10.1038/nature.2014.15135>.
- [13] Sengupta S, Chattopadhyay MK, Grossart H-P. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in Microbiology* 2013;4:47. <https://doi.org/10.3389/fmicb.2013.00047>.
- [14] Ventola CL. The antibiotic resistance crisis: part 2: management strategies and new agents. *P & T : A Peer-Reviewed Journal for Formulary Management* 2015;40:344–52.
- [15] Shier WT. On the origin of antibiotics and mycotoxins. *Toxin Reviews* 2011;30:6–30. <https://doi.org/10.3109/15569543.2011.550862>.

- [16] Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, et al. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *Journal of Bacteriology* 2008;190:4050–60. <https://doi.org/10.1128/JB.00204-08>.
- [17] Gaynes R. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Diseases* 2017;23:849–53. <https://doi.org/10.3201/eid2305.161556>.
- [18] Džidić S, Šušković J, Kos B. Antibiotic resistance mechanisms in bacteria: Biochemical and genetic aspects. *Food Technology and Biotechnology* 2008;46:11–21.
- [19] Kapoor G, Saigal S, Elongavan A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *Journal of Anaesthesiology, Clinical Pharmacology* 2017;33:300–5. https://doi.org/10.4103/joacp.JOACP_349_15.
- [20] Abraham EP, Chain E. An enzyme from bacteria able to destroy penicillin. *Nature* 1940;146:837. <https://doi.org/10.1038/146837a0>.
- [21] Levy SB, Bonnie M, Marshall B. Antibacterial resistance worldwide: Causes, challenges and responses. *Nature Medicine* 2004;10:S122–9. <https://doi.org/10.1038/nm1145>.
- [22] Sengupta S, Chattopadhyay MK, Grossart HP. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in Microbiology* 2013;4:47. <https://doi.org/10.3389/fmicb.2013.00047>.
- [23] Paterson DL. Resistance in gram-negative bacteria: Enterobacteriaceae. *American Journal of Infection Control* 2006;34:20–8. <https://doi.org/10.1016/j.ajic.2006.05.238>.
- [24] Rupp M, Fey PD, Rupp ME, Fey PD. Extended Spectrum β -Lactamase (ESBL)-Producing Enterobacteriaceae Considerations for Diagnosis, Prevention and Drug Treatment. *Drugs* 2003;63:353–65.
- [25] Service PH, Services H. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *American Journal of Infection Control* 2004;32:470–85. <https://doi.org/10.1016/j.ajic.2004.10.001>.
- [26] Huttner B, Jones M, Rubin MA, Neuhauser MM, Gundlapalli A, Samore M. Drugs of last resort? the use of polymyxins and tigecycline at us veterans affairs medical centers, 2005-2010. *PLoS ONE* 2012;7:e36649. <https://doi.org/10.1371/journal.pone.0036649>.
- [27] McCarthy M. CDC calls for urgent action to combat rise of drug resistant pathogens. *BMJ (Clinical Research Ed)* 2013;347:5649. <https://doi.org/10.1136/bmj.f5649>.
- [28] Davies J, Davies D. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*: MMBR 2010;74:417–33. <https://doi.org/10.1128/MMBR.00016-10>.
- [29] Directorate. VM. UK One Health Report - Joint report on antibiotic use and antibiotic resistance, 2013–2017. *Veterinary Medicines Directorate* 2019:64.
- [30] Falagas ME, Michalopoulos A. Polymyxins: Old antibiotics are back. *Lancet* 2006;367:633–4. [https://doi.org/10.1016/S0140-6736\(06\)68241-X](https://doi.org/10.1016/S0140-6736(06)68241-X).

- [31] Falagas ME, Rafailidis PI, Matthaïou DK, Virtzili S, Nikita D, Michalopoulos A. Pandrug-resistant *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* infections: Characteristics and outcome in a series of 28 patients. *International Journal of Antimicrobial Agents* 2008;32:450–4. <https://doi.org/10.1016/j.ijantimicag.2008.05.016>.
- [32] Amabile-Cuevas C. Antibiotic resistance in Mexico: a brief overview of the current status and its causes. *Journal of Infection in Developing Countries* 2010;4:126–31. <https://doi.org/10.3855/jidc.427>.
- [33] Van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin SA, et al. Global antibiotic consumption 2000 to 2010: An analysis of national pharmaceutical sales data. *The Lancet Infectious Diseases* 2014;14:742–50. [https://doi.org/10.1016/S1473-3099\(14\)70780-7](https://doi.org/10.1016/S1473-3099(14)70780-7).
- [34] Li J, Nation RL, Turnidge JD, Milne RW, Coulthard K, Rayner CR, et al. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *The Lancet Infectious Diseases* 2006;6:589–601. [https://doi.org/10.1016/S1473-3099\(06\)70580-1](https://doi.org/10.1016/S1473-3099(06)70580-1).
- [35] Falagas ME, Rafailidis PI, Ioannidou E, Alexiou VG, Matthaïou DK, Karageorgopoulos DE, et al. Colistin therapy for microbiologically documented multidrug-resistant Gram-negative bacterial infections: a retrospective cohort study of 258 patients. *International Journal of Antimicrobial Agents* 2010;35:194–9. <https://doi.org/10.1016/j.ijantimicag.2009.10.005>.
- [36] Bos MP, Robert V, Tommassen J. Biogenesis of the Gram-Negative Bacterial Outer Membrane. *Annual Review of Microbiology* 2007;61:191–214. <https://doi.org/10.1146/annurev.micro.61.080706.093245>.
- [37] Koebnik R, Locher KP, Van Gelder P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular Microbiology* 2000;37:239–53. <https://doi.org/10.1046/j.1365-2958.2000.01983.x>.
- [38] Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* 2018;8. <https://doi.org/10.1128/ecosalplus.ESP-0001-2018>.
- [39] Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* 2018;8. <https://doi.org/10.1128/ecosalplus.esp-0001-2018>.
- [40] Nikaldo H, Nakae T. *The Outer Membrane of Gram-negative Bacteria*. vol. 20. 1980. [https://doi.org/10.1016/S0065-2911\(08\)60208-8](https://doi.org/10.1016/S0065-2911(08)60208-8).
- [41] Delcour AH. Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta - Proteins and Proteomics* 2009;1794:808–16. <https://doi.org/10.1016/j.bbapap.2008.11.005>.
- [42] Whitfield C, Trent MS. Biosynthesis and Export of Bacterial Lipopolysaccharides. *Annual Review of Biochemistry* 2014;83:99–128. <https://doi.org/10.1146/annurev-biochem-060713-035600>.
- [43] Klein G, Raina S. Regulated assembly of LPS, its structural alterations and cellular response to LPS defects. *International Journal of Molecular Sciences* 2019;20:356. <https://doi.org/10.3390/ijms20020356>.

- [44] Cryz SJ, Pitt TL, Fürer E, Germanier R. Role of lipopolysaccharide in virulence of *Pseudomonas aeruginosa*. *Infection and Immunity* 1984;44:508–13. <https://doi.org/10.1128/iai.44.2.508-513.1984>.
- [45] Zhang L, Rozek A, Hancock REW. Interaction of Cationic Antimicrobial Peptides with Model Membranes* 2001. <https://doi.org/10.1074/jbc.M104925200>.
- [46] Hancock REW, Diamond G. The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology* 2000;8:402–10. [https://doi.org/10.1016/S0966-842X\(00\)01823-0](https://doi.org/10.1016/S0966-842X(00)01823-0).
- [47] Ciumac D, Gong H, Hu X, Lu JR. Membrane targeting cationic antimicrobial peptides. *Journal of Colloid and Interface Science* 2019;537:163–85. <https://doi.org/10.1016/j.jcis.2018.10.103>.
- [48] Glukhov E, Stark M, Burrows LL, Deber CM. Basis for selectivity of cationic antimicrobial peptides for bacterial versus mammalian membranes. *Journal of Biological Chemistry* 2005;280:33960–7. <https://doi.org/10.1074/jbc.M507042200>.
- [49] Wang J, Dou X, Song J, Lyu Y, Zhu X, Xu L, et al. Antimicrobial peptides: Promising alternatives in the postfeeding antibiotic era. *Medicinal Research Reviews* 2018. <https://doi.org/10.1002/med.21542>.
- [50] Vardanyan R, Hruby V. Chapter 30 – Antibiotics. *Synthesis of Best-Seller Drugs*, Academic Press; 2016, p. 573–643. <https://doi.org/10.1016/B978-0-12-411492-0.00030-4>.
- [51] Li J, Nation RL, Milne RW, Turnidge JD, Coulthard K. Evaluation of colistin as an agent against multi-resistant Gram-negative bacteria. *International Journal of Antimicrobial Agents* 2005;25:11–25. <https://doi.org/10.1016/j.ijantimicag.2004.10.001>.
- [52] Poirel L, Jayol A, Nordmann P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. *Clinical Microbiology Reviews* 2017;30:557–96. <https://doi.org/10.1128/CMR.00064-16>.
- [53] Pristovšek P, Kidrič J. Solution structure of polymyxins B and E and effect of binding to lipopolysaccharide: An NMR and molecular modeling study. *Journal of Medicinal Chemistry* 1999;42:4604–13. <https://doi.org/10.1021/jm991031b>.
- [54] Mares J, Kumaran S, Gobbo M, Zerbe O. Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy. *Journal of Biological Chemistry* 2009;284:11498–506. <https://doi.org/10.1074/jbc.M806587200>.
- [55] Yang Q, Li M, Spiller OB, Andrey DO, Hinchliffe P, Li H, et al. Balancing mcr-1 expression and bacterial survival is a delicate equilibrium between essential cellular defence mechanisms. *Nature Communications* 2017;8:2054. <https://doi.org/10.1038/s41467-017-02149-0>.
- [56] Lim LM, Ly N, Anderson D, Yang JC, Macander L, Jarkowski A, et al. Resurgence of colistin: A review of resistance, toxicity, pharmacodynamics, and dosing. vol. 30. Wiley-Blackwell; 2010. <https://doi.org/10.1592/phco.30.12.1279>.
- [57] Trimble MJ, Mlynářčík P, Kolář M, Hancock REW. Polymyxin: Alternative Mechanisms of Action and Resistance. *Cold Spring Harbor Perspectives in Medicine* 2016;6:a025288. <https://doi.org/10.1101/CSHPERSPECT.A025288>.

- [58] Nation RL, Li J. Colistin in the 21st century. *Current Opinion in Infectious Diseases* 2009;22:535–43. <https://doi.org/10.1097/QCO.0b013e328332e672>.
- [59] Biswas S, Brunel JM, Dubus JC, Reynaud-Gaubert M, Rolain JM. Colistin: An update on the antibiotic of the 21st century. *Expert Review of Anti-Infective Therapy* 2012;10:917–34. <https://doi.org/10.1586/eri.12.78>.
- [60] Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *The Lancet Infectious Diseases* 2016;16:161–8. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7).
- [61] Ye H, Li Y, Li Z, Gao R, Zhang H, Wen R, et al. Diversified mcr-1-harboring plasmid reservoirs confer resistance to colistin in human gut microbiota. *MBio* 2016;7:e00177. <https://doi.org/10.1128/mBio.00177-16>.
- [62] Martti V. Polymyxins and their novel derivatives. *Curr Opin Microbiol* 2010;13:574–81. <https://doi.org/10.1016/j.mib.2010.09.002>.
- [63] Gales AC, Jones RN, Sader HS. Global assessment of the antimicrobial activity of polymyxin B against 54 731 clinical isolates of Gram-negative bacilli: report from the SENTRY antimicrobial surveillance programme (2001–2004). *Clinical Microbiology and Infection* 2006;12:315–21. <https://doi.org/10.1111/J.1469-0691.2005.01351.X>.
- [64] Bakthavatchalam YD, Pragasam AK, Biswas I, Veeraraghavan B. Polymyxin susceptibility testing, interpretative breakpoints and resistance mechanisms: An update. *Journal of Global Antimicrobial Resistance* 2018;12:124–36. <https://doi.org/10.1016/j.jgar.2017.09.011>.
- [65] Zhou Z, Ribeiro AA, Lin S, Cotter RJ, Miller SI, Raetz CRH. Lipid A modifications in polymyxin-resistant *Salmonella typhimurium*: PmrA-dependent 4-amino-4-deoxy-L-arabinose, and phosphoethanolamine incorporation. *Journal of Biological Chemistry* 2001;276:43111–21. <https://doi.org/10.1074/jbc.M106960200>.
- [66] Bishop RE, Gibbons HS, Guina T, Trent MS, Miller SI, Raetz CRH. Transfer of palmitate from phospholipids to lipid A in outer membranes of Gram-negative bacteria. *EMBO Journal* 2000;19:5071–80. <https://doi.org/10.1093/emboj/19.19.5071>.
- [67] Yan A, Guan Z, Raetz CRH. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. *Journal of Biological Chemistry* 2007;282:36077–89. <https://doi.org/10.1074/jbc.M706172200>.
- [68] Mitrophanov AY, Groisman EA. Signal integration in bacterial two-component regulatory systems. *Genes and Development* 2008;22:2601–11. <https://doi.org/10.1101/gad.1700308>.
- [69] Chang C, Stewart RC. The two-component system: Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiology* 1998;117:723–31. <https://doi.org/10.1104/pp.117.3.723>.
- [70] Cannatelli A, Giani T, D'Andrea MM, Di Pilato V, Arena F, Conte V, et al. MgrB Inactivation Is a Common Mechanism of Colistin Resistance in KPC-Producing *Klebsiella pneumoniae* of Clinical Origin 2014;58. <https://doi.org/10.1128/AAC.03110-14>.

- [71] Jeannot K, Bolard A, Plésiat P. Resistance to polymyxins in Gram-negative organisms. *International Journal of Antimicrobial Agents* 2017;49:526–35. <https://doi.org/10.1016/j.ijantimicag.2016.11.029>.
- [72] Olaitan AO, Morand S, Rolain J-MM. Mechanisms of polymyxin resistance: Acquired and intrinsic resistance in bacteria. vol. 5. *Frontiers*; 2014. <https://doi.org/10.3389/fmicb.2014.00643>.
- [73] Baron S, Hadjadj L, Rolain JM, Olaitan AO. Molecular mechanisms of polymyxin resistance: knowns and unknowns. *International Journal of Antimicrobial Agents* 2016;48:583–91. <https://doi.org/10.1016/j.ijantimicag.2016.06.023>.
- [74] Hinchliffe P, Yang QE, Portal E, Young T, Li H, Tooke CL, et al. Insights into the Mechanistic Basis of Plasmid-Mediated Colistin Resistance from Crystal Structures of the Catalytic Domain of MCR-1. *Scientific Reports* 2017;7:39392. <https://doi.org/10.1038/srep39392>.
- [75] Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *The Lancet Infectious Diseases* 2016;16:161–8. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7).
- [76] Raetz CRHH, Reynolds CM, Trent MS, Bishop RE. Lipid A Modification Systems in Gram-Negative Bacteria. *Annual Review of Biochemistry* 2007;76:295–329. <https://doi.org/10.1146/annurev.biochem.76.010307.145803>.
- [77] Lesho E, Yoon EJ, Mcgann P, Snesrud E, Kwak Y, Milillo M, et al. Emergence of colistin-resistance in extremely drug-resistant acinetobacter baumannii containing a novel pmrCAB operon during colistin therapy of wound infections. *Journal of Infectious Diseases* 2013;208:1142–51. <https://doi.org/10.1093/infdis/jit293>.
- [78] Wright MS, Suzuki Y, Jones MB, Marshall SH, Rudin SD, van Duijn D, et al. Genomic and transcriptomic analyses of colistin-resistant clinical isolates of *Klebsiella pneumoniae* reveal multiple pathways of resistance. *Antimicrobial Agents and Chemotherapy* 2015;59:536–43. <https://doi.org/10.1128/AAC.04037-14>.
- [79] Cox AD, Wright JC, Li J, Hood DW, Moxon ER, Richards JC. Phosphorylation of the lipid A region of meningococcal lipopolysaccharide: identification of a family of transferases that add phosphoethanolamine to lipopolysaccharide. *Journal of Bacteriology* 2003;185:3270–7. <https://doi.org/10.1128/JB.185.11.3270-3277.2003>.
- [80] Fage CD, Brown DB, Boll JM, Keatinge-Clay AT, Trent MS, IUCr. Crystallographic study of the phosphoethanolamine transferase EptC required for polymyxin resistance and motility in *Campylobacter jejuni*. *Acta Crystallographica Section D: Biological Crystallography* 2014;70:2730–9. <https://doi.org/10.1107/S1399004714017623>.
- [81] Anjum MF, Duggett NA, AbuOun M, Randall L, Nunez-Garcia J, Ellis RJ, et al. Colistin resistance in *Salmonella* and *Escherichia coli* isolates from a pig farm in Great Britain. *Journal of Antimicrobial Chemotherapy* 2016;71:2306–13. <https://doi.org/10.1093/jac/dkw149>.
- [82] Arcilla MS, van Hattem JM, Matamoros S, Melles DC, Penders J, de Jong MD, et al. Dissemination of the mcr-1 colistin resistance gene. *The Lancet Infectious Diseases* 2016;16:147–9. [https://doi.org/10.1016/S1473-3099\(15\)00541-1](https://doi.org/10.1016/S1473-3099(15)00541-1).

- [83] Rhouma M, Beaudry F, Thériault W, Letellier A, Thériault W, Letellier A, et al. Colistin in Pig Production: Chemistry, Mechanism of Antibacterial Action, Microbial Resistance Emergence, and One Health Perspectives. *Frontiers in Microbiology* 2016;7:1–22. <https://doi.org/10.3389/fmicb.2016.01789>.
- [84] Shen Z, Wang Y, Shen Y, Shen J, Wu C. Early emergence of *mcr-1* in *Escherichia coli* from food-producing animals. *The Lancet Infectious Diseases* 2016;16:293. [https://doi.org/10.1016/S1473-3099\(16\)00061-X](https://doi.org/10.1016/S1473-3099(16)00061-X).
- [85] Skov RL, Monnet DL. Plasmid-mediated colistin resistance (*mcr-1* gene): Three months later, the story unfolds. *Eurosurveillance* 2016;21:1–6. <https://doi.org/10.2807/1560-7917.ES.2016.21.9.30155>.
- [86] Walsh TR, Wu Y. China bans colistin as a feed additive for animals. *The Lancet Infectious Diseases* 2016;16:1102–3. [https://doi.org/10.1016/S1473-3099\(16\)30329-2](https://doi.org/10.1016/S1473-3099(16)30329-2).
- [87] Fernandes MMR, Moura Q, Sartori L, Silva KKC, Cunha MMP, Esposito F, et al. Silent dissemination of colistin-resistant *Escherichia coli* in South America could contribute to the global spread of the *mcr-1* gene. *Eurosurveillance* 2016;21:30214. <https://doi.org/10.2807/1560-7917.ES.2016.21.17.30214>.
- [88] Park KS, Lee JH, Park M, Ko KS, Lee SH. How many *mcr-1*-harbouring bacteria were spreading geographically? *Biomedical Research (India)* 2017;28:1659–63.
- [89] Nang SC, Li J, Velkov T. The rise and spread of *mcr* plasmid-mediated polymyxin resistance. *Critical Reviews in Microbiology* 2019;45:131–61. <https://doi.org/10.1080/1040841X.2018.1492902>.
- [90] Ellem JA, Ginn AN, Chen SC-A, Ferguson J, Partridge SR, Iredell JR. Locally acquired *mcr-1* in *Escherichia coli*, Australia, 2011 and 2013. *Emerging Infectious Diseases* 2017;23:1160–3. <https://doi.org/10.3201/eid2307.161638>.
- [91] Poirel L, Nordmann P. Emerging plasmid-encoded colistin resistance: The animal world as the culprit? *Journal of Antimicrobial Chemotherapy* 2016;71:2326–7. <https://doi.org/10.1093/jac/dkw074>.
- [92] Liakopoulos A, Mevius DJ, Olsen B, Bonnedahl J. The colistin resistance *mcr-1* gene is going wild: Table 1. *Journal of Antimicrobial Chemotherapy* 2016;71:2335–6. <https://doi.org/10.1093/jac/dkw262>.
- [93] Ruzauskas M, Vaskeviciute L. Detection of the *mcr-1* gene in *Escherichia coli* prevalent in the migratory bird species *Larus argentatus*. *Journal of Antimicrobial Chemotherapy* 2016;71:2333–4. <https://doi.org/10.1093/jac/dkw245>.
- [94] Nang SC, Li J, Velkov T. The rise and spread of *mcr* plasmid-mediated polymyxin resistance. vol. 45. 2019. <https://doi.org/10.1080/1040841X.2018.1492902>.
- [95] Aghapour Z, Gholizadeh P, Ganbarov K, Bialvaei AZ, Mahmood SS, Tanomand A, et al. Molecular mechanisms related to colistin resistance in Enterobacteriaceae. *Infection and Drug Resistance* 2019;12:965–75. <https://doi.org/10.2147/IDR.S199844>.
- [96] Di Pilato V, Arena F, Tascini C, Cannatelli A, Henrici De Angelis L, Fortunato S, et al. *mcr-1.2*, a New *mcr* Variant Carried on a Transferable Plasmid from a Colistin-Resistant KPC Carbapenemase-Producing *Klebsiella pneumoniae* Strain of Sequence Type 512.

- [97] Cannatelli A, D'Andrea MM, Giani T, Di Pilato V, Arena F, Ambretti S, et al. In vivo emergence of colistin resistance in *Klebsiella pneumoniae* producing KPC-type carbapenemases mediated by insertional inactivation of the PhoQ/PhoP mgrB regulator. *Antimicrobial Agents and Chemotherapy* 2013;57:5521–6. <https://doi.org/10.1128/AAC.01480-13>.
- [98] Xiaomin S, Yiming L, Yuying Y, Zhangqi S, Yongning W, Shaolin W. Global impact of mcr-1-positive Enterobacteriaceae bacteria on “one health.” *Critical Reviews in Microbiology* 2020;46:565–77. <https://doi.org/10.1080/1040841X.2020.1812510>.
- [99] Quan J, Li X, Chen Y, Jiang Y, Zhou Z, Zhang H, et al. Prevalence of mcr-1 in *Escherichia coli* and *Klebsiella pneumoniae* recovered from bloodstream infections in China: a multicentre longitudinal study. *The Lancet Infectious Diseases* 2017;17:400–10. [https://doi.org/10.1016/S1473-3099\(16\)30528-X](https://doi.org/10.1016/S1473-3099(16)30528-X).
- [100] Shen Y, Wu Z, Wang Y, Zhang R, Zhou H-W, Wang S, et al. Heterogeneous and Flexible Transmission of mcr-1 in Hospital-Associated *Escherichia coli* Downloaded from n.d.;9:943–61. <https://doi.org/10.1128/mBio.00943-18>.
- [101] Shen Y, Zhou H, Xu J, Wang Y, Zhang Q, Walsh TR, et al. Anthropogenic and environmental factors associated with high incidence of mcr-1 carriage in humans across China. *Nature Microbiology* 2018;3:1054–62. <https://doi.org/10.1038/s41564-018-0205-8>.
- [102] Yang Y, Miao M, Mediavilla JR, Chen L, Tang Y-W, Du H, et al. Complete Sequences of mcr-1 -Harboring Plasmids from Extended-Spectrum- β -Lactamase- and Carbapenemase-Producing Enterobacteriaceae. *Antimicrobial Agents and Chemotherapy* 2016;60:4351–4. <https://doi.org/10.1128/aac.00550-16>.
- [103] Sun J, Zhang H, Liu YH, Feng Y. Towards Understanding MCR-like Colistin Resistance. *Trends in Microbiology* 2018;26:794–808. <https://doi.org/10.1016/j.tim.2018.02.006>.
- [104] Carroll LM, Gaballa A, Guldemann C, Sullivan G, Henderson LO, Wiedmann M. Identification of novel mobilized colistin resistance gene mcr-9 in a multidrug-resistant, colistin-susceptible salmonella enterica serotype typhimurium isolate. *MBio* 2019;10. <https://doi.org/10.1128/mBio.00853-19>.
- [105] Teale C, Anjum MF, Kirchner M, Stubberfield EJ, Nunez-Garcia J, Lemma F, et al. mcr-1 and mcr-2 (mcr-6.1) variant genes identified in *Moraxella* species isolated from pigs in Great Britain from 2014 to 2015. *Journal of Antimicrobial Chemotherapy* 2017;72:2745–9. <https://doi.org/10.1093/jac/dkx286>.
- [106] Zhang J, Chen L, Wang J, Butaye P, Huang K, Qiu H, et al. Molecular detection of colistin resistance genes (mcr-1 to mcr-5) in human vaginal swabs. *BMC Research Notes* 2018;11:143. <https://doi.org/10.1186/s13104-018-3255-3>.
- [107] Hadjadj L, Baron SA, Olaitan AO, Morand S, Rolain JM. Co-occurrence of Variants of mcr-3 and mcr-8 Genes in a *Klebsiella pneumoniae* Isolate From Laos. *Frontiers in Microbiology* 2019;10:2720. <https://doi.org/10.3389/fmicb.2019.02720>.
- [108] Rebelo AR, Bortolaia V, Kjeldgaard JS, Pedersen SK, Leekitcharoenphon P, Hansen IM, et al. Multiplex PCR for detection of plasmid-mediated colistin resistance

- determinants, mcr-1, mcr-2, mcr-3, mcr-4 and mcr-5 for surveillance purposes. *Eurosurveillance* 2018;23:17–00672. <https://doi.org/10.2807/1560-7917.ES.2018.23.6.17-00672>.
- [109] Rebelo AR, Bortolaia V, Kjeldgaard JS, Pedersen SK, Leekitcharoenphon P, Hansen IM, et al. Multiplex PCR for detection of plasmid-mediated colistin resistance determinants, mcr-1, mcr-2, mcr-3, mcr-4 and mcr-5 for surveillance purposes. *Eurosurveillance* 2018;23:17–00672. <https://doi.org/10.2807/1560-7917.ES.2018.23.6.17-00672>.
- [110] Stojanoski V, Sankaran B, Prasad BVV, Poirel L, Nordmann P, Palzkill T. Structure of the catalytic domain of the colistin resistance enzyme MCR-1. *BMC Biology* 2016;14:81. <https://doi.org/10.1186/s12915-016-0303-0>.
- [111] Wanty C, Anandan A, Piek S, Walshe J, Ganguly J, Carlson RW, et al. The structure of the neisserial lipooligosaccharide phosphoethanolamine transferase A (LptA) required for resistance to polymyxin. *Journal of Molecular Biology* 2013;425:3389–402. <https://doi.org/10.1016/j.jmb.2013.06.029>.
- [112] Hanson SR, Best MD, Wong CH. Sulfatases: Structure, mechanism, biological activity, inhibition, and synthetic utility. *Angewandte Chemie - International Edition* 2004;43:5736–63. <https://doi.org/10.1002/anie.200300632>.
- [113] Galperin MY, Jedrzejas MJ. Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes. *Proteins: Structure, Function and Genetics* 2001;45:318–24. <https://doi.org/10.1002/prot.1152>.
- [114] Stojanoski V, Sankaran B, Prasad BVV, Poirel L, Nordmann P, Palzkill T, et al. Structure of the catalytic domain of the colistin resistance enzyme MCR-1. *BMC Biology* 2016;14:81. <https://doi.org/10.1186/s12915-016-0303-0>.
- [115] Coates K, Walsh TR, Spencer J, Hinchliffe P. 1.12 Å resolution crystal structure of the catalytic domain of the plasmid-mediated colistin resistance determinant MCR-2. *Acta Crystallographica Section:F Structural Biology Communications*, vol. 73, International Union of Crystallography; 2017, p. 443–9. <https://doi.org/10.1107/S2053230X17009669>.
- [116] Millán JL. Alkaline phosphatases. *Purinergic Signalling* 2006;2:335–41. <https://doi.org/10.1007/s11302-005-5435-6>.
- [117] Barrozo A, Duarte F, Bauer P, Carvalho ATP, Kamerlin SCL. Cooperative Electrostatic Interactions Drive Functional Evolution in the Alkaline Phosphatase Superfamily. *Journal of the American Chemical Society* 2015;137:9061–76. <https://doi.org/10.1021/jacs.5b03945>.
- [118] Pabis A, Kamerlin SCL. Promiscuity and electrostatic flexibility in the alkaline phosphatase superfamily. *Current Opinion in Structural Biology* 2016;37:14–21. <https://doi.org/10.1016/j.sbi.2015.11.008>.
- [119] Stec B, Holtz KM, Kantrowitz ER. A revised mechanism for the alkaline phosphatase reaction involving three metal ions. *Journal of Molecular Biology* 2000;299:1303–11. <https://doi.org/10.1006/JMBI.2000.3799>.
- [120] Lythell E, Lythell E, Suardiaz R, Suardiaz R, Suardiaz R, Suardiaz R, et al. Resistance to the “last resort” antibiotic colistin: A single-zinc mechanism for phosphointermediate

- formation in MCR enzymes. *Chemical Communications* 2020;56:6874–7. <https://doi.org/10.1039/d0cc02520h>.
- [121] Anandan A, Evans GL, Condic-Jurkic K, O'Mara ML, John CM, Phillips NJ, et al. Structure of a lipid A phosphoethanolamine transferase suggests how conformational changes govern substrate binding. *Proceedings of the National Academy of Sciences* 2017;114:2218–23. <https://doi.org/10.1073/pnas.1612927114>.
- [122] Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research* 2017;45:D313–9. <https://doi.org/10.1093/nar/gkw1132>.
- [123] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research* 2018;46:W296–303. <https://doi.org/10.1093/nar/gky427>.
- [124] Heinig M, Frishman D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research* 2004;32:W500–2. <https://doi.org/10.1093/nar/gkh429>.
- [125] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 2019;47:W636–41. <https://doi.org/10.1093/nar/gkz268>.
- [126] Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G v, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Research* 2021;49:D437–51. <https://doi.org/10.1093/nar/gkaa1038>.
- [127] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>.
- [128] Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, et al. CATH: Increased structural coverage of functional space. *Nucleic Acids Research* 2021;49:D266–73. <https://doi.org/10.1093/nar/gkaa1079>.
- [129] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 2014;42:W320–4. <https://doi.org/10.1093/nar/gku316>.
- [130] Holm L, Rosenström P. Dali server: Conservation mapping in 3D. *Nucleic Acids Research* 2010;38:W545-9. <https://doi.org/10.1093/nar/gkq366>.
- [131] Aslam N, Nadeem A, Babar ME, Pervez MT, Aslam M, Naveed N, et al. The accuracy of protein structure alignment servers. *Electronic Journal of Biotechnology* 2016;20:9–13. <https://doi.org/10.1016/j.ejbt.2016.01.005>.
- [132] Holm L. DALI and the persistence of protein shape. *Protein Science* 2020;29:128–40. <https://doi.org/10.1002/pro.3749>.
- [133] Beceiro A, Llobet E, Aranda J, Bengoechea JA, Doumith M, Hornsey M, et al. Phosphoethanolamine modification of lipid A in colistin-resistant variants of *Acinetobacter baumannii* mediated by the pmrAB two-component regulatory system.

- Antimicrobial Agents and Chemotherapy 2011;55:3370–9.
<https://doi.org/10.1128/AAC.00079-11>.
- [134] Berrow NS, Alderton D, Owens RJ. The Precise Engineering of Expression Vectors Using High-Throughput In-Fusion™ PCR Cloning 2009. <https://doi.org/10.1007/978-1-59745-196-3>.
- [135] LAEMMLI UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 1970;227:680–5. <https://doi.org/10.1038/227680a0>.
- [136] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, Totowa, NJ: Humana Press; 2005, p. 571–607. <https://doi.org/10.1385/1-59259-890-0:571>.
- [137] Johnson KA, Goody RS. The original Michaelis constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry* 2011;50:8264–9. <https://doi.org/10.1021/bi201284u>.
- [138] Lineweaver H, Burk D. The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society* 1934;56:658–66. <https://doi.org/10.1021/ja01318a036>.
- [139] Caniaux I, van Belkum A, Zambardi G, Poirel L, Gros MF. MCR: modern colistin resistance. *European Journal of Clinical Microbiology and Infectious Diseases* 2017;36:415–20. <https://doi.org/10.1007/s10096-016-2846-y>.
- [140] Lythell E, Lythell E, Suardíaz R, Suardíaz R, Suardíaz R, Suardíaz R, et al. Resistance to the “last resort” antibiotic colistin: A single-zinc mechanism for phosphointermediate formation in MCR enzymes. *Chemical Communications* 2020;56:6874–7. <https://doi.org/10.1039/d0cc02520h>.
- [141] David L, Cheah E, Cygler M, Dijkstra B, Frolow F, Sybille M, et al. The α/β hydrolase fold. *Protein Engineering, Design and Selection* 1992;5:197–211. <https://doi.org/10.1093/protein/5.3.197>.
- [142] Jonas S, Hollfelder F. Mapping catalytic promiscuity in the alkaline phosphatase superfamily. *Pure and Applied Chemistry* 2009;81:731–42. <https://doi.org/10.1351/PAC-CON-08-10-20>.
- [143] Mohamed MF, Hollfelder F. Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer. *Biochimica et Biophysica Acta - Proteins and Proteomics* 2013;1834:417–24. <https://doi.org/10.1016/j.bbapap.2012.07.015>.
- [144] Duarte F, Amrein BA, Kamerlin SCL. Modeling catalytic promiscuity in the alkaline phosphatase superfamily. *Physical Chemistry Chemical Physics* 2013;15:11160–77. <https://doi.org/10.1039/c3cp51179k>.
- [145] Sharma U, Pal D, Prasad R. Alkaline phosphatase: An overview. *Indian Journal of Clinical Biochemistry* 2014;29:269–78. <https://doi.org/10.1007/s12291-013-0408-y>.
- [146] Galperin MY, Bairoch A, Koonin E v. A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Science* 1998;7:1829–35. <https://doi.org/10.1002/pro.5560070819>.

- [147] Jesse G. Zalatan §, Timothy D. Fenn II, Axel T. Brunger II and, Daniel Herschlag* §, I. Structural and Functional Comparisons of Nucleotide Pyrophosphatase/Phosphodiesterase and Alkaline Phosphatase: Implications for Mechanism and Evolution†,‡ 2006. <https://doi.org/10.1021/BI060847T>.
- [148] Schmidt B, Selmer T, Ingendoh A, Figurat K von. A novel amino acid modification in sulfatases that is defective in multiple sulfatase deficiency. *Cell* 1995;82:271–8. [https://doi.org/10.1016/0092-8674\(95\)90314-3](https://doi.org/10.1016/0092-8674(95)90314-3).
- [149] López-Canut V, Roca M, Bertrán J, Moliner V, Tuñón I. Promiscuity in alkaline phosphatase superfamily. Unraveling evolution through molecular simulations. *Journal of the American Chemical Society* 2011;133:12050–62. <https://doi.org/10.1021/ja2017575>.
- [150] Moss DW. Perspectives in Alkaline Phosphatase Research. *Clinical Chemistry* 1992;38:2486–92. <https://doi.org/10.1093/clinchem/38.12.2486>.
- [151] van Loo B, Jonas S, Babbie AC, Benjdia A, Berteau O, Hyvönen M, et al. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:2740–5. <https://doi.org/10.1073/pnas.0903951107>.
- [152] Bhadra R, Srinivasan N, Pandit SB. A new domain family in the superfamily of alkaline phosphatases. *In Silico Biology* 2005;5:379–87.
- [153] Orengo CA, Jones DT, Thornton JM. Protein superfamilies and domain superfolds. *Nature* 1994;372:631–4. <https://doi.org/10.1038/372631a0>.
- [154] Theobald DL, Wuttke DS. Divergent evolution within protein superfolds inferred from profile-based phylogenetics. *Journal of Molecular Biology* 2005;354:722–37. <https://doi.org/10.1016/j.jmb.2005.08.071>.
- [155] White LJ, Sutton G, Shechonge A, Day JJ, Dasmahapatra KK, Pownall ME. Adaptation of the carbamoyl-phosphate synthetase enzyme in an extremophile fish. *Royal Society Open Science* 2020;7:201200. <https://doi.org/10.1098/rsos.201200>.
- [156] Plewczynski D, Pas J, von Grotthuss M, Rychlewski L. Comparison of proteins based on segments structural similarity. *Acta Biochimica Polonica* 2004;51:161–72. https://doi.org/10.18388/abp.2004_3608.
- [157] Chang AB, Lin R, Studley WK, Tran C v., Saier, Jr MH, Saier MH. Phylogeny as a guide to structure and function of membrane transport proteins. *Molecular Membrane Biology* 2004;21:171–81. <https://doi.org/10.1080/09687680410001720830>.
- [158] Chang AB, Lin R, Studley WK, Tran C V., Saier MH. Phylogeny as a guide to structure and function of membrane transport proteins. *Molecular Membrane Biology* 2004;21:171–81. <https://doi.org/10.1080/09687680410001720830>.
- [159] Balaji S, Srinivasan N. Comparison of sequence-based and structure-based phylogenetic trees of homologous proteins: Inferences on protein evolution. *Journal of Biosciences* 2007;32:83–96. <https://doi.org/10.1007/s12038-007-0008-1>.
- [160] Lakshmi B, Mishra M, Srinivasan N, Archunan G. Structure-based phylogenetic analysis of the Lipocalin superfamily. *PLoS ONE* 2015;10:1–18. <https://doi.org/10.1371/journal.pone.0135507>.

- [161] van Loo B, Bayer CD, Fischer G, Jonas S, Valkov E, Mohamed MF, et al. Balancing Specificity and Promiscuity in Enzyme Evolution: Multidimensional Activity Transitions in the Alkaline Phosphatase Superfamily. *Journal of the American Chemical Society* 2019;141:370–87. <https://doi.org/10.1021/jacs.8b10290>.
- [162] Gonzalez JM. Structure-guided phylogenetic reconstruction of the superfamily of proteins sharing the metallo- β -lactamase fold. *BioRxiv* 2020:2020.04.16.045138. <https://doi.org/10.1101/2020.04.16.045138>.
- [163] Zalatan JG, Fenn TD, Herschlag D. Comparative Enzymology in the Alkaline Phosphatase Superfamily to Determine the Catalytic Role of an Active-Site Metal Ion. *Journal of Molecular Biology* 2008;384:1174–89. <https://doi.org/10.1016/j.jmb.2008.09.059>.
- [164] Wei W, Srinivas S, Lin J, Tang Z, Wang S, Ullah S, et al. Defining ICR-Mo, an intrinsic colistin resistance determinant from *Moraxella osloensis*. *PLOS Genetics* 2018;14:e1007389. <https://doi.org/10.1371/journal.pgen.1007389>.
- [165] Poirel L, Kieffer N, Fernandez-Garayzabal JF, Vela AI, Larpin Y, Nordmann P. MCR-2-mediated plasmid-borne polymyxin resistance most likely originates from *Moraxella pluranimalium*. *Journal of Antimicrobial Chemotherapy* 2017;72:2947–9. <https://doi.org/10.1093/jac/dkx225>.
- [166] Stogios PJ, Cox G, Zubyk HL, Evdokimova E, Wawrzak Z, Wright GD, et al. Substrate Recognition by a Colistin Resistance Enzyme from *Moraxella catarrhalis*. *ACS Chemical Biology* 2018;13:acschembio.8b00116. <https://doi.org/10.1021/acscchembio.8b00116>.
- [167] Zhang H, Wei W, Huang M, Umar Z, Feng Y. Definition of a Family of Nonmobile Colistin Resistance (NMCR-1) Determinants Suggests Aquatic Reservoirs for MCR-4. *Advanced Science* 2019;6. <https://doi.org/10.1002/adv.201900038>.
- [168] Telke AA, Rolain JM. Functional genomics to discover antibiotic resistance genes: The paradigm of resistance to colistin mediated by ethanolamine phosphotransferase in *Shewanella algae* MARS 14. *International Journal of Antimicrobial Agents* 2015;46:648–52. <https://doi.org/10.1016/j.ijantimicag.2015.09.001>.
- [169] Son SJ, Huang R, Squire CJ, Leung IKH. MCR-1: a promising target for structure-based design of inhibitors to tackle polymyxin resistance. *Drug Discovery Today* 2018;00:206–16. <https://doi.org/10.1016/j.drudis.2018.07.004>.
- [170] Wang X, Lu Q, Qi J, Chai Y, Wang Y, Gao GF. Structural and functional insights into MCR-2 mediated colistin resistance. *Science China Life Sciences* 2018;61:1432–6. <https://doi.org/10.1007/s11427-018-9363-4>.
- [171] Li H, Wang Y, Chen Q, Xia X, Shen J, Wang Y, et al. Identification of Functional Interactome of Colistin Resistance Protein MCR-1 in *Escherichia coli*. *Frontiers in Microbiology* 2021;11:3632. <https://doi.org/10.3389/fmicb.2020.583185>.
- [172] Lundin D, Poole AM, Sjöberg BM, Högbom M. Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *Journal of Biological Chemistry* 2012;287:20565–75. <https://doi.org/10.1074/jbc.M112.367458>.
- [173] Leman M. Lecture notes in artificial intelligence. vol. 1317. 1997.

- [174] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* 1995;247:536–40. [https://doi.org/10.1016/S0022-2836\(05\)80134-2](https://doi.org/10.1016/S0022-2836(05)80134-2).
- [175] Holm L, Kääriäinen S, Wilton C, Plewczynski D. Using Dali for Structural Comparison of Proteins. *Current Protocols in Bioinformatics* 2006;14:1–24. <https://doi.org/10.1002/0471250953.bi0505s14>.
- [176] Holm L, Laakso LM. Dali server update. *Nucleic Acids Research* 2016;44:W351–5. <https://doi.org/10.1093/nar/gkw357>.
- [177] Holm L, Kääriäinen S, Rosenström P, Schenkel A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* 2008;24:2780–1. <https://doi.org/10.1093/bioinformatics/btn507>.
- [178] Anderson AC, Burnett AJN, Hiscock L, Maly KE, Weadge JT. The Escherichia coli cellulose synthase subunit G (BcsG) is a Zn²⁺-dependent phosphoethanolamine transferase. *The Journal of Biological Chemistry* 2020;295:6225–35. <https://doi.org/10.1074/jbc.RA119.011668>.
- [179] Fan J, Petersen EM, Hinds TR, Zheng N, Miller SI. Structure of an inner membrane protein required for phopq-regulated increases in outer membrane cardiolipin. *MBio* 2020;11. <https://doi.org/10.1128/mBio.03277-19>.
- [180] Nishimasu H, Okudaira S, Hama K, Mihara E, Dohmae N, Inoue A, et al. Crystal structure of autotaxin and insight into GPCR activation by lipid mediators. *Nature Structural and Molecular Biology* 2011;18:205–13. <https://doi.org/10.1038/nsmb.1998>.
- [181] RL F, TJ R, JJ T. Structure of Francisella tularensis AcpA: prototype of a unique superfamily of acid phosphatases and phospholipases C. *The Journal of Biological Chemistry* 2006;281:30289–98. <https://doi.org/10.1074/JBC.M606391200>.
- [182] Fan J, Petersen EM, Hinds TR, Zheng N, Miller SI. Structure of an inner membrane protein required for phopq-regulated increases in outer membrane cardiolipin. *MBio* 2020;11. <https://doi.org/10.1128/mBio.03277-19>.
- [183] Clairfeuille T, Buchholz KR, Li Q, Verschueren E, Liu P, Sangaraju D, et al. Structure of the essential inner membrane lipopolysaccharide–PbgA complex. *Nature* 2020;584:479–83. <https://doi.org/10.1038/s41586-020-2597-x>.
- [184] Edgar RJ, van Hensbergen VP, Ruda A, Turner AG, Deng P, le Breton Y, et al. Discovery of glycerol phosphate modification on streptococcal rhamnose polysaccharides. *Nature Chemical Biology* 2019;15:463–71. <https://doi.org/10.1038/s41589-019-0251-4>.
- [185] Johnsen U, Schönheit AP. Characterization of cofactor-dependent and cofactor-independent phosphoglycerate mutases from Archaea. *Extremophiles* 2007;11:647–57. <https://doi.org/10.1007/s00792-007-0094-x>.
- [186] van Loo B, Jonas S, Babbie AC, Benjdia A, Berteau O, Hyvönen M, et al. An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:2740–5. <https://doi.org/10.1073/pnas.0903951107>.

- [187] Anderson AC, Burnett AJNN, Hiscock L, Maly KE, Weadge JT. The Escherichia coli cellulose synthase subunit G (BcsG) is a Zn²⁺-dependent phosphoethanolamine transferase. *Journal of Biological Chemistry* 2020;295:6225–35. <https://doi.org/10.1074/jbc.RA119.011668>.
- [188] Uhlhorn-Dierks G, Kolter T, Sandhoff K. How does nature cleave sulfuric acid esters? A novel posttranslational modification of sulfatases. *Angewandte Chemie - International Edition* 1998;37:2453–5. [https://doi.org/10.1002/\(SICI\)1521-3773\(19981002\)37:18<2453::AID-ANIE2453>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1521-3773(19981002)37:18<2453::AID-ANIE2453>3.0.CO;2-N).
- [189] de Backer MME, McSweeney S, Lindley PF, Hough E. Ligand-binding and metal-exchange crystallographic studies on shrimp alkaline phosphatase. *Acta Crystallographica Section D: Biological Crystallography* 2004;60:1555–61. <https://doi.org/10.1107/S0907444904015628>.
- [190] Ma L, Tibbitts TT, Kantrowitz ER. Escherichia coli alkaline phosphatase: X-ray structural studies of a mutant enzyme (His-412 → Asn) at one of the catalytically important zinc binding sites. *Protein Science* 1995;4:1498–506. <https://doi.org/10.1002/pro.5560040807>.
- [191] Holtz KM, Stec B, Kantrowitz ER. A model of the transition state in the alkaline phosphatase reaction. *Journal of Biological Chemistry* 1999;274:8351–4. <https://doi.org/10.1074/jbc.274.13.8351>.
- [192] Hung H-C. Differentiation of the slow-binding mechanism for magnesium ion activation and zinc ion inhibition of human placental alkaline phosphatase. *Protein Science* 2001;10:34–45. <https://doi.org/10.1110/ps.35201>.
- [193] Wei P, Song G, Shi M, Zhou Y, Liu Y, Lei J, et al. Substrate analog interaction with MCR-1 offers insight into the rising threat of the plasmid-mediated transferable colistin resistance. *FASEB Journal* 2018;32:1085–98. <https://doi.org/10.1096/fj.201700705R>.
- [194] Chen AY, Adamek RN, Dick BL, Credille C v., Morrison CN, Cohen SM. [ASAP] Targeting Metalloenzymes for Therapeutic Intervention. *Chemical Reviews* 2018;119:acs.chemrev.8b00201. <https://doi.org/10.1021/acs.chemrev.8b00201>.
- [195] Kato K, Nishimasu H, Oikawa D, Hirano S, Hirano H, Kasuya G, et al. Structural insights into cGAMP degradation by Ecto-nucleotide pyrophosphatase phosphodiesterase 1. *Nature Communications* 2018;9:4424. <https://doi.org/10.1038/s41467-018-06922-7>.
- [196] Huang J, Zhu Y, Han M-LL, Li M, Song J, Velkov T, et al. Comparative analysis of phosphoethanolamine transferases involved in polymyxin resistance across 10 clinically relevant Gram-negative bacteria. *International Journal of Antimicrobial Agents* 2018;51:586–93. <https://doi.org/10.1016/J.IJANTIMICAG.2017.12.016>.
- [197] Gabale U, Peña Palomino PA, Kim HA, Chen W, Ressler S. The essential inner membrane protein YejM is a metalloenzyme. *Scientific Reports* 2020;10:17794. <https://doi.org/10.1038/s41598-020-73660-6>.
- [198] Lu D, Wörmann ME, Zhang X, Schneewind O, Gründling A, Freemont PS. Structure-based mechanism of lipoteichoic acid synthesis by Staphylococcus aureus LtaS. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:1584–9. <https://doi.org/10.1073/pnas.0809020106>.

- [199] Gao R, Hu Y, Li Z, Sun J, Wang Q, Lin J, et al. Dissemination and Mechanism for the MCR-1 Colistin Resistance. *PLoS Pathogens* 2016;12:e1005957. <https://doi.org/10.1371/journal.ppat.1005957>.
- [200] Zhang H, Hou M, Xu Y, Srinivas S, Huang M, Liu L, et al. Action and mechanism of the colistin resistance enzyme MCR-4. *Communications Biology* 2019;2:36. <https://doi.org/10.1038/s42003-018-0278-1>.
- [201] Suardíaz R, Lythell E, Hinchliffe P, van der Kamp M, Spencer J, Fey N, et al. Catalytic mechanism of the colistin resistance protein MCR-1. *Organic & Biomolecular Chemistry* 2021;19:3813–9. <https://doi.org/10.1039/d0ob02566f>.
- [202] Klein G, Müller-Loennies S, Lindner B, Kobylak N, Brade H, Raina S. Molecular and structural basis of inner core lipopolysaccharide alterations in *Escherichia coli*: incorporation of glucuronic acid and phosphoethanolamine in the heptose region. *The Journal of Biological Chemistry* 2013;288:8111–27. <https://doi.org/10.1074/jbc.M112.445981>.
- [203] Cullen TW, O'Brien JP, Hendrixson DR, Giles DK, Hobb RI, Thompson SA, et al. EptC of *Campylobacter jejuni* Mediates Phenotypes Involved in Host Interactions and Virulence. *Infection and Immunity* 2013;81:430–40. <https://doi.org/10.1128/IAI.01046-12>.
- [204] Russell RB. Detection of protein three-dimensional side-chain patterns: New examples of convergent evolution. *Journal of Molecular Biology* 1998;279:1211–27. <https://doi.org/10.1006/jmbi.1998.1844>.
- [205] Kwon S-K, Kim SK, Lee D-H, Kim JF. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Scientific Reports* 2015;5:16076. <https://doi.org/10.1038/srep16076>.
- [206] Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, et al. Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences* 2008;105:14371–6. <https://doi.org/10.1073/pnas.0804090105>.
- [207] Schlegel S, Rujas E, Ytterberg AJ, Zubarev RA, Luirink J, de Gier J-W, et al. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microbial Cell Factories* 2013;12:24. <https://doi.org/10.1186/1475-2859-12-24>.
- [208] Terpe K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* 2006;60:523–33. <https://doi.org/10.1007/s00253-002-1158-6>.
- [209] Demain AL, Vaishnav P. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances* 2009;27:297–306. <https://doi.org/10.1016/j.biotechadv.2009.01.008>.
- [210] Studier FW, Moffatt BA. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 1986;189:113–30. [https://doi.org/10.1016/0022-2836\(86\)90385-2](https://doi.org/10.1016/0022-2836(86)90385-2).
- [211] Miroux B, Walker JE. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.

- [212] Schlegel S, Hjelm A, Baumgarten T, Vikström D, de Gier J-W. Bacterial-based membrane protein production. *Biochimica et Biophysica Acta* 2014;1843:1739–49. <https://doi.org/10.1016/j.bbamcr.2013.10.023>.
- [213] Zhang J, Greasham R. Chemically defined media for commercial fermentations. *Applied Microbiology and Biotechnology* 1999;51:407–21. <https://doi.org/10.1007/s002530051411>.
- [214] Seddon AM, Curnow P, Booth PJ. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochimica et Biophysica Acta - Biomembranes* 2004;1666:105–17. <https://doi.org/10.1016/j.bbamem.2004.04.011>.
- [215] Burgess RR. A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting. *Protein Expression and Purification* 2018;150:81–5. <https://doi.org/10.1016/j.pep.2018.05.007>.
- [216] Oliva A, Llabrés M, Fariña JB. Comparative study of protein molecular weights by size-exclusion chromatography and laser-light scattering. *Journal of Pharmaceutical and Biomedical Analysis* 2001;25:833–41. [https://doi.org/10.1016/S0731-7085\(01\)00359-4](https://doi.org/10.1016/S0731-7085(01)00359-4).
- [217] Some D, Amartely H, Tsadok A, Lebendiker M. Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *Journal of Visualized Experiments* 2019. <https://doi.org/10.3791/59615>.
- [218] Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* 2007;1:2527–35. <https://doi.org/10.1038/nprot.2006.204>.
- [219] Rosano GLGL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology* 2014;5:172. <https://doi.org/10.3389/fmicb.2014.00172>.
- [220] Hannig G, Makrides SC. Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* 1998;16:54–60. [https://doi.org/10.1016/S0167-7799\(97\)01155-4](https://doi.org/10.1016/S0167-7799(97)01155-4).
- [221] Schlegel S, Löfblom J, Lee C, Hjelm A, Klepsch M, Strous M, et al. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *Journal of Molecular Biology* 2012;423:648–59. <https://doi.org/10.1016/j.jmb.2012.07.019>.
- [222] Gordon E, Horsefield R, Swarts HGP, de Pont JJHM, Neutze R, Snijder A. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expression and Purification* 2008;62:1–8. <https://doi.org/10.1016/j.pep.2008.07.005>.
- [223] Studier FFW. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 2005;41:207–34. <https://doi.org/10.1016/j.pep.2005.01.016>.
- [224] Studier FW. *Structural Genomics*. vol. 1091. Totowa, NJ: Humana Press; 2014. <https://doi.org/10.1007/978-1-62703-691-7>.

- [225] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. The Proteomics Protocols Handbook. 2005. <https://doi.org/10.1385/1592598900>.
- [226] Sivashanmugam A, Murray V, Cui C, Zhang Y, Wang J, Li Q. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Science: A Publication of the Protein Society* 2009;18:936–48. <https://doi.org/10.1002/pro.102>.
- [227] Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification* 2004;37:203–6. <https://doi.org/10.1016/j.pep.2004.04.025>.
- [228] Arechaga I, Miroux B, Karrasch S, Huijbregts R, De Kruijff B, Runswick MJ, et al. Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F1F(o) ATP synthase. *FEBS Letters* 2000;482:215–9. [https://doi.org/10.1016/S0014-5793\(00\)02054-8](https://doi.org/10.1016/S0014-5793(00)02054-8).
- [229] Wagner S, Bader ML, Drew D, de Gier J-W. Rationalizing membrane protein overexpression. *Trends in Biotechnology* 2006;24:364–71. <https://doi.org/10.1016/j.tibtech.2006.06.008>.
- [230] Mus-Veteau I. Membrane proteins production for structural analysis. *Membrane Proteins Production for Structural Analysis* 2014:1–425. <https://doi.org/10.1007/978-1-4939-0662-8>.
- [231] Kwan TOC, Reis R, Siligardi G, Hussain R, Cheruvara H, Moraes I. Selection of Biophysical Methods for Characterisation of Membrane Proteins. *International Journal of Molecular Sciences* 2019;20. <https://doi.org/10.3390/ijms20102605>.
- [232] Some D, Amartely H, Tsadok A, Lebendiker M. Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (Sec-mals). *Journal of Visualized Experiments* 2019;2019:e59615. <https://doi.org/10.3791/59615>.
- [233] Ogawa T, Hirokawa N. Multiple analyses of protein dynamics in solution. *Biophysical Reviews* 2018;10:299–306. <https://doi.org/10.1007/s12551-017-0354-7>.
- [234] Some D, Amartely H, Tsadok A, Lebendiker M. Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (Sec-mals). *Journal of Visualized Experiments* 2019;2019:e59615. <https://doi.org/10.3791/59615>.
- [235] Wyatt PJ. Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* 1993;272:1–40. [https://doi.org/10.1016/0003-2670\(93\)80373-S](https://doi.org/10.1016/0003-2670(93)80373-S).
- [236] Slotboom DJ, Durkens RH, Olieman K, Erkens GB. Static light scattering to characterize membrane proteins in detergent solution. *Methods* 2008;46:73–82. <https://doi.org/10.1016/j.ymeth.2008.06.012>.
- [237] Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 2007;1:2876–90. <https://doi.org/10.1038/nprot.2006.202>.
- [238] Corrêa D, Ramos C. The use of circular dichroism spectroscopy to study protein folding, form and function. *African J Biochem Res* 2009;3:164–73.

- [239] Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function and Bioinformatics* 2012;80:374–81. <https://doi.org/10.1002/prot.23188>.
- [240] Tusnady GE, Dosztanyi Z, Simon I. Transmembrane proteins in the Protein Data Bank: identification and classification. *Bioinformatics* 2004;20:2964–72. <https://doi.org/10.1093/bioinformatics/bth340>.
- [241] Xu Y, Wei W, Lei S, Lin J, Srinivas S, Feng Y. An evolutionarily conserved mechanism for intrinsic and transferable polymyxin resistance. *MBio* 2018;9:1–18. <https://doi.org/10.1128/mBio.02317-17>.
- [242] Xu Y, Lin J, Cui T, Srinivas S, Feng Y. Mechanistic insights into transferable polymyxin resistance among gut bacteria. *Journal of Biological Chemistry* 2018;293:4350–65. <https://doi.org/10.1074/jbc.RA117.000924>.
- [243] Tripathi N. High Yield Production of Heterologous Proteins with *Escherichia coli*. *Defence Science Journal* 2009;59:137–46. <https://doi.org/10.14429/dsj.59.1501>.
- [244] Zhang H, Hou M, Xu Y, Srinivas S, Huang M, Liu L, et al. Action and mechanism of the colistin resistance enzyme MCR-4. *Communications Biology* 2019;2. <https://doi.org/10.1038/s42003-018-0278-1>.
- [245] Hu M, Guo J, Cheng Q, Yang Z, Chan EWC, Chen S, et al. Crystal Structure of *Escherichia coli* originated MCR-1, a phosphoethanolamine transferase for Colistin Resistance. *Scientific Reports* 2016;6:38793. <https://doi.org/10.1038/srep38793>.
- [246] Li H, Yang L, Liu Z, Yin W, Liu D, Shen Y, et al. Molecular insights into functional differences between mcr-3- and mcr-1-mediated colistin resistance. *Antimicrobial Agents and Chemotherapy* 2018;62. <https://doi.org/10.1128/AAC.00366-18>.
- [247] Raynal B, Lenormand P, Baron B, Hoos S, England P. Quality assessment and optimization of purified protein samples: Why and how? *Microbial Cell Factories* 2010;13. <https://doi.org/10.1186/s12934-014-0180-6>.
- [248] Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *The Lancet Infectious Diseases* 2016;16:161–8. [https://doi.org/10.1016/S1473-3099\(15\)00424-7](https://doi.org/10.1016/S1473-3099(15)00424-7).
- [249] Nordmann P, Jayol AA, Poirel L. A universal culture medium for screening polymyxin-resistant gram-negative isolates. *Journal of Clinical Microbiology* 2016;54:1395–9. <https://doi.org/10.1128/JCM.00446-16>.
- [250] Humphries RM. Susceptibility testing of the polymyxins: Where are we now? *Pharmacotherapy* 2015;35:22–7. <https://doi.org/10.1002/phar.1505>.
- [251] Escalante EG, Condor KY, di Conza JA, Gutkind GO. Phenotypic detection of plasmid-mediated colistin resistance in Enterobacteriaceae. *Journal of Clinical Microbiology* 2020;58. <https://doi.org/10.1128/JCM.01555-19>.
- [252] Furniss CRD, Kostrzewa M, Mavridou DAI, Larrouy-Maumus G. The clue is in the lipid A: Rapid detection of colistin resistance. *PLoS Pathogens* 2020;16:e1008331. <https://doi.org/10.1371/journal.ppat.1008331>.

- [253] Reynolds CM, Kalb SR, Cotter RJ, Raetz CRH. A phosphoethanolamine transferase specific for the outer 3-deoxy-D-manno-octulosonic acid residue of *Escherichia coli* lipopolysaccharide: Identification of the *eptB* gene and Ca²⁺ hypersensitivity of an *eptB* deletion mutant. *Journal of Biological Chemistry* 2005;280:21202–11. <https://doi.org/10.1074/jbc.M500964200>.
- [254] Nowicki EM, O'Brien JP, Brodbelt JS, Trent MS. Extracellular zinc induces phosphoethanolamine addition to *Pseudomonas aeruginosa* lipid A via the ColRS two-component system. *Molecular Microbiology* 2015;97:166–78. <https://doi.org/10.1111/mmi.13018>.
- [255] Cullen TW, Trent MS. A link between the assembly of flagella and lipooligosaccharide of the gram-negative bacterium *Campylobacter jejuni*. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107:5160–5. <https://doi.org/10.1073/pnas.0913451107>.
- [256] Anderson AC, Burnett AJN, Hiscock L, Maly KE, Weadge JT. The *Escherichia coli* cellulose synthase subunit G (BcsG) is a Zn²⁺-dependent phosphoethanolamine transferase. *Journal of Biological Chemistry* 2020;295:6225–35. <https://doi.org/10.1074/jbc.RA119.011668>.
- [257] Bisswanger H. Enzyme assays. *Perspectives in Science* 2014;1:41–55. <https://doi.org/10.1016/j.pisc.2014.02.005>.
- [258] Dimitrijević M, Grujić-injac B, Lajšić S. The Synthesis of a New Phospholipid from the Koilin Glandular Layer of Chicken Gizzard. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie* 1979;360:477–80. <https://doi.org/10.1515/bchm2.1979.360.1.477>.
- [259] Hernández I, Fernández JA, Niell FX. A comparative study of alkaline phosphatase activity in two species of *Gelidium* (Gelidiales, Rhodophyta). *European Journal of Phycology* 1995;30:69–77. <https://doi.org/10.1080/09670269500650811>.
- [260] Wachino J ichi, Kanechi R, Nishino E, Mochizuki M, Jin W, Kimura K, et al. 4-Amino-2-sulfanylbenzoic acid as a potent subclass B3 metallo-β-lactamase-specific inhibitor applicable for distinguishing metallo-β-lactamase subclasses. *Antimicrobial Agents and Chemotherapy* 2019;63. <https://doi.org/10.1128/AAC.01197-19>.
- [261] Williams PAM, Barrio DA, Etcheverry SB. Interactions of vanadyl(IV) with dithiothreitol and thioglycolic acid. Their action on alkaline phosphatase activity. *Journal of Inorganic Biochemistry* 1999;75:99–104. [https://doi.org/10.1016/S0162-0134\(99\)00039-2](https://doi.org/10.1016/S0162-0134(99)00039-2).
- [262] Diehl Harvey. The chelate rings. *Chemical Reviews* 1937;21:39–111. <https://doi.org/10.1021/cr60068a003>.
- [263] Esposito F, Fernandes MR, Lopes R, Muñoz M, Sabino CP, Cunha MP, et al. Detection of colistin-resistant *mcr-1*-positive *Escherichia coli* by use of assays based on inhibition by EDTA and zeta potential. *Journal of Clinical Microbiology* 2017;55:3454–65. <https://doi.org/10.1128/JCM.00835-17>.
- [264] Dean RL. Kinetic studies with alkaline phosphatase in the presence and absence of inhibitors and divalent cations. *Biochemistry and Molecular Biology Education* 2002;30:401–7. <https://doi.org/10.1002/bmb.2002.494030060138>.

- [265] Anderson AC, Burnett AJN, Hiscock L, Maly KE, Weadge JT. The Escherichia coli cellulose synthase subunit G (BcsG) is a Zn²⁺-dependent phosphoethanolamine transferase. *Journal of Biological Chemistry* 2020;295:6225–35. <https://doi.org/10.1074/jbc.RA119.011668>.
- [266] Fage CD, Brown DB, Boll JM, Keatinge-Clay AT, Trent MS. Crystallographic study of the phosphoethanolamine transferase EptC required for polymyxin resistance and motility in *Campylobacter jejuni*. *Acta Crystallographica Section D Biological Crystallography* 2014;70:2730–9. <https://doi.org/10.1107/S1399004714017623>.
- [267] Pendleton JN, Gorman SP, Gilmore BF. Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-Infective Therapy* 2013;11:297–308. <https://doi.org/10.1586/eri.13.12>.
- [268] Peleg AY, Seifert H, Paterson DL. *Acinetobacter baumannii*: Emergence of a successful pathogen. vol. 21. *American Society for Microbiology Journals*; 2008. <https://doi.org/10.1128/CMR.00058-07>.
- [269] Glew RH, Moellering RC, Kunz LJ. Infections with acinetobacter calcoaceticus (herellea vaginicola): Clinical and laboratory studies. *Medicine (United States)* 1977;56:79–97. <https://doi.org/10.1097/00005792-197703000-00001>.
- [270] Bouvet PJM, Grimont PAD. Taxonomy of the genus *Acinetobacter* with the recognition of *Acinetobacter baumannii* sp. nov., *Acinetobacter haemolyticus* sp. nov., *Acinetobacter johnsonii* sp. nov., and *Acinetobacter junii* sp. nov. and emended descriptions of *Acinetobacter calcoaceticus* a. *International Journal of Systematic Bacteriology* 1986;36:228–40. <https://doi.org/10.1099/00207713-36-2-228>.
- [271] Bernardts AT, Dijkshoorn L, van der Toorn J, Bochner BR, van Boven CPAA. Phenotypic characterisation of *Acinetobacter* strains of 13 DNA-DNA hybridisation groups by means of the Biolog system. *Journal of Medical Microbiology* 1995;42:113–9. <https://doi.org/10.1099/00222615-42-2-113>.
- [272] Maragakis LL, Perl TM. *Acinetobacter baumannii*: Epidemiology, antimicrobial resistance, and treatment options. *Clinical Infectious Diseases* 2008;46:1254–63. <https://doi.org/10.1086/529198>.
- [273] Shlaes DM, Bradford PA. Antibiotics—From There to Where? *Pathogens and Immunity* 2018;3:19. <https://doi.org/10.20411/pai.v3i1.231>.
- [274] Services USD of H and H. Antibiotic resistance threats in the United States. *Centers for Disease Control and Prevention* 2019:1–113. <https://doi.org/10.15620/cdc:82532>.
- [275] Nguyen M, Joshi SG. Carbapenem resistance in *Acinetobacter baumannii*, and their importance in hospital-acquired infections: a scientific review. *Journal of Applied Microbiology* 2021;jam.15130. <https://doi.org/10.1111/jam.15130>.
- [276] Morris FC, Dexter C, Kostoulias X, Uddin MI, Peleg AY. The Mechanisms of Disease Caused by *Acinetobacter baumannii*. *Frontiers in Microbiology* 2019;10:1601. <https://doi.org/10.3389/fmicb.2019.01601>.
- [277] Anstey NM, Currie BJ, Hassell M, Palmer D, Dwyer B, Seifert H. Community-acquired bacteremic *Acinetobacter pneumonia* in tropical Australia is caused by diverse strains of *Acinetobacter baumannii*, with carriage in the throat in at-risk groups. *Journal of Clinical Microbiology* 2002;40:685–6. <https://doi.org/10.1128/JCM.40.2.685-686.2002>.

- [278] Anstey NM, Currie BJ, Withnall KM. Community-acquired acinetobacter pneumonia in the Northern Territory of Australia. *Clinical Infectious Diseases* 1992;14:83–91. <https://doi.org/10.1093/clinids/14.1.83>.
- [279] Leung WS, Chu CM, Tsang KY, Lo FH, Lo KF, Ho PL. Fulminant community-acquired *Acinetobacter baumannii* pneumonia as a distinct clinical syndrome. *Chest* 2006;129:102–9. <https://doi.org/10.1378/chest.129.1.102>.
- [280] Lee CR, Lee JH, Park M, Park KS, Bae IK, Kim YB, et al. Biology of *Acinetobacter baumannii*: Pathogenesis, antibiotic resistance mechanisms, and prospective treatment options. *Frontiers in Cellular and Infection Microbiology* 2017;7:55. <https://doi.org/10.3389/fcimb.2017.00055>.
- [281] Jones A, Morgan D, Walsh A, Turton J, Livermore D, Pitt T, et al. Importation of multidrug-resistant *Acinetobacter* spp infections with casualties from Iraq. *The Lancet Infectious Diseases* 2006;6:317–8. [https://doi.org/10.1016/S1473-3099\(06\)70471-6](https://doi.org/10.1016/S1473-3099(06)70471-6).
- [282] Coelho JM, Turton JF, Kaufmann ME, Glover J, Woodford N, Warner M, et al. Occurrence of carbapenem-resistant *Acinetobacter baumannii* clones at multiple hospitals in London and Southeast England. *Journal of Clinical Microbiology* 2006;44:3623–7. <https://doi.org/10.1128/JCM.00699-06>.
- [283] Walsh TR, Toleman MA, Poirel L, Nordmann P. Metallo-Lactamases: the Quiet before the Storm? *CLINICAL MICROBIOLOGY REVIEWS* 2005;18:306–25. <https://doi.org/10.1128/CMR.18.2.306-325.2005>.
- [284] Mosqueda N, Espinal P, Cosgaya C, Viota S, Plasencia V, Alvarez-Lerma F, et al. Globally expanding carbapenemase finally appears in Spain: nosocomial outbreak of *acinetobacter baumannii* producing plasmid-encoded OXA-23 in Barcelona, Spain. *Antimicrobial Agents and Chemotherapy* 2013;57:5155–7. <https://doi.org/10.1128/AAC.01486-13>.
- [285] Alcántar-Curiel MD, Rosales-Reyes R, Jarillo-Quijada MD, Gayosso-Vázquez C, Fernández-Vázquez JL, Toledano-Tableros JE, et al. Carbapenem-Resistant *Acinetobacter baumannii* in Three Tertiary Care Hospitals in Mexico: Virulence Profiles, Innate Immune Response and Clonal Dissemination. *Frontiers in Microbiology* 2019;10:2116. <https://doi.org/10.3389/fmicb.2019.02116>.
- [286] Morfin-Otero R, Alcántar-Curiel MD, Rocha MJ, Alpuche-Aranda CM, Santos-Preciado JI, Gayosso-Vázquez C, et al. *Acinetobacter baumannii* infections in a tertiary care Hospital in Mexico over the past 13 years. *Chemotherapy* 2013;59:57–65. <https://doi.org/10.1159/000351098>.
- [287] Cai Y, Chai D, Wang R, Liang B, Bai N. Colistin resistance of *Acinetobacter baumannii*: Clinical reports, mechanisms and antimicrobial strategies. *Journal of Antimicrobial Chemotherapy* 2012;67:1607–15. <https://doi.org/10.1093/jac/dks084>.
- [288] Andersson DI, Nicoloff H, Hjort K. Mechanisms and clinical relevance of bacterial heteroresistance. *Nature Reviews Microbiology* 2019;17:479–96. <https://doi.org/10.1038/s41579-019-0218-1>.
- [289] Haeili M, Kafshdouz M, Feizabadi MM. Molecular mechanisms of colistin resistance among pandrug-resistant isolates of *acinetobacter baumannii* with high case-fatality rate in intensive care unit patients. *Microbial Drug Resistance* 2018;24:1271–6. <https://doi.org/10.1089/mdr.2017.0397>.

- [290] Trebosc V, Gartenmann S, Tötzl M, Lucchini V, Schellhorn B, Pieren M, et al. Dissecting Colistin Resistance Mechanisms in Extensively Drug-Resistant *Acinetobacter baumannii* Clinical Isolates. *MBio* 2019;10. <https://doi.org/10.1128/mBio.01083-19>.
- [291] Machado D, Antunes J, Simões A, Perdigão J, Couto I, McCusker M, et al. Contribution of efflux to colistin heteroresistance in a multidrug resistant *Acinetobacter baumannii* clinical isolate. *Journal of Medical Microbiology* 2018;67:740–9. <https://doi.org/10.1099/jmm.0.000741>.
- [292] Hawley JS, Murray CK, Jorgensen JH. Colistin heteroresistance in *Acinetobacter* and its association with previous colistin therapy. *Antimicrobial Agents and Chemotherapy* 2008;52:351–2. <https://doi.org/10.1128/AAC.00766-07>.
- [293] Li J, Rayner CR, Nation RL, Owen RJ, Spelman D, Tan KE, et al. Heteroresistance to colistin in multidrug-resistant *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy* 2006;50:2946–50. <https://doi.org/10.1128/AAC.00103-06>.
- [294] Bakour S, Olaitan AO, Ammari H, Touati A, Saoudi S, Saoudi K, et al. Emergence of Colistin- and Carbapenem-Resistant *Acinetobacter baumannii* ST2 Clinical Isolate in Algeria: First Case Report. *Microbial Drug Resistance* 2015;21:279–85. <https://doi.org/10.1089/mdr.2014.0214>.
- [295] Herrera ME, Mobilia LN, Posse GR. Comparative evaluation of the sensitivity of *Acinetobacter* to colistin, using the prediffusion and minimum inhibitory concentration methods. detection of heteroresistant isolates. *Revista Argentina de Microbiología* 2011;43:115–9. <https://doi.org/10.1590/S0325-75412011000200009>.
- [296] Leite GC, Oliveira MS, Perdigão-Neto LV, Rocha CKD, Guimarães T, Rizek C, et al. Antimicrobial combinations against pan-resistant *Acinetobacter baumannii* isolates with different resistance mechanisms. *PLoS ONE* 2016;11. <https://doi.org/10.1371/journal.pone.0151270>.
- [297] Al-Agamy MH, Khalaf NG, Tawfick MM, Shibl AM, el Kholy A. Molecular characterization of carbapenem-insensitive *Acinetobacter baumannii* in Egypt. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 2014;22:49–54. <https://doi.org/10.1016/j.ijid.2013.12.004>.
- [298] Rolain JM, Roch A, Castanier M, Papazian L, Raoult D. *Acinetobacter baumannii* resistant to colistin with impaired virulence: A case report from France. *Journal of Infectious Diseases* 2011;204:1146–7. <https://doi.org/10.1093/infdis/jir475>.
- [299] Göttig S, Gruber TM, Higgins PG, Wachsmuth M, Seifert H, Kempf VAJ. Detection of pan drug-resistant *Acinetobacter baumannii* in Germany. *Journal of Antimicrobial Chemotherapy* 2014;69:2578–9. <https://doi.org/10.1093/jac/dku170>.
- [300] Oikonomou O, Sarrou S, Papagiannitsis CC, Georgiadou S, Mantzaris K, Zakyntinos E, et al. Rapid dissemination of colistin and carbapenem resistant *Acinetobacter baumannii* in Central Greece: Mechanisms of resistance, molecular identification and epidemiological data. *BMC Infectious Diseases* 2015;15:559. <https://doi.org/10.1186/s12879-015-1297-x>.
- [301] Taneja N, Singh G, Singh M, Sharma M. Emergence of tigecycline & colistin resistant *Acinetobacter baumannii* in patients with complicated urinary tract infections in north India. *The Indian Journal of Medical Research* 2011;133:681–4.

- [302] Modarresi F, Azizi O, Shakibaie MR, Motamedifar M, Valibeigi B, Mansouri S. Effect of iron on expression of efflux pump (adeABC) and quorum sensing (luxI, luxR) genes in clinical isolates of *Acinetobacter baumannii*. *APMIS* 2015;123:959–68. <https://doi.org/10.1111/apm.12455>.
- [303] Agodi A, Voulgari E, Barchitta M, Quattrocchi A, Bellocchi P, Poulou A, et al. Spread of a carbapenem- and colistin-resistant *Acinetobacter baumannii* ST2 clonal strain causing outbreaks in two Sicilian hospitals. *Journal of Hospital Infection* 2014;86:260–6. <https://doi.org/10.1016/j.jhin.2014.02.001>.
- [304] Tojo M, Mawatari M, Hayakawa K, Nagamatsu M, Shimada K, Mezaki K, et al. Multidrug-resistant *Acinetobacter baumannii* isolated from a traveler returned from Brunei. *Journal of Infection and Chemotherapy : Official Journal of the Japan Society of Chemotherapy* 2015;21:212–4. <https://doi.org/10.1016/j.jiac.2014.08.029>.
- [305] Al-Sweih NA, Al-Hubail MA, Rotimi VO. Isolated from Patients in Kuwait Hospitals. *Journal of Chemotherapy* 2013;23:13–6. <https://doi.org/10.1179/joc.2011.23.1.13>.
- [306] Quinteira S, Grosso F, Ramos H, Peixe L. Molecular epidemiology of imipenem-resistant *Acinetobacter haemolyticus* and *Acinetobacter baumannii* isolates carrying plasmid-mediated OXA-40 from a Portuguese hospital [6]. *Antimicrobial Agents and Chemotherapy* 2007;51:3465–6. <https://doi.org/10.1128/AAC.00267-07>.
- [307] Moisoiu A, Ionitã M, Sârbu L, Stoica C, Grigoriu L. [Antibiotic resistance of *Acinetobacter baumannii* strains isolated from clinical specimens in the "Marius Nasta" Pneumology Institute, Bucharest]. *Pneumologia (Bucharest, Romania)* 2014;63:109–11.
- [308] Baadani AM, Thawadi SI, El-Khizzi NA, Omrani AS. Prevalence of colistin and tigecycline resistance in *Acinetobacter baumannii* clinical isolates from 2 hospitals in Riyadh Region over a 2-year period. *Saudi Medical Journal* 2013;34:248–353.
- [309] Lee SY, Shin JH, Park KH, Kim JH, Shin MG, Suh SP, et al. Identification, genotypic relation, and clinical features of colistin-resistant isolates of *Acinetobacter* genomic species 13BJ/14TU from bloodstreams of patients in a university hospital. *Journal of Clinical Microbiology* 2014;52:931–9. <https://doi.org/10.1128/JCM.02868-13>.
- [310] López-Rojas R, Jiménez-Mejías ME, Lepe JA, Pachón J. *Acinetobacter baumannii* resistant to colistin alters its antibiotic resistance profile: A case report from Spain. *Journal of Infectious Diseases* 2011;204:1147–8. <https://doi.org/10.1093/infdis/jir476>.
- [311] Chang K-C, Lin M-F, Lin N-T, Wu W-J, Kuo H-Y, Lin T-Y, et al. Clonal spread of multidrug-resistant *Acinetobacter baumannii* in eastern Taiwan. *Journal of Microbiology, Immunology and Infection* 2012;45:37–42. <https://doi.org/10.1016/J.JMII.2011.09.019>.
- [312] Jaidane N, Chaouech C, Messaoudi A, Boujaafar N, Bouallegue O. Colistin-resistant *Acinetobacter baumannii*: A case report and literature review. *Reviews in Medical Microbiology* 2015;26:78–83. <https://doi.org/10.1097/MRM.000000000000029>.
- [313] Cikman A, Gulhan B, Aydin M, Ceylan MR, Parlak M, Karakecili F, et al. In vitro activity of colistin in combination with tigecycline against carbapenem-resistant *Acinetobacter baumannii* strains isolated from patients with ventilator-associated pneumonia. *International Journal of Medical Sciences* 2015;12:695–700. <https://doi.org/10.7150/ijms.11988>.

- [314] Qureshi ZA, Hittle LE, O'Hara JA, Rivera JI, Syed A, Shields RK, et al. Colistin-resistant acinetobacter baumannii: Beyond carbapenem resistance. *Clinical Infectious Diseases* 2015;60:1295–303. <https://doi.org/10.1093/cid/civ048>.
- [315] Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* 2018;8. <https://doi.org/10.1128/ecosalplus.esp-0001-2018>.
- [316] Moffatt JH, Harper M, Harrison P, Hale JDFF, Vinogradov E, Seemann T, et al. Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. *Antimicrobial Agents and Chemotherapy* 2010;54:4971–7. <https://doi.org/10.1128/AAC.00834-10>.
- [317] Li J, Nation RL, Owen RJ, Wong S, Spelman O, Franklin C. Antibigrams of multidrug-resistant clinical *Acinetobacter baumannii*: Promising therapeutic options for treatment of infection with colistin-resistant strains. *Clinical Infectious Diseases* 2007;45:594–8. <https://doi.org/10.1086/520658>.
- [318] Da Silva GJ, Domingues S. Interplay between colistin resistance, virulence and fitness in acinetobacter baumannii. *Antibiotics* 2017;6:1–11. <https://doi.org/10.3390/antibiotics6040028>.
- [319] Nhu NTK, Riordan DW, Nhu TDH, Thanh DP, Thwaites G, Lan NPH, et al. The induction and identification of novel Colistin resistance mutations in *Acinetobacter baumannii* and their implications. *Scientific Reports* 2016;6:28291. <https://doi.org/10.1038/srep28291>.
- [320] Arroyo LA, Herrera CM, Fernandez L, Hankins J v, Trent MS, Hancock REW. The pmrCAB operon mediates polymyxin resistance in *Acinetobacter baumannii* ATCC 17978 and clinical isolates through phosphoethanolamine modification of lipid A. *Antimicrobial Agents and Chemotherapy* 2011;55:3743–51. <https://doi.org/10.1128/AAC.00256-11>.
- [321] Lee H, Hsu FF, Turk J, Groisman EA. The PmrA-regulated pmrC gene mediates phosphoethanolamine modification of lipid A and polymyxin resistance in *Salmonella enterica*. *Journal of Bacteriology* 2004;186:4124–33. <https://doi.org/10.1128/JB.186.13.4124-4133.2004>.
- [322] Nurtop E, Bayındır Bilman F, Menekse S, Kurt Azap O, Gönen M, Ergonul O, et al. Promoters of Colistin Resistance in *Acinetobacter baumannii* Infections. *Microbial Drug Resistance* 2019;25:997–1002. <https://doi.org/10.1089/mdr.2018.0396>.
- [323] Quan J, Li X, Chen Y, Jiang Y, Zhou Z, Zhang H, et al. Prevalence of mcr-1 in *Escherichia coli* and *Klebsiella pneumoniae* recovered from bloodstream infections in China: a multicentre longitudinal study. *The Lancet Infectious Diseases* 2017;17:400–10. [https://doi.org/10.1016/S1473-3099\(16\)30528-X](https://doi.org/10.1016/S1473-3099(16)30528-X).
- [324] Ma F, Shen C, Zheng X, Liu Y, Chen H, Zhong L, et al. Identification of a novel plasmid carrying mcr-4.3 in an acinetobacter baumannii strain in China. *Antimicrobial Agents and Chemotherapy* 2019;63. <https://doi.org/10.1128/AAC.00133-19>.
- [325] Martins-Sorenson N, Snesrud E, Xavier DE, Cacci LC, Iavarone AT, McGann P, et al. A novel plasmid-encoded mcr-4.3 gene in a colistin-resistant *Acinetobacter baumannii* clinical strain. *Journal of Antimicrobial Chemotherapy* 2020;75:60–4. <https://doi.org/10.1093/jac/dkz413>.

- [326] Hameed F, Khan MA, Muhammad H, Sarwar T, Bilal H, Rehman TU. Plasmid-mediated *mcr-1* gene in *Acinetobacter baumannii* and *Pseudomonas aeruginosa*: first report from Pakistan. *Revista Da Sociedade Brasileira de Medicina Tropical* 2019;52. <https://doi.org/10.1590/0037-8682-0237-2019>.
- [327] Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* 1995;6:494–500. [https://doi.org/10.1016/0958-1669\(95\)80082-4](https://doi.org/10.1016/0958-1669(95)80082-4).
- [328] Importance C, Relevance I. *The Biology of Acinetobacter*. Boston, MA: Springer US; 1991. <https://doi.org/10.1007/978-1-4899-3553-3>.
- [329] Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology* 2014;5:172. <https://doi.org/10.3389/fmicb.2014.00172>.
- [330] Bolanos-Garcia VM, Davies OR. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA) - General Subjects* 2006;1760:1304–13. <https://doi.org/10.1016/J.BBAGEN.2006.03.027>.
- [331] Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* 2012;11:56. <https://doi.org/10.1186/1475-2859-11-56>.
- [332] Fathi-Roudsari M, Akhavian-Tehrani A, Maghsoudi N. Comparison of three *Escherichia coli* strains in recombinant production of reteplase. *Avicenna Journal of Medical Biotechnology* 2016;8:16–22.
- [333] Lin FH, Sun T, Fletcher GL, Davies PL. Thermolabile antifreeze protein produced in *Escherichia coli* for structural analysis. *Protein Expression and Purification* 2012;82:75–82. <https://doi.org/10.1016/j.pep.2011.11.013>.
- [334] de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC Biotechnology* 2007;7:32. <https://doi.org/10.1186/1472-6750-7-32>.
- [335] Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics* 2005;1751:119–39. <https://doi.org/10.1016/j.bbapap.2005.06.005>.
- [336] Sun J, Xu Y, Gao R, Lin J, Wei W, Srinivas S, et al. Deciphering MCR-2 Colistin Resistance. *MBio* 2017;8:e00625-17. <https://doi.org/10.1128/mBio.00625-17>.
- [337] Newman J, Egan D, Walter TS, Meged R, Berry I, Jelloul M ben, et al. Biological Crystallography Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy n.d. <https://doi.org/10.1107/S0907444905024984>.
- [338] Wooh JW, Kidd RD, Martin JL, Kobe B, IUCr. Comparison of three commercial sparse-matrix crystallization screens. *Acta Crystallographica Section D Biological Crystallography* 2003;59:769–72. <https://doi.org/10.1107/S0907444903002919>.

- [339] Gorrec F, IUCr. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography* 2009;42:1035–42. <https://doi.org/10.1107/S0021889809042022>.
- [340] Fazio VJ, Peat TS, Newman J, IUCr. A drunken search in crystallization space. *Acta Crystallographica Section F Structural Biology Communications* 2014;70:1303–11. <https://doi.org/10.1107/S2053230X1401841X>.
- [341] Moffatt JH, Harper M, Harrison P, Hale JDFF, Vinogradov E, Seemann T, et al. Colistin resistance in *Acinetobacter baumannii* is mediated by complete loss of lipopolysaccharide production. *Antimicrobial Agents and Chemotherapy* 2010;54:4971–7. <https://doi.org/10.1128/AAC.00834-10>.
- [342] Gerson S, Betts JW, Lucaßen K, Nodari CS, Wille J, Josten M, et al. Investigation of Novel pmrB and eptA Mutations in Isogenic *Acinetobacter baumannii* Isolates Associated with Colistin Resistance and Increased Virulence In Vivo. *Antimicrobial Agents and Chemotherapy* 2019;63:1–15. <https://doi.org/10.1128/AAC.01586-18>.
- [343] Adams MD, Nickel GC, Bajaksouzian S, Lavender H, Murthy AR, Jacobs MR, et al. Resistance to colistin in *Acinetobacter baumannii* associated with mutations in the PmrAB two-component system. *Antimicrobial Agents and Chemotherapy* 2009;53:3628–34. <https://doi.org/10.1128/AAC.00284-09>.
- [344] Beceiro A, Llobet E, Aranda J, Bengoechea JA, Doumith M, Hornsey M, et al. Phosphoethanolamine modification of lipid A in colistin-resistant variants of *Acinetobacter baumannii* mediated by the pmrAB two-component regulatory system. *Antimicrobial Agents and Chemotherapy* 2011;55:3370–9. <https://doi.org/10.1128/AAC.00079-11>.
- [345] Nurtop E, Bayındır Bilman F, Menekse S, Kurt Azap O, Gönen M, Ergonul O, et al. Promoters of Colistin Resistance in *Acinetobacter baumannii* Infections. *Microbial Drug Resistance* 2019;25:997–1002. <https://doi.org/10.1089/mdr.2018.0396>.
- [346] Feng Y. Transferability of MCR-1/2 Polymyxin Resistance: Complex Dissemination and Genetic Mechanism. *ACS Infectious Diseases* 2018;4:291–300. <https://doi.org/10.1021/acsinfecdis.7b00201>.
- [347] Nang SC, Morris FC, McDonald MJ, Han M-LL, Wang J, Strugnell RA, et al. Fitness cost of mcr-1-mediated polymyxin resistance in *Klebsiella pneumoniae*. *Journal of Antimicrobial Chemotherapy* 2018;73:1604–10. <https://doi.org/10.1093/jac/dky061>.
- [348] Coppi M, Cannatelli A, Antonelli A, Baccani I, di Pilato V, Sennati S, et al. A simple phenotypic method for screening of MCR-1-mediated colistin resistance. *Clinical Microbiology and Infection* 2018;24:201.e1-201.e3. <https://doi.org/10.1016/j.cmi.2017.08.011>.
- [349] Li H, Wang Y, Meng Q, Wang Y, Xia G, Xia X, et al. Comprehensive proteomic and metabolomic profiling of mcr-1-mediated colistin resistance in *Escherichia coli*. *International Journal of Antimicrobial Agents* 2019;53:795–804. <https://doi.org/10.1016/j.ijantimicag.2019.02.014>.
- [350] Esposito F, Fernandes MR, Lopes R, Muñoz M, Sabino CP, Cunha MP, et al. Detection of colistin-resistant mcr-1-positive *Escherichia coli* by use of assays based on inhibition by EDTA and zeta potential. *Journal of Clinical Microbiology* 2017;55:3454–65. <https://doi.org/10.1128/JCM.00835-17>.

- [351] Lobstein J, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* 2012;11:56. <https://doi.org/10.1186/1475-2859-11-56>.
- [352] Zhao Y, Meng Q, Lai Y, Wang L, Zhou D, Dou C, et al. Structural and mechanistic insights into polymyxin resistance mediated by EptC originating from Escherichia coli. *FEBS Journal* 2019;286:750–64. <https://doi.org/10.1111/febs.14719>.
- [353] Wang W, Nema S, Teagarden D. Protein aggregation-Pathways and influencing factors. *International Journal of Pharmaceutics* 2010;390:89–99. <https://doi.org/10.1016/j.ijpharm.2010.02.025>.
- [354] Antunes LCS, Visca P, Towner KJ. *Acinetobacter baumannii*: evolution of a global pathogen. *Pathogens and Disease* 2014;71:292–301. <https://doi.org/10.1111/2049-632X.12125>.
- [355] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The Protein Data Bank. *Acta Crystallographica Section D, Biological Crystallography* 2002;58:899–907. <https://doi.org/10.1107/s0907444902003451>.
- [356] Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification* 2004;37:203–6. <https://doi.org/10.1016/j.pep.2004.04.025>.
- [357] Zhang H, Hou M, Xu Y, Srinivas S, Huang M, Liu L, et al. Action and mechanism of the colistin resistance enzyme MCR-4. *Communications Biology* 2019;2:36. <https://doi.org/10.1038/s42003-018-0278-1>.
- [358] Lenhard JR, Bulman ZP, Tsuji BT, Kaye KS. Shifting gears: The future of polymyxin antibiotics. *Antibiotics* 2019;8:42. <https://doi.org/10.3390/antibiotics8020042>.
- [359] Vaara M. Polymyxins and their potential next generation as therapeutic antibiotics. *Frontiers in Microbiology* 2019;10:1689. <https://doi.org/10.3389/fmicb.2019.01689>.

Supplementary data:

Appendix A.E. *coli* MCR-1 DNA and amino acid sequence.

DNA sequence.

```
>ATGATGCAGCATACTTCTGTGTGGTACCGACGCTCGGTCAGTCCGTTTGTCTTGTGG
CGAGTGTTGCCGTTTTCTTGACCGCGACCGCCAATCTTACCTTTTTTGATAAAATCAGC
CAAACCTATCCCATCGCGGACAATCTCGGCTTTGTGCTGACGATCGCTGTCGTGCTCTT
TGGCGCGATGCTACTGATCACCACGCTGTTATCATCGTATCGCTATGTGCTAAAGCCTG
TGTTGATTTTGTCTATTAATCATGGGCGCGGTGACCAGTTATTTTACTGACACTTATGGCA
CGGTCTATGATACGACCATGCTCCAAAATGCCCTACAGACCGACCAAGCCGAGACCAA
GGATCTATTAACGCAGCGTTTATCATGCGTATCATTGGTTTGGGTGTGCTACCAAGTTT
GCTTGTGGCTTTTTGTTAAGGTGGATTATCCGACTTGGGGCAAGGGTTTGTATGCGCCGA
TTGGGCTTGATCGTGGCAAGTCTTGCCTGATTTTACTGCCTGTGGTGGCGTTCAGCA
GTCATTATGCCAGTTTCTTTCGCGTGCATAAGCCGCTGCGTAGCTATGTCAATCCGATC
ATGCCAATCTACTCGGTGGGTAAGCTTGCCAGTATTGAGTATAAAAAAGCCAGTGCGCC
AAAAGATACCATTTATCACGCCAAAGACGCGGTACAAGCAACCAAGCCTGATATGCGTA
AGCCACGCCTAGTGGTGTTCGTCGTCGGTGAGACGGCACGCGCCGATCATGTCAGCTT
CAATGGCTATGAGCGCGATACTTTCCACAGCTTGCCAAGATCGATGGCGTGACCAATT
TTAGCAATGTCACATCGTGCGGCACATCGACGGCGTATTCTGTGCCGTGTATGTTTACG
TATCTGGGCGCGGATGAGTATGATGTCGATACCGCAAATACCAAGAAAATGTGCTGG
ATACGCTGGATCGCTTGGGCGTAAGTATCTTGTGGCGTGATAATAATTCGGACTCAAAA
GGCGTGATGGATAAGCTGCCAAAAGCGCAATTTGCCGATTATAAATCCGCGACCAACA
ACGCCATCTGCAACACCAATCCTTATAACGAATGCCGCGATGTCGGTATGCTCGTTGGC
TTAGATGACTTTGTCGCTGCCAATAACGGCAAAGATATGCTGATCATGCTGCACCAAAT
GGGCAATCACGGGCGCTGCGTATTTTAAGCGATATGATGAAAAGTTTGCCAAATTCACGC
CAGTGTGTGAAGGTAATGAGCTTGCCAAGTGCGAACATCAGTCCTTGATCAATGCTTAT
GACAAATGCCTTGCTTGCCACCGATGATTTTCATCGCTCAAAGTATCCAGTGGCTGCAGAC
GCACAGCAATGCCTATGATGTCTCAATGCTGTATGTCAGCGATCATGGCGAAAGTCTGG
GTGAGAACGGTGTCTATCTACATGGTATGCCAAATGCCTTTGCACCAAAAAGAACAGCGC
AGTGTGCCTGCATTTTTCTGGACGGATAAGCAAATGGCATCACGCCAATGGCAACCG
ATACCGTCCTGACCCATGACGCGATCACGCCGACATTATTAAGCTGTTTGTATGTCACC
GCGGACAAAGTCAAAGACCGCACCGCATTTCATCCGCTGA
```

Amino Acid Sequence.

```
>>MMQHTSVWYRRSVSPFVLVASVAVFLTATANLTFDKISQTYPIADNLGFVLTIAVVLFGA
MLLITLLSSYRYVLKPVLLILLIMGAVTSYFTDTYGTVDTTMLQNALQTDQAETKDLLNAAF
IMRIIGLVLPSSLVAFVKVDYPTWGKGLMRRLGLIVASLALILLPVVAFSSHYASFVRVHKPL
RSYVNPIMPIYSVGKLASIEYKKASAPKDTIYHAKDAVQATKPDMRKPRLVVFVGETARAD
HVSFNGYERDTPQLAKIDGVTNFSNVTSCGTSTAYSVPCMFSYLGADEYDVDTAKYQEN
VLDLDRDLGVSILWRDNNSDSKGVMDKLPKAQFADYKSATNNAICNTNPYNECRDVGMLV
GLDDFVAANNGKDMMLIMLHQMGNHGPAYFKRYDEKFAKFTPVCEGNELAKCEHQSLINAY
DNALLATDDFIAQSIQWLQTHSNAYDVSMLYVSDHGESLGENGVYLVHGMNPAFAPKEQRS
VPAFFWTDKQTGITPMATDVTVLTHDAITPTLLKLFVDVTADKVKDRTA FIRHHHHHH
```

Appendix B. *E. coli* Full-length MCR-1 Properties.

Amino Acids	547
Molecular Weight	60946.48 Da
Isoelectric Point (pI)	6.52
Extinction Coefficient	
Cys fully reduced	66240.00 M-1cm-1
Abs 0.1% (1 g/l)	1.087
Cys fully oxidized	66615.00 M-1cm-1
Abs 0.1% (1 g/l)	1.093

Appendix C. Full-length MCR-1 protein identification by mass fingerprint analysis.

Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
MCR-1	MCR-1 Full Length Sequence	1555.05	57.86	31	541	60.1
A0A140N6W0	Elongation factor Tu OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=tuf PE=1 SV=1 - [A0A140N6W0_ECOBD]	81.80	39.34	12	394	43.3
A0A140NDD7	Sulfatase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_4069 PE=4 SV=1 - [A0A140NDD7_ECOBD]	79.27	26.00	14	577	66.5
A0A140N9F7	Uncharacterized protein OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2032 PE=4 SV=1 - [A0A140N9F7_ECOBD]	62.18	36.06	14	502	54.7
A0A140NFK2	30S ribosomal protein S2 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rpsB PE=3 SV=1 - [A0A140NFK2_ECOBD]	49.55	34.02	6	241	26.7
A0A140NCU2	Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=IpxA PE=3 SV=1 - [A0A140NCU2_ECOBD]	34.72	14.50	2	262	28.1
A0A140NHS0	ATP synthase subunit beta OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=atpD PE=3 SV=1 - [A0A140NHS0_ECOBD]	29.97	13.91	5	460	50.3
A0A140NF95	Sulfatase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_3917 PE=4 SV=1 - [A0A140NF95_ECOBD]	24.82	12.25	6	547	61.6
A0A140ND72	ATP synthase subunit alpha OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=atpA PE=3 SV=1 - [A0A140ND72_ECOBD]	22.36	13.06	6	513	55.2
A0A140NA80	Succinate dehydrogenase flavoprotein subunit OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2937 PE=3 SV=1 - [A0A140NA80_ECOBD]	22.04	9.86	5	588	64.4

Appendix D. Full-length MCR-1 Mass fingerprint analysis of peptide modifications by tryptic and chymotryptic digestion.

Description	Score	Coverage	# PSMs
MCR-1 Full Length Sequence	1555.05	57.86	680
Confidence	Sequence	# PSMs	Modifications
High	ADHVSFNGYER	32	
High	AQFADYK	19	
High	DAVQATKPDmR	50	M10(Oxidation)
High	DAVQATKPDmR	7	
High	DLLNAAFImR	18	M9(Oxidation)
High	DLLNAAFIMR	18	
High	DmLImLHQmGNHGPAYFK	19	M2(Oxidation); M5(Oxidation); M9(Oxidation)
High	DNNSDSKGVmDK	2	M10(Oxidation)
High	DNNSDSKGVmDKLPK	3	M10(Oxidation)
High	DTFPQLAK	8	
High	DTIYHAK	27	
High	DVGmLVGLDDFVAANNGK	49	M4(Oxidation)
High	DVGMLVGLDDFVAANNGK	8	
High	FAKFTPVcEGNELAK	2	C8(Carbamidomethyl)
High	FTPVcEGNELAK	22	C5(Carbamidomethyl)
High	GVmDKLPK	43	M3(Oxidation)
High	GVMdkLPK	1	
High	IDGVTNFSNVTS _c GTSTAYSVP _c mFSYLGADEYDVDTAK	1	C13(Carbamidomethyl); C23(Carbamidomethyl); M24(Oxidation)
High	IIGLGVLPSLLVAFVK	2	
High	LASIEYK	20	
High	LASIEYKK	40	
High	LFDVTADK	22	
High	LFDVTADKVK	30	
High	LFDVTADKVKDR	17	
High	LGVSILWR	8	
High	LVVFVGETAR	23	
High	mmQHTSVWYR	16	M1(Oxidation); M2(Oxidation)
High	QTGITPmATDVLTHDAITPTLL K	29	M7(Oxidation)
High	QTGITPMATDVLTHDAITPTLL K	14	
High	SATNNAIcNTNPYNEcR	68	C8(Carbamidomethyl); C16(Carbamidomethyl)
High	SVPAFFWTDK	8	
High	SYVNPImPIYSVGK	21	M7(Oxidation)
High	SYVNPIMPIYSVGK	7	
High	SYVNPImPIYSVGKLASIEYK	2	M7(Oxidation)
High	TAFIR	4	
High	VDYPTWGK	10	
High	YQENVLDTLDR	10	

PSMs: Number of unique identified Peptide Sequences.

Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
MCR-1	MCR-1 Full Length Sequence	1531.38	53.23	31	541	60.1
A0A140NAX3	FAD-dependent pyridine nucleotide-disulphide oxidoreductase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2492 PE=4 SV=1 - [A0A140NAX3_ECOBD]	52.20	26.73	9	434	47.3
A0A140N3C3	Ribosomal RNA large subunit methyltransferase G OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rlmG PE=3 SV=1 - [A0A140N3C3_ECOBD]	44.71	26.19	8	378	42.3
A0A140NAA3	Riboflavin biosynthesis protein RibD OS=Escherichia coli (strain B / BL21- DE3) OX=469008 GN=ECBD_3247 PE=3 SV=1 - [A0A140NAA3_ECOBD]	35.12	19.35	6	367	40.2
A0A140NDE9	MOSC domain containing protein OS=Escherichia coli (strain B / BL21- DE3) OX=469008 GN=ECBD_2648 PE=4 SV=1 - [A0A140NDE9_ECOBD]	17.31	8.40	2	369	40.6
A0A140N964	tRNA pseudouridine synthase D OS=Escherichia coli (strain B / BL21- DE3) OX=469008 GN=truD PE=3 SV=1 - [A0A140N964_ECOBD]	16.17	18.91	4	349	39.1
A0A140N3D6	Transcriptional regulator, Crp/Fnr family OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_0391 PE=4 SV=1 - [A0A140N3D6_ECOBD]	15.23	24.29	5	210	23.6
A0A140NFK2	30S ribosomal protein S2 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rpsB PE=3 SV=1 - [A0A140NFK2_ECOBD]	4.12	9.54	2	241	26.7
A0A140N598	50S ribosomal protein L13 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rplM PE=3 SV=1 - [A0A140N598_ECOBD]	3.90	11.97	2	142	16.0

Description	Score	Coverage	# Unique Peptides
MCR-1 Full Length Sequence	1531.38	53.23	31
Confidence	Sequence	# PSMs	Modifications
High	ADHVSFNGYER	40	
High	AQFADYK	12	
High	AQFADYKSATNNAIcNTNPYNEcR	1	C15(Carbamidomethyl); C23(Carbamidomethyl)
High	ASAPKDTIYHAK	3	
High	DAVQATKPDmR	48	M10(Oxidation)
High	DAVQATKPDmR	9	
High	DLLNAAFImR	2	M9(Oxidation)
High	DmLImLHQmGNHGPAFYK	22	M2(Oxidation); M5(Oxidation); M9(Oxidation)
High	DmLImLHQmGNHGPAFYK	19	M2(Oxidation); M5(Oxidation)
High	DNNSDSKGVmDK	2	M10(Oxidation)
High	DNNSDSKGVmDKLPK	5	M10(Oxidation)
High	DTFPQLAK	12	
High	DTIYHAK	11	
High	DVGmLVGLDDFVAANNGK	50	M4(Oxidation)
High	DVGMLVGLDDFVAANNGK	9	
High	FAKFTPVcEGNELAK	1	C8(Carbamidomethyl)
High	FTPVcEGNELAK	17	C5(Carbamidomethyl)
High	GVmDKLPK	31	M3(Oxidation)
High	IDGVTNFSNVTSccGTSTAYSVPcmF SYLGADEYDVDTAK	1	C13(Carbamidomethyl); T17(Phospho); C23(Carbamidomethyl); M24(Oxidation)
High	IDGVTNFSNVTSccGTSTAYSVPcm FSYLGADEYDVDTAK	1	C13(Carbamidomethyl); C23(Carbamidomethyl); M24(Oxidation)
High	LASIEYK	16	
High	LASIEYKK	50	
High	LFDVTADK	16	
High	LFDVTADKVK	36	
High	LFDVTADKVKDR	52	
High	LGVSILWR	9	
High	LVFVVGATAR	24	
High	QTGITPmATDVLTHDAITPTLLK	40	M7(Oxidation)
High	QTGITPMATDVLTHDAITPTLLK	11	
High	SATNNAIcNTNPYNEcR	45	C8(Carbamidomethyl); C16(Carbamidomethyl)
High	SVPAFFWTDK	7	
High	SVPAFFWTDKQTGITPmATDVLTHDAITPTLLK	12	M17(Oxidation)
High	SYVNPImPIYSVGK	6	M7(Oxidation)
High	SYVNPIMPISVGK	3	
High	TAFIR	3	
High	YDEKFAK	1	
High	YQENVLDTLDR	11	

Appendix E.Truncated MCR-1 protein identification by mass fingerprint analysis.

Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
MCR-1	Truncated MCR-1	1491.07	53.23	31	541	60.1
A0A140NAX3	FAD-dependent pyridine nucleotide-disulphide oxidoreductase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2492 PE=4 SV=1 - [A0A140NAX3_ECOBD]	50.15	26.73	9	434	47.3
A0A140N3C3	Ribosomal RNA large subunit methyltransferase G OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rlmG PE=3 SV=1 - [A0A140N3C3_ECOBD]	44.71	26.19	8	378	42.3
A0A140NAA3	Riboflavin biosynthesis protein RibD OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_3247 PE=3 SV=1 - [A0A140NAA3_ECOBD]	35.12	19.35	6	367	40.2
A0A140NDE9	MOSC domain containing protein OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_2648 PE=4 SV=1 - [A0A140NDE9_ECOBD]	17.31	8.40	2	369	40.6
A0A140N964	tRNA pseudouridine synthase D OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=truD PE=3 SV=1 - [A0A140N964_ECOBD]	16.17	18.91	4	349	39.1
A0A140N3D6	Transcriptional regulator, Crp/Fnr family OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_0391 PE=4 SV=1 - [A0A140N3D6_ECOBD]	13.37	26.67	5	210	23.6
A0A140NFK2	30S ribosomal protein S2 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rpsB PE=3 SV=1 - [A0A140NFK2_ECOBD]	4.12	9.54	2	241	26.7
A0A140N598	50S ribosomal protein L13 OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=rpIM PE=3 SV=1 - [A0A140N598_ECOBD]	3.90	11.97	2	142	16.0

Appendix F. *E. coli* cMCR-1 DNA and amino acid sequence.

DNA sequence:

```
>>GATACCATTTATCACGCCAAAGACGCGGTACAAGCAACCAAGCCTGATATGCGTAAG
CCACGCCTAGTGGTGTTCGTCGTCGGTGAGACGGCACGCGCCGATCATGTCAGCTTCA
ATGGCTATGAGCGCGATACTTTCCCACAGCTTGCCAAGATCGATGGCGTGACCAATTTT
AGCAATGTCACATCGTGCGGCACATCGACGGCGTATTCTGTGCCGTGTATGTTTCAGCT
ATCTGGGCGCGGATGAGTATGATGTGATACCGCCAAATACCAAGAAAATGTGCTGGA
TACGCTGGATCGCTTGGGCGTAAGTATCTTGTGGCGTGATAATAATTCGGACTCAAAG
GCGTGATGGATAAGCTGCCAAAAGCGCAATTTGCCGATTATAAATCCGCGACCAACAAC
GCCATCTGCAACACCAATCCTTATAACGAATGCCGCGATGTGCGGTATGCTCGTTGGCTT
AGATGACTTTGTGCGCTGCCAATAACGGCAAAGATATGCTGATCATGCTGCACCAAATGG
GCAATCACGGGCCTGCGTATTTTAAGCGATATGATGAAAAGTTTGCCAAATTCACGCCA
GTGTGTGAAGGTAATGAGCTTGCCAAGTGCGAACATCAGTCCTTGATCAATGCTTATGA
CAATGCCTTGCTTGCCACCGATGATTTTCATCGCTCAAAGTATCCAGTGGCTGCAGACGC
ACAGCAATGCCTATGATGTCTCAATGCTGTATGTCAGCGATCATGGCGAAAGTCTGGGT
GAGAACGGTGTCTATCTACATGGTATGCCAAATGCCTTTGCACCAAAGAAGCAGCGCAG
TGTGCCTGCATTTTTCTGGACGGATAAGCAAAGTGGCATCACGCCAATGGCAACCGATA
CCGTCCTGACCCATGACGCGATCACGCCGACATTATTAAGCTGTTTGTATGTCACCGC
GGACAAAGTCAAAGACCGCACCGCATTTCATCCGCTGA
```

Amino Acid sequence:

```
DTIYHAKDAVQATKPDMRKPRLLVVFVGETARADHVSFNGYERDTFPQLAKIDGVTNFSNV
TSCGTSTAYSVPCMFSYLGADEYDVTAKYQENVLDTLDRLGVSILWRDNNSDSKGVMDK
LPKAQFADYKSATNNAICNTNPYNECRDVGMLVGLDDFVAANNGKDMLIMLHQMGNHHPA
YFKRYDEKFAKFTPVCEGNELAKCEHQSLINAYDNALLATDDFIAQSIQWLQTHSNAYDVSM
LYVSDHGESLGENGVYLVHGMPNAFAPKEQRSVPAFFWTDKQTGITPMATDTVLTHDAITPT
LLKLFVTDKVKDRTA FIR
```

Appendix G. *E. coli* cMCR-1 Properties.

Amino Acids	324
Molecular Weight	36112.23 Da
Isoelectric Point (pI)	5.08
Extinction Coefficient	
Cys fully reduced	37360.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.035
Cys fully oxidized	37735.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.045

Appendix H. *E. coli* cMCR-2 DNA and amino acid sequence.

DNA sequence:

```
>>ACCATCTATCATGCCAAAGACGCCGTGCAGACCACCAAGCCGAGCGAGCGTAAGCC
ACGCCTAGTGGTGTTCGTCGTCGGTGAGACGGCGCGTGCTGACCATGTGCAGTTCAAT
GGCTATGGCCGTGAGACTTTCCCGCAGCTTGCCAAAGTTGATGGCTTGGCGAATTTTA
GCCAAGTGACATCGTGTGGCACATCGACGGCGTATTCTGTGCCGTGTATGTTTCAGCTA
TTTGGGTCAAGATGACTATGATGTTCGATACCGCCAAATACCAAGAAAATGTGCTAGATA
CGCTTGACCGCTTGGGTGTGGGTATCTTGTGGCGTGATAATAATTCAGACTCAAAGGC
GTGATGGATAAGCTACCTGCCACGCAGTATTTTGATTATAAATCAGCAACCAACAATAC
CATCTGTAACACCAATCCCTATAACGAATGCCGTGATGTCGGTATGCTTGTGGGCTAG
ATGACTATGTCAGCGCCAATAATGGCAAAGATATGCTCATCATGCTACACCAAATGGGC
AATCATGGGCCGGCGTACTTTAAGCGTTATGATGAGCAATTTGCCAAATTCACCCCGT
GTGCGAAGGCAACGAGCTTGCCAAATGCGAACACCAATCACTCATCAATGCCTATGAC
AATGCGCTACTTGCGACTGATGATTTTATCGCCAAAAGCATCGATTGGCTAAAAACGCA
TGAAGCGAACTACGATGTCGCCATGCTCTATGTCAGTGACCACGGCGAGAGCTTGGGC
GAAAATGGTGTCTATCTGCATGGTATGCCAAATGCCTTTGCACCAAAGAACAGCGAGC
TGTGCCTGCGTTTTTTTTGGTCAAATAATACGACATTCAAGCCAACTGCCAGCGATACTG
TGCTGACGCATGATGCGATTACGCCAACACTGCTTAAGCTGTTTGTATGTCACAGCGGG
CAAGGTCAAAGACCGCGCGGCATTTATCCAGTAA
```

Amino acid sequence:

```
>>TIYHAKDAVQTTKPSERKPRLVVFVGETARADHVQFNGYGRETFFQLAKVDGLANFSQ
VTSCGTSTAYSVPCMFSYLGQDDYDVDTAKYQENVLDTLDRLGVGILWRDNNSDSKGVMD
KLPATQYFDYKSATNNTICNTNPYNECRDVGMLVGLDDYVSANNGKDMLIMLHQMGNHGP
AYFKRYDEQFAKFTPVCEGNELAKCEHQSLINAYDNALLATDDFIAKSIDWLKTHEANYDVA
MLYVSDHGESLGENGVYLVHGMPNAFAPKEQRAVPAFFWSNNTTFKPTASDTVLTDAITPT
LLKLFDVTAGKVKDRAAFIQ
```

Appendix I. *E. coli* cMCR-2 Properties.

Amino Acids	322
Molecular Weight	35888.91 Da
Isoelectric Point (pI)	5.16
Extinction Coefficient	
Cys fully reduced	40340.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.124
Cys fully oxidized	40715.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.134

Appendix J. cMCR-2 mass fingerprint analysis.

Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
MCR-2 Catalytic domain	MCR-2 Catalytic domain	2085.27	97.20	70	322	35.9
P0ACE7	Purine nucleoside phosphoramidase OS=Escherichia coli (strain K12) OX=83333 GN=hinT PE=1 SV=1 - [HINT_ECOLI]	29.78	22.69	2	119	13.2
C4ZQV9	3,4-dihydroxy-2-butanone 4-phosphate synthase OS=Escherichia coli (strain K12 / MC4100 / BW2952) OX=595496 GN=ribB PE=3 SV=1 - [RIBB_ECOBW]	13.79	21.20	5	217	23.3
B1XEN9	Probable cytosol aminopeptidase OS=Escherichia coli (strain K12 / DH10B) OX=316385 GN=pepA PE=3 SV=1 - [AMPA_ECODH]	9.70	11.93	4	503	54.8
P0AB89	Adenylosuccinate lyase OS=Escherichia coli (strain K12) OX=83333 GN=purB PE=1 SV=1 - [PUR8_ECOLI]	8.10	7.46	4	456	51.5
P0A9K9	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD OS=Escherichia coli (strain K12) OX=83333 GN=slyD PE=1 SV=1 - [SLYD_ECOLI]	8.05	26.02	3	196	20.8
P0AF90	Regulator of ribonuclease activity B OS=Escherichia coli (strain K12) OX=83333 GN=rraB PE=1 SV=1 - [RRAB_ECOLI]	7.67	23.19	2	138	15.6
P0AFG6	Dihydrolipoyllysine-residue succinyltransferase component of 2- oxoglutarate dehydrogenase complex OS=Escherichia coli (strain K12) OX=83333 GN=sucB PE=1 SV=2 - [ODO2_ECOLI]	6.83	8.89	3	405	44.0
C4ZYT7	S-ribosylhomocysteine lyase OS=Escherichia coli (strain K12 / MC4100 / BW2952) OX=595496 GN=luxS PE=3 SV=1 - [LUXS_ECOBW]	6.20	20.47	3	171	19.4
P0CE47	Elongation factor Tu 1 OS=Escherichia coli (strain K12) OX=83333 GN=tufA PE=1 SV=1 - [EFTU1_ECOLI]	5.41	6.09	2	394	43.3

Appendix K.cMCR-2 Mass fingerprint analysis of peptide modifications.

Description	Score	Coverage	# Unique Peptides
MCR-2 Catalytic domain	2085.27	97.20	70
Confidence	Sequence	# PSMs	Modifications
High	AKcEHQSL	1	C3(Carbamidomethyl)
High	AKVDGLANF	4	
High	APKEQRAVPAF	1	
High	APKEQRAVPAFF	3	
High	APKEQRAVPAFFW	2	
High	ATDDFIAKSIDW	3	
High	ATDDFIAKSIDWL	1	
High	DEQFAKF	3	
High	DNALLATDDF	27	
High	DRLGVGIL	15	
High	DRLGVGILW	13	
High	DTLDRLGVGIL	3	
Medium	DVDTAKY	1	
High	DVDTAKYQENVLDTL	1	
High	DVTAGKVKDRAAF	11	
High	DVTAGKVKDRAAFIQ	2	
High	DYKSATNNTIcNTNPY	2	C11(Carbamidomethyl)
High	DYKSATNNTIcNTNPYNEcRD VGML	1	C11(Carbamidomethyl); C19(Carbamidomethyl)
High	FKRYDEQF	1	
High	FWSNNTTF	1	
High	GQDDYDVDTAKY	2	
High	GRETFPQL	4	
High	GRETFPQLAKVDGL	28	
High	HAKDAVQTTKPSERKPRLVVF	9	
High	HGMPNAF	4	
High	HGMPNAFAPKEQRAVPAFF	2	
High	HQmGNHGPAY	1	M3(Oxidation)
High	HQMGNHGPAY	1	
High	HQmGNHGPAYF	32	M3(Oxidation)
High	HQMGNHGPAYF	53	
High	IAKSIDW	3	
Medium	INAYDNALL	2	
High	KLFDVTAGKVKDRAAF	13	
High	KPTASDTV L	21	
High	KPTASDTV LTHDAITPTLL	5	
Medium	KRYDEQF	6	
High	KRYDEQFAKF	5	
High	KSATNNTIcNTNPY	11	C9(Carbamidomethyl)
Medium	KTHEANY	1	
High	LATDDFIAKSIDW	3	
High	LGQDDYDVDTAKY	17	
Medium	LHGmPNAF	1	M4(Oxidation)
High	LHGMPNAF	3	
High	LKTHEANY	7	

High	LKTHEANYDVAML	2	
High	NEcRDVGmL	29	C3(Carbamidomethyl); M8(Oxidation)
High	NEcRDVGML	25	C3(Carbamidomethyl)
High	NEcRDVGmLVGLDDY	2	C3(Carbamidomethyl); M8(Oxidation)
High	NEcRDVGMLVGLDDY	2	C3(Carbamidomethyl)
High	PQLAKVDGL	1	
High	PQLAKVDGLANF	4	
Medium	QENVLDTL	2	
High	QENVLDTLDRl	10	
High	RDNNSDSKGVmDKLPATQY	39	M11(Oxidation)
High	RDNNSDSKGVMDKLPATQY	27	
High	RDNNSDSKGVmDKLPATQYF	5	M11(Oxidation)
High	RDNNSDSKGVMDKLPATQYF	14	
High	SNNTTFKPTASDTVl	2	
Medium	SQVTScGTStAY	1	C6(Carbamidomethyl); T10(Phospho)
High	SQVTScGTSTAY	5	C6(Carbamidomethyl)
Medium	SYLGQDDY	1	
High	THDAITPTL	3	
High	THDAITPTLL	70	
High	TPVcEGNEL	7	C4(Carbamidomethyl)
High	TPVcEGNELAKcEHQSL	19	C4(Carbamidomethyl); C12(Carbamidomethyl)
High	TPVcEGNELAKcEHQSLINAY	3	C4(Carbamidomethyl); C12(Carbamidomethyl)
High	VGLDDYVSANNGKdML	24	M15(Oxidation)
High	VGLDDYVSANNGKdML	26	
High	VSANNGKdML	19	M9(Oxidation)
High	VSANNGKdML	16	
High	VSANNGKdMLmL	2	M9(Oxidation); M12(Oxidation)
High	VSANNGKdMLmL	8	M12(Oxidation)
High	VSANNGKdMLIML	1	
High	VSDHGESLGENGVY	100	
High	VSDHGESLGENGVYL	36	
High	VGETARADHVQF	25	
High	VGETARADHVQFNgy	4	
High	WRDNNSDSKGVmDKLPATQY	1	M12(Oxidation)
High	WRDNNSDSKGVMDKLPATQY	15	
Medium	WSNNTTF	3	
High	WSNNTTFKPTASDTVl	2	
High	YVSDHGESL	2	
High	YVSDHGESLGENGVY	13	

Appendix L. PDB25 data set obtained from Dali Server.

No:	PDB	Chain	Z-Score	rmsd	lali	nres	%id	PDB Description
1	5lm	A	60.8	0	325	325	100	PHOSPHATIDYLETHANOLAMINE TRANSFERASE MCR-1;
2	5fql	A	25.6	2.7	277	507	15	IDURONATE-2-SULFATASE;
3	6s21	A	24.2	3.1	273	486	14	ENDO-4-O-SULFATASE;
4	4upl	A	23.7	3	272	555	15	SULFATASE FAMILY PROTEIN;
5	5g2v	A	23.7	3	272	494	15	N-ACETYLGLUCOSAMINE-6-SULFATASE;
6	3b5q	B	23.4	2.9	265	467	12	PUTATIVE SULFATASE YIDJ;
7	6uss	A	23.1	3.1	263	485	13	SULFATASE;
8	4miv	D	22.9	3.1	266	480	12	N-SULPHOGLUCOSAMINE SULPHOHYDROLASE;
9	1p49	A	22	3.2	253	549	15	STERYL-SULFATASE;
10	6xlp	A	21.3	3.1	255	586	12	LPS BINDING PROTEIN;
11	3lxq	A	20.8	3.3	270	409	17	UNCHARACTERIZED PROTEIN VP1736;
12	6c01	A	20.3	3.2	250	819	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTERASE
13	5olt	A	19.9	3.3	258	383	12	CELLULOSE BIOSYNTHESIS PROTEIN BCSG;
14	3igy	B	19.6	3.2	243	549	12	COFACTOR-INDEPENDENT PHOSPHOGLYCERATE MUTASE;
15	2w5q	A	19.6	3.4	260	424	13	PROCESSED GLYCEROL PHOSPHATE LIPOTEICHOIC ACID
16	5xwk	A	19.5	3.2	257	530	12	ALKALINE PHOSPHATASE PHOK;
17	4cys	A	19.2	3.4	261	534	13	ARYLSULFATASE;
18	6dgm	B	17.9	3.4	254	382	13	PHOSPHOGLYCEROL TRANSFERASE GACH;
19	7kwd	A	16.5	3.3	236	473	8	ALKALINE PHOSPHATASE;
20	4n7t	A	16.3	2.9	217	402	17	PHOSPHOPENTOMUTASE;
21	6qsq	A	16.2	3.8	234	503	13	ALKALINE PHOSPHATASE;
22	3szy	A	15.7	3.5	227	413	9	PHOSPHONOACETATE HYDROLASE;
23	2zkt	B	15.7	2.8	216	381	10	2,3-BISPHOSPHOGLYCERATE-INDEPENDENT PHOSPHOGLYCER
24	4kjg	A	14.9	3.4	227	486	11	INTESTINAL-TYPE ALKALINE PHOSPHATASE 1;
25	2d1g	A	14.5	3.3	216	481	12	ACID PHOSPHATASE;
26	2l82	A	7.4	3.5	125	162	6	DESIGNED PROTEIN OR32;
27	5xkt	A	5.7	3.7	136	199	13	UREASE ACCESSORY PROTEIN UREG;
28	1xjq	B	5.6	4	137	590	7	BIFUNCTIONAL 3'-PHOSPHOADENOSINE 5'-PHOSPHOSULFAT
29	1de0	A	5	4.2	137	289	5	NITROGENASE IRON PROTEIN;
30	2mr5	A	4.8	3.2	97	136	14	DE NOVO DESIGNED PROTEIN OR457;
31	6l6z	A	4.8	4.1	154	562	8	CTP SYNTHASE;
32	2j37	W	4.7	3.5	124	479	11	SIGNAL RECOGNITION PARTICLE 19 KDA PROTEIN
33	1m8p	A	4.7	3.6	120	573	10	SULFATE ADENYLYLTRANSFERASE;
34	5n0s	B	4.7	4.1	128	405	12	PEPTIDE N-METHYLTRANSFERASE;
35	6qp0	A	4.6	3.9	120	188	5	PUTATIVE CHROMATIN BINDING PROTEIN;

36	5dmh	A	4.6	3.4	118	420	8	UNCHARACTERIZED PROTEIN CONSERVED IN BACTERIA;
37	6qu3	A	4.6	4.7	86	465	7	ATP-DEPENDENT 6- PHOSPHOFRUCTOKINASE;
38	1wde	A	4.5	3.9	114	289	10	PROBABLE DIPHTHINE SYNTHASE;
39	1grq	A	4.4	4.4	129	178	9	CHLORAMPHENICOL 3-O PHOSPHOTRANSFERASE;
40	4jyc	C	4.4	3.9	136	324	10	METHYLMALONYL-COA MUTASE ACCESSORY PROTEIN;
41	2vli	B	4.4	4.4	130	173	8	ANTIBIOTIC RESISTANCE PROTEIN;
42	2bdt	A	4.4	3.8	122	172	10	BH3686;
43	2qv5	A	4.4	5	110	246	12	UNCHARACTERIZED PROTEIN ATU2773;
44	3fmi	A	4.4	3.8	135	232	10	DETHIOBIOTIN SYNTHETASE;
45	3nd1	B	4.2	4.1	115	250	6	PRECORRIN-6A SYNTHASE/COBF PROTEIN;
46	3mle	B	4.2	4.2	144	224	8	DETHIOBIOTIN SYNTHETASE;
47	1pjq	B	4.2	4.4	122	456	11	SIROHEME SYNTHASE;
48	3cmu	A	4.1	4.9	141	###	10	DNA (5'-
49	3czp	A	4.1	4.5	136	466	7	PUTATIVE POLYPHOSPHATE KINASE 2;
50	2rhm	C	4.1	5	133	190	8	PUTATIVE KINASE;
51	6u05	A	4	3.8	130	413	11	TRNA LIGASE;
52	6ouv	A	3.9	4.5	119	595	5	1-DEOXY-D-XYLULOSE-5- PHOSPHATE SYNTHASE;
53	1xu4	A	3.9	4.2	142	318	6	DNA REPAIR AND RECOMBINATION PROTEIN RADA;
54	3rhf	A	3.9	3.7	142	283	8	PUTATIVE POLYPHOSPHATE KINASE 2 FAMILY PROTEIN;
55	6ub8	A	3.9	3.6	94	249	4	GLYCO_HYDRO_CC DOMAIN- CONTAINING PROTEIN;
56	4xfr	A	3.9	4.6	125	401	7	UNCHARACTERIZED PROTEIN;
57	3czq	A	3.9	4.2	145	288	7	PUTATIVE POLYPHOSPHATE KINASE 2;
58	6aok	A	3.9	3.5	115	213	4	CEG4;
59	5jlp	A	3.8	4.1	123	397	17	PHOSPHOSERINE PHOSPHATASE;
60	4ohz	A	3.8	3.9	130	400	6	PROTEIN CLPF-1;
61	6m48	A	3.8	3.4	110	813	14	SPAC;
62	2ln3	A	3.8	2.2	65	83	6	DE NOVO DESIGNED PROTEIN OR135;
63	6rjb	B	3.8	4.9	119	622	7	TRANSKETOLASE;

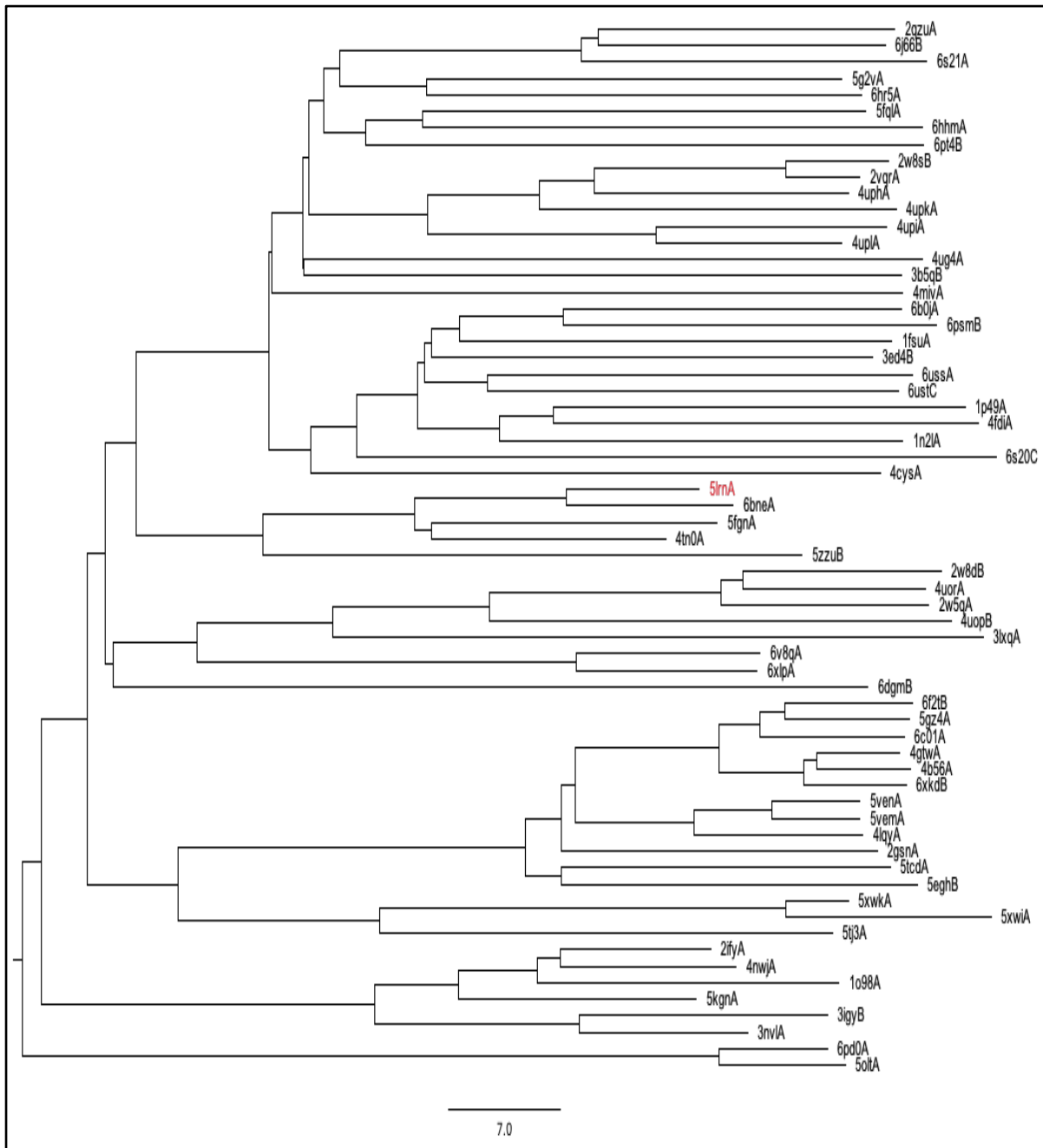
No:	PDB	Chain	Z-score	rmsd	lali	nres	%id	PDB Description
1	5lm	A	60.8	0	325	325	100	PHOSPHATIDYLETHANOLAMINE TRANSFERASE MCR-1;
2	6bne	A	52.5	1	324	336	67	PHOSPHOETHANOLAMINE TRANSFERASE;
3	5fgn	A	42.7	2	315	536	40	LIPOOLIGOSACCHARIDE PHOSPHOETHANOLAMINE TRANSFERASE
4	4tn0	A	41.6	1.6	296	308	41	UPF0141 PROTEIN YJDB;
5	5zzu	B	32.4	2.4	293	357	20	PHOSPHOETHANOLAMINE TRANSFERASE EPTC;
6	5fql	A	25.6	2.7	277	507	15	IDURONATE-2-SULFATASE;
7	2qzu	A	25.5	3.2	277	465	15	PUTATIVE SULFATASE YIDJ;
8	6hhm	A	25.2	2.9	273	459	14	ARYLSULFATASE;
9	2w8s	B	25.1	2.7	271	513	13	PHOSPHONATE MONOESTER HYDROLASE;
10	4uph	A	24.8	2.7	272	504	14	SULFATASE (SULFURIC ESTER HYDROLASE) PROTEIN;
11	4upk	A	24.7	2.8	270	506	15	PHOSPHONATE MONOESTER HYDROLASE;
12	6pt4	B	24.7	2.9	269	472	16	EXO-2S-IOTA CARRAGEENAN S1 SULFATASE;
13	6j66	B	24.7	3.2	276	476	19	CHONDROITIN SULFATE/DERMATAN SULFATE 4-O-ENDOSULF
14	4ug4	A	24.3	3	275	515	16	CHOLINE SULFATASE;
15	6s21	A	24.2	3.1	273	486	14	ENDO-4-O-SULFATASE;
16	5g2v	A	23.7	3	271	494	16	N-ACETYLGALACTOSAMINE-6-SULFATASE;
17	4upl	A	23.7	3	272	555	15	SULFATASE FAMILY PROTEIN;
18	3b5q	B	23.4	3	267	467	13	PUTATIVE SULFATASE YIDJ;
19	6uss	A	23.1	3.1	263	485	13	SULFATASE;
20	4miv	A	23	3	262	484	11	N-SULPHOGLUCOSAMINE SULPHOHYDROLASE;
21	6ust	C	22.8	3.2	265	464	13	N-ACETYLGALACTOSAMINE 6-SULFATE SULFATASE;
22	3ed4	B	22.4	10.7	271	474	11	ARYLSULFATASE;
23	6hr5	A	22.4	3	256	376	15	ALPHA-L-RHAMNOSIDASE/SULFATASE (GH78);
24	1p49	A	22	3.2	253	549	15	STERYL-SULFATASE;
25	6psm	B	21.8	3	255	447	13	EXO-4S-KAPPA CARRAGEENAN S1 SULFATASE;
26	6b0j	A	21.8	3.2	261	453	13	IOTA-CARRAGEENAN SULFATASE;
27	4fdi	A	21.7	3.2	255	494	13	N-ACETYLGALACTOSAMINE-6-SULFATASE;
28	6v8q	A	21.7	3.1	253	569	11	INNER MEMBRANE PROTEIN YEJM;
29	1fsu	A	21.7	3	255	475	13	N-ACETYLGALACTOSAMINE-4-SULFATASE;
30	6xlp	A	21.3	3.1	255	586	13	LPS BINDING PROTEIN;
31	4uop	B	21	3.6	266	409	12	LIPOTEICHOIC ACID PRIMASE;
32	1n2l	A	20.9	3.1	251	483	12	ARYLSULFATASE A;
33	3lxq	A	20.8	3.3	270	409	17	UNCHARACTERIZED PROTEIN VP1736;
34	6s20	C	20.5	3.3	256	480	11	N-ACETYLGALACTOSAMINE-6-O-SULFATASE;
35	6pd0	A	20.5	3.4	262	366	13	CELLULOSE BIOSYNTHESIS PROTEIN BCSG;

36	6c01	A	20.3	3.2	250	819	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE
37	5olt	A	19.9	3.3	258	383	12	CELLULOSE BIOSYNTHESIS PROTEIN BCSG;
38	5tcd	A	19.9	3	241	388	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE
39	5vem	A	19.9	3	244	383	16	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE
40	2gsn	A	19.9	3.1	246	382	15	PHOSPHODIESTERASE-NUCLEOTIDE PYROPHOSPHATASE;
41	3igy	B	19.6	3.2	243	549	12	COFACTOR-INDEPENDENT PHOSPHOGLYCERATE MUTASE;
42	6aek	A	19.6	3.2	249	726	12	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE
43	2w5q	A	19.6	3.4	260	424	13	PROCESSED GLYCEROL PHOSPHATE LIPOTEICHOIC ACID
44	5xwk	A	19.5	3.2	257	530	12	ALKALINE PHOSPHATASE PHOK;
45	4cys	A	19.2	3.4	261	534	13	ARYLSULFATASE;
46	1o98	A	19	3	228	509	11	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT
47	5egh	B	18.4	3	240	393	15	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE
48	5tj3	A	18.3	3.2	258	520	14	ALKALINE PHOSPHATASE PAFA;
49	5kgn	A	18.2	3.4	231	530	15	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
50	6dgm	B	17.9	3.5	258	382	13	PHOSPHOGLYCEROL TRANSFERASE GACH;
51	5vpu	A	17.4	3.2	236	509	14	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
52	2x98	A	16.9	3	225	430	12	ALKALINE PHOSPHATASE;
53	7kwd	A	16.5	3.3	236	473	8	ALKALINE PHOSPHATASE;
54	2w5v	A	16.5	3.2	223	346	11	ALKALINE PHOSPHATASE;
55	4n7t	A	16.3	2.9	217	402	17	PHOSPHOPENTOMUTASE;
56	6qsq	A	16.2	3.8	234	503	13	ALKALINE PHOSPHATASE;
57	3a52	A	16.1	3.2	222	400	13	COLD-ACTIVE ALKALINE PHOSPHATASE;
58	3szy	A	15.7	3.5	227	413	9	PHOSPHONOACETATE HYDROLASE;
59	2zkt	B	15.7	2.8	216	381	10	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
60	4kjj	A	14.9	3.4	227	486	11	INTESTINAL-TYPE ALKALINE PHOSPHATASE 1;
61	1k7h	A	14.6	3.4	227	476	12	ALKALINE PHOSPHATASE;
62	2d1g	A	14.5	3.3	216	481	12	ACID PHOSPHATASE;
63	3kd8	A	11.3	3.1	181	367	12	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT
64	3nkm	A	10.1	3.2	234	805	12	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIEST ERASE

No:	PDB	Chain	Z-Score	rmsd	lali	nres	%id	PDB Description
1	5lm	-A	60.8	0	325	325	100	PHOSPHATIDYLETHANOLAMINE TRANSFERASE MCR-1;
2	6bne	-A	52.5	1	324	336	67	PHOSPHOETHANOLAMINE TRANSFERASE;
3	5fgn	-A	42.7	2	315	536	40	LIPOOLIGOSACCHARIDE PHOSPHOETHANOLAMINE TRANSFERASE
4	4tn0	-A	41.6	1.6	296	308	41	UPF0141 PROTEIN YJDB;
5	5zzu	-B	32.4	2.4	293	357	20	PHOSPHOETHANOLAMINE TRANSFERASE EPTC;
6	5fql	-A	25.6	2.7	277	507	15	IDURONATE-2-SULFATASE;
7	2qzu	-A	25.5	3.2	277	465	15	PUTATIVE SULFATASE YIDJ;
8	6hhm	-A	25.2	2.9	273	459	14	ARYLSULFATASE;
9	2w8s	-B	25.1	2.7	270	513	13	PHOSPHONATE MONOESTER HYDROLASE;
10	4uph	-A	24.8	2.7	272	504	14	SULFATASE (SULFURIC ESTER HYDROLASE) PROTEIN;
11	4upk	-A	24.7	2.8	270	506	15	PHOSPHONATE MONOESTER HYDROLASE;
12	6j66	-B	24.7	3.2	276	476	19	CHONDROITIN SULFATE/DERMATAN SULFATE 4-O-ENDOSULF
13	6pt4	-B	24.7	2.9	269	472	16	EXO-2S-IOTA CARRAGEENAN S1 SULFATASE;
14	2vqr	-A	24.6	2.8	270	512	13	PUTATIVE SULFATASE;
15	4ug4	-A	24.3	3	275	515	16	CHOLINE SULFATASE;
16	6s21	-A	24.2	3.1	273	486	14	ENDO-4-O-SULFATASE;
17	4upi	-A	23.8	2.7	270	548	15	SULFATASE FAMILY PROTEIN;
18	5g2v	-A	23.7	3	272	494	15	N-ACETYLGLUCOSAMINE-6-SULFATASE;
19	4upl	-A	23.7	3	272	555	15	SULFATASE FAMILY PROTEIN;
20	3b5q	-B	23.4	3	267	467	13	PUTATIVE SULFATASE YIDJ;
21	6uss	-A	23.1	3.1	263	485	13	SULFATASE;
22	4miv	-A	23	3	262	484	11	N-SULPHOGLUCOSAMINE SULPHOHYDROLASE;
23	6ust	-C	22.8	3.2	265	464	13	N-ACETYL GALACTOSAMINE 6-SULFATE SULFATASE;
24	3ed4	-B	22.4	10.7	271	474	11	ARYLSULFATASE;
25	6hr5	-A	22.4	3	256	376	15	ALPHA-L-RHAMNOSIDASE/SULFATASE (GH78);
26	1p49	-A	22	3.2	253	549	15	STERYL-SULFATASE;
27	6b0j	-A	21.8	3.2	261	453	13	IOTA-CARRAGEENAN SULFATASE;
28	6psm	-B	21.8	3	255	447	13	EXO-4S-KAPPA CARRAGEENAN S1 SULFATASE;
29	1fsu	-A	21.7	3	255	475	13	N-ACETYL GALACTOSAMINE-4-SULFATASE;
30	6v8q	-A	21.7	3.1	253	569	11	INNER MEMBRANE PROTEIN YEJM;
31	4fdi	-A	21.7	3.2	255	494	13	N-ACETYL GALACTOSAMINE-6-SULFATASE;
32	6xlp	-A	21.3	3.1	255	586	13	LPS BINDING PROTEIN;
33	4uop	-B	21	3.6	266	409	12	LIPOTEICHOIC ACID PRIMASE;
34	1n2l	-A	20.9	3.1	251	483	12	ARYLSULFATASE A;
35	3lxq	-A	20.8	3.4	271	409	16	UNCHARACTERIZED PROTEIN VP1736;
36	6s20	-C	20.5	3.3	256	480	11	N-ACETYL GALACTOSAMINE-6-O-SULFATASE;

37	6f2t	-B	20.5	3.2	249	722	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
38	6pd0	-A	20.5	3.4	262	366	13	CELLULOSE BIOSYNTHESIS PROTEIN BCSG;
39	6c01	-A	20.3	3.2	250	819	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
40	4lqy	-A	20	3	243	379	15	BIS(5'-ADENOSYL)-TRIPHOSPHATASE ENPP4;
41	5ven	-A	20	3.2	245	382	16	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
42	5tcd	-A	19.9	3.1	243	388	11	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
43	2w8d	-B	19.9	3.4	263	421	15	PROCESSED GLYCEROL PHOSPHATE LIPOTEICHOIC ACID SY
44	5vem	-A	19.9	3	244	383	16	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
45	5olt	-A	19.9	3.3	258	383	12	CELLULOSE BIOSYNTHESIS PROTEIN BCSG;
46	2gsn	-A	19.9	3.1	246	382	15	PHOSPHODIESTERASE-NUCLEOTIDE PYROPHOSPHATASE;
47	4gtw	-A	19.8	3.2	250	706	12	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
48	6xkd	-B	19.8	3.1	248	681	12	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
49	4b56	-A	19.7	3.3	250	816	12	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
50	4uor	-A	19.6	3.4	257	418	14	LIPOTEICHOIC ACID SYNTHASE;
51	3igy	-B	19.6	3.2	243	549	12	COFACTOR-INDEPENDENT PHOSPHOGLYCERATE MUTASE;
52	2w5q	-A	19.6	3.4	260	424	13	PROCESSED GLYCEROL PHOSPHATE LIPOTEICHOIC ACID
53	5xwk	-A	19.5	3.2	257	530	12	ALKALINE PHOSPHATASE PHOK;
54	3nvl	-A	19.4	3.2	240	549	13	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
55	5xwi	-A	19.4	3.2	256	503	13	ALKALINE PHOSPHATASE PHOK;
56	4cys	-A	19.2	3.4	261	534	13	ARYLSULFATASE;
57	2ify	-A	19.2	3.2	227	508	13	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT
58	5gz4	-A	19.1	3.1	246	804	12	SNAKE VENOM PHOSPHODIESTERASE (PDE);
59	1o98	-A	19	3.2	229	509	11	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT
60	4nwj	-A	18.8	3.1	228	504	11	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
61	5egh	-B	18.4	3	240	393	15	ECTONUCLEOTIDE PYROPHOSPHATASE/PHOSPHODIESTE RASE
62	5tj3	-A	18.3	3.2	258	520	14	ALKALINE PHOSPHATASE PAFA;
63	5kgn	-A	18.2	3.4	231	530	15	2,3-BISPHOSPHOGLYCERATE- INDEPENDENT PHOSPHOGLYCER
64	6dgm	-B	17.9	3.4	254	382	13	PHOSPHOGLYCEROL TRANSFERASE GACH;

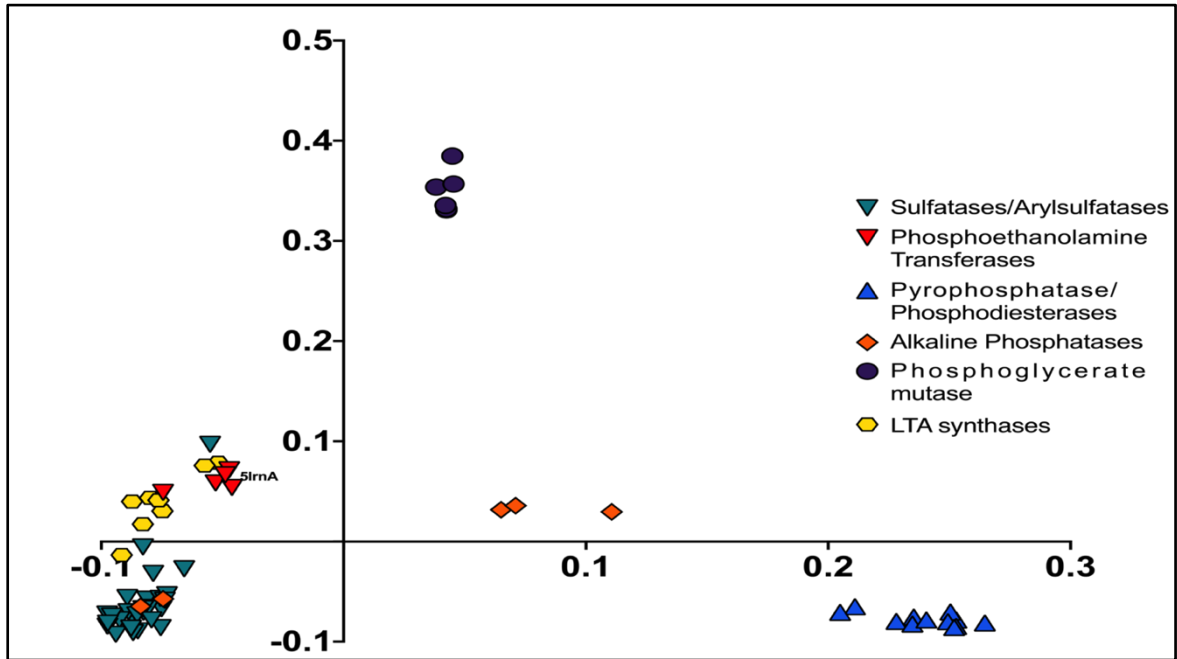
Appendix O. Structural dendrogram of the PDB90 data set



Structural dendrogram of the PDB90 data set.

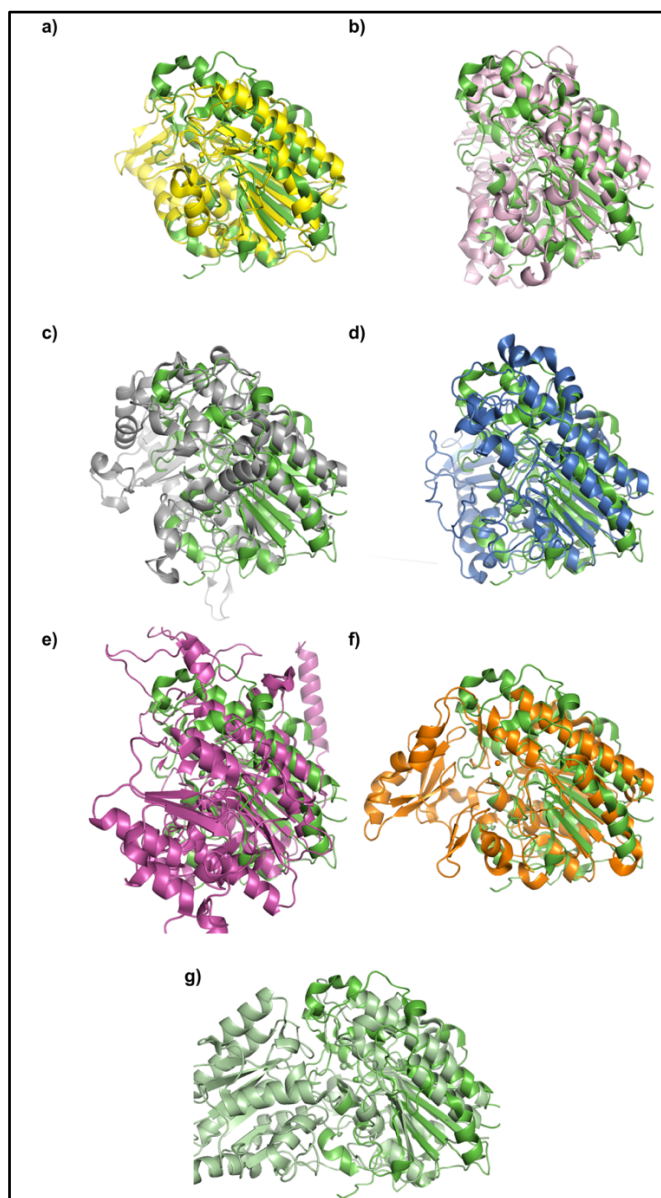
MCR-1 (5lmA Labelled in red for easier location) along with other phosphoethanolamine transferases (6bneA, 5fgnA, 4tn0A and 5zzuB) are located in the clade of the monometallic Arylsulphatases/sulphatases. The second major clade (2w8B and bellow) corresponds to the multimeric alkaline phosphatases.

Appendix P. Correspondence analysis showing the principal groups from the PDB90 data set.



Correspondence analysis showing the principal groups from the PDB90 data set.

Six different enzyme families were confirmed by correspondence analysis. The sulphatases/Arylsulphatases (green triangles pointing downwards),). The Phosphoethanolamine transferases (shown as red triangles pointing downwards), the alkaline phosphatases (orange diamonds), Phosphoglycerate mutases (purple circles), the Pyrophosphatases/Phosphodiesterase family (the blue triangle pointing upwards) and the family of Lipoteichoic acid synthases. MCR-1 has been labelled for easy location. The phosphoethanolamine transferases seem to be a part of the sulfatase/arylsulfatase due to their proximity to this group



Superposition of the representative structures for comparison against MCR-1

The mono-zinc structure of MCR-1 (5LRN in green) was super imposed against the mono metalloenzymes a) Cellulose synthase subunit G (6PD0 in yellow), b) Lipoteichoic acid synthase LtaS (2W5Q in lilac), c) Phosphonate Monoester Hydrolase (2W8S in silver grey) and d) Arylsulfatase(3ED4 in navy blue) as well as the multi-metalloenzymes e) Alkaline phosphatase (1K7H in pink), f) Nucleotide Pyrophosphatase/Phosphodiesterase (2GSN in orange) and g) Phosphoglycerate Mutase (1O98 in light green)

DNA Sequence:

```
>>GTTATCTATGGTCAGGATGCACATCAAGTCCAGCGTGTGCAGAAGAACCTTCCGAAA  
CTGATGATTCTGGTTGTTGGTGAGACAGCTCGTGCGGAGTCTTTTAGCCTGAACGGCTA  
CGCGAAGAATACGAACCCGGAGCTGAGCAAACAGGACATCTTCAACTTTTCACAGGTC  
TCTAGCTGTGGAACCGCGACCGCGGTGAGCGTGCCGTGTATGTTTAGCGGTATGCCGA  
GAGTTGATTACGACGAACAGCTGGCGAGCCACCGTGAAGGTCTGTTGGACATCGCGAA  
GCGCGCTGGTTATCAGGTGACCTGGATTGATAACAACAGCGGCTGCAAAGGTGCGTGC  
GACCGCGTGGAACAATACCAGATCCCGGAGAACCTTAAGAAGAAGTGGTGTAAAGACG  
GCGAGTGCTATGACGACATTTTGATTGACAGCCTGAAGCAATACCTGTCTACGATCGCC  
AAGGATGACGATCGTCCGCGTCTGATTGTTCTGCATCAGGTTGGCTCGCACGGTCCGG  
CATACTACAAACGTGCGCCTGAGGCATATCAGCCATTTAAACCGACGTGCGACACCAAT  
GCAATTCAGGGTTGCTCCCAAACGGAGTTGCTGAATAGCTATGACAACACCATCGTGTA  
TACCGATCATGTTTTGAGTCAAATGATTAACACTCTGAAAGAAATCTCGAAATACCAAAC  
CGGTCTGTGGTATCTGTCTGATCACGGTGAGTCTACAGGTGAGCACGGCCTGTATCTG  
CACGGCTCCCCGATGCTATCGCCCCAAGCCAACAGACCCACGTCCCGATGATCATGT  
GGTTTTCGGAATCTTGAAACAGCATAATCTGGCTCAGGTAAACTGCCTGTCCCAACAG  
ACGAAACAGAAGTTGTCCCAAGATAACCTGTTCCCGAGCTTGCTGTGCTTGCTGGACG  
TAACCACCCAGGTGATCAATCCGCAACTGGACATGCTGCACAGCTGCGCTCATGTGAA  
C
```

Amino Acid Sequence:

```
>>VIYGQDAHQVQRVQKNLPKLMILVVGETARAESFSLNGYAKNTNPESKQDIFNFSQVSS  
CGTATAVSVPCMFSGMPRVDYDEQLASHREGLLDIAKRAGYQVTWIDNNSGCKGACDRV  
EQYQIPENLKKKWCKDGEYDDILIDSLKQYLSTIAKDDDRPRLIVLHQVGSHPAYYKRAP  
EAYQPFKPTCDTNAIQGCSQTELLNSYDNTIVYTDHVL SQMINTLKEISKYQTGLWYLSDHG  
ESTGEHGLYLHGSPYAIAPSQQTHVPMIMWFSESWKQHNLAQVNCLSQQTKQKLSQDNLF  
PSLLSLLDVTTQVINPQLDMLHSCAHVN
```

Appendix S.A. *baumannii* ScPmrC properties.

Amino Acids	331
Molecular Weight	37269.81 Da
Isoelectric Point (pI)	5.93
Extinction Coefficient	
Cys fully reduced	51340.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.378
Cys fully oxidized	51965.00 M ⁻¹ cm ⁻¹
Abs 0.1% (1 g/l)	1.394

PmrC Multi Sequence Alignment (MSA) analysis

Phosphoethanolamine Transferases amino acid sequences used for *A. baumannii* PmrC Multi Sequence Alignment (MSA) analysis

>ATD51479.1 PmrC[*Acinetobacter baumannii*]

MLNFFSTLRNKQISLFMFNLIIAIWLGAILNIGFYHQVHTLTPYFGVKAILFLAATLILVATYYAVLQILNW
KWTAKIFAILLIFIGGFSSYFVNTLGVIIISPDQIQNMVQTDVSEVTDLISLRFVLWTIFFVILPILITQVKFK
QEKVSRLLLLKVFSLVASLAVVGVLLFTYYVDFAAIFREHRDLKGMISPNQNSISSLSMYYHKKAPKKNL
PLVIYGGQDAHQVQRVQKNLPLKMLILVGETARAESFSLNGYAKNTNPELSKQDIFNFSQVSSCGTATA
VSVPCMFSGMPRVYDEQLASHREGLLDIAKRAGYQVTWIDNNSGCKGACDRVEQYQIPENLKKKW
CKDGECYDDILIDSLKQYLSTIAKDDDRPRLIVLHQVGSHPAYYKRAPEAYQPFKPTCDTNAIQGCS
QTELLNSYDNTIVYTDHVLVLSQMINTLKEISKYQTGLWYLSDHGESTGEHGLYLHGSPYAIAPSQQTHV
PMIMWFSESWKQHNLAVNCLSQQTKQKLSQDNLFPSLLSLLDVTTQVINPQLDMLHSCAHVN

>CDK37739.1 PmrC protein [*Klebsiella pneumoniae*]

MSELLPLRRPVVSRTTYLILFACYIGIFLNLAIFYRQVFPLLPVNSLHNWLVFLSMPVIVASVMNLTTLASFL
KLDRLVISLIFILLSASAQYFIWNFGVVIDRSMITNILDTPAESFALLSGEMIAVLGLSGVLAVFVAWWWK
IRKPATRWGAAMRLLNIAVSALLIILVAALFYKDYASVFRNNKELVKSLSPSNSIVAVNSWYAHHRMD
NLPLVKIGEDATQKAVMHNAPRKNLTIVVLGETSRADNFSLGGYSRDTNPLMRQDGVYIFPHTTSCG
TATAVSVPCMFSDMPRAHYDEELAHHQEGVLDILQRAGIRVLWVNDNDGGCKGACDRVPHQNVTDL
KLTGQCIDGECYDDVLFHNLDSYIDNLQQDGIIVLHTIGSHGPTYYNRYPAFRKFTPTCDTNEIQGCT
REQLTNTYDNTILYVDYVVDKAIKLLQSKQDKFTTSLVYLSDHGESLGEDGVYVYHGLPYAIAPDTQKHV
PMALWLSADYQQRYGISAHCLQQRRAQKENYSQDNLFSTLLGLLGVSTREYQAADDILTPCREAG

>UNIPROT:EPTA_ECOLI EptA Polymyxin resistance protein PmrC

MLKRLLKRPSLNLLAWLLLAIFYISICLNIAFFKQVLQALPLDSLHNVLVFLSMPVVAFSVINIVLTLSSFL
WLNRLPLACFILVGAQAQYFIMTYGIVIDRSMIANIIDTTPAESYALMTPQMMLTLGFSGLVLAALIAWIKI
KPATSRLRSVLFGRANILVSVLLILLVAALFYKDYASLFRNNKELVKSLSPSNSIVASWSWYSHQRLAN
LPLVRIGEDAHRNPLMQNEKRKNLTILIVGETSRAENFSLNGYPRETNPRLAKDNVYVFPNTASCGTA
TAVSVPCMFSDMPREHYKEELAQQHQEGVLDIIQRAGINVLWVNDNDGGCKGACDRVPHQNVLTALNLP
DQCINGECYDEVLFHGLEEYINNLLQGDGVIVLHTIGSHGPTYYNRYPPQFRKFTPTCDTNEIQTCTKE
QLVNTYDNTLVYVDYIVDKAINLLKEHQDKFTTSLVYLSDHGESLGENGIYVYHGLPYAIAPDSQKQVPM
LLWLSSEYQKRYQVDQNCQKQQAQTQHYSQDNLFSTLLGLTGVETKYYQAADDILTQTCRRVSE

>UNIPROT:EPTA_SALTY P36555 Polymyxin resistance protein PmrC

MLKRFLKRPLVGLQIAWLLLSFYIAVCLNIAFYKQVLQDLPLNSLRNVLVFISMPVVAFSVNSVLTLAS
FIWLNRLACVFILVGAQAQYFILTGYIHDRSMIANMMMDTTPAETFALMTPQMMLTLGFSGLVLAALIAFW
VKIRPATPRLRSGLYRLASVLISILLVILVAAFFYKDYASLFRNNKQLIKALSPSNSIVASWSWYSHQRLA
NLPLVRIGEDAHRNPLMLKGDNRKNLTILIVGETSRGDDFSLGGYPRDTNPRLAKDDVIYFPHTTSCGT
ATAISVPCMFSDMPRKHYDEELAHHQEGLLDIIQRAGINVLWVNDNDGGCKGACDRVPHQNVTELNLP
GQCIDGECYDEVLFHGLEEYIDHLKGDGVIVLHTIGSHGPTYYNRYPPQFKFTPTCDTNEIQNCSQE
QLINTYDNTLVYVDYIVDKAINLLKSHQDKFTTSLVYLSDHGESLGENGVYVYHGLPYAIAPDTQKHVPM
LIWLSKDYQQRYQVDQAACLQKRASTLDYSQDNLFSTMLGLTGVQTTYYQAADDILQPCRRLSE

>UNIPROT:MCR1_ECOLX A0A0R6L508 Polymyxin resistance protein MCR-1

MMQHTSVWYRRSVSPFVLVASVAVFLTATANLTFDFKISQTYPIADNLGFVLTIAVVLFGAMLLITLTLSS
SYRYVLKPVLLLLIMGAVTSYFTDYGTVYDITMLQNALQTDQAETKDLLNAAFIMRIIGLVPLSLLV
AFVKVDYPTWGKGLMRRGLIVASLALILLPVVAFSSHYASFFRVHKPLRSYVNPIMPIYSVGKLASIEY
KKASAPKDTIYHAKDAVQATKPDMRKPRLLVVFVGETARADHVSFNGYERDTFPQLAKIDGVTNFSN
VTSCGTSTAYSVPCMFSYLGADEYDVTAKYQENVLDTLDRLGVSILWRDNNSDSKGVMDKLPKAAQ
FADYKSATNNAICNTNPYNECRDVGMLVGLDDFVAANNKGDMLIMLHQMGNHGPAYFKRYDEKFAK
FTPVCEGNELAKCEHQSLINAYDNALLATDDFIAQSIQWLQTHSNAYDVSMLYVSDHGESLGENGVY
LHGMPNAFAPKEQRSVPAFFWTDKQGTGITPMATDTVLTHDAITPTLLKLFVDVTADKVKDRTAFIG

>UNIPROT:EPTA_HELPY O24867 Phosphoethanolamine transferase EptA

MASLFHLRFLKPLSCLQAGLLYSLIFGVLYHFPLFVYVYKESNQVSFIAMMVVLFVNGALFLALGLIS
ASLMRWSAIVFSLNSVAFYFISAYKVFLNKSMMGNVLTNTTHEVLGFLSVKLFVFIWVFGVLPGYVIY
KIPLKNSSKKAPFLAILALVFIFIASALANTKNWLWFDKHAKFIGGLILPFAYSVNAFRVSALKFFAPTIP
LPLFSPNHSHSFVVLVIGESARKHNYALYGYQKPTTPrLSKRLADNELTLFNATSCATYTTASLECILD
SSFKNAYENLPTYLTkAGIKVFWYSANDGEKNVKVTSYLKNYELIQKPCNCEAIAPYDESLLYNLPDL
LKEHSNENVLLILHLAGSHGPNYDNKVPLNFRVFKPYCSSADLSSCSKESLINAYDNTIFYNDYLLDKII
SMLENAKQPALMIYLSDHGESLGEEAFYHLHGIPKSIAPKEQYEIPFIVYANEPFKEKHSIIQTQTPINQNV
IFHSVLGVFLDFKNPSVVYRPSLDLLKHKKE

>UNIPROT:EPTB_ECOLI P37661 Phosphoethanolamine transferase EptB

MRYIKSITQQKLSFLLAIYIGLFMNGAVFYRRFGSYAHDFTVWKGISAVVELAATVLTFFLLRLLSLFG
RRSWRILASLVVLFsAGASYYMTFLNVVIGYGIIASVMTTDidLSKEVVGLNFILWLIASVLPILIIWNN
RCRYTLLRQLRTPGQRIRSLAVVLAGIMVWAPIRLLDIQQKKVERATGVDLPsYGGVVANSYLPsNW
LSALGLYAWARVDESSDNNsLLNPAKKFTYQAPQNVDDTYVVFIIGETTRWDHMGIFGYERNTPKL
AQEKNLAAFRGYSCDTATKLSLRCMFVRQGAEDNPQRTLKEQNIFAVLKQLGFSSDLYAMQSEMw
FYSNTMADNIAYREQIGAEPNRNGKPVDDMLLVDEMQQSLGRNPDGKHLIILHTKGSHFNyTQRyPR
SFAQWKPECIGVDSGCTKAQMINSYDNSVTYVDHFISsVIDQVRDKKAIvFYAADHGESINEREHLHG
TPRELAPPEQFRVPMVWMSDKYLENPANAQAFaQLKKEADMKVPRRHVELYDTIMGCLGYTSPD
GGINENNNWCHIPQAKEAAAN

>UNIPROT:EPTB_SALTY P43666 Phosphoethanolamine transferase EptB

MRYIKSMTQQKLSFLLALYIGLFMNCaVFYRRFGSYAQEFtIWKGLSAVVELGATVLTFFLLRLLSLF
GRRVWRVLATLVVLFsAGASYYMTFLNVVIGYGIIASVMTTDidLSKEVVGLHFVLWLIASVLPILFIW
SNHCRYTLLRQLRTPGQRFRSAAVVLAGVMVWAPIRLLDIQQKKVERATGIDLPsYGGVVANSYLP
SNWLSALGLYAWAQVDESSDNNsLINPARKFTYVAPKDGDDTYVVFIIGETTRWDHMGIFGYERNTP
PKLAQEKNLAAFRGYSCDTATKLSLRCMFVREGGADNNPQRTLKEQNVFAVLKQLGFSSDLYAMQs
EMWFYSNTMADNISYREQIGAEPNRNGKTVDDMLLIDEMQNSLAQNPEGKHLIILHTKGSHFNyTQR
YPRSYAQWKPECIGVDSGCTKAQMINSYDNSVTYVDHFITSVFDQLRDKKAIvFYAADHGESINEREH
LHGTPRNMAPPEQFRVPMVWMSDKYLASPQHAQMFAHLKQQAeIKVPRRHVELYDTIMGCLGYTs
PNGGINQNNNWCHIPDVQKVAAK

>UNIPROT:YHBX_ECOLI P42640 Putative transferase YhbX

MTVFNKFARTFKSHWLLYLCVIVFGITNLVASSGAHMVQRLLFFVLtILVVKRISSLPLRLLVAAPFVLLT
AADMSISLYSWCTFGTTFNDFGFAISVLQSDPDEVVKMLGMYIPYLCAFAFLSLLFLAVIIKYDVSLPTKK
VTGILLLIVISGSLFSACQFAYKDAKNKKAfSPYILASRFATYTPFFNLNYFALAAKEHQRLLSIANTVPY
FQLSVRDTGIDTYVLIVGESVRVDNMSLYGYTRSTTPQVEAQRKQIKLFNQAISGAPYTALSVPPLSLTA
DSVLSHDIHNYPDNIINMANQAGFQTFWLSSQSAFRQNGTAVTSIAMRAMETVYVRGFDELLPHLS
QALQQNTQQKLVHLNLSHEPACSAYPQSSAVFQPQDDQDACYDNSIHytDSLLGQVFELLKDRR
ASVMYFADHGLERDPTKKNVYFHGGREASQQAyHVPMFIWYSPVLGDGVDRRTENNIFSTAYNNYLI
NAWMGVTKPEQPQTLEEVIAHYKGDSRVVDANHDVFDYVMLRKEFTEDKQGNPTPEGQG

>UNIPROT:CPTA_SALTY Q7CPC0 Phosphoethanolamine transferase CptA

MQSTLLQTKPAFSWKALGWALLYFWFFSTLLQAIYLTGYSGTNGLRDSLLYSSLWLIPVFLFPGRIRVI
AAVIGVVWLAASLAALSYVVIYGQEFsQSVLFVMFETNANEASEYLSQYFSLKIVLVALAYTVAAILLWT
RLRPVYIPSPWRYLVSFALLYGLILHPIAMNTFIKHKSMEKTLDSLAsRMEPAAPWQFITGYyQYRLQL
ASLNKLLNENDALPPLANFQDHSGDAPRTLVLVIGESTQRGRMSLYGYPRETTPELDALHKTDpGLT
VFNNVVTsRPYTIEILQQALTFADeKNPDWYLTkPSLMNMKQAGYKTFWITNQQTMTARNTMLTVF
SKQTDKQFYMNQQRTQSAREYDSNVLAPFKAVLADPAPKKFIIVHLLGTHIKYKFRYPENQgKFDGKT
DHVPPGLSSDELESYNDYDNANLYNDYVVASLIKDYKATDPNGFLLYFSDHGEEVYDTPPHKTQGRN
EDSPTRHMYTVPFLWTSEKWQAaHPRDFSQDVDRKYSSSELIHTWSDLAGLTYDGYDPTRSITNP
QFKETTRWIGNPYKKNALIDYDTLPYGDQVGNQ

>UNIPROT:EPTC_ECOLI P0CB39 Phosphoethanolamine transferase EptC

MHSTEVQAKPLFSWKALGWALLYFWFFSTLLQAIYISGYSGTNGIRDSLLFSSLWLIPVFLFPKRIKIIA
AVIGVVLWAASLAALCYVYVYGQEFSSQSVLFVFMFETNTNEASEYLSQYFSLKIVLIALAYTAVAVLLWTR
LRPVYIPKPWRYVVSFALLYGLILHPIAMNTFIKKNKPFKTLNLSRMEPAAPWQFLTGYQQYRQQL
NSLTKLLNENNALPPLANFKDESGNEPRTLVLVIGESTQRGRMSLYGYPRETTPELDALHKTDPNLTV
FNNVVTSRPYTIEILQQALTFANEKPNLDLYLTQPPLMNMKQAGYKTFWITNQQTMTARNTMLTVFS
RQTDKQYYMNQQRQSAREYDTNVLKPFQEVLDNDPAPKKLIIVHLLGTHIKYKYRYPENQGGKFDGNT
DHVPPGLNAEELESYNDYDNANLYNDHVVASLIKDFKAANPNGFLVYFSDHGEEVYDTPPHKTQGR
NEDNPTRHMYTIPFLLWTSEKWQATHPRDFSQDVDRKYSLAELIHTWSDLAGLSYDGYDPTRSVNV
PQFKETTRWIGNPYKKNALIDYDTLPYGDQVGNQ

>UNIPROT:Y1005_HAEIN P44974 Putative phosphoethanolamine transferase HI_1005

MMFGKKIIHSKTIWFYLLFLFSIFTNMGLGNVLSNIPVSDYAILFVITAFYFTPWLGRILITVFLSAL
FYCSAGFFYGIPSLGIIASVYETNINETLEYTTIPFWIFLQASYFILFVALMKLSFHVKKQAFKWLNTA
VISLLLVSYNQFSDWNYGYKFRFYPVLFYSEFDRMNDLYLEQRDFLNQSANAPSQWDIQSFMPKYK
NYILIIGESMRKDYMSLYGFPLKTTTPFLERVKGTVFENYYSAAPNTQPSLQLTLYRAEKGETVYTDNIIS
LAKKAGVKTYWISNQGKIGEFDTIASRIGQSADETIFMKPLGYNSKKVYDDEMLPVLDKALKENISNSK
LIVHLIGSHPAFCERLPYEVKNYFINQSMSCYLESIKYTDQFLEKLNSQLVAQNEPYSVIYFSDHGLAH
YEDSNGLSLHPNNLYKQDYEIPFIMFSSDSQKVEKIKTPQSAFNFVYGFADWMGIKEKHLQGVDFFHP
EKQEIKVFDWNNVNVKELADDPKLPETVQ

>UNIPROT:OPGE_ECOLI P75785 Phosphoethanolamine transferase OpgE

MNLTLESLVTRSRVFPWTAIFYFLQSLINLGLGYPFSLLYTAAFTAILLLLWRTLPRVQKVLVGVSSL
VAACYFPFAQAYGAPNFNTLLALHSTNMEESTEILTIFPWYSYLVGLFIFALGVIAIRKKENEKARWNT
FDSLCLVFSVATFFVAPVQNLAWGGVFKLKDGTGYPVFRFAKDVIVNNNEVIEEQERMAKLSGMKDTW
TVTAVKPKYQTYVVVIGESARRDALGAFGGHWDNTPFASSVNGLIFADYIAASGSTQKSLGLTLNRV
DGKPFQDNFVTLANRAGFQTWWFSNQGQIGEYDTAIAIAKRADEVYFLKEGNFEADKNTKDEALL
DMTAQVLAQEHSQPQLIVLHLMGSHPQACDRTQGKYETFVQSKETSCYLYTMTQTDDLLRKLYDQL
RNSGSSFSLVYFSDHGLAFKERGKDVQYLAHDDKYQQNFQVPMVISDDKAHRVIKARRSANDFL
GFFSQWTGIKAKEINIKYPFISEKKAGPIYITNFQLQKVDYNHLGTDIFDPKP

Appendix U. ScPmrC protein identification by mass fingerprint analysis.

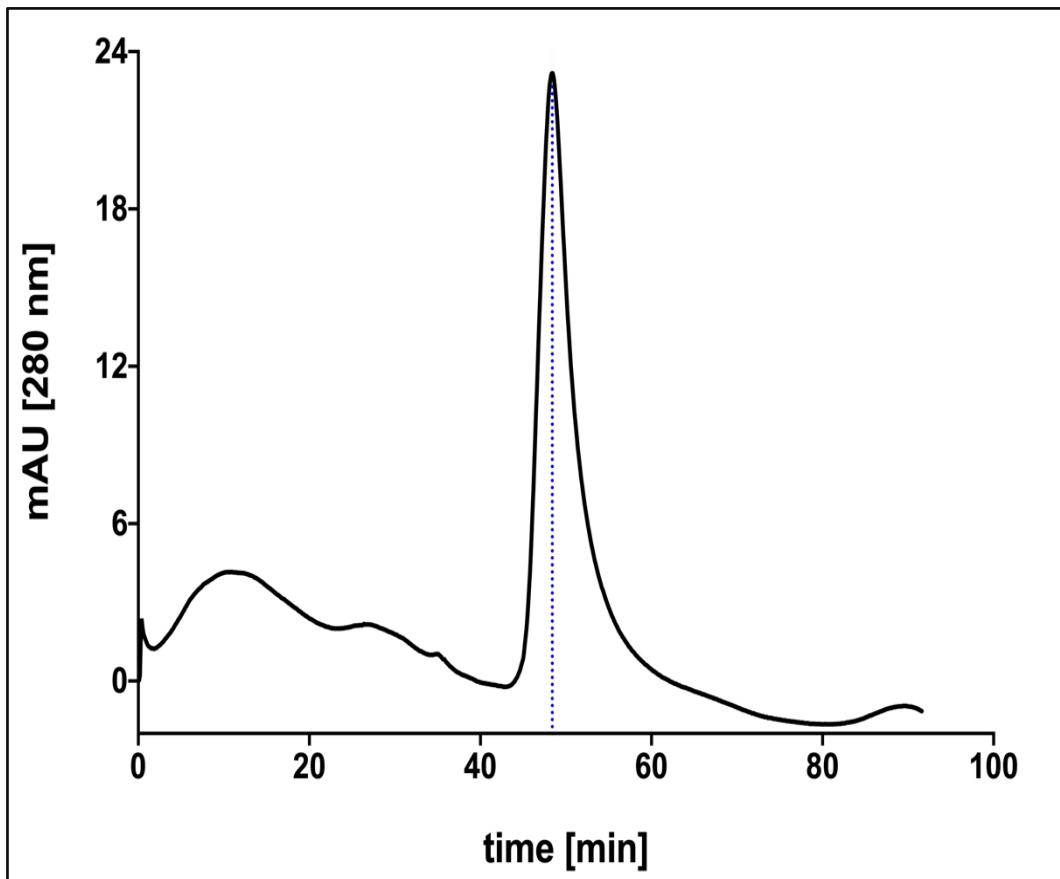
Accession	Description	Score	Coverage	# Unique Peptides	# AAs	MW [kDa]
His-ScPmrC	His-ScPmrC His-ScPmrC	1394.96	86.86	59	350	39.4
A0A140N587	Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=arnA PE=3 SV=1 - [A0A140N587_ECOBD]	197.24	60.76	45	660	74.2
A0A140NEB2	4-hydroxy-tetrahydrodipicolinate reductase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=dapB PE=3 SV=1 - [A0A140NEB2_ECOBD]	10.21	9.52	2	273	28.7
A0A140N855	Tyrosine recombinase XerD OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=xerD PE=3 SV=1 - [A0A140N855_ECOBD]	9.00	12.42	3	298	34.2
A0A140NF57	Transcriptional regulator, LysR family OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_4063 PE=3 SV=1 - [A0A140NF57_ECOBD]	7.64	23.61	6	305	34.3
A0A140N9B6	Succinylornithine transaminase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=astC PE=3 SV=1 - [A0A140N9B6_ECOBD]	5.40	5.91	2	406	43.6
A0A140N6V1	Peptidyl-prolyl cis-trans isomerase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_0400 PE=3 SV=1 - [A0A140N6V1_ECOBD]	5.19	15.31	2	196	20.8
A0A140NEC0	Aspartokinase OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=ECBD_4013 PE=3 SV=1 - [A0A140NEC0_ECOBD]	4.10	5.35	2	449	48.5
A0A140NAY3	Histidine biosynthesis bifunctional protein HisB OS=Escherichia coli (strain B / BL21-DE3) OX=469008 GN=hisB PE=3 SV=1 - [A0A140NAY3_ECOBD]	3.81	12.68	3	355	40.2

Appendix V.ScPmrC Mass fingerprint analysis of peptide modifications.

Description	Score	Coverage	# Unique Peptides
His-ScPmrC His-	1394.96	86.86	59
Confidence Level	Sequence	Modifications	
High	AIAPSQQTHVPmImW	M12(Oxidation); M14(Oxidation)	
High	AIAPSQQTHVPmIMW	M12(Oxidation)	
High	AIAPSQQTHVPMImW	M14(Oxidation)	
High	AIAPSQQTHVPMIMW		
High	AIAPSQQTHVPmImWF	M12(Oxidation); M14(Oxidation)	
High	AIAPSQQTHVPMImWF	M14(Oxidation)	
Medium	AIAPSQQTHVPmImWFSESW	M12(Oxidation); M14(Oxidation)	
High	AKNTNPELSKQDIF		
High	AKNTNPELSKQDIFNF		
High	ASHREGLL		
High	ASHREGLLDIAKRAGY		
Medium	cKDGEcYDDLIDSL	C1(Carbamidomethyl); C6(Carbamidomethyl)	
High	DDLIDSL		
High	DDLIDSLKQY		
High	DEQLASHREGL		
High	DEQLASHREGLL		
High	DIAKRAGYQVTW		
High	DIAKRAGYQVTWIDNNSGcKGAcDR	T11(Phospho); C19(Carbamidomethyl); C23(Carbamidomethyl)	
High	VEQY		
High	DIAKRAGYQVTWIDNNSGcKGAcD	C19(Carbamidomethyl); C23(Carbamidomethyl)	
High	RVEQY		
High	DNTIVYTDHVL		
High	DVTTQVINPQLDmL	M13(Oxidation)	
High	DVTTQVINPQLDML		
High	DVTTQVINPQLDmLHScAHVN	M13(Oxidation); C17(Carbamidomethyl)	
High	DVTTQVINPQLDMLHScAHVN	C17(Carbamidomethyl)	
High	GQDAHQVQRVQKNLPKlmIL	M18(Oxidation)	
High	HQVGSHGPAYY		
High	HQVGSHGPAYYKRAPEAY		
High	IDNNSGcKGAcDRVEQY	C7(Carbamidomethyl); C11(Carbamidomethyl)	
High	IDNNSGcKGAcDRVEQYQIPENL	C7(Carbamidomethyl); C11(Carbamidomethyl)	
High	IDSLKQY		
High	IVLHQVGSHGPAY		
High	KEISKY		
High	KPTcDTNAIQGcSQTEL	C4(Carbamidomethyl); C12(Carbamidomethyl)	
High	KPTcDTNAIQGcSQTELLNSY	C4(Carbamidomethyl); C12(Carbamidomethyl)	
High	KQHNLAQVNcL	C10(Carbamidomethyl)	
High	KQHNLAQVNcLSQQTkQKL	C10(Carbamidomethyl)	
High	KQHNLAQVNcLSQQTkQKL	C10(Carbamidomethyl)	
High	KRAPEAY		
High	L DIAKRAGY		
Medium	L DIAKRAGY		
High	LSDHGESTGEHGL		
High	LSDHGESTGEHGLY		
High	LSTIAKDDDRPRL		
High	LSTIAKDDDRPRLVL		

High	mILVVGETARAESF	M1(Oxidation)
Medium	NFsQVSScGTATAVSVpCmF	S3(Phospho); C8(Carbamidomethyl); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVsScGTATAVSVpCmF	S6(Phospho); C8(Carbamidomethyl); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVsScGTATAVSVpCmF	S6(Phospho); C8(Carbamidomethyl); C18(Carbamidomethyl)
High	NFSQVSScGTATAVSVpCmF	S7(Phospho); C8(Carbamidomethyl); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVSScGtATAVSVpCmF	C8(Carbamidomethyl); T10(Phospho); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVSScGTAtAVSVpCmF	C8(Carbamidomethyl); T12(Phospho); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVSScGTAtAVSVpCmF	C8(Carbamidomethyl); T12(Phospho); C18(Carbamidomethyl)
High	NFSQVSScGTATAVSVpCmF	C8(Carbamidomethyl); S15(Phospho); C18(Carbamidomethyl); M19(Oxidation)
High	NFSQVSScGTATAVSVpCmF	C8(Carbamidomethyl); S15(Phospho); C18(Carbamidomethyl)
High	NGYAKNTNPELSKQDIF	
High	NSYDNTIVY	
High	QIPENLKKKW	
High	QVtWIDNNSGcKGAcDRVEQY	T3(Phospho); C11(Carbamidomethyl); C15(Carbamidomethyl)
High	QVTWIDNNSGcKGAcDRVEQY	C11(Carbamidomethyl); C15(Carbamidomethyl)
High	SDHGESTGEHGLY	
High	SGmPRVDY	M3(Oxidation)
High	SGMPRVDY	
High	SGmPRVDYDEQL	M3(Oxidation)
High	SGMPRVDYDEQL	
High	SGmPRVDYDEQLASHREGL	M3(Oxidation)
High	SGMPRVDYDEQLASHREGL	
High	SLNGYAKNTNPEL	
Medium	SQmINTL	M3(Oxidation)
High	SQmINTLKEISKY	M3(Oxidation)
High	SQMINTLKEISKY	
High	SQVsScGTATAVSVpCmF	S4(Phospho); C6(Carbamidomethyl); C16(Carbamidomethyl); M17(Oxidation)
High	SQVsScGTATAVSVpCmF	S4(Phospho); C6(Carbamidomethyl); C16(Carbamidomethyl)
High	SQVSScGTATAVSVpCmF	S5(Phospho); C6(Carbamidomethyl); C16(Carbamidomethyl); M17(Oxidation)
High	SQVSScGTATAVSVpCmF	S5(Phospho); C6(Carbamidomethyl); C16(Carbamidomethyl)
High	SQVSScGtATAVSVpCmF	C6(Carbamidomethyl); T8(Phospho); C16(Carbamidomethyl); M17(Oxidation)
High	SQVSScGtATAVSVpCmF	C6(Carbamidomethyl); T8(Phospho); C16(Carbamidomethyl)
High	SQVSScGTAtAVSVpCmF	C6(Carbamidomethyl); T10(Phospho); C16(Carbamidomethyl); M17(Oxidation)
High	SQVSScGTATAVSVpCmF	C6(Carbamidomethyl); C16(Carbamidomethyl); M17(Oxidation)
High	STIAKDDDRPRL	
High	STIAKDDDRPRLIVL	
High	TDHVLSQmINTL	M8(Oxidation); T11(Phospho)
High	TDHVLSQmINTL	M8(Oxidation)
High	TDHVLSQmINTL	T11(Phospho)
High	TDHVLSQmINTL	
High	TDHVLSQmINTLKEISKY	M8(Oxidation)
High	TDHVLSQmINTLKEISKY	
High	YKRAPEAYQPF	
High	YLSDHGESTGEHGL	

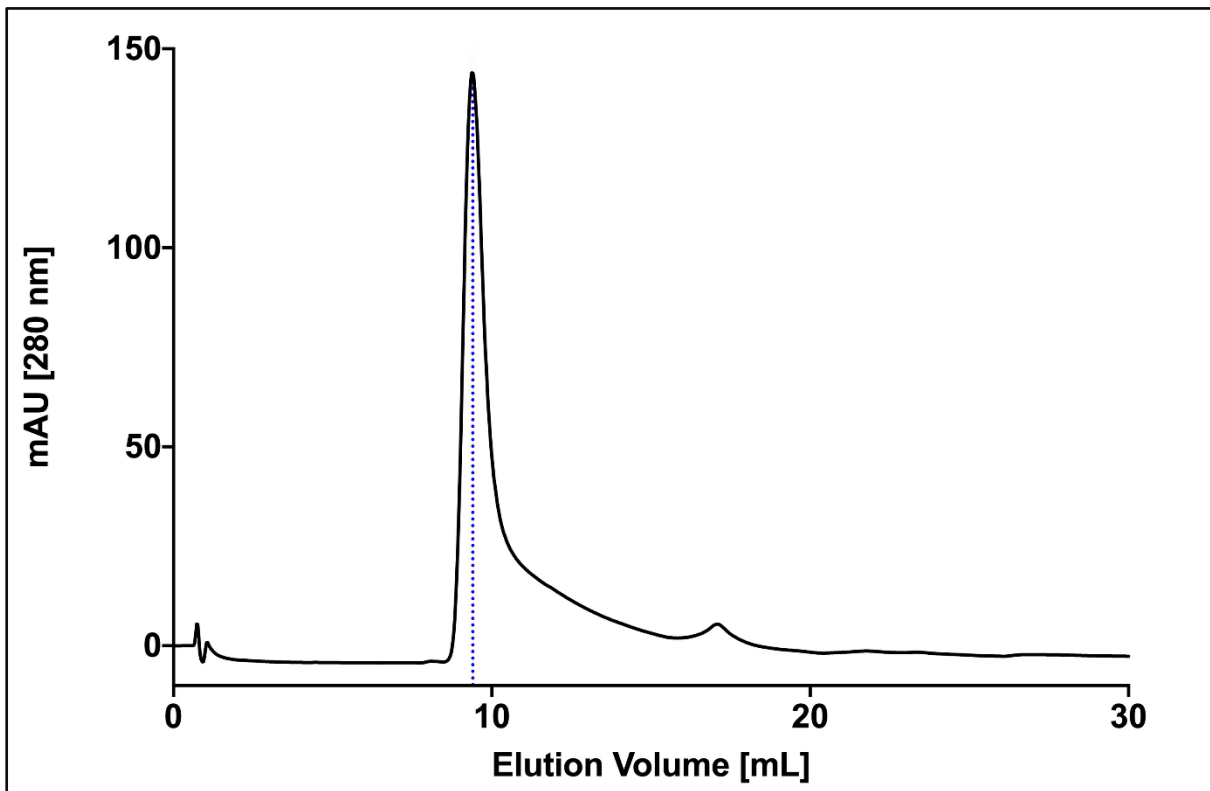
Appendix W. Size Exclusion Chromatography S75 column void volume estimation using Blue Dextran 2000.



Blue dextran calibration chromatogram of Superdex 75 column.

Blue dextran is a collection of dextran molecules with a covalently attached blue dye. Dextran molecules are large and irregular in shape thus incapable of interact with the resin beads. The volume at which the first peak of A280 from the blue dextran emerges from the system is the void volume, V_0 , of the column. For this case since the flow was established as $1 \text{ mL} \cdot \text{min}^{-1}$, the V_0 for the Superdex 75 column is 48 mL.

Appendix X. Size Exclusion Chromatography S200 column void volume estimation using Blue Dextran 2000.



Blue dextran calibration chromatogram of Superdex S200 column.

The volume at which the first peak of A280 from the blue dextran emerges from the system is the void volume, V_0 , of the column. For this case since the flow was established as $0.5 \text{ mL} \cdot \text{min}^{-1}$, the V_0 for the Superdex S200 column is 9.4 mL.