

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

UNIVERSITY OF SOUTHAMPTON

Faculty of Natural and Environmental Sciences

School of Biological Sciences

**Molecular mechanisms of nitric oxide induced *Pseudomonas aeruginosa* biofilm
dispersal**

by

Andrew James Hutchin

Thesis for the degree of Doctor of Philosophy

April 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Biological Sciences

Thesis for the degree of Doctor of Philosophy

MOLECULAR MECHANISMS OF NITRIC OXIDE INDUCED *PSEUDOMONAS AERUGINOSA* BIOFILM DISPERSAL

Andrew James Hutchin

Bacteria are able to transition between a single celled, planktonic lifestyle and a communal, antibiotic tolerant phenotype known as a biofilm. The ability of bacteria to transition between resistant and virulent behaviours underpins a number of chronic infections and is regulated by levels of the secondary messenger cyclic dimeric guanosine monophosphate (c-di-GMP). A number of external stimuli are able to modulate the levels of intracellular c-di-GMP and induce biofilm formation or dispersal. Nitric oxide (NO) is one such stimulus, able to induce biofilm dispersal in *Pseudomonas aeruginosa*, although the nature of NO detection, and how this couples to reduced levels of c-di-GMP is currently unknown.

Chapter 1 describes the current knowledge available in literature about biofilm dispersal, potential sensors of NO and enzymatic domains that regulate c-di-GMP concentrations. Chapter 2 details the methodologies used within this work to structurally and biochemically characterise potential NO sensors. Chapter 3 describes attempts to crystallise the transmembranous sensor protein Mucoid alginate regulator (MucR). Structural characterisation of the c-di-GMP catabolising enzyme domain within MucR, and contrast to equivalent domains in other proteins, offers new insight into how this process may be regulated. Chapter 4 describes the spectroscopic characterisation of NO binding to the protein Biofilm dispersal locus A (BdIA). Crystal structures of the apo-form of BdIA are used to propose models of how NO binding may go on to induce biofilm dispersal. Chapter 5 details how bioinformatics strategies, coupled to established knowledge of enzyme architectures and related phenotypes, can be used to identify further proteins that may serve to bind NO. Understanding the molecular details of NO-induced biofilm dispersal could lead to the development of novel therapies to treat biofilm infections when used in conjugation with conventional antibiotics.

Table of Contents

Table of Contents.....	i
List of Figures.....	v
List of Tables.....	ix
DECLARATION OF AUTHORSHIP	xi
Acknowledgements	xiii
Definitions and Abbreviations.....	xv
Chapter 1: Introduction	1
1.1 Introduction	1
1.2 Biofilms and cyclic dimeric guanosine monophosphate	1
1.2.1 Biofilms	1
1.2.2 The biofilm lifecycle	2
1.2.3 Biofilm extracellular polymeric substance	4
1.2.4 Biofilms are tolerant to antimicrobials	5
1.2.5 Biofilm formation and dispersal can be controlled by external stimuli, including nitric oxide.....	7
1.2.6 C-di-GMP regulation in <i>P. aeruginosa</i>	8
1.2.7 C-di-GMP receptors.....	11
1.2.8 Other nucleotides involved in biofilm regulation	13
1.3 Structural and functional details of c-di-GMP regulating proteins	14
1.3.1 GGDEF domain diguanylate cyclases	14
1.3.2 EAL domain phosphodiesterases	21
1.3.3 HD-GYP domain phosphodiesterases.....	25
1.3.4 Oligoribonuclease phosphodiesterases	28
1.4 Nitric oxide detection within <i>Pseudomonas aeruginosa</i>	30
1.4.1 PAS domains.....	31
1.4.2 Haem-binding PAS domains	33
1.4.3 BdIA	40
1.4.4 MHYT domains	42
1.4.5 MucR and NbdA	43

Chapter 2:	Materials and Methods	47
2.1	Experimental methods	47
2.1.1	High-throughput molecular biology at Oxford Protein Production Facility (OPPF-UK).....	47
2.1.2	Small scale expression of transmembranous protein fragments at OPPF-UK52	
2.1.3	Large scale expression of transmembranous MucR fragments at Diamond Membrane Protein Lab (MPL).....	53
2.1.4	Membrane harvesting at MPL.....	53
2.1.5	Fluorescence size-exclusion chromatography at MPL.....	54
2.1.6	Purification of MucR ⁷⁻²⁶⁶ at MPL.....	54
2.1.7	Fluorescence quantification at MPL.....	55
2.1.8	SEC-Multi-angle light scattering (MALS) at MPL.....	55
2.1.9	Thermostability analysis at MPL.....	56
2.1.10	Cloning of soluble protein fragments.....	56
2.1.11	Expression of soluble protein fragments.....	58
2.1.12	Expression of MucR ^{EAL}	58
2.1.13	Purification of soluble MucR and NbdA protein fragments	58
2.1.14	Analysis of MucR ^{GGDEF-EAL} and MucR ^{EAL} PDE activity	59
2.1.15	Analysis of MucR ^{GGDEF-EAL} DGC activity.....	60
2.1.16	Crystallisation of MucR ^{EAL}	60
2.1.17	Purification of BdIA fragments	60
2.1.18	BdIA haem co-purification through anion exchange.....	61
2.1.19	BdIA haem co-purification through SEC	61
2.1.20	Spectroscopic analysis of BdIA ¹⁻¹²⁷ -NO interaction	61
2.1.21	Crystallisation of BdIA.....	62
2.2	Macromolecular crystallography.....	63
2.2.1	The requirement for X-ray crystallography	63
2.2.2	Crystal growth	63
2.2.3	X-ray diffraction.....	65
2.2.4	Structure factors.....	69
2.2.5	Refinement.....	71

Chapter 3:	Structural and functional studies of the MYHT domain proteins MucR and NbdA.....	73
3.1	Purification and analysis of MYHT domains	73
3.1.1	Gene fragment design and cloning at OPPF-UK.....	73
3.1.2	Fluorescence size-exclusion chromatography	76
3.1.3	Protein purification	79
3.1.4	Optimisation of expression conditions	80
3.1.5	Optimised protein purification and SEC-MALS	81
3.1.6	Thermostability analysis.....	84
3.2	Analysis of MucR and NbdA cytosolic domains.....	86
3.2.1	The EAL domain of MucR is inactive in isolation, but is an active PDE when expressed with the GGDEF domain	86
3.2.2	Crystal structure of the MucR EAL domain	91
3.2.3	The product bound state of PA3825 may demonstrate further regulation of EAL PDEs	100
3.3	Further work required to characterise regulation of c-di-GMP metabolism in MucR and NbdA.....	104
Chapter 4:	Spectroscopic and structural characterisation of the NO binding protein BdlA.....	107
4.1	Chemotaxis within <i>P. aeruginosa</i>	107
4.2	Protein purification and optimisation of haem incorporation	114
4.3	Haem-containing BdlA interacts with NO	120
4.4	Crystallisation of BdlA	124
4.5	Future work to characterise the role of BdlA in nitric oxide induced biofilm dispersal.....	136
Chapter 5:	Identification of haem-binding PAS domains within the PAO1 genome	137
5.1	Analysis of cofactor binding on the basis of conserved evolution	138
5.2	Assignment of cofactor binding based on conservation of interacting residues	143
5.3	Candidate proteins involved in NO-induced biofilm dispersal.....	146
5.4	Future work.....	152

Chapter 6: Conclusions	155
Appendix A	159
Appendix B	161
Appendix C	165
List of References.....	169

List of Figures

Figure 1.1 - A biofilm growing on a metal surface, imaged using scanning electron microscopy.	2
Figure 1.2 - Each stage of biofilm development, represented as a cartoon with an associated photomicrograph image.	4
Figure 1.3 - Mechanisms underpinning increased antimicrobial resistance of bacteria within a biofilm.....	7
Figure 1.4 – Overview of c-di-GMP signalling.	13
Figure 1.5 - The GGDEF domain of <i>C. crescentus</i> PleD (PDB 2V0N).	15
Figure 1.6 - The GGDEF domain active half-site, or A-site, of PleD (PDB 2V0N).	16
Figure 1.7 - The proposed mechanism of GGDEF catalysed diguanylate cyclisation.	17
Figure 1.8 - Activation of PleD.	19
Figure 1.9 - The structure of c-di-GMP inhibited PleD (PDB 2V0N).	20
Figure 1.10 - The structure of <i>Thiobacillus denitrificans</i> TBD1264 ^{EAL} (PDB 3N3T).	21
Figure 1.11 - EAL-PDE conserved residues within <i>T. denitrificans</i> TBD1265 (PDB 3N3T).	22
Figure 1.12 - Two-metal catalysed hydrolysis of c-di-GMP by EAL domains.	23
Figure 1.13 - EAL domain dimerisation is required for PDE activity.	25
Figure 1.14 - The structure of the HD-GYP domain of <i>Persephonella marina</i> GH (PDB 4MDZ)..	26
Figure 1.15 - Coordination of c-di-GMP and metal ions within the HD-GYP of <i>P. marina</i> GH.	27
Figure 1.16 - Proposed mechanism for HD-GYP catalysed c-di-GMP hydrolysis.	28
Figure 1.17 - The structure of <i>X. campestris</i> Orn (PDB 2GBZ).	29
Figure 1.18 - Reaction mechanism proposed to be shared between <i>E. coli</i> DNA polymerase III and Orn.	30
Figure 1.19 - The PAS domain-fold of <i>Bradyrhizobium japonicum</i> protein FixL (PDB 1DRM).....	33

Figure 1.20 - Regulation of Dos activity by changes in the redox state of the iron ion within PAS domain associated haem.	34
Figure 1.21 - Conformational changes in FixL upon oxygen binding.	36
Figure 1.22 - The haem PAS domain containing protein Aer2.	38
Figure 1.23 - Proposed BdlA processing and signalling pathways.	42
Figure 1.24 - Proposed effect of NO or altered redox potential on MucR.	45
Figure 2.1 – Graphical overview of designed MucR and NbdA gene fragments cloned at OPPF-UK and further detailed in Table 2.1.	49
Figure 2.2 – Vector map of the pOPINE-3C-GFP vector.	50
Figure 2.3 - The phase diagram.	65
Figure 2.4 - Miller planes.	67
Figure 2.5 - Graphical representation of Bragg's law.	68
Figure 2.6 - The Ewald sphere.	69
Figure 3.1 - High throughput fragment design and expression of MucR and NbdA.	75
Figure 3.2 - FSEC analysis of MucR ⁷⁻²⁶⁶	77
Figure 3.3 - FSEC analysis of MucR ⁴⁻⁶⁸⁵	78
Figure 3.4 - Purification of MucR ⁷⁻²⁶⁶	79
Figure 3.5 - Relative fluorescence of MucR ⁷⁻²⁶⁶ expression in KRX cells.	81
Figure 3.6 - Relative fluorescence of MucR ⁷⁻²⁶⁶ expression in Lemo21(DE3) cells.	81
Figure 3.7 - Improvement in protein yield through optimised purification.	82
Figure 3.8 - Analysis of MucR ⁷⁻²⁶⁶ oligomerisation states with SEC-MALS.	83
Figure 3.9 - Identification of transmembranous cysteine residues in MucR ⁷⁻²⁶⁶	84
Figure 3.10 - First derivative thermostability profiles of CPM-MucR ⁷⁻²⁶⁶ solutions given in Table 3.1.	86
Figure 3.11 - Purification of cytosolic c-di-GMP metabolising domains from MucR and NbdA.	88

Figure 3.12 - Purification of MucR ^{EAL}	89
Figure 3.13 - Purification of MucR ^{GGDEF-EAL}	90
Figure 3.14 - Enzymatic activity of MucR ^{EAL} and MucR ^{GGDEF-EAL}	91
Figure 3.15 - A MucR ^{EAL} crystal, diffracted at ID23-2, ESRF.	92
Figure 3.16 - The structure of MucR ^{EAL} in complex with c-di-GMP.	94
Figure 3.17 - Coordination of c-di-GMP and magnesium ions within MucR ^{EAL}	95
Figure 3.18 - The dimerisation interface of MucR ^{EAL} contrasts to that of PA3825 ^{EAL}	97
Figure 3.19 - The M3 metal site is concurrent with bound product within EAL-PDEs.	101
Figure 3.20 - Proposed regulatory steps involved in hydrolysis of c-di-GMP by EAL-PDEs.....	102
Figure 3.21 - Reaction mechanism of nuclease P1.	103
Figure 4.1 - Chemotaxis pathways of <i>P. aeruginosa</i>	113
Figure 4.2 - UV-Vis spectra of purified BdIA fragments.	115
Figure 4.3 - UV-Vis spectra of BdIA ¹⁻⁴¹⁷ expressed in the presence of haem precursors.	116
Figure 4.4 - UV-Vis spectra of haem-bound BdIA.	118
Figure 4.5 - UV-Vis spectra of BdIA ¹⁻⁴¹⁷ , scaled by absorbance at 280 nm, showing improved haem incorporation.	119
Figure 4.6 - UV-Vis spectra of BdIA ¹⁻¹²⁷ in the presence and absence of NO.	121
Figure 4.7 - FTIR spectra of interactions between BdIA ¹⁻¹²⁷ and NO.....	123
Figure 4.8 - FTIR spectra of interactions between BdIA ¹⁻¹²⁷ and NO over time.	124
Figure 4.9 - A crystal of BdIA ¹⁻⁴¹⁷ diffracted at Diamond Light Source I03.	125
Figure 4.10 - The structure of BdIA ¹⁻⁴¹⁷	129
Figure 4.11 – BdIA ¹⁻⁴¹⁷ coloured by B-factor and oriented similarly to Figure 4.10.	129
Figure 4.12 - Phosphorylation of Tyr238 is likely to alter interaction within the MA domain..	130
Figure 4.13 - The BdIA cleavage site.	132
Figure 4.14 - His42 may act as the axial ligand for haem coordination in PAS 1.	133

Figure 4.15 - Disulphide bond within BdlA PAS 1.	134
Figure 4.16 - Gly31 may be involved in the formation of PAS tetramers.	135
Figure 5.1 - Maximum likelihood phylogenetic analysis of all PAS domains within PAO1 with PAS domains that have been characterised structurally in complex with their physiological ligand.	140
Figure 5.2 - Example assignment of cofactor binding based on conservation of interacting residues.	145
Figure 5.3 - NO binding to the haem PAS of PA0176 (Aer2) induces a change in motility.....	148
Figure 5.4 - Proposed model of how PA0285 and PA1181 may function in NO-induced biofilm dispersal.	150
Figure 5.5 - Binding of NO to the haem PAS domains of PA5442 could induce dispersal by inducing homodimerisation.....	151

List of Tables

Table 1.1 - The DGC and PDE proteins with <i>P. aeruginosa</i> PAO1.	10
Table 2.1 - PCR conditions used to clone transmembranous MucR and NbdA fragments	51
Table 2.2 - Composition of CPM thermostability reactions	56
Table 2.3 - PCR conditions used to amplify soluble protein fragments from genomic PAO1 genome.....	57
Table 3.1 - Composition of CPM thermostability reactions	85
Table 3.2 - Crystallographic data and refinement statistics for MucR EAL.	93
Table 3.3 - Dimerisation architectures of EAL domains.	98
Table 3.4 - Classification of all EAL domain structures in the PDB to date.	99
Table 4.1 - Crystallographic data and refinement statistics for BdIA ¹⁻⁴¹⁷	126
Table 4.2 – Crystallographic data statistics for the BdIA ¹⁻⁴¹⁷ native dataset, divided into resolution shells.	128
Table 5.1 - Bacterial PAS domains structurally characterised with their natural ligand	139
Table 5.2 - PAS domain cofactor binding identified on the basis of conserved interacting residues	144
Table 6.1 - PCR primers used to amplify genes, as described in section 2.1.10.	159
Table 6.2 - Composition of Morpheus Custom Optimisation Hanging Drop Screen.	161
Table 6.3 - Components of Morpheus crystallisation screen stock solutions used to produce Morpheus Custom Optimisation Hanging Drop Screen.....	163
Table 6.4 - All PAS domains within the PAO1 genome, used for analysis in chapter 5.....	165

DECLARATION OF AUTHORSHIP

I, Andrew Hutchin declare that this thesis, entitled “Molecular mechanisms of nitric oxide induced *Pseudomonas aeruginosa* biofilm dispersal,” and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:
Bellini *et al.* 2017

Signed:

Date:

Acknowledgements

I would like to thank Ivo Tews, Martin Walsh and Jeremy Webb for the supervision and tireless support offered to me throughout the project. I am grateful to the University of Southampton and Diamond Light Source for funding all of the work enclosed hereafter. I would also like to thank the Oxford Protein Production Facility (in particular Louise Bird, Ray Owens and Heather Rada) and the Membrane Protein Lab (Isabel Moraes, James Birch, Maria Rosa, Shilpi Sheth) for technical advice and provision of consumables during work at these facilities.

I would like to thank Yuming Cai for continuous discussion on the topic of nitric oxide induced biofilm dispersal, performance of gene knockout experiments and access to unpublished physiological data included in Chapter 5. I would like to thank all members of the Tews lab group (Jack Craddock, Jordana Griffiths, Moritz Machellet, Christian Orr, Curtis Phippen, Matthew Rodrigues and Emma Sutton) for advice on numerous crystallographic and lab techniques. Further thanks are also offered to members of the Doyle, East, Werner and Williamson groups at the University of Southampton for frequent assistance and discussion. I would like to thank Christopher Holes and Stuart Findlow for their work in the provision of the crystallography facility at the University of Southampton, as well as Neville Wright for access and instruction with the University of Southampton biophysics facility. I would like to thank project students Bhavik Barokhia, Sally Frost, Nick Tsimplis and Antonia Turberville for the assistance they have each provided in the lab that has contributed to this work. I would also like to thank the Walsh group at Diamond Light Source, particularly Dom Bellini and Sam Horrell for advice on numerous occasions and for facilitating collaboration on the EAL domain work carried out in chapter 3.

I would like to thank numerous beamline scientists and support staff at ESRF and Diamond Light Source. I would like to thank Neil Hunt and Niall Simpson for supervision whilst performing FTIR measurements and assistance in analysing these spectra.

I would like to thank Mike Hough at the University of Essex, Mark Shepherd at the University of Kent, Gwyndaff Evans at Diamond Light Source and Bill Keevil at the University of Southampton for advice at various points throughout the project throughout the project.

Finally, I would like to thank my wife, family and friends for their constant support and encouragement.

Definitions and Abbreviations

2D-PAGE	Two-dimensional polyacrylamide gel electrophoresis
ALA	δ -aminolevulinic acid
A₆₀₀	Absorbance of 600 nm light, as used to determine cell optical density
BdIA	The protein encoded by Biofilm dispersal locus A, PA1423
Bis-pNPP	Bis(p-nitrophenyl) phosphate
CAMP	Cyclic adenosine monophosphate
CD	Circular dichroism
Che	Chemotaxis
C-di-AMP	Cyclic dimeric adenosine monophosphate
C-di-GMP	Bis-(3'-5') cyclic dimeric guanosine monophosphate
CPM	N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide
DEA NONOate	Diethylammonium (Z)-1-(N,N-diethylamino)diazene-1,2-diolate
DipA	Dispersion-induced-phosphodiesterase A
DGC	Diguanylate cyclase
DMSO	Dimethyl sulphoxide
dNTPs	An equal mix of adenine, cytosine, guanine and thymine deoxyribonucleotide triphosphates
Dos	Direct oxygen sensor
EAL	Protein domain named after characteristic amino acid motif Glu-Ala-Leu
EPR	Electron paramagnetic resonance
EPS	Extracellular polymeric substance
FPLC	Fast protein liquid chromatography
FTIR	Fourier transform infrared spectroscopy

GFP	Green fluorescent protein
GGDEF	Protein domain named after characteristic amino acid motif Gly-Gly-Asp-Glu-Phe
HAMP	Protein domain named after the systems in which it is found: Histidine kinase, Adenylyl cyclase, Methyl-accepting chemotaxis protein and Phosphatases
HD-GYP	Domain named after two characteristic amino acid motifs, His-Asp and Gly-Tyr-Pro
HPLC	High-pressure liquid chromatography
HNOX	Haem-nitric oxide/oxygen-binding
IMAC	Immobilised metal-affinity chromatography
IR	Infrared
ITC	Isothermal titration calorimetry
Lap	Large adhesion protein
LB	Lysogeny broth
LPS	Lipopolysaccharide
MA	Methyl-accepting chemotaxis domain
MAHMA NONOate	6-(-2-Hydroxy-1-methyl-2-nitrosohydrazine)-N-methyl-1-hexanamine
MALS	Multi-angle light scattering
MCP	Methyl-accepting chemotaxis protein
MPD	2-methyl-2,4-pentanediol
MPL	Membrane Protein Lab, located within Diamond Light Source
MucR	Mucoid alginate regulator
NbdA	Nitric oxide-induced biofilm dispersion locus A
NO	Nitric oxide
NTA	Nitrilotriacetic acid, a metal chelating ligand cross-linked to agarose

OPPF-UK	Oxford Protein Production Facility UK
(p)ppGpp	Guanosine tetraphosphate and guanosine pentaphosphate,
PAS	Protein domain named after systems in which it was originally identified: PERIOD (Per) - Aryl hydrocarbon nuclear translocator (ARNT) - Single-minded (SIM)
PDE	Phosphodiesterase
PEG	Polyethylene glycol
Pel	Polysaccharide produced by proteins encoded by the <i>pel</i> (pellicle deficient) operon
PSI	Pounds per square inch
Psl	Polysaccharide produced by the polysaccharide synthesis locus
pGpG	5'-Phosphoguanylyl-(3'-5')-guanosine
RbdA	Regulation of biofilm disposal A
RCaH	Research Complex at Harwell
REC	Response regulator receiver protein domain
SAD	Single-wavelength anomalous dispersion
SAXS	Small-angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SNP	Sodium nitroprusside
SOC	Super optimal broth with catabolite repression
TPP	Thiamine pyrophosphate
UV-Vis	Ultraviolet-visible light spectroscopy
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Chapter 1: Introduction

1.1 Introduction

The Gram-negative bacterium *Pseudomonas aeruginosa* has recently been identified as a species of bacteria urgently requiring new antibiotics¹. *P. aeruginosa* is capable of growth in a wide range of different conditions and habitats, including soil, coastal marine habitats as well as on plant and animal tissues^{2,3}. This diversity in viable habitat differs from the widely studied *Escherichia coli*, which is the predominant aerobe in the vertebrate gut but is not viable in most habitats outside of the host⁴. The capacity for *P. aeruginosa* to survive in such a range of environments is thought to be facilitated by high levels of genetic complexity and its large genome of 6.2 million base pairs (1.7 million base pairs larger than that of *Escherichia coli*)^{2,3,5}. Amongst other environmental niches, *P. aeruginosa* is an opportunistic human pathogen, able to infect patients suffering from burns, immunosuppression and cystic fibrosis². One way in which *P. aeruginosa* can adapt to environmental pressures is to adopt a communal, surface-attached biofilm phenotype. Within a biofilm *P. aeruginosa* is refractory to treatment with antibiotics^{6–12}, making removal of *P. aeruginosa* infection difficult and often causing chronic infection. Reduced pulmonary function caused by chronic infection of *P. aeruginosa* is the largest cause of mortality in cystic fibrosis patients^{2,3,13}. The clinical benefits of controlling *P. aeruginosa* biofilm dispersal have made the bacteria a model organism for studies into biofilm regulation.

An understanding of biofilm regulation has implications not only for the clinical outcome of chronic infections, but also for the production of biofuels, water contamination and pipe corrosion^{14–20}. This work focuses on how *Pseudomonas aeruginosa* biofilms are regulated, with a particular focus upon how these biofilms disperse upon perception of nitric oxide, a non-toxic diatomic gas that could be used in combination with antibiotics as part of a novel anti-biofilm therapeutic course^{21–23}.

1.2 Biofilms and cyclic dimeric guanosine monophosphate

1.2.1 Biofilms

Biofilms, such as that shown in Figure 1.1, are a sessile community of microbial cells attached to one another or to an interface⁶. Bacteria within biofilms are embedded in a matrix of extracellular polymeric substances of their own production and display a distinct phenotype with altered gene transcription and growth rate⁶.

Confocal laser scanning microscopy (CLSM) has revealed that mature biofilms are a heterogeneous collection of bacterial cells (approximately 15% of total volume) and extracellular matrix (approximately 85%). The cells act as microcolonies located in “mushroom”-shaped towers surrounded by the matrix and are spaced around open-water channels. With few differences, this biofilm structure is conserved across different species and conditions^{6,24–26}.

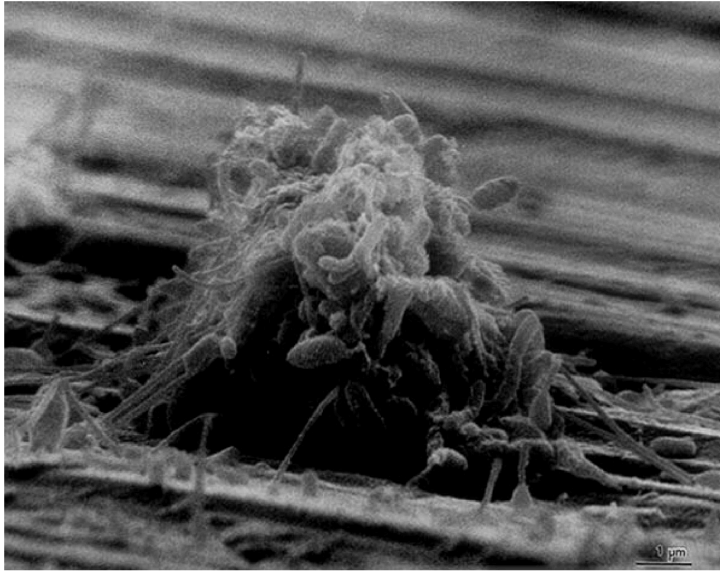


Figure 1.1 - A biofilm growing on a metal surface, imaged using scanning electron microscopy.

Taken from Donlan and Costerton, 2002⁶.

1.2.2 The biofilm lifecycle

Studies using proteomics and microscopy into the transition between planktonic and biofilm bacterial lifestyle break this process into five separate stages: reversible attachment, irreversible attachment, maturation-1, maturation-2 and dispersion, with dispersed cells able to return to the planktonic phenotype²⁷. The five-stages of biofilm development are shown in Figure 1.2.

Reversible attachment is the initial step in biofilm formation, occurring when the cell pole contacts transiently with an interface, which can be observed using a light microscope²⁷. Analysis of *Pseudomonas aeruginosa* protein extract through two-dimensional polyacrylamide gel-electrophoresis (2D-PAGE) demonstrates no change in levels of protein expression between reversibly attached and planktonic bacteria²⁷. Microscopically, cells can be observed to reorient, begin attachment to one another and to cluster during irreversible attachment, which follows reversible attachment. As part of irreversible attachment the flagella are lost and flagellum-mediated motility ceases. The transition between reversible and irreversible attachment is caused by alterations in gene expression²⁷. Irreversible attachment is so called because bacteria can no longer readily detach from the surface, and instead go on to form a biofilm²⁸.

Maturation-1 follows irreversible attachment, when cells can be observed to form layers and cell clusters reach a thickness in excess of 10 μm and the extracellular matrix is produced. Maturation-1 is also associated with alterations in gene expression, with 2D-PAGE identifying upregulated expression of 540 proteins and downregulated expression of 45 proteins, of the 1884 *P. aeruginosa* proteins detected²⁹. Amongst these upregulated genes are a number of genes associated with anaerobic respiration, suggesting that during maturation-1 oxygen levels within the biofilm start to become limiting. Also upregulated are numerous transcription regulators and secreted factors, promoting the transition into maturation-2^{27,29}. During maturation-2 the biofilm reaches its maximum thickness (up 100 μm thick), with cells remaining non-motile and cell clusters reaching their maximum size. It is worth noting that during maturation-2 the majority of the cells in the biofilm are not directly attached to the surface of initial attachment and are instead attached to other bacteria in the biofilm, observed as removal of cell clusters from the surface²⁷. In maturation-2 more than 50% of proteins detected through 2D-PAGE have altered levels of expression when compared to the planktonic form, and approximately 585 proteins demonstrate a more than 6-fold change in the level of expression in comparison to maturation-1²⁷. Upregulation of numerous proteins involved in small molecule uptake, the Krebs cycle and amino acid metabolism suggests that there are further changes in energy and carbon metabolism during maturation-2^{27,29}.

The final step in the biofilm lifecycle is dispersal, where bacteria in the middle of cell clusters become motile once more and swim from the cell cluster into the bulk liquid, leaving other non-motile cells surrounding hollow centres. This can be observed through the rearrangement of cell clusters and the formation of pores within the biofilm²⁷. It is thought that this dispersal allows the dispersing bacteria to gain access to more nutrients (which could range from carbon and nitrogen sources to oxygen) whilst also increasing the nutrient exposure of the bacteria remaining in the cell cluster. Once dispersed, bacteria can then revert to the planktonic behaviour once more. It is therefore not surprising that the levels of gene-expression detected by 2D-PAGE from dispersion stage biofilms more closely resemble those of planktonic bacteria than maturation-2²⁷.

5 stages of Biofilm Development

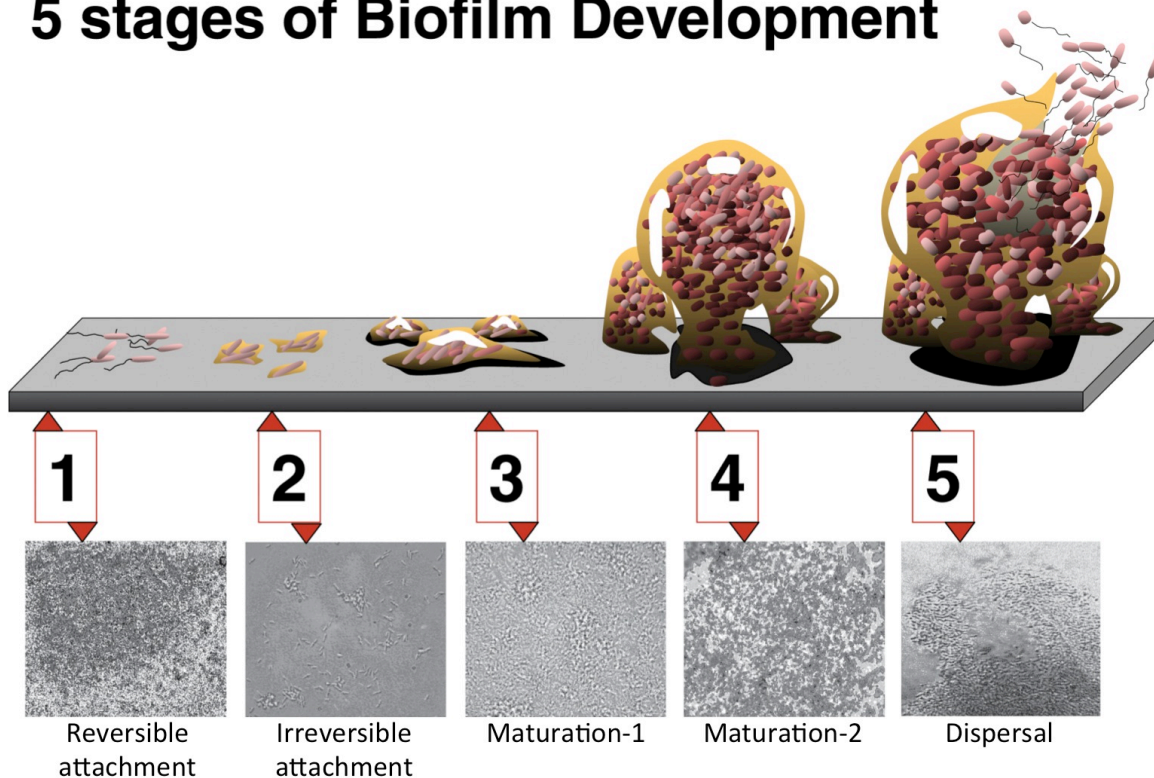


Figure 1.2 - Each stage of biofilm development, represented as a cartoon with an associated photomicrograph image. Image credit to D. Davis, taken from Monroe, 2007³⁰.

1.2.3 Biofilm extracellular polymeric substance

One of the most distinctive features of a biofilm is the presence of an extracellular polymeric substance (EPS) or matrix, which accounts for over 90% of the mass of most dry biofilms (in contrast to the bacteria themselves)³¹. The EPS is essential for surface colonisation and is self-produced by the bacteria. The shape and composition of the EPS is dependant upon the stage of biofilm development, nutrient availability, social competition and physical environmental factors like shear stress^{6,20,27,32-34}. Although a high energetic cost is required to establish and adapt the EPS³⁵, the matrix itself can contribute to capture and digestion of resources as well as offering protection from desiccation and antimicrobial agents (discussed further in section 1.2.4)^{20,32,36-40}.

During the early stages of *Pseudomonas* biofilm development polysaccharides are the main constituents of the EPS⁴¹. The main forms of polysaccharide within the biofilm EPS are capsular polysaccharides, such as alginate and levan, and aggregative polysaccharides, such as Psl, Pel and cellulose⁴¹. Capsular polysaccharides provide structure to the biofilm, as well as participating in water and nutrient retention⁴²⁻⁴⁴. Aggregative polysaccharides play a role in the adhesion of cells

within the biofilm to a surface and between cells^{41,45–48}. Both varieties of polysaccharides are also implicated in protecting the biofilm from the host immune system^{41,46,47,49}.

As *Pseudomonas* biofilms mature, nucleic acids make up a greater proportion of the EPS, with DNA making up a large component of the matrix as biofilms progress from maturation-1 into maturation-2, but late into maturation-2 DNA within the matrix is localised to the caps within the biofilm⁵⁰. Due to the structural similarities between DNA and alginate it is possible that DNA within the EPS has a similar role to capsular polysaccharides, particularly as it has been implicated in maintenance of biofilm structure and cell-cell adhesion, but all roles of extracellular DNA in a biofilm are yet to be determined^{50,51}.

A number of other polysaccharide-containing components are also present throughout the lifetime of the *Pseudomonas* biofilm; including lipopolysaccharide (LPS) and cyclic β -glucans. Although many of these polysaccharide-containing components have undefined functions it is thought that LPS may serve to mediate attachment and dispersal while cyclic glucans may have a role in resistance to aminoglycoside antibiotics^{52–55}. Numerous proteins are also present within the EPS, which have roles in the regulation of biofilm structure^{53,56}. Rhamnolipids are also produced within biofilms, although their function is yet to be elucidated^{41,57}. As the exact composition of the EPS varies depending on the bacterial species present within the biofilm and the conditions under which the biofilm is formed, studies are now on-going to examine the EPS and behaviour of bacteria within multi-species biofilms, which may be more representative of naturally formed biofilms^{6,20,41,43,44,58–60}.

1.2.4 Biofilms are tolerant to antimicrobials

Biofilms act to protect the enclosed bacteria from antibacterial agents, with biofilms 10-100 times more tolerant to many antimicrobials than planktonic bacteria^{6–12}. This tolerance to antimicrobials is facilitated through a number of mechanisms including slow growth rates, altered protein expression and decreased antibiotic diffusion rates⁶¹.

Numerous studies have demonstrated a reduced rate of antibiotic perfusion throughout biofilms, but the extent of this delayed perfusion is not the same across all types of antimicrobial. There are two suggested mechanisms of how the biofilm EPS may alter diffusion rates of antibiotics, either by providing a barrier that restricts diffusion or by directly reacting with the antibiotic^{6,62–67}. Despite altered susceptibility and bacterial exposure to certain antibiotics and antimicrobials, the EPS has been demonstrated to not present an impenetrable barrier to most antimicrobials^{7,68,69}.

Chapter 1

Slow growth of bacteria within a biofilm is thought to lead to reduced antibiotic susceptibility. For many antibiotics, effective use has been demonstrated to be dependant upon rapid proliferation^{70–73}. It is therefore logical that the reduced growth rate within a biofilm can confer tolerance to some antibiotics. This theory has been tested by growing planktonic *E. coli* at an equivalent rate to that observed in a biofilm, with both biofilm and slow-growing planktonic cells displaying equivalent susceptibility to the quaternary ammonium antiseptic cetrимide⁷⁴. Similar observations have been made with different bacteria and different antibiotics^{75–78}.

The final way in which biofilms display increased antibiotic tolerance is through altered metabolism and gene expression. This can include activating the stress response (which lowers nutrient and antibiotic uptake), increasing expression of antibiotic degrading enzymes (such as β -lactamases) and changing the expression levels of outer membrane proteins including drug efflux pumps, capable of actively reducing the concentrations of cytotoxic drugs within the bacteria to non-lethal levels^{7,79–84}. The main mechanisms for increased antimicrobial tolerance within biofilms are summarised in Figure 1.3²⁰.

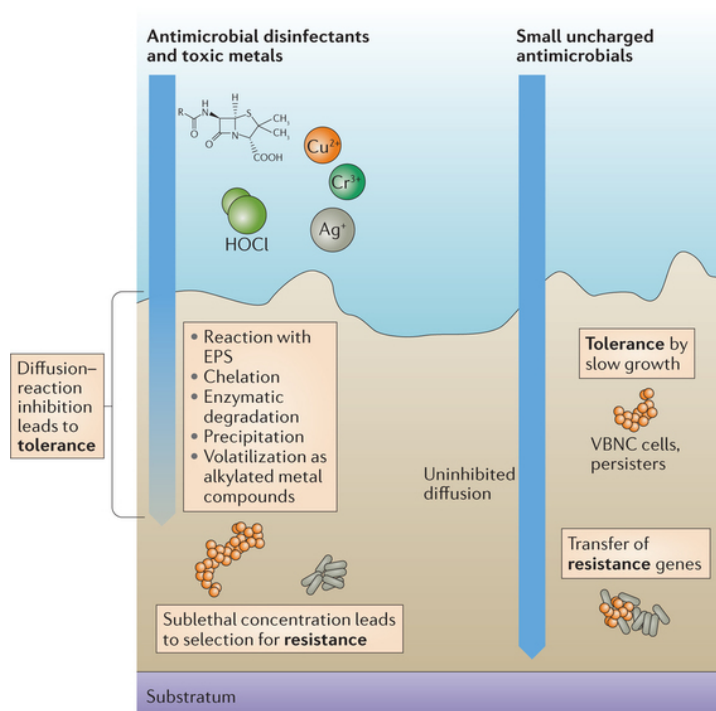


Figure 1.3 - Mechanisms underpinning increased antimicrobial resistance of bacteria within a biofilm. The EPS (shown in brown) acts as a diffusion barrier to numerous antimicrobial agents that interact with the EPS directly or are degraded by extracellular enzymes within the EPS. The EPS does not inhibit the diffusion of all antimicrobial agents, however other mechanisms such as slow growth can confer tolerance to these antimicrobials. As bacteria within a biofilm are exposed to antibiotics without lethal effect, mutations conferring resistance are allowed to develop, and to rapidly transfer to other bacteria within the densely populated biofilm. Taken from Flemming et al., 2016²⁰.

1.2.5 Biofilm formation and dispersal can be controlled by external stimuli, including nitric oxide

The transition between virulent and biofilm phenotypes can be induced by a number of environmental stimuli, with the nature of the altered behaviour varying between species. For example, increased growth medium osmolarity was demonstrated to induce biofilm formation in *Escherichia coli*, but can inhibit biofilm formation in *P. fluorescens*^{85,86}. As *Pseudomonads* are known to be able to deal with larger variation of environmental osmolarity than *E. coli*⁸⁷, it is possible that when *E. coli* is forced into a self-preserving biofilm phenotype *P. fluorescens* is able to make the most of the reduced competition and looks to colonise new environments. However, many environmental stimuli are believed to induce similar phenotypes across environmentally diverse bacterial species. For example in several species of *Enterobacteriaceae*, biofilm formation is repressed by high glucose levels, while biofilm dispersal can be induced within the more

environmentally diverse *P. aeruginosa* by increasing carbon substrate availability^{88,89}. Another example is the observation that carbon repression and storage can be manipulated to influence biofilm regulation in *E. coli* and *P. aeruginosa*, demonstrating the shift in nutrient degradation pathways that occurs within a biofilm^{90,91}. Further stimuli that can regulate biofilm behaviour include high concentrations of tryptophan, which is associated with bacteria in the stationary phase and restricts *P. aeruginosa* biofilm formation^{92–94}. *P. aeruginosa* biofilm formation is also influenced by iron availability, a characteristic exploited by the human innate immune response⁹⁵.

Regardless of the stimulus and the species, biofilm behaviour is controlled by intrinsic levels of bis-(3'-5') cyclic dimeric guanosine monophosphate, c-di-GMP^{96,97}. Initially identified as activating cellulose synthase in *Gluconacetobacter xylinus*, c-di-GMP is now implicated in virulence, adhesion and motility in a number of bacterial species through action at transcriptional, translational and posttranslational levels^{98–101}. High levels of c-di-GMP induce the change from a planktonic lifestyle to a biofilm phenotype. Upon dispersal, intracellular levels of c-di-GMP decrease and bacteria adopt a dispersal phenotype, a behaviour common to all bacterial species^{89,96,98,102,103}.

Non-lethal (nanomolar) doses of nitric oxide (NO) donors, have been shown to reduce the population of bacteria in biofilms, identifying NO as a stimulus able to induce biofilm dispersal in *P. aeruginosa*²¹. Using sodium nitroprusside (SNP) as a model nitric oxide donor, the extent of biofilm dispersion correlated to the concentration of NO donor between 25 nM and 500 nM. However, upon exposure of millimolar concentrations of SNP the biofilm biomass increased compared to the control sample, with a corresponding decrease in planktonic population²¹. At the time of the investigation it was suggested this might be due to the conversion of NO into NO₂⁻ and NO₃⁻, which would promote anaerobic metabolism within the biofilm. The presence of NO would therefore mark cells within the biofilm as oxygen limited, making it advantageous for them to disperse from the biofilm. Unsurprisingly, quantification of c-di-GMP extracted from biofilm *P. aeruginosa* cells and cells 1 hour after NO-induced dispersal demonstrated that NO induced biofilm dispersion is linked to intracellular c-di-GMP levels¹⁰⁴.

1.2.6 C-di-GMP regulation in *P. aeruginosa*

Levels of c-di-GMP within the bacterial cell are regulated by diguanylate cyclases (DGCs) and phosphodiesterases (PDEs). One class of domain with DGC activity, capable of synthesising c-di-GMP from two GTP molecules, has been identified and is known as the GGDEF domain, named after the characteristic amino acid sequence motif required for enzymatic activity^{105–107}. A subset of GGDEF domains have also recently been implicated in the formation of the

dinucleotides cyclic di-AMP and cyclic AMP-GMP¹⁰⁸. There are two main types of domain that possess PDE activity capable of c-di-GMP hydrolysis (also named after characteristic amino acid sequence motifs), the EAL domain, able to degrade c-di-GMP to the linear molecule pGpG, and the HD-GYP domain, able to catabolise c-di-GMP or pGpG to GMP^{109–111}. The linear hydrolysis product pGpG can also be degraded to GMP by oligoribonucleases^{112,113}. All of these c-di-GMP metabolising domains are ubiquitous between bacterial kingdoms and are discussed in more detail in section 1.3⁹⁸.

It is often found that proteins possess both DGC and PDE domains, although in a number of cases one of these is inactive and degenerate (frequently functioning as sensor domain, discussed in section 1.2.7)¹¹⁴. Many of these proteins contain at least one signal-receptor domain or membrane-localising hydrophobic regions, including the majority of DGC and PDE domain containing proteins in *P. aeruginosa*^{114,115}. It is hypothesised that these sensory domains receive an environmental input and this causes the phenotypic change¹¹⁶. This would allow the bacteria to adapt to a biofilm state when required, but be able to revert to a virulent lifestyle when any threat to this phenotype is not present.

Within the genome of *P. aeruginosa* strain PAO1, 41 genes were identified that encoded proteins with GGDEF domains, EAL domains, both GGDEF and EAL domains or HD-GYP domains, shown in Table 1.1¹¹⁴. Transposon insertion mutants have been used to knock out the functionality of GGDEF and EAL domain containing proteins¹¹⁴. Many of the transposon mutants directly affected biofilm formation, either preventing the formation of biofilms or causing the bacteria to produce pellicles (biofilms at the liquid-air interface¹¹⁷), a sign of hyper-biofilm behaviour. By comparing c-di-GMP metabolism with PDE and DGC overexpression the formation of pellicles was shown as correlating to high levels of c-di-GMP¹¹⁴. However, the different mutants did show distinct phenotypes, even for proteins with the same architecture. This is consistent with similar experiments showing different phenotypes for gene knockout mutants of two active DGCs which display distinct regulation of flagellar motility and polysaccharide production¹¹⁸. These experiments demonstrate that c-di-GMP metabolising enzymes do not provide a high level of redundancy, but instead supports a model of different DGC and PDE proteins regulating localised c-di-GMP concentrations responsible for distinct behaviours rather than global levels across the cell. This is further supported by mutations to certain DGCs not affecting biofilm formation despite being previously shown to produce high levels of c-di-GMP¹¹⁴. Such an observation may explain why two different classes of c-di-GMP degrading PDEs exist within the same species, as they may modulate distinct phenotypes.

Table 1.1 - The DGC and PDE proteins with *P. aeruginosa* PAO1.

	Gene	Protein name	DGC motif	PDE motif	DGC-Activity	PDE-Activity
1.	PA0169	SiaD	GGEEF		—	
2.	PA0285	-	GGDEF	ESL	—	—
3.	PA0290	-	GGEEF		—	
4.	PA0338	-	GGEEF		—	
5.	PA0575	-	GGDEF	EAL	—	—
6.	PA0847	-	GGDEF		16	
7.	PA0861	RbdA	GGDEF	ELL	—	—
8.	PA1107	RoeA	GGEEF		35	
9.	PA1120	TbpB / YfiN	GGDEF		134	
10.	PA1181	YegE	GGDEF	ELL	—	—
11.	PA1433	-	RGGEF	KVL	—	—
12.	PA1727	MucR	GGDEF	EAL	24	—
13.	PA1851	-	GGEEF		—	
14.	PA2072	-	GGDEF	EAL	—	—
15.	PA2133	-		ETL	—	71
16.	PA2200	-		EAL		61
17.	PA2567	-	SPTRF	EAL	—	—
18.	PA2572	-		YN-GYP		
19.	PA2771	-	GGEEF		—	
20.	PA2818	Arr		EAL		—
21.	PA2870	-	GGEEF		96	
22.	PA3177	-	GGEEF		—	
23.	PA3258	-	GGDDF	EAL	—	—
24.	PA3311	NbdA	AGDEF	EAL	—	—
25.	PA3343	-	GGEEF		22	
26.	PA3702	WspR	GGEEF		445	
27.	PA3825	-		EVL		27
28.	PA3947	RocR		EVL		100
29.	PA4108	-		HD-GYP		
30.	PA4332	SadC	GGEEF		—	
31.	PA4367	BifA	GGDQF	EAL	—	—
32.	PA4396	-	DEQHF		—	
33.	PA4601	MorA	GGDEF	EAL	—	—
34.	PA4781	-		HD-GYP		
35.	PA4843	GcbA	GGEEF		—	
36.	PA4929	-	GGEEF		—	
37.	PA4959	FimX	GDSIF	EVL	—	—
38.	PA5017	DipA	ASNEF	EAL	—	—
39.	PA5295	-	GSDEF	EAL	—	—
40.	PA5442	-	AGDEF	EAL	—	—
41.	PA5487	-	GGEEF		109	

DGC activity is measured as pmol of c-di-GMP produced per mg of wet cell weight and PDE activity is measured as percent of c-di-GMP degraded upon addition of exogenous nucleotide.

Adapted from Kulasakara et al., 2006¹¹⁴.

1.2.7 C-di-GMP receptors

A number of different c-di-GMP recognition domains have been identified to date. The first of these was the PilZ domain, first identified as common to a number of c-di-GMP regulated systems through bioinformatics studies¹¹⁹ before experimental confirmation of c-di-GMP binding^{120,121}. PilZ domains are the highest affinity c-di-GMP receptors identified to date, with K_d values of 8.4 μM ¹²¹, 5.2 μM ¹²¹, 0.84 μM ¹²⁰, 0.31 μM ¹²² and less than 0.05 μM ¹²³ previously reported from a range of different proteins. Within PilZ domains, c-di-GMP is observed to bind between two distinct sequence motifs, RXXR and (D/N)X(S/A)XXG¹²⁴. Binding of c-di-GMP to PilZ domains brings together the two protein regions containing these motifs and induces downstream signalling¹²⁴. PilZ domains have been observed as stand alone proteins and as part of multi-domain proteins, found in complex with (amongst others) c-di-GMP metabolising domains, chemotaxis signalling domains and DNA-binding domains^{103,119}. Single domain PilZ containing proteins are thought to form complexes following c-di-GMP binding, whilst binding of c-di-GMP to PilZ domains in multi-domain proteins is believed to be a mechanism of allosteric activation. The exact nature of signal propagation from PilZ domains is unclear, with observations of PilZ domains in numerous oligomeric states suggesting that multiple methods may be utilised throughout different signalling pathways^{124–128}.

Degenerate c-di-GMP metabolising domains are also commonly observed as c-di-GMP binding domains, particularly within the context of tandem DGC-PDE domains¹²⁴. Binding of c-di-GMP to a degenerate DGC domain has been shown to regulate PDE activity of *Caulobacter crescentus* CC3396¹²⁹. Degenerate c-di-GMP metabolising domains can also regulate signalling between proteins in response to c-di-GMP binding. Binding of c-di-GMP to the degenerate PDE domain of *P. fluorescens* large adhesion protein D (LapD) propagates conformational change throughout the protein facilitating interactions with the periplasmic protease LapG, preventing LapG from breaking the connections between the cell and the surface of attachment and thus allowing biofilm formation^{130–132}. A further mechanism of c-di-GMP binding exhibited by c-di-GMP metabolising enzymes is that of the autoinhibitory site in DGC domains, which is discussed in section 1.3.1¹³³.

Allosteric regulation of protein activity through c-di-GMP binding has been observed in numerous proteins, even those lacking a distinct c-di-GMP binding domain^{134–147}. One such example is that of protein *C. crescentus* CckA, which in the absence of c-di-GMP acts as a kinase. Binding of c-di-GMP to CckA stabilises the protein in a conformation compatible with phosphatase activity, facilitating reciprocal activity to that in the absence of c-di-GMP¹⁴⁷.

Chapter 1

Two different classes of riboswitch have also been observed to bind c-di-GMP^{148–150}. Riboswitches are non-coding segments of RNA, often found on mRNA, which bind to a distinct small molecule known as an aptamer. The secondary structure of the riboswitch is rearranged upon aptamer binding, altering gene expression. Numerous riboswitches across a range of species have been identified as putatively c-di-GMP binding, with genes regulated by c-di-GMP binding riboswitches thought to include those involved in c-di-GMP metabolism, motility and further regulation of transcription regulation^{124,149,151}.

It is worth noting that co-localisation of c-di-GMP receptors with specific DGCs could provide a model for how different c-di-GMP metabolising enzymes can control different facets of biofilm behaviour¹²⁴. For example, *P. aeruginosa* gene knock-out mutants of either Alg44 or Mucoid Alginate Regulator (MucR), result in reduced levels of alginate production^{152,153}. Anchored to the cytoplasmic membrane, Alg44 is thought to act as a subunit of the alginate polymerisation complex, and contains a c-di-GMP binding PilZ domain required for activity^{152–154}. MucR is also located at the cytoplasmic membrane and contains an active DGC domain^{114,153,155}. It is therefore logical to hypothesise that upon perception of a suitable stimulus, the DGC activity of MucR is increased, increasing the local concentration of c-di-GMP which can then bind to the PilZ domain of Alg44 and promote alginate synthesis^{153,156}. An overview of the range of different phenotypes that are regulated through co-localised c-di-GMP metabolising enzymes and receptors is given in Figure 1.4.

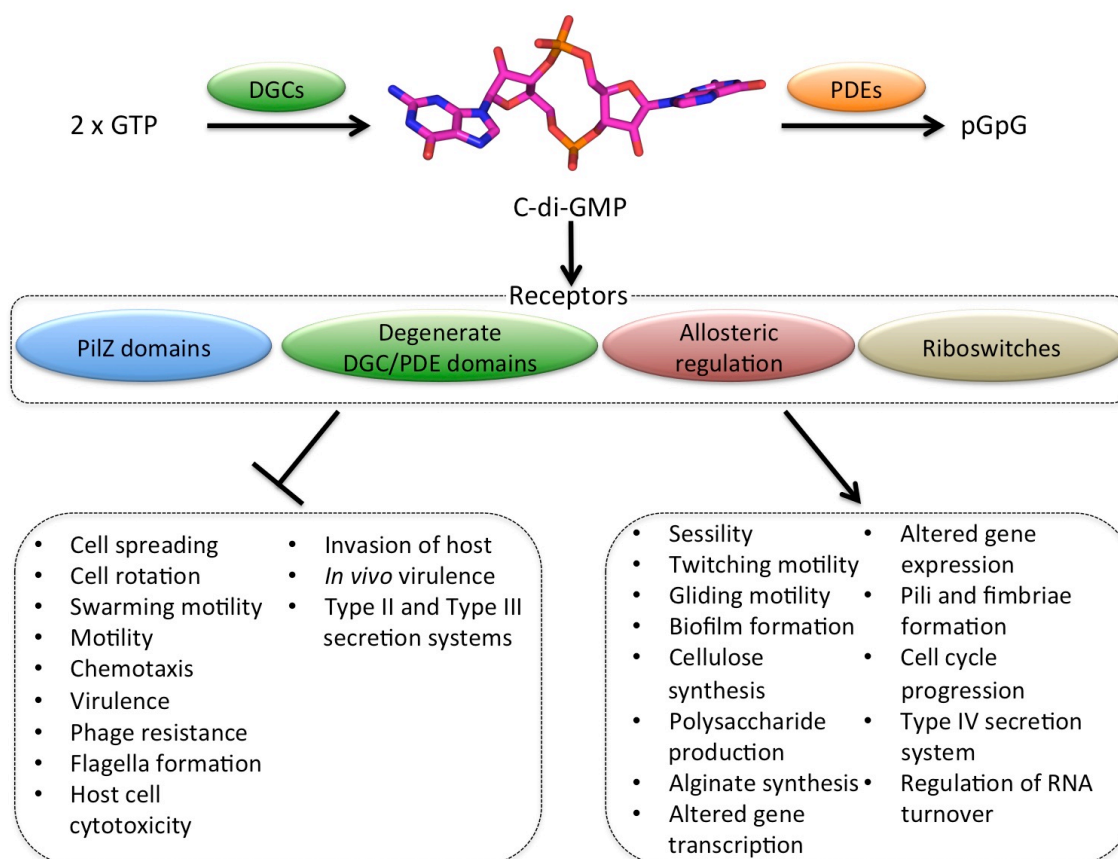


Figure 1.4 – Overview of c-di-GMP signalling. Once synthesised from two GTP molecules, c-di-GMP can bind to receptors located within the proximity of the DGC, each altering distinct phenotypes within the bacteria. Biofilm formation requires signalling to numerous DGC domains, each of which can regulate a specific behaviour to induce large changes in cell lifestyle. Adapted from Römmling 2013¹²⁴ and Sonderman 2012¹⁵⁷.

1.2.8 Other nucleotides involved in biofilm regulation

Although regulation of biofilm behaviour is predominantly tied to levels of intracellular c-di-GMP, other nucleotide signalling systems have been proposed to play a role in the regulation of c-di-GMP metabolism and biofilm formation. One such system is that of cyclic AMP (cAMP), which is produced in response to limited carbon availability within the cell and can alter gene transcription to induce the synthesis of pathways capable of catabolising alternative carbon sources^{158–160}. As part of this regulation of gene transcription, cAMP is thought to regulate the expression of a number of c-di-GMP metabolising enzymes, thus exerting a level of control over c-di-GMP signalling within the cell^{160,161}. CAMP signalling can also directly regulate a range of other cellular systems that overlap with c-di-GMP signalling, such as virulence, flagellum synthesis and even biofilm formation^{88,160,162–165}.

Another nucleotide system linked to biofilm regulation is that of guanosine tetraphosphate and guanosine pentaphosphate, collectively known as (p)ppGpp. The (p)ppGpp signalling network is utilised in instances of intracellular nutrient limitation, as detected through the presence of tRNAs lacking amino acids (indicative of amino acid starvation)^{160,166,167}. Much like cAMP, (p)ppGpp signalling can dramatically alter the gene transcription within the cell, doing so through control of the RNA polymerase σ^5 subunit, which as part of the stress response promotes the transcription of DGCs and the formation of curli fimbriae, an extracellular fibre involved in biofilm formation, as well as regulating chemotaxis and flagella synthesis^{161,168–174}. As such, both of these nucleotides demonstrate an involvement in biofilm regulation, however these are part of more complex signalling cascades, of which biofilm regulation contributes only a small portion.

Recent studies have identified a potential involvement for cyclic di-AMP in biofilm formation within *Streptococcus mutans* and *Staphylococcus aureus*^{175,176}, however further study is required to determine if this behaviour is consistent across all bacterial species and to determine the involvement of cyclic di-AMP with c-di-GMP. Similarly, a role for cyclic GMP in biofilm regulation has also been suggested, although this prospect also requires further investigation¹⁷⁷. Although all of these nucleotides are suggested to have an involvement in the development of certain aspects of biofilm development, intracellular c-di-GMP concentration is still believed to be the overriding determinant of biofilm regulation.

1.3 Structural and functional details of c-di-GMP regulating proteins

1.3.1 GGDEF domain diguanylate cyclases

The first structure of a GGDEF domain was elucidated from the *Caulobacter crescentus* protein PleD¹⁰⁵. The GGDEF domain is formed from a five-stranded β -sheet surrounded by four α -helices in an arrangement of $\beta\alpha\alpha\beta\beta\alpha\beta$, shown in Figure 1.5. The GGDEF motif is located on a loop between the second and third β -strands (also referred to as the β 2- β 3 hairpin) and is part of the active half-site^{105,178–180}.

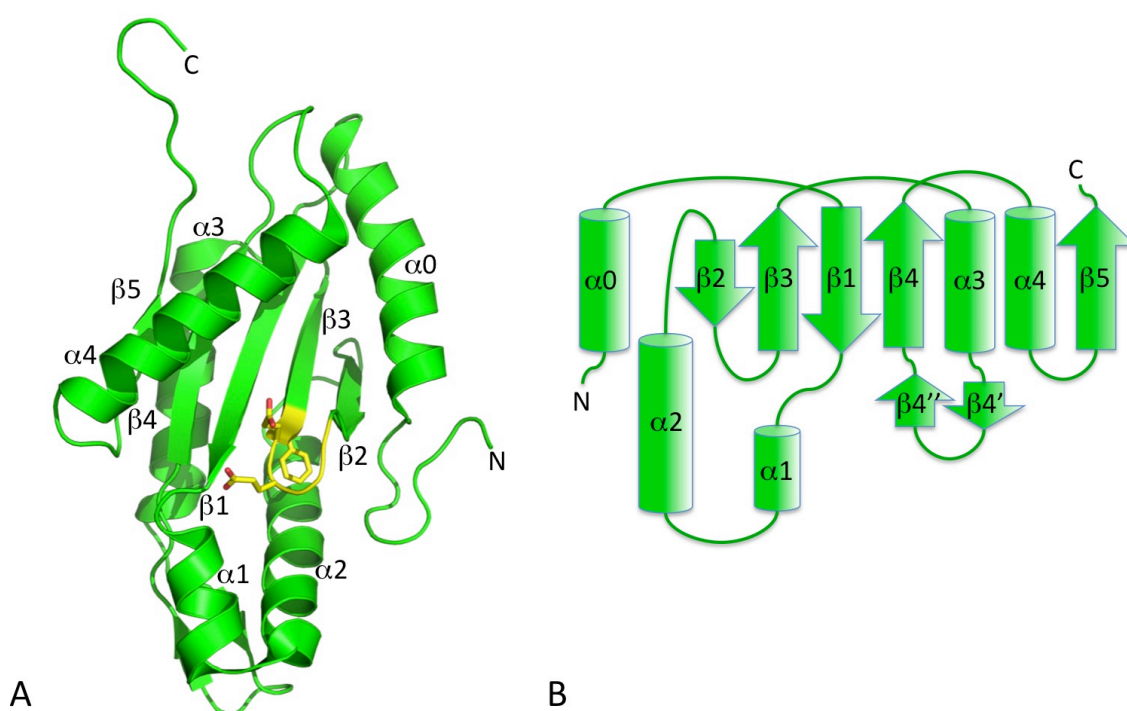


Figure 1.5 - The GGDEF domain of *C. crescentus* PleD (PDB 2V0N). A) Cartoon representation of PleD^{GGEEF}, with the GGEEF motif (equivalent to the GGDEF motif) highlighted as sticks in yellow. B) Secondary structure topology of the PleD GGDEF domain.

Substrate binding within the active half-site of the GGDEF domain, shown in Figure 1.6, was identified using the non-hydrolysable substrate analogue GTP α S¹⁷⁹. Two magnesium ions are coordinated within this binding site. One of these magnesium ions is coordinated by the first glutamic acid residue from the GGEEF motif, E370 in Figure 1.6 (equivalent to the canonical GGDEF motif) and an aspartic acid residue from the first β -strand, D327 in Figure 1.6. Point-mutations have demonstrated that both D327 and E370 are required for activity, although the first glutamic acid in the GGEEF motif can be replaced by an aspartic acid, as seen in the more typical GGDEF^{133,179}. This magnesium ion is close to the α -phosphate at a point where the phosphodiester bond between the α - and β -phosphates can be cleaved. It is worth noting that within the GTP α S bound GGDEF domain structure shown in Figure 1.6 this magnesium ion is not suitably positioned to coordinate the α -phosphate, but that adjustments of the internal torsion angles within the substrate could allow coordination of the α -phosphate by this magnesium without altering the binding observed of the nucleoside or the other phosphates¹⁷⁹. This difference in α -phosphate position has been suggested to be due to the substitution of sulphur in the place of the oxygen within the native substrate¹⁷⁹. An alternative explanation is that the presence of a second substrate (required for c-di-GMP formation) may alter the α -phosphate position to allow magnesium ion coordination¹⁷⁹. The second magnesium ion is also coordinated by these two residues and through interactions with the peptide backbone between β 1 and α 1¹⁷⁹.

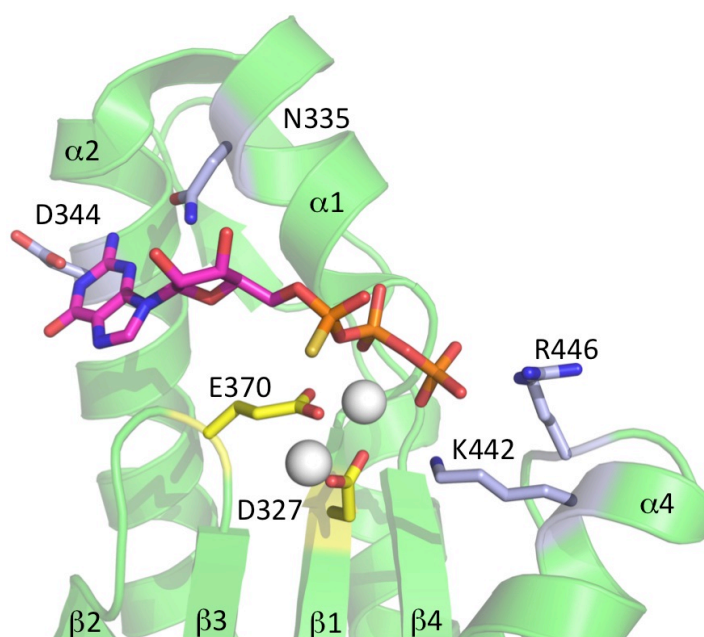


Figure 1.6 - The GGDEF domain active half-site, or A-site, of PleD (PDB 2V0N). Residues highlighted in yellow are involved in the coordination of magnesium ions, shown as grey spheres. Residues highlighted in grey are involved in coordination the GMP analogue GTP α S, shown in magenta.

The GTP α S molecule is coordinated through its guanine base and phosphates. The guanine base is hydrogen bonded to an asparagine on the α 1-helix (N335 in Figure 1.6) and an aspartic acid on the α 2-helix (D344 in Figure 1.6). The γ -phosphate is coordinated by conserved lysine and arginine residues from the loop between α 4 and β 5 (K442 and R446 in Figure 1.6), the magnesium-coordinating glutamic acid of the GGEEF motif (E370), the peptide backbone within the loop between β 1 and α 1 and the second magnesium ion. Only a single GTP binding site can be observed in the structure of the GGDEF domain, so it is logical that for diguanylate cyclase activity a dimer of these domains must form.

Although such a complex has yet to be structurally determined, from analogy to other cyclases it has been suggested that in the catalytically active dimer the two GGDEF monomers align antiparallel, bringing the 3'-hydroxyls of each GTP into the proximity of the α -phosphate from the other GTP^{180–183}. In this arrangement the phosphodiester bond between the α - and β -phosphates could be subjected to nucleophilic attack from the ribose 3'-hydroxyl, allowing the diguanylate cyclase reaction to follow the two-metal-assisted phosphodiester formation mechanism suggested for adenylate cyclases^{179,180,183,184}. The predicted DGC reaction mechanism for condensation of two GTP molecules to form c-di-GMP is shown in Figure 1.7, with amino acids numbered consistent with PleD. Within this dimer Lys332 is suitably positioned to stabilise the transition state, a theory supported by a 100-fold reduction in activity when this is mutated to

alanine¹⁸⁵. The arrangement of nucleotide in this way would fit well with the c-di-GMP bound structure of DgcZ^{GGDEF}, which is thought to represent the product-associated state after catalysis (PDB code 3TVK)¹⁸³. However, in such a model the cause for 3'-hydroxyl deprotonation is unclear. Within adenylate cyclases one of the bound metal ions acts as a Lewis acid to help generate the nucleophile, but within PleD the metal ions are positioned near the phosphates and are likely to be too far away to have such activity, as shown in Figures 1.6 and 1.7^{179,180,183,184}. The catalytic parameters of a number of DGCs have been determined, k_{cat} is predominantly in the order of 1 min⁻¹, although extremes of 0.054 min⁻¹ and 72 min⁻¹ have been observed^{179,180,183,185–188}.

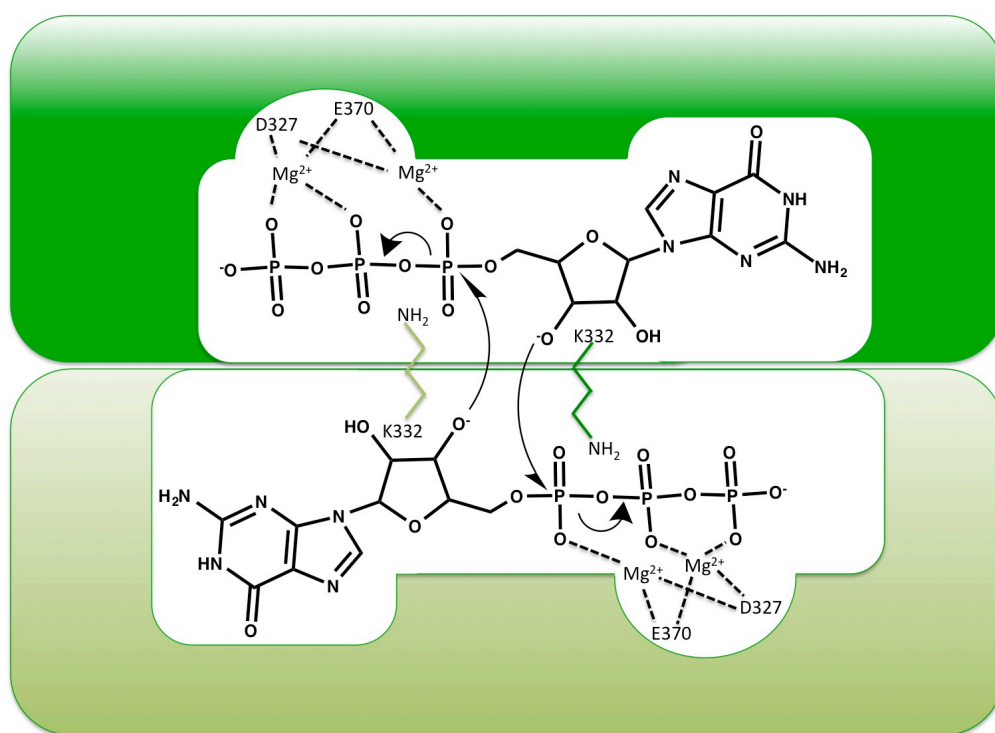


Figure 1.7 - The proposed mechanism of GGDEF catalysed diguanylate cyclisation. The two GGDEF domains (shown as rectangles) align anti-parallel, bringing the 3'-OH from one GTP molecule into close proximity with the α -phosphate of the GTP bound to the other GGDEF domain. The deprotonated 3'-OH can then perform a nucleophilic attack on the α -phosphate, forming a pentavalent phosphoryl transition state, stabilised by magnesium ions and the side chain of Lys332 contributed by the other monomer. The pentavalent phosphoryl transition state is resolved by cleavage of the α - β -phosphodiester bond, resulting in one c-di-GMP molecule and two pyrophosphate molecules thought to be stabilised by the magnesium ions¹⁷⁹. Magnesium ions and coordinating residues are indicated. The side chain of Lys332 is depicted as green sticks with the terminal amide group denoted.

Chapter 1

As many GGDEF domains are found in larger proteins with N-terminal dimerisation domains it is widely perceived that these domains are responsible for recognising a sensory input and then modulating their own dimerisation to control the dimerisation of the GGDEF domain^{178,182,189}. In such a model DGC activity would therefore be regulated through signal perception and dimerisation of the sensor domain. PleD, consisting of two N-terminal CheY-like receiver domains (a form of signalling domain which is phosphorylated upon activation) and a C-terminal GGDEF domain, has been used to study DGC activation. The initially obtained structure of PleD depicted the protein in a non-activated form, with a loose dimer formed between the CheY-like domains of each protein, but the active half-site of the each GGDEF domain on opposite sides of the dimer¹⁰⁵. However, when PleD is activated through phosphorylation of the N-terminal CheY-like domain (mimicked in structural studies by berylllofluoride modification of a phosphoryl accepting residue) there is a significant increase in surface area dimerisation between the two CheY-like domains. In solution, this modification causes a 50-fold increase in DGC activity, with a decrease in the dimer dissociation constant from 100 μM to around 3 μM , supporting the concept of phosphorylation aiding in the formation of a stable dimer^{179,182}. These observations have lead to the proposal that dimerisation of the CheY-like domains regulates DGC activity of PleD using the mechanism shown in Figure 1.8.

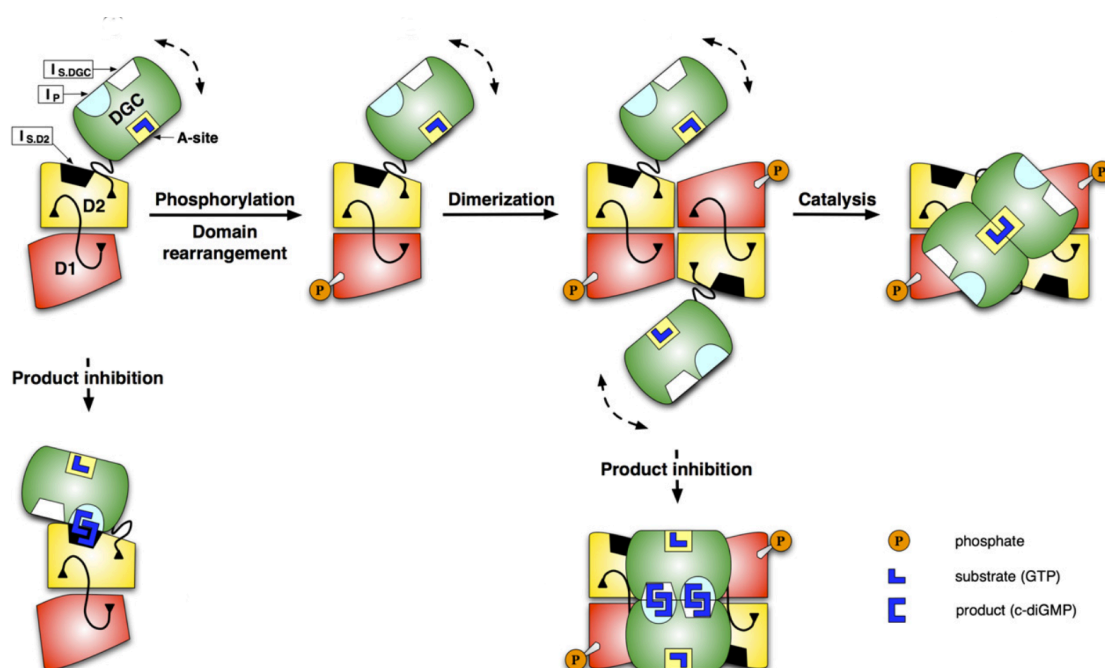


Figure 1.8 - Activation of PleD. Phosphorylation of the N-terminal CheY-like domain (labelled D1, shown in red) causes a structural rearrangement in relation to the other CheY-like domain (in yellow, labelled D2). The rearranged complex is then able to dimerise, bringing the GGDEF domain (shown in green, labelled DGC) of each molecule into close proximity. Once within close proximity the two GGDEF domains can dimerise to form the active DGC. C-di-GMP dimers can bind to the I-site when the protein is monomeric or dimeric and prevent the active half-sites of the DGC domains reorienting for catalysis. Taken from Wassman et al., 2007¹⁷⁹.

Regulation of DGC activity through dimerisation is further supported by studies of the *P. aeruginosa* protein WspR, the response regulator of the Wsp chemosensory system. As well as observing a dimerisation requirement for DGC activity with wild-type protein, a hybrid protein was also created using the GGDEF domain of WspR and a leucine zipper domain for dimerisation¹⁹⁰. This hybrid protein retained DGC catalytic activity, suggesting that dimerisation produces a functionally active GGDEF complex, but the dimerisation domains themselves are interchangeable¹⁹⁰. However, such a model does not fit with the GGDEF domain containing *E. coli* protein DosC which, despite being constitutively dimeric, requires binding of O₂ to an N-terminal globin domain for DGC activity^{186,191}. Similarly, the *E. coli* protein DgcZ (also known as YdeH) is dimeric when active and when inhibited through zinc-binding to an N-terminal CZB domain¹⁸³. A requirement for dimerisation and specific local interactions may be rationalised through observation that, despite large variation in the nature of N-terminal signal input domains in DGC proteins (with Bhr¹⁹², GAF¹⁸⁵, PAS^{187,193} and REC¹⁸² domains previously observed), a C-terminal helix capable of forming a parallel coiled coil within dimeric arrangements is conserved¹⁸⁰.

Rearrangement around this helix could therefore offer a second regulatory level to controlling c-di-GMP synthesis, with perception of appropriate stimuli allowing GGDEF domains to adopt the conformation of a catalytic dimer¹⁸⁰.

Many GGDEF domain proteins are inhibited in the presence of c-di-GMP at concentrations greater than 1 μM ^{105,133}. This prevents excess GTP usage by setting an upper limit on local c-di-GMP levels. The structural mechanism of this inhibition was elucidated from the crystal structure of PleD. C-di-GMP intercalates as a dimer (with all four guanine base stacking) and cross-links the two GGDEF domains, locking them in an inactive conformation¹⁰⁵. There are two sites involved in this inhibition, known as the primary inhibition site (I_p , which features a characteristic RXXD sequence motif, with X representing any amino acid, within the GGDEF domain) and the secondary inhibition site (I_s), shown in Figure 1.9. Both of these inhibitory sites are located on the opposite side of the protein from the GGDEF active half-site. Within the non-activated PleD structure the I_p site located on the GGDEF domain coordinates a c-di-GMP dimer with an I_s site located within one of the CheY-like domains inside the same protein monomer, but in the activated structure the bound c-di-GMP dimer links the I_p site of one GGDEF domain to the I_s site of the GGDEF domain present within the other monomer^{105,179}. This method of negative autoregulation is widely conserved among GGDEF domain proteins, with more than 60% of GGDEF domains containing I_p site RxxD motifs¹³³.

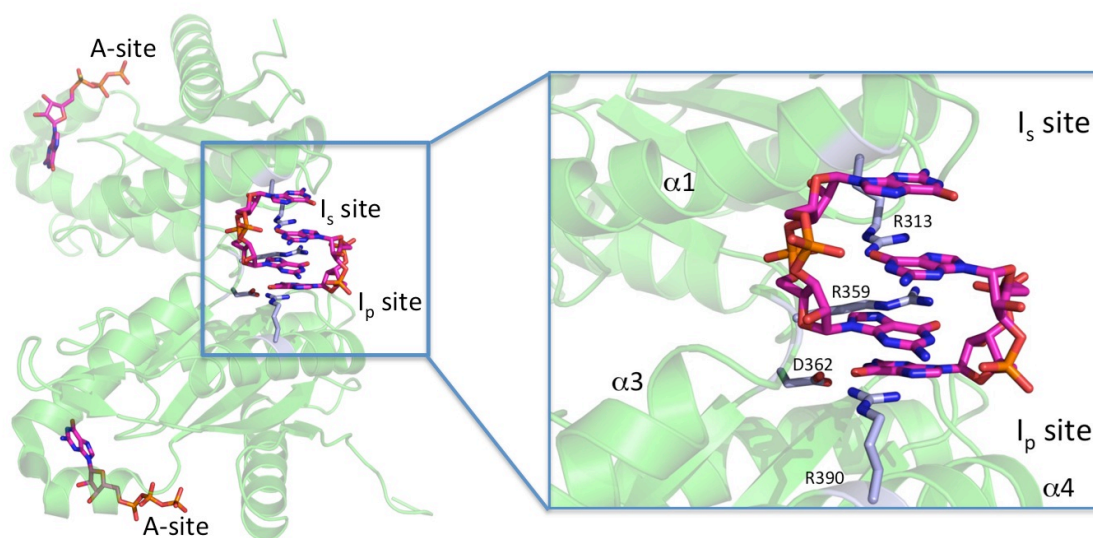


Figure 1.9 - The structure of c-di-GMP inhibited PleD (PDB 2V0N). The binding of c-di-GMP to the I_s -site locks the two GGDEF domains in a conformation where the two A-sites cannot come together to catalyse diguanylate cyclisation. C-di-GMP forms a dimer through base stacking interactions and interacts with the I_p site on one monomer and the I_s site on the other. Residues that interact with the c-di-GMP dimer (magenta) are highlighted in grey.

1.3.2 EAL domain phosphodiesterases

EAL domains are c-di-GMP specific phosphodiesterases, which require Mg^{2+} or Mn^{2+} for activity^{110,129,194}. EAL-PDE catalysed c-di-GMP hydrolysis has been determined to take place with a k_{cat} in the order of 1 sec^{-1} and K_m values between $0.06 \mu\text{M}$ and $35 \mu\text{M}$ ^{110,111,129,194–198}. Structurally, EAL domains consist of 8 α -helices and 8 β -strands arranged to form a triosephosphate isomerase (TIM)-barrel like fold, shown in Figure 1.10. The EAL domain barrel is arranged $\alpha\beta(\beta/\alpha)_6\beta$, with $\beta 1$ antiparallel to the other 7 strands that make up the β -barrel core^{199–201}. The two N-terminal β -strands and two α -helices contribute to the formation of a lobe adjacent to the barrel²⁰⁰.

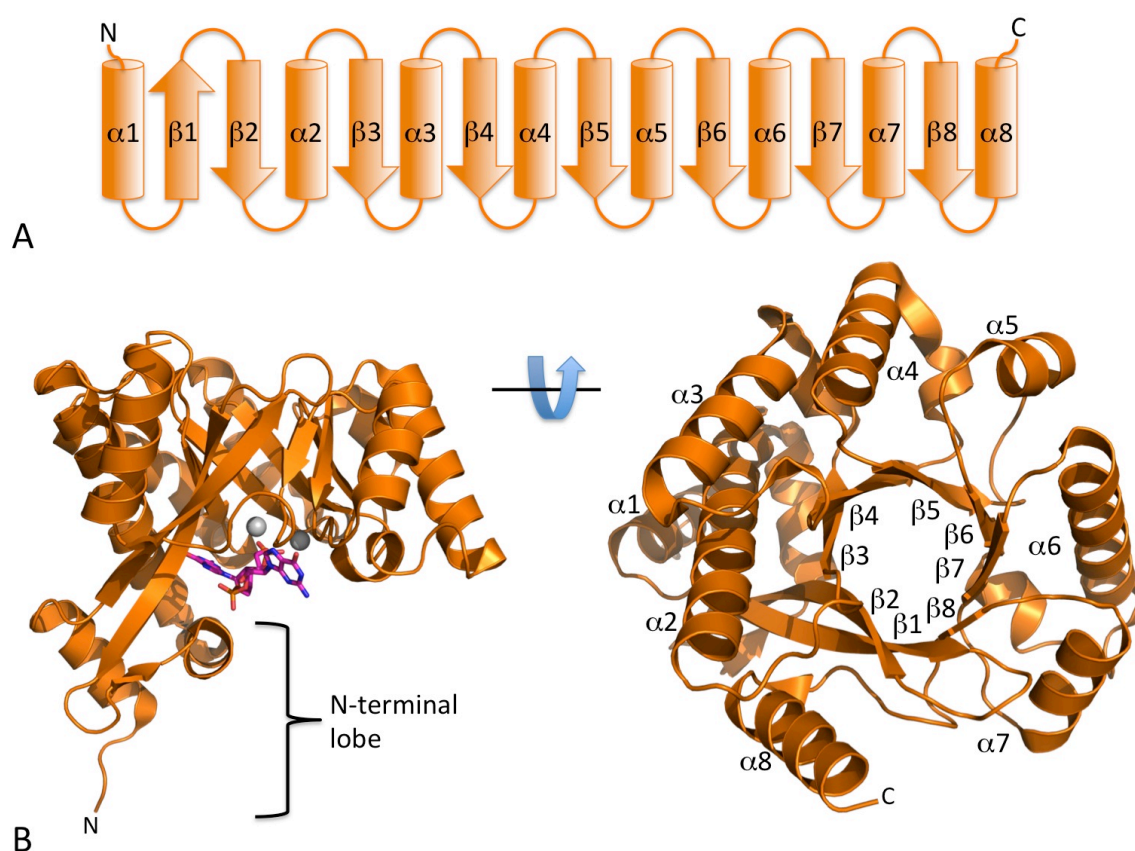


Figure 1.10 - The structure of *Thiobacillus denitrificans* TBD1264^{EAL} (PDB 3N3T). A) Secondary structure topology of EAL domains. B) Cartoon representation of the EAL domain, shown in two orientations. Bound c-di-GMP and divalent cations can be seen within the active site as magenta sticks and grey spheres respectively.

c-di-GMP binds in an extended conformation to the bottom of the EAL domain barrel, at the C-terminal end of strands $\beta 2$ - $\beta 8$, as shown in Figure 1.10B^{198,199,201}. Although the position of c-di-GMP is consistent between different proteins, nucleotide binding interactions vary^{196,197,199,201,202}. The reaction product of c-di-GMP hydrolysis, pGpG, has been observed to compete directly for the c-di-GMP binding site¹¹². Despite differences in the nature of substrate

binding, in all determined structures substrate is observed in the presence of bound metal ions. The presence of bound metal ions within the active site is concurrent with a requirement of Mg^{2+} or Mn^{2+} for activity and Ca^{2+} -induced inhibition^{111,129,196,198}. The majority of substrate-bound structures (whether bound to Mg^{2+} , Mn^{2+} or Ca^{2+}) display two metal ions, labelled M1 and M2 in Figure 1.11^{196,197,199,201,202}. The M1 metal is coordinated by the side chains of the first Asp residue within the conserved DDFGTG motif (labelled D646 in Figure 1.11, with numbering corresponding to *T. denitrificans* TBD1265), the Glu residue of the signature EAL motif (E523 in Figure 1.11), as well as side chains of conserved Asn and Glu residues (N584 and E616 in Figure 1.11), a c-di-GMP non-bridging phosphate oxygen and a water. There is slight variation in the coordination of M2 metal, a non-bridging c-di-GMP phosphate oxygen and the side chain of the second Asp residue in the conserved DDFGTG motif (D647 in Figure 1.11) are always involved, with the first Asp of the DDFGTG motif (E646 in Figure 1.11) and a conserved Glu side chain (E703 in Figure 1.11) also frequently observed in this coordination sphere^{196,197,199,201,202}.

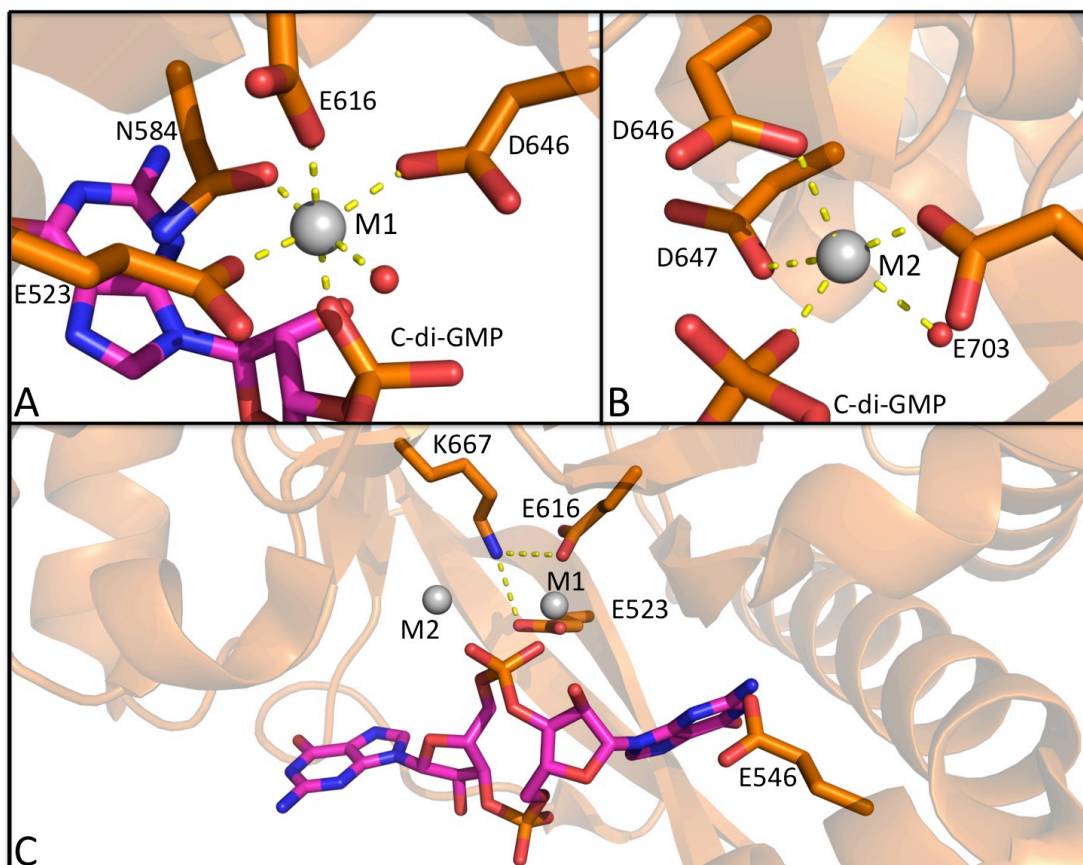


Figure 1.11 - EAL-PDE conserved residues within *T. denitrificans* TBD1265 (PDB 3N3T). A) The M1 metal site, displaying strictly conserved coordination. B) The M2 metal site, to which c-di-GMP and E647 always contribute. C) The other conserved EAL-PDE residues required for function, K667 and E546.

Within the EAL domain of *Thiobacillus denitrificans* protein TBD1265, single alanine mutations of all conserved metal-coordinating residues (three glutamic acid residues, two aspartic acid residues and an asparagine, E523, E616, E703, D646, D647 and N584 respectively in Figure 1.11) displayed no c-di-GMP hydrolysis, suggesting a requirement for both metal ions in catalysis²⁰¹. Further mutations of residues conserved between EAL domains demonstrated a lysine residue (K667 in Figure 1.11), which interacts with M1 binding glutamic acid residues (E523 and E616 in Figure 1.11), and a glutamic acid residue of unknown function (E546 in Figure 1.11) were required for PDE activity²⁰¹. On the back of these findings a two-metal ion catalytic mechanism of c-di-GMP hydrolysis was proposed²⁰¹. A conserved lysine (K667 in Figure 1.11) and the two metal ions activate a water molecule to generate a hydroxide nucleophile. This nucleophile attacks the nearby phosphorus atom, with the O3' leaving group protonated by a nearby water molecule and then coordinated by a conserved aspartic acid^{198,201}. The reaction mechanism for two-metal catalysed EAL domain is shown in Figure 1.12.

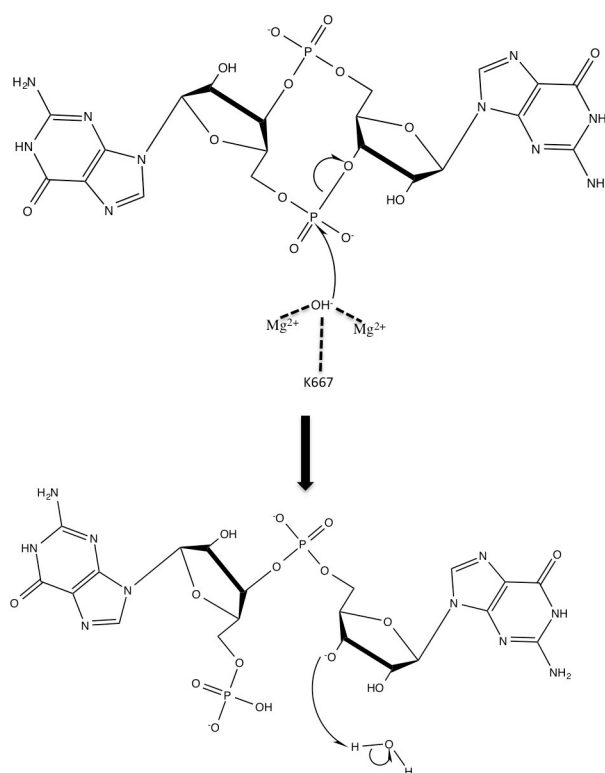


Figure 1.12 - Two-metal catalysed hydrolysis of c-di-GMP by EAL domains. Hydroxide nucleophile coordination is shown with dashed lines.

Much like DGC activity within GGDEF domains, PDE activity of EAL domains is dependent on the formation of catalytically competent dimers, even though each EAL domain monomer is able to bind substrate^{196,197,202,203}. A dimerisation interface, created through the formation of a compound helix made up of α 5-helical segments from each monomer and with the highly conserved α 6-helix from each running anti-parallel, was first identified from the *Klebsiella pneumoniae* protein BlrP1¹⁹⁸ (shown in Figure 1.13) but has subsequently been observed in the crystal structures of other EAL-domain-containing proteins^{196–199,201,202}. Site directed mutagenesis to prevent the formation of this dimer has been demonstrated to reduce PDE activity without altering substrate binding affinity¹⁹⁶. The functional role of this dimerisation interface has been determined through observation that, in contrast to monomeric substrate-free structures, dimeric crystal structures of the *P. aeruginosa* protein MorA EAL domain possess a shorter α 5 helix, allowing the now-extended β 5- α 5 loop to position the catalytic aspartate residues of the DDFGTG motif into the active site for metal ion and substrate binding²⁰², as depicted in Figure 1.13. This regulatory mechanism has been demonstrated to be used within larger multi-domain proteins in solution¹⁹⁷. BlrP1 of *K. pneumoniae* consists of an EAL domain PDE regulated by a light-sensing LOV domain, and although BlrP1 is in a predominantly dimeric state independent of photo-sensation by the LOV domain, hydrogen-deuterium exchange experiments demonstrate that activation of the LOV-domain and substrate-binding stabilise the dimerisation interface between EAL domains, increasing activity¹⁹⁷.

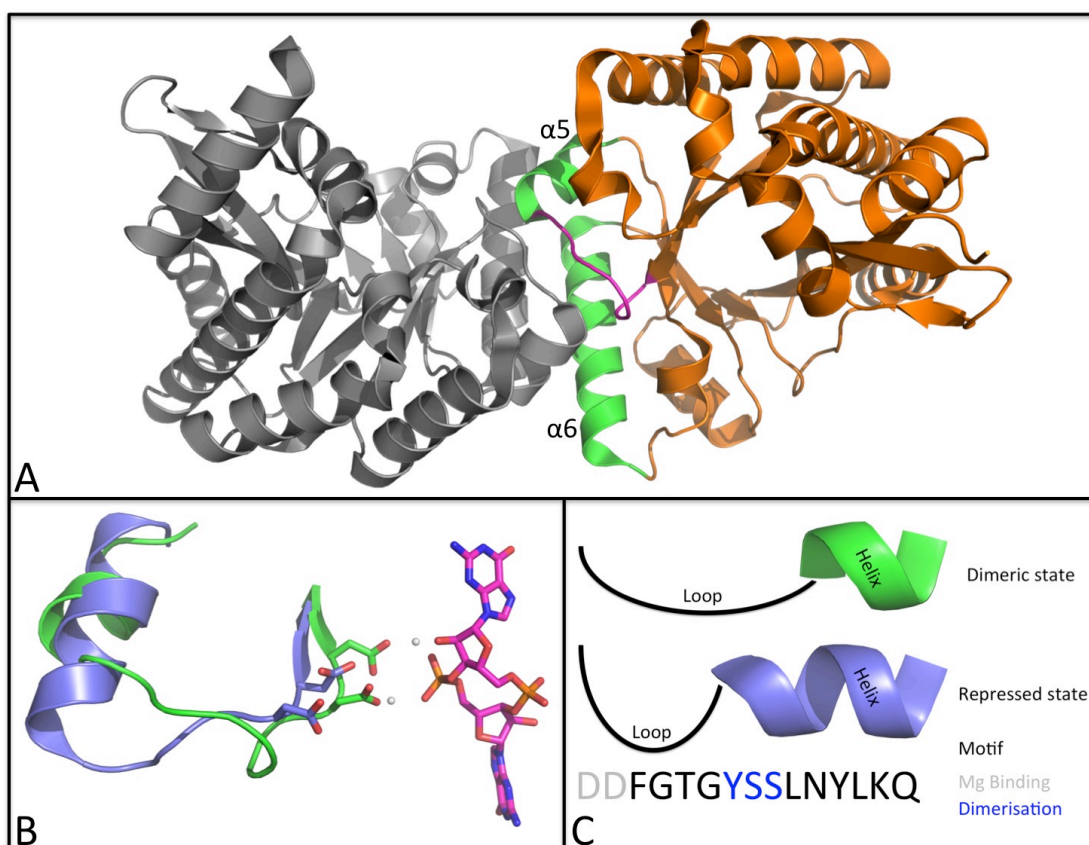


Figure 1.13 - EAL domain dimerisation is required for PDE activity. A) EAL domain dimerisation around the $\alpha 5$ and $\alpha 6$ helices (shown in green) induces a conformational change in the $\beta 5$ - $\alpha 5$ loop (shown in magenta). B) In the repressed state (blue) the $\alpha 5$ helix is extended, preventing catalytic Asp residues (shown as sticks) from coordinating metal ions (shown as grey spheres), however this is not true of the dimeric state (shown in green). C) Dimerisation contacts to the YSS motif at the end of the $\alpha 5$ helix can cause extension of the $\beta 5$ - $\alpha 5$ loop, facilitating metal coordination. Adapted from Phippen et al., 2014²⁰².

1.3.3 HD-GYP domain phosphodiesterases

HD-GYP domains (named after two conserved sequence motifs) are part of the HD superfamily of metal-dependant phosphohydrolases and have been demonstrated to cleave c-di-GMP into a pair of GMP molecules through an intermediate of pGpG^{109,204,205}, although in some cases catalysis has been reported to cease at pGpG^{206,207}. A model of sequential, independent phosphodiester bond cleavage is supported by enzymatic and structural data, which is consistent with observations that pGpG can be used as substrate for hydrolysis^{208,209}.

The HD-GYP domain is made up of 7 α -helices, and like other members of the HD superfamily these are arranged around a 5-helix core, which forms a claw-like structure through α 2- α 3, α 4- α 5 and α 5- α 6 loops (shown in Figure 1.14)^{208,210,211}. The HD and GYP motifs are found on the α 2-helix and the α 4- α 5-loop respectively. A single HD-GYP PDE has been structurally characterised (*Persephonella marina* protein GH), and has been observed in holo, GMP bound and c-di-GMP bound states²⁰⁸. Within the c-di-GMP bound structure, the substrate is distorted to form a V-shape (shown in in Figure 1.14), as opposed to the extended conformation of the molecule when bound to an EAL domain, (shown in Figure 1.11). The c-di-GMP is buried within the pocket of the claw, coordinated by the side chains of the tyrosine within the GYP motif (Y285 in Figure 1.15), Arg314 and Lys317, as well as the protein backbone (labelled K235 in Figure 1.15).

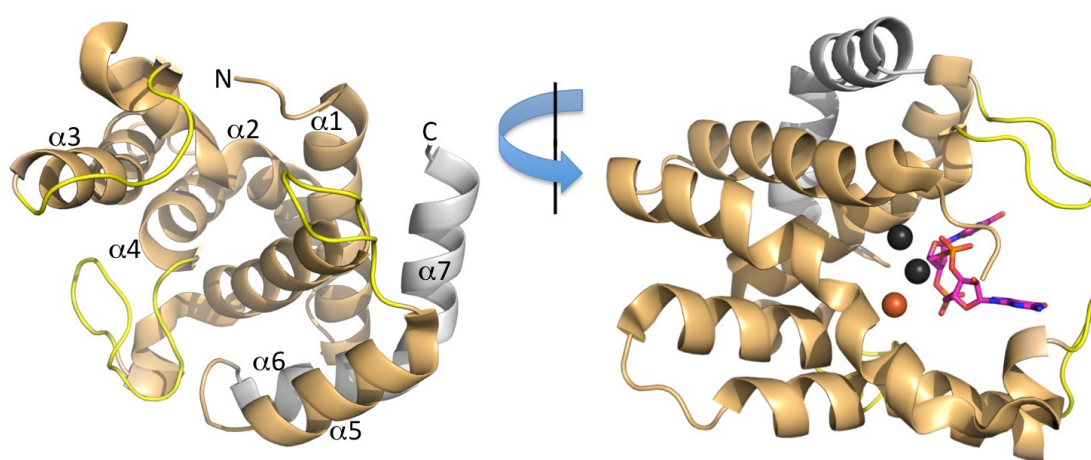


Figure 1.14 - The structure of the HD-GYP domain of *Persephonella marina* GH (PDB 4MDZ). The structure is shown in two orientations with c-di-GMP (magenta) and metal binding sites shown as spheres. The metal site occupied in holo, c-di-GMP bound and GMP bound structures is shown in brown, with metal binding sites identified in holo and GMP bound structures but unoccupied in the c-di-GMP bound structure shown in black. The 5-helix core of the HD superfamily is shown in wheat, with α 2- α 3, α 4- α 5 and α 5- α 6 loops shown in yellow.

Holo and GMP-loaded complexes of *P. marina* GH show three irons within the active site. Although the c-di-GMP bound structure shows only one iron (labelled as M1 in Figure 1.15), superposition of c-di-GMP into a fully occupied trinuclear centre highlights a proximity between the substrate and M2 metal, shown in Figure 1.15, which has been suggested to be indicative of a role in substrate coordination²⁰⁸. In holo and GMP-loaded *P. marina* GH complexes, all three iron atoms are coordinated through buffer molecules and conserved HD-GYP domain residues: Glu185 and His189 on the α 1-helix, the HD motif (His221 and Asp222), His250 on α 3, His276 and His277 on α 4 and Asp305 on α 5, shown in Figure 1.15²⁰⁸.

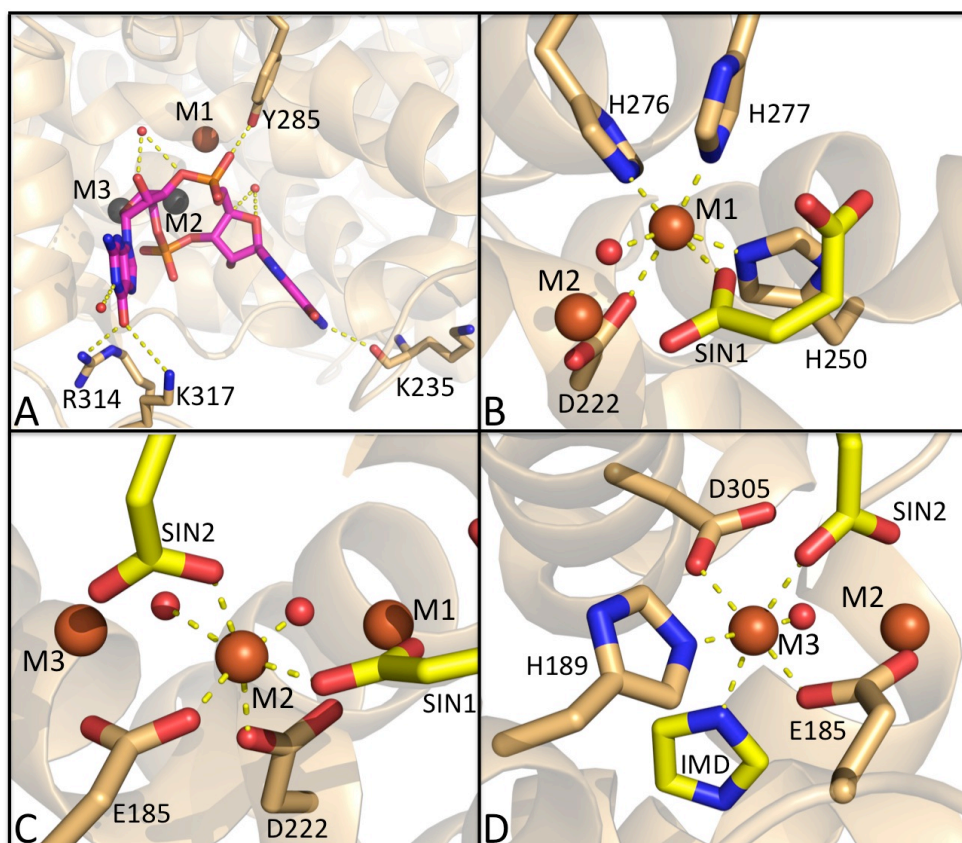


Figure 1.15 - Coordination of c-di-GMP and metal ions within the HD-GYP of *P. marina* GH. A) Coordination of c-di-GMP within HD-GYP domains. One iron is present within the *P. marina* GH c-di-GMP complex (PDB 4MDZ), labelled M1 (shown in brown), however as M2 and M3 metals are thought to be involved in catalysis the superposed position of these are displayed as black spheres. B) Coordination of M1 iron within the *P. marina* GH holo structure. C) Coordination of the M2 iron within the *P. marina* GH holo structure. D) Coordination of the M3 iron within the *P. marina* GH holo structure. Irons are shown as brown spheres, buffer molecules in yellow (labelled SIN1 and SIN2 for succinic acid and IMD for imidazole) and waters as red spheres.

As demonstrated by a loss of activity upon soaking with EDTA²⁰⁸, HD-GYP PDE activity is dependant on metal ions within the active site, although either Fe²⁺ or Mn²⁺ can fulfil this requirement²¹². Variation in the number of metal ions present within the active site has been used to explain observations of hydrolysis ceasing at pGpG²¹². While single-point mutants that prevent occupation of the M2 metal site completely abolish PDE activity, similar mutations preventing coordination of the M3 metal allows c-di-GMP hydrolysis, but prevents degradation of pGpG to GMP²¹². Upon this evidence, a reaction mechanism for HD-GYP PDEs has been proposed (shown in Figure 1.16) with c-di-GMP cleaved through nucleophilic attack upon the P1 phosphate by a water bridging M1 and M2 metals. The newly formed pGpG then dissociates and reassociates with a

phosphodiester bond of the P2 phosphate hydrolysed by a nucleophile generated from the tri-metal centre²¹².

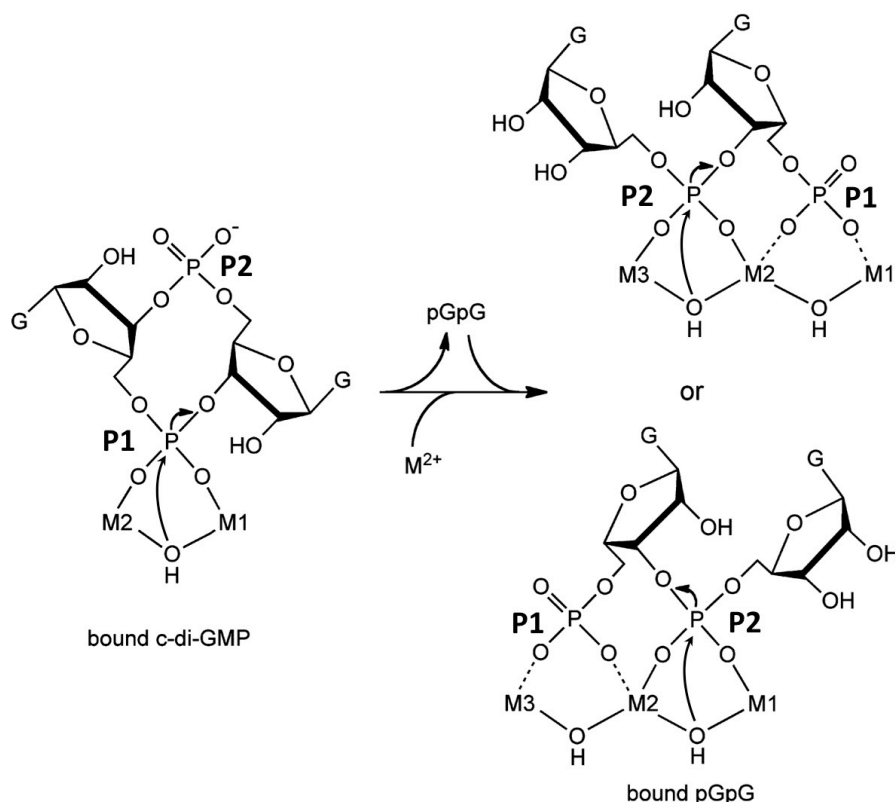


Figure 1.16 - Proposed mechanism for HD-GYP catalysed c-di-GMP hydrolysis. M1 and M2 metals coordinate a hydroxide, which carries out nucleophilic attack on the P1 phosphate. The hydrolysis product, pGpG, then leaves and reassociates along with a metal ion, which occupies the M3 position. The P2 phosphate of pGpG is then subject to nucleophilic attack by a hydroxide radical coordinated by either M2 and M3 or M1 and M2. Adapted from Miner and Kurtz, 2016²¹².

1.3.4 Oligoribonuclease phosphodiesterases

Not all species that use c-di-GMP signalling contain genes encoding for an HD-GYP domain^{112,124}, suggesting other strategies must be required for catabolism of pGpG. Two studies have recently identified oligoribonucleases (Orn) as capable of pGpG hydrolysis^{112,113}. Highly conserved among bacterial species, Orn has previously been characterised as a manganese dependent 3'-5' exonuclease involved in degrading short length RNA oligomers^{213,214}. When assayed *In vitro*, *Vibrio cholerae* and *P. aeruginosa* Orn (PA4951 in PAO1) have been demonstrated to degrade pGpG, with no observable c-di-GMP binding^{112,113}. *In vivo* gene knockout studies further underline the significance of Orn as the primary enzyme responsible for catabolism of pGpG within

P. aeruginosa, although low levels of pGpG degradation are still observed in Δorn strains due to the presence of HD-GYP PDEs^{112,113}.

Orn belongs to the DEDDh subfamily of exoribonucleases, named after the conserved acidic active site motif, and a further conserved histidine²¹⁵. *P. aeruginosa* Orn shows a high level of sequence homology to the structurally characterised *Xanthomonas campestris* Orn^{112,216}, which has a $\beta_3\alpha_4\beta_2\beta_3\alpha_3$ topology, shown in Figure 1.17. Single alanine point mutation of active site residues Asp11, Glu13, Asp111, His157 and Asp162, shown in Figure 1.17, render *P. aeruginosa* Orn unable to hydrolyse pGpG without altering binding¹¹². These results are similar to equivalent analysis upon another DEDDh exoribonuclease, *E. coli* RNase T²¹⁷.

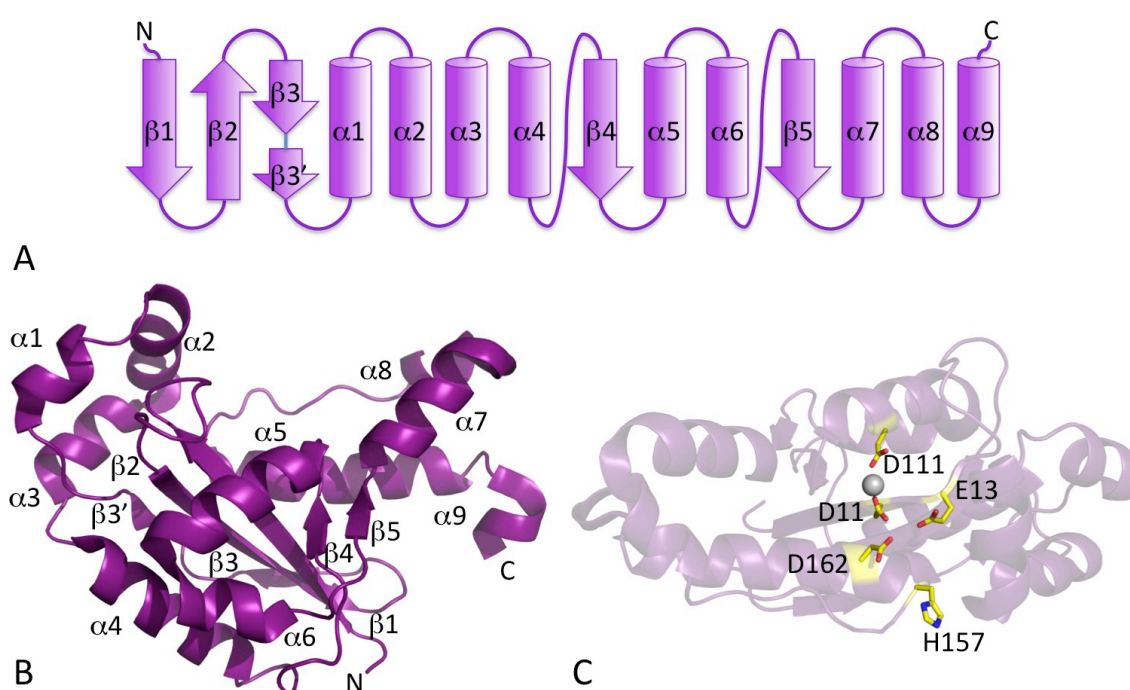


Figure 1.17 - The structure of *X. campestris* Orn (PDB 2GBZ). A) Secondary structure topology of Orn. B) The three dimensional structure of Orn. C) Active site residues required for Orn activity. The magnesium ion crystallised within the Orn active site is shown as a silver sphere.

Although a mechanism for pGpG degradation by Orn has not yet been determined, Orn is predicted to hydrolyse nucleotides using a similar reaction mechanism to the ϵ subunit of *E. coli* DNA polymerase III (also a member of the DEDDh exoribonuclease subfamily, shown in Figure 1.18)²¹⁸. Within the DNA polymerase III ϵ subunit, two Mn^{2+} ions are coordinated by side chains of three aspartic acids (labelled D12, D103 and D167 in Figure 1.18) and a glutamic acid residue (labelled E14), thought to be equivalent to Asp11, Glu13, Asp111 and Asp162 in *P. aeruginosa* Orn, as well as four waters. His162 (equivalent to His157 in *P. aeruginosa* Orn) deprotonates a

water also coordinated by the Mn^{2+} ion in the M1 position, which then acts as a nucleophile to attack the phosphodiester bond joining the 3'-terminal nucleoside to the remaining DNA. The newly formed five-coordinate phosphorane transition state is then further deprotonated by Glu14, resulting in elimination of the penultimate deoxyribose 3' nucleotide. The 3'-OH leaving group of the eliminated DNA is then protonated, presumably through a water molecule coordinated to the Mn^{2+} ion in the M2 position and hydrogen bonded to Asp103^{219–221}. Utilisation of this reaction mechanism by Orn is supported by loss of activity when metal-coordinating residues are mutated or in the absence of metal ions^{112,218}. A requirement for dimerisation has been observed amongst other DEDD family exoribonucleases, so it is possible the Orn may also be regulated through control of oligomerisation state, although this has not yet been observed²²².

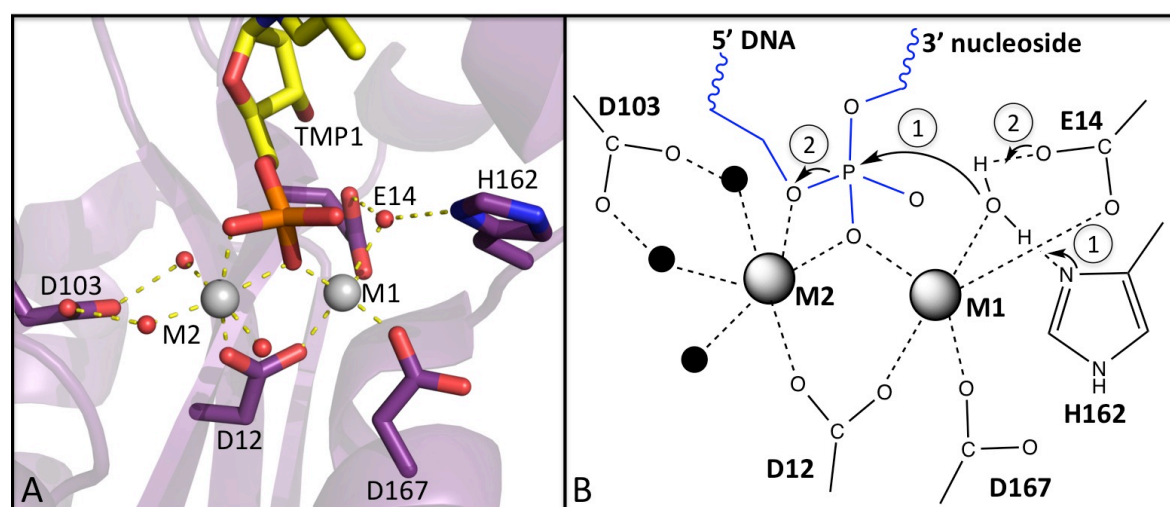


Figure 1.18 - Reaction mechanism proposed to be shared between *E. coli* DNA polymerase III and Orn. A) Structural representation of DNA polymerase III in complex with two Mn^{2+} ions (silver spheres labelled M1 and M2), and TMP, which occupies the position of the 3'-terminal nucleoside. Waters are shown as red spheres. B) The reaction mechanism of DNA polymerase III. Substrate DNA is highlighted in blue, with directionality indicated through positions of the 3' and 5' ends. Mn^{2+} ions are shown as large grey circles, with small black circles denoting waters. Circled labels denote the sequential steps of phosphodiester bond cleavage.

1.4 Nitric oxide detection within *Pseudomonas aeruginosa*

P. aeruginosa biofilms disperse in response to sub-lethal doses of NO, presumably as part of a physiological dispersal response to limiting oxygen levels within the cell²¹. A novel approach to combatting antibiotic tolerant *P. aeruginosa* biofilms could therefore be to prescribe NO in addition to conventional antibiotics, however before such a clinical strategy is implemented it would be advantageous to understand the molecular basis of NO induced biofilm dispersal within

P. aeruginosa. A paradigm of direct coupling between NO detection and c-di-GMP signalling (and thus biofilm regulation) is provided from a study within *Shewanella woodyi*, where gene knockouts of a haem-nitric oxide/oxygen-binding (H-NOX) domain, cause a reduction in biofilm formation and intracellular levels of c-di-GMP²²³. H-NOX proteins have previously been demonstrated to selectively bind NO^{224–226}. Pull-down assays suggest that this protein directly interacts with the protein SwDGC, which has both DGC and PDE activity. Recombinant expression in *E. coli* shows that when the two-proteins are co-expressed SwDGC produces c-di-GMP and promotes biofilm growth. However, just as in the wild-type *S. woodyi*, when nitric oxide was added to *E. coli* recombinantly expressing these proteins, a decrease in c-di-GMP concentrations consistent with increased PDE activity and decreased DGC activity is observed²²³. This study suggests that DGC activity of SwDGC is triggered by association with SwH-NOX in the absence of ligand, but NO binding to SwH-NOX downregulates DGC activity and upregulates PDE activity.

However, no H-NOX proteins have been identified within the *P. aeruginosa* genome. Instead, attention has focused on NO sensation through haem-binding PAS domains and through the predicted copper-binding MHYT domain, which could then interact with DGC or PDE domains to induce a reduction in intracellular c-di-GMP levels.

1.4.1 PAS domains

PAS domains were first identified as a conserved domain between the fly clock protein PERIOD, the vertebrate aryl hydrocarbon nuclear translocator (ARNT) and the fly developmental regulator single-minded (SIM), combined to give Per-Arnt-Sim or PAS²²⁷. Occurring in all biological kingdoms, PAS domains have been predominantly characterised as sensory domains involved in regulation of a number of biological processes²²⁸, however an enzymatic PAS domain has recently been identified²²⁹. PAS domains are highly diverse; many PAS domains bind cofactors, ranging from ions to more complex molecules such as flavin nucleotides and haem, allowing the PAS domain architecture to be utilised to detect a number of sensory inputs, such as NO. Most PAS domains however do not have an associated cofactor or known ligand, suggesting that they may act as sensors through protein-protein interactions, or may have a function independent of sensation²²⁸.

The PAS domain architecture referred to throughout this work (consistent with more recent studies of this domain architecture) consists of what has previously been denoted as separate PAS and PAC motifs, or S1 and S2 boxes^{230–232}. PAS domains are approximately 100 amino acids in length and contain an archetypal fold consisting of an antiparallel, five-stranded β -sheet with a 2-1-5-4-3 strand order flanked by numerous α -helices that create a pocket for ligand binding,

displayed in Figure 1.19. Although the number and length of PAS domain α -helices varies^{228,233}, cytoplasmic PAS typically display an $\beta\beta\alpha_4\beta_3$ topology, with extracytoplasmic PAS domains (also referred to as PDC domains^{234,235}) exhibiting an $\alpha_2\beta_2\alpha$ (or α_2) $\beta_3\alpha$ topology²³⁵.

In bacteria, PAS domains are commonly associated with nucleotide cyclases, PDEs, methyl-accepting chemotaxis proteins, and most commonly with histidine kinases. These PAS domains are often found N-terminal to the output domains, and many proteins contain more than one N-terminal PAS domain^{236,237}. It is believed that these PAS domains detect a sensory input and use this to modulate the activity of the associated enzymatic domain. Irrespective of bound cofactor, many PAS domains have been observed to promote formation of higher-order oligomers^{227,228,238,239}. Several different dimeric arrangements have been observed, with most dimerisation interactions mediated through hydrophobic residues on the outer surface of the β -sheet^{116,240–245}. It has previously been proposed that PAS domains modulate formation of dimers and higher order oligomers between proteins, with sensory PAS domains regulating these interactions in response to specific signals²²⁸. As the activity of DGC domains and that of EAL-type PDE domains is regulated by dimerisation (as discussed in section 1.3), it is possible that signalling through associated PAS domains could be used to induce changes in the dimerisation state of c-di-GMP metabolising enzymes, regulating biofilm behaviour.

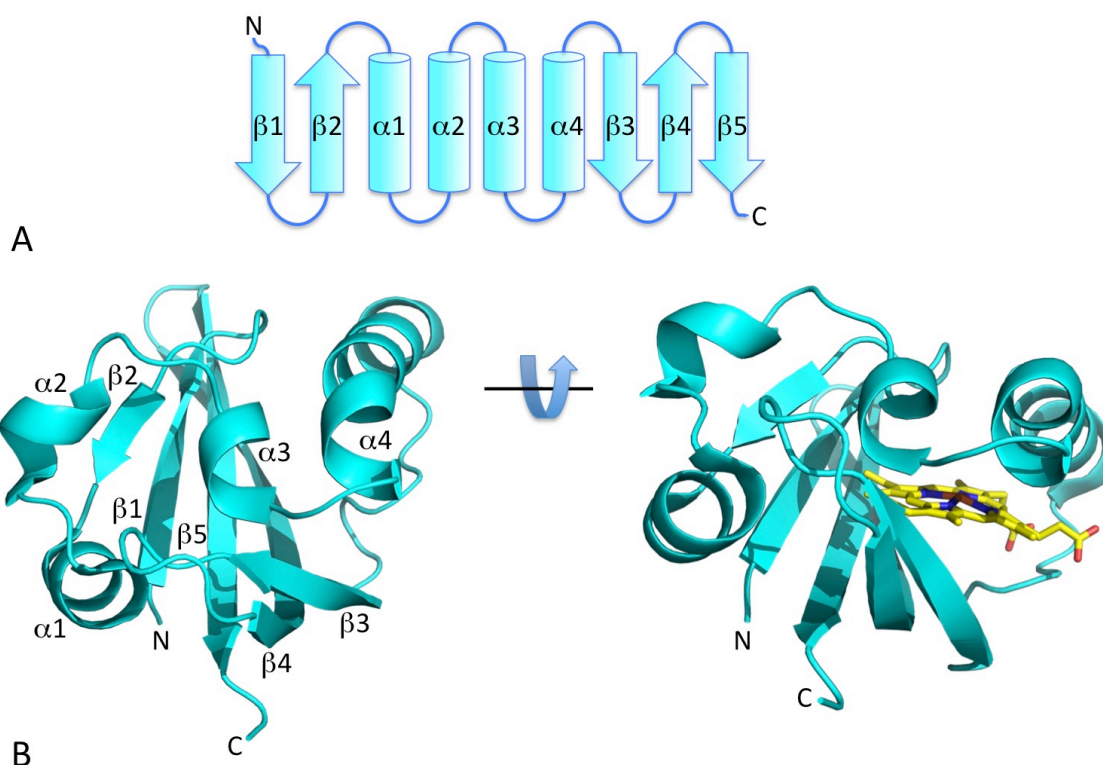


Figure 1.19 - The PAS domain-fold of *Bradyrhizobium japonicum* protein FixL (PDB 1DRM). A) Secondary structure topology of cytoplasmic PAS domains. B) Cartoon representation of the cytoplasmic *B. japonicum* FixL PAS domain, with a haem cofactor shown in yellow.

1.4.2 Haem-binding PAS domains

NO has been demonstrated to interact with PAS domains which contain a haem cofactor¹¹⁵. It is therefore possible that binding of NO to a haem-PAS domain could propagate signalling within the cell to induce biofilm dispersal. Two different forms of haem-binding PAS domain have been identified, cytoplasmic PAS domains that have a haem-*b* cofactor (a type of haem cofactor with no covalent bonding to the surrounding protein) and PAS domains that bind haem-*c* (a type of haem cofactor that forms thioether links to the protein), which reside in the periplasm^{246,247}.

A haem-*b* binding PAS domain from the *E. coli* protein Dos (direct oxygen sensor) already provides a paradigm for direct haem-mediated regulation of PDE activity, coupling iron reduction to EAL domain dimerisation and activation^{115,116}. The Dos protein exists as a homodimer, with the haem cofactors running parallel and their iron centres approximately 30 Å apart. This dimer is stabilised by hydrophobic interactions through short N-terminal helices¹¹⁶.

The haem cofactor within the Dos PAS domain contains a hexacoordinated iron, with axial ligands provided by an imidazole nitrogen on His77 and a water molecule whilst in the Fe³⁺ state. When

the iron is reduced to an Fe^{2+} state the coordinating water is displaced by Met95, with no movement of the haem or obvious effect on His77 reported¹¹⁶. Met95 is part of the loop between the α 4-helix and the β 3-strand near the dimer interface, which is highly flexible when the iron is oxidised, but iron reduction causes this loop to become more rigid, not only through the direct interactions of Met95 on the iron, but also due to the propionate oxygens of the haem interacting with the main chains of Gly94 and Met95. This rigidity of the α 4- β 3-loop brings together Phe113 from one subunit and Arg131 of the other, pulling the dimer closer together¹¹⁶. The movement at the dimer interface between the PAS domains results in a 3° rotation around a hinge within the PAS domain, pulling the phosphodiesterase domains close enough together to become active¹¹⁶. Regulation of Dos through iron oxidation state is illustrated in Figure 1.20.

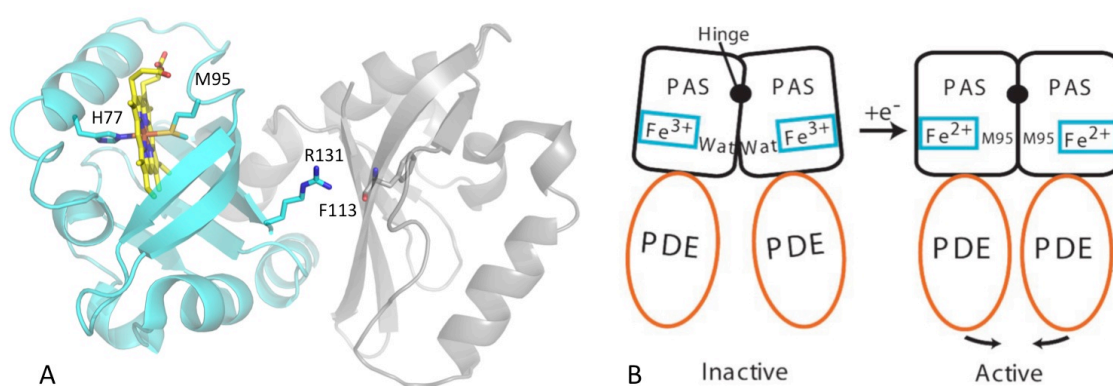


Figure 1.20 - Regulation of Dos activity by changes in the redox state of the iron ion within PAS domain associated haem. A) The PAS domain homodimer of Dos with bound Fe^{2+} (PDB 1V97). His77 coordinates the iron ion in the ferric (Fe^{3+}) and ferrous (Fe^{2+}) oxidation states (PDB 1V9Z), but reduction induces binding to Met95. In the ferrous form hydrogen bonding between haem and the protein backbone of Gly94 and Met95 can also occur. The altered interactions with the haem cofactor allow Arg131 of one subunit to interact with Phe113 of the other, altering the dimer interface. Residues involved in the induction of conformational change within the PAS domain are shown as sticks, with the haem cofactor of one subunit shown in yellow. B) Conformational change within the PAS domain of Dos is predicted to promote dimerisation of PDE domains within the protein, increasing catalytic activity. Adapted from Kurokawa et al, 2004¹¹⁶.

Two other types of protein containing haem-*b* binding PAS domains have been characterised, FixL (with structures determined for both *Bradyrhizobium japonicum*^{246,248,249} and *Sinorhizobium meliloti*²⁵⁰ FixL) and *P. aeruginosa* Aer2^{187,251,252}. In both *B. japonicum* and *S. meliloti*, FixL is a histidine kinase that regulates the expression of nitrogen-fixation genes in response to low oxygen levels by phosphorylating the response regulator FixJ^{246,248,249,253–257}.

Similar to Dos, the iron within the haem of the FixL PAS domain is axially coordinated by a histidine imidazole group. However, within FixL there is no distal methionine to coordinate the central haem iron, leaving the iron pentacoordinated and resulting in a puckered haem not observed in Dos^{116,246}. The distal positions around the haem are instead occupied by a hydrophobic triad (Ile215, Leu236 and Ile238 in *B. japonicum* FixL and Ile209, Leu230 and Val232 in *S. meliloti* FixL), which contribute to the hydrophobic haem binding pocket. The propionate hydroxyls of the haem in FixL form salt bridges with a histidine and an arginine, as well as forming a number of hydrogen bonds to backbone nitrogens^{250,253,258}.

When O₂ is bound to the haem PAS domain, the kinase domain of FixL is autoinhibited, with the kinase domain activated upon dissociation of O₂^{258–262}. Based on the differences in structure between O₂, CN[−] and ligand free structures of the *B. japonicum* FixL PAS domain, a mechanism of conformational change upon ligand binding has been proposed^{246,248,249,257}. Binding of O₂ to the haem induces a displacement of the haem 7 propionate carboxylate group (shown in Figure 1.21), replacing the salt bridge it forms to the side chain of Arg220 (numbered in accordance to *B. japonicum* FixL, equivalent to Arg214 in *S. meliloti* FixL) with a salt bridge to the side chain of Arg206 (again, numbered according to *B. japonicum* FixL, equivalent to Arg220 in *S. meliloti* FixL). Arg220 then moves into the haem-binding pocket where it hydrogen bonds to the bound O₂, clashing with Ile215 (*B. japonicum* FixL, equivalent to Ile209 in *S. meliloti* FixL) and inducing a shift in the α4-β3 loop (also referred to as the FG loop) of approximately 2 Å. Movement of Arg206 to bind the haem 7 propionate carboxylate would also shift the position of Tyr207 (Tyr201 in *S. meliloti* FixL), which in turn would shift Glu240 (Glu234 in *S. meliloti* FixL) in the β4 strand, an observation supported by resonance Raman spectroscopy²⁵⁸. Movement in the β4 and β5 strands could also be induced by displacement of the hydrophobic triad distal to the haem^{246,249,258,263}. As no structure of full-length FixL has yet been determined, homology modelling of FixL onto the PAS-histidine kinase proteins *Thermotoga maritimas* ThkA and *Erythrobacter litoralis* EL346 has been used to suggest that movement of the β3 and β4 strands (shown in Figure 1.21) within the PAS domain upon ligand binding alters the nature of the interaction interface to the kinase domain, preventing kinase activity^{258,261,264}.

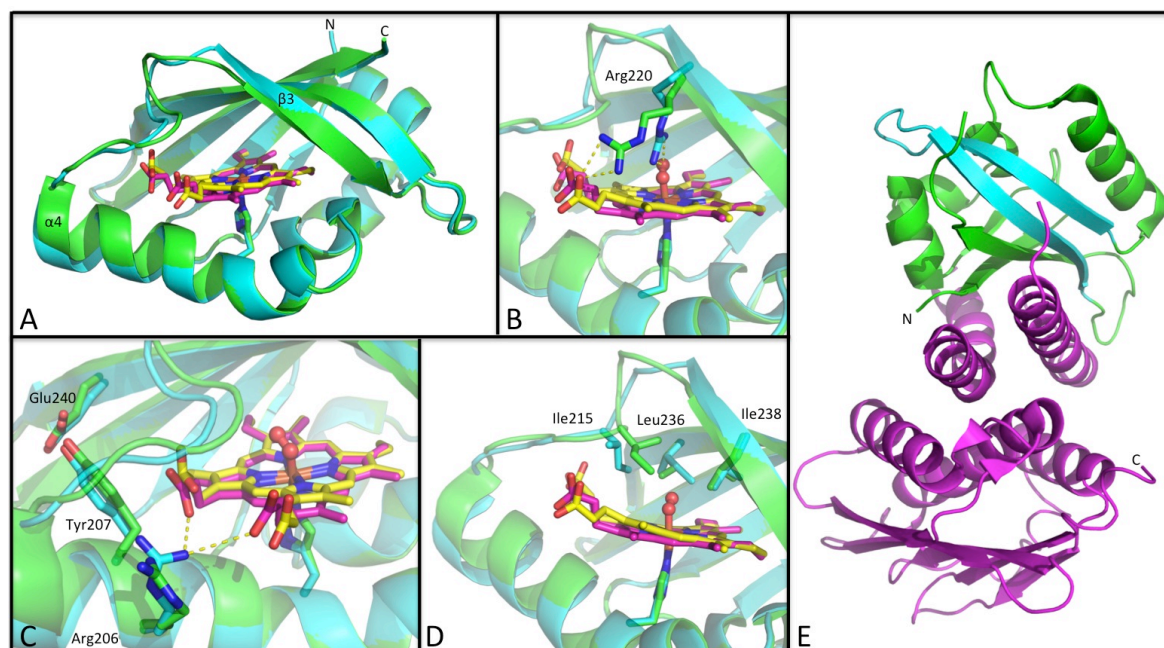


Figure 1.21 - Conformational changes in FixL upon oxygen binding. The ligand-free structure of *B. japonicum* FixL (PDB 1DRM) is shown in green (with the haem in yellow), and the oxygen bound structure (PDB 1DP6) is shown in blue (with the haem in magenta). A) Oxygen binding induces a large shift in the α 4- β 3 loop. B) Arg220 hydrogen bonds to propionate 7 of the haem group in the deoxy state, but to the ligand upon oxygen binding. C) Upon oxygen binding Arg206 reorientates to bind the haem propionate 7, also shifting Tyr207 and Glu240. D) The distal hydrophobic residues of Ile215 and Leu236 are reoriented upon ligand binding. E) The structure of EL346 (PDB 4R3A), with the PAS domain shown in green, and the kinase domain in purple. The β 3 and β 4 strands in close proximity to the kinase domain are highlighted in blue.

The physiological role of Aer2 (also known as McpB and TlpG1) is unclear, Aer2 has been suggested to act as a sensor for O_2 in *P. aeruginosa* aerotaxis (the movement of the bacterium towards oxygen), but experimental evidence is highly contradictory^{265–270}. Aer2 consists of three N-terminal HAMP domains (histidine kinase-adenylyl cyclase-methyl-accepting chemotaxis protein-phosphatase domains, a commonly used signalling domain²⁷¹), a haem-binding PAS domain, two further HAMP (HAMP 4 and 5) domains and a C-terminal methyl-accepting chemotaxis (MA) domain^{252,265}. The haem-*b* within Aer2 is bound predominantly through hydrophobic interactions, with further coordination only provided from a histidine imidazole group as a proximal axial ligand for the haem-bound iron, and hydrogen bonding between a histidine side chain and a haem propionate group²⁶⁶. Differences in observed crystal structures of Aer2 in CN and ferric, ligand-free forms have led to the suggestion that binding of ligand to the haem cofactor induces a 90° rotation in the ligand coordinating tryptophan residue (Trp283) and an accompanying shift of a nearby leucine residue (Leu264) within the haem-binding pocket,

shown in Figure 1.22²⁵². This propagates a conformation change throughout the β -sheet, the N-terminal α -helix and the loop C-terminal to the β 5 strand, which is thought to cause a rearrangement of the PAS dimer, or shift the monomer-dimer equilibrium²⁵².

When expressed in isolation, the PAS domain of Aer2 is found in a predominantly monomeric state, with the HAMP domains of Aer2 thought to hold the PAS domains in a dimeric arrangement in the context of full-length protein, a concept supported by the requirement of the N-terminal HAMP domains for signalling, even though they do not lie between the signal input and output domains^{252,265}. The weak dimerisation interaction between PAS domains may facilitate the far larger conformational changes upon ligand binding than those observed for Dos or FliL^{116,252,258,261,264}. Small-angle X-ray scattering (SAXS) studies have demonstrated that the HAMP domains of Aer2 are arranged linearly, preventing direct interaction between the PAS domain and the MA domain, supporting a role for HAMP 4 and 5 propagation of signal to the C-terminal kinase control domain²⁵². Through analogy to other chemotaxis systems (discussed further in section 4.1 within the context of *Pseudomonas aeruginosa*), the MA domain of Aer2 is thought to form a stable complex with CheA2 histidine kinase and the CheW2 docking protein^{265,272–275}. Within this Aer2-CheA2-CheW2 complex, signals from the MA domain of Aer2 could modify the autophosphorylation rate of CheA2, which can then regulate the phosphorylation levels of CheY2^{265,276–278}. The direction of the flagellar motor can then be reversed from anti-clockwise to clockwise through direct interaction with phosphorylated CheY2^{265,279,280}. The proposed signalling cascade for Aer2 is shown in Figure 1.22. Such a cascade has previously been demonstrated for the chemotaxis Che pathway, however Aer2 is part of the orthologous chemotaxis Che2 pathway, where attempts to identify similar behaviour have been unsuccessful^{265,267,281}. It is therefore possible that Aer2 may also be involved in a different signalling pathway.

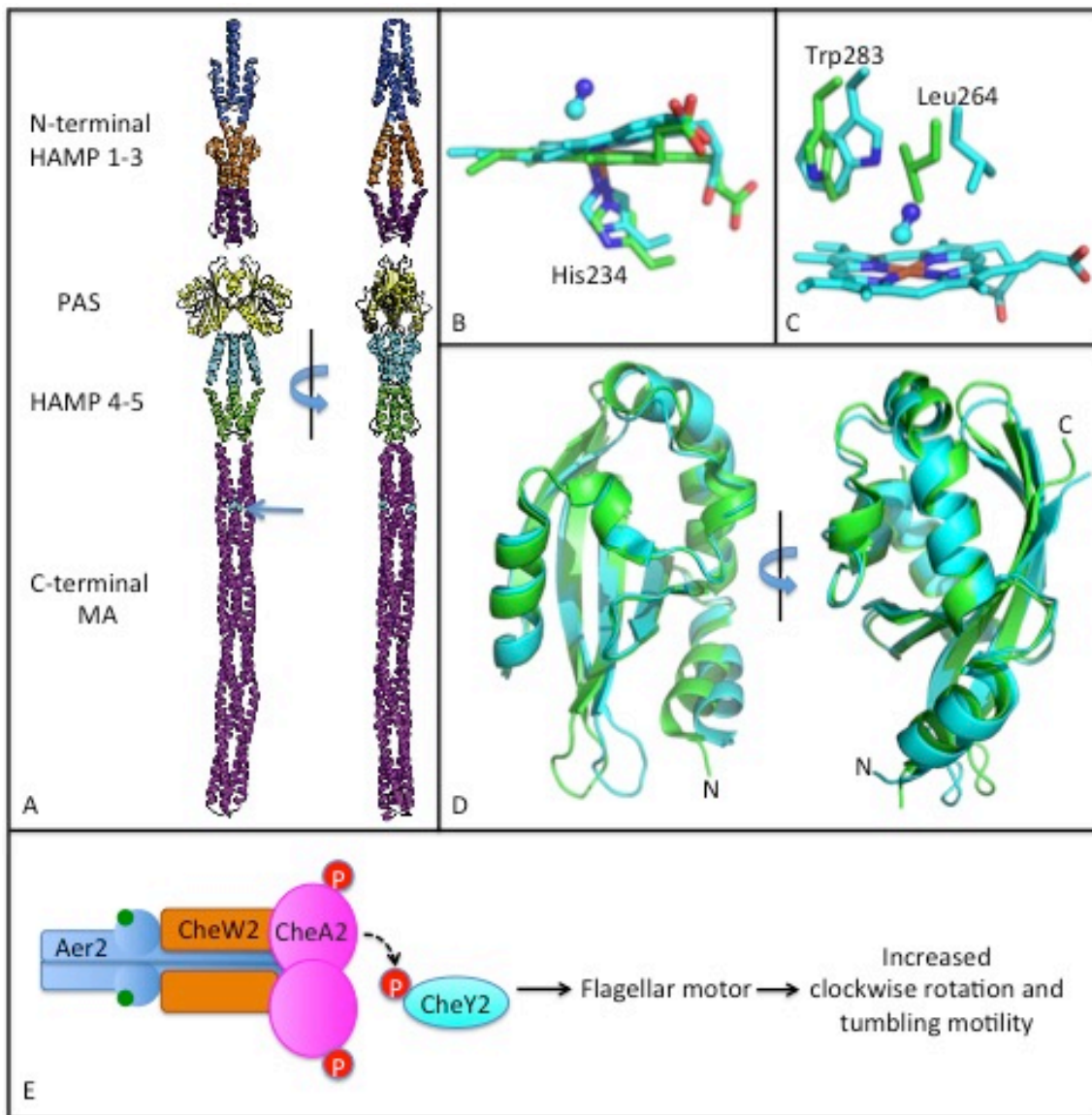


Figure 1.22 - The haem PAS domain containing protein Aer2. A) The structure of the Aer2 dimer, as determined with SAXS from the N-terminus to the point denoted with a blue arrow. The remainder of the MA domain is modelled on the crystal structure of *Thermotoga maritima* 1143 (PDB 2CH7). Taken from Airola *et al.* 2013²⁵². B) Binding of cyanide (used as a model of other diatomic ligands) causes a shift in the puckering of the haem, also shifting the proximal His234 residue. C) Binding of cyanide also induces large rotations in Trp283 and Leu264. D) Reorientations upon cyanide binding shift the overall positions of many secondary structure elements within the PAS domain. The ligand free structure of Aer2 (PDB 4HI4) is modelled in green in B, C and D, with the cyanide bound structure (3VOL) modelled in blue. E) The binding of diatomic gas to the haem PAS domain of Aer2 increases the autophosphorylation, of CheA2, which then phosphorylates CheY2. Phosphorylated CheY2 then interacts with the flagellar motor to increase clockwise rotation of the flagellum and induce tumbling motility.

Further *b*-type haem binding PAS domains have been identified within YybT proteins from *Bacillus subtilis* and *Geobacillus thermodenitrificans* through absorption spectroscopy²⁸².

Ultraviolet-visible light (UV-Vis) spectroscopy can be used to demonstrate the presence of porphyrin rings. Porphyrin rings contain a system of aromatic π -electrons which can transition to a π^* state due to an electron dipole movement. This can be detected by UV-Vis spectroscopy through an absorption peak at around 400 nm, known as the Soret peak^{283–285}. Despite the presence of Soret peaks characteristic of haem binding when the PAS domain of YybT from *G. thermodenitrificans* is expressed in isolation, sequence alignment with previously identified haem-binding PAS domains shows an absence of the haem coordinating histidine within the PAS domains of both proteins. These are proposed to be the first of a new type of haem-*b* binding PAS domains, which coordinate their cofactor through a new mode of binding yet to be identified²⁸².

The periplasmic, *c*-type haem binding PAS domains have been structurally characterised for two transmembranous proteins involved in regulation of chemotaxis (the movement of the bacterium towards or away a chemical stimulus), *Geobacter sulfurreducens* proteins GSU0582 and GSU0935²⁴⁷. In crystal structures of both proteins the C-terminal β -sheet does not show the binding pocket typically associated with a PAS-domain, instead consisting of 2 N-terminal α -helices and a 4-stranded β -sheet²⁴⁷. Despite being purified as monomers, within both crystals the proteins are arranged as swapped dimers with the two N-terminal helices stacking against the β -sheet of the other monomer. This study suggests that the swapped dimer formation may be used as a mechanism of signal transduction, but this conclusion isn't supported by any other data and the structure may be an artefact of crystal packing²⁴⁷.

The haem is covalently bound by a helical segment between the third and fourth strand of the β -sheet, which is in close proximity to the loop between the N-terminal helices of the other monomer. The swapped dimer is held together around the haem binding site by a salt bridge between the monomers. Within the dimeric crystals of GSU0582, axial coordination of each haem iron is provided by a histidine residue and a water molecule, the latter of which is held in place by main-chain interactions with a methionine residue from the other monomer. Both haems within GSU0582 are high-spin, which has been demonstrated to be concurrent with an oxidised iron in these proteins through optical absorption spectroscopy²⁴⁷. In the crystal of GSU0935 one monomer has similar haem coordination (and is high-spin), whilst the other is low-spin (associated with the iron in a reduced form) and the iron is coordinated by a histidine from the CXXCH motif (conserved amongst *c*-type haem binding proteins, where X is any amino acid) from one monomer and a methionine from the loop between the two N-terminal helices in the other monomer. In this alternative haem-binding conformation no salt bridge between the two monomers is observed²⁴⁷.

The c-type haem binding PAS domains also provide a framework for nitric oxide sensing. NMR and optical spectroscopy studies demonstrated that GSU0582 and GSU0935 can bind NO to the iron when the ion is in both reduced and oxidised states, but upon NO binding the iron enters a low-spin state. This is suggested to be due to the NO displacing the ferric iron coordinating water molecule and displacing the coordinating methionine when the iron is ferrous²⁴⁷.

In addition to Aer2, one other haem binding PAS domain has been identified in *P. aeruginosa*, this is within the protein BdlA (biofilm dispersal locus A, PA1423). BdlA does not contain the CXXCH motif characteristic of haem-c coordinating proteins, and sequence alignment with *E. coli* Dos and *B. japonicum* FixL does not display conservation of the histidine which serves to axially coordinate the haem iron, suggesting that BdlA may bind haem through a method similar to YybT proteins discussed above, which is yet to be structurally determined.

1.4.3 BdlA

BdlA was identified as essential for nutrient-induced biofilm dispersal through a screen of functional knock-out mutations of chemotaxis transducer proteins within *P. aeruginosa*²⁸⁶. Subsequent knockout mutants identified that BdlA is also required for nitric oxide induced biofilm dispersal¹⁰⁴. Biofilms consisting of *bdlA* knock-out mutants display increased levels of c-di-GMP as well as reduced swimming motility and enhanced adhesiveness in comparison to biofilms of wild-type *P. aeruginosa*²⁸⁶. However, despite the clear link between BdlA and intracellular c-di-GMP concentration, BdlA does not contain any enzymatic domains directly involved in c-di-GMP metabolism. Instead, BdlA consists of two PAS domains and a methyl-accepting (MA) domain involved in chemotaxis signalling^{286,287}. The first of these PAS domains was identified as binding haem through UV-Vis spectroscopy²⁸⁸. It has been proposed on receipt of a stimulus the PAS domains of BdlA dimerise with PAS domains of proteins with phosphodiesterase activity, namely DipA (dispersion-induced phosphodiesterase A) or RbdA (regulation of biofilm disposal A), and that this dimerisation promotes PDE activity to reduce levels of c-di-GMP and induce dispersal^{287–290}.

BdlA is inactive in its native state, with post-translational modifications required for signalling²⁸⁷. Use of LC-MS/MS analysis identified small fragments of BdlA present after dispersion as the first PAS domain on its own and a truncated form of BdlA consisting of just the second PAS domain and the MA domain. A strategy of co-immunoprecipitation and 2D/PAGE identified PA0459 (ClpD) as a potential binding partner of BdlA with overexpression of ClpD resulting in reduced levels of full-length BdlA and ClpD knockouts containing more full-length BdlA, but these were insensitive to nutrient-induced biofilm dispersal²⁸⁷. ClpD is a homologue of *E. coli* chaperone proteins ClpA

and ClpB, the former of which is involved in the regulation of the ClpAP protease, responsible for degradation of protein aggregates. A homologue of ClpP encoded close to ClpD in the PAO1 genome (PA0451) is also essential for BdlA cleavage, suggesting PA0451 may be the protease responsible for BdlA processing. Site-directed mutagenesis of Met130 and Ala131, a sequence motif targeted by *E. coli* ClpP protease²⁹¹ and located between the two PAS domains, alters the cleavage of BdlA, with mutants of this site unable to restore biofilm dispersion in BdlA KO mutants, suggesting this as a potential sight for ClpD interaction²⁸⁷.

Complementation assays demonstrated that both fragments of cleavage (the first PAS domain and the second PAS domain with the MA domain) are required for dispersion, and pull-down assays and cross-linking demonstrated the two fragments interacted through PAS domain dimerisation. Site-directed mutagenesis demonstrated the requirement of Tyr238 for BdlA processing and biofilm dispersion, with LC-MS/MS identifying this as a phosphorylation site²⁸⁷. As phosphorylation of Tyr238 is required for cleavage of BdlA and for subsequent signalling in biofilm dispersal, phosphorylation of Tyr238 is suggested to be the signal that initiates BdlA processing.

It has recently been determined that the post-translational modifications of BdlA are induced by an increase in local c-di-GMP concentration, potentially controlled by GcbA (PA4843). GcbA contains a response regulator receiver (REC) domain and a DGC domain, but complementation of gene-knock-out mutants with an inactive DGC in GcbA failed to restore BdlA processing to wild-type levels, with other active DGCs unable to compensate for GcbA, suggesting the DGC activity of GcbA has a role in BdlA phosphorylation and cleavage²⁹². GcbA is suggested to increase c-di-GMP concentrations upon the transition from reversible attachment to irreversible attachment, triggering the processing of BdlA only within developing and developed biofilms, ready to activate dispersion in response to environmental stimuli^{287,292}. A diagrammatic representation of the post-translational processing of BdlA is shown in Figure 1.23.

Lifestyle:

Planktonic

Biofilm

Dispersed

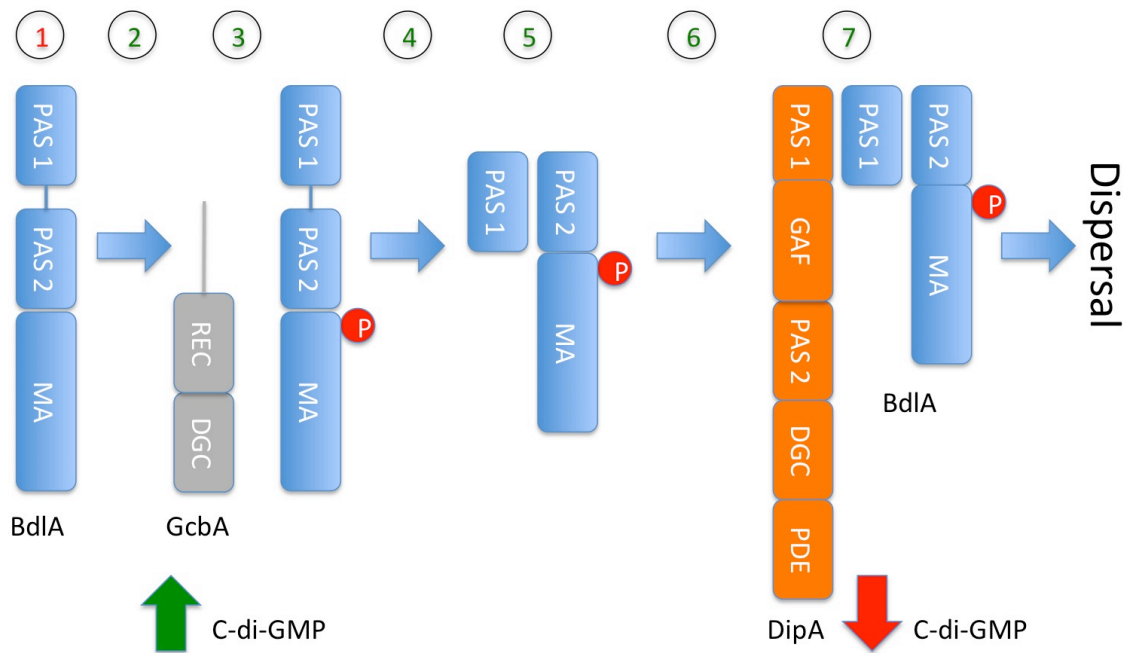


Figure 1.23 - Proposed BdlA processing and signalling pathways. 1) BdlA is expressed by bacteria in a planktonic phenotype as a multi-domain protein consisting of two PAS domains (PAS 1 and PAS 2) and a MA chemotaxis domain (MA). 2) Biofilms begin to form as the bacteria irreversibly attaches to a surface. 3) During irreversible attachment, the local concentration of c-di-GMP is increased by DGCs including GcbA, triggering phosphorylation of BdlA at Tyr238 by an unknown kinase. 4) A combination of ClpD and ClpP recognise phosphorylated BdlA and catalyse the cleavage of BdlA at Met130. 5) The two fragments of BdlA dimerise. 6) BdlA processing ceases while the bacteria remains in a biofilm. 7) Dispersal stimuli induce interaction between BdlA and DipA. This interaction activates the PDE of DipA, reducing the concentration of c-di-GMP and inducing dispersal.

1.4.4 MHYT domains

First identified through sequence analysis, MYHT domains (so called after a characteristic amino acid motif) are a type of predicted redox sensory domain²⁰⁴. The MHYT domain is predicted to consist of 6 transmembrane helices connected by short arginine-rich cytoplasmic loops and highly charged periplasmic loops. Within the second, fourth and sixth transmembrane helices are conserved residue motifs of MHYTXM (where X represents any amino acid) near the periplasmic face, and it is suggested that these motifs and some of the surrounding conserved residues may

coordinate metal ions within the transmembrane region²⁰⁴. From the high levels of methionine and histidine conservation it is suggested that these residues contribute to the formation of a binuclear copper centre, with each metal held in a three- or four-coordinated metal-binding site, however the nature(s) of the bound metals is yet to be determined²⁰⁴.

If the MHYT does indeed bind copper ions then this domain may serve to sense changes in redox state²⁰⁴. Redox sensation could be achieved through conformational changes brought about by the oxidation state of the copper ions changing directly in response to the change of environmental redox state, or through the copper ions binding O₂, CO or NO as differing quantities of these are present when the cell is in different oxidation states. If the latter of these two models for detecting changes in oxidation state is true then conformational changes could be elicited directly upon NO binding, however either model could be appropriate for transmitting a perceived change in periplasmic redox state to the cytoplasm or to subsequent domains within the protein²⁰⁴.

Two MHYT domain containing proteins have been identified within *P. aeruginosa*, MucR (Mucoid alginate regulator, PA1727) and NbdA (NO-induced biofilm dispersion locus A, PA3311), both of which are coupled to c-di-GMP metabolising domains²⁰⁴.

1.4.5 MucR and NbdA

Gene knockout studies have demonstrated that MucR and NbdA are required for NO-induced *P. aeruginosa* biofilm dispersion¹⁵⁵. Both MucR and NbdA contain a GGDEF domain and an EAL domain. Although the GGDEF domain of NbdA has not been demonstrated to display DGC activity, this domain has been suggested to be involved in regulation of PDE activity via the EAL domain within NbdA^{153,155,156}. Hydrophobicity analysis suggests that both proteins have 7 transmembranous N-terminal helices, with the first through to the sixth involved in the formation of an MHYT domain. Cloning with reporter enzymes has been used to demonstrate that MucR is localised to the cytoplasmic membrane, with the protein N-terminus in the periplasm and the C-terminal GGDEF and EAL domains exposed to the cytosol¹⁵³. It is assumed that NbdA has similar localisation and orientation, and it has been suggested that from this position the MHYT domain of each protein may act as a NO sensor¹⁵⁵.

MucR has previously been implicated as being involved in regulation of alginate synthesis, with Δ *mucR* gene knockout studies in PDO300 (a mutant *P. aeruginosa* strain of PAO1) showing a 38-fold decrease in alginate production. This reduction in alginate production was not recoverable by overexpressing another active DGC (WspR), whilst overexpression of an active PDE (PA3947) in the knockout mutant completely abolished alginate synthesis¹⁵³. As the C-terminal cytosolic

region was demonstrated to be essential for alginate synthesis, this evidence supports a model where MucR regulates local levels of c-di-GMP around the cytosolic membrane, rather than regulating the c-di-GMP levels across the whole organism. These high levels of local c-di-GMP are thought to interact with the PilZ domain of Alg44, an alginate synthesising protein also localised to the cytoplasmic membrane¹⁵³.

Nitroactive intermediates from cell denitrification, such as nitrate, have previously been linked to suppressed alginate production in *P. aeruginosa*^{156,293}. When grown on nitrate-containing media PDO300 alginate production was suppressed, with the presence of excess alginate precursor (GDP-mannuronic acid) suggesting that this suppressed production is due to inhibited polymerisation. As alginate polymerisation is normally catalysed by the inner membrane proteins Alg8 and Alg44 in response to high levels of c-di-GMP, and MucR is thought to modulate the levels of c-di-GMP that control Alg44, it is possible that MucR may act as a sensor for nitroactive intermediates^{152–154,294,295}.

The theory of MucR acting as a nitrate sensor for alginate production is supported by recent observations of unchanged alginate yield in response to nitrate when MucR is knocked-out, a change which is restored when MucR is reintroduced¹⁵⁶. Point mutations to catalytic residues have been used to identify both the GGDEF and EAL domain of MucR as required for nitrate-induced suppression of alginate production¹⁵⁶. It is possible that nitrate is denitrified to release NO and shift redox potential, either of which may be detected by the MHYT domain of MucR. This theory is supported by single, double and triple His-Ala mutants of the MHYT motifs in MucR, which demonstrate that if the second MHYT motif or any two of the three MHYT motifs are mutated to MAYT then nitrate-induced suppression of alginate production is abolished¹⁵⁶. These His-Ala mutants within the MHYT motif of MucR may also provide evidence for MYHT domains binding two copper ions, but with only one bound copper required for activity. If a single copper ion is bound to the first and second MHYT motifs, and a second copper ion is bound between the second and third MHYT motifs then mutating the second motif may abolish the binding of both copper ions, explaining the loss of nitrate-induced suppression of alginate-production. If the first and third MHYT motifs contribute to the binding of a single copper ion each then this may explain why a similar phenotype is observed when the first and third MHYT motifs are mutated in combination but not in isolation. However, more extensive characterisation is required to test this theory.

The detection of NO or an altered redox potential by MucR could then alter c-di-GMP metabolism through the GGDEF and EAL domains, most likely in the form of reducing c-di-GMP production of the GGDEF domain or increasing the rate of EAL domain catalysed c-di-GMP catabolism, which

can then suppress alginate production (illustrated in Figure 1.24)¹⁵⁶. It is also possible that if the MHYT domain of MucR does detect NO then it may be able to translate this perception into biofilm dispersal.

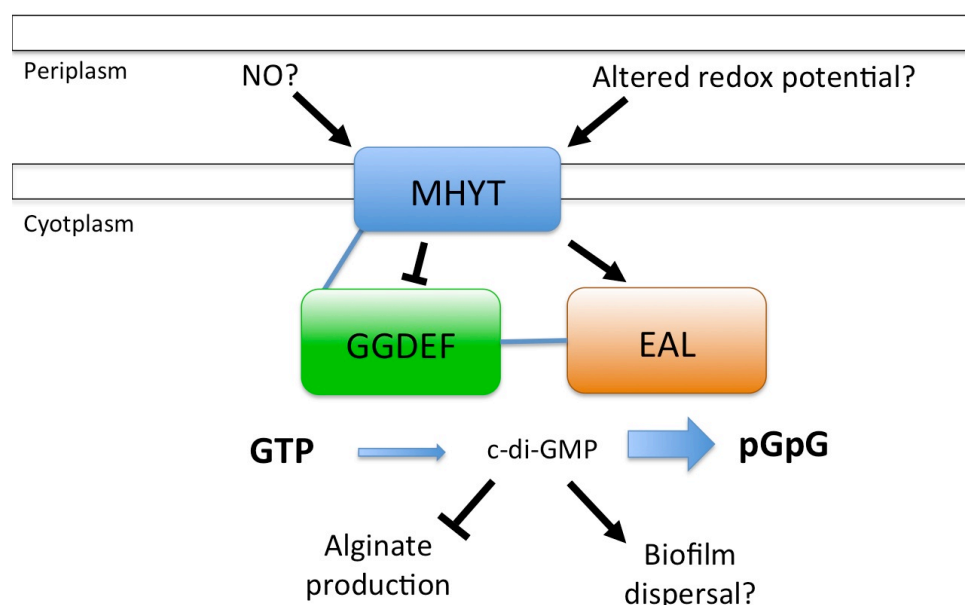


Figure 1.24 - Proposed effect of NO or altered redox potential on MucR. A shift in redox potential or NO is detected directly by the MHYT domain, which then acts to reduce the local concentration of c-di-GMP either through inhibition of GGDEF domain DGC activity or stimulation of EAL domain PDE activity, or both. A reduction in the local concentration of c-di-GMP inhibits alginate production. The same sensory mechanism may be used during NO-induced biofilm dispersal.

A role for NbdA in NO-induced biofilm dispersal has also been suggested, consistent with a lack of NO-induced dispersal in $\Delta nbdA$ biofilms¹⁵⁵. Real-time PCR has been used to demonstrate that transcript levels of NbdA are 4-times higher in biofilms than in planktonic cells. Transcript levels of NbdA were increased further in cells from NO-dispersed biofilms, but a significantly smaller effect was observed in cells from biofilms dispersed by other stimuli. The addition of tobramycin (an inhibitor of translation) to the biofilm before NO exposure prevented changes in c-di-GMP levels within the cell¹⁵⁵. Although there is clearly a strict level of transcriptional control over NbdA, it is possible that there may also be a post-translational interaction between NO and NbdA to increase phosphodiesterase activity.

Chapter 2: Materials and Methods

2.1 Experimental methods

2.1.1 High-throughput molecular biology at Oxford Protein Production Facility (OPPF-UK)

Transmembranous fragments of MucR and NbdA (detailed in Figure 2.1 and Table 2.1) were amplified from the genome of *P. aeruginosa* PAO1 using the polymerase chain reaction (PCR), confirmed by resolution of a single appropriately sized band upon gel electrophoresis in a 1.8% agarose gel. Due to wide variety in size and base-composition of the selected protein fragments, three different conditions for the PCR reaction were required to amplify DNA of all fragments. Details of each reaction are given below, with the reaction that was confirmed to exclusively amplify the fragment in question listed in Table 2.1. Four fragments were not amplified exclusively and were gel purified according to a Qiagen gel purification kit, these fragments are denoted gel purified in Table 2.1.

During PCR 1, 3 μ L each of fragment specific forward and reverse primer (at 10 μ M) were added to 0.5 μ L of genomic *P. aeruginosa* PAO1 DNA (at 1 μ g per μ L), 25 μ L of Phusion flash polymerase master mix (Thermo Scientific) and 18.5 μ L of water, to give a 50 μ L final reaction volume. Reactions were initially heated to 98°C for 5 minutes to instigate an initial DNA denaturation, before a reaction cycle of DNA denaturation at 98°C for 1 second, primer annealing at 60°C for 5 seconds and DNA extension at 72°C for 45 seconds was implemented 40 times. A further DNA extension step at 72 °C for 1 minute was used to complete the reaction.

During PCR 2, 3 μ L each of fragment specific forward and reverse primer (at 10 μ M) were added to 0.5 μ L of genomic *P. aeruginosa* PAO1 DNA (at 1 μ g per μ L), 25 μ L of 2x KOD Xtreme Hot Start DNA polymerase sample buffer (Novagen), 10 μ L of mixed dNTPs (2 mM), 1 μ L KOD Xtreme Hot Start DNA polymerase (Novagen) and 7.5 μ L of water, to give a 50 μ L final reaction volume. Reactions were initially heated to 94°C for 2 minutes to instigate an initial DNA denaturation, before a reaction cycle of DNA denaturation at 98°C for 10 seconds, primer annealing at 65°C for 7 seconds and DNA extension at 68°C for 150 seconds was implemented 40 times. A further DNA extension step at 68 °C for 3 minutes was used to complete the reaction.

During PCR 3, 3 μ L each of fragment specific forward and reverse primer (at 10 μ M) were added to 0.5 μ L of genomic *P. aeruginosa* PAO1 DNA (at 1 μ g per μ L), 12.5 μ L of 2x CloneAmp HiFi PCR

Chapter 2

Premix (Clontech) and 9 μL of water, to give a 25 μL final reaction volume. PCR 3 reactions progressed using the same thermocycling protocol as PCR 2.

PCR products were subsequently purified using AMPure XP magnetic bead purification. 90 μL of AMPure XP magnetic bead solution (Agencourt/ Beckman) was added to each PCR reaction before thorough mixing and incubation at room temperature for 5 minutes, allowing DNA fragments larger than 100 bp to bind to the magnetic beads. After binding, a magnet is placed under each reaction for 5 minutes to allow the DNA-bound beads to pellet before removal of solution and unbound PCR reactants. Pelleted magnetic beads were washed twice with 300 μL of 70% ethanol, followed by removal of all ethanol by pipetting and evaporation for 10-20 minutes. Amplified DNA was subsequently eluted from the magnetic beads using 30 μL of elution buffer (10 mM Tris at pH 8.0) and transferred to a separate 96-well plate.

Purified PCR products were inserted into pOPINE-3C-GFP vectors using ligation-independent In-Fusion cloning. A vector map of pOPINE-3C-GFP is shown in Figure 2.2. 1 μL of linearised pOPINE-3C-GFP (100 ng), 2 μL of purified PCR product and 7 μL of water were added to lyophilised In-Fusion enzyme (Clontech). The pOPINE-3C-GFP vector encodes for a methionine immediately N-terminal to the gene of interest, as well as a C-terminal tag consisting of a 3C protease cleavage site, a green fluorescent protein (GFP) and six histidines, with expression of this tagged protein fragment under the control of a lactose inducible promoter. The In-Fusion reaction was then allowed to proceed at 42°C for 30 minutes, before each reaction was deactivated through the addition of 40 μL of TE buffer. Ligation-independent cloning with In-Fusion requires a sequence of 15 base pairs on each terminus of the target gene which are homologous to equivalent regions on the linearised vector. The In-Fusion enzyme degrades these sequences of base pairs into single stranded 15-nucleotide overhangs, which are complimentary between the insert and the vector and subsequently anneal to one another. The expression plasmid is circularised within *E. coli* following transformation.

Plasmids were subsequently transformed into OmniMaxII competent cells (ThermoFisher) using heat-shock. 5 μL of inactivated In-Fusion reaction was added to 50 μL of OmniMaxII competent cells before incubation on ice for 30 minutes, heat shock at 42°C for 30 seconds, a further 2 minute incubation on ice, addition of 300 μL of SOC media and incubation at 37°C for 1 hour. Following incubation, 25 μL of each transformation was spread onto 1 mL lysogeny broth (LB) agar plates, supplemented with 50 $\mu\text{g}/\text{ml}$ carbenicillin, 500 μM IPTG and 0.02% 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal). 5 μL of each transformation was added to 20 μL of LB media before spreading onto further 1 mL LB agar plates, supplemented as before. LB agar plates were incubated at 37°C overnight.

Colonies grown on agar plates were inoculated into 1.2 mL Power Broth which was then incubated at 37°C overnight. Cells were harvested by centrifugation at 5,000 g for 15 minutes in an Avanti J-20 XPI centrifuge before plasmid DNA was extracted using a miniprep kit (Promega) and a Qiagen BioRobot 8000, according to manufacturers instruction. Cloning was verified through PCR using a protocol similar to PCR 1, except with an annealing temperature of 60°C, 1.5 µL of purified plasmid used as template DNA, a reverse primer specific for each inserted fragment and a forward primer for the T7 forward priming sequence encoded by the pOPINe-3C-GFP vector.

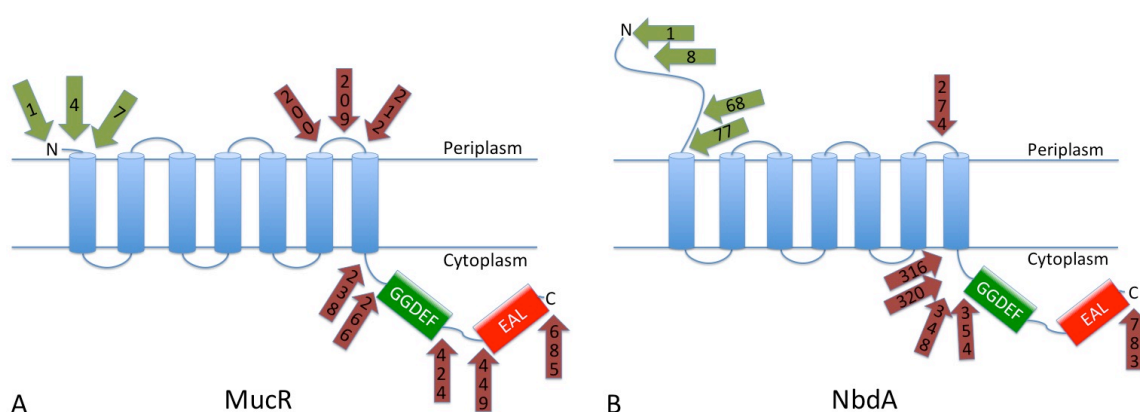


Figure 2.1 – Graphical overview of designed MucR and NbdA gene fragments cloned at OPPF-UK and further detailed in Table 2.1. Transmembrane helices of the MHYT domain are shown as blue cylinders, DGC and PDE domains are shown as green and red boxes respectively. Green arrows denote N-terminal construct residues and red arrows denote C-terminal residues.

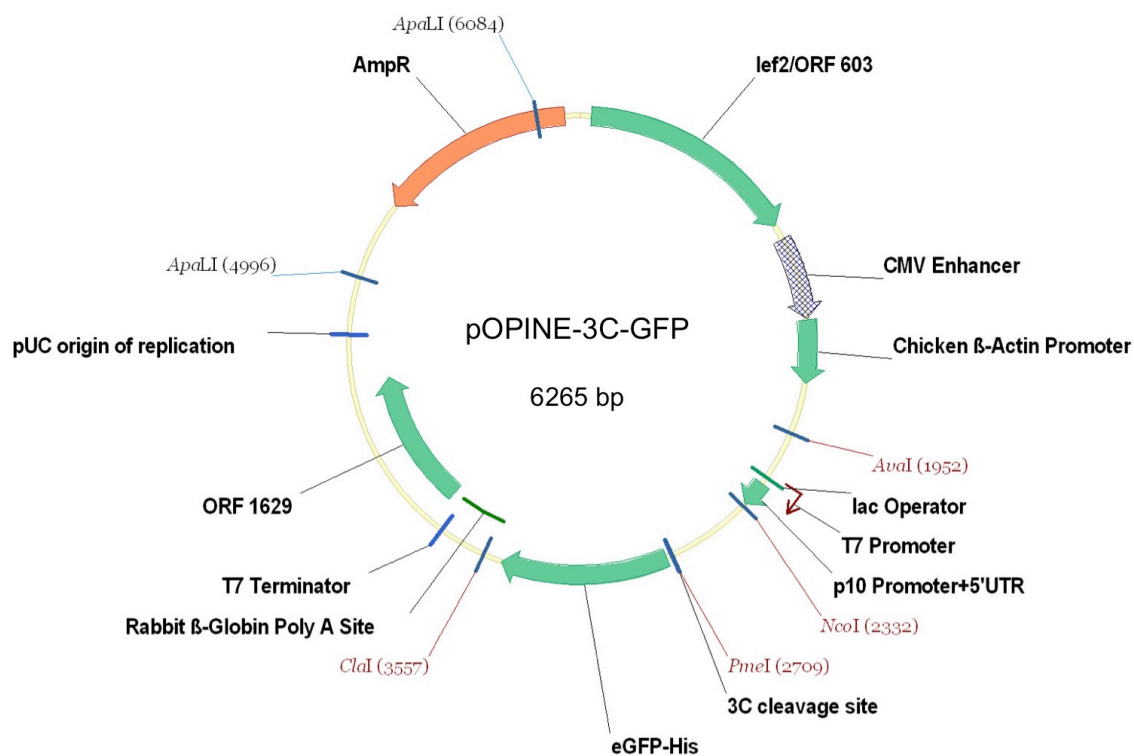


Figure 2.2 – Vector map of the pOPINE-3C-GFP vector. Restriction enzyme recognition sites are denoted by the number of the base cleaved in the 5'-3' direction. Vector map obtained from addgene.org, under the original name of pOPINE-3C-eGFP²⁹⁶.

Table 2.1 - PCR conditions used to clone transmembranous MucR and NbdA fragments

Gene name	N-terminal AA	C-terminal AA	Confirmed PCR reaction
MucR	1	200	2
MucR	1	209	2
MucR	1	212	2
MucR	1	238	2
MucR	1	266	2
MucR	1	424	2
MucR	1	449	Gel purified 2
MucR	1	685	Gel Purified 2
MucR	4	200	2
MucR	4	209	2
MucR	4	212	2
MucR	4	238	2
MucR	4	266	2
MucR	4	424	2
MucR	4	449	Gel purified 2
MucR	4	685	Gel Purified 2
MucR	7	200	2
MucR	7	209	2
MucR	7	212	2
MucR	7	238	2
MucR	7	266	2
MucR	7	424	2
MucR	7	449	1
MucR	7	685	2
NbdA	1	274	2
NbdA	1	316	2
NbdA	1	320	2
NbdA	1	348	2
NbdA	1	354	2
NbdA	1	783	3
NbdA	8	274	2
NbdA	8	316	2
NbdA	8	320	1
NbdA	8	348	3
NbdA	8	354	3
NbdA	8	783	3
NbdA	68	274	3
NbdA	68	316	2
NbdA	68	320	2
NbdA	68	348	2
NbdA	68	354	1
NbdA	68	783	2
NbdA	77	274	2
NbdA	77	316	2
NbdA	77	320	2
NbdA	77	348	2
NbdA	77	354	2
NbdA	77	783	2

2.1.2 Small scale expression of transmembranous protein fragments at OPPF-UK

Cloned plasmids were transformed into four different *E. coli* expression cell lines, single step KRX competent cells (Promega), Lemo21(DE3) (NEB), C41(DE3) pLysS and C43(DE3) pLysS (Sigma). For each fragment and each cell line, 3 μ L of expression plasmid was added to 50 μ L of competent cells, before a 30 minute incubation on ice, a 30 second heat shock at 42°C and a further 2 minute incubation on ice. Each transformation was then supplemented with 300 μ L of Power Broth (Molecular Dimensions) before a 1 hour incubation at 37°C and spreading 30 μ L onto 1 mL LB agar plates. All LB agar plates were supplemented with 50 μ g/ mL carbenicillin, with agar plates for Lemo21(DE3), C41(DE3) pLysS and C43(DE3) pLysS also supplemented with 34 μ g/ mL chloramphenicol. Each transformation was incubated at 37°C overnight. Expression starter cultures were grown from picked bacterial colonies incubated in 750 μ L of Power Broth (supplemented with the same antibiotics as were used with LB agar) at 37°C overnight. 150 μ L of each starter culture was added to 3 mL of Power Broth, supplemented with antibiotic as before, and incubated in a shaking incubator at 37°C for 6 hours before expression was induced through the addition of 1 mM IPTG. Growths for Lemo21(DE3) cells were also supplemented with 25 μ M L-rhamnose as starter culture was added, whilst 2.5 mM L-rhamnose was added to KRX cultures at the time of induction. Following induction, cultures were transferred to a shaking incubator at 20°C and incubated overnight.

1 mL of each expression culture was harvested by centrifugation at 6,000 g in an Avanti J-20 XPI centrifuge for 10 minutes before removal of supernatants and storage at -80°C. Harvested cells were thawed at room temperature for 20 minutes before resuspension in 200 μ L of lysis buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, 1% v/v Tween 20, pH 8.0) and 20 μ L of enzyme solution (1 mg/ mL lysozyme, 400 units/ mL DNase I and 10 μ L/ mL protease inhibitor), followed by shaking incubation at room temperature for 15 minutes. 25 μ L of 10% n-dodecyl β -D-maltoside (DDM) was then added to each sample before a further 1 hour incubation at 4°C and centrifugation in an Avanti J-20 XPI centrifuge at 5,000 g for 30 minutes. 10 μ L of each centrifugation supernatant was mixed with 10 μ L of SDS PAGE gel loading buffer (100 mM Tris, 4% w/v SDS, 0.2% w/v bromophenol blue, 10% v/v β -mercaptoethanol, 20% glycerol, pH 6.8), with 10 μ L of this mixture resolved through SDS-polyacrylamide gel electrophoresis (SDS-PAGE). Expected molecular mass of each fragment was calculated using the ProtParam web server²⁹⁷. Expression of GFP tagged protein was identified by examining in-gel fluorescence, with excitation at 395 nm and detection at 509 nm.

2.1.3 Large scale expression of transmembranous MucR fragments at Diamond Membrane Protein Lab (MPL)

MucR⁴⁻⁶⁸⁵ was transformed into Lemo21(DE3) cells (Sigma) and MucR⁷⁻²⁶⁶ was transformed into single step KRX competent cells (Promega) using the heat shock transformation protocol detailed in section 2.1.2. Following transformation, colonies of each fragment were grown in 20 mL of Power Broth with antibiotic (50 µg/ mL carbenicillin for KRX cells, 50 µg/ mL carbenicillin and 34 µg/ mL chloramphenicol for Lemo21(DE3) cells) at 37°C overnight. 20 mL starter cultures were each added to 1 L of Power Broth supplemented with 10 µM copper chloride and appropriate antibiotic. The culture of MucR⁴⁻⁶⁸⁵ in Lemo21(DE3) cells was also supplemented with 250 µM L-rhamnose. Both cultures were incubated in a New Brunswick Innova 43/43R shaking incubator at 37°C, 250 rpm. Upon reaching an optical density (as measured through absorbance at 600 nm, A₆₀₀) of 1.4 the culture of MucR⁴⁻⁶⁸⁵ in Lemo21(DE3) cells was transferred to a 25°C incubator shaking at 250 rpm for 10 minutes, before induction with 1 mM IPTG. The culture of MucR⁷⁻²⁶⁶ in KRX cells was grown to an O.D. of between 1.2 and 1.8 before transferring to a 25°C incubator shaking at 250 rpm for 10 minutes and subsequent induction with 1 mM IPTG and 2 mM L-rhamnose. Cultures were incubated overnight before harvesting through centrifugation in a Sorval evolution RC centrifuge with an F8 rotor at 8,000 g, 4°C, with supernatant discarded and pellet retained for membrane harvesting.

2.1.4 Membrane harvesting at MPL

Following expression, each cell pellets harvested from 1 L of cell cultures was resuspended in 50 mL of breaking buffer and kept on ice. Breaking buffer was composed of 1x phosphate buffered saline (PBS, consisting of 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 2.7 mM KCl 137 mM NaCl, pH 7.4), 1 mM MgCl₂, 10 µM CuCl₂, 0.5 mg/ mL DNase 1 and 1 complete EDTA free protease inhibitor tablet per 200 mL of buffer. Cells were lysed using a cell press at 25,000 PSI, 4°C, followed by removal of unlysed cells through centrifugation at 24,000 g for 12 minutes at 4°C, using an Avanti J-26 XPI centrifuge with a JLA-16.250 rotor. Centrifugation supernatant was then further centrifuged at 190,000g for 1 hour. Pellets from high-speed centrifugation were resuspended in 10 mL of 1x PBS, followed by homogenisation and subsequently flash freezing in liquid nitrogen and storage at -80°C.

2.1.5 Fluorescence size-exclusion chromatography at MPL

Membranes harvested and resuspended according to 2.1.4 were solubilised in 2% 6-cyclohexylhexyl β -D-maltoside (CYMAL-6), 1% n-decyl- β -D-maltoside (DM), 1% DDM, 1% Fos-choline-12 (FC-12), 1% lauryldimethylamine *N*-oxide (LDAO) or 2% octyl glucoside (OG). Resolubilisation was achieved by adding 500 μ L of harvested membrane to 50 μ L of 10x PBS, 3 μ L of 100 mM CuCl_2 and 100 μ L of 10 % detergent (in the cases of CYMAL-6 and OG, 200 μ L of detergent were used), with volume adjusted to 1 mL using water. Membrane-detergent mixtures were incubated for 2 hours at 4°C, before centrifugation at 150,000 g, 4°C for 45 minutes, using a Beckman Coulter Optima L-100XP ultracentrifuge with a 26/4 rotor. 120 μ L of each centrifugation supernatant was resolved on a Superose 6 10/300 GL size exclusion column, attached to Äkta Purifier Fast Protein Liquid Chromatography (FPLC) system run at 0.3 mL/min in a buffer of 20 mM Tris, 150 mM NaCl, 10 μ M CuCl_2 , 0.03% DDM, pH 7.5, with elution of green-fluorescent protein (GFP) monitored by examining fluorescence through excitation at 488 nm and emission at 512 nm.

2.1.6 Purification of MucR⁷⁻²⁶⁶ at MPL

Membranes harvested and resuspended according to 2.1.4 were added to MPL purification buffer (1x PBS, 150 mM NaCl, 10 μ M CuCl_2 , 1% DDM and 1 complete EDTA free protease inhibitor tablet per 400 mL), and incubated stirring at 4°C for 2 hours. Resolubilised membranes were isolated by centrifugation at 100,000 g, 4°C, using a Beckman Coulter Optima L-100XP ultracentrifuge with a 26/4 rotor. Resolubilised protein was purified with immobilised metal affinity chromatography (IMAC). Centrifugation supernatant was added to 10 mM imidazole and 10 mL of Ni-NTA resin before stirring incubation at 4°C for 2 hours. Following incubation, contaminating proteins were washed from the nickel resin using 10 mM, 20 mM and 30 mM imidazole in buffer A (1x PBS, 150 mM NaCl, 10 μ M CaCl_2 , 0.1% DDM), with 50 mL of each solution sequentially washed over the resin. MucR⁷⁻²⁶⁶ was eluted from the Ni-NTA resin using 25 mL of buffer A supplemented with 250 mM imidazole. Imidazole was subsequently removed from the sample through dialysis against buffer B (20 mM Tris, pH 7.5, 150 mM NaCl, 0.03% DDM and 10 μ M CuCl_2), using an 8,000 Da molecular weight cut off dialysis membrane. 4 mg of His-tagged 3C protease was also added to cleave the C-terminal tag and GFP-His tag was cleaved from the expressed MucR⁷⁻²⁶⁶ fragment. Cleavage and dialysis were allowed to proceed overnight at 4°C.

Dialysed sample was centrifuged at 3,900 g, 4°C in a bench top centrifuge for 10 minutes to remove precipitation, before supernatant was loaded onto a 5 mL HisTrap column for reverse phase immobilised metal affinity chromatography (reverse IMAC). As MucR⁷⁻²⁶⁶ was no longer

nickel affinity tagged, flow-through was collected and concentrated in a 100,000 Da molecular weight cut off spin concentrator to a final volume of 500 μ L. Sample was then centrifuged at 18,000 g, 4°C in a bench top micro-centrifuge for 10 minutes before further purification through size exclusion chromatography, using a Superdex 200 10/300 column in buffer B attached to an Äkta Purifier FPLC system at 0.4 mL/ min.

2.1.7 Fluorescence quantification at MPL

Samples of 100 μ L were taken before and after 100,000 g centrifugation of solubilised membranes, from IMAC flow-through, from IMAC wash fractions and from IMAC eluate. Samples were added to a clear bottomed, 96 well block and measured in a plate reading fluorimeter, recording fluorescence of GFP with excitation at 488 nm and excitation at 512 nm.

2.1.8 SEC-Multi-angle light scattering (MALS) at MPL

MucR⁷⁻²⁶⁶, purified according to 2.1.6 was concentrated to 36.4 μ M (as determined by absorbance of light at 280 nm, using a NanoDrop spectrophotometer) in a volume of 160 μ L before resolution on a Superdex 200 Increase 10/300 column in buffer B at 0.5 mL/ min, attached to a Viscotek TDMax system (Malvern) SEC-MALS system. The Viscotek TDMax system contains a triple detector array (TDA-305) that measures refractive index, multi-angle laser light scattering (MALS) and a detector of absorbance at 280 nm in series. Maria Rosa at MPL operated the SEC-MALS system.

2.1.9 Thermostability analysis at MPL

MucR⁷⁻²⁶⁶, purified according to 2.1.6 was concentrated to 229.7 μ M (as determined by absorbance of light at 280 nm, using a NanoDrop spectrophotometer), before addition of 0.2 mg/ mL N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) dye and buffer B in various ratios, listed in Table 2.2. CPM is a fluorochrome dye that is not fluorescent in solution, but fluoresces upon binding to thiol groups. MucR⁷⁻²⁶⁶ contains 3 cysteines, two of which are predicted to be located within transmembranous helices^{298,299}. As these cysteines are inaccessible to CPM while the protein is folded, increase in fluorescence caused by binding of CPM to the cysteine residues can be used to determine the temperature of protein denaturation and can therefore be used to screen a range of conditions to identify those that help stabilise the protein. The fluorescence of protein-CPM mixtures was determined across a linear temperature gradient between 25°C and 95°C using a Q PCR machine.

Table 2.2 - Composition of CPM thermostability reactions

Well	Component (μ L)		
	Protein	CPM dye	Buffer B
A1	0.5	5	44.5
A2	0.5	7	42.5
A3	1	5	44
A4	1	7	42
A6	2	7	41
A7	4	5	41

2.1.10 Cloning of soluble protein fragments

Genes encoding for MucR^{GGDEF}, MucR^{GGDEF-EAL}, MucR^{EAL}, NbdA^{GGDEF-EAL}, NbdA^{EAL}, BdlA¹⁻⁴¹⁷, BdlA¹⁻¹²⁷ and BdlA⁸⁻¹²⁷ were amplified from genomic *P. aeruginosa* PAO1 genome, extracted from cultures grown at the University of Southampton. 2.5 μ L each of fragment specific forward and reverse primers (detailed in Table 2.3, used at a concentration of 10 μ M) were added to 1 μ L of template DNA (20 μ g/ mL), 1 μ L of mixed dNTPS (10 mM) 4 μ L of DMSO, 10 μ L of 5x Phusion high-fidelity polymerase buffer, 0.5 μ L of Phusion high-fidelity polymerase (NEB) and 28.5 μ L of water to give 50 μ L reactions.

All reactions were initiated by denaturing DNA at 98°C for 180 seconds, before cycling 30 times between a denaturation step of 98°C for 10 seconds, an annealing step for 15 seconds at a temperature specific to the primers (listed in Table 2.3) and an extension step at 72°C for a length of time specific to the fragment length (listed in Table 2.3), followed by a final extension step at

72°C for 5 minutes. PCR products were analysed through gel electrophoresis on a 1% agarose gel run for 35 minutes at a constant voltage of 120 V.

Table 2.3 - PCR conditions used to amplify soluble protein fragments from genomic PAO1 genome.

Gene	Forward PCR primer	Reverse PCR primer	Annealing temperature (°C)	Extension time (seconds)
MucR ^{GGDEF}	MucR 266 fwd	MucR 425 rev	56	15
MucR ^{GGDEF-EAL}	MucR 235 fwd	MucR 685 rev	60	43
MucR ^{EAL}	MucR 425 fwd	MucR 685 rev	58	12
NbdA ^{GGDEF-EAL}	NbdA 332 fwd	NbdA 772 rev	54	43
NbdA ^{EAL}	NbdA 506 fwd	NbdA 772 rev	52	30
BdIA ¹⁻⁴¹⁷	BdIA 1 fwd	BdIA 417 rev	70	38
BdIA ¹⁻¹²⁷	BdIA 1 fwd	BdIA 127 rev	64	12
BdIA ⁸⁻¹²⁷	BdIA 8 fwd	BdIA 127 rev	65	11

Numbers within the PCR primer label denote fragment terminal amino acids (N-terminal for forward primers and C-terminal for reverse primers). The nucleic acid composition of each primer is given in Appendix A, Table 6.1.

Amplified genes were cloned into pET28a expression vectors using ligation-dependent cloning. The pET28a vector contains a lactose inducible promoter of expression, allowing the protein of interest to be expressed with the addition of an N-terminal 6-histidine tag. To insert amplified genes into pET28a, both vector and gene were separately digested with HindIII and NdeI in the presence of a buffer suitable for the activity of both enzymes (supplied by NEB). Digestion consisted of a 2-hour incubation at 37°C followed by a denaturation at 74°C for 25 minutes. After the first hour of 37°C incubation 1 µL of thermosensitive alkaline phosphatase (TSAP, obtained from Promega), at a concentration of 1 mg/mL was added to the 20 µL reaction to prevent vector relegation by removing the 5' phosphate groups from the DNA. PCR-amplified genes were ligated into vectors using T4 DNA ligase and the supplied buffer (obtained from Promega).

To amplify DNA and determine the success of ligation, plasmids were transformed into *E. coli* strain DH5α. 2 µL of plasmid DNA was transformed into 20 µL of competent cells using a heat shock protocol consisting of incubating the cells on ice for 15 minutes, heat shocking at 42°C for 45 seconds and then incubating on ice for a further 2 minutes before adding 200 µL of LB and incubating at 37°C for 1 hour. The DH5α were then plated onto LB agar with 50 µg/ mL kanamycin, selecting only for DH5α cells that had taken up the plasmid and thus the gene for antibiotic resistance. Plates were then incubated at 37°C overnight. Following amplification of vectors by DH5α, plasmid DNA was removed from the bacteria using a QIAGEN miniprep kit and then the presence of inserted DNA was examined using an analytical digest with the same

restriction enzymes used for initial digestion. The results of analytical digestion were determined through gel electrophoresis using a 1% agarose gel run at a constant 120 V for 35 minutes.

2.1.11 Expression of soluble protein fragments

Cloned plasmids were transformed into *E. coli* strain BL21(DE3) for expression, using the heat shock protocol detailed in 2.1.10. Following transformation, expression starter cultures were incubated overnight at 37°C, using 100 ml of LB with 50 µg/ mL kanamycin of the expression. 8 mL of starter culture was added to each 1 litre LB with 50 µg/ mL kanamycin and incubated in a New Brunswick Innova 4330 shaking incubator at 37°C until an A_{600} of between 0.3 and 0.4 was reached. The temperature of the shaking incubator was then reduced to 18°C, with cultures induced with 1 mM IPTG after cells had grown to an A_{600} of between 0.6 and 0.8, before further incubation overnight at 18°C. Cultures were harvested following expression by centrifugation at 4,000 g, 4°C for 20 minutes, using an Avanti J-20 XPI centrifuge, JLA 8.1000 rotor. Following centrifugation cell pellets were resuspended in lysis buffer (50 mM Tris, 200 mM NaCl, 2 mM β-mercaptoethanol, 5% glycerol, pH 7.5).

2.1.12 Expression of MucR^{EAL}

MucR^{EAL} was transformed into *E. coli* strain BL21(DE3) for expression using the heat shock protocol detailed in 2.1.10, with subsequent expression starter cultures grown as detailed in section 2.1.11. 8 mL of starter culture was added to each 1 litre LB with 50 µg/ mL kanamycin and incubated in a New Brunswick Innova 4330 shaking incubator at 37°C until an A_{600} of between 0.6 and 0.8 was reached. Cultures were then induced through the addition of 1 mM IPTG to each litre, and incubated for a further 2 hours at 37°C, before cell harvesting as detailed in 2.1.11.

2.1.13 Purification of soluble MucR and NbdA protein fragments

Harvested cell pellets were lysed through sonication for 5 minutes. Unbroken cells and cell membranes were removed by centrifugation at 30,000 g (Optima XPN-80 Ultracentrifuge, 45 Ti rotor), 4°C for 40 minutes. Desired protein fragments were purified using IMAC. 5 mL of Ni-NTA resin was incubated with centrifugation supernatant for 1 hour in lysis buffer (section 2.1.11) supplemented with 20 mM imidazole. Contaminant proteins were washed from the nickel resin using 20 mL of buffer C (50 mM HEPES, 300 mM NaCl, 2 mM β-mercaptoethanol, pH 7.5) supplemented with 5% glycerol and 80 mM imidazole. Desired protein fragments were eluted from Ni-NTA resin using 25 mL of buffer C supplemented with 5% glycerol and 300 mM imidazole. IMAC eluate was concentrated to a volume of 1.2 mL before centrifugation at 18,000 g for 15

minutes in a bench top micro-centrifuge and further purification through size exclusion chromatography (SEC), using a Superdex 75 16/600 column in buffer C at 1 mL/ min, attached to an Akta Prime plus FPLC system. SEC fractions showing absorbance of UV light at 280 nm were analysed with SDS-PAGE to determine respective protein compositions. Expected molecular mass and molar extinction coefficient of each fragment was calculated using the ProtParam web server²⁹⁷. Samples were added to 2x reducing SDS sample buffer (250 mM Tris, pH 6.8, 8% SDS, 20% glycerol, 600 mM, β -mercaptoethanol, 4 mg/ mL bromophenol blue), before resolving on a 10% tris-tricine gel run at 160 volts for 45 minutes and stained with coomassie blue staining.

2.1.14 Analysis of MucR^{GGDEF-EAL} and MucR^{EAL} PDE activity

Following purification and concentration PDE activity of MucR^{GGDEF-EAL} and MucR^{EAL} was assayed using the protocol previously established by Sundriyal *et al*¹⁹⁶. 3 μ M MucR^{GGDEF-EAL} was added to EAL-PDE buffer, consisting of 25 mM MgCl₂ (used to provide catalytic magnesium ions), 50 μ M thiamine pyrophosphate (TPP, used as an internal standard during nucleotide separation), 50 mM NaCl and 50 mM Tris pH 9.35 (used to stabilise the protein at a pH conducive to maximal activity). PDE reactions (1 mL in total volume) were initiated with the addition of c-di-GMP (200 μ M, 100 μ M or 20 μ M) and incubated at room temperature, with 70 μ L samples withdrawn periodically (every 60 minutes in 200 μ M c-di-GMP reactions, every 30 minutes in 100 μ M reactions and every 6 minutes for 20 μ M reactions) and reactions stopped with the addition of 182 mM CaCl₂, as calcium has been observed to prevent EAL domain catalysed PDE hydrolysis^{110,129,198}. It is worth noting that at room temperature, Tris is not advised as a buffer above pH 9³⁰⁰, and while this work was designed to be consistent with previous studies^{196,301}, future studies using this technique may benefit from a change in buffer. Samples were diluted with 5 mM ammonium bicarbonate (623 μ L ammonium bicarbonate added to 77 μ L of stopped reaction), before 500 μ L of this solution was loaded onto a 1 mL Resource-Q column pre-equilibrated in 5 mM ammonium bicarbonate. Substrate (c-di-GMP) and product (pGpG) were separated across an 18 column volume ammonium bicarbonate gradient (from 5 mM to 1 M), with elution monitored by absorbance of 254 nm light. Nucleotide concentration was determined by integration of nucleotide specific peaks (as identified by comparable retention times to standards, Biolog) using UNICORN software. PDE reactions were repeated in triplicate.

PDE activity of MucR^{EAL} was assayed using the same conditions as MucR^{GGDEF-EAL}, with a substrate concentration of 20 μ M, repeated in triplicate. Samples were taken every 30 minutes for 2 hours, before identical separation as used for MucR^{GGDEF-EAL}.

2.1.15 Analysis of MucR^{GGDEF-EAL} DGC activity

Following purification and concentration, 3 μ M MucR^{GGDEF-EAL} was added to DGC buffer (25 mM MgCl₂, 50 mM NaCl, 50 mM Tris, pH 7.5). Reactions, repeated in triplicate, were initiated with the addition of 50 μ M GTP. 70 μ L of reaction was removed every 30 minutes and the enzyme denatured by heating at 99 °C for 2 minutes. 630 μ L of 5 mM ammonium bicarbonate was added to each 70 μ L sample, before 500 μ L of this solution was loaded onto a 1 mL Resource-Q column for nucleotide separation according to 2.1.14.

2.1.16 Crystallisation of MucR^{EAL}

Following purification, MucR^{EAL} was concentrated to 442 μ M (as determined by absorbance of light at 280 nm, using a NanoDrop spectrophotometer), before centrifugation at 18,000 g for 15 minutes in a bench top micro centrifuge. 0.1 μ L sitting drops crystallisation trials were established using an Art Robbins Instruments Gryphon, with crystallisation screens JCSG+, Morpheus and PACT Premier (Molecular Dimensions). Crystals grew at 21°C in Morpheus crystallisation screen wells D2 (100 mM Morpheus Buffer System 1 pH 6.5, 120 mM Morpheus Alcohols and 50% of an ethylene glycol, PEG 8,000 mix) and E4 (100 mM Morpheus Buffer System 1 pH 6.5, 120 mM Morpheus ethylene glycols and 50% of a MPD, PEG 1000 and PEG 3350 mix), as well as JCSG+ wells H2 (100 mM BIS-Tris pH 5.5, 1 M ammonium sulphate and 1 % PEG 3350) and H3 (100 mM BIS-Tris pH 5.5, 25% PEG 3350). Highest resolution data was collected on I04 (Diamond Light Source) at 100 K, using crystals grown in Morpheus condition E4, with data collection statistics given in Table 3.1.

2.1.17 Purification of BdIA fragments

Harvested cell pellets were lysed through sonication for 5 minutes. Unbroken cells and cell membranes were removed by centrifugation at 100,000 g, 4°C (Optima XPN-80 ultracentrifuge, 45 Ti rotor) for 40 minutes. Desired protein fragments were purified using IMAC. 5 mL of Ni-NTA resin was incubated with centrifugation supernatant for 1 hour in lysis buffer (section 2.1.11) supplemented with 20 mM imidazole. During purification of BdIA⁸⁻¹²⁷ and BdIA¹⁻⁴¹⁷, contaminant proteins were washed from the nickel resin using 20 mL of buffer D (50 mM Tris, 300 mM NaCl, 2 mM β -mercaptoethanol, pH 8.0) supplemented with 5% glycerol and 80 mM imidazole. Desired protein fragments were eluted from Ni-NTA resin using 15 mL of buffer D supplemented with 5% glycerol and 300 mM imidazole. IMAC eluate was concentrated to a volume of 1.2 mL before centrifugation at 18,000 g for 15 minutes in a bench top micro-centrifuge and further purification through size exclusion chromatography (SEC), using a Superdex 75 16/600 column in buffer D at

1 mL/ min, attached to an Akta Prime plus FPLC system. SEC fractions showing absorbance of UV light at 280 nm were analysed with SDS-PAGE to determine respective protein compositions. Samples were added to 2x reducing SDS sample buffer (250 mM Tris, pH6.8, 8% SDS, 20% glycerol, 600 mM, β -mercaptoethanol, 4 mg/ mL bromophenol blue), before resolving on a 10% tris-tricine gel run at 160 volts for 45 minutes and stained with coomassie blue staining. Expected molecular mass and molar extinction coefficient of each fragment was calculated using the ProtParam web server²⁹⁷.

Purification of BdIA¹⁻¹²⁷ was performed using identical protocols to BdIA⁸⁻¹²⁷ and BdIA¹⁻⁴¹⁷, except buffer E (50 mM Tris, 300 mM NaCl, pH 9.0) was used in the place of buffer D.

2.1.18 BdIA haem co-purification through anion exchange

Harvested cell pellets were lysed and centrifuged as detailed in 2.1.17. Centrifugation supernatant was loaded onto a 5 mL HiTrap quaternary ammonium ion column (Q column, GE) attached to an Äkta Prime FPLC system, pre-equilibrated in low salt buffer (50 mM Tris, 20 mM NaCl, 2 mM β -mercaptoethanol, 5% glycerol, pH 7.5). Following sample addition, the Q column was re-equilibrated in low salt buffer before selectively eluting proteins across a 100 mL (20 column volume) gradient between low salt buffer and 50% high salt buffer (50 mM Tris, 1M NaCl, 2 mM β -mercaptoethanol, 5% glycerol, pH 7.5), with elution monitored through absorbance of light at 280 nm. A 10 mL gradient between 50% high salt buffer and 100% high salt buffer was then used to elute any remaining bound protein and DNA, before a further wash with 25 mL of high salt buffer.

2.1.19 BdIA haem co-purification through SEC

BdIA purified according to 2.1.17 was incubated with 10 x the protein concentration of hemin (dissolved in 1.2 mL DMSO) at 4°C overnight. Samples were centrifuged at 18,000 g for 15 minutes in a bench top micro-centrifuge before concentration to a volume of 1.2 mL. Sample was then loaded onto a Superdex 75 16/600 column to remove unbound hemin.

2.1.20 Spectroscopic analysis of BdIA¹⁻¹²⁷-NO interaction

Interactions between BdIA¹⁻¹²⁷ and NO were monitored using UV-visible light spectroscopy (UV-Vis) and Fourier transform infrared (FTIR) spectroscopy. For analysis with UV-Vis, 120 μ L of 25 mM NO donor (diethylammonium (Z)-1-(N,N-diethylamino)diazen-1-ium-1,2-diolate, DEA NONOate) in buffer E was added to 1.2 mL of 50 μ M BdIA¹⁻¹²⁷, purified according to 2.1.19 with concentration determined by absorbance of light at 280 nm. Every 15 minutes absorbance of

light between 300 and 700 nm was determined using an Agilent 8453 UV-visible spectroscopy system. Controls of buffer E, 1.2 mL BdIA¹⁻¹²⁷ with 120 μ L buffer E, and buffer E with 2.5 mM DEA NONOate were also monitored at equivalent time points to discount degradation of buffer or protein, or interaction between DEA NONOate and buffer from contributing to observed spectroscopic changes.

FTIR spectra were acquired between 800 and 4000 cm^{-1} using a Bruker Vertex 70 FTIR with a 100 μ m pathlength and a spectral resolution of 1 cm^{-1} , at the University of Strathclyde. BdIA¹⁻¹²⁷ at 7.11 mM (purified according to 2.1.19) was exchanged into a deuterated buffer E (i.e. 50 mM Tris, 300 mM NaCl, pH 9, dissolved in deuterium oxide and with pH adjusted using deuterium chloride), through repeated dilution and concentration. Spectra of the empty FTIR cell (with CaF_2 windows and PTFE spacer), 79 μ L deuterated buffer E and 79 μ L deuterated buffer E with 1% water were obtained to estimate the effects of water in the protein sample. Following collection of FTIR spectra without NO addition, 72 μ L of BdIA¹⁻¹²⁷ was added to 7 μ L of 20 mM NO donor (6-(-2-Hydroxy-1-methyl-2-nitrosohydrazine)-N-methyl-1-hexanamine, MAHMA NONOate) under anaerobic conditions before incubation for 45 minutes at room temperature. Sample was then transferred to an FTIR cell before spectra were collected every 10 minutes for 140 minutes.

2.1.21 Crystallisation of BdIA

Following purification according to 2.1.17, BdIA¹⁻⁴¹⁷ was concentrated to 442 μ M (as determined by absorbance of light at 280 nm, using a NanoDrop spectrophotometer), before centrifugation at 50,000 g for 30 minutes at 4°C in an Optima MAX-XP ultracentrifuge, TLA-55 rotor. 0.1 μ L sitting drops crystallisation trials were established using an Art Robbins Instruments Gryphon, with crystallisation screens JCSG+, Morpheus and PACT Premier (Molecular Dimensions).

Crystals were grown at 21°C in Morpheus crystallisation screen wells D6 (100 mM Morpheus buffer system 2 pH 7.5, 120 mM Morpheus alcohols and 50% of an ethylene glycol, PEG 8,000 mix), E6 (100 mM Morpheus buffer system 2 pH 7.5, 120 mM Morpheus ethylene glycols and 50% of an ethylene glycol, PEG 8,000 mix), B10 (100 mM Morpheus buffer system 3 pH 8.5, 90 mM Morpheus Halogens and 50% of an ethylene glycol, PEG 8,000 mix), C10 (100 mM Morpheus buffer system 3 pH 8.5, 90 mM Morpheus nitrate phosphate sulphate and 50% of an ethylene glycol, PEG 8,000 mix), D10 (100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus ethylene glycols and 50% of an ethylene glycol, PEG 8,000 mix), E10 (100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus alcohols and 50% of an ethylene glycol, PEG 8,000 mix), F10 (100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus monosaccharides and 50% of an ethylene glycol, PEG 8,000 mix), G10 (100 mM Morpheus buffer

system 3 pH 8.5, 100 mM Morpheus carboxylic acids and 50% of an ethylene glycol, PEG 8,000 mix) and H10 (100 mM Morpheus Buffer System 3 pH 8.5, 120 mM Morpheus amino acids and 50% of an ethylene glycol, PEG 8,000 mix). These crystal forming conditions were used in the design of the Morpheus Custom Optimisation Hanging Drop Screen, which is detailed in Appendix B. Vapour diffusion hanging drop plates were established using this screen with 1 μ L of concentrated protein added to 1 μ L of precipitant above a 500 μ L reservoir. Highest resolution data was collected on crystals grown in Morpheus Custom Optimisation Hanging Drop Screen well D6, with data collection statistics given in Table 4.1. Experimental phases were determined on crystals grown in Morpheus Custom Optimisation Hanging Drop Screen well D6, which were soaked in 16.5 mM thimerosal for 20 hours, before back soaking in 1 μ L of reservoir solution for 1 hour.

2.2 Macromolecular crystallography

Atomic structures of MucR^{EAL} and BdlA¹⁻⁴¹⁷ were modelled using data obtained from single crystal macromolecular crystallography.

2.2.1 The requirement for X-ray crystallography

All direct imaging techniques can only resolve two points independently if the distance between them is equal to or greater than half the wavelength of the light used³⁰². The atoms within a protein are separated by distances on the Ångstrom (10^{-10} m) scale, making it therefore necessary to use light within the X-ray range to distinguish individual atoms within a protein structure.

Due to challenges in making optics capable of precisely focusing X-rays, current resolution limit of X-ray microscopes is roughly 10 nm³⁰³, which is not sufficient to individually resolve the positions of atoms. It is therefore necessary to utilise X-ray diffraction, with the X-ray diffraction pattern providing information on the structure of the protein.

2.2.2 Crystal growth

X-ray diffraction is dependent on a sample consisting of regular repeating units, a criterion fulfilled through the formation of crystals. The growth of protein crystals is dependent upon the formation of an ordered three dimensional lattice as proteins aggregate out of solution. As it is impossible to predict the conditions under which this crystallisation process will occur exclusively from knowledge of the amino acid sequence, crystallisation experiments typically involve incubating purified protein within a range of different precipitant conditions³⁰⁴. In the vapour diffusion technique of crystallisation (used throughout this work), purified protein is mixed with a

precipitant buffer and held in a sealed environment with a larger volume of the precipitant buffer, referred to as a reservoir. As the reservoir contains the precipitant buffer at a higher concentration than that present within the protein sample it possess a lower vapour pressure, resulting in water being slowly drawn from the protein sample to the reservoir. As water is removed from the protein sample the concentration of the protein is increased. Once the concentration of the protein has increased above the intrinsic solubility, the protein will either precipitate or begin to crystallise³⁰⁵. By performing this crystallisation experiment with a wide variety of precipitant buffers, one would hope to identify a condition within which the nucleation of crystals is favoured above precipitation^{305,306}. Once the nucleus of a crystal is formed, the crystal can grow if further protein molecules from solution bind³⁰⁴. This can be represented graphically using the phase diagram, shown in Figure 2.3. Although it is not possible to predict conditions under which a protein can crystallise, crystallisation of a protein (assuming identical sample preparation) should be entirely reproducible. Therefore conditions under which growth of small or poorly diffracting crystals is observed could be optimised (for example by slightly reducing the precipitant concentration to reduce the number of nucleation events) to obtain larger or more ordered crystals, capable of providing more information on the protein structure within a diffraction experiment^{306,307}.

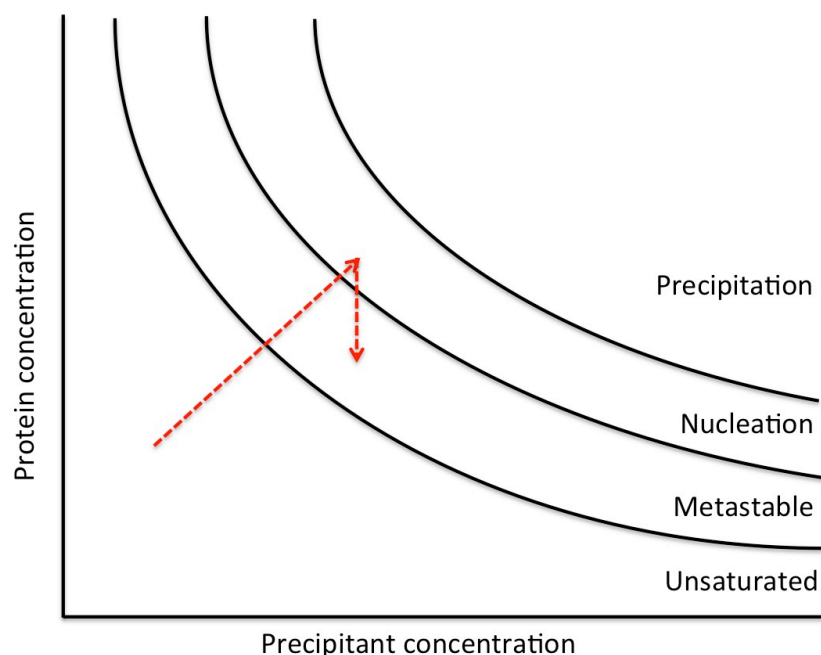


Figure 2.3 - The phase diagram. At high concentrations of protein and precipitant, protein precipitates out of solution. However at lower concentrations of protein or precipitant the protein can instead nucleate to form the core of crystals. Further protein molecules can then add to the crystal nuclei to grow the crystal. The path of crystallisation through the phase diagram during a vapour diffusion experiment is shown with red arrows.

2.2.3 X-ray diffraction

Following growth, crystals can be isolated by mounting within a loop and are cryo-cooled to 100 K³⁰⁸. The next stage of the X-ray diffraction experiment is to subject the crystal to X-ray radiation. All diffraction data discussed within this work was collected at synchrotron radiation sources. Within a synchrotron, electrons produced using an electron gun (which is similar to a cathode ray tube) are accelerated to close to the speed of light, before circulating in a low-pressure storage ring^{309,310}. Despite appearances, synchrotrons are not circular and are instead polygons composed of numerous straight sections^{309,311}. As the accelerated electrons are pushed around the corners of the polygon by bending magnets, they emit energy in the form of X-ray radiation^{310,311}. Classically, it is this X-ray radiation that is then used for the diffraction experiment. Modern synchrotrons also emit X-rays through the action of undulating magnets, also known as insertion devices. Typically, undulators consist of two rows of magnets that sit either side of the electron path. As the electron passes through an undulator, the multiple magnets force the electron to follow an undulating trajectory. The radiation emitted from each bend interferes with that emitted from the other bends, with this interference giving rise to a more focussed radiation

beam than that generated by a single bending magnet. Another significant advantage over the bending magnet is that the gap between the two rows of magnets within the undulator can be adjusted to tune the wavelength of the emitted X-ray, whereas bending magnets emit a continuous spectrum of radiation not only limited to X-rays but also including ultraviolet, visible, infrared and microwave radiation^{310,311}. The beam of produced X-rays is then narrowed by collimation before further reduction in beam diameter using focussing mirrors.

The protein crystal is then exposed to the focussed X-rays, resulting in interaction between the X-rays and electrons within the protein result in the scattering of X-rays. X-ray diffraction from a protein crystal obeys Bragg's law, in that the diffracted beams can be computed by treating diffraction as if from a series of equivalent parallel planes^{312,313}.

The protein crystal can be considered to be composed of a series of identical tessellating boxes that generate the three-dimensional lattice of a crystal through translational symmetry. Such a box is referred to as the unit cell^{313,314}. The unit cell can be described by the length of three unique edges, **a** (the length along the x-axis), **b** (the length along the y-axis) and **c** (the length along the z-axis), as well as three unique angles, α , β and γ . Each face of the unit cell can be thought of as a plane parallel to equivalent planes on other unit cells throughout the crystal. Further planes may also exist that intersect edges of the unit cell. Each of these planes is assigned a three number Miller index (*hkl*) or lattice index, with each number of the Miller index corresponding to the number of parallel planes that exist within the given axis of a unit cell, or put another way the number of sections they divide the unit cell into³¹⁴. For example the planes defined by the Miller index (2,3,4) will divide the unit cell edge **a** in two, **b** into three and **c** into four, as shown in Figure 2.4A. Another example would be the planes defined by Miller index (1,0,0), which divides the unit cell edge **a** once per unit cell and does not divide the **b** and **c** edges, so corresponds to the face between **b** and **c**, shown in Figure 2.4B. Each of these sets of parallel planes, known as Miller planes, can act as a source of diffraction, producing a single reflection³¹⁴.

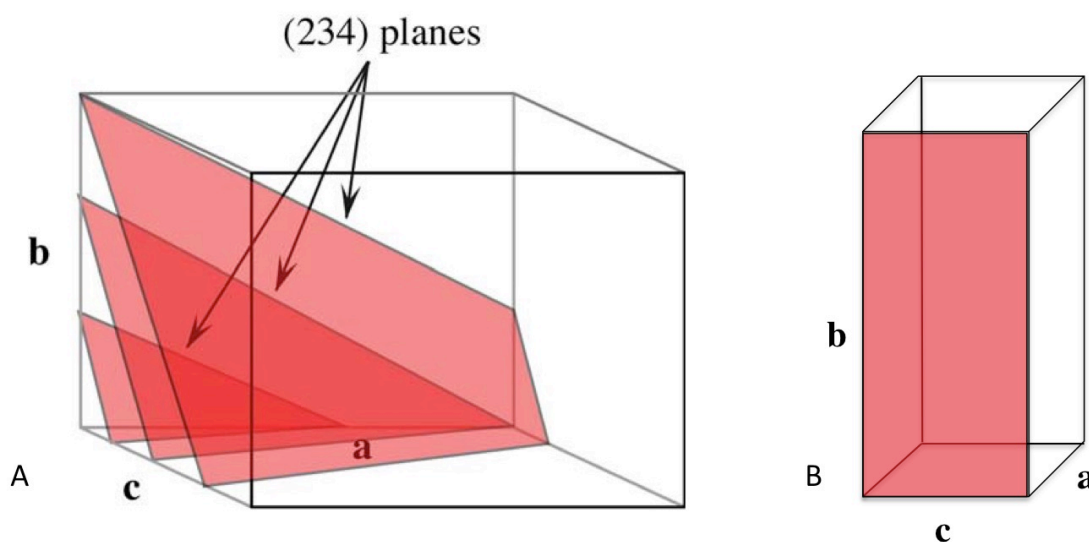


Figure 2.4 - Miller planes. A) The (2,3,4) Miller planes, shown in red, divide the **a**, **b** and **c** axes into 2, 3 and 4 respectively. Taken from Rhodes et al. 2006³¹⁴. B) The (1,0,0) plane, shown in red, divides the **a** axis once, without dividing the **b** or **c** axes and so corresponds to the **bc** plane.

Bragg's law, given by equation 1 and shown in Figure 2.5, states the spatial requirement needed to generate constructive interference between diffracted X-rays, with X-rays experiencing destructive interference destroyed before the intensity of the diffracted X-ray beam can be measured³¹².

$$n\lambda = 2d\sin\theta$$

equation 1

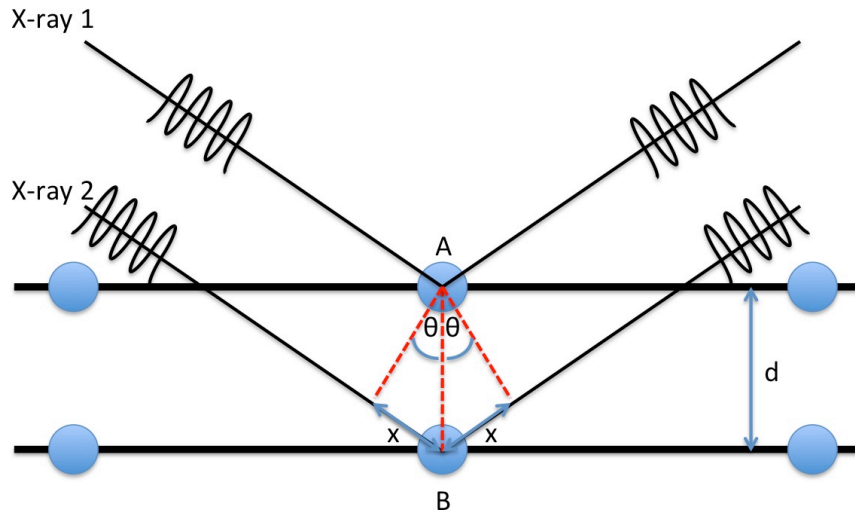


Figure 2.5 - Graphical representation of Bragg's law. Where n represents an integer, λ is the wavelength of the X-rays, d is the distance between planes and θ is the angle made between the incident X-ray beam and the planes. Both X-ray 1 and X-ray 2 are in phase prior to diffraction from the crystal. X-ray 1 is diffracted at point A, whilst X-ray 2 is diffracted from point B. X-ray 2 has to travel $2x$ further than X-ray 1, with x equal to $d\sin\theta$. For diffracted X-ray 2 to travel this extra distance and still be in phase with X-ray 1, this distance must be equal to an integer of the X-ray wavelength ($n\lambda$). In all instances when θ is changed so that $2d\sin\theta$ does not equal $n\lambda$, the two diffracted X-rays are now out of phase and are cancelled out through destructive interference.

Bragg's law can be rearranged in the form shown in equation 2. Arranged in this way, Bragg's law demonstrates that d is inversely proportional to $\sin\theta$. From this it can be inferred that when the distance between diffracting planes is increased the observed diffraction pattern will be compressed.

$$\sin\theta = \left(\frac{n\lambda}{2}\right)\left(\frac{1}{d}\right) \quad \text{equation 2}$$

From equation 2 it can also be observed that $\sin\theta$ is directly proportional to $1/d$. The index of the observed reflections hkl , can therefore be considered to be equal to $1/\text{distance between } (hkl) \text{ planes}$. Due to the nature of this relationship, it can be useful to introduce the concept of reciprocal space. In reciprocal space, each real space diffracting plane is represented by a vector perpendicular to the plane, with a length equal to the real space distance between the planes d . If all of these vectors are plotted about an origin and a marker denoted at the end of each vector then a lattice is generated, known as the reciprocal lattice. Each point on the reciprocal lattice

thus represents a plane that diffracts in real space. As the distance between points hkl in the reciprocal lattice is equal to the reciprocal of the distance between (hkl) planes, Bragg's law can be rearranged as equation 3.

$$d_{hkl} = \frac{1}{d_{(hkl)}} = \frac{2\sin\theta}{n\lambda} \quad \text{equation 3}$$

Therefore, reflections are only observed when the reciprocal lattice vector is equal to $2\sin\theta$ divided by $n\lambda$. This is often considered through the construction of sphere of reflection, or the Ewald sphere³¹⁵, shown in Figure 2.6. From the construction of the Ewald sphere it can be observed that few reciprocal points contribute to diffraction for each orientation of the crystal, making it necessary to rotate the crystal to observe diffraction from each set of Miller planes.

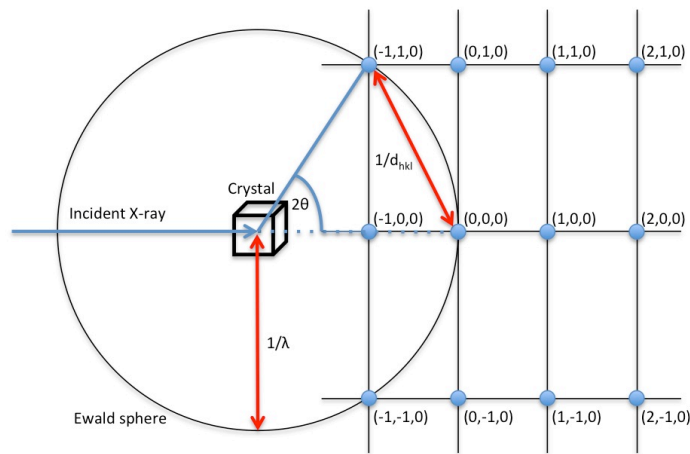


Figure 2.6 - The Ewald sphere. The incident X-ray is diffracted by the crystal (located at the origin of the sphere), with diffraction only observed when Bragg's law is satisfied. The origin of the reciprocal lattice lies on the Ewald sphere, directly in-line with the path of the incident X-ray. Through the construct of a sphere (here represented as circle) of radius $1/\lambda$, conditions that satisfy Bragg's law (i.e. when $2\sin\theta$ is equal to the reciprocal of the distance between the observed Miller indices) can be easily identified. As very few points lie on the Ewald sphere, in this two dimensional example only $(-1,1,0)$ and $(-1,-1,0)$, it is necessary to rotate the crystal to obtain diffraction from other Miller planes.

2.2.4 Structure factors

Each diffracted X-ray can be described mathematically by the complex structure factor (\mathbf{F}_{hkl}), which describes three physical parameters of the X-ray wave: direction, phase and amplitude. The direction can be determined by calculating the scattering angle, which is dependant upon the distances between diffracting planes, described by the Miller indices (hkl) , and the orientation of

the crystal. The amplitude of the X-ray is proportional to the square root of the intensity measured in a diffraction experiment, with the intensity of each reflection dependent upon the density and distribution of electrons along the diffracting planes, contributed by all of the atoms within the unit cell. Mathematically, the amplitude of the X-ray is given by the structure factor amplitude (F_{hkl})³¹⁴, calculated using equation 4. The phase of the X-ray, ϕ , cannot be determined directly from the X-ray diffraction experiment. Two methods have been used to determine the initial phases of X-rays during the course of this work, molecular replacement and single-wavelength anomalous dispersion (SAD). Initial phase estimates generated using these methods were improved through refinement, discussed below.

$$\mathbf{F}_{\mathbf{hkl}} = F_{hkl} e^{2\pi i \phi_{hkl}} \quad \text{equation 4}$$

The molecular replacement strategy of phase determination is dependant upon knowing the structure of a homologous protein. Typically, protein homology is determined on the basis of amino acid sequence, however it will vary between individual cases as to what level of sequence homology is sufficient to find a structurally homologous model, as in some cases even 100% sequence homology does not confer sufficient structural homology. By squaring the structure factor amplitudes and performing a Fourier transform a Patterson map of inter-atomic vectors can be constructed. An equivalent Patterson map can also be generated from the known homologous structure and the two Patterson maps compared. Similar positioning and orientation within the unit cell between unknown structure and homologous model would result in similar Patterson maps. The homologous model structure is rotated and translated, with model Patterson maps from each alteration compared to those derived from the unknown structure until an optimal solution is found^{316,317}.

The SAD strategy of phase determination requires the presence of heavy atoms, which may be achieved by the soaking of crystals with at least one heavy atom derivative. In the case of BdlA¹⁻⁴¹⁷, crystals were soaked with the Hg derivative thimerosal, allowing Hg²⁺ ions to bind to cysteine residues on the surface of the protein. As well as scattering X-rays, elements absorb X-rays, with this absorption sharply decreasing at element specific wavelengths (known as absorption edges). Absorption edges of light elements within the unit cell are far removed from the X-ray wavelengths typically used in X-ray crystallography, however this is not true of heavy atoms. When the X-ray wavelength is close to the absorption edge of a heavy metal, a fraction of the X-rays are absorbed by the heavy atom and reemitted with a different phase³¹⁸. Absorption of X-rays in this way results in the breakdown of an equivalence between intensities of Friedel pairs (h,k,l and $-h,-k,-l$, which according to Friedel's law would otherwise be equal), allowing the

positions of absorbing heavy atoms within the unit cell to be determined through Patterson methods and initial estimates of phase for each structure can be derived^{318–321}.

Using the Fourier transform given in equation 5, it is possible to calculate the electron density of the unit cell contents from the determined scattering angle, phase and amplitude. Determining the density of the electrons (p) at Cartesian coordinates x, y, z requires multiplying the reciprocal volume of the unit cell (V) by the sum of all complex structure factors (F_{hkl}). As the complex structure factor of (F_{hkl}) for each X-ray is distinct, it is necessary to observe the diffraction of all Miller planes. Throughout this work, all lattice points were observed by rotating the crystal in the path of the X-ray, capturing small wedges of data (typically less than 1°) from the diffracting crystal, which were then combined to describe all possible orientations³¹⁴.

$$p(x, y, z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| e^{-2\pi i \phi_{hkl}(hx+ky+lz)} \quad \text{equation 5}$$

2.2.5 Refinement

Equation 5 describes how it possible to determine an electron density map based upon amplitudes observed in a diffraction experiment and estimated phases. As this map is generated on the basis of observed data it is referred to as the Fo map. It is possible to calculate an alternative map based entirely on the atomic model, known as the Fc map. The Fc map is calculated by determining the contribution of each atom in the unit cell to the observed diffraction from each set of Miller planes, thus making it possible to calculate the diffraction observed from the whole model. The complex structure factor (**Fc**), calculated in equation 6, states how each set of Miller planes in the model unit cell contributes to diffraction.

$$\mathbf{Fc} = G \cdot \sum_j n_j f_j e^{2\pi i (hx_j + ky_j + lz_j)} \cdot e^{\frac{-B_j \sin^2 \theta}{\lambda^2}} \quad \text{equation 6}$$

Equation 6 is a Fourier sum, with one term contributed by each atom (j) in the model. G is a scale factor to shift all Fcs onto a convenient numerical scale. The occupancy of atom j within the unit cell is given by n_j , which states the likelihood a specific atom is observed in the described location and is equal to 1 if all molecules in the crystal are perfectly identical. The structure factor amplitude of atom j is given by f_j , whilst x_j, y_j and z_j define the location of the atom. The final variable included is the temperature factor, B_j , which describes the range of motion of the atom around the defined position and can explain smearing of electron density.

The quality of the model in relation to the observed data can be examined by fitting the model into an electron difference density map, calculated by subtracting Fc from Fo (known as an Fo-Fc map). Improperly placed atoms are surrounded by negative electron density in Fo-Fc maps, and

similarly positive electron density in Fo-Fc maps is indicative of a region requiring further model building and placement of more atoms. As Fo-Fc maps display only misplaced or missing atoms in the current model it lacks the appearance typical of many proteins. It is therefore more common for models to be built into a 2Fo-Fc map. In a 2Fo-Fc map the map component generated by observed electron densities is greater than the subtracted equivalent calculated from the model, resulting in an observable electron density surface that has a greater resemblance to the model placed within but is not as heavily influenced by the model as an Fo map. During the course of this work models were built using weighted variants of Fo-Fc (*m*Fo-*D*Fc) and 2Fo-Fc (*2m*Fo-*D*Fc) maps, known as sigma-A weighted maps³²². Within sigma-A maps the Fo component is weighted according to an approximate measure of the phase quality (*m*, which is also known as the figure of merit, and is calculated for each reflection). The Fc component in sigma-A weighted maps is weighted according to the sigma-A weighting factor (*D*), which is an estimate of the accuracy within the current model, in combination with the model completeness³²². Following model building, Fc and Fo can then be recalculated through the process of refinement.

For a protein structure to be considered as fully modelled it must both explain the experimental observations of electron density and conform to the geometry and chemistry expected of atoms within a protein, which will limit parameters such as available bond lengths and angles. Whilst deviations from ideal bond lengths and angles are given explicitly as root mean square deviations (RMSD), the quality of fit between the model and experimentally determined data is given through calculation of the R_{work} , using equation 7. An equally important metric to consider during refinement is that of the R_{free} ³²³. Just like R_{work} , R_{free} is calculated using equation 7, but using a subset of amplitudes, typically 5%, which have been omitted from refinement. Rounds of model building and refinement are iteratively repeated until the protein displays realistic chemistry and it is no longer possible to decrease the value of R_{free} , at which point model building can be considered complete.

$$R_{work} = \frac{\sum ||F_{observed}| - |F_{calculated}||}{\sum |F_{observed}|} \quad \text{equation 7}$$

Chapter 3: Structural and functional studies of the MHYT domain proteins MucR and NbdA

Pseudomonas aeruginosa PAO1 requires the proteins MucR and NbdA for nitric oxide induced biofilm dispersal¹⁵⁵. Both MucR and NbdA contain an MHYT domain, a transmembranous arrangement of 7 helices thought to act as a redox sensor or to directly bind diatomic gases, as well as DGCs and PDEs^{153,155,204}. We hypothesised that the MHYT domain may serve to bind NO and regulate the activity of the DGC or PDE domains to facilitate biofilm dispersal. To investigate this hypothesis we sought to establish whether MucR or NbdA are amenable to structural and biochemical characterisation.

3.1 Purification and analysis of MYHT domains

3.1.1 Gene fragment design and cloning at OPPF-UK

MHYT domains are predicted to consists of 7 transmembrane helices, and are thought to act as sensors of redox state^{153,204}. As membrane proteins are notoriously difficult to study through X-ray crystallography^{324,325}, and MHYT domains have not previously been structurally characterised, a high-throughput approach to cloning and expression was utilised. Use of a high-throughput approach allows a variety of MucR and NbdA (the MHYT domain-containing proteins of PAO1) protein fragments with different N- and C-termini to be expressed on a small scale in parallel and fragments demonstrating expression of correctly folded protein can be further investigated.

Fragments were designed to encode for the MHYT domain in isolation, as well as fragments that included the DGC and PDE domains. As proteins containing 7 transmembrane (TM) helices are infrequent in bacteria in comparison to 6 and 8 TM helix proteins³²⁶, we hypothesised that the seventh helix of the MHYT domain may not run across the membrane, so fragments were also designed that didn't feature this helix. Within the confines of creating fragments to feature these different domains and secondary structure elements 24 variants each of MucR and NbdA were designed, with varying N- and C-termini, as depicted in Figure 3.1A and 3.1B.

Designed fragments were amplified from the PAO1 genome through PCR and were subsequently cloned into pOPINE-3C-GFP vectors using InFusion cloning, as detailed in section 2.1.1. Following confirmation of successful cloning through colony PCR, these plasmids were transformed into *E. coli* cell lines C41(DE3), C43(DE3), Lemo21(DE3) and KRX, and were expressed on a small scale. Harvested cells were lysed by rapid freezing and membranes were resolubilised in 1% DDM.

Chapter 3

Resolubilised samples were resolved through SDS-PAGE and examined for GFP fluorescence in-gel. The presence of a C-terminal GFP-tag on all fragments allows the presence of expressed, correctly folded protein to be observed, as misfolded proteins won't feature folded GFP and thus will not fluoresce. Following resolution by SDS-PAGE, comparison of a GFP control to the C-terminal GFP in each fragment allows qualitative assessment of expression levels and protein integrity (as less stable proteins may lose their GFP during expression and these initial purification steps) without the requirement of further purification steps³²⁷. Representative results of in-gel fluorescence for MucR and NbdA fragments are shown in Figure 3.1C and 3.1D, respectively. Multiple fluorescent bands can be observed in several lanes, which are likely to be indicative of protein degradation.

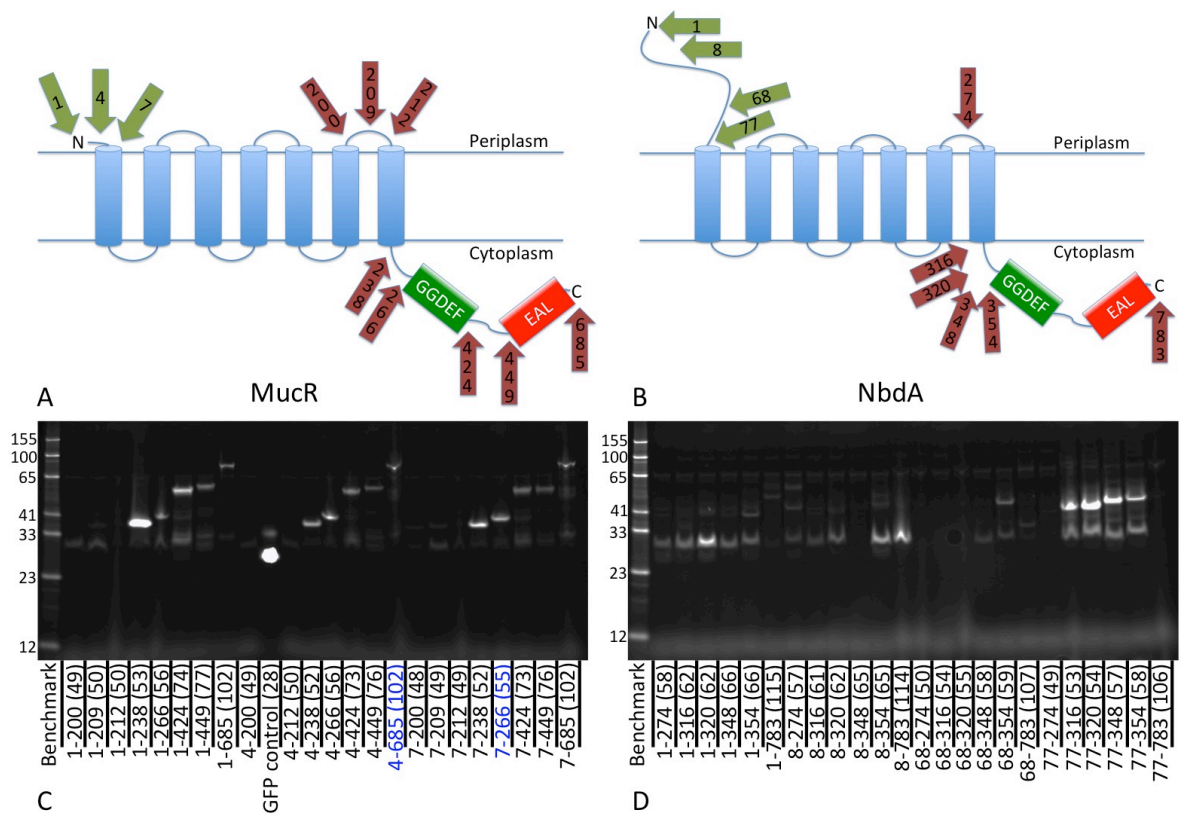


Figure 3.1 - High throughput fragment design and expression of MucR and NbdA. A) Overview of MucR fragments designed. B) Overview of NbdA fragments designed.

Transmembrane helices of the MHYT domain are shown as blue cylinders, DGC and PDE domains are shown as green and red boxes respectively, green arrows denote N-terminal residues and red arrows denote C-terminal residues. C) High-throughput expression of MucR fragments, resolved by SDS-PAGE and imaged through in-gel fluorescence. D) High-throughput expression of NbdA fragments, resolved by SDS-PAGE and imaged through in-gel fluorescence. Protein fragment termini are denoted for each band, with predicted molecular mass (in kDa) given in parenthesis. The molecular mass of each fluorescent protein standard is given in kDa to the left of the associated band. Fragments selected for further study are labelled in blue in panel C.

3.1.2 Fluorescence size-exclusion chromatography

Two fragments of MucR (and accompanying expression cell lines) were selected for further examination on the basis of small-scale expression, MucR⁷⁻²⁶⁶, which encodes for the MHYT domain but neither of the cytosolic domains and displays a high level of fluorescence for the desired protein fragment without fluorescent contaminant, and MucR⁴⁻⁶⁸⁵, which encodes for all of the domains of MucR, so is likely to behave differently in crystallisation trials and may be amenable to functional characterisation. As all fragments were cloned with a C-terminal GFP tag, fluorescence size-exclusion chromatography (FSEC) can be used to monitor separation of free-GFP, fluorescent aggregate and fluorescent resolubilised protein. As FSEC separates out these fluorescent species, protein monodispersity and stability can rapidly be assessed following resolubilisation in a range of different detergents without the requirement of purification³²⁸. To this end, both fragments were analysed with FSEC following resolubilisation in 6 different detergents that have previously been used for successful crystallisation at MPL: 6-cyclohexyl-β-D-maltoside (CYMAL6); n-decyl-β-D-maltoside (DM); n-dodecyl β-D-maltoside (DDM); fos-choline-12 (FC12); lauryldimethylamine *N*-oxide (LDAO) and octyl glucoside (OG).

From Figure 3.2 it can be seen that MucR⁷⁻²⁶⁶ displays low proportions of aggregate and free-GFP when resolubilised with CYMAL6, DDM, DM or FC12. FC12 is a harsh detergent well-known for offering poor protein stability despite efficient extraction from the membrane^{324,329,330}, so was not selected for further protein purification. Proportionally lower aggregate peaks are observed in the presence of DDM than samples resolubilised in CYMAL6 or DM, therefore DDM was selected as detergent for resolubilisation and further purification. It is worth noting the shift in elution time of the protein peak, which is likely to be due to differences in the size of micelles formed by the different detergents.

From Figure 3.3 it can be seen that MucR⁴⁻⁶⁸⁵ shows a low percentage of soluble GFP-tagged protein when resolubilised with CYMAL6, DM, DDM, LDAO and OG. When analysed with FSEC, GFP-tagged MucR⁴⁻⁶⁸⁵ (shown in Figure 3.3B) resolubilised in DDM does not present the expected Gaussian distribution observed of the equivalent peak in other detergents (Figure 3.3A and 3.3C), suggesting MucR⁴⁻⁶⁸⁵ is polydisperse or prone to aggregation within this mixture^{328,331}. No subsequent purification and characterisation of MucR⁴⁻⁶⁸⁵ was performed, as the only other examined detergent, FC12, is unsuitable for further purification. Whilst a broader range of detergents could have been examined to identify any that may have resolubilised the MucR⁴⁻⁶⁸⁵ fragment, it may be more efficient to investigate a different fragment from high-throughput screening that contains the same domains with FSEC.

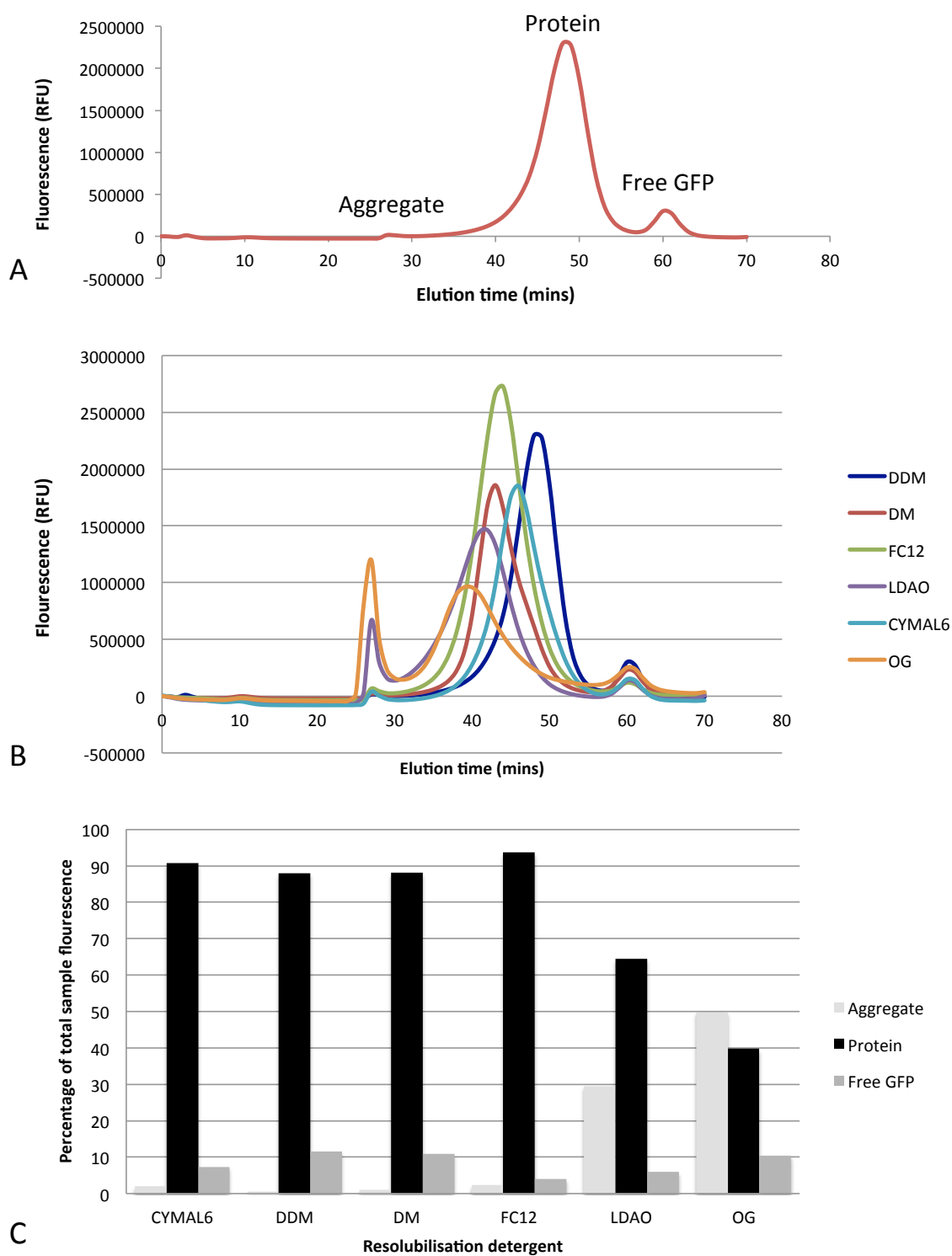


Figure 3.2 - FSEC analysis of MucR⁷⁻²⁶⁶. A) FSEC profile of MucR⁷⁻²⁶⁶ resolubilised in DDM. Peaks corresponding to fluorescent protein aggregate, fluorescent protein and free GFP are all labelled. B) FSEC profiles of MucR⁷⁻²⁶⁶ when resolubilised in OG, CYMAL6, DDM, DM and LDAO. C) Proportions of aggregate, protein and free GFP present in each FSEC sample, as determined by peak amplitude.

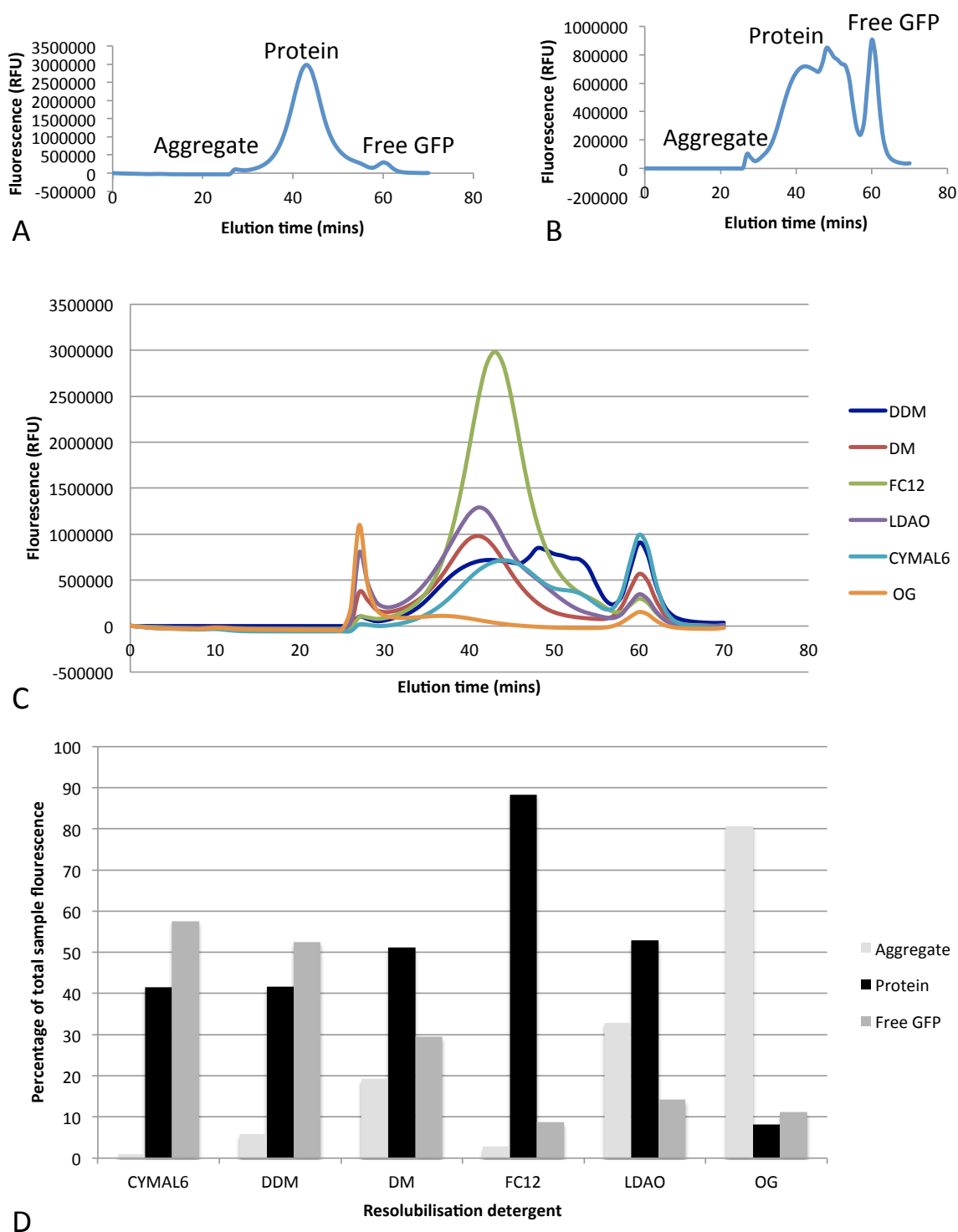


Figure 3.3 - FSEC analysis of MucR⁴⁻⁶⁸⁵. A) FSEC profile of MucR⁴⁻⁶⁸⁵ resolubilised in FC12, with peaks corresponding to fluorescent protein aggregate, fluorescent protein and free GFP labelled. B) FSEC profile of MucR⁴⁻⁶⁸⁵ resolubilised in DDM, showing a polydisperse protein peak. C) FSEC profiles of MucR⁴⁻⁶⁸⁵ when resolubilised in OG, CYMAL6, DDM, DM and LDAO. D) Proportions of aggregate, protein and free GFP present in each FSEC sample, as determined by peak amplitude.

3.1.3 Protein purification

MucR⁷⁻²⁶⁶ was expressed in *E. coli* KRX cells grown in 6 L of Power Broth. As MHYT domains are thought to bind copper, all expression and purification buffers were supplemented with 10 μ M copper chloride. Expressed protein was resolubilised in 1% DDM before purification by immobilised metal-affinity chromatography (IMAC) as detailed in 2.1.6. The C-terminal 6xHis-GFP was cleaved from MucR⁷⁻²⁶⁶ using His-tagged 3C protease overnight. During dialysis overnight a significant portion of protein solution appeared to precipitate, so precipitant was removed through centrifugation. MucR⁷⁻²⁶⁶ was further purified, and separated from the His-GFP tag and 3C protease, through reverse phase metal affinity chromatography (reverse-IMAC) and size exclusion chromatography (SEC), as detailed in 2.1.6. Protein purity was assessed following purification using SDS-PAGE as imaged with in-gel fluorescence and with coomassie staining, shown in Figure 3.4. Following concentration, only 0.158 mg of purified MucR⁷⁻²⁶⁶ remained, which was not sufficient for crystallisation.

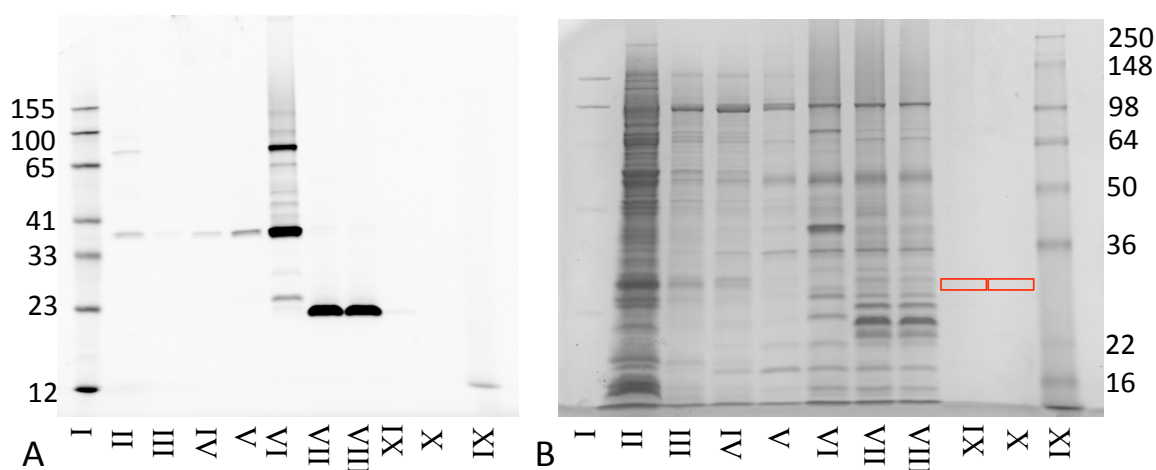


Figure 3.4 - Purification of MucR⁷⁻²⁶⁶. Protein purity was assessed through in-gel fluorescence (A) and coomassie staining (B). Lanes correspond to: benchmark fluorescent protein standard (I); IMAC flow-through, 10 mM imidazole wash, 20 mM imidazole wash, 30 mM imidazole wash and eluate (II to VI); sample after proteolysis and dialysis before and after centrifugation (VII and VIII); reverse-IMAC flow-through (IX), dialysis buffer (X) and SeeBlue Plus 2 protein ladder (XI). Atomic weights of benchmark fluorescent protein standard and SeeBlue Plus 2 protein ladder are listed in kDa on the left and right-hand boundaries of A and B respectively, and the positions of expected bands for MucR⁷⁻²⁶⁶ in lanes XI and X are denoted with red boxes.

3.1.4 Optimisation of expression conditions

As insufficient protein was purified for crystallisation, conditions for expression were optimised. Using 3 mL growths, 96 expression conditions were screened in parallel using 10 mL 24-well blocks with an air-permeable seal. Expression was screened across 2 different *E. coli* cell lines (KRX and Lemo21(DE3)), which were grown as single overnight starter cultures (i.e. one starter culture of KRX cells was used to inoculate the appropriate 48 expression growths, with an equivalent culture used for Lemo21(DE3) cells). Both strains were screened in 2 different types of media (Power Broth and Terrific Broth), with expression allowed to proceed at 2 different temperatures overnight (18°C and 25°C after initial growth at 37°C). Within KRX and Lemo21(DE3) cells, protein expression is not only controlled by the *lac* repressor, but is also finely regulated by the addition of L-rhamnose. Within KRX cells, expression of T7 RNA polymerase is directly regulated by the rhamnose binding promoter rhaBAD, however in Lemo21(DE3) cells addition of rhamnose induces the expression of lysozyme, a natural inhibitor of T7 RNA polymerase. As T7 RNA polymerase is required for expression of encoded gene fragments, different concentrations of L-rhamnose were screened (0 mM, 0.5 mM, 1 mM and 2 mM added at the start of growth of Lemo21(DE3) cells, whilst 1 mM, 2 mM, 4 mM and 8 mM were added at the point of induction in KRX cells) and concentration of IPTG used for induction was also investigated (1mM, 2mM and 4mM).

After harvesting, cells were resuspended in 100 μ L of phosphate buffered saline (PBS) and the total level of fluorescence measured. As cell fluorescence is representative of expression and is thus likely to correlate to final protein yield, fluorescence of each expression condition was compared to that of KRX cells grown in Power Broth at 25°C and induced with 1 mM IPTG, 2mM L-rhamnose, the expression conditions previously used for protein purification, shown in Figures 3.5 and 3.6. The largest increase in fluorescence over previous expression conditions (54%) was observed for MucR⁷⁻²⁶⁶ grown in Lemo21(DE3) cells, at 25°C in Terrific Broth, in the presence of no L-rhamnose and induced with 2 mM IPTG, so these conditions were selected for large scale protein expression.

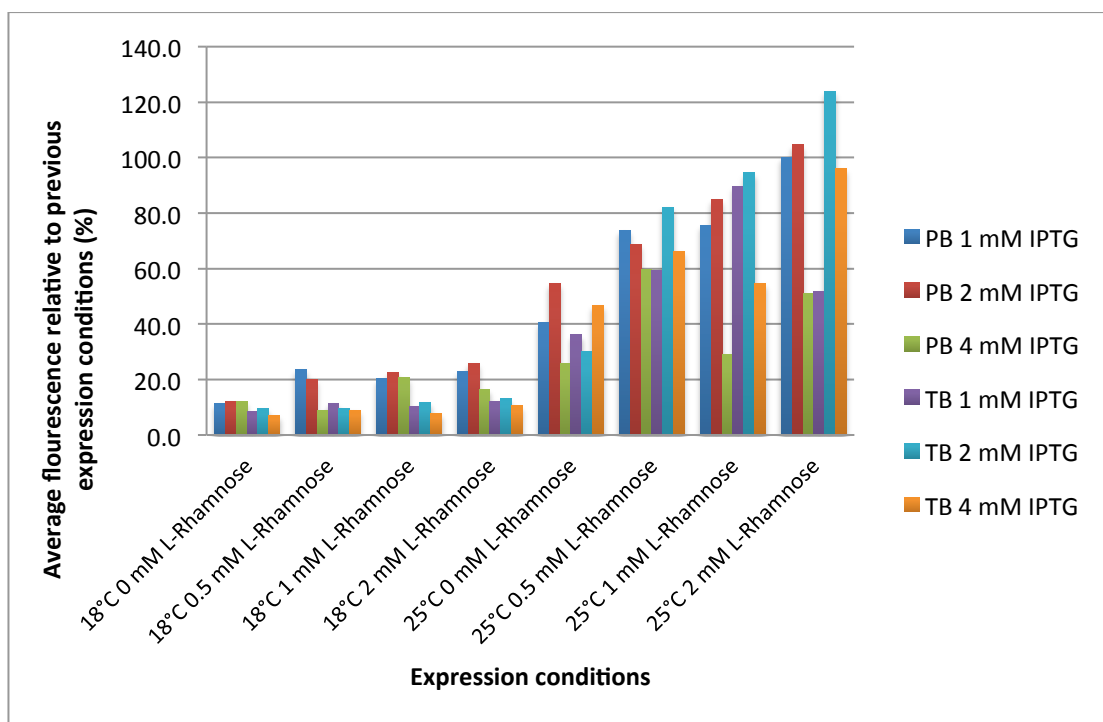


Figure 3.5 - Relative fluorescence of MucR⁷⁻²⁶⁶ expression in KRX cells.

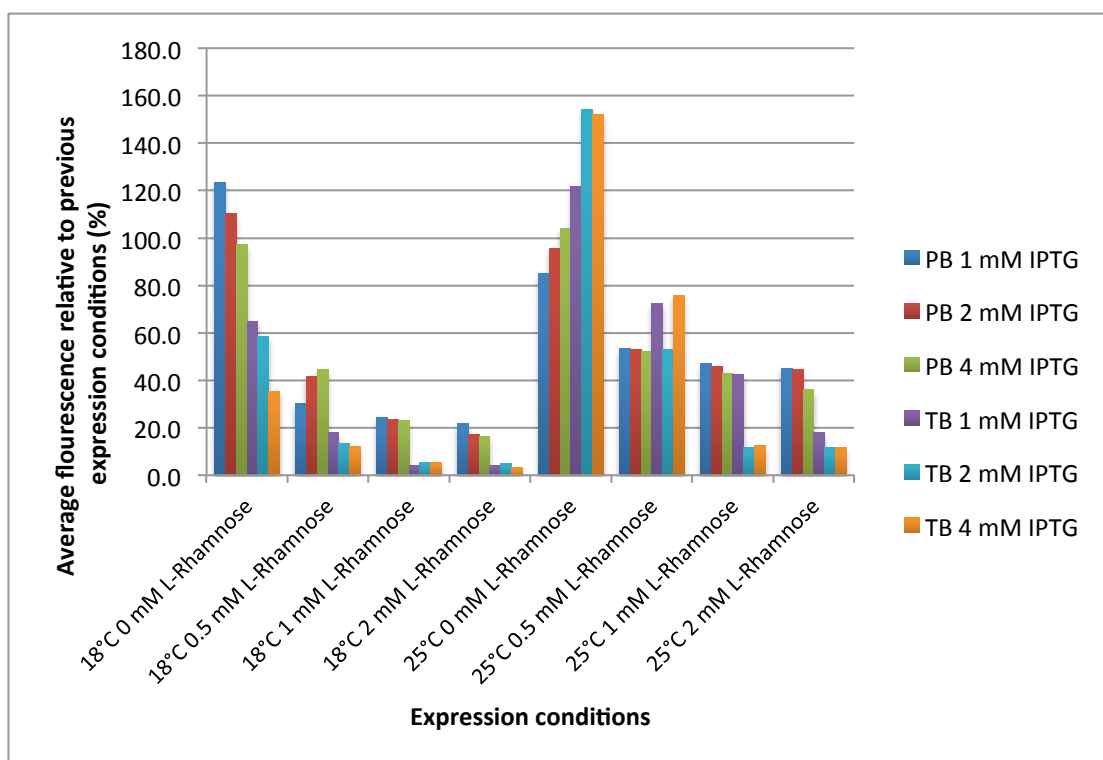


Figure 3.6 - Relative fluorescence of MucR⁷⁻²⁶⁶ expression in Lemo21(DE3) cells.

3.1.5 Optimised protein purification and SEC-MALS

As well as improving levels of expression, efficiency of each step of protein purification was also increased to improve the final protein yield of MucR⁷⁻²⁶⁶. Until the point of GFP-cleavage, the loss

of protein at each stage of purification can be estimated from the remaining fluorescence, as shown in Figure 3.7. To improve protein retention, changes were made to resolubilisation and IMAC protocols. Specifically, efficiency of resolubilisation was increased by doubling the volume of solubilisation buffer and IMAC efficiency was improved by extending the period of nickel-protein incubation from 2 hours to 4 hours. By optimising expression and purification, final protein yield was increased approximately 4-fold from 0.158 mg to 0.632 mg.

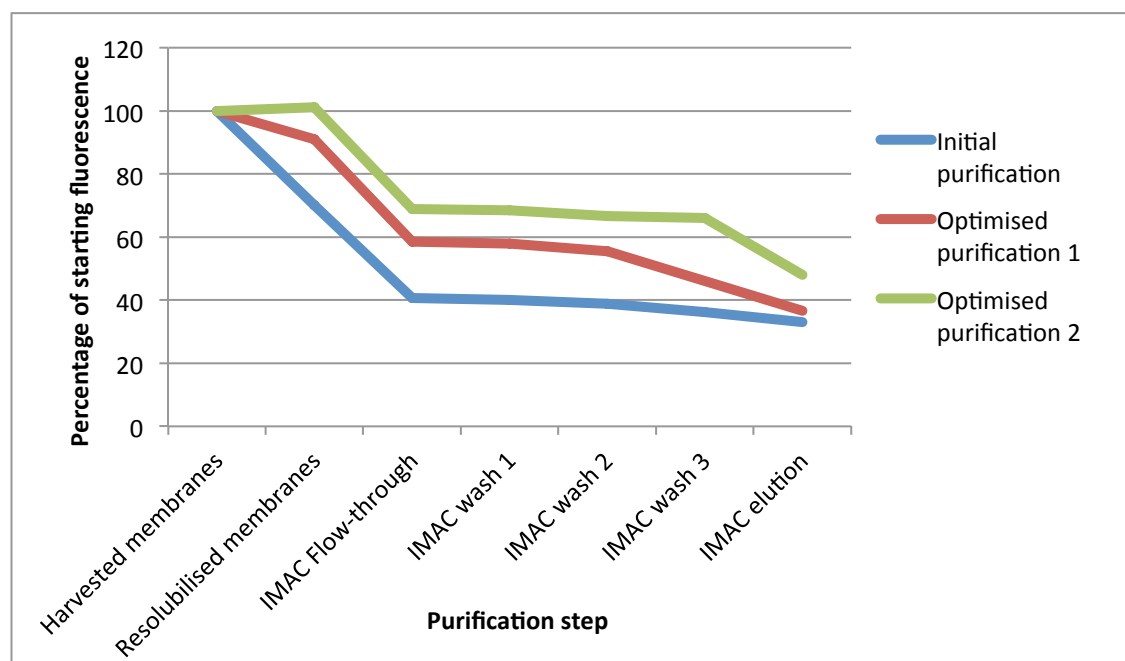


Figure 3.7 - Improvement in protein yield through optimised purification. At each stage of purification fluorescence of 100 μ L of MucR⁷⁻²⁶⁶ sample was determined, as detailed in 2.1.16. As sample volume changes over the course of purification, fluorescence measurements were normalised by volume, with normalised fluorescence plotted relative to the fluorescence of harvested cell membranes. IMAC washes 1, 2 and 3 correspond to 10 mM, 20 mM and 30 mM imidazole.

Purification of MucR⁷⁻²⁶⁶ by SEC (as shown in purple in Figure 3.8), demonstrates that the protein is highly polydisperse, possibly existing in a number of oligomerisation states or undergoing degradation^{328,331}. As all protein within a crystal should be structurally identical, crystallisation is likely to be impaired by MucR⁷⁻²⁶⁶ occupying multiple oligomerisation states. To determine which peak of the SEC profile corresponds to monomeric protein, and thus identify how large a proportion of the protein exists as a monomer, MucR⁷⁻²⁶⁶ was analysed by multi-angled light scattering coupled to size-exclusion chromatography (SEC-MALS). SEC-MALS identifies a monodisperse peak corresponding to resolubilised monomeric protein (labelled monomer is 3.8A and further analysed in 3.8B). This peak corresponds not only to monomeric protein, but also to DDM micelles, as evidenced by the large increase in refractive index and right angle scattering

relative to UV absorbance at this position. As the 280 nm UV absorbance peak is only 31.66% the height of the refractive index and right angle scattering peaks (labelled Wt Fr (A) in Figure 3.8), protein only accounts for 31.66% of the observed molecular mass (79,179 Da, listed as Mw in Figure 3.8), corresponding to 25,068 Da. This calculated mass is equal to 91% of the theoretical molecular weight of MucR⁷⁻²⁶⁶ (27,431 Da), suggesting that the analysed peak corresponds to monomeric protein.

However, a large proportion of the purified MucR⁷⁻²⁶⁶ exists as higher oligomers, and further analysis of protein could be hindered by equivalency in size between the resolubilised monomeric protein and DDM micelles.

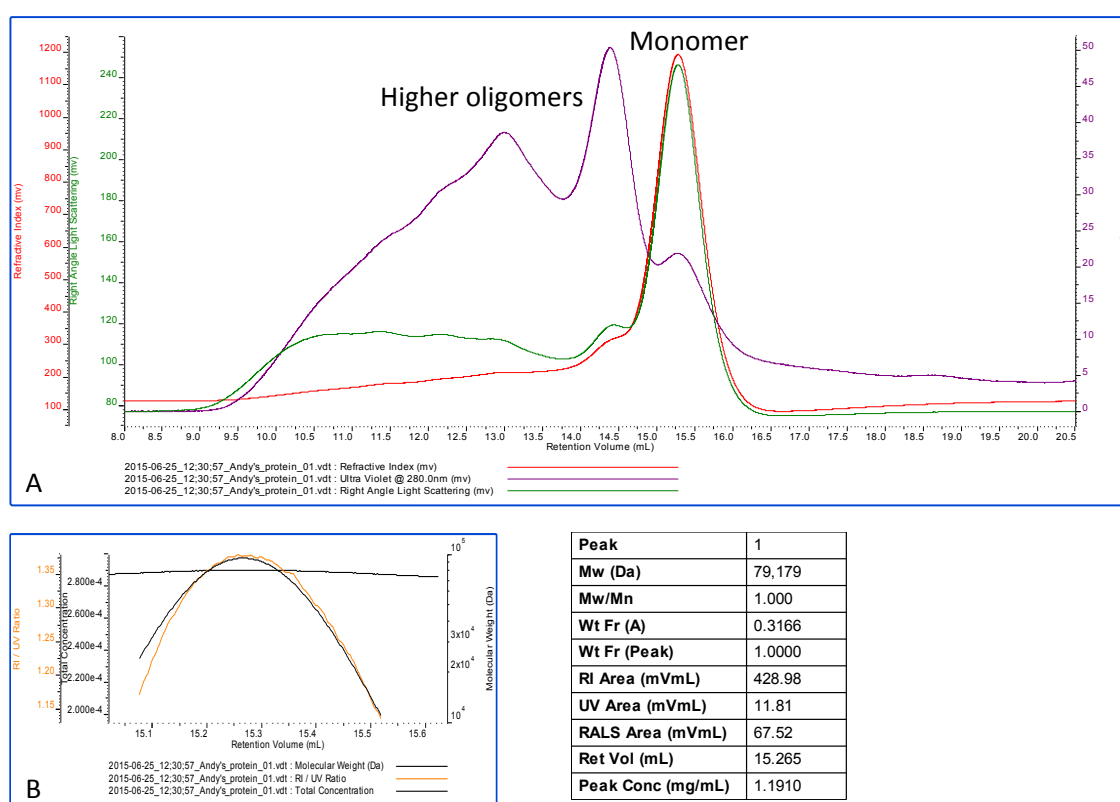


Figure 3.8 - Analysis of MucR⁷⁻²⁶⁶ oligomerisation states with SEC-MALS. A) The SEC-MALS profile of MucR⁷⁻²⁶⁶. UV light absorbance at 280 nm is shown in purple, refractive index in red and right angle scattering in green. Peaks corresponding to monomeric protein and higher oligomeric states are labelled. B) Further analysis of the monomeric protein peak, showing a high level of monodispersity. Statistics from analysis of the monomer peak are given to the right of B.

3.1.6 Thermostability analysis

As more stable proteins are often more likely exist in a single species and to crystallise^{332–334}, the stability of purified MucR^{7–266} in different detergents and buffers was examined. The fluorochrome N-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide (CPM) is not fluorescent in solution, but fluoresces upon specific binding to thiol groups, so can be used to assess thermostability of proteins that contain cysteines located within transmembranous helices³³⁵. MucR^{7–266} contains 3 cysteines, two of which are predicted by TMHMM2.0^{298,299} to be located in transmembrane helices, as depicted in Figure 3.9.

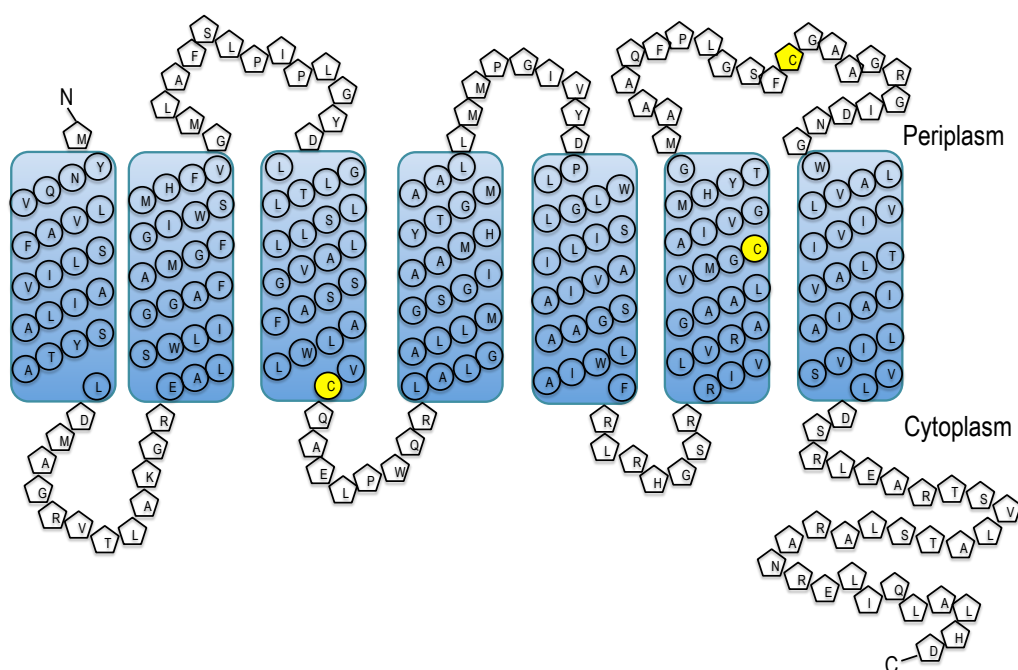


Figure 3.9 - Identification of transmembranous cysteine residues in MucR^{7–266}. Amino acids that are predicted to form transmembrane helices by TMHMM 2.0^{298,299} are denoted in circles, with periplasmic and cytoplasmic amino acids shown in pentagons. Cysteine residues are highlighted in yellow. It is worth noting that the transmembrane topology of MucR with a periplasmic N-terminus and a cytoplasmic C-terminus is consistent not only amongst membrane protein prediction algorithms^{299,336} but the presence of a cytoplasmic C-terminus is also supported by biochemical evidence¹⁵³.

As MucR⁷⁻²⁶⁶ also contains a solvent exposed cysteine residue, which would be likely to react with CPM whether the protein is correctly folded or denatured, it was considered necessary to determine what proportions of protein, dye and buffer display sufficient signal for the temperature of protein denaturation to be calculated. Six different ratios of protein: dye: buffer were screened, detailed in section 2.1.9 and given again in Table 3.1, however none of the conditions screened show sufficient signal for the temperature of denaturation to be determined, Figure 3.10. As a denaturation temperature could not be determined using this assay for MucR⁷⁻²⁶⁶ in 1% DDM as purified, it is impossible to compare this to other detergent and buffer mixtures without ambiguity, making this assay unusable for optimising MucR⁷⁻²⁶⁶ protein stability.

Table 3.1 - Composition of CPM thermostability reactions

Well	Component (μL)		
	Protein	CPM dye	Buffer B
A1	0.5	5	44.5
A2	0.5	7	42.5
A3	1	5	44
A4	1	7	42
A6	2	7	41
A7	4	5	41

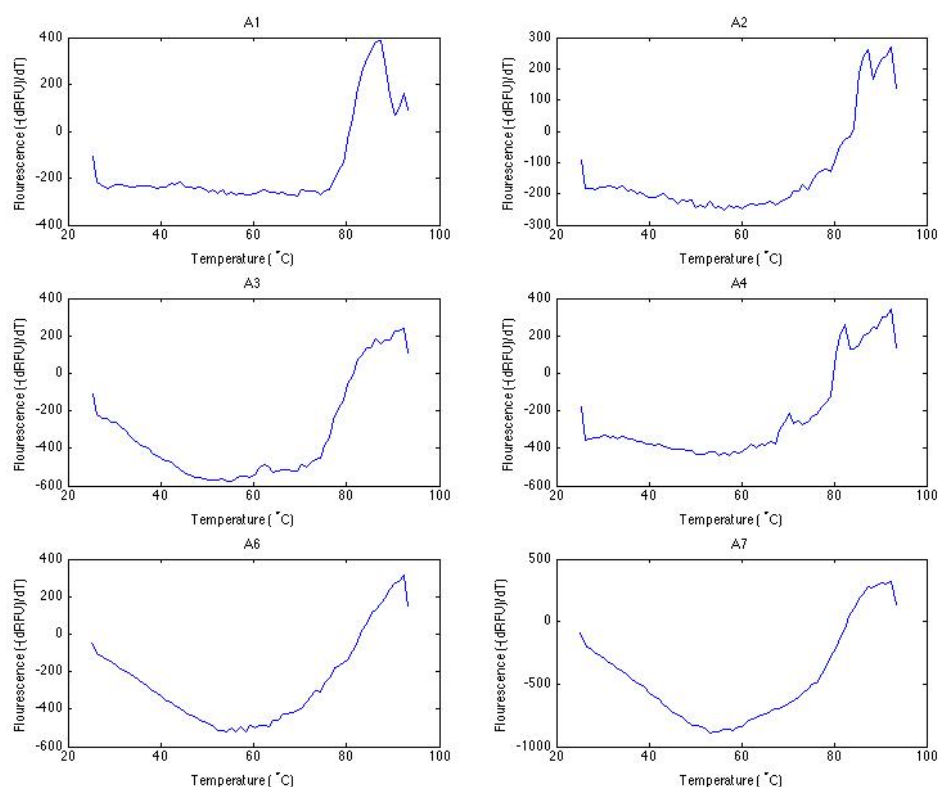


Figure 3.10 - First derivative thermostability profiles of CPM-MucR⁷⁻²⁶⁶ solutions given in Table 3.1.

The first derivative of change in CPM fluorescence is plotted against temperature for MucR⁷⁻²⁶⁶. Melting temperature (T_m), at which 50% of the protein is unfolded, would correspond to the nadir of such curves, but as no clearly defined nadir is apparent T_m cannot be determined.

3.2 Analysis of MucR and NbdA cytosolic domains

MucR and NbdA are multi-domain proteins, featuring cytosolic DGC and PDE domains. While it was not possible to structurally characterise the MHYT domains of MucR or NbdA, the cytosolic domains of both proteins may offer insight into how c-di-GMP metabolism could be regulated upon NO binding to MHYT domains, if MHYT domains function as NO sensors.

3.2.1 The EAL domain of MucR is inactive in isolation, but is an active PDE when expressed with the GGDEF domain

Gene fragments encoding for the DGC, PDE and the tandem DGC-PDE domains of MucR and NbdA were cloned into *E. coli* BL21(DE3) and expressed, as detailed in 2.1.10, 2.1.11 and 2.1.12. As the DGC domain of NbdA contains a degenerate GGDEF motif and has previously been demonstrated as inactive¹⁵⁵, this domain was not characterised as any role in NO-induced biofilm dispersal is

likely to be dependant on larger protein fragments. While no conditions could be determined that resulted in expression of NbdA^{EAL} and MucR^{GGDEF} and NbdA^{GGDEF-EAL} with sufficient stability for purification beyond IMAC (as shown in Figure 3.11), MucR^{EAL} and the MucR^{GGDEF-EAL} tandem domain fragment were purified through IMAC and SEC, shown in Figures 3.12 and 3.13 respectively.

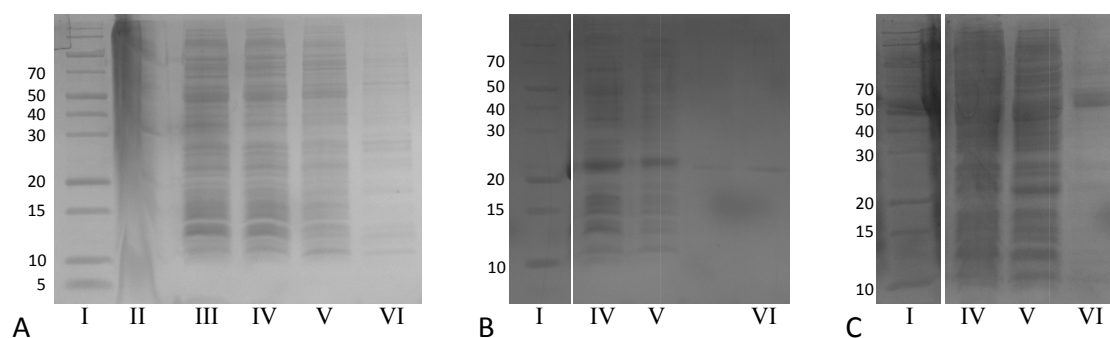


Figure 3.11 - Purification of cytosolic c-di-GMP metabolising domains from MucR and NbdA.

A) SDS-PAGE analysis of NbdA^{EAL} purification, with no band corresponding to NbdA^{EAL} (approximately 32 kDa) present in the soluble cell fraction. B) SDS-PAGE analysis of MucR^{GGDEF} purification, with a clear band corresponding to MucR^{GGDEF} present in IMAC-eluate (approximately 20 kDa), however protein precipitated before any further purification could take place. C) SDS-PAGE analysis of NbdA^{AGDEF-EAL} purification. A band corresponding to NbdA^{AGDEF-EAL} is present in IMAC-eluate (approximately 53 kDa), however the protein was not sufficiently stable for further purification. Lanes correspond to: PAGE ruler protein standard (I); cell lysate (II), soluble cell fraction (III), IMAC flow through, wash and eluate (IV-VI). Molecular weights of protein standards are shown to the left of each gel in kDa.

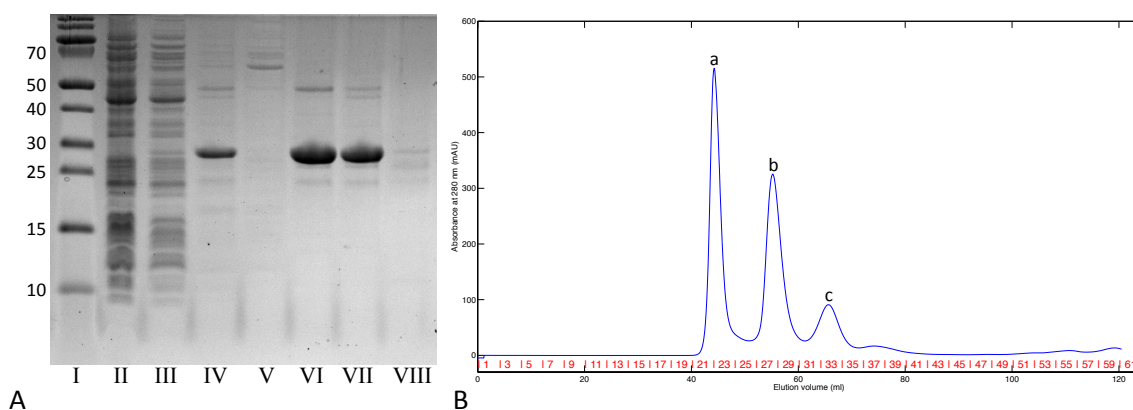


Figure 3.12 - Purification of MucR^{EAL}. A) SDS-PAGE showing MucR^{EAL} (approximately 30 kDa) is purified through IMAC, with purity increased still further through SEC. Lanes correspond to: PAGE ruler protein standard (I); IMAC flow through, wash and eluate (II-IV); SEC fractions 23, 28, 29 and 33 (V-VIII). Molecular weights of protein standards are shown to the left in kDa. B) Absorbance at 280 nm during purification MucR^{EAL} through SEC, with collected fractions labelled in red. SEC fractions collected in B correspond to the SDS-PAGE shown in A. When compared to typical results obtained for calibration of an S75 16/600 column (obtained from GE healthcare), peak 'a' corresponds to the column void volume, peak 'b' to approximately 55 kDa and peak 'c' to approximately 30 kDa. The dominant species in peak 'b' is therefore likely to be dimeric MucR^{EAL}, whilst the dominant species in peak 'c' is likely to be monomeric MucR^{EAL}.

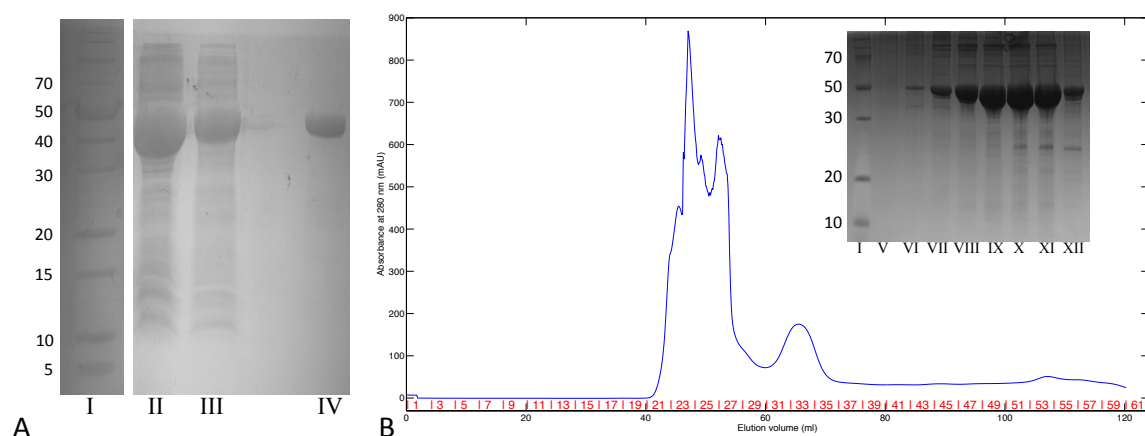


Figure 3.13 - Purification of MucR^{GGDEF-EAL}. A) SDS-PAGE showing purification of MucR^{GGDEF-EAL} (approximately 52 kDa) through IMAC. B) Purification of MucR^{GGDEF-EAL} by SEC, absorbance at 280 nm is denoted by a blue line, with corresponding SEC fractions (labelled in red) resolved through SDS-PAGE, shown inset. Lanes correspond to: PAGE ruler protein standard (I); IMAC flow through, wash and eluate (II-IV); SEC fractions 21, 22, 23, 24, 25, 26, 27 and 28 (V-XII). Molecular weights of protein standards are shown to the left of each gel in kDa.

When incubated with c-di-GMP in conditions suitable for PDE activity (i.e. at pH 9 in the presence of magnesium chloride^{196,198,201}) MucR^{EAL} does not degrade c-di-GMP to pGpG, however MucR^{GGDEF-EAL} does, as shown in Figure 3.14. Consistent with these observations, a tandem DGC-PDE fragment of MorA from *P. aeruginosa* has previously been demonstrated to show increased PDE activity in comparison to the PDE domain is expressed in isolation²⁰². MucR^{GGDEF-EAL} was also demonstrated to have DGC activity, capable of producing c-di-GMP when incubated with GTP.

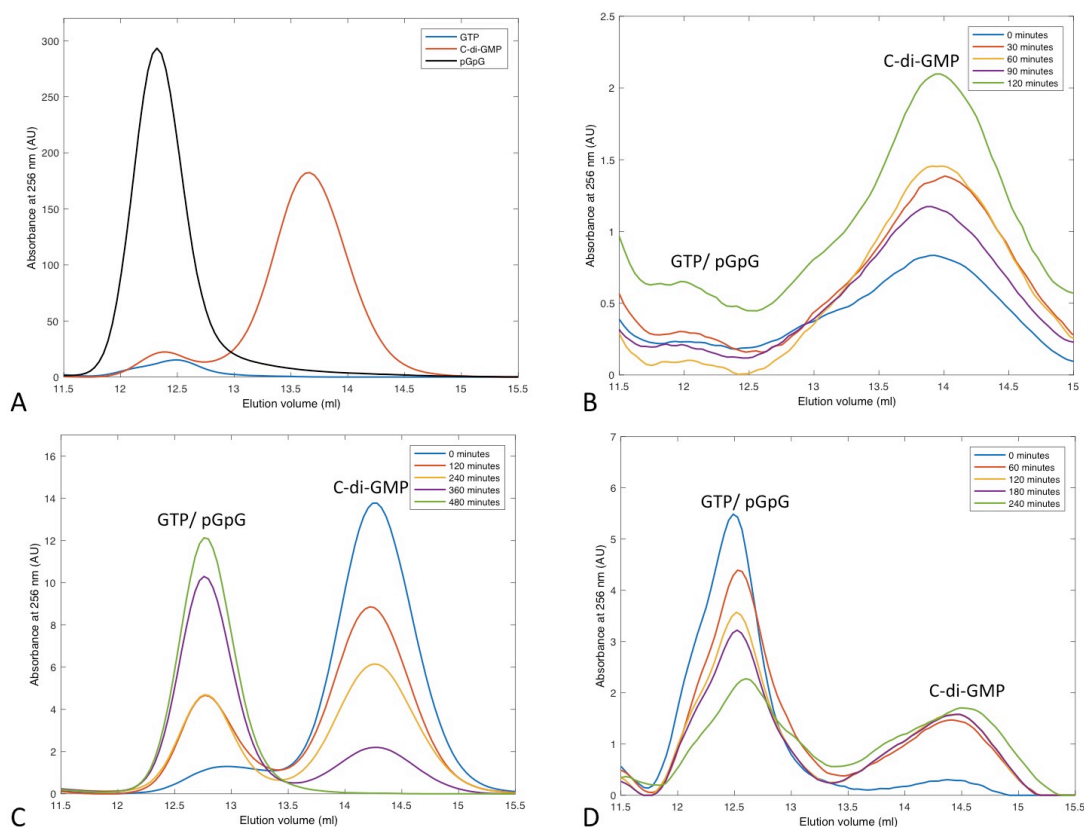


Figure 3.14 - Enzymatic activity of MucR^{EAL} and MucR^{GGDEF-EAL}. A) Resource Q elution profiles of nucleotide standards. B) Resource Q elution profiles of MucR^{EAL} incubated with c-di-GMP at 30 minute time points for two hours, with no peak corresponding to pGpG observed at any point. C) Resource Q elution profiles of MucR^{GGDEF-EAL} incubated with c-di-GMP in conditions compatible with PDE activity. Over time, levels of c-di-GMP decrease and levels of pGpG increase. D) Resource Q elution profiles of MucR^{GGDEF-EAL} incubated with GTP in conditions compatible with DGC activity. Levels of c-di-GMP increase over time but remain roughly constant after the first two hours.

3.2.2 Crystal structure of the MucR EAL domain

To determine if any structural differences are responsible for the low level of PDE activity, MucR^{EAL} and MucR^{GGDEF-EAL} were subjected to crystallisation trials, with successful crystallisation of MucR^{EAL} following 2-3 weeks incubation at 21°C in Morpheus crystallisation screen conditions D2, E3 and E4, as well as JCSG+ screen conditions H3 and H4. Crystals, shown in Figure 3.15, were diffracted on ID23-1 at ESRF on 29/11/2013 to a resolution of 2.5 Å, which was used to solve the structure of MucR^{EAL}, however further refinement was performed on data collected from crystals diffracted on I04 at Diamond Light Source on 28/04/2014 to a resolution of 2.3 Å. Data was

processed using the xia2 3dii pipeline^{337,338}, which uses XDS for indexing and integration of images³³⁹, followed by scaling with AIMLESS and removal of 5% of reflections for cross-validation^{340–342}. The structure of MucR^{EAL} was solved to a resolution of 2.27 Å using the PDB 2R6O as a molecular replacement model using the MOLREP software package³¹⁷. Model building was performed in Coot³⁴³, and REFMAC was used for structure refinement³⁴⁴, with MolProbity used to validate the final structure³⁴⁵. Full data processing and refinement statistics are presented in Table 3.2.

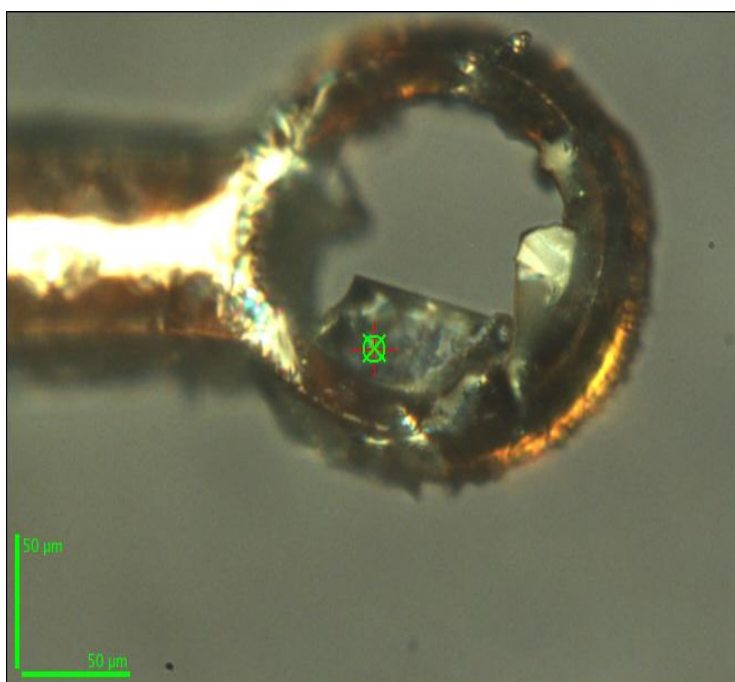


Figure 3.15 - A MucR^{EAL} crystal, diffracted at ID23-2, ESRF.

Table 3.2 - Crystallographic data and refinement statistics for MucR EAL.

Structure	MucR EAL
Data Collection	
Space Group	P2 ₁
Cell Axes (Å)	a = 46.4 b = 116.1 c = 52.1
Angles (°)	$\alpha = \gamma = 90$ $\beta = 102.5$
Beamline	I04
Wavelength (Å)	0.9795
Resolution (Å)	45.27-2.27 (2.33-2.27)
Unique Reflections (#)	24883 (1831)
Measured Reflections	170715 (12632)
Redundancy	6.9 (6.9)
R _{pim} (%)	6.9 (27.6)
I/ σ (I)	11.3 (2.9)
CC _{1/2}	0.99 (0.85)
Completeness (%)	99.9 (99.8)
Refinement	
Molecules/AU	2
R _{work} /R _{free} (%)	19.5/23.8
Rmsd	
Bond Length (Å)	0.011
Bond Angles (°)	1.440
Average B Factor (Å ²)	
Protein	36.6
Nucleotide	27.8
Metals	22.0
Water	36.0
Ramachandran (%)	
Favoured Regions	98.0
Allowed Regions	2.0
PDB Code	5M1T

MucR^{EAL} adopts an $\alpha\beta(\beta\alpha)_7$ -barrel fold, similar to other EAL-PDE structures, shown in Figure 3.16. Despite no ligands being soaked into the protein at any point during purification or crystallisation, MucR^{EAL} crystallised with c-di-GMP and two metal ions bound in the active site, as shown in Figure 3.16. It is assumed that the bound c-di-GMP associated with MucR^{EAL} during overexpression and is bound with sufficient affinity to remain associated during purification. Within MucR^{EAL}, the c-di-GMP P1 phosphate is coordinated by the side chains of Glu469, Asn528 Asp590 and Asp591, as well as both metal ions and 4 waters, whilst the P2 phosphate interacts exclusively with Arg473. One of the c-di-GMP guanine bases interacts with side-chain of Glu650 and the main chain of Phe669, whilst the side chain of Gln533, the main chain of Leu472 and three waters coordinate the other guanine, as shown in Figure 3.17A. The only other interaction

coordinating c-di-GMP within MucR^{EAL} is between a water and the 2'OH bridging a ribose to the P1 phosphate.

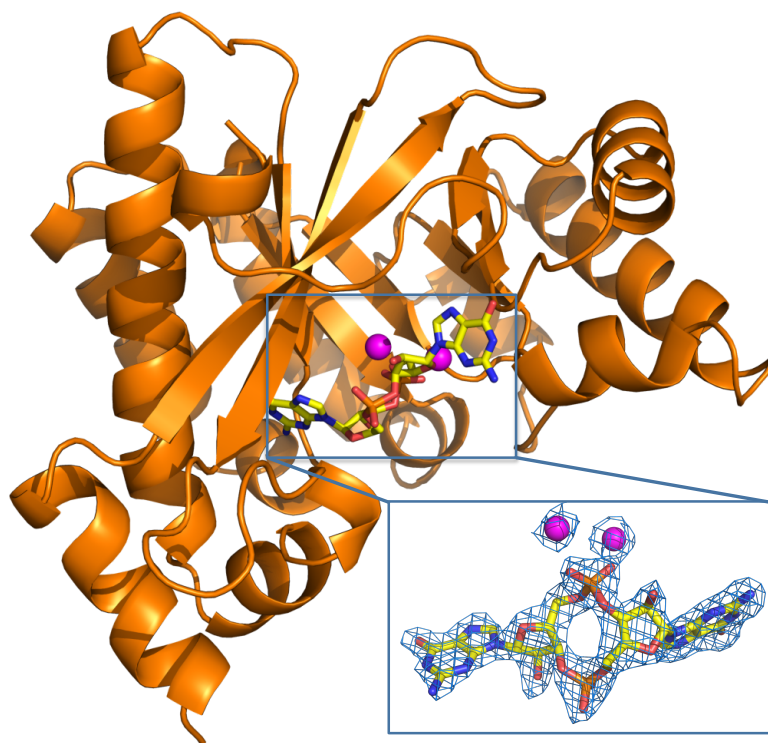


Figure 3.16 - The structure of MucR^{EAL} in complex with c-di-GMP. MucR^{EAL} is shown in cartoon representation, with c-di-GMP shown as sticks and magnesium ions shown as magenta spheres. The zoomed image of c-di-GMP and metal ions shows electron density of nucleotide and metals within a $2mFo-DFc$ electron density map, contoured to 2.0σ .

Metal ions within the EAL domain active site were assigned as magnesium ions, as magnesium chloride was present in all purification buffers (as detailed in section 2.1.13). One of these magnesium ions, labelled M1 in Figure 3.17B, is coordinated in an octahedral geometry by the side chains of Glu496, Asn528, Glu560, Asp590, as well as a non-bridging oxygen from c-di-GMP phosphate P1 and a water. The other magnesium ion, labelled M2 in Figure 3.17C, is coordinated by side chains of Asp590, Asp591, Glu647, a c-di-GMP P1 phosphate non-bridging oxygen and a water.

All of the residues involved in binding magnesium ions within MucR^{EAL} (Glu469, Asn528, Glu560, Asp590, Asp591 and Glu647), have previously been identified as conserved between EAL PDEs, with individual mutation of any of these residues abolishing PDE activity²⁰¹. However, the residues that directly bind c-di-GMP but not magnesium are not essential for function. Phe669 is

not conserved between EAL domains, and individual point mutation of residues equivalent to Gln455 and Glu650 to alanine have previously been shown to reduce PDE activity by less than 50%, whilst mutations equivalent to an Arg473Ala mutation reduced activity by approximately 70%²⁰¹. These previous mutation studies suggest that there is variation in possible binding of c-di-GMP to active EAL domains whilst still retaining catalytic activity, but magnesium ion coordination requires strict residue conservation. As coordination of metal ions within MucR^{EAL} is conserved with previous observations in active enzymes and variations in c-di-GMP binding are compatible with activity^{198,199,201}, differences in PDE activity of MucR^{EAL} and previously characterised EAL-PDEs are unlikely to be due to altered coordination of catalytic metals and substrate within the active site.

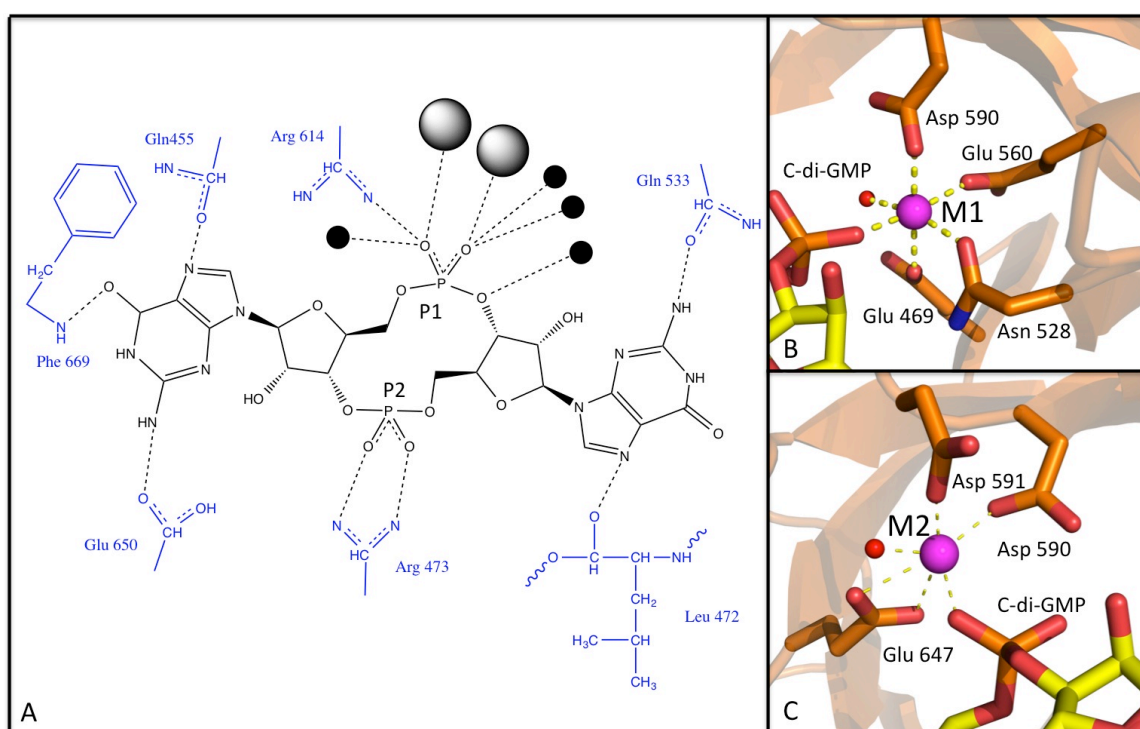


Figure 3.17 - Coordination of c-di-GMP and magnesium ions within MucR^{EAL}. A) C-di-GMP is shown in black with coordinating bonds from amino acids (shown in blue), metal ions (shown as grey circles) and water molecules (shown as black circles), displayed as dashed lines. B) Coordination of the M1 magnesium ion with MucR^{EAL}. C) Coordination of the M2 magnesium ion with MucR^{EAL}. Magnesium ions are shown as magenta spheres within B and C, with c-di-GMP shown as yellow sticks, water as a red spheres and bonds as yellow dashed lines.

Within the crystal, MucR^{EAL} forms a dimerisation interface around the $\alpha 5$ and $\alpha 6$ helices, shown in Figure 3.18A, and reported as the “classic dimer” in Table 3.3. Dimerisation around the $\alpha 5$ and $\alpha 6$ helices has previously been reported as a requirement for catalysis^{196,197,202}. Dimerisation reduces the length of the $\alpha 5$ helix and extends the $\beta 5$ - $\alpha 5$ loop, allowing conserved aspartic acid residues

located within the $\beta 5$ - $\alpha 5$ loop (corresponding to the Mg^{2+} binding Asp 590 and Asp 591 in MucR^{EAL}, which are part of the catalytic DDFGTG motif) to enter the active site and coordinate metal ions^{196,202}. MucR^{EAL} displays a short $\alpha 5$ helix (6 residues rather than 9) and extended $\beta 5$ - $\alpha 5$ loop (9 residues rather than 6) consistent with EAL-PDEs activated through this mechanism of dimerisation-induced metal coordination.

Comparison to other EAL-PDEs demonstrates that this dimerisation-induced change around the $\alpha 5$ helix is required for substrate binding, even though the dimerisation interface may vary. The EAL domain of PAO1 protein PA3825, as studied by collaborators Dom Bellini and Sam Horrell, forms a different dimer to that of MucR^{EAL}, referred to as the “closed dimer” in Table 3.3. PA3825^{EAL} is observed in this “closed dimer” when substrate-bound (Figure 3.18B) and when substrate-free (Figure 3.18C). However, a significant shift within the $\beta 5$ - $\alpha 5$ loop can be observed between these two activation states, facilitating active site formation and metal coordination via conserved aspartic acid residues of the DDFGTG motif. Structural rearrangement of the $\beta 5$ - $\alpha 5$ loop to facilitate a functional role in metal coordination within PA3825^{EAL} indicates that this method of dimerisation-induced activation is conserved, regardless of the dimer interface. All structurally characterised EAL-PDEs (listed in Table 3.4), regardless of the dimerisation interface, display rearrangement of the $\beta 5$ - $\alpha 5$ loop when substrate bound, as shown in Table 3.3, highlighting this as a requirement for substrate binding.

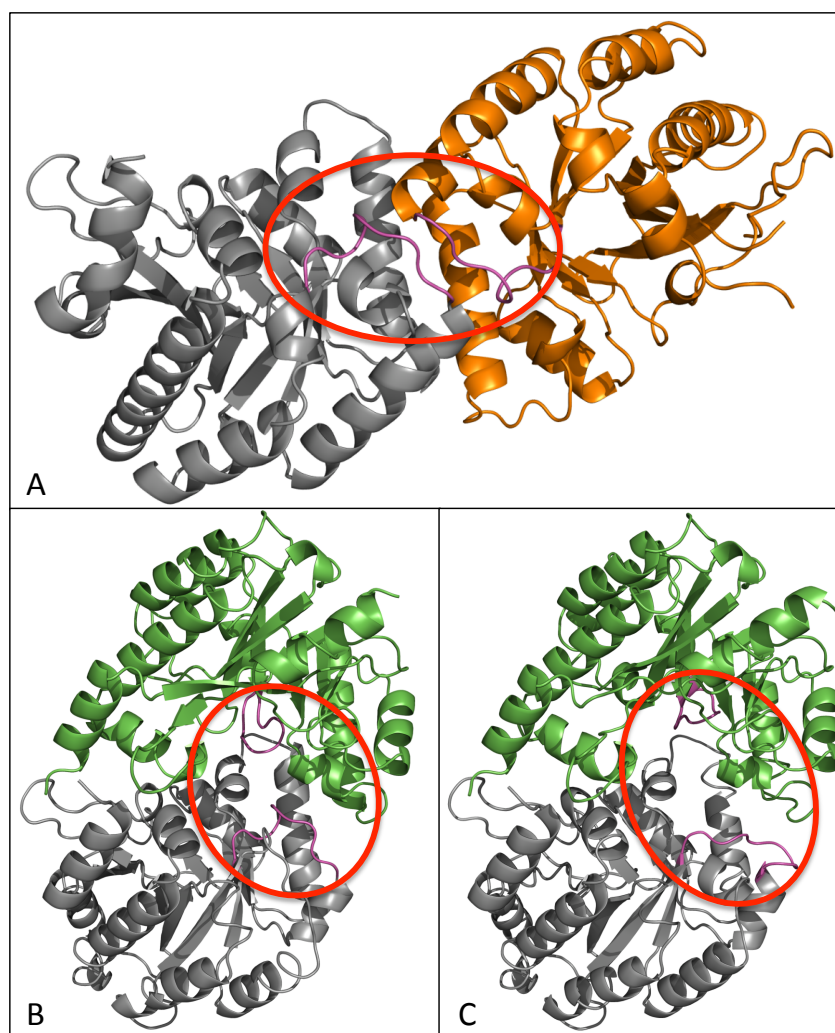


Figure 3.18 - The dimerisation interface of MucR^{EAL} contrasts to that of PA3825^{EAL}. A) MucR^{EAL} (one chain shown as an orange cartoon, the other as a grey cartoon) dimerises around the $\alpha 6$ helix, resulting in extension of the $\beta 5$ - $\alpha 5$ loop, shown in magenta. B) PA3825^{EAL} in complex with c-di-GMP dimerises around a different interface to that seen in MucR^{EAL}. The grey chain of PA3825^{EAL} superposes with the equivalent of MucR^{EAL} shown in A. The $\beta 5$ - $\alpha 5$ loop is similarly positioned within the MucR^{EAL} in complex with c-di-GMP and PA3825^{EAL} in complex with c-di-GMP, as evidenced by a C α RMSD of 1.186 Å between the two $\beta 5$ - $\alpha 5$ loops, with the two structures capable of being superposed with equivalent C α s positioned at an RMSD of 1.189 Å, according to the PDBeFold server³⁴⁶. C) PA3825^{EAL} forms a similar complex in the absence of substrate to that observed in the presence of substrate, but the $\beta 5$ - $\alpha 5$ loop (shown in magenta), undergoes a significant change in position. Despite it being possible to superpose the PA3825^{EAL} apo and PA3825^{EAL} c-di-GMP bound structures with a C α RMSD of 0.75 Å³⁴⁶, the $\beta 5$ - $\alpha 5$ loops within the superposed structures have equivalent C α s positioned with an RMSD of 7.542 Å.

Chapter 3

Table 3.3 - Dimerisation architectures of EAL domains.

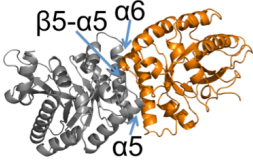
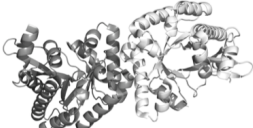
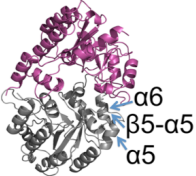

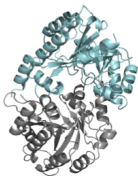
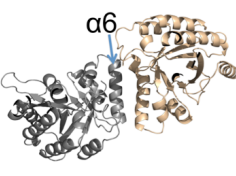
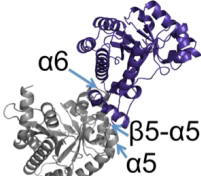
Type of dimer	Classic dimer		Closed dimer			Offset dimers	
							
Structure	MucR – Mg ²⁺ c-di-GMP (5M1T)	CC3396 – Mg ²⁺ pGpG (3U2E)	PA3825 – Apo (4Y9M)	PA3825 – Mg ²⁺ c-di-GMP (5MF5)	PA3825 – Mn ²⁺ pGpG (5MFU)	PA3825 – Ca ²⁺ c-di-GMP (4Y9P)	CC3396 – Ca ²⁺ c-di-GMP (4HJF)
α5 helix	6 residues	6 residues	9 residues	6 residues	6 residues	Not modelled	6 residues
β5-α5 loop	9 residues	9 residues	6 residues	9 residues	9 residues	Not modelled	9 residues
Site 1 metal	✓	✓	—	✓	✓	✓	✓
Site 2 metal	✓	✓	—	✓	—	✓	✓
Site 3 metal	—	✓	—	—	✓	—	—
Similar PDBs	3GG0 (Mn ²⁺ c-di-GMP), 3GG1 (Ca ²⁺ c-di-GMP), 3GFZ (Mn ²⁺ c-di-GMP), 3GFX (Ca ²⁺ c-di-GMP), 4RNH (Mg ²⁺ c-di-GMP), 3SY8 (Mg ²⁺ apo), 3N3T (Mg ²⁺ c-di-GMP), 2R6O (Mg ²⁺ apo)		4LJ3 (Ca ²⁺ c-di-GMP)			4Y8E (Ca ²⁺ apo) 4LYK (apo)	none

Table 3.4 - Classification of all EAL domain structures in the PDB to date.

No metal C-di-GMP	No metal No substrate	1 metal (M1) No substrate	1 metal (M1) C-di-GMP	2 metals (M1 & M2) C-di-GMP	pGpG
Monomeric <i>FimX-EAL (3HV8)</i> <i>FimX-EAL (4FOJ)</i> <i>FimX-EAL (4FOU)</i> <i>FimX-EAL (4FOK)</i> <i>FimX-EAL (4F3H)</i> <i>FimX-EAL (4F48)</i> <i>LapD-EAL (3PJW)</i> <i>LapD-EAL (3PJX)</i> <i>LapD-EAL (3PJU)</i> Dimeric <i>LapD-EAL (3PJT)</i>	Monomeric DosP-EAL (4HU3) <i>LapD-EAL (3PFM)</i> <i>FimX-EAL (3HV9)</i> CC3396-EAL (3S83) Dimeric PA3825-EAL (4Y9M) Ykul (2BAS) <i>YdiV-EAL (3TLQ)</i> Imo0131-EAL (4Q6J) MorA-EAL (4RNJ) MorA-EAL (4RNI) <i>FimX-EAL (4AG0)</i>	Dimeric Mg²⁺ TBD1265-EAL (2R6O) Ca²⁺ PA3825-EAL (4Y8E) Tetrameric Mg²⁺ RocR (3SY8)	Dimeric Mg²⁺ MorA-EAL (4RNH) Ca²⁺ Blrp1_B (3GFX) Ykul_AB (2W27)	Dimeric Mn²⁺ Blrp1 (3GFZ) Blrp1 (3GG0) Mg²⁺ PA3825-EAL (5MF5) MucR-EAL (5M1T) TBD1265_EAL (3N3T) Ca²⁺ PA3825-EAL (4Y9P) CC3396-EAL (4HJF) Blrp1_A (3GFX) Blrp1 (3GG1) YahA (4LI3)	Dimeric Mn²⁺ PA3825-EAL (5MFU) Mg²⁺ CC3396-EAL (3U2E) no metal <i>FimX-EAL (4AFY)</i>

Inactive EAL domains are shown in italics.

3.2.3 The product bound state of PA3825 may demonstrate further regulation of EAL PDEs

The observation of putatively active dimers of MucR^{EAL} and of catalytically active EAL-PDEs, listed in Table 3.4, maintaining bound substrate through the course of crystallisation suggests that other mechanisms may be involved in regulating activity of EAL-PDEs. The structure of PA3825^{EAL} bound to the hydrolysis product pGpG, determined by Dom Bellini and Sam Horrell, may offer new insight into what form this regulation may take. Co-crystallisation of PA3825^{EAL} with c-di-GMP and Mn²⁺ resulted in the pGpG product complex, shown in Figure 3.19A. When in complex with pGpG, one metal ion is found in identical position to the M1 metal within the substrate bound complex of PA3825^{EAL} and of MucR^{EAL}, shown in Figure 3.19B. However, no metal is observed in a position superposable to M2, with corresponding interactions involved in coordinating the nucleotide P1 phosphate replaced by interactions between the phosphate and Glu217 (contributed by the β 7- α 7 loop, in close vicinity to the β 5- α 5 loop). A second metal ion is observed close to the hydrolysed phosphodiester bond of pGpG within the active site of PA3825^{EAL}, but this occupies a distinct metal binding site, designated as M3 in Figure 3.19. Anomalous data collection above the manganese absorption edge confirmed the metal within the M1 site as Mn²⁺, but displayed no signal in the M3 site. The M3 metal is tetrahedrally coordinated by the side chain of Asp160, the O2' ribose oxygen, a P1 hydroxyl oxygen of the pGpG terminal phosphate and a water molecule. Consistent with a tetrahedral coordination geometry and absence of anomalous signal for Mn²⁺, the metal within the M3 site was refined as sodium as this is the most abundant metal in the crystallisation buffer and would fit the observed geometry³⁴⁷.

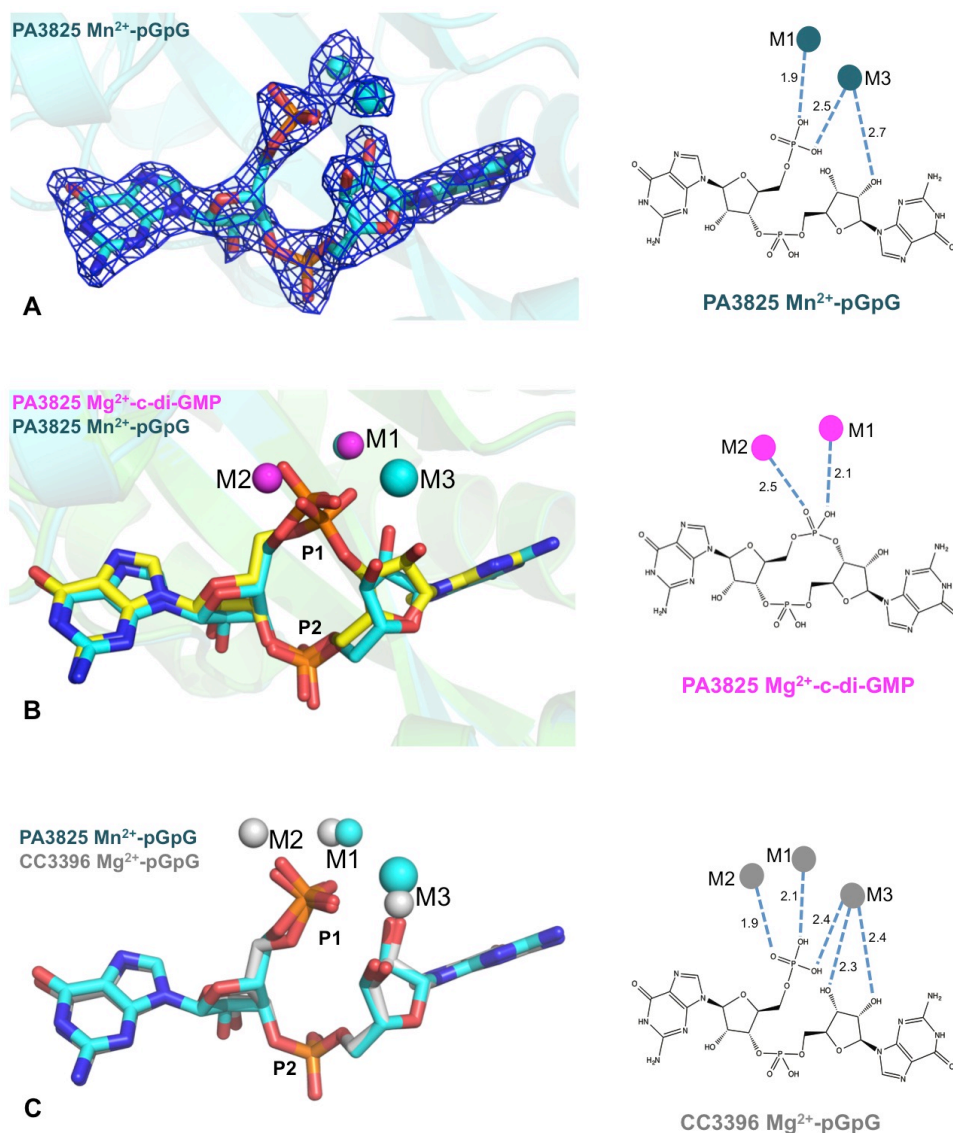


Figure 3.19 - The M3 metal site is concurrent with bound product within EAL-PDEs. A) The PA3825^{EAL} pGpG complex displays metals in the M1 and M3 positions. Electron density for pGpG and metal ions is shown, contoured to 1.3 σ . B) Superposition of PA3825^{EAL} substrate and product complexes. The substrate c-di-GMP is shown as yellow sticks, with Mg²⁺ ions (shown as magenta spheres) occupying the M1 and M2 metal sites. The product pGpG is shown as blue sticks, with Mn²⁺ and Na⁺ ions (shown as blue spheres) occupying the M1 and M3 positions. C) Overlay of PA3825^{EAL} and CC3396^{EAL} product complexes. Within the CC3396^{EAL} complex, Mg²⁺ ions (shown as silver spheres) occupy the M1, M2 and M3 positions. Interactions between bound metal ions and nucleotide are shown through the schematics on the right.

Two other structures of EAL domains in complex with pGpG have been uploaded to the PDB (Table 3.4). One of these, the *P. aeruginosa* protein FimX³⁴⁸ (PDB code 4AFY), lacks the DDFGTG motif for c-di-GMP hydrolysis and is an inactive PDE^{194,200,349}. Within the FimX-pGpG complex the positioning of the guanine bases differs from the observed equivalents in all c-di-GMP EAL-PDE

complexes, including MucR^{EAL} and PA3825^{EAL}, which overlay well with the positioning of bases in the product-bound form of PA3825^{EAL}. Within the other pGpG-bound EAL domain structure, that of CC3396^{EAL} from *Caulobacter crescentus* (PDB code 3U2E), the pGpG adopts a similar position to its equivalent in the PA3825^{EAL} complex. Three metal ions (modelled as Mg²⁺) are observed within the active site of CC3396^{EAL} in complex with pGpG, these metal ions occupy the M1 and M2 binding sites of substrate bound EAL domains, as well as the M3 site observed in the PA3825^{EAL} pGpG complex. Coordination of the M3 metal site differs between PA3825^{EAL} and CC3396^{EAL}, interactions with the conserved aspartate in PA3825^{EAL} are replaced within CC3396^{EAL} by interactions to the O3' oxygen of the pGpG ribose sugar and water. The presence of metal ions in M1, M2 and M3 metal binding sites within CC3396^{EAL} suggests the M3 metal site identified in the PA3825^{EAL} pGpG complex is not simply the migration of a metal from the M2 position and that a metal in the M3 site discriminates between substrate and product bound states, as summarised in Figure 3.20.

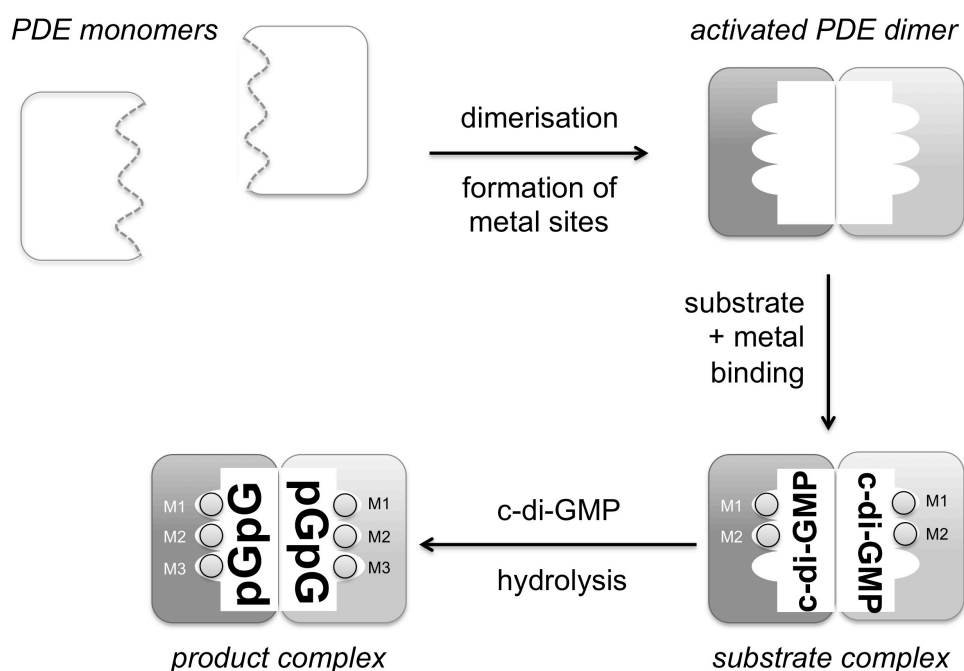


Figure 3.20 - Proposed regulatory steps involved in hydrolysis of c-di-GMP by EAL-PDEs.

Monomeric EAL-PDEs are inactive, with dimerisation inducing the formation of metal binding sites within the active site. The now formed active is able to bind c-di-GMP and metals within the M1 and M2 positions. Hydrolysis of c-di-GMP may be further regulated by metal binding to the M3 metal site.

While a requirement for a metal ion within the M3 site during c-di-GMP needs confirmation, any functional role is likely to differ from those proposed for metals within M1 and M2. Metals in the M1 and M2 sites are thought to be required for activation of a hydroxide nucleophile^{184,198}, however the M3 site is not suitably positioned for such a role and is more likely to function in stabilising the negatively charged reaction intermediate formed during hydrolysis of c-di-GMP to pGpG. Quantum mechanics calculations support a similar role for a third metal ion within nuclease P1³⁵⁰ and endonuclease IV³⁵¹, both of which had previously been structurally determined to bind three Zn²⁺ ions^{352–354}. The reaction mechanism of nuclease P1 is shown in Figure 3.21. Kinetic analysis of T5 endonuclease across a range of magnesium ion concentrations demonstrates a requirement for three metal ions in catalysis³⁵⁵, suggesting this phenomenon may also apply for nucleases using magnesium ions. A similar functional role and requirement for three metals within EAL-PDEs may explain the abundance of substrate-bound structures similar to MucR^{EAL} within the PDB. A functional role for a third metal ion may also be linked to why MucR^{EAL} is able to bind c-di-GMP but not complete substrate hydrolysis despite conservation of residues required for catalysis if MucR^{EAL} were unable to coordinate a metal in the M3 position. Such a theory requires further experimentation, but it is possible that within the context of the full-length protein (and within the MucR^{GGDEF-EAL} fragment) that another domain is suitably positioned to contribute to M3 coordination. If a metal in the M3 site is a functional requirement this could also offer another method for regulating c-di-GMP hydrolysis by EAL-PDEs. It is therefore conceivable that any stimulatory signal perceived by the MHYT domain of MucR (such as NO, although this requires confirmation) could regulate c-di-GMP metabolism by propagating conformational change to facilitate M3 site formation, although differences in the M3 metal coordination sphere between the structures of PA3825^{EAL} and CC3396^{EAL} make it difficult to speculate upon exactly what these conformational changes may entail.

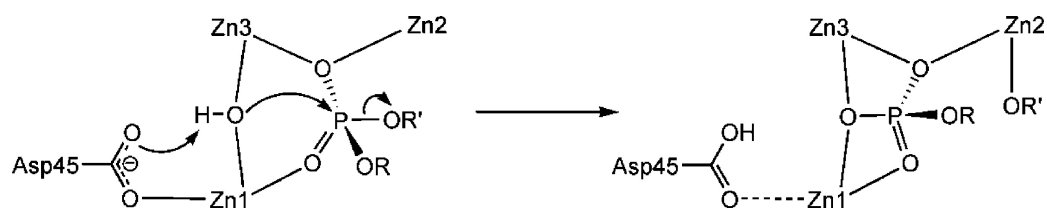


Figure 3.21 - Reaction mechanism of nuclease P1. Taken from Liao *et al.* 2010³⁵⁰.

3.3 Further work required to characterise regulation of c-di-GMP metabolism in MucR and NbdA

A number of approaches could be pursued to determine if MucR and NbdA directly regulate c-di-GMP turnover in response to NO during NO-induced biofilm formation. Attempts to crystallise MucR⁷⁻²⁶⁶ may be improved by examining thermostability of the protein in different detergents and buffers, with stability measured by circular dichroism (CD) or SEC-MALS; however either of these approaches would be likely to require large quantities of protein. Another possible approach to structural characterisation of MYHT domains could be to select some of the different fragments discussed in section 3.1.1, and carry out equivalent work to that performed on MucR⁷⁻²⁶⁶, as a different fragment may be more monodisperse when resolved by SEC and thus be a more suitable crystallisation target. An alternative to this could be to look at MYHT domain proteins from other species, as a more stable MYHT domain protein, such as one expressed in thermophilic bacteria, may be more likely to crystallise. Alternative approaches that could be implemented to structurally characterise MYHT domains include the use of solid-state NMR and cryo-electron microscopy. The MYHT domain in isolation (MucR⁷⁻²⁶⁶ is approximately 27 kDa in mass), and even the full-length protein of MucR (approximately 74 kDa), are below the current minimum size-requirement of cryo-electron microscopes (approximately 150 kDa)³⁵⁶. However, as both DGC and PDE domains require dimerisation for activity, discussed in section 1.3.1 and 1.3.2 respectively, it may be possible to use non-hydrolysable substrate analogues to lock full-length MucR into a dimeric arrangement, which may be amenable to microscopy. Solid-state NMR could be used to characterise the MYHT domain in isolation or MYHT-domain containing proteins in full. However both cryo-EM and NMR would still require purified, monodisperse protein, which has not been obtained through this work.

Numerous questions still remain about MYHT domains that could be answered without the need for structural characterisation. As metal binding has not been experimentally determined for MYHT domains, MYHT domains could be purified from source and analysed with mass spectrometry to identify any bound metals, if present. Alternatively, the thermostability of MYHT domains from recombinant expression and purification could be examined in a range of different metal ions, as an increase in stability may be indicative of physiological binding. It may be possible to further characterise interactions between any bound metal, if present, and nitric oxide through vibrational spectroscopic techniques, such as infrared and Raman spectroscopy, but such a possibility would depend heavily on the identity of the associated metal. Another experiment that could directly answer whether MucR and NbdA directly regulate c-di-GMP metabolism in

response to NO would be to add NO to the full-length protein and test for any change in DGC or PDE activity.

As regulation of EAL-PDEs is likely to be linked to NO-induced biofilm dispersal (even if MucR and NbdA don't serve as NO sensors), any functional role for a third metal in EAL-should be further examined. One experimental method could be to measure the enzyme kinetics of EAL-PDEs across a range of metal ion concentrations, as has been performed on T5 flap endonuclease³⁵⁵. Alternatively, quantum mechanics calculations could be applied to determine if a metal positioned within the M3 site would reduce the activation energy of the reaction and clarify whether formation of the M3 site may be a regulatory mechanism in EAL-PDEs.

Chapter 4: Spectroscopic and structural characterisation of the NO binding protein BdlA

Gene knockout studies within the *Pseudomonas aeruginosa* strain PAO1 demonstrate a requirement for the putative chemotaxis protein BdlA in NO-induced biofilm dispersal¹⁰⁴. BdlA is a cytoplasmic protein predicted to consist of two PAS domains N-terminal to a methyl-accepting chemotaxis (MA) domain²⁸⁶. Despite not possessing DGC or PDE domains, a role for BdlA in NO-induced biofilm dispersal can be somewhat rationalised through the observation that the first PAS domain of BdlA binds a haem-*b* cofactor, making it possible that BdlA might function as a sensor for NO, and then propagating this signal through the cell to enzymes capable of reducing levels of c-di-GMP²⁸⁸. It is possible that BdlA could signal to c-di-GMP regulating domains through altered PAS domain dimerisation²²⁸, with NO-bound BdlA promoting dimerisation of PDE proteins also containing a PAS domain or by inhibiting dimerisation of proteins contain both a PAS domain and a DGC domain. An alternative possibility is that the binding of NO to the haem cofactor within BdlA may alter signalling through the MA domain. To understand the function of MA domains, and how the MA domain within BdlA may propagate the signal of NO binding to induce biofilm dispersal, it is necessary to discuss the nature of chemotaxis within *P. aeruginosa*.

4.1 Chemotaxis within *P. aeruginosa*

Chemotaxis is the ability of bacteria to alter behaviour in response to changes in the concentration of chemicals in the surrounding environment, with chemotaxis shown to play a key role in *P. aeruginosa* animal infection^{357,358} as well as the development and dispersal of *P. aeruginosa* biofilms^{359–361}. Chemotaxis signalling is initiated by the binding of chemotactic ligands to cell surface chemoreceptors. Chemotactic ligands range from small molecules like oxygen^{269,362}, to larger compounds such as amino acids^{363,364} and fatty acids³⁶⁵. Chemoreceptors are referred to methyl-accepting chemotaxis proteins (MCPs), and are so-called because they contain MA domains³⁶⁶. Chemotactic signals are thereafter generated by the MA domains which, via a series of chemotaxis (Che) proteins, regulate the flagellar motor.

The *P. aeruginosa* PAO1 genome contains 26 genes that encode for MCP-like proteins, as well as more than 20 genes predicted to encode for Che proteins, with the 20 *che* genes encoded in five distinct clusters (Che, Che2, Pil-Chp, Wsp and *cheVR*)³⁶⁷. These chemotaxis genes have been linked to four systems within *P. aeruginosa*. The Che and *cheVR* clusters (also known as cluster I and cluster V) are thought to work together to form the flagella-mediated Che chemotactic

pathway^{281,368}, which is distinct from the other form of flagella-mediated chemotaxis regulated by the Che2 cluster of *che* genes (cluster II)²⁶⁸. Flagella-mediated chemotaxis is dependent upon the Che pathway, whilst the role of the Che2 pathway is less clear^{267,281}. The Pil-Chp cluster of *che* genes (also known as cluster IV) are believed to be involved in a pili-mediated method of chemotaxis, regulating twitching motility^{369–371}. The Wsp cluster of *che* genes (cluster III) are involved in biofilm formation³⁷². Due to the complex nature of the overlapping chemotaxis systems within *P. aeruginosa*, understanding of flagella-mediated *P. aeruginosa* chemotaxis is derived through homology to that of *E. coli*, which contains only a single chemotaxis pathway^{367,373,374}.

Within the archetypal *E. coli* chemotaxis system, ligands are detected through binding to the MCP, which contains an MA domain. Within *P. aeruginosa* most of the predicted MCPs are transmembranous, featuring a periplasmic ligand binding receptor, with cytosolic HAMP (a commonly used signalling domain²⁷¹) and MA domains. The HAMP domain(s) is thought to transduce the signal perceived by the ligand binding domain into altered behaviour of the MA domain^{268,375}, with the MA domain signalling within the cell. However, it is worth noting that this domain architecture is not the case for all MCPs, as exemplified by the aerotaxis sensor proteins Aer and Aer2²⁶⁸. Aer contains a cytosolic PAS domain responsible for ligand binding, separated from the MA domain by two transmembrane helices²⁶⁹, whilst Aer2 does not contain any transmembranous regions, instead consisting of 3 N-terminal HAMP domains, a haem-binding PAS domain (responsible for ligand perception), a further 2 HAMP domains and a C-terminal MA domain (shown in Figure 1.22)^{252,265,268}. In a number of bacterial species, MCPs have been observed localised to the poles, however within PAO1 the MCPs contributing to different pathways have been observed to form distinct signal transduction complexes³⁷⁶.

Through analogy to *E. coli*, the MA domains of MCPs in the Che pathway are thought to form stable ternary complexes with the coupling protein CheW and the histidine kinase CheA^{272–275}. Conformation changes within the MCP upon ligand binding are propagated to the MA domain and alter the autophosphorylation activity of CheA^{272,377}. In *E. coli*, the binding of an attractant ligand to the MCP decreases the autophosphorylation activity of CheA, with the opposite effect observed upon the binding of repellent²⁷⁶. However, the binding of attractant to MCP in *Bacillus subtilis* enhances CheA autophosphorylation activity³⁷⁸. It is currently unclear whether binding of attractant to MCPs in *P. aeruginosa* results in increased or decreased CheA autokinase activity³⁷⁴. The role of CheW in this process is unknown, but the presence of CheW has been demonstrated to be a requirement for CheA autophosphorylation²⁷².

Within *E. coli*, the phosphoryl group of autophosphorylated CheA is transferred to onto a regulatory protein, with CheY and CheB competing to receive it. Once phosphorylated, CheY goes on to interact with FliM and FliN in the flagellar switch complex^{379–381}. To date, little is known about the *P. aeruginosa* flagella motor, so it is therefore hypothesised to function similarly to that of *E. coli*, with anticlockwise rotation the predominant driving force. Upon the binding of phosphorylated CheY, rotation of the flagellar motor is altered to increase the probability of clockwise rotation and cause the bacterium to tumble^{272,279}. CheY can then be dephosphorylated by the CheZ phosphatase³⁷⁷.

Once phosphorylated, CheB functions to “tune” the chemotaxis response. CheB is a methylesterase specific to the MA domains of MCPs, working in opposition to the methyltransferase CheR. Using S-adenosylmethionine as a methyl donor, CheR methylates specific glutamate residues within the MA domain³⁸². Once phosphorylated, CheB acts to demethylate these methylglutamate residues^{383,384}. Demethylation of the MA domain increases the levels of CheA autokinase activity, subsequently increasing the levels of CheY phosphorylation and promoting the tumbling behaviour of the bacterium. These changes in MA domain methylation state occur much slower than the phosphorylation/ dephosphorylation steps occurring within the chemotaxis pathway³⁷⁴, and differences in the levels of receptor methylation allows comparison between the current concentration and previously detected concentrations of chemotactic signalling ligand, facilitating the perception of a chemical gradient^{367,374}. Perception of a chemical gradient can guide the bacterium to moving in to high concentrations of attractant, or to distance itself from repellent ligand.

Observations of increased methylation upon binding of chemotactic attractants within *P. putida*³⁸⁵ suggest that there is a decrease in phosphorylation of CheB in response to binding of chemotactic attractants, making it likely that binding of attractants to MCPs in *Pseudomonas* species reduces the autokinase activity of CheA, reducing tumbling activity, following the paradigm of chemotaxis provided by *E. coli*²⁷⁶. It is worth noting however that while *E. coli* contains only a single *cheR* gene, 4 paralogues are present within the PAO1 genome: CheR, CheR2, PilK (also known as CheR3) and WspC³⁸⁶. Isothermal titration calorimetry (ITC) studies have identified a GWEEF pentapeptide, common to the C-terminus of many MA domains, as a binding site for CheR2 but not for CheR, PilK or WspC³⁸⁶. The same study suggests that the different CheR paralogues may be responsible for methylation specific to certain MCPs. As CheR has been demonstrated to be part of the Che pathway and to be involved in biofilm development³⁶⁰ it is possible that BdIA could act as an MCP within the Che pathway, promoting flagella-mediated motility and promote biofilm dispersal in response to NO²¹.

Chapter 4

Despite suggestion that it may be involved in aerotaxis, the role of the Che2 pathway in *P. aeruginosa* is currently undetermined^{265–268,270,387}. The Che2 pathway is thought to be orthologous to the Che pathway, with Aer2 fulfilling the role of an MCP, perceiving signal through a haem PAS domain and propagating this signal onto CheA2, with CheW2 involved in stabilising this complex. CheA2 is believed to be a histidine kinase capable of phosphorylating CheB2 and CheY2^{265,276–278}. As the Che2 pathway has not yet been implicated in a distinct phenotype it is unclear what role phosphorylated CheY2 may play within the cell. CheR2 has been observed to specifically interact with Aer2, and in combination with CheB2 is thought to regulate the methylation state of the Aer2 MA domain³⁸⁶. Recombinant expression of Aer2 in an *E. coli* mutant lacking MCPs demonstrates that Aer2 is able to utilise the *E. coli* Che pathway in response to O₂. These mutant *E. coli* cells displayed increased tumbling upon O₂ exposure, the opposite of the reduced tumbling and increased swimming observed within the wild-type *E. coli* aerotaxis response²⁶⁵. This observation raises the possibility that substrate binding to Aer2 results in increased CheA2 autokinase activity, signalling in the reverse manner to that expected of MCPs in the Che pathway. As both Aer2 and BdlA contain haem PAS domains as well as MA domains, one could speculate that both proteins may be involved in coordinating the Che2 pathway, although no evidence for this theory has been observed to date.

The Pil-Chp pathway regulates the production of the type IV pilus and twitching motility (a flagella independent form of motion used in host colonisation and biofilm formation³⁸⁸) within *P. aeruginosa*^{369–371}. Although the Pil-Chp pathway has not been studied as extensively as the Che pathway of flagella-mediated chemotaxis, the Pil-Chp pathway contains homologues of CheA (ChpA), CheB (ChpB), CheR (PilK), CheW (Pill and ChpC), CheY (PilG and PilH) and a MCP (PilJ), leading to the proposal of pili-chemotaxis models analogous to those of flagella-mediated chemotaxis^{371,388,389}. Localised to both poles within PAO1, PilJ is presumed to act as an MCP controlling twitching motility, as evidenced by the requirement of PilJ for type IV pili retraction and extension³⁹⁰. PilJ is thought to form a complex with ChpA, Pill and ChpC, with binding of chemotactic ligand to PilJ altering the autophosphorylation of ChpA. ChpA phosphorylates PilG and PilH, with phosphorylated PilG expected to interact with a complex including the PilB and PilZ ATPases, as well as the diguanylate cyclase FimX, serving to mediate pilus extension³⁸⁹. Phosphorylated PilH is thought to interact with the PilT and PilU ATPases to mediate pilus retraction. It is currently unknown whether PilB and PilT compete for phosphorylation by ChpA. Autophosphorylated ChpA is also thought to phosphorylate ChpB, with phosphorylated ChpB acting as a methyltransferase upon the MA domain of PilJ, countering the methyltransferase activity of PilK to generate a temporal marker of chemotaxis activation^{389,391}. Of all the chemotaxis pathways within *P. aeruginosa*, any involvement of BdlA in the Pil-Chp pathway is perhaps the

hardest to rationalise. However, twitching motility, regulated by the Pil-Chp pathway, is involved in biofilm development³⁹², making it therefore possible that BdlA may be involved in modulating this pathway to promote biofilm dispersal.

The Wsp pathway contains a CheB-like methyltransferase (WspF), a CheR-like methyltransferase (WspC), two homologues of CheW (WspB and WspD), a MCP (WspA), with the CheA and CheY replaced by the histidine-kinase response regulator WspE and the GGDEF domain containing response regulator WspR, respectively⁹⁷. All of the proteins within the Wsp pathway are predicted to form a complex around the transmembranous WspA³⁹³. Unlike the MCPs within the Che pathway, WspA has not been observed to localise at the cell poles, instead being distributed across the whole cell periphery^{393,394}. The C-terminal MA domain within WspA shows low levels of amino acid identity to MA domains present in other PAO1 MCPs (less than 14% identity in comparison to PctA, PctB, PctC and PA2652), with chimeric fusion proteins demonstrating that this unique MA domain is required for the distribution of WspA³⁹³. The Wsp pathway is activated in response to growth upon a surface, rather than in response to the chemical gradients responsible for canonical chemotaxis behaviour³⁹⁴. However, it is worth noting that the physical or atomic nature of the Wsp activatory signal is currently unknown³⁹³. Activation of WspA is proposed to result in autophosphorylation of WspE and subsequent phosphorylation of WspR, activating WspR to increase c-di-GMP production and contribute to biofilm formation^{393,394}.

10 of the 26 MCPs encoded within the PAO1 genome have been functionally characterised^{3,367,395}. The MCPs PctA, PctB and PctC mediate chemotaxis towards the 20 amino acids commonly found in proteins^{363,364}, whilst also mediating a repellent response away from chloroform and methylthiocyanate^{396,397}. CtpH and CtpL mediate chemotaxis towards inorganic phosphate³⁹⁸, whilst PA2652, TlpQ and McpA mediate chemotaxis towards malate, ethylenes and chloroethylenes respectively^{397,399,400}. Aer is responsible for chemotaxis towards oxygen (known as aerotaxis³⁸⁷, for which a possible involvement of Aer2 is a matter of debate²⁶⁵⁻²⁷⁰). Whilst Aer2 is hypothesised to act as an exclusive input for the Che2 pathway, PilJ to input into the Pil-Chp pathway, and WspA to input into the Wsp pathway, all of the other identified MCPs are thought to signal into the Che pathway^{97,267,386,390}.

The chemotaxis pathways in *P. aeruginosa* are summarised in Figure 4.1. The complex nature of the multiple chemotaxis pathways within *P. aeruginosa* makes it difficult to speculate on a role for the MA domain of BdlA, as BdlA has not yet been associated to any single chemotactic pathway. As BdlA is involved in biofilm dispersal one could postulate that BdlA may be involved with regulation of the Wsp pathway, serving to inhibit the DGC activity of WspR and reduce levels of

c-di-GMP to induce biofilm dispersal in response to NO sensation, however as it could be involved in any of the four pathways they all require consideration.

Further complication is added by the post-translational modification of BdlA required for signalling²⁸⁷. Increased DGC activity of GcbA upon the formation of a biofilm induces phosphorylation of Tyr238 within BdlA, by an as yet unknown kinase^{287,292}. Following phosphorylation, BdlA is cleaved between the two PAS domains at Met130, with hydrolysis thought to be catalysed by ClpP (PA0451), separating the haem-binding PAS1 domain from the remaining PAS2-MA protein fragment²⁸⁷. The now isolated PAS1 domain of BdlA then dimerises with the PAS2 domain located in the PAS2-MA fragment²⁸⁷.

To investigate whether binding of NO to the haem cofactor within BdlA PAS1 induces conformational changes, and to identify how BdlA may further signal within the cell we sought to structurally characterise BdlA.

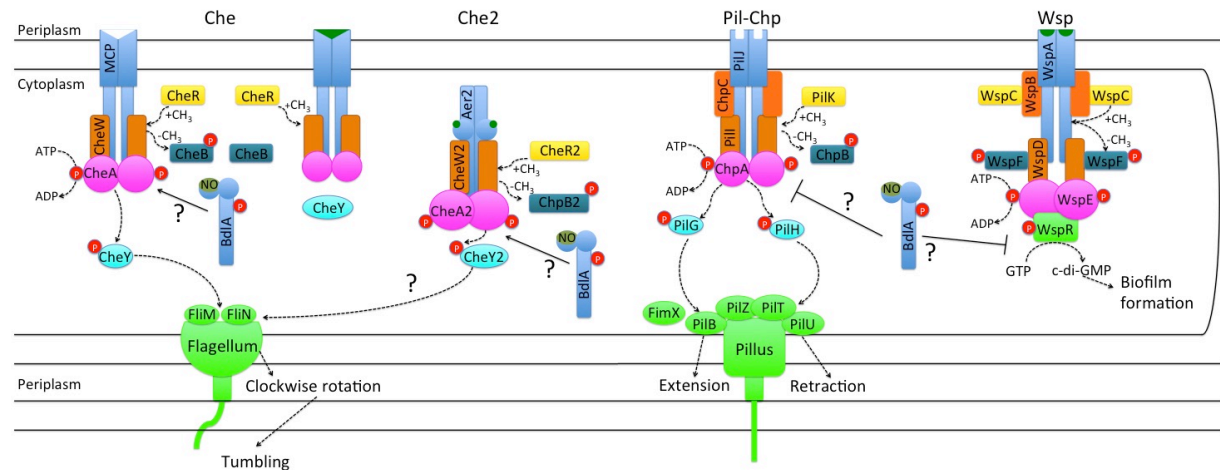


Figure 4.1 - Chemotaxis pathways of *P. aeruginosa*. In the absence of chemotactic attractant (shown in green), CheA (in magenta) autophosphorylates, as well as phosphorylating CheB (dark blue) and CheY (teal). Phosphorylated CheY binds to FliM and FliN (in light green), inducing clockwise rotation of the flagellum and twitching motility. Phosphorylated CheB demethylates the MA domain of the MCP, working in opposition to the methylating activity of CheR (yellow). Substrate binding to the MCP induces a conformational change resulting in reduced phosphorylation activity of CheA, preventing CheY and CheB from being phosphorylated, resulting in predominantly anticlockwise rotation of the flagellum (conductive to swimming activity) and increased methylation of the MA domain. The Che2 pathway is thought to function similarly to Che, although it is not clear whether signal perception by Aer2 results in increased CheA2 autokinase activity and the physiological output of the Che2 pathway in *P. aeruginosa* is also undetermined. Equivalent machinery to that observed in Che mediates pillus extension and retraction in the Pil-Chp pathway. Unlike the other three chemotactic pathways, binding of ligand to WspA increases phosphorylation in the Wsp pathway, resulting in increased c-di-GMP levels and biofilm formation. Homologous proteins are coloured consistently to the Che pathway. Although no involvement for BdlA in any single chemotactic pathway has yet been identified, methods through which BdlA may influence each chemotactic pathway to induce biofilm dispersal upon binding of NO are included.

4.2 Protein purification and optimisation of haem incorporation

As binding of haem-*b* to BdlA is exclusive to the first PAS domain, BdlA⁸⁻¹²⁷ (predicted by the secondary structure predictor PSIPRED to define the first PAS domain^{401,402}) was cloned into *E. coli* BL21(DE3) and expressed, as detailed in 2.1.10 and 2.1.11. Despite expression of soluble protein, purified BdlA⁸⁻¹²⁷ did not display the presence of a Soret peak, characteristic of a bound haem cofactor, when analysed with Ultra-violet visible light (UV-Vis) spectroscopy, as shown in Figure 4.2.

We hypothesised that the lack of haem incorporation may have been due to misfolding or requirement of another portion of the protein for haem loading. To test this theory, the full-length protein (BdlA¹⁻⁴¹⁷) was cloned, expressed and purified using the protocols given in sections 2.1.10, 2.1.11 and 2.1.17. Much like BdlA⁸⁻¹²⁷, UV-Vis spectroscopic analysis of purified BdlA¹⁻⁴¹⁷ did not demonstrate the presence of a haem cofactor Figure 4.2. The absence of a bound haem cofactor from purified BdlA¹⁻⁴¹⁷ provides evidence against protein misfolding being responsible for poor cofactor incorporation or cofactor dissociation.

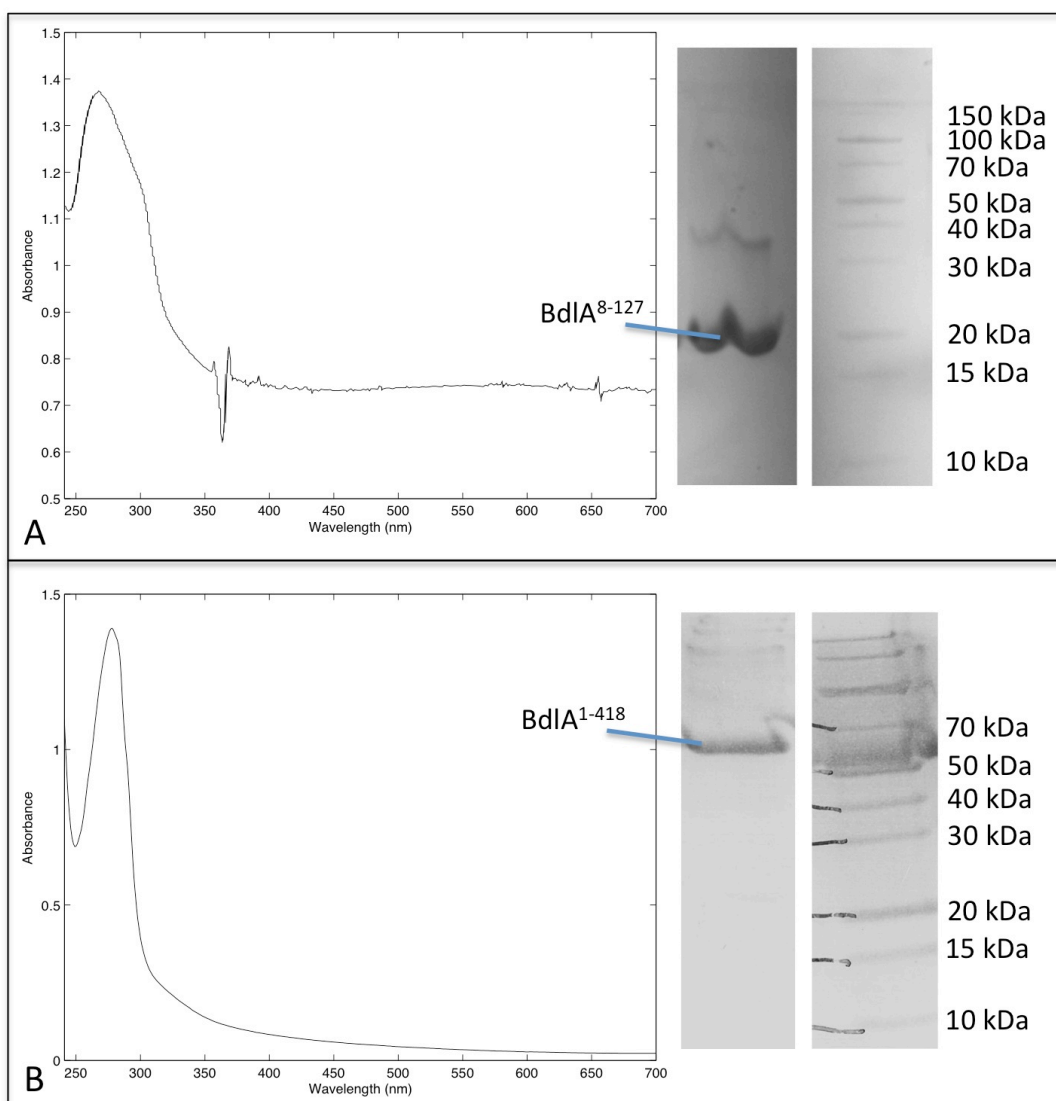


Figure 4.2 - UV-Vis spectra of purified BdlA fragments. A) No Soret peak is observed in UV-Vis spectroscopy of BdlA⁸⁻¹²⁷, suggesting no haem is bound. B) Purified BdlA¹⁻⁴¹⁷ also does not contain bound-haem as shown by UV-Vis spectroscopy through the absence of a Soret-peak.

Previous studies on the haem-binding PAS domain of *E. coli* DOS, have demonstrated purification of haem-bound PAS domains when expression media is supplemented with 0.45 mM δ -aminolevulinic acid (ALA), a haem precursor⁴⁰³. To test whether absence of haem was due to insufficient cellular levels of haem during expression, media was supplemented with 1 mM ALA during expression of BdlA¹⁻⁴¹⁷. However, analysis of the purified protein with UV-Vis spectroscopy again did not identify bound haem cofactor. A similar lack of haem was observed when expression conditions were supplemented with 45 mM ALA and 1.8 mM iron sulphate Figure 4.3.

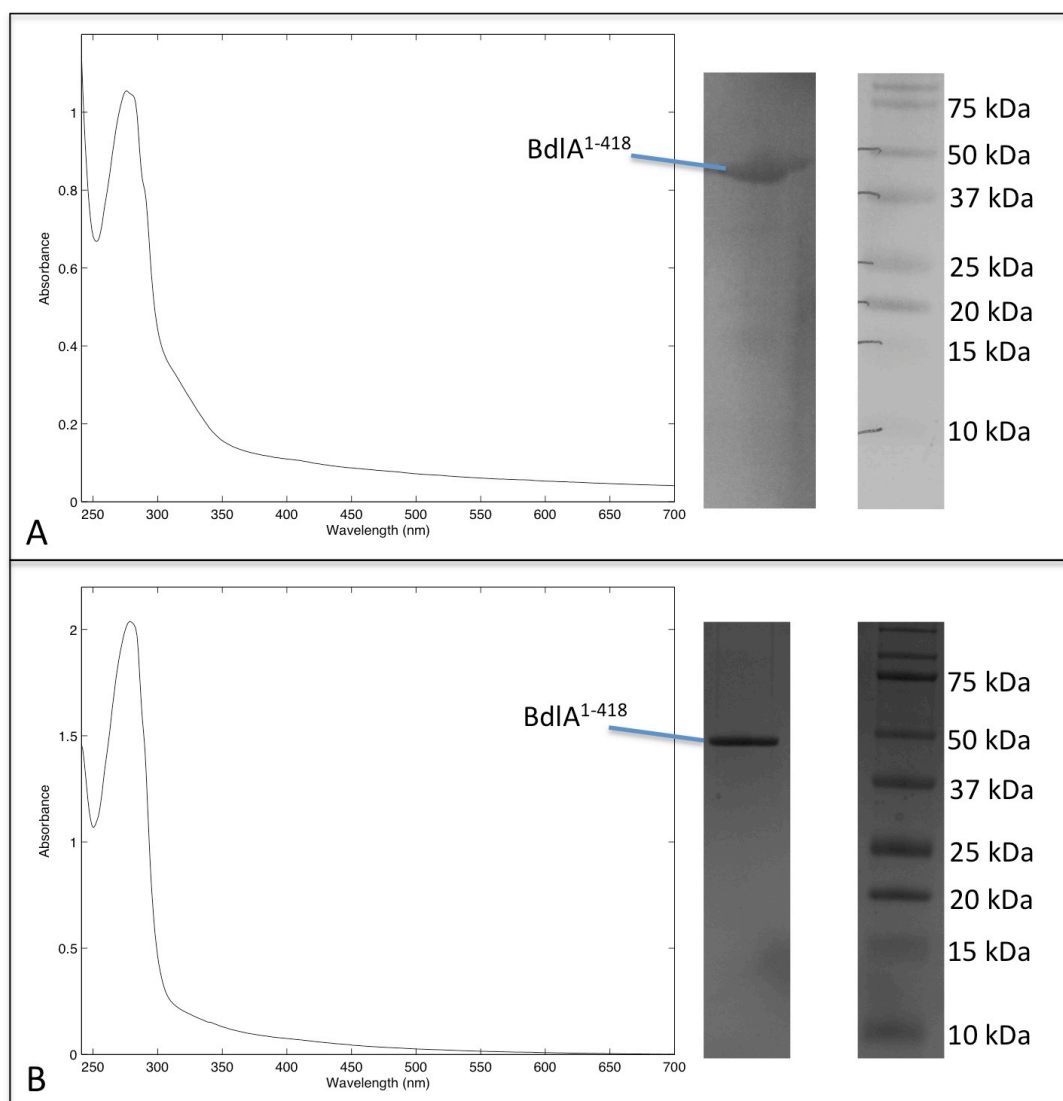


Figure 4.3 - UV-Vis spectra of BdlA¹⁻⁴¹⁷ expressed in the presence of haem precursors. A) Purified BdlA¹⁻⁴¹⁷, expressed in the presence of 1 mM ALA. B) Purified BdlA¹⁻⁴¹⁷, expressed in the presence of 45 mM ALA and 1.8 mM iron sulphate. No Soret peak is observed of either sample when examined with UV-Vis spectroscopy.

An alternative approach to supplementing expression is to add haem to the protein during purification. Following expression, lysis and centrifugation, 20 mg of hemin solubilised in DMSO was added to the BdlA¹⁻⁴¹⁷ lysate before purification through anion exchange chromatography (detailed in 2.1.18) and SEC (using protocols given in 2.1.17). Purified BdlA¹⁻⁴¹⁷ was analysed through SDS-PAGE and UV-Vis spectroscopy, shown in Figure 4.4. As can be seen from Figure 4.4, UV-Vis spectroscopy shows a Soret peak is present at 408 nm that corresponds to protein-bound haem, as is evidenced by the shift in Soret peak in comparison to the hemin reference sample Figure 4.4C. However, even after purification steps, the sample of BdlA¹⁻⁴¹⁷ still retained other proteins, as shown in Figure 4.4A, which may be responsible for the observed haem binding. To obtain pure protein, BdlA¹⁻⁴¹⁷ (expressed in 2 L of LB) was purified through IMAC (according to protocols detailed in 2.1.17) before the addition of 15 mg of hemin (dissolved in DMSO), and subsequent separation of BdlA¹⁻⁴¹⁷ from unbound hemin using anion exchange chromatography. UV-Vis spectroscopy of BdlA¹⁻⁴¹⁷ purified in this method demonstrates a Soret peak consistent with protein-bound haem, while analysis through SDS-PAGE shows only the presence of BdlA¹⁻⁴¹⁷ Figure 4.4B.

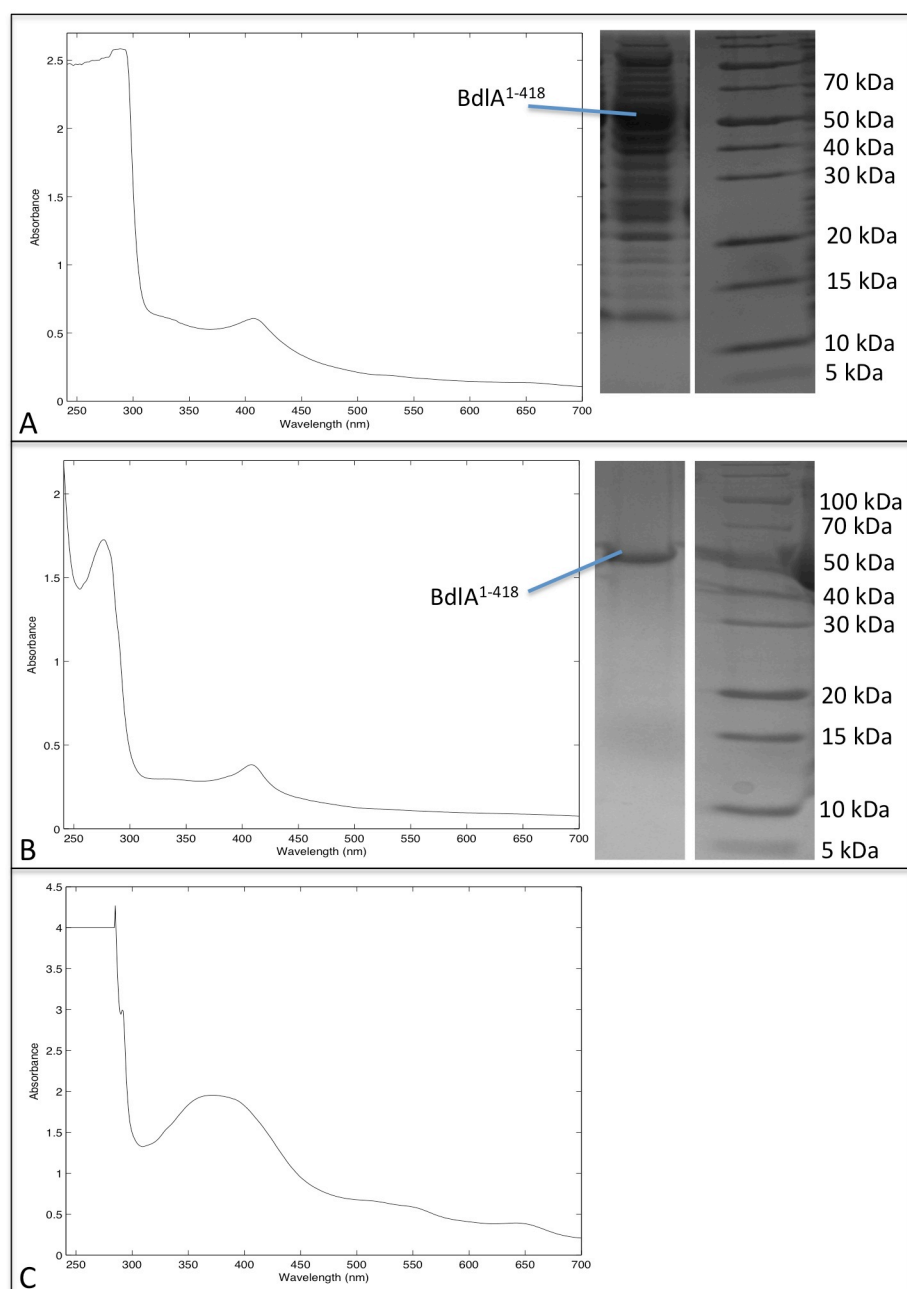


Figure 4.4 - UV-Vis spectra of haem-bound BdIA. A) BdIA soaked with hemin, purified by anion exchange chromatography and gel filtration. A Soret peak at a wavelength of around 408 nm is observable in the UV-Vis spectra, however the tested sample contains numerous proteins other than BdIA¹⁻⁴¹⁷, as shown by SDS-PAGE. B) BdIA soaked with hemin, purified by IMAC and anion exchange chromatography. A similar Soret peak is observed as that in A, suggesting bound haem and SDS-PAGE demonstrates this sample contains far fewer other proteins. C) Reference UV-Vis spectrum of 5 μM hemin. The crest of the Soret peak is around 380 nm. The shift in wavelength of the Soret peak between the reference hemin sample and the purified BdIA protein samples suggests a difference in porphyrin ring binding, likely to be due to hemin being protein-bound in the protein samples and in solution in the reference sample.

Through application of the Beer-Lambert law ($A = \epsilon cd$, where A is the measured absorbance, ϵ is the molar extinction coefficient, c is the concentration of the sample and d is the pathlength) to observed UV-Vis spectra, molecules that absorb light can be quantified. As the bound haem and BdlA¹⁻⁴¹⁷ have distinct absorption peaks within UV-Vis spectra, corresponding to the Soret peak at 408 nm and absorbance of tryptophan at 280 nm respectively, it is possible to quantify both of these species and determine stoichiometry of the purified complex. Following purification through IMAC and anion exchange chromatography (shown in Figure 4.5), BdlA¹⁻⁴¹⁷ is present at a concentration 41 μ M, however only 6 μ M haem is present, a BdlA:haem stoichiometry of roughly 6:1. If each copy of the first BdlA PAS domain were to bind a haem cofactor, a 1:1 stoichiometry of purified protein to haem would be expected. Therefore the protocol for purification of haem-bound BdlA¹⁻⁴¹⁷ was further optimised by increasing the time hemin was incubated with IMAC purified BdlA¹⁻⁴¹⁷ (from an hour to overnight), with subsequent removal of unbound hemin through SEC. Using this protocol, 48 μ M BdlA¹⁻⁴¹⁷ was purified bound to 45 μ M haem, providing an almost 1:1 stoichiometry. A comparison between the two protocols is shown in Figure 4.5, with the optimised protocol used for subsequent purifications.

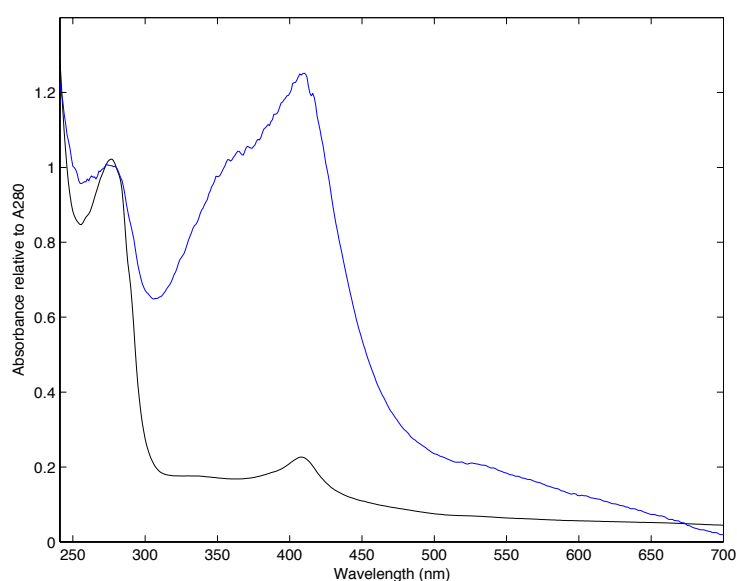


Figure 4.5 - UV-Vis spectra of BdlA¹⁻⁴¹⁷, scaled by absorbance at 280 nm, showing improved haem incorporation. BdlA¹⁻⁴¹⁷ purified through IMAC, a 1 hour hemin soak and anion exchange chromatography is shown in black. BdlA¹⁻⁴¹⁷ purified through IMAC, an overnight hemin soak and SEC is shown in blue. The haem Soret peak observed at 408 nm is far larger in relation to the 280 nm tryptophan peak when BdlA¹⁻⁴¹⁷ is purified using the optimised purification protocol.

4.3 Haem-containing BdlA interacts with NO

The haem-binding PAS domain of BdlA is thought to be cleaved from the rest of the protein during biofilm formation. To investigate whether direct interaction between NO and the haem-PAS domain of BdlA is possible and to determine a function in biofilm dispersal, a fragment of the haem-binding PAS domain of BdlA (BdlA¹⁻¹²⁷) was cloned, expressed and purified according to section 2.1.17. It is worth noting that although the purification protocol optimised for haem incorporation gave an almost 1:1 BdlA¹⁻¹²⁷:haem stoichiometry, application of the same protocol resulted in an approximately 5:4 BdlA¹⁻¹²⁷:haem stoichiometry.

UV-VIS spectroscopy can be used to probe alterations in the state of haem cofactors. Within the central iron atom and the porphyrin ring of the haem, d-electrons and π -electrons can absorb radiation in the range of UV and visible light, exciting them to a higher energy state. Photons of light are only absorbed if they are of an equivalent energy to that required for shifting electrons to an excited state. As only certain wavelengths of the light exposed to the sample will fulfil this criteria, specific absorption spectra can be observed and the observed absorption profile can provide information on the arrangement of electrons⁴⁰⁴.

In the absence of an NO donor, BdlA¹⁻¹²⁷ displays a Soret peak at 404 nm. However, incubation with the NO donor DEA NONOate results in a shift of the position of this peak to 383 nm, indicating that the arrangement of electrons within the haem cofactor has been altered, as shown in Figure 4.6. The observed changes in spectroscopic peak suggests that NO is directly interacting with the bound haem cofactor. The observed shift in Soret peak position is consistent with observations in cytochrome c' of the transition between a ferric iron and a ferrous, NO bound iron in the haem cofactor^{405,406}.

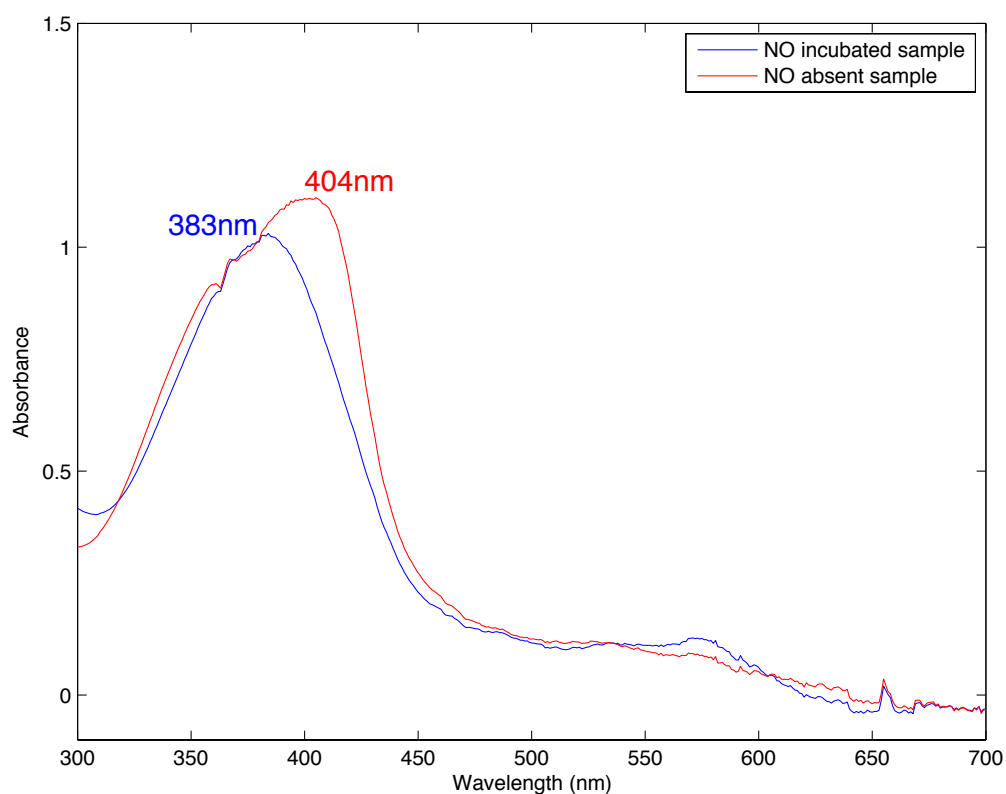


Figure 4.6 - UV-Vis spectra of BdlA¹⁻¹²⁷ in the presence and absence of NO. The Soret peak (labelled according to the wavelength of the peak), shifts in response to addition of NO.

To further confirm that observed shifts in UV-Vis spectra were caused by direct haem-NO interaction, BdlA¹⁻¹²⁷ was incubated with the NO donor MAHMA NONOate before study through Fourier Transform Infrared (FTIR) spectroscopy. During FTIR spectroscopy the sample is exposed to a range of infrared radiation, with wavelengths capable of exciting vibrations between atoms within the sample selectively absorbed. As the nature and rate of bond vibration is dependent upon the environment of the atom, including the nature of bonds joining it to other atoms, the bonds joining different atoms absorb different wavelengths of IR radiation. FTIR spectroscopy therefore makes it possible to identify bonds specific to two defined atoms if their environment is sufficiently unique within the sample⁴⁰⁷.

FTIR spectra of BdlA¹⁻¹²⁷ in the presence and absence of MAHMA NONOate are shown in Figure 4.7. As every bond within a sample examined with FTIR contributes to the observed spectra, bonds that are commonplace throughout the sample represent a large proportion of the observed spectra. For example, vibrations within the solvent molecule deuterium oxide are responsible for the signal-saturating peak present at approximately 2200-2700 cm⁻¹ observed in Figure 4.7. In comparison, only a relatively small number of bonds between haem and NO would be expected within the sample, making it therefore necessary to examine differences in FTIR

spectra by normalising spectra on a common peak (in this case the peak contributed by vibrations of water, at approximately 3412 cm^{-1}) and subtracting the spectra of the NO absent sample from that of the NO present sample. The resultant difference spectra are shown in Figure 4.7. Within the difference spectra, peaks can be observed at 1462 cm^{-1} , 1623 cm^{-1} , 1652 cm^{-1} , 2873 cm^{-1} and 2947 cm^{-1} . Peaks observed at 1462 cm^{-1} , 2873 cm^{-1} and 2947 cm^{-1} are also observed when an equivalent concentration of MAHMA NONOate was added to deuterated buffer E (detailed in section 2.1.20), suggesting these peaks do not represent protein specific interactions. However, peaks observed at 1623 cm^{-1} and 1652 cm^{-1} are unique to BdlA¹⁻¹²⁷ in the presence of MAHMA NONOate (shown in Figure 4.7), suggesting these are likely to be NO specific interactions. The peak observed at 1623 cm^{-1} is consistent with previously identified interactions between Fe^{2+} and NO^{408} , but the nature of the interaction contributing the peak at 1652 cm^{-1} is currently unknown. It is worth highlighting that the peak observed at 1623 cm^{-1} is consistent with interactions between NO and Fe^{2+} , suggesting that binding of NO results in reduction of the haem iron. It is possible that the peak observed at 1652 cm^{-1} is contributed by a second binding mode of NO or from a conformational change induced upon NO binding.

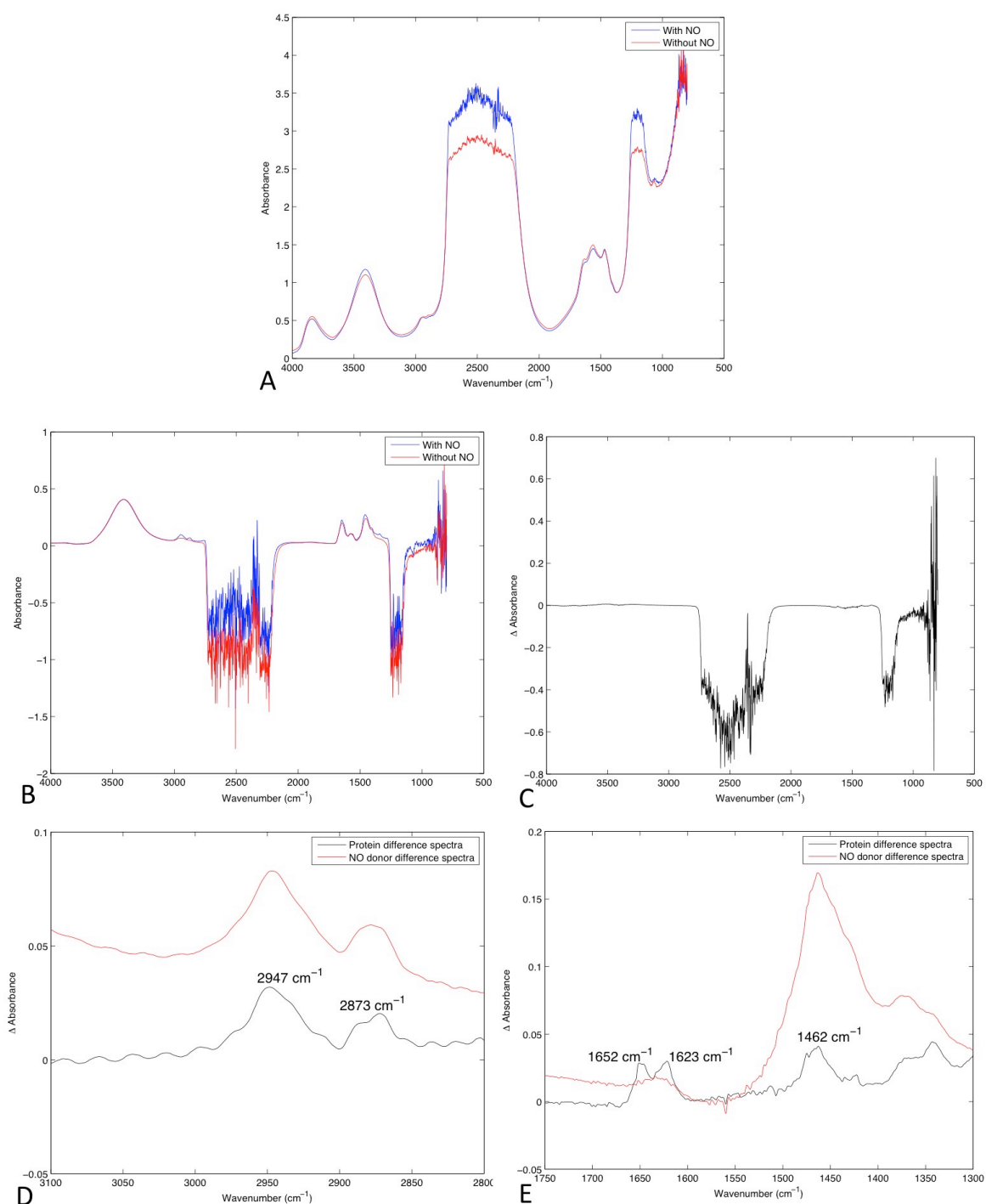


Figure 4.7 - FTIR spectra of interactions between BdIA¹⁻¹²⁷ and NO. A) FTIR spectra of BdIA¹⁻¹²⁷ in the presence and absence of NO without any further processing. B) FTIR spectra of BdIA¹⁻¹²⁷ in the presence and absence of NO with the contribution of deuterated buffer E subtracted and scaled according to the water peak 3412 cm⁻¹. C) Difference spectra generated by subtracting the NO absent sample in B from the NO present sample. D and E) Zoomed view of the difference spectra shown in C showing NO dependant peaks at 1462 cm⁻¹, 1623 cm⁻¹, 1652 cm⁻¹, 2873 cm⁻¹ and 2947 cm⁻¹. The FTIR spectra of NO donor in deuterated buffer E is also shown, identifying peaks not specific to interaction between BdIA¹⁻¹²⁷ and NO.

To examine whether the nature of NO interaction with BdlA¹⁻¹²⁷ changed over time, spectra were acquired every subsequent 10 minutes for a further 140 minutes, shown in Figure 4.8. No differences in non-saturated spectral regions were observed between the first and last spectra, suggesting that interaction between NO and BdlA¹⁻¹²⁷ does not change over this time period.

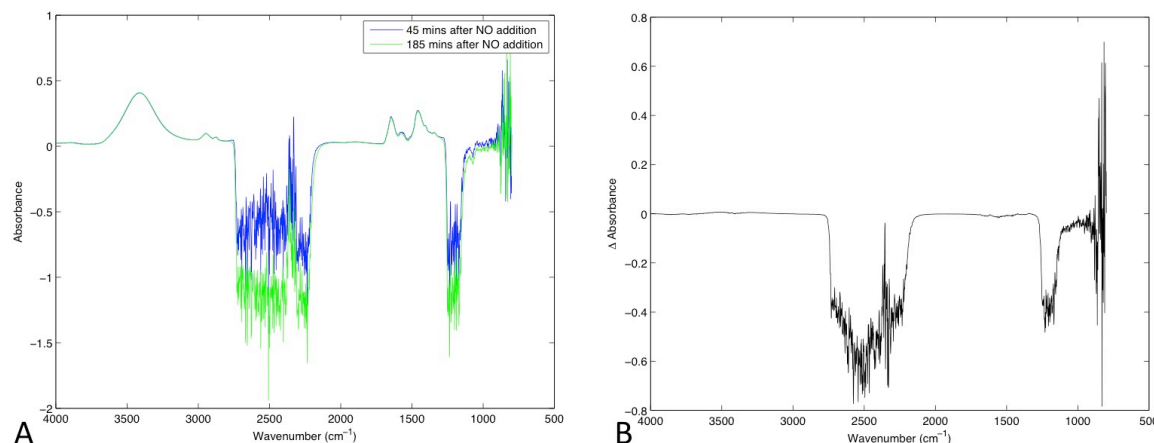


Figure 4.8 - FTIR spectra of interactions between BdlA¹⁻¹²⁷ and NO over time. A) Buffer subtracted and scaled spectra of BdlA¹⁻¹²⁷ at the extremes of the time-course used to determine interactions BdlA¹⁻¹²⁷ and NO. B) Difference spectra obtained by subtracting the spectra obtained 45 minutes after NO addition from that obtained 185 minutes from NO addition.

Having determined that BdlA¹⁻¹²⁷ is able to bind NO, BdlA was subjected to structural studies to examine the nature of conformational changes induced upon NO binding.

4.4 Crystallisation of BdlA

Despite successful copurification of BdlA¹⁻¹²⁷ and BdlA¹⁻⁴¹⁷ with bound hemin, crystallisation trials of these complexes proved unsuccessful. However, crystals of BdlA¹⁻⁴¹⁷ were obtained of the protein in the apo-state. Crystals grew in a range of conditions (detailed in section 2.1.21), and initially diffracted to 4 Å. Attempts to optimise away from the manufacturer provided crystallisation conditions were also unsuccessful, so Morpheus Custom Optimisation Hanging Drop Screen was established, which yielded crystals (shown in Figure 4.9) in condition D5 that diffracted to 2.6 Å at Diamond Light Source I03 on 09/09/2016, with a wavelength of 0.9794 Å. The structure of BdlA¹⁻⁴¹⁷ was solved through single wavelength anomalous dispersion using a heavy metal soak with thimerosal. Anomalous data was collected on I04 at Diamond Light Source using a wavelength of 0.9919 Å on 08/08/2016 to a resolution of 4 Å. Images were indexed and integrated with XDS³³⁹, before scaling and the removal of 5% of data for cross validation using AIMLESS³⁴¹. Starting phases were obtained using the Crank2 software pipeline⁴⁰⁹, consisting of

phase estimation using SHELXC and SHELXD⁴¹⁰, density modification with Parrot⁴¹¹, automated model building with Buccaneer⁴¹² and refinement with REFMAC³⁴⁴. Model building was then performed in Coot, with subsequent refinement using REFMAC^{343,344}. Due to significant anisotropy within the data, the native dataset (indexed and integrated using the xia2 3dii pipeline and scaled with AIMLESS)^{317,337–339,341} was processed using the STARANISO server⁴¹³. The model generated from the anomalous dataset was then used as a molecular replacement solution for the native data (using the MOLREP package), before subsequent model building with Coot and refinement with BUSTER^{317,343,414}. Data collection and refinement statistics are given in Table 4.1. It is worth noting that data scaled without a resolution cut off was submitted to the STARANISO server (in accordance with the instructions provided), with STARANISO thereafter defining the resolution cut off. The high-resolution cut-off chosen by STARANISO does not conform to convention of cutting data to only predominantly complete resolution shells, as evidenced by the selection of a high resolution cut-off with only 2.4% completeness. This resolution cut-off also reduces the average completeness of the data and affects the calculated R_{pim} , as can be seen from Table 4.2 which gives data collection statistics for the native dataset according to resolution shells. Further refinement of the generated BdIA¹⁻⁴¹⁷ model is required to improve the observed geometries, however it is possible to draw conclusions from the model presented hereafter.

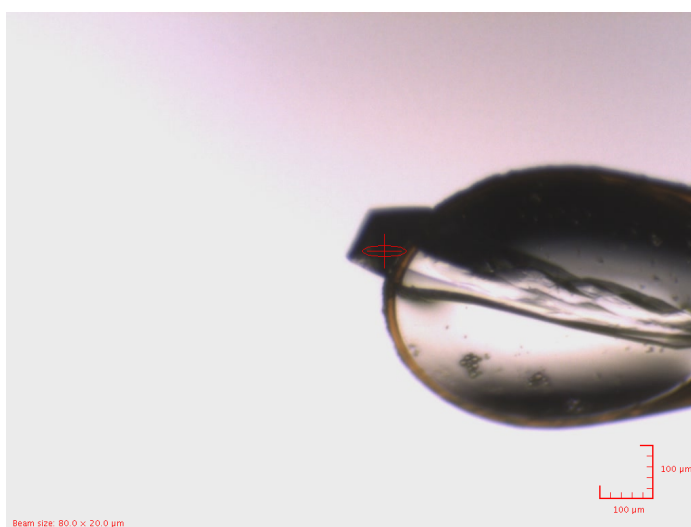


Figure 4.9 - A crystal of BdIA¹⁻⁴¹⁷ diffracted at Diamond Light Source I03.

Table 4.1 - Crystallographic data and refinement statistics for BdIA¹⁻⁴¹⁷.

BdIA ¹⁻⁴¹⁷ dataset	Anomalous	Native
Data Collection		
Space Group	C2	C2
Cell Axes (Å)	a = 150.08 b = 49.68 c = 140.00	a = 148.2 b = 49.29 c = 140.67
Angles (°)	$\alpha = \gamma = 90$ $\beta = 99.54$	$\alpha = \gamma = 90$ $\beta = 98.33$
Beamline	I04	I03
Wavelength (Å)	0.9919	0.9794
Resolution (Å)	47.09-4.00 (4.47-4.00)	73.32-2.50 (2.6-2.50)
Unique Reflections (#)	8904 (2499)	22699 (166)
Measured Reflections	116809 (33687)	42398 (101)
Redundancy	13.1 (13.5)	1.9 (1.6)
R _{pim} (%)	1.5 (2.8)	3.3 (59.1)
I/ σ (I)	35.6 (22.8)	12.0 (0.8)
CC _{1/2}	0.99 (0.99)	0.99 (0.37)
Completeness (%)	99.8 (99.9)	63.3 (2.4)
Anomalous completeness (%)	99.8 (99.9)	-
Anomalous multiplicity	6.9 (7.0)	-
Refinement		
Molecules/AU	2	2
R _{work} /R _{free} (%)		0.2064/0.2713
Rmsd		
Bond Length (Å)		0.009
Bond Angles (°)		1.17
Average Protein B Factor (Å ²)		106.5
Ramachandran (%)		
Favoured Regions		93.9
Allowed Regions		4.52
Outliers		1.6

The crystal structure of BdlA is shown in Figure 4.10. Within the crystal, BdlA¹⁻⁴¹⁷ is observed in a linear conformation, with both PAS domains in-line with the MA domain. In this linear conformation BdlA¹⁻⁴¹⁷ is approximately 213 Å long, with the majority of this length contributed by the 138 Å long MA domain. Due to the high average of B-factor of the structure, the crystal structure of BdlA¹⁻⁴¹⁷ is also shown in Figure 4.11, coloured by B-factor, highlighting the regions of greatest motion within the crystal. From Figure 4.11 it can be observed that the central β -sheets in each PAS domain have the lowest B-factors, identifying these regions as being the most consistently ordered throughout the crystal. Far greater disorder is observed of a number of loops within the first PAS domain of the A chain, both PAS domains in the B chain and the inter-helix loops within the MA domain in both chains, unsurprisingly identifying these regions as highly flexible. With the exception of differences in the loops of the PAS 2 domain, both models in the asymmetric unit display a similar distribution of B-factors.

Table 4.2 – Crystallographic data statistics for the BdIA¹⁻⁴¹⁷ native dataset, divided into resolution shells.

Resolution (Å)	Measured reflections	Unique reflections	Shell completeness (%)	Cumulative completeness (%)	Multiplicity	Rpim (%)	Mn(I/sd)	CC _{1/2}
12.72	1455	833	99.2	99.2	1.7	1.6	34.6	0.999
7.35	2618	1432	100	99.7	1.8	2.4	27.7	0.998
5.69	3388	1826	100	99.8	1.9	2.7	21.3	0.996
4.81	4031	2151	100	99.9	1.9	2.8	22.6	0.997
4.24	4541	2417	100	99.9	1.9	3.7	17.6	0.996
3.84	5084	2688	99.9	99.9	1.9	6	11.2	0.988
3.53	5400	2875	99.9	99.9	1.9	10.5	6.7	0.974
3.29	5830	3088	99.2	99.8	1.9	19.4	4	0.936
3.09	4945	2599	79.6	96.6	1.9	28.9	2.7	0.891
2.92	2694	1448	40.9	88.4	1.9	34.5	2.3	0.809
2.78	1473	804	21.6	79.5	1.8	42	2	0.629
2.65	773	437	11.1	71.2	1.8	64.2	1.2	0.292
2.54	166	101	2.4	63.3	1.6	59.1	0.8	0.369
Overall	42398	22699	63.3	63.3	1.9	3.3	12	0.999

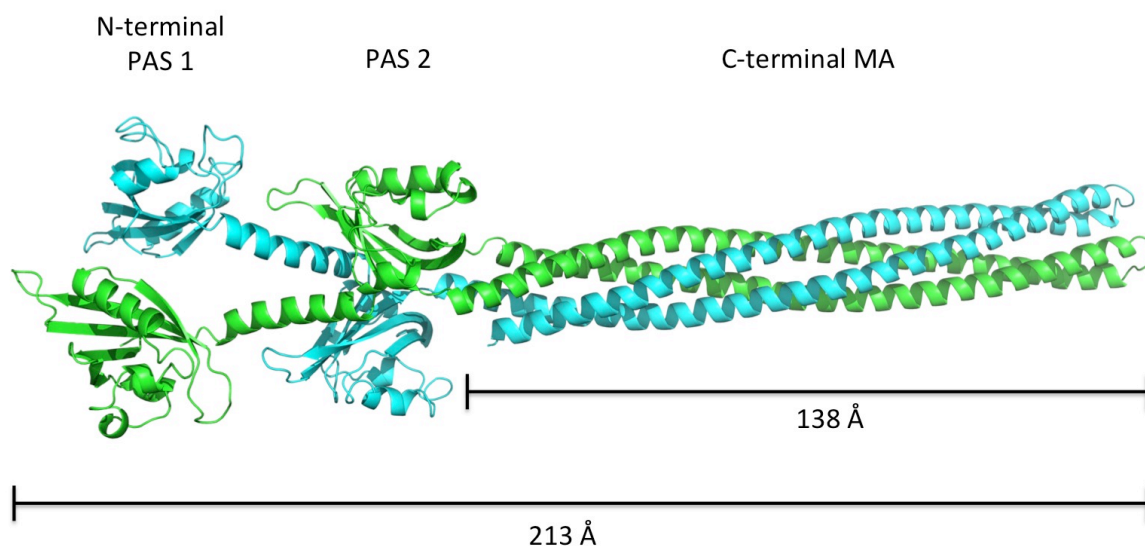


Figure 4.10 - The structure of BdIA¹⁻⁴¹⁷.

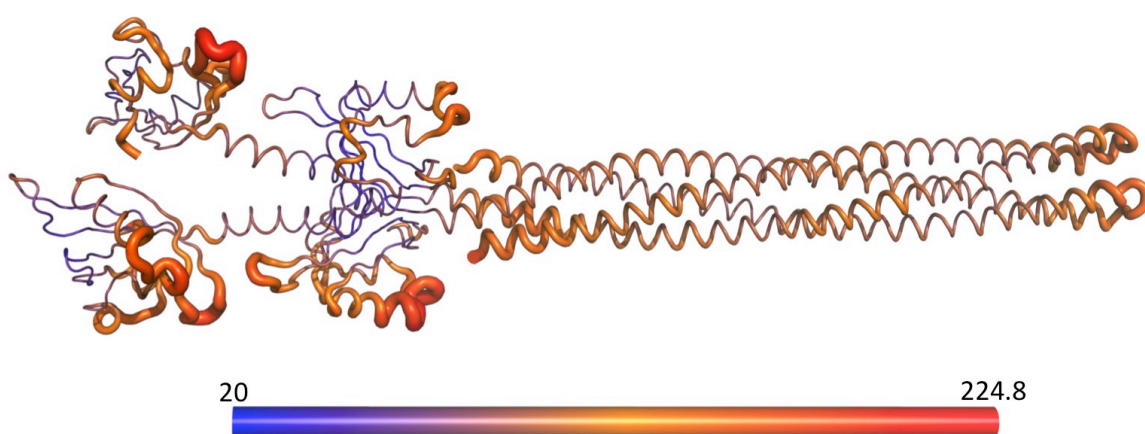


Figure 4.11 – BdIA¹⁻⁴¹⁷ coloured by B-factor and oriented similarly to Figure 4.10. BdIA¹⁻⁴¹⁷ is modelled as “putty” and coloured in a continuous spectrum of blue to red (in accordance with the scale bar), with higher B-factor regions highlighted in red as a thicker “putty”.

The observed linear conformation of BdIA¹⁻⁴¹⁷ facilitates the formation of a dimer, with a dimerisation interface consisting of intercoiled MA domains and the PAS 2 domain of BdIA. In this dimer the distance between BdIA PAS 2 domains is roughly equivalent to that observed between PAS domains in DosP, suggesting that the BdIA PAS 2 dimer may be appropriately positioned to function in signal transduction. The distance between BdIA PAS 1 domains is far greater (15.7 Å between β -sheets of each monomer, rather than the 8.6 Å observed between the β sheets in PAS 2), suggesting that in this conformation there is no interaction between the PAS 1 present in each monomer.

Phosphorylation of BdlA at Tyr238 and subsequent cleavage at Met130 is required for nutrient induced biofilm dispersal and is thought to be also required for NO-induced biofilm dispersal²⁸⁷. From Figure 4.12 it can be seen that Tyr238 is located at the top of the MA domain, and is ideally situated to stabilise interaction between the α helices that constitute the MA domain. It is also possible that from this location Tyr238 could serve to stabilise dimerisation between MA domains. The addition of a bulky, negatively charged phosphate group is likely to dramatically alter the nature of interactions possible for Tyr238. Phosphorylation may promote a dimer stabilising interaction between Tyr238 of one monomer and Arg234 of the other, but it may also prevent Tyr238 from interacting with the C-terminus of BdlA¹⁻⁴¹⁷ and could result in destabilisation of the MA domain dimer, or push apart the two MA domain helices.

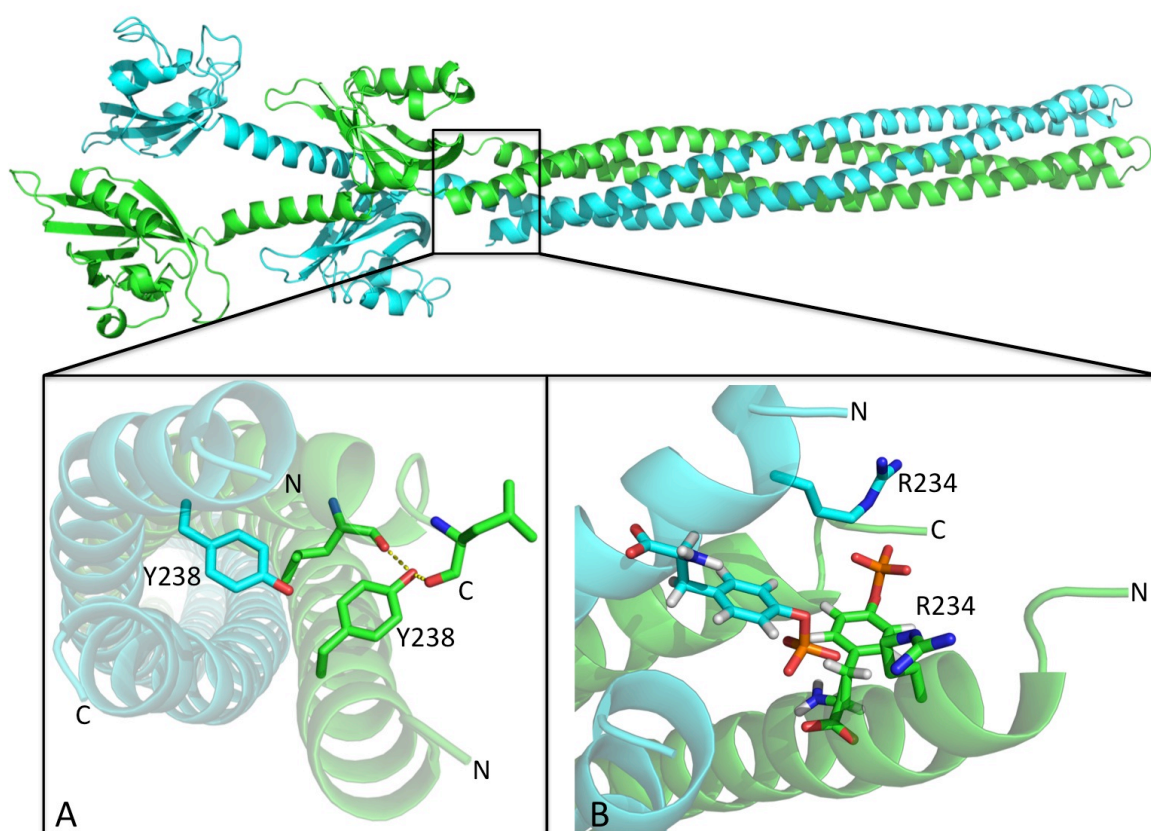


Figure 4.12 - Phosphorylation of Tyr238 is likely to alter interaction within the MA domain. A) In the observed structure of BdlA¹⁻⁴¹⁷, Tyr238 is suitably positioned to interact with the second helix of the MA domain, an interaction that may be able to stabilise MA dimerisation.

B) Phosphorylation of Tyr238 (modelled by overlaying a small molecule phosphotyrosine, with three letter code PTR, to the observed side chain of Tyr238) could position the phosphate to interact with Arg234, but the apparent steric clashes induced by phosphorylation of Tyr238 highlight the requirement for other conformational changes.

Conformational change upon phosphorylation of Tyr238 could result in exposing Met130, which is buried in the observed BdlA structure, as shown in Figure 4.13, facilitating cleavage. Exposure of Met130 could be achieved by altering the position of PAS 2, with the altered MA domain interaction either pushing PAS 2 away or pulling it in closer, or through the dissociation of the BdlA dimer. Alternatively, as MA domain containing proteins are often found co-localised³⁷⁶, it is possible that reduced dimerisation between MA domains could result in dissociation of BdlA from a cluster of MA domain containing proteins, exposing the Met130 site for cleavage by ClpP (PA0451).

Following cleavage, BdlA PAS 1 is thought to reassociate with the remaining fragment of BdlA in an interaction believed to be mediated through PAS domain dimerisation²⁸⁷. PAS domain dimerisation has been mostly commonly observed through interaction between the external surface of the β -sheet²²⁸. In the observed structure of BdlA¹⁻⁴¹⁷, the PAS 2 domains from each monomer form such a dimer. Any PAS domain dimer formed between cleaved PAS 1 and the PAS 2-MA domain fragment would therefore have to take place through an unusual PAS domain dimerisation interface, or would require dissociation of the observed dimer. A model of dimer dissociation upon phosphorylation of Tyr238 could leave PAS 2 suitably positioned to interact with the now cleaved PAS 1 and accommodate further interaction between the MA domain and the α -helix previously connecting PAS 1 to PAS 2, shown in Figure 4.13C. BdlA is thought to remain in this reassociated state until perception of a biofilm dispersal signal.

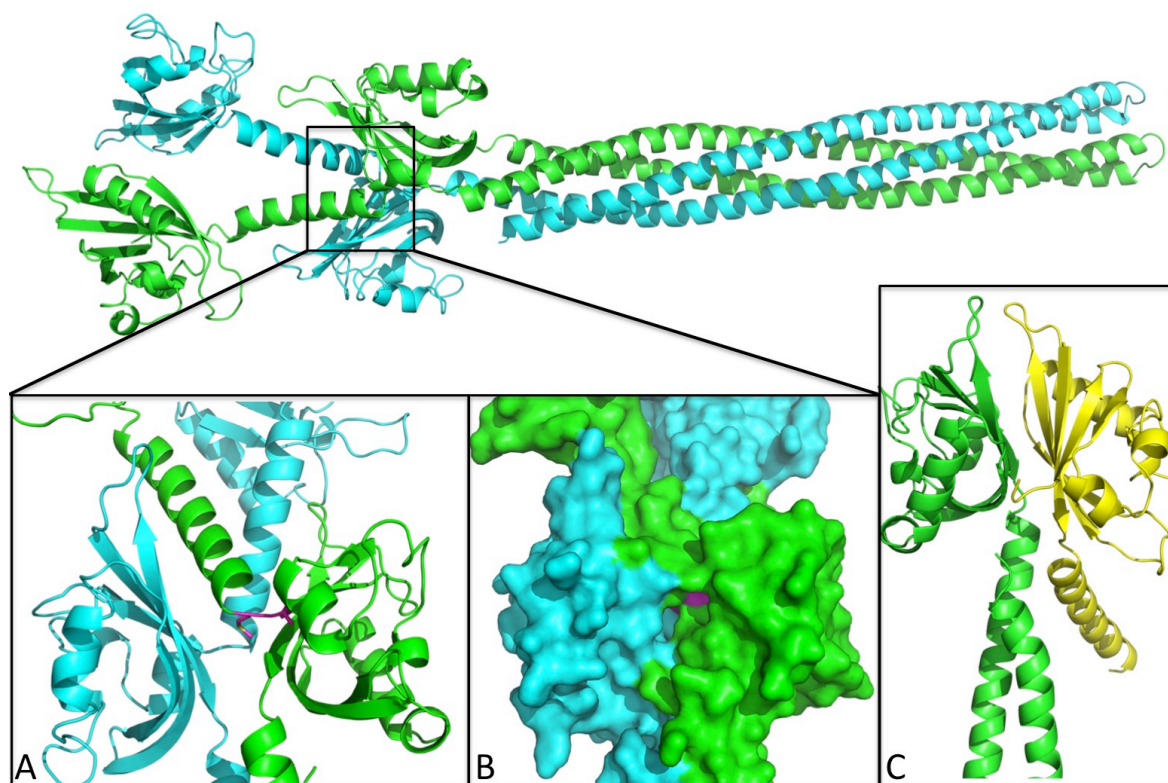


Figure 4.13 - The BdIA cleavage site. A) The BdIA cleavage site, Met130-Ala131 and highlighted in magenta, is located at the N-terminus of PAS 2, at the end of the α -helix between PAS 1 and PAS 2. B) Surface representation of A, demonstrating that within the observed dimeric conformation the BdIA cleavage site is likely to be inaccessible to proteases. C) Proposed PAS dimer interaction to allow reassociation of the cleaved PAS 1 fragment (shown in yellow) to the remaining fragment of BdIA.

Although BdIA¹⁻⁴¹⁷ was not crystallised with a bound haem cofactor, it may be possible to infer possible modes of haem binding through superposition with structurally characterised haem bound PAS domains. The *E. coli* Dos haem PAS domain (PDB 1V9Y) shares 17% sequence identity to BdIA PAS 1 and an ideal superposition of the two structures has equivalent C α s positioned with an RMSD of 2.26 Å, as determined using the PDBeFold server³⁴⁶. From structure superposition of the *E. coli* Dos haem PAS domain (1V9Y) with BdIA PAS1, shown in Figure 4.14, it can be observed that His42, located on the α 3 helix, would be well positioned to act as an axial ligand for haem coordination. Within the haem binding PAS domains of *E. coli* Dos and FixL proteins from *B. japonicum* and *S. meliloti* axial haem coordination is provided through a histidine present on the α 4 helix^{116,250,415}. However, a model of haem coordination through a histidine located within the α 3 helix is supported by the structure of the *P. aeruginosa* Aer2 haem-bound PAS domain, where axial haem coordination is provided by a similarly positioned histidine. The *P. aeruginosa* Aer2 PAS domain (PDB 4HI4) shares 14% sequence identity with BdIA PAS 1 and equivalent C α s are positioned with an RMSD of 2.0 Å, as determined using the PDBeFold server³⁴⁶.

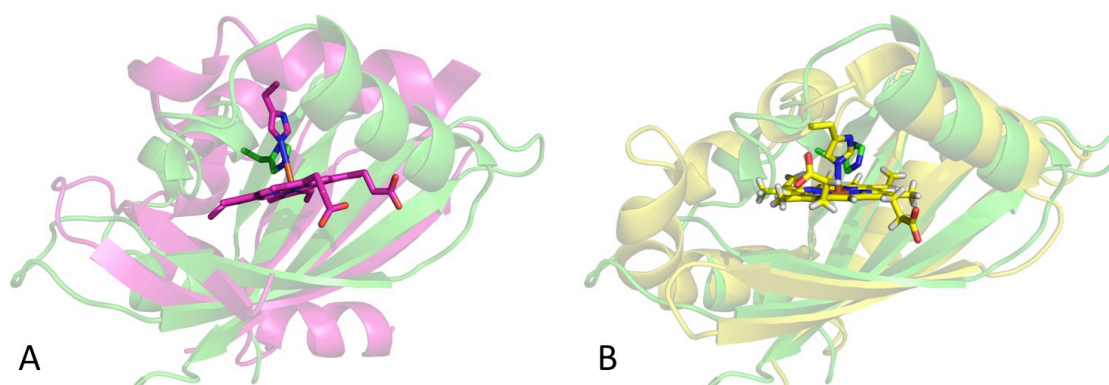


Figure 4.14 - His42 may act as the axial ligand for haem coordination in PAS 1. A) Structural superposition of Dos (PDB 1V9Y, magenta) to PAS 1 of BdlA (green). B) Structural superposition of Aer2 (PDB 4HI4, yellow) with BdlA PAS 1. BdlA His42 and haem axial histidines are shown as sticks.

Given the cytosolic location of BdlA within the cell, and the presence of reducing agents throughout protein purification, both monomers of BdlA PAS 1 were surprisingly observed to contain a disulphide bond between Cys46 and Cys73, shown in Figure 4.15. The presence of such a disulphide bond within a cytosolic protein suggests that it may play a sensory role, classifying it as an allosteric disulphide bond⁴¹⁶. Allosteric disulphide bonds are commonly found in the -RHstaple conformation, similar to that observed between Cys46 and Cys73⁴¹⁶. Adoption of -RHstaple geometry by disulphide bonds results in high values of dihedral strain energy (on average 18.1 kJ mol^{-1} , in contrast to the 14.8 kJ mol^{-1} observed across all disulphides⁴¹⁶), with these high-energy disulphide bonds more easily cleaved than those with a lower energy⁴¹⁶⁻⁴¹⁸. When quantified using the disulphide bond dihedral angle energy server, an average disulphide strain energy of 48.8 kJ mol^{-1} is observed of the disulphide bond between Cys46 and Cys73, further suggesting that this may be an allosteric disulphide bond^{416,419}. Although further investigation is required to identify what role such an allosteric disulphide bond plays within BdlA, cysteine-NO interaction has previously been observed to complement haem-NO interaction in regulation of *Homo sapiens* soluble guanylate cyclase⁴²⁰. Binding of NO to the haem within soluble guanylate cyclase increases activity 10 fold, whilst binding of NO to the haem and to cysteines increases activity 200 fold⁴²⁰. It is possible that cleavage of this disulphide bond accompanies NO binding to BdlA and that one of the cysteines could directly interact with NO. An alternative possibility is that binding of NO to the haem distal site repositions the $\alpha 3$ helix, which contain both the predicted axial haem coordinating His42 and Cys46, resulting in cleavage of the allosteric disulphide bond. Cleavage of this disulphide bond could allow repositioning of Cys73, which is located at the end of the $\beta 3$ -strand and well situated to couple disulphide bond cleavage with PAS domain dimerisation across the β -sheet.

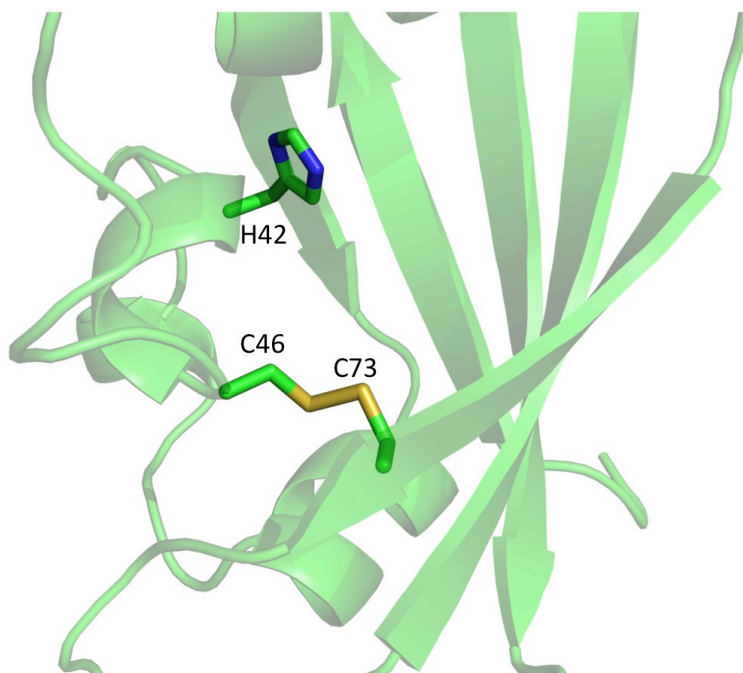


Figure 4.15 - Disulphide bond within BdlA PAS 1. A disulphide bond is formed between Cys46 and Cys73. As Cys46 is located in close proximity to His42, which may serve as the axial haem coordinating ligand, it is possible that conformational changes upon ligand binding to the haem result in cleavage of this disulphide bond.

As BdlA does not contain c-di-GMP metabolising domains, it is unclear how conformational changes within BdlA upon NO binding are propagated to induce biofilm dispersal. It has previously been suggested that once BdlA detects a dispersal stimulus it then goes on to interact with c-di-GMP metabolising enzymes that also contain PAS domains, such as DipA^{287–290}. Levels of c-di-GMP could then be reduced by promoting dimerisation of PDEs or inhibiting dimerisation of DGCs. In this model it is unclear whether PAS 1 or PAS 2, or both, would contact these c-di-GMP metabolising enzymes. Binding of a BdlA PAS domain to a PAS domain within a c-di-GMP catabolising enzyme could serve to increase PDE activity if binding resulted in the formation of higher order oligomers, such as trimers or tetramers, with increased proximity between PDE domains likely to result in increased formation of catalytic PDE dimers. This theory is supported by studies showing that Gly31Ala mutation in BdlA PAS1 results in increased interaction with BdlA and a hyperdispersive phenotype²⁸⁸. Gly31 is located at the end of the α 1 helix in BdlA PAS 1, so is far removed from the typical PAS domain dimerisation interface, but may be involved in the formation of PAS domain tetramers, as shown in Figure 4.16.

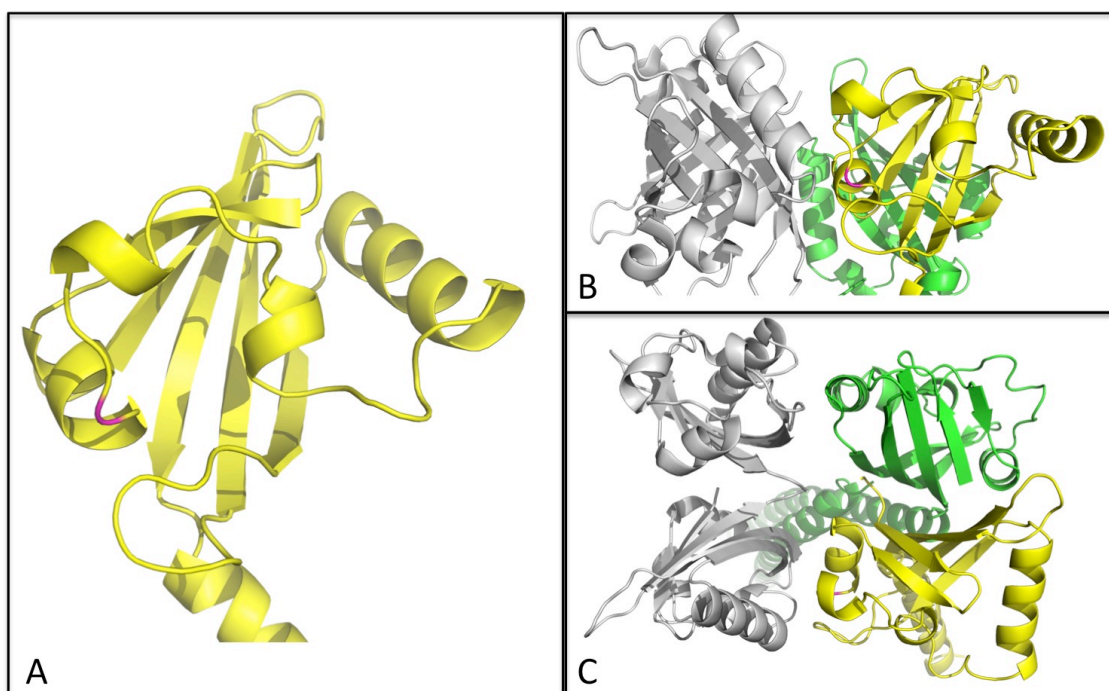


Figure 4.16 - Gly31 may be involved in the formation of PAS tetramers. A) Gly31 (highlighted in magenta) is located at the end of the $\alpha 1$ helix in BdIA PAS 1. B) Predicted tetramer orientation of PAS domains formed upon NO binding to BdIA. The BdIA dimer (in yellow and green, with Gly31 highlighted in magenta) is the same as that predicted in Figure 4.13, modelled in complex with another PAS dimer (that of *Rhodobacter sphaeroides* PpsR, PDB 4HH2), shown in grey. C) The tetramer predicted in B shown from a different angle.

An alternative theory is that a BdIA PAS domain could bind to a PAS domain within a c-di-GMP synthesising enzyme and disrupt a previously established active DGC complex, reducing levels of c-di-GMP synthesis and thus reducing the local concentration of c-di-GMP. The role of the MA domain in BdIA is also worthy of consideration. The MA domain of BdIA lacks the methylation sites commonly associated with MA domains, suggesting that it does not function in a chemotaxis pathway²⁸⁷. The MA domain may instead be required to localise BdIA to other chemotaxis systems, which could accommodate a model of BdIA Tyr238 phosphorylation by a chemotaxis kinase. An attractive model therefore is that BdIA could localise with the Wsp chemotaxis system and that activation of the Wsp chemotaxis system upon surface binding results in phosphorylation of BdIA through the activity of WspE, initiating the BdIA cleavage process and leaving BdIA suitably prepared to propagate biofilm dispersal stimuli. It is worth noting that one caveat to this model is that WspE has previously been suggested to function as a histidine kinase, although no documented investigations have explored the possibility that WspE may also act as a tyrosine kinase^{393,394}.

4.5 Future work to characterise the role of BdlA in nitric oxide induced biofilm dispersal

While this work identifies BdlA as NO binding and provides some insight into BdlA processing, conformational changes upon NO binding can only be inferred. Therefore, future work could focus on characterisation of these conformational changes. As haem-bound BdlA was not amenable to crystallisation a more efficient strategy may be to examine conformational changes of the protein in solution. The highest resolution technique to identify such conformational changes is likely to be using SAXS to examine the structure of BdlA in the presence and absence of NO. An alternative to this could be to use CD, which could be used to identify any conformational changes resulting in a shift in proportions of α helix and β strand and could also identify changes in haem coordination upon NO binding.

Future work on BdlA should also focus on identifying how BdlA signals to induce biofilm formation. Using BdlA as bait, interaction partners could be identified through a pull-down assay coupled to mass spectrometry. Such an experiment could not only identify c-di-GMP metabolising proteins that interact with BdlA but may also determine the identity of the kinase responsible for Tyr238 phosphorylation.

Two further experiments that would logically follow on from this work would be to examine the physiological effects of site-directed mutation to His42, Cys46 and Cys73. If His42 serves as a haem axial ligand then mutation of His42 would be expected to result in a haem-free form of BdlA incapable of binding NO and thus prevented from signalling to induce biofilm dispersal. If the disulphide bond between Cys46 and Cys73 is allosteric in nature then mutation of either residue (preventing the formation of the disulphide bond) could result in a constitutively activated protein, which may be represented by a hyperdispersive phenotype.

Chapter 5: Identification of haem-binding PAS domains within the PAO1 genome

Many bacterial signalling pathways display a high level of redundancy. With biofilm dispersal known to be modulated by a range of different sensory and effector proteins (discussed in a number of recent reviews^{124,421,422}), it is possible that proteins other than those discussed in chapters 3 and 4 (namely, MucR, NbdA and BdlA) may be involved in the perception of NO during NO-induced *P. aeruginosa* biofilm dispersal.

While there are known to be only two MHYT domain-containing proteins within the PAO1 genome^{155,204}, both of which are discussed above in sections 1.3.5 and chapter 3, the number of haem-binding PAS domains is more difficult to determine. The low sequence homology of PAS domains (less than 20%)^{228,236,423}, prevents sequence alignment within the PAO1 genome from identifying all genes coding for this protein domain, let alone discrimination into bound cofactors. However, homology detection by iterative hidden Markov model – hidden Markov model comparison has previously been used to identify 70 PAS and PDC domain-coding genes within the PAO1 genome²³⁵. As numerous proteins contain multiple PAS or PDC domains, which although structurally similar may bind different cofactors, protein sequences were separated into those of the individual PAS/PDC domains. Using the domain boundaries predicted by this study and those given from analysis of these proteins with architecture prediction^{424,425}, a shortlist of 112 PAS/PDC domains across 70 different proteins within PAO1 (listed in Table 6.4 in Appendix C) can be analysed. As PDC domains are structurally similar²³⁵ to PAS domains, and were subjected to identical analysis based on such similarity, PDC domains are referred to as PAS domains throughout this chapter. PAS domains within the PAO1 genome were analysed to identify the nature of any bound cofactor, with the intention of identifying haem-binding PAS domains that may function in NO-induced biofilm dispersal.

5.1 Analysis of cofactor binding on the basis of conserved evolution

Prediction of a protein's functional role on the basis of primary amino acid sequence is common practice during the annotation of newly sequenced genomes. Typically, the function of a newly sequenced gene is assigned on the basis of homology to previously studied genes or proteins⁴²⁶. Despite poor sequence conservation between all PAS domains, analysis of sequence homology may highlight shared evolution between PAS domains within PAO1 and those that have previously been characterised, which could be indicative of a conserved function and conserved ligand binding. To determine conserved evolution, phylogeny of all PAS domains encoded within the PAO1 and reference PAS domains was analysed using the maximum likelihood method⁴²⁷, with quantitation provided through 100 replicates of the bootstrap statistical method using the MEGA 7 package^{428–430}. All bacterial PAS domains that have been structurally characterised with their physiological ligand were included for reference (listed in Table 5.1), with the constructed phylogenetic tree shown in Figure 5.1.

Table 5.1 - Bacterial PAS domains structurally characterised with their natural ligand, used as reference structures for comparison throughout the course of this chapter.

Protein	Species	Cofactor	Reference
CitA	<i>Klebsiella pneumoniae</i>	Carboxylic acids	Reinelt 2003 ⁴³¹
DctB	<i>Sinorhizobium meliloti</i>		Zhou 2008 ⁴³²
DctB	<i>Vibrio cholerae</i>		Cheung 2008 ²³⁴
DcuS	<i>Escherichia coli</i>		Cheung 2008 ²³⁴
PhoQ	<i>Escherichia coli</i>	Divalent cations	Marina 2001 ⁴³³
PhoQ	<i>Salmonella typhimurium</i>		Cho 2006 ⁴³⁴
MmoS	<i>Methylococcus capsulatus</i>	FAD	Ukaegbu 2009 ⁴³⁵
NifL	<i>Azotobacter vinelandii</i>		Key 2007 ²⁴²
EI222	<i>Erythrobacter litoralis</i>	FMN	Nash 2011 ⁴³⁶
LOVHK	<i>Brucella melitensis</i>		Rinaldi 2012 ⁴³⁷
LOVK	<i>Rhodobacter Sphaeroides</i>		Conrad 2013 ⁴³⁸
SB1LOV	<i>Pseudomonas putida</i>		Rollen 2016 ⁴³⁹
YtvA	<i>Bacillus subtilis</i>		Möglich 2007 ⁴⁴⁰
Aer2	<i>Pseudomonas aeruginosa</i>	Haem-b	Sawai 2012 ²⁶⁶
Dos	<i>Escherichia coli</i>		Kurokawa 2004 ¹¹⁶
FixL	<i>Bradyrhizobium japonicum</i>		Gong 1998 ²⁴⁶
GSU0582	<i>Geobacter sulfurreducens</i>	Haem-c	Pokkuluri 2008 ²⁴⁷
GSU0935	<i>Geobacter sulfurreducens</i>		Pokkuluri 2008 ²⁴⁷
Ppr	<i>Rhodospirillum centenum</i>	Hydroxycinnamic acid	Rajagopal 2003 ⁴⁴¹
PYP	<i>Halorhodospira halophila</i>		Borgstahl 1995 ⁴⁴²

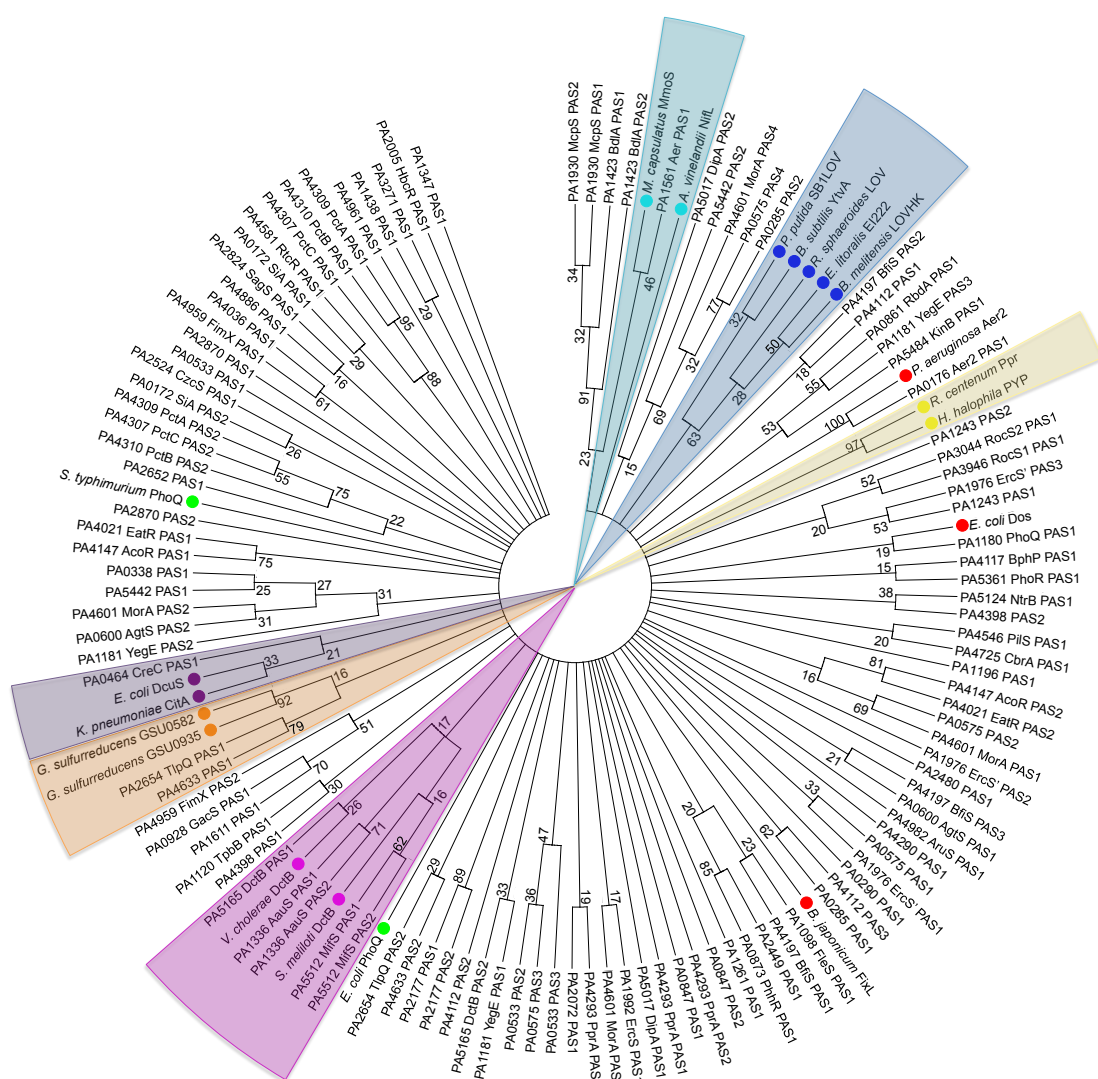


Figure 5.1 - Maximum likelihood phylogenetic analysis of all PAS domains within PAO1 with PAS domains that have been characterised structurally in complex with their physiological ligand. The percentage of bootstrap replicates that reproduced each branch is given, with branches corresponding to less than 15% bootstrap replicates collapsed for clarity. BLUF (FAD binding) domains are denoted with a light blue circle, LOV (FMN binding) domains with a dark blue circle, divalent cation binding PAS domains with a green circle, *b*-type haem binding PAS domains with a light red circle, *c*-type haem binding PAS with an orange circle, DctB-like carboxylic acid binding PAS domains with a pink circle, DcuS/CitA-like carboxylic acid binding PAS domains with a lilac circle, hydroxycinnamic acid binding PAS domains with a yellow circle. Like coloured arcs are used to denote PAS domains that can be identified as similar to all reference PAS domains that bind a specific ligand.

Many of the reference PAS domains display closer homology to each other than to any PAO1 PAS domains, as shown by grouping of these PAS domains in Figure 5.1. For example all hydroxycinnamic acid binding PAS domains and FMN binding PAS domains are found clustered. Clustering of known cofactors in this fashion demonstrates sequence conservation between these PAS domains, but provides no further information as to which PAO1 PAS domains bind these cofactors.

However, maximum likelihood phylogenetic analysis does group some PAO1 genomes with references structures. Within Figure 5.1, both PAS domains of PA1336, the first PAS domain of PA5165 and both PAS domains of PA5512 are found in close proximity to the carboxylic acid binding PAS domains of *Sinorhizobium meliloti* DctB and *Vibrio cholerae* DctB (highlighted in pink). PA5165 (DctB) has previously been identified as homologous to the *S. meliloti* protein DctB (49% sequence identity), which acts as part of the Dct system of C₄-dicarboxylate import⁴⁴³. Within this system, DctB forms the sensory part of a two-component system, with binding of C₄-dicarboxylates causing a phosphotransfer to DctD (the regulatory component of the Dct two-component system), which in turn promotes the transcription of C₄-dicarboxylate transporting proteins⁴⁴³. Two component systems are common bacterial signalling pathways used to respond to environmental changes and consist of two proteins, a sensor kinase that contains at least one signal recognition domain (such as a PAS domain) coupled to an autokinase domain, and a response regulator⁴⁴⁴. Upon signal perception by the sensor domain, the autokinase domain is activated, hydrolysing ATP to phosphorylate a conserved histidine residue, and subsequently transferring this phosphate group to a conserved aspartate residue in the receiver domain of the response regulator⁴⁴⁵. The now phosphorylated response regulator can then go on to induce a phenotypic response within the bacteria. 123 two-component signalling pathways are encoded within the PAO1 genome to facilitate responses to a range of stimuli⁴⁴⁶. Gene knock-out studies of PA5165 further validate the assignment of PA5165 as a homologue of DctB within PAO1⁴⁴³. Binding of carboxylic acids to the first PAS domain of PA5165 could induce conformational changes within the kinase domain, resulting in phosphorylation of PA5166 (the PAO1 homologue of DctD), which could then go on to induce upregulation of C₄-dicarboxylate import.

PA5512 (MifS) has also been previously identified as the sensory protein in a two-component system, with gene knock outs of PA5512 unable to utilise C₅-dicarboxylates (α -ketoglutarate and glutarate) as a sole carbon source, a phenotype which can be recovered by overexpression of the expression of the C₅-dicarboxylate transporter PA5530^{447–449}. As the sensor kinase of the MifSR two-component system, PA5512 is thought to directly sense C₅-dicarboxylates and subsequently phosphorylate MifR to active induction of PA5530⁴⁴⁸. Direct binding of carboxylic acids to either the first or second PAS domain of PA5512 would therefore seem a logical method of signal input,

supporting the assignment of the first PAS domain of PA5512 as a carboxylic acid binding PAS domain.

The assignment of both PA1336 (AauS) PAS domains as binders of carboxylic acids is supported by studies of the orthologous *P. putida* protein PP1067⁴⁵⁰. PP1067 is believed to be the sensor kinase in a two-component signalling pathway also consisting of PP1066 (orthologous to PA1335 in PAO1, AauR)⁴⁴⁶. Deletion mutants of either of these genes rendered bacteria unable to efficiently use aspartate, glutamate and glutamine as sole sources of carbon and nitrogen⁴⁵⁰. One could therefore speculate that a carboxylic acid binding PAS domain within PA1336 may serve to bind the amino acids aspartate, glutamate and glutamine, with the subsequent two-component signalling pathway responsible for upregulation of genes responsible for catabolism of these amino acids.

The first PAS domain of PA0464 (CreC) is grouped with *E. coli* DcuS and *Klebsiella pneumoniae* CitA carboxylic acid binding PAS domains (highlighted in lilac) in Figure 5.1. PA0464 has previously been identified as homologous to the *E. coli* protein CreC (which was itself formerly labelled as PhoM), which is the sensor kinase in a two-component signalling pathway, regulating the expression of a number of genes in response to the presence of carbon sources in the growth medium^{451,452}. A carboxylic acid binding PAS domain within PA0464 may provide a homologous role to that observed the homologous *E. coli* CreC protein.

A similar instance of experimental evidence supporting cofactor assignment based on maximum likelihood phylogenetic analysis to reference structures can be seen in the case of PA1561 (Aer). Within Figure 5.1 the PAS domain of PA1561 is grouped with the FAD binding PAS domains of *Methylococcus capsulatus* MmoS and *Azotobacter vinelandii* NifL (highlighted in light blue). PA1561 displays 45% sequence homology to *E. coli* Aer, with gene knockouts of PA1561 highlighting a similar involvement in aerotaxis²⁶⁹. As *E. coli* Aer directly monitors oxygen levels and facilitates aerotaxis through a bound FAD cofactor, it is logical that an equivalent function of PA1561 in PAO1 would also utilise a bound FAD cofactor.

Despite evidence suggesting maximum likelihood phylogenetic analysis to reference structures can successfully identify the nature of cofactor binding within PA1336, PA5165, PA5512, PA0464 and PA1561, this methodology is not useable for identification of haem binding PAS domains within the PAO1 genome. The broad distribution of the *b*-type haem binding PAS domains of *E. coli* Dos, *B. japonicum* FixL and *P. aeruginosa* Aer2 throughout Figure 5.1 (all highlighted with red circles) demonstrates that this method does not identify the necessary characteristics for *b*-type haem binding PAS domains, an observation further supported by none of the reference *b*-type haem binding PAS domains being found in close proximity to the first PAS domain of BdIA,

which has been experimentally determined to bind *b*-type haem²⁸⁸. Further complications are also present in the identification of *c*-type haem binding PAS domains. Although *Geobacter sulfurreducens* GSU0582 and GSU0935 are clustered with PAS 1 of PA2654 and PAS 1 of PA4633 in Figure 5.1 (highlighted in orange), closer inspection of the amino acid sequences of these PAO1 PAS domains shows that they both lack the CXXCH motif required for binding a *c*-type haem. It is therefore necessary to use other methods to identify haem binding PAS domains within the PAO1 genome.

5.2 Assignment of cofactor binding based on conservation of interacting residues

An alternative approach to identify the nature of cofactors bound by PAS domains within the PAO1 genome, and thus identify haem binding PAS domains, is to discriminate based upon the presence of residues involved in side-chain specific cofactor coordination. Conservation of amino acids that form side-chain specific interactions with the cofactor between PAS domains with known mechanisms of cofactor binding and uncharacterised PAO1 PAS domains may be indicative of conserved cofactor binding. To identify bound cofactors in this method, the amino acid sequences of all PAS domains within PAO1 were individually aligned to a group of reference PAS domains known to bind the same cofactor (for example *E. coli* Dos and *B. japonicum* FixL were used as reference PAS domains for *b*-type haem binding PAS domains). This process was then repeated to analyse each PAS domain against all of the types of structurally characterised cofactor binding PAS domains (namely divalent cation binding, FAD binding, FMN binding, *b*-type haem binding, *c*-type haem binding, hydroxycinnamic acid binding, Dcus/ CitA-like carboxylic acid binding and DctB like carboxylic acid binding), using the reference structures listed in Table 5.1. PAS domains that displayed conservation of greater than 66% of cofactor binding residues (or residues with similar physical properties) were assigned binding cofactors and are listed below in Table 5.2, with example alignments shown in Figure 5.2.

Table 5.2 - PAS domain cofactor binding identified on the basis of conserved interacting residues

Gene name	PAS domain	Bound ligand/ cofactor	Known physiological role
PA0176 Aer2	PAS1	Haem- <i>b</i>	Suggested involvement in aerotaxis ^{252,266,453}
PA0285	PAS1	Haem- <i>b</i>	Biofilm formation, swimming and swarming motilities ⁴⁵⁴
	PAS2	FAD	
PA0575	PAS2	Hydroxycinnamic acid	Unknown ^{104,114,289,454}
	PAS4	FAD	
PA0873 PhhR	PAS1	Carboxylic acids	Regulation of genes involved in amino acid catabolism ^{455–458}
PA0928 GacS	PAS1	Carboxylic acids	Regulation of virulence factors and proteins involved in secretion and biofilm formation ^{459–461}
PA1098 FleS	PAS1	Hydroxycinnamic acid	Transcription of genes responsible for flagellar biogenesis ^{462,463}
PA1181 YegE	PAS2	Haem- <i>b</i>	Biofilm dispersal ^{104,289}
PA1336 AauS	PAS1	Carboxylic acids	Use of aspartate, glutamate and glutamine ^{446,450}
PA1347	PAS1	Hydroxycinnamic acid	Swarming motility ⁴⁶⁴
PA1438	PAS1	FAD	Unknown
PA1561 Aer	PAS1	FAD	Aerotaxis ^{269,387}
PA1611	PAS1	Haem- <i>b</i>	Biofilm formation, swimming motility, type III secretion ^{461,465}
PA2654 TlpQ	PAS1	FAD	Chemotaxis towards ethylene ^{367,400}
PA2870	PAS1	Haem- <i>b</i>	Unknown ^{114,118}
	PAS2	Divalent cations	
PA3946 RocS1	PAS1	Haem- <i>b</i>	Fimbriae expression and biofilm formation ^{349,466–468}
PA4021 EatR	PAS1	Haem- <i>b</i>	Ethanolamine catabolism ⁴⁶⁹
PA4036	PAS1	Haem- <i>b</i>	Unknown
PA4147 AcoR	PAS1	Hydroxycinnamic acid	Regulation of genes involved in ethanolamine catabolism ^{469–471}
PA4290	PAS1	Haem- <i>b</i>	Unknown
PA4309 PctA	PAS2	Divalent cations	Sensor for chemotaxis towards amino acids ^{363,364}
PA4725 CbrA	PAS1	Carboxylic acids	Regulation of genes involved in degradation of amino acids and sugars ^{472,473}
PA4886	PAS1	Haem- <i>b</i>	Unknown
PA5017 DipA	PAS2	FAD	Biofilm dispersal, growth, chemotaxis ^{289,474}
PA5165 DctB	PAS1	Carboxylic acids	Use of succinate, malate or fumarate ⁴⁴³
PA5442	PAS1	Haem- <i>b</i>	Growth and swarming motility ⁴⁵⁴
	PAS2	Haem- <i>b</i>	
PA5512 MifS	PAS1	Carboxylic acids	Use of α -ketoglutarate and glutarate as carbon sources. ^{447–449}

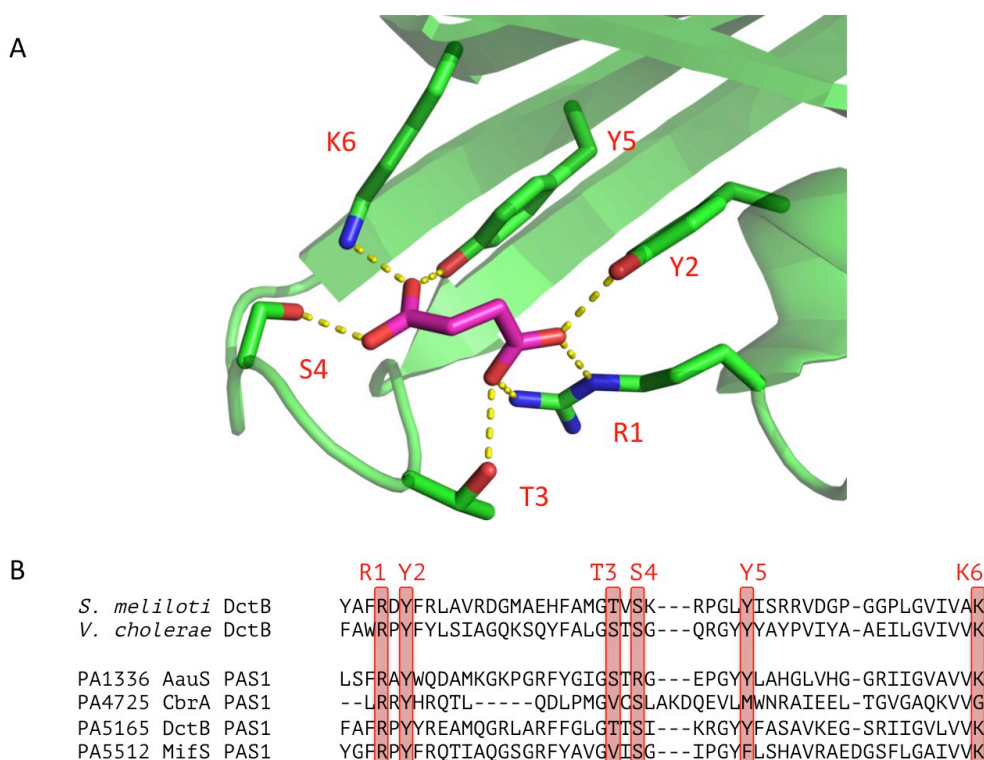


Figure 5.2 - Example assignment of cofactor binding based on conservation of interacting residues. A) *Sinorhizobium meliloti* DctB in complex with succinate (PDB 3E4O), with residues involved in the coordination of malonate and succinate labelled and shown as sticks. B) Composite sequence alignment of reference DctB PAS domains from *S. meliloti* and *Vibrio cholerae* and putative DctB-like PAS domains identified in the PAO1 genome through conservation of labelled residues.

In a number of cases, previously determined experimental evidence supports the assignment of cofactor binding based upon conservation of coordinating residues, supporting the validity of this method. For example, the PA0873 (PhhR) PAS domain and the PAS domain of PA0928 (GacS) were both assigned as DcuS/ CitA-like carboxylic acid binding. PA0873 has been characterised as regulating the transcription of a number of protein involved in catabolism of phenylalanine and tyrosine^{455–458}. A carboxylic acid binding PAS domain within PA0873 could allow direct sensation of increased intracellular phenylalanine and tyrosine levels, with subsequent conformational changes inducing increased transcription of amino acid catabolic machinery. PA0928 has previously been identified as the sensor of a two-component system responsible for regulating the expression of virulence factors, proteins involved in secretion and biofilm formation^{459–461}. Increased levels of citrate have been identified to induce the dispersal of *P. aeruginosa* biofilms⁸⁹. A carboxylic acid binding PAS domain within PA0928 could be capable of directly binding citrate and altering the functionality of the two component system, downregulating the expression of DGCs or other proteins involved in biofilm formation.

Further examples of previous experimental evidence supporting cofactor assignment are provided through the determination of DctB-like carboxylic acid binding PAS domains. As well as identifying the first PAS domain of PA1336 (AauS), the first PAS domain of PA1565 and the first PAS domain of PA5512 as DctB-like binders of carboxylic acids (all of which are discussed in further detail in section 5.1 as they were also identified through phylogeny analysis), the first PAS domain of PA4725 (CbrA) was also identified as binding carboxylic acids through a DctB-like interactions. PA4725 is the sensor kinase of a two-component signalling pathway involved in regulating the expression of catabolic pathways responsible for degrading numerous amino acids (including arginine, histidine and proline) and sugars (including glucose, citrate, mannitol and pyruvate)^{472,473}. A carboxylic acid binding PAS domains within PA4725 may be capable of directly binding amino acids or sugars, with binding inducing a conformational change resulting in upregulated expression of the systems responsible for catabolism of the bound substrate.

The only *b*-type haem binding PAS domain within PAO1 to have been previously structurally characterised, that of Aer2, is also successfully identified using assignment based upon conservation of coordinating residues within amino acid alignment, even though the axial histidine for haem coordination in Aer2 has been identified to be located on the $\alpha 3$ helix, rather than on the $\alpha 4$ helix observed in structures of *B. japonicum* FixL and *E. coli* Dos which was searched for here^{116,250,266,415}. It is worth noting that the first PAS domain of BdlA is not identified as haem binding with this methodology, which contrasts to experimental observation²⁸⁸. This contradiction may be indicative of BdlA putatively coordinating haem through an axial histidine located on the $\alpha 3$ helix (as detailed in section 4.4). This technique is also not suitable for identifying haem binding PAS domains similar in nature to YybT, as haem coordinating residues have not yet been determined within these proteins^{282,475}.

As a number of PAS domains are identified as binding haem-*b*, being able to place them in the context of full-length proteins may help to identify proteins involved in NO-induced biofilm dispersal.

5.3 Candidate proteins involved in NO-induced biofilm dispersal

Predicted domain topology and previously determined *in vivo* roles can be used to further guide which of the proteins identified to contain *b*-type haem binding PAS domains may be involved in NO-induced biofilm dispersal. Four proteins that may play a sensory role in NO induced biofilm dispersal are short-listed for further investigation.

PA0176 (Aer2) has been experimentally determined to bind *b*-type haem and directly bind O₂^{252,265–268}. A role for PA0176 in aerotaxis has previously been suggested based upon

recombinant expression in *E. coli*, with binding of NO to PA0176 observed to increase tumbling behaviour²⁶⁵ (rapid bacterial reorientation due to flagellar rotation, a motility inactivated during chemotaxis towards an attractant to allow straight-line swimming^{476–478}). As tumbling behaviour is part of an early stage response to colonisation of a new area, it is therefore possible that the haem PAS domain of PA0176 acts to sense NO and conformational changes within PA0176 upregulate tumbling as part of a biofilm dispersal response.

Binding of diatomic gas to the haem PAS domain of PA0176 has been modelled using cyanide and comparison to a ligand-free state^{252,266}. Binding of a diatomic gas molecule results in a repositioning of the haem distal Trp293 and Leu 264 residues and movement of the PAS domain β -sheet, the N-terminal helix and the C-terminal connector to the HAMP4 domain, as shown in Figure 5.3. As small-angle X-ray scattering (SAXS), electron paramagnetic resonance (EPR) and pull-down studies all support a linear model of PA0176 (shown in Figure 1.22 and Figure 5.3), as opposed to a direct interface between the PAS domain and MA domain, HAMP domains C-terminal to the PAS domain are thought to serve a role in the propagation of ligand binding signal to alter behaviour of the MA domain²⁵². Through analogy to other chemotaxis systems, the MA domain of PA0176 is thought to interact with CheA2 and CheW2^{265,273–275}. Conformational changes upon NO binding are thought to alter the autophosphorylation of CheA2 and subsequent phosphorylation of CheY2, with the phosphorylated CheY2 then able to interact with the flagellar motor and induce tumbling behaviour, reorienting the bacterium to facilitate travel to a more favourable environment^{265,276–280}. It is worth noting that the increased tumbling behaviour in response to increased oxygen concentration observed of *E. coli* expressing PA0176 is the opposite to the classic aerotaxis response in *E. coli*, in which tumbling motility is increased as oxygen concentrations decrease^{265,479}.

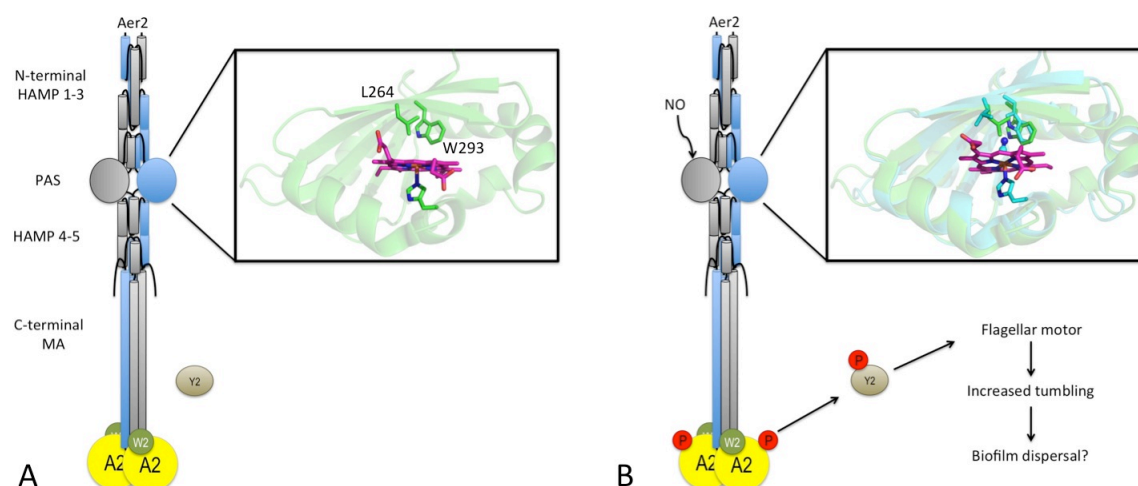


Figure 5.3 - NO binding to the haem PAS of PA0176 (Aer2) induces a change in motility. A) CheW2 (green circles) and CheA2 (yellow circles) form a complex with the MA domain of PA0176. B) Upon NO binding to the haem located within the PAS domain, Leu264 and Trp293 are displaced, shifting the PAS domain. This signal is propagated through the HAMP 4-5 domains, and the C-terminal MA domain, allowing CheA2 to autophosphorylate and phosphorylate CheY2. CheY2 then goes on to interact with the flagellar motor to increase tumbling, which may have a role in biofilm dispersal. Adapted from Watts *et al.* 2011²⁶⁵. The ligand free haem PAS domain of Aer2 is shown as a green cartoon in both A and B (PDB 4HI4), and the cyanide bound PAS domain (used as a model for NO binding) is shown as a blue cartoon in panel B (PDB 3VOL).

It is worth noting that although PA0176 has been observed to play a role in chemotaxis when expressed in *E. coli*, attempts to show a similar role in *P. aeruginosa* have been unsuccessful^{267,268}. It is therefore possible that PA0176 may have an alternative and potentially complementary role as a sensor of diatomic gases for biofilm regulation, inducing dispersal in response to NO. As both PA0176 and BdlA contain haem-PAS domains and MA domains, it is possible that PA0176 may signal using similar mechanisms to BdlA (section 4.4), interacting with other PAS domains to promote PDE domain dimerisation.

Gene knockout studies of PA0285 in PAO1 display altered motility, reduced growth and reduced biofilm formation⁴⁵⁴. PA0285 is predicted to contain a transmembrane anchor, DGC and PDE domains, as well as two PAS domains, predicted by this work to bind *b*-type haem and FAD respectively. If PA0285 does contain a *b*-type haem binding PAS domain it may function to detect diatomic gas molecules such as NO, and directly alter c-di-GMP metabolism, contributing to NO-induced biofilm dispersal.

PA1181 (YegE) is predicted to contain eight transmembrane helices, two PAS domains, a DGC and a putatively degenerate PDE domain. PA1181 has been demonstrated to have a role in NO-induced and glutamate-induced biofilm dispersal in PA14, with gene knock-outs showing 40-66% reduced, but not abolished dispersal^{104,289}. Assuming a similar phenotype is observed for gene knockouts of PA1181 in PAO1 to those observed in PA14, a *b*-type haem binding PAS domain within PA1181 may serve to bind NO, inducing conformational changes that are propagated through the DGC and PDE domains to reduce the local pool of c-di-GMP and contribute to biofilm dispersal.

Gene knock-out experiments of PA5442 in PAO1 display reduced cell growth and altered swarming motility in response to nitric oxide⁴⁵⁴. PA5442 is predicted to contain a transmembranous domain featuring five transmembrane helices, two PAS domains, a degenerate DGC domain (with an AGDEF motif rather than GGDEF) and a PDE domain; although activity of neither DGC nor PDE were observed through *in vitro* overexpression assays¹¹⁴. As high levels of c-di-GMP have previously been shown to repress swarming motility^{421,480,481}, it is possible that a haem-binding PAS domain within PA5442 may function to sense NO and modulate PDE activity to increase swarming motility, as part of a biofilm dispersal response.

It is possible to speculate upon the possible mechanisms linking binding of NO to haem cofactors within the PAS domains of PA0285, PA1181 and PA5442 to biofilm dispersal using models provided from *E. coli* Dos and FixL proteins (also discussed in section 1.4.2). Within *E. coli* Dos, ligand binding shifts the position of the distal iron coordinating residue, shifting the position of the α 4- β 3 loop which in turn alters the PAS domain dimerisation interface, predicted to prevent dimerisation of C-terminal PDE domains^{115,116,403}. Within FixL proteins, ligand binding causes the bound haem cofactor to shift, altering the salt bridges formed by propionate carboxyl groups of the haem and allowing the α 4- β 3 loop, β 3 and β 4 strands to reposition as an arginine side chain previously involved in coordinating the haem propionate carboxyl now coordinates bound ligand and displaces the distal hydrophobic residues^{246,249,258,263}. Repositioning of the β 3 and β 4 strands is thought to subsequently alter the nature of the interface between the haem-PAS domain in FixL and the C-terminal kinase domain, inhibiting kinase activity²⁵⁸.

The position of the distal residues with relation to secondary structure elements varies between FixL and Dos, so without structural characterisation it is difficult to speculate on the atomic changes that might be induced by NO binding, however as in both FixL and Dos rearrangement in the α 4- β 3 loop is thought to be a key feature in signal transduction, this feature may be conserved within PA0285, PA1181 and PA5442 signalling. It is worth noting however that as PA0285 and PA1181 both possess DGC and PDE domains, and as both DGC and PDE domains are

regulated through dimerisation^{179,196,197,202}, conformational changes upon NO binding are unlikely to shift the overall monomer-dimer equilibrium and are instead more likely to alter the interaction interface with one of the enzymatic domains (as is the case in FixL²⁵⁸). It could therefore be speculated that direct interaction between the PAS domain $\alpha 4$ - $\beta 3$ and either the DGC or PDE domain may be responsible for propagating the signal of NO binding through to altered c-di-GMP metabolism, as represented in Figure 5.4. As the DGC domain in PA5442 is degenerate and putatively inactive, NO binding to one or both of the PA5442 haem PAS domains may signal in a similar way to PA0285 and PA1181, or it may increase PDE activity by promoting dimerisation, as represented in Figure 5.5.

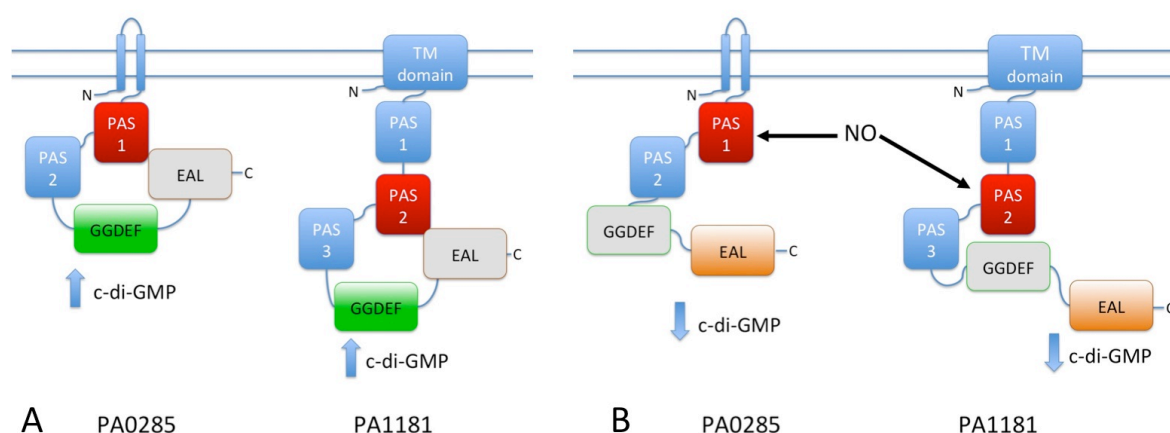


Figure 5.4 - Proposed model of how PA0285 and PA1181 may function in NO-induced biofilm dispersal. A) In the biofilm state PAS1 of PA0285 and PAS2 of PA1181 (coloured red) may interact directly with their respective EAL domains, holding them in an inactive state (shown in grey), but allowing their GGDEF domains (green) to dimerise and synthesise c-di-GMP. B) NO binding to the haem PAS domains of PA0285 and PA1181 induces a conformational change allowing their respective EAL domains (orange) to dimerise and catabolise c-di-GMP, inducing biofilm dispersal. An equally likely model is that the haem PAS domains in PA0285 and PA1181 directly interact the GGDEF domains, holding them in an active form in the absence of NO, but preventing the formation of the active GGDEF domain dimer in the presence of NO.

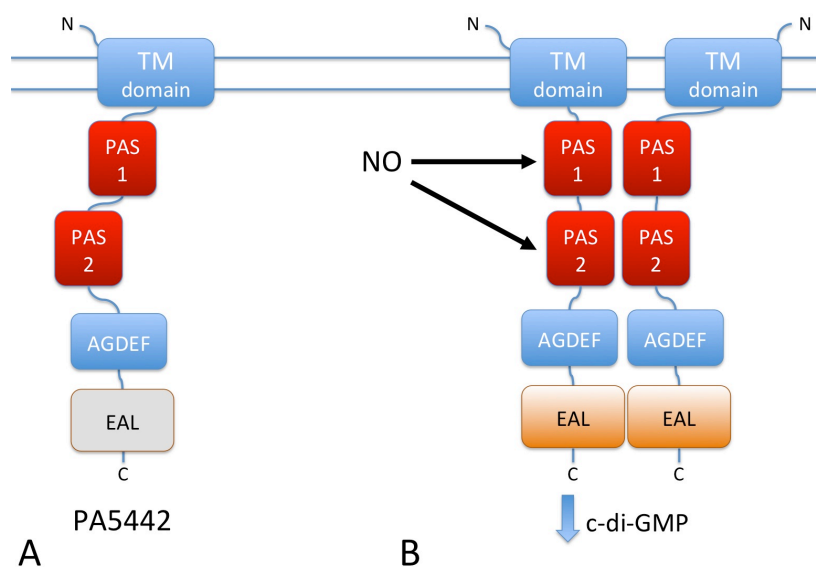


Figure 5.5 - Binding of NO to the haem PAS domains of PA5442 could induce dispersal by inducing homodimerisation. A) In the monomeric form, the EAL domain of PA5442 is inactive (grey). B) Binding of NO to the haem PAS domains (red) induces conformational changes promoting dimerisation, leading to the formation of the active EAL domain dimer (orange), catabolism of c-di-GMP and biofilm dispersal. Such a shift in monomer-dimer equilibrium could also be concurrent with a model of direct PAS domain-EAL domain interaction as outlined in Figure 5.4.

5.4 Future work

Whilst the bioinformatics approach to determination of ligand binding presented within this chapter may help to guide future studies of NO sensation within PAO1 (as well as other PAS domain mediated phenotypes), this work could be further improved and conclusions experimentally validated.

The process of predicting cofactor binding on the basis of conserved coordinating residues could be improved with the inclusion of secondary structure prediction. PAS domains are defined by their structural similarity despite poor sequence homology, therefore an alignment based on conserved secondary structure elements (a topology of $\beta\beta\alpha_4\beta_3$ in the case of cytoplasmic PAS domains and an $\alpha_2\beta_2\alpha$ (or α_2) $\beta_3\alpha$ topology for extracellular PAS domains)²³⁵ and that could accommodate differing lengths of loop between these secondary structure elements may better reflect equivalent positioning of cofactor coordinating residues than alignment based on amino acid sequence. Proteins discussed in section 5.3 were evaluated for positioning of conserved amino acids with respect to protein secondary structure elements using the secondary structure prediction PSIPRED web-server^{401,402}, suggesting that results may be similar to alignment based upon amino acid sequence, however a hierarchical alignment strategy based upon protein secondary structure then examining for conservation of cofactor binding residues may identify further haem-binding PAS domains that may be candidates for involvement in NO-induced biofilm dispersal.

To fully assess the accuracy of ligand binding predictions of PAS domains given here, these proteins could be studied *in vivo* and *in vitro*. Whilst phenotypes have been determined from gene knockout studies of a number of these proteins, and in some cases these would appear complimentary to the predictions of ligand binding (for example the assignment of PA1181 as binding haem-*b*), a more targeted approach with removal of putative ligand binding residues may provide further insight as to the accuracy of these predictions.

An alternative strategy to determine ligand interactions would be to overexpress the proteins and analyse them *in vitro*. Protein overexpression within PAO1 would be likely to be the most informative method of determining the native ligand, and may prevent optimisation of ligand binding conditions, such as those detailed for haem-binding to BdlA in chapter 4. As this strategy may not always be possible, ligand binding could also be studied through overexpression in other systems followed by purification and, if no native ligand is present, assessment of ligand binding with isothermal calorimetry.

In the context of NO-induced biofilm dispersal, the *in vitro* studies described above could be further extended to structural characterisation of haem-*b* binding within the first PAS domain of PA0285, the second PAS domain of PA1181 and both PAS domains of PA5442. Structures in the presence and absence of NO in the context of full-length protein could provide atomic understanding of how NO perception corresponds to regulation of c-di-GMP levels and thus biofilm dispersal. However as PA0285, PA1181 and PA5442 all possess N-terminal transmembranous regions, it may be more experimentally viable to construct shorter protein fragments not including these N-terminal domains. *In vitro* reconstitution of each protein could also be used to directly monitor if addition of NO could result in either a reduction in c-di-GMP production (in the case of PA0285 and PA1181) or an increase in c-di-GMP catabolism (which could apply to PA0285, PA1181 and PA5442).

Binding of haem-*b* to PA0176 has previously been determined^{252,265–268}, and structural changes upon the binding of NO could be modelled upon the previously observed structure of the PA0176 haem PAS domain in complex with CN, however any novel conformational changes upon NO binding could be identified by soaking NO into PA0176 PAS domain crystals, using fragments and crystallisation conditions previously identified^{252,266}. A more prominent experiment with regards to the role of PA0176 may be to identify interaction partners within PAO1 using a protein pull-down, with interaction partners identified with mass-spectrometry. Such a pull-down experiment could be performed in the presence and absence of NO, potentially identifying binding partners specific to the NO-bound form of PA0176 and suggesting the signalling pathways PA0176 could be involved in.

Chapter 6: Conclusions

Non-lethal doses of NO can induce dispersal of antibiotic tolerant *P. aeruginosa* biofilms, returning the bacteria to an antibiotic susceptible phenotype²¹. The aim of this work was to identify systems within *P. aeruginosa* for sensation of NO, and to determine how these systems can then regulate biofilm dispersal. While conformational changes induced upon NO binding to *P. aeruginosa* proteins were not observed, this work does provide insight on how c-di-GMP degrading PDEs are regulated, as well as confirming NO binding to a putative sensor protein required for NO-induced biofilm dispersal and identifying other candidate NO sensors for NO-induced biofilm dispersal.

Chapter 3 describes enzymatic characterisation of the protein MucR, required for NO-induced biofilm dispersal^{155,156}, as well as structural characterisation of the MucR^{EAL} domain. While attempts to structurally characterise the putative NO sensing domain of MucR (and that of the related protein NbdA) were unsuccessful, further trials to identify conditions yielding monodisperse MucR⁷⁻²⁶⁶ could result in purified protein suitable for crystallisation, building upon the optimisation of expression and purification conditions detailed within this chapter. Similarly, other gene fragments of MucR and NbdA cloned during this work could be subjected to the same methods of optimisation and purification, which could result in protein more suitable for crystallisation.

Although the exact nature of signal transduction from the MHYT domains to the enzymatic domains in MucR and NbdA could not be determined, structural characterisation of the c-di-GMP catabolising MucR^{EAL} domain does provide some insight into how PDE activity may be regulated. The structure of substrate-bound MucR^{EAL} contrasts to that of substrate-bound PA3825^{EAL} with respect to the dimerisation interface. However, conformational changes of the $\beta 5$ - $\alpha 5$ loop required for substrate binding are conserved, suggesting that regulation of dimerisation around the $\alpha 5$ helix (in close proximity to the $\beta 5$ - $\alpha 5$ loop) may regulate EAL-PDE activity across different dimerisation interfaces. Such a regulatory model may explain how the opposing activities of tandem DGC-PDE proteins are regulated. As both DGC and EAL domain PDEs require dimerisation for catalysis it is possible that c-di-GMP hydrolysis only occurs when dimerisation around the EAL domain $\alpha 5$ helix take place, allowing the tandem DGC-PDE protein to maintain a dimeric quaternary state and c-di-GMP synthesis to occur in the absence of this specific interaction.

Further insight into the mechanism of c-di-GMP hydrolysis is offered through observation of PA3825^{EAL} in complex with pGpG. Within this product complex metal ions are located in different positions to the substrate complex, occupying the M1 and M3 sites in the product complex but

the M1 and M2 sites in the substrate complex. Comparison to the structurally characterised complex between CC3396^{EAL} and pGpG identifies the observation of a metal in M3 position as concurrent with bound pGpG. Although a catalytic requirement for a metal in the M3 position has not yet been identified, a metal positioned as such would facilitate PDE activity through a similar mechanism to nuclease P1 and endonuclease IV^{350–354}. A three metal mechanism of c-di-GMP hydrolysis would raise the possibility that EAL-PDE activity could be further regulated by controlling the binding of a metal to the M3 position, although the nature of how this would take place is currently unclear. It is therefore possible that binding of NO to the MHYT domains of MucR or NbdA may propagate conformational changes throughout the protein resulting in dimerisation around the EAL domain α 5 helix and positioning of a metal in the M3 site, acting to increase EAL-PDE activity and promote biofilm dispersal.

The spectroscopic characterisation of NO binding to BdlA^{1–127} in chapter 4 identifies BdlA as a sensor for NO within *P. aeruginosa*. Interactions between NO and the haem cofactor of BdlA^{1–127} were observed using UV-Vis and FTIR spectroscopy. FTIR spectra not only identify interaction between NO and Fe²⁺, but also a second interaction unique to NO binding. Although the nature of this interaction is not yet clear this could be indicative of a second NO binding state, or of a conformational change induced upon NO binding.

Structural characterisation of BdlA^{1–417} in a cofactor free state provides insight with regards to the molecular details of how this protein is processed within *P. aeruginosa*. Tyr238 is located at the top of the first α helix of the MA domain and is suitable positioned to hydrogen bond to the second α helix of the MA domain or to hydrogen bond to the MA domain of another BdlA monomer. Phosphorylation of Tyr238, part of the processing of BdlA during biofilm formation, is likely to alter the nature of this interaction, with subsequent conformational changes propagated throughout the protein, exposing the BdlA cleavage site at Met130²⁸⁷. The newly cleaved fragments of BdlA are then thought to reassociate through PAS domain dimerisation. Inferences with regards to haem binding and signal propagation can also be made based upon the determined structure of BdlA^{1–417}. The positioning of His42 within the cofactor binding pocket suggests this could function as an axial haem ligand. A disulphide bond between Cys46 and Cys73 was also identified that could function to propagate conformational changes associated with NO binding across the dimerisation interface of BdlA PAS 1.

The bioinformatics approaches detailed in chapter 5 were used to identify haem binding PAS domains within *P. aeruginosa* that could function as NO sensors. The proteins PA0285, PA1181 (YegE) and PA5442 all contain c-di-GMP metabolising domains in addition to haem PAS domains, suggesting that binding of NO to the haem PAS could be directly coupled to regulation of

c-di-GMP levels and biofilm dispersal. The other protein identified as being of particular interest for NO-induced biofilm dispersal is PA0176 (Aer2). Conformational changes induced upon binding of NO to the haem PAS domain within PA0176 have already been characterised, however the physiological role of PA0176 has not yet been fully elucidated²⁶⁵. It is therefore worth exploring the possibility that PA0176 may serve as a sensor for NO and induce biofilm dispersal.

Biofilm dispersal requires coordinated alteration of several distinct behaviours through reductions in pools of local c-di-GMP. It is therefore likely that many proteins may serve to sense NO and induce biofilm dispersal, with each NO sensor regulating a particular facet of dispersal behaviour. This work is therefore hoped to lay foundations of research into how conformational changes induced upon NO binding to one or all of BdlA, MucR, NbdA, PA0176, PA0285, PA1181 and PA5442 can induce biofilm dispersal. The understanding gained from such research into NO induced biofilm dispersal could be used in the design of therapeutics to treat biofilm infections when used in combination with conventional antibiotics.

Appendix A

Table 6.1 - PCR primers used to amplify genes, as described in section 2.1.10.

PCR primer	Sequence
MucR 235 fwd	AACTT CATATG TCGGTGCTGGACTCG
MucR 266 fwd	AACTT CATATG GACAACCTGACCAAGCT
MucR 425 fwd	AACTT CATATG TCGATGAACGCCA
NbdA 332 fwd	AACTT CATATG AGCGAAAACAGCCG
NbdA 506 fwd	AACTT CATATG AGCCGGCAACTGC
BdIA 1 fwd	AACTT CATATG GCGGCCCTGGAC
BdIA 8 fwd	AACTT CATATG GCGCGCGTGG
MucR 425 rev	AACTT AAGCTT TCAACTCTCGAAGAAGCAG
MucR 685 rev	AACTT AAGCTT CAGGCGACGCTG
NbdA 772 rev	AACTT AAGCTT CAGTCGAAGCGGT
BdIA 127 rev	AACTT AAGCTT CAGGACAGGGCATCCA
BdIA 417 rev	AACTT AAGCTT CTAGAGATCGGCGTTGAGGG

PCR primers are labelled according to the terminal amino acids of the gene fragment they encode (N-terminal for forward primers, abbreviated to fwd, and C-terminal for reverse primers, abbreviated to rev). Restriction endonuclease recognition sites are highlighted in bold.

Appendix B

Table 6.2 - Composition of Morpheus Custom Optimisation Hanging Drop Screen. Each condition was generated by mixing together stock conditions from the Morpheus crystallisation screen (Molecular dimensions and detailed in Table 6.3)⁴⁸², here labelled according to the wells they contribute in a 96 well plate.

	1	2	3	4	5	6
A	100% A10	83% A10, 17% A6	67% A10, 33% A6	50% A10, 50% A6	33% A10, 67% A6	17% A10, 83% A6
B	100% B10	83% B10, 17% B6	67% B10, 33% B6	50% B10, 50% B6	33% B10, 67% B6	17% B10, 83% B6
C	100% C10	83% C10, 17% C6	67% C10, 33% C6	50% C10, 50% C6	33% C10, 67% C6	17% C10, 83% C6
D	100% D10	83% D10, 17% D6	67% D10, 33% D6	50% D10, 50% D6	33% D10, 67% D6	17% D10, 83% D6
E	100% E10	83% E10, 17% E6	67% E10, 33% E6	50% E10, 50% E6	33% E10, 67% E6	17% E10, 83% E6
F	100% F10	83% F10, 17% F6	67% F10, 33% F6	50% F10, 50% F6	33% F10, 67% F6	17% F10, 83% F6
G	100% G10	83% G10, 17% G6	67% G10, 33% G6	50% G10, 50% G6	33% G10, 67% G6	17% G10, 83% G6
H	100% H10	83% H10, 17% H6	67% H10, 33% H6	50% H10, 50% H6	33% H10, 67% H6	17% H10, 83% H6

Table 6.3 - Components of Morpheus crystallisation screen stock solutions used to produce Morpheus Custom Optimisation Hanging Drop Screen.

Stock	Chemical composition
A10	100 mM Morpheus Buffer System 3 pH 8.5, 60 mM Morpheus Divalents and 50% of an ethylene glycol, PEG 8,000 mix
A6	100 mM Morpheus buffer system 2 pH 7.5, 60 mM Morpheus Divalents and 50% of an ethylene glycol, PEG 8,000 mix
B10	100 mM Morpheus buffer system 3 pH 8.5, 90 mM Morpheus Halogens and 50% of an ethylene glycol, PEG 8,000 mix
B6	100 mM Morpheus buffer system 2 pH 7.5, 90 mM Morpheus Halogens and 50% of an ethylene glycol, PEG 8,000 mix
C10	100 mM Morpheus buffer system 3 pH 8.5, 90 mM Morpheus Nitrate Phosphate Sulphate and 50% of an ethylene glycol, PEG 8,000 mix
C6	100 mM Morpheus buffer system 2 pH 7.5, 90 mM Morpheus Nitrate Phosphate Sulphate and 50% of an ethylene glycol, PEG 8,000 mix
D10	100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus Ethylene Glycols and 50% of an ethylene glycol, PEG 8,000 mix
D6	100 mM Morpheus buffer system 2 pH 7.5, 120 mM Morpheus Ethylene Glycols and 50% of an ethylene glycol, PEG 8,000 mix
E10	100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus Alcohols and 50% of an ethylene glycol, PEG 8,000 mix
E6	100 mM Morpheus buffer system 2 pH 7.5, 120 mM Morpheus Alcohols and 50% of an ethylene glycol, PEG 8,000 mix
F10	100 mM Morpheus buffer system 3 pH 8.5, 120 mM Morpheus Monosaccharides and 50% of an ethylene glycol, PEG 8,000 mix
F6	100 mM Morpheus buffer system 2 pH 7.5, 120 mM Morpheus Monosaccharides and 50% of an ethylene glycol, PEG 8,000 mix
G10	100 mM Morpheus buffer system 3 pH 8.5, 100 mM Morpheus Carboxylic Acids and 50% of an ethylene glycol, PEG 8,000 mix
G6	100 mM Morpheus buffer system 2 pH 7.5, 100 mM Morpheus Carboxylic Acids and 50% of an ethylene glycol, PEG 8,000 mix
H10	100 mM Morpheus Buffer System 3 pH 8.5, 120 mM Morpheus Amino Acids and 50% of an ethylene glycol, PEG 8,000 mix
H6	100 mM Morpheus buffer system 2 pH 7.5, 100 mM Morpheus Amino Acids and 50% of an ethylene glycol, PEG 8,000 mix

Appendix C

Table 6.4 - All PAS domains within the PAO1 genome, used for analysis in chapter 5.

Gene number	Protein name	PAS domain	Start residue	End residue
PA0172	SiaA	PAS 1	62	196
		PAS 2	198	313
PA0176	Aer2	PAS 1	163	396
PA0285	None	PAS 1	81	200
		PAS 2	201	320
PA0290	None	PAS 1	31	160
PA0338	None	PAS 1	41	180
PA0464	CreC	PAS 1	86	180
PA0533	None	PAS 1	1	155
		PAS 2	132	270
		PAS 3	242	400
PA0575	None	PAS 1	311	430
		PAS 2	438	550
		PAS 3	561	680
		PAS 4	681	800
PA0600	AgtS	PAS 1	320	457
		PAS 2	447	637
PA0847	None	PAS 1	142	284
		PAS 2	441	560
PA0861	RbdA	PAS 1	243	384
PA0873	PhhR	PAS 1	58	229
PA0928	GacS	PAS 1	43	161
PA1098	FleS	PAS 1	58	180
PA1120	TpbB	PAS 1	46	152
PA1180	PhoQ	PAS 1	33	161
PA1181	YegE	PAS 1	300	420
		PAS 2	421	550
		PAS 3	551	680
PA1196	None	PAS 1	1	149
PA1243	None	PAS 1	1	185
		PAS 2	336	563
PA1261	None	PAS 1	1	120
PA1336	AauS	PAS 1	73	210
		PAS 2	196	315
PA1347	None	PAS 1	23	134
PA1423	BdIA	PAS 1	1	127
		PAS 2	130	218
PA1438	None	PAS 1	41	167

Appendix C

Gene number	Protein name	PAS domain	Start residue	End residue
PA1561	Aer	PAS 1	1	128
PA1611	None	PAS 1	38	169
PA1930	McpS	PAS 1	1	138
		PAS 2	134	319
PA1976	ErcS ¹	PAS 1	101	224
		PAS 2	225	344
		PAS 3	339	456
PA1992	ErcS	PAS 1	36	164
PA2005	HbcR	PAS 1	1	161
PA2072	None	PAS 1	301	420
PA2177	None	PAS 1	58	200
		PAS 2	183	318
PA2449	None	PAS 1	69	232
PA2480	None	PAS 1	30	150
PA2524	CzcS	PAS 1	34	171
PA2652	None	PAS 1	35	207
PA2654	TlpQ	PAS 1	103	242
		PAS 2	243	360
PA2824	SagS	PAS 1	56	172
PA2870	None	PAS 1	97	211
		PAS 2	241	348
PA3044	RocS2	PAS 1	101	243
PA3271	None	PAS 1	601	800
PA3946	RocS1	PAS 1	546	713
PA4021	EatR	PAS 1	80	185
		PAS 2	213	356
PA4036	None	PAS 1	420	555
PA4112	None	PAS 1	336	469
		PAS 2	484	625
		PAS 3	601	748
PA4117	BphP	PAS 1	23	123
PA4147	AcoR	PAS 1	82	192
		PAS 2	215	347
PA4197	BfiS	PAS 1	152	271
		PAS 2	253	388
		PAS 3	381	516
PA4290	None	PAS 1	411	538
PA4293	PprA	PAS 1	303	419
		PAS 2	424	555
		PAS 3	554	700
PA4307	PctC	PAS 1	65	189
		PAS 2	192	279
PA4309	PctA	PAS 1	61	176
		PAS 2	177	279
PA4310	PctB	PAS 1	61	176

Gene number	Protein name	PAS domain	Start residue	End residue
PA4310	PctB	PAS 2	177	278
PA4398	None	PAS 1	50	154
		PAS 2	280	423
PA4546	PilS	PAS 1	186	319
PA4581	RtcR	PAS 1	52	165
PA4601	MorA	PAS 1	291	411
		PAS 2	581	710
		PAS 3	711	847
		PAS 4	825	970
PA4633	None	PAS 1	101	238
		PAS 2	239	349
PA4725	CbrA	PAS 1	622	766
PA4886	None	PAS 1	69	166
PA4959	FimX	PAS 1	266	381
		PAS 2	389	502
PA4961	None	PAS 1	142	254
PA4982	AruS	PAS 1	50	170
PA5017	DipA	PAS 1	9	200
		PAS 2	341	460
PA5124	NtrB	PAS 1	1	116
PA5165	DctB	PAS 1	64	204
		PAS 2	203	293
PA5361	PhoR	PAS 1	95	212
PA5442	None	PAS 1	258	400
		PAS 2	401	520
PA5484	KinB	PAS 1	244	400
PA5512	MifS	PAS 1	59	194
		PAS 2	186	292

List of References

Structural figures were prepared using the UCSF chimera package or PyMOL version 1.3r1^{483,484}.

1. World Health Organization. WHO publishes list of bacteria for which new antibiotics are urgently needed. *WHO* <http://www.who.int/mediacentre/news/releases/2017/> (2017). Available at: <http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/>. (Accessed: 6th March 2017)
2. Hardalo, C. & Edberg, S. C. *Pseudomonas aeruginosa*: assessment of risk from drinking water. *Crit. Rev. Microbiol.* **23**, 47–75 (1997).
3. Stover, C. K. *et al.* Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–64 (2000).
4. Groisman, E. A. & Winfield, M. D. Role of nonhost environments in the lifestyles of *Salmonella* and *E. coli*. *Appl. Environ. Microbiology* **69**, 3687–3694 (2003).
5. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
6. Donlan, R. & Costerton, J. Biofilms : Survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.* **15**, 167–193 (2002).
7. Mah, T. F. & O'Toole, G. a. Mechanisms of biofilm resistance to antimicrobial agents. *Trends Microbiol.* **9**, 34–9 (2001).
8. Prosser, B. L., Taylor, D., Dix, B. A. & Cleeland, R. Method of evaluating effects of antibiotics on bacterial biofilm. *Antimicrob. Agents Chemother.* **31**, 1502–1506 (1987).
9. Nickel, J. C., Ruseska, I., Wright, J. B. & Costerton, J. W. Tobramycin resistance of *Pseudomonas aeruginosa* cells growing as a biofilm on urinary catheter material. *Antimicrob. Agents Chemother.* **27**, 619–624 (1985).
10. Evans, R. C. & Holmes, C. J. Effect of vancomycin hydrochloride on *Staphylococcus epidermidis* biofilm associated with silicone elastomer. *Antimicrob. Agents Chemother.* **31**, 889–894 (1987).
11. Ceri, H. *et al.* The Calgary biofilm device : New technology for rapid determination of antibiotic susceptibilities of bacterial biofilms. *J. Clin. Microbiol.* **37**, 1771–1776 (1999).

Bibliography

12. Vorachit, M., Lam, K., Jayanetra, P. & Costerton, J. W. Resistance of *Pseudomonas pseudomallei* growing as a biofilm on silastic disks to ceftazidime and cotrimoxazole. *Antimicrob. Agents Chemother.* **37**, 2000–2002 (1993).
13. Lyczak, J. B., Cannon, C. L. & Pier, G. B. Lung infections associated with Cystic Fibrosis. *Clin. Microbiol. Rev.* **15**, 194–222 (2002).
14. Costerton, J. W. *et al.* Bacterial biofilms in nature and disease. *Ann. Rev. Microbiol.* **41**, 435–464 (1987).
15. Wingender, J. & Flemming, H. C. Biofilms in drinking water and their role as reservoir for pathogens. *Int. J. Hyg. Environ. Health* **214**, 417–423 (2011).
16. Halan, B., Buehler, K. & Schmid, A. Biofilms as living catalysts in continuous chemical syntheses. *Trends Biotechnol.* **30**, 453–465 (2012).
17. Little, B. J. & Lee, J. S. Microbiologically influenced corrosion: an update. *Int. Mater. Rev.* **59**, 384–393 (2014).
18. de Vos, W. M. Microbial biofilms and the human intestinal microbiome. *npj Biofilms Microbiomes* **1**, 15005 (2015).
19. Meckenstock, R. U. *et al.* Biodegradation: Updating the concepts of control for microbial cleanup in contaminated aquifers. *Environ. Sci. Technol.* **49**, 7073–7081 (2015).
20. Flemming, H.-C. *et al.* Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).
21. Barraud, N. *et al.* Involvement of nitric oxide in biofilm dispersal of *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**, 7344–53 (2006).
22. Barraud, N. *et al.* Cephalosporin-3'-diazeniumdiolates: targeted NO-donor prodrugs for dispersing bacterial biofilms. *Angew. Chem. Int. Ed. Engl.* **51**, 9057–60 (2012).
23. Barraud, N., Kelso, M. J., Rice, S. A. & Kjelleberg, S. Nitric Oxide : A key mediator of biofilm dispersal with applications in infectious diseases. *Curr. Pharm. Des.* **21**, 31–42 (2015).
24. de Beer, D., Stoodley, P. & Lewandowski, Z. Liquid flow in heterogeneous biofilms. *Biotechnol. Bioeng.* **44**, 636–641 (1994).
25. Lawrence, J. R., Korber, D. R., Hoyle, B. D., Costerton, J. W. & Caldwell, D. E. Optical sectioning of microbial biofilms. *J. Bacteriol.* **173**, 6558–6567 (1991).

26. Stoodley, P., Lewandowski, Z., Boyle, J. D. & Lappin-Scott, H. M. Oscillation characteristics of biofilm streamers in turbulent flowing water as related to drag and pressure drop. *Biotechnol. Bioeng.* **57**, 536–544 (1998).
27. Sauer, K., Camper, A. K., Ehrlich, G. D., Costerton, J. W. & Davies, D. G. *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *J. Bacteriol.* **184**, 1140–54 (2002).
28. Hinsa, S. M., Espinosa-Urgel, M., Ramos, J. L. & O'Toole, G. A. Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Mol. Microbiol.* **49**, 905–918 (2003).
29. Park, A. J. *et al.* A temporal examination of the planktonic and biofilm proteome of whole cell *Pseudomonas aeruginosa* PAO1 using quantitative mass spectrometry. *Mol. Cell. Proteomics* **13**, 1095–105 (2014).
30. Monroe, D. Looking for chinks in the armor of bacterial biofilms. *PLoS Biol.* **5**, e307 (2007).
31. Flemming, H. & Wingender, J. The biofilm matrix. *Nat. Rev. Microbiol.* **8**, 623–33 (2010).
32. Schmitt, J., Nivens, D., White, D. & Flemming, H.-C. Changes of biofilm properties in response to sorbed substances — An ftir-atr study. *Water Sci. Technol.* **32**, 149–155 (1995).
33. Whitfield, G. B., Marmont, L. S. & Howell, P. L. Enzymatic modifications of exopolysaccharides enhance bacterial persistence. *Front. Microbiol.* **6**, 1–21 (2015).
34. Nadell, C. D., Drescher, K. & Foster, K. R. Spatial structure, cooperation and competition in biofilms. *Nat. Rev. Microbiol.* **14**, 589–600 (2016).
35. Saville, R. M., Rakshe, S., Haagensen, J. A. J., Shukla, S. & Spormann, A. M. Energy-dependent stability of *Shewanella oneidensis* MR-1 biofilms. *J. Bacteriol.* **193**, 3257–3264 (2011).
36. Tielen, P. *et al.* Interaction between extracellular lipase LipA and the polysaccharide alginate of *Pseudomonas aeruginosa*. *BMC Microbiol.* **13**, 159–171 (2013).
37. Worm, J., Jensen, L. E., Hansen, T. S., Sondergaard, M. & Nybroe, O. Interactions between proteolytic and non-proteolytic *Pseudomonas fluorescens* affect protein degradation in a model community. *FEMS Microbiol. Ecol.* **32**, 103–109 (2000).

Bibliography

38. Roberson, E. B. & Firestone, M. K. Relationship between desiccation and exopolysaccharide production in a soil *Pseudomonas* sp. *Appl. Environ. Microbiol.* **58**, 1284–1291 (1992).
39. Chang, W. S. *et al.* Alginate production by *Pseudomonas putida* creates a hydrated microenvironment and contributes to biofilm architecture and stress tolerance under water-limiting conditions. *J. Bacteriol.* **189**, 8290–8299 (2007).
40. Knowles, E. J. & Castenholz, R. W. Effect of exogenous extracellular polysaccharides on the desiccation and freezing tolerance of rock-inhabiting phototrophic microorganisms. *FEMS Microbiol. Ecol.* **66**, 261–270 (2008).
41. Mann, E. E. & Wozniak, D. J. *Pseudomonas* biofilm matrix composition and niche biology. *FEMS Microbiol. Rev.* **36**, 893–916 (2012).
42. Laue, H. *et al.* Contribution of alginate and levan production to biofilm formation by *Pseudomonas syringae*. *Microbiology* **152**, 2909–18 (2006).
43. Sutherland, I. Biofilm exopolysaccharides: a strong and sticky framework. *Microbiology* **147**, 3–9 (2001).
44. Sutherland, I. The biofilm matrix – an immobilized but dynamic microbial environment. *Trends Microbiol.* **9**, 222–227 (2001).
45. Ma, L. *et al.* Assembly and development of the *Pseudomonas aeruginosa* biofilm matrix. *PLoS Pathog.* **5**, e1000354 (2009).
46. Mishra, M. *et al.* *Pseudomonas aeruginosa* Psl polysaccharide reduces neutrophil phagocytosis and the oxidative response by limiting complement-mediated opsonization. *Cell. Microbiol.* **14**, 95–106 (2012).
47. Colvin, K. M. *et al.* The pel polysaccharide can serve a structural and protective role in the biofilm matrix of *Pseudomonas aeruginosa*. *PLoS Pathog.* **7**, e1001264 (2011).
48. Friedman, L. & Kolter, R. Two genetic loci produce distinct carbohydrate-rich structural components of the *Pseudomonas aeruginosa* biofilm matrix. *J. Bacteriol.* **186**, 4457–4465 (2004).
49. Boucher, J. C., Yu, H., Mudd, M. H. & Deretic, V. Mucoid *Pseudomonas aeruginosa* in cystic fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infect. Immun.* **65**, 3838–46 (1997).

50. Allesen-Holm, M. *et al.* A characterization of DNA release in *Pseudomonas aeruginosa* cultures and biofilms. *Mol. Microbiol.* **59**, 1114–28 (2006).
51. Gloag, E. S. *et al.* Self-organization of bacterial biofilms is facilitated by extracellular DNA. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11541–6 (2013).
52. Davey, M. E. & O'toole, G. a. Microbial biofilms: from ecology to molecular genetics. *Microbiol. Mol. Biol. Rev.* **64**, 847–67 (2000).
53. Nett, J. *et al.* Putative role of beta-1,3 glucans in *Candida albicans* biofilm resistance. *Antimicrob. Agents Chemother.* **51**, 510–20 (2007).
54. Spiers, A. J. & Rainey, P. B. The *Pseudomonas fluorescens* SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity. *Microbiology* **151**, 2829–39 (2005).
55. Coulon, C., Vinogradov, E., Filloux, A. & Sadovskaya, I. Chemical analysis of cellular and extracellular carbohydrates of a biofilm-forming strain *Pseudomonas aeruginosa* PA14. *PLoS One* **5**, e14220 (2010).
56. Tielker, D. *et al.* *Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* **151**, 1313–23 (2005).
57. Abdel-Mawgoud, A. M., Lépine, F. & Déziel, E. Rhamnolipids: diversity of structures, microbial origins and roles. *Appl. Microbiol. Biotechnol.* **86**, 1323–36 (2010).
58. Flemming, H.-C., Neu, T. R. & Wozniak, D. J. The EPS matrix: the 'house of biofilm cells'. *J. Bacteriol.* **189**, 7945–7 (2007).
59. Klausen, M. *et al.* Biofilm formation by *Pseudomonas aeruginosa* wild type, flagella and type IV pili mutants. *Mol. Microbiol.* **48**, 1511–1524 (2003).
60. Meervenne, E. Van *et al.* Biofilm models for the food industry: Hot spots for plasmid transfer? *Pathog. Dis.* **70**, 332–338 (2014).
61. Costerton, J., Lewandowski, Z., Caldwell, D., Korber, D. & Lappin-Scott, H. Microbial biofilms. *Annu. Rev. ...* **49**, 711–745 (1995).
62. Suci, P. A., Mittelman, M. W., Yu, F. P. & Geesey, G. G. Investigation of ciprofloxacin penetration into *Pseudomonas aeruginosa* biofilms. *Antimicrob. Agents Chemother.* **38**, 2125–2133 (1994).

Bibliography

63. Hatch, R. A. & Schiller, N. L. Alginate lyase promotes diffusion of aminoglycosides through the extracellular polysaccharide of mucoid *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **42**, 974–977 (1998).
64. Souli, M. & Giamarellou, H. Effects of slime produced by clinical isolates of Coagulase-negative *Staphylococci* on activities of various antimicrobial agents. *Antimicrob. Agents Chemother.* **42**, 939–941 (1998).
65. Gordon, C. a, Hodges, N. a & Marriott, C. Antibiotic interaction and diffusion through alginate and exopolysaccharide of cystic fibrosis-derived *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **22**, 667–74 (1988).
66. Duguid, I. G., Evans, E., Brown, M. R. & Gilbert, P. Effect of biofilm culture upon the susceptibility of *Staphylococcus epidermidis* to tobramycin. *J. Antimicrob. Chemother.* **30**, 803–10 (1992).
67. Hoyle, B. D., Wong, C. K. W. & Costerton, J. W. Disparate efficacy of tobramycin on Ca^{2+} , Mg^{2+} , and HEPES-treated *Pseudomonas aeruginosa* biofilms. *Can. J. Microbiol.* **38**, 1214–1218 (1992).
68. Dunne, W. M., Mason, E. O. & Kaplan, S. L. Diffusion of rifampin and vancomycin through a *Staphylococcus epidermidis* biofilm. *Antimicrob. Agents Chemother.* **37**, 2522–2526 (1993).
69. Anderl, J. N., Franklin, M. J. & Stewart, P. S. Role of antibiotic penetration limitation in *Klebsiella pneumoniae* biofilm resistance to ampicillin and ciprofloxacin. *Antimicrob. Agents Chemother.* **44**, 1818–1824 (2000).
70. Brown, M. R. W. & Williams, P. Influence of substrate limitation and growth phase on sensitivity to antimicrobial agents. *J. Antimicrob. Chemother.* **15**, 7–14 (1985).
71. Tuomanen, E., Durack, D. T. & Tomasz, A. Antibiotic tolerance among clinical isolates of bacteria. *Antimicrob. Agents Chemother.* **30**, 521–527 (1986).
72. Tuomanen, E., Cozens, R., Tosch, W., Zak, O. & Tomasz, A. The rate of killing of *Escherichia coli* by beta-lactam antibiotics is strictly proportional to the rate of bacterial growth. *J. Gen. Microbiol.* **132**, 1297–304 (1986).
73. Brown, M. R., Allison, D. G. & Gilbert, P. Resistance of bacterial biofilms to antibiotics: a growth-rate related effect? *J. Antimicrob. Chemother.* **22**, 777–783 (1988).

74. Evans, D. J., Allison, D. G., Brown, M. R. & Gilbert, P. Effect of growth-rate on resistance of gram-negative biofilms to ceftrimide. *J. Antimicrob. Chemother.* **26**, 473–8 (1990).
75. Desai, M., Buhler, T., Weller, P. & Brown, M. Increasing resistance of planktonic and biofilm cultures of *Burkholderia cepacia* to ciprofloxacin and ceftazidime during exponential growth. *J. Antimicrob. Chemother.* **42**, 153–160 (1998).
76. Chuard, C., Vaudaux, P., Waldvogel, F. A. & Lew, D. P. Susceptibility of *Staphylococcus aureus* growing on fibronectin-coated surfaces to bactericidal antibiotics. *Antimicrob. Agents Chemother.* **37**, 625–632 (1993).
77. Amorena, B. et al. Antibiotic susceptibility assay for *Staphylococcus aureus* in biofilms developed in vitro. *J. Antimicrob. Chemother.* **44**, 43–55 (1999).
78. Duguid, I. G., Evans, E., Brown, M. R. & Gilbert, P. Growth-rate-independent killing by ciprofloxacin of biofilm-derived *Staphylococcus epidermidis*; evidence for cell-cycle dependency. *J. Antimicrob. Chemother.* **30**, 791–802 (1992).
79. Cochran, W. L., McFeters, G. A. & Stewart, P. S. Reduced susceptibility of thin *Pseudomonas aeruginosa* biofilms to hydrogen peroxide and monochloramine. *J. Appl. Microbiol.* **88**, 22–30 (2000).
80. Adams, J. L. & McLean, R. J. C. Impact of rpoS deletion on *Escherichia coli* biofilms. *Appl. Environ. Microbiol.* **65**, 4285–4287 (1999).
81. Liu, X. & Ferenci, T. Regulation of porin-mediated outer membrane permeability by nutrient limitation in *Escherichia coli*. *J. Bacteriol.* **180**, 3917–3922 (1998).
82. Bagge, N. et al. Dynamics and spatial distribution of β -lactamase expression in *Pseudomonas aeruginosa* biofilms. *Antimicrob. Agents Chemother.* **48**, 1168–1174 (2004).
83. Zhang, L. & Mah, T. F. Involvement of a novel efflux system in biofilm-specific resistance to antibiotics. *J. Bacteriol.* **190**, 4447–4452 (2008).
84. Liao, J., Schurr, M. J. & Sauera, K. The merR-like regulator BrIR confers biofilm tolerance by activating multidrug efflux pumps in *Pseudomonas aeruginosa* biofilms. *J. Bacteriol.* **195**, 3352–3363 (2013).
85. Prigent-Combaret, C. & Brombacher, E. Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the csgD gene. *J. Bacteriol.* **183**, 7213–7223 (2001).

Bibliography

86. O'Toole, G. a & Kolter, R. Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Mol. Microbiol.* **28**, 449–61 (1998).
87. Çetiner, U. *et al.* Tension-activated channels in the mechanism of osmotic fitness in *Pseudomonas aeruginosa*. *J Gen Physiol.* **149**, 595–609 (2017).
88. Jackson, D., Simecka, J. & Romeo, T. Catabolite repression of *Escherichia coli* biofilm formation. *J. Bacteriol.* **184**, 3406–3410 (2002).
89. Sauer, K. *et al.* Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *J. Bacteriol.* **186**, 7312–7326 (2004).
90. O'Toole, G. a, Gibbs, K. a, Hager, P. W., Phibbs, P. V & Kolter, R. The global carbon metabolism regulator Crc is a component of a signal transduction pathway required for biofilm development by *Pseudomonas aeruginosa*. *J. Bacteriol.* **182**, 425–31 (2000).
91. Jackson, D. & Suzuki, K. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. *J. Bacteriol.* **184**, 290–301 (2002).
92. Lam, H. *et al.* D-amino acids govern stationary phase cell wall remodeling in bacteria. *Science* **325**, 1552–1555 (2009).
93. Kolodkin-Gal, I. *et al.* D-amino acids trigger biofilm disassembly. *Science* **328**, 627–629 (2010).
94. Brandenburg, K. S. *et al.* Tryptophan inhibits biofilm formation by *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* **57**, 1921–5 (2013).
95. Singh, P. K., Parsek, M. R., Greenberg, E. P. & Welsh, M. J. A component of innate immunity prevents bacterial biofilm development. *Nature* **417**, 552–5 (2002).
96. Thormann, K. M. *et al.* Control of formation and cellular detachment from *Shewanella oneidensis* MR-1 biofilms by cyclic di-GMP. *J. Bacteriol.* **188**, 2681–2691 (2006).
97. Hickman, J. W., Tifrea, D. F. & Harwood, C. S. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14422–7 (2005).
98. Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–73 (2009).

99. Ross, P., Weinhouse, H. & Aloni, Y. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature* **325**, 279–281 (1987).
100. Römling, U., Gomelsky, M. & Galperin, M. Y. C-di-GMP: the dawning of a novel bacterial signalling system. *Mol. Microbiol.* **57**, 629–39 (2005).
101. Wolfe, A. J. & Visick, K. L. Get the message out: cyclic-Di-GMP regulates multiple levels of flagellum-based motility. *J. Bacteriol.* **190**, 463–75 (2008).
102. Simm, R., Morr, M., Kader, A., Nimtz, M. & Römling, U. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. *Mol. Microbiol.* **53**, 1123–1134 (2004).
103. Jenal, U. & Malone, J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu. Rev. Genet.* **40**, 385–407 (2006).
104. Barraud, N. *et al.* Nitric oxide signaling in *Pseudomonas aeruginosa* biofilms mediates phosphodiesterase activity, decreased cyclic di-GMP levels, and enhanced dispersal. *J. Bacteriol.* **191**, 7333–42 (2009).
105. Chan, C. *et al.* Structural basis of activity and allosteric control of diguanylate cyclase. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17084–9 (2004).
106. Ryjenkov, D. A., Tarutina, M., Moskvina, O. V & Gomelsky, M. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria : insights into biochemistry of the GGDEF protein domain. *J. Bacteriol.* **187**, 1792–1798 (2005).
107. Paul, R. *et al.* Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes Dev.* **18**, 715–27 (2004).
108. Hallberg, Z. F. *et al.* Hybrid promiscuous (Hypr) GGDEF enzymes produce cyclic AMP-GMP (3', 3'-cGAMP). *Proc. Natl. Acad. Sci.* **113**, 1790–1795 (2016).
109. Ryan, R. P. *et al.* Cell-cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6712–7 (2006).
110. Tamayo, R., Tischler, A. D. & Camilli, A. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *J. Biol. Chem.* **280**, 33324–30 (2005).

Bibliography

111. Schmidt, A. J., Ryjenkov, D. A. & Gomelsky, M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *J. Bacteriol.* **187**, 4774–4781 (2005).
112. Orr, M. W. *et al.* Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc. Natl. Acad. Sci.* 201507245 (2015). doi:10.1073/pnas.1507245112
113. Cohen, D. *et al.* Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* 201421450 (2015). doi:10.1073/pnas.1421450112
114. Kulasakara, H. *et al.* Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2839–44 (2006).
115. Delgado-Nixon, V. M., Gonzalez, G. & Gilles-Gonzalez, M. a. Dos, a heme-binding PAS protein from *Escherichia coli*, is a direct oxygen sensor. *Biochemistry* **39**, 2685–91 (2000).
116. Kurokawa, H. *et al.* A redox-controlled molecular switch revealed by the crystal structure of a bacterial heme PAS sensor. *J. Biol. Chem.* **279**, 20186–93 (2004).
117. Armitano, J., Méjean, V. & Jourlin-Castelli, C. Gram-negative bacteria can also form pellicles. *Environ. Microbiol. Rep.* **6**, 534–544 (2014).
118. Merritt, J. H. *et al.* Specific control of *Pseudomonas aeruginosa* surface-associated behaviors by two c-di-GMP diguanylate cyclases. *MBio* **1**, e00183-10 (2010).
119. Amikam, D. & Galperin, M. Y. PilZ domain is part of the bacterial c-di-GMP binding protein. *Bioinformatics* **22**, 3–6 (2006).
120. Ryjenkov, D. a, Simm, R., Römling, U. & Gomelsky, M. The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. *J. Biol. Chem.* **281**, 30310–4 (2006).
121. Merighi, M., Lee, V. T., Hyodo, M., Hayakawa, Y. & Lory, S. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **65**, 876–95 (2007).

122. Pratt, J. T., Tamayo, R., Tischler, A. D. & Camilli, A. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*. *J. Biol. Chem.* **282**, 12860–12870 (2007).
123. Christen, M. *et al.* DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci.* **104**, 4112–4117 (2007).
124. Römling, U., Galperin, M. Y. & Gomelsky, M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol. Mol. Biol. Rev.* **77**, 1–52 (2013).
125. Ko, J. *et al.* Structure of PP4397 reveals the molecular basis for different c-di-GMP binding modes by pilz domain proteins. *J. Mol. Biol.* **398**, 97–110 (2010).
126. Habazettl, J., Allan, M. G., Jenal, U. & Grzesiek, S. Solution structure of the PilZ domain protein PA4608 complex with cyclic di-GMP identifies charge clustering as molecular readout. *J. Biol. Chem.* **286**, 14304–14314 (2011).
127. Li, T.-N. *et al.* A novel tetrameric PilZ domain structure from Xanthomonads. *PLoS One* **6**, e22036 (2011).
128. Gentner, M., Allan, M. G., Zaehring, F., Schirmer, T. & Grzesiek, S. Oligomer formation of the bacterial second messenger c-di-GMP: Reaction rates and equilibrium constants indicate a monomeric state at physiological concentrations. *J. Am. Chem. Soc.* **134**, 1019–1029 (2012).
129. Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J. Biol. Chem.* **280**, 30829–37 (2005).
130. Chatterjee, D. *et al.* Mechanistic insight into the conserved allosteric regulation of periplasmic proteolysis by the signaling molecule cyclic-di-GMP. *Elife* **3**, e03650 (2014).
131. Newell, P. D., Monds, R. D. & O'Toole, G. A. LapD is a bis-(3',5')-cyclic dimer GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0 – 1. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3461–3466 (2009).
132. Navarro, M. V. A. S. *et al.* Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. *PLOS Biol.* **9**, e1000588 (2011).

Bibliography

133. Christen, B. *et al.* Allosteric control of cyclic di-GMP signaling. *J. Biol. Chem.* **281**, 32015–24 (2006).
134. Chin, K.-H. *et al.* The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas campestris*. *J. Mol. Biol.* **396**, 646–62 (2010).
135. Krasteva, P. V *et al.* *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. *Science* **327**, 866–8 (2010).
136. Tao, F., He, Y.-W., Wu, D.-H., Swarup, S. & Zhang, L.-H. The cyclic nucleotide monophosphate domain of *Xanthomonas campestris* global regulator Clp defines a new class of cyclic di-GMP effectors. *J. Bacteriol.* **192**, 1020–9 (2010).
137. Srivastava, D., Harris, R. C. & Waters, C. M. Integration of cyclic di-GMP and quorum sensing in the control of vpsT and aphA in *Vibrio cholerae*. *J. Bacteriol.* **193**, 6331–6341 (2011).
138. Tuckerman, J. R., Gonzalez, G. & Gilles-Gonzalez, M. A. Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J. Mol. Biol.* **407**, 633–639 (2011).
139. Li, W. & He, Z. G. LtmA, a novel cyclic di-GMP-responsive activator, broadly regulates the expression of lipid transport and metabolism genes in *Mycobacterium smegmatis*. *Nucleic Acids Res.* **40**, 11292–11307 (2012).
140. Parvatiyar, K. *et al.* The helicase DDX41 recognizes the bacterial secondary messengers cyclic di-GMP and cyclic di-AMP to activate a type I interferon immune response. *Nat. Immunol.* **13**, 1155–1161 (2012).
141. Shu, C., Yi, G., Watts, T., Kao, C. C. & Li, P. Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat. Struct. Mol. Biol.* **19**, 722–724 (2012).
142. Steiner, S., Lori, C., Boehm, A. & Jenal, U. Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein–protein interaction. *EMBO J.* **32**, 354–368 (2012).
143. Baraquet, C. & Harwood, C. S. Cyclic diguanosine monophosphate represses bacterial flagella synthesis by interacting with the Walker A motif of the enhancer-binding protein FleQ. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18478–83 (2013).

144. An, S. qi *et al.* Novel Cyclic di-GMP Effectors of the YajQ Protein Family Control Bacterial Virulence. *PLoS Pathog.* **10**, (2014).
145. Lolicato, M. *et al.* Cyclic dinucleotides bind the C-linker of HCN4 to control channel cAMP responsiveness. *Nature Chemical Biology* (2014).
146. Lori, C. *et al.* Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* **523**, 236–239 (2015).
147. Dubey, B. N. *et al.* Cyclic di-GMP mediates a histidine kinase/phosphatase switch by noncovalent domain cross-linking. *Sci. Adv.* **2**, e1600823 (2016).
148. Smith, K. D., Shanahan, C. a, Moore, E. L., Simon, A. C. & Strobel, S. a. Structural basis of differential ligand recognition by two classes of bis-(3'-5')-cyclic dimeric guanosine monophosphate-binding riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7757–62 (2011).
149. Sudarsan, N. *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411–3 (2008).
150. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science.* **329**, 845–848 (2010).
151. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35**, 4809 (2007).
152. Remminghorst, U. & Rehm, B. H. a. Alg44, a unique protein required for alginate biosynthesis in *Pseudomonas aeruginosa*. *FEBS Lett.* **580**, 3883–3888 (2006).
153. Hay, I. D., Remminghorst, U. & Rehm, B. H. a. MucR, a novel membrane-associated regulator of alginate biosynthesis in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol.* **75**, 1110–20 (2009).
154. Oglesby, L. L., Jain, S. & Ohman, D. E. Membrane topology and roles of *Pseudomonas aeruginosa* Alg8 and Alg44 in alginate polymerization. *Microbiology* **154**, 1605–15 (2008).
155. Li, Y., Heine, S., Entian, M., Sauer, K. & Frankenberg-Dinkel, N. NO-induced biofilm dispersion in *Pseudomonas aeruginosa* is mediated by an MHYT domain-coupled phosphodiesterase. *J. Bacteriol.* **195**, 3531–42 (2013).
156. Wang, Y., Hay, I. D., Rehman, Z. U. & Rehm, B. H. a. Membrane-anchored MucR mediates nitrate-dependent regulation of alginate production in *Pseudomonas aeruginosa*. *Appl. Microbiol. Biotechnol.* **99**, 7253–7265 (2015).

Bibliography

157. Sondermann, H., Shikuma, N. J. & Yildiz, F. H. You've come a long way: c-di-GMP signaling. *Curr. Opin. Microbiol.* **15**, 140–6 (2012).
158. Botsford, J. L. & Harman, J. G. Cyclic AMP in prokaryotes. *Microbiol Rev* **56**, 100–122 (1992).
159. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
160. Pesavento, C. & Hengge, R. Bacterial nucleotide-based second messengers. *Curr. Opin. Microbiol.* **12**, 170–176 (2009).
161. Hengge-Aronis, R. Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. *Microbiol. Mol. Biol. Rev.* **66**, 373–395 (2002).
162. Yokota, T. & Gots, J. S. Requirement of adenosine 3', 5'-cyclic phosphate for flagella formation in *Escherichia coli* and *Salmonella typhimurium*. *J. Bacteriol.* **103**, 513–516 (1970).
163. Liang, W., Pascual-Montano, A., Silva, A. J. & Benitez, J. A. The cyclic AMP receptor protein modulates quorum sensing, motility and multiple genes that affect intestinal colonization in *Vibrio cholerae*. *Microbiology* **153**, 2964–2975 (2007).
164. Fong, J. C. N. & Yildiz, F. H. Interplay between cyclic AMP-cyclic AMP receptor protein and cyclic di-GMP signaling in *Vibrio cholerae* biofilm formation. *J. Bacteriol.* **190**, 6646–6659 (2008).
165. Kalivoda, E. J., Brothers, K. M., Stella, N. A., Schmitt, M. J. & Shanks, R. M. Q. Bacterial cyclic AMP-phosphodiesterase activity coordinates biofilm formation. *PLoS One* **8**, e71267 (2013).
166. Magnusson, L. U., Farewell, A. & Nyström, T. ppGpp: A global regulator in *Escherichia coli*. *Trends Microbiol.* **13**, 236–242 (2005).
167. Potrykus, K. & Cashel, M. (p)ppGpp: Still Magical? *Annu. Rev. Microbiol.* **62**, 35–51 (2008).
168. Austin, J. W., Sanders, G., Kay, W. W. & Collinson, S. K. Thin aggregative fimbriae enhance *Salmonella enteritidis* biofilm formation. *FEMS Microbiol. Lett.* **162**, 295–301 (1998).
169. Barnhart, M. M. & Chapman, M. R. Curli biogenesis and function. *Annu Rev Microbiol* **60**, 131–147 (2006).

170. Girgis, H. S., Liu, Y., Ryu, W. S. & Tavazoie, S. A comprehensive genetic characterization of bacterial motility. *PLoS Genet.* **3**, 1644–1660 (2007).
171. Durfee, T., Hansen, A. M., Zhi, H., Blattner, F. R. & Ding, J. J. Transcription profiling of the stringent response in *Escherichia coli*. *J. Bacteriol.* **190**, 1084–1096 (2008).
172. Pesavento, C. *et al.* Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. *Genes Dev.* **22**, 2434–2446 (2008).
173. Chávez de Paz, L. E., Lemos, J. A., Wickström, C. & Sedgley, C. M. Role of (p)ppGpp in biofilm formation by *Enterococcus faecalis*. *Appl. Environ. Microbiol.* **78**, 1627–1630 (2012).
174. Sugisaki, K. *et al.* Role of (p)ppGpp in biofilm formation and expression of filamentous structures in *Bordetella pertussis*. *Microbiol. (United Kingdom)* **159**, 1379–1389 (2013).
175. Corrigan, R. M., Abbott, J. C., Burhenne, H., Kaeffer, V. & Gründling, A. C-di-AMP is a new second messenger in *Staphylococcus aureus* with a role in controlling cell size and envelope stress. *PLoS Pathog.* **7**, e1002217 (2011).
176. Peng, X., Zhang, Y., Bai, G., Zhou, X. & Wu, H. Cyclic di-AMP mediates biofilm formation. *Mol. Microbiol.* **99**, 945–959 (2016).
177. An, S.-Q. *et al.* A cyclic GMP-dependent signalling pathway regulates bacterial phytopathogenesis. *EMBO J.* **32**, 2430–2438 (2013).
178. Schirmer, T. & Jenal, U. Structural and mechanistic determinants of c-di-GMP signalling. *Nat. Rev. Microbiol.* **7**, 724–35 (2009).
179. Wassmann, P. *et al.* Structure of BeF₃⁻-modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure* **15**, 915–27 (2007).
180. Schirmer, T. C-di-GMP Synthesis: Structural aspects of evolution, catalysis and regulation. *J. Mol. Biol.* **428**, 3683–3701 (2016).
181. Tews, I. *et al.* The structure of a pH-sensing mycobacterial adenylyl cyclase holoenzyme. *Science* **308**, 1020 LP-1023 (2005).
182. Paul, R. *et al.* Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization. *J. Biol. Chem.* **282**, 29170–7 (2007).

Bibliography

183. Zähnger, F., Lacanna, E., Jenal, U., Schirmer, T. & Boehm, A. Structure and signaling mechanism of a zinc-sensory diguanylate cyclase. *Structure* **21**, 1149–1157 (2013).
184. Tesmer, J. J. *et al.* Two-metal-ion catalysis in adenylyl cyclase. *Science* **285**, 756–760 (1999).
185. Oliveira, M. C. *et al.* Cooperative substrate binding by a diguanylate cyclase. *J. Mol. Biol.* **427**, 415–432 (2015).
186. Tarnawski, M., Barends, T. R. M. & Schlichting, I. Structural analysis of an oxygen-regulated diguanylate cyclase. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 2158–2177 (2015).
187. Sawai, H. *et al.* Molecular oxygen regulates the enzymatic activity of a heme-containing diguanylate cyclase (HemDGC) for the synthesis of cyclic di-GMP. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 166–172 (2010).
188. Rao, F. *et al.* Enzymatic synthesis of c-di-GMP using a thermophilic diguanylate cyclase. *Anal. Biochem.* **389**, 138–142 (2009).
189. Boyd, C. D. & O'Toole, G. A. Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. *Annu. Rev. Cell Dev. Biol.* **28**, 439–62 (2012).
190. De, N., Navarro, M. V. A. S., Raghavan, R. V. & Sondermann, H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. *J. Mol. Biol.* **393**, 619–633 (2009).
191. Kitanishi, K. *et al.* Important roles of Tyr43 at the putative heme distal side in the oxygen recognition and stability of the Fe(II)-O₂ complex of YddV, a globin-coupled heme-based oxygen sensor diguanylate cyclase. *Biochemistry* **49**, 10381–10393 (2010).
192. Schaller, R. A., Ali, S. K., Klose, K. E. & Kurtz, D. M. A bacterial hemerythrin domain regulates the activity of a *Vibrio cholerae* diguanylate cyclase. *Biochemistry* **51**, 8563–8570 (2012).
193. Giardina, G. *et al.* Investigating the allosteric regulation of YfiN from *Pseudomonas aeruginosa*: clues from the structure of the catalytic domain. *PLoS One* **8**, e81324 (2013).
194. Rao, F., Yang, Y., Qi, Y. & Liang, Z.-X. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. *J. Bacteriol.* **190**, 3622–31 (2008).

195. Tarnawski, M., Barends, T. R. M., Hartmann, E. & Schlichting, I. Structures of the catalytic EAL domain of the *Escherichia coli* direct oxygen sensor. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1045–1053 (2013).
196. Sundriyal, A. *et al.* Inherent regulation of EAL domain-catalyzed hydrolysis of second messenger cyclic di-GMP. *J. Biol. Chem.* **289**, 6978–90 (2014).
197. Winkler, A. *et al.* Characterization of elements involved in allosteric light regulation of phosphodiesterase activity by comparison of different functional BlrP1 states. *J. Mol. Biol.* **426**, 853–868 (2014).
198. Barends, T. R. M. *et al.* Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* **459**, 1015–8 (2009).
199. Minasov, G. *et al.* Crystal structures of YkuL and its complex with second messenger cyclic Di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains. *J. Biol. Chem.* **284**, 13174–84 (2009).
200. Navarro, M. V. A. S., De, N., Bae, N., Wang, Q. & Sondermann, H. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX. *Structure* **17**, 1104–1116 (2009).
201. Tchigvintsev, A. *et al.* Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases. *J. Mol. Biol.* **402**, 524–38 (2010).
202. Phippen, C. W. *et al.* Formation and dimerization of the phosphodiesterase active site of the *Pseudomonas aeruginosa* MorA, a bi-functional c-di-GMP regulator. *FEBS Lett.* **588**, 4631–4636 (2014).
203. Rao, F. *et al.* The functional role of a conserved loop in EAL domain-based cyclic di-GMP-specific phosphodiesterase. *J. Bacteriol.* **191**, 4722–31 (2009).
204. Galperin, M. Y., Gaidenko, T. a, Mulkidjanian, a Y., Nakano, M. & Price, C. W. MHYT, a new integral membrane sensor domain. *FEMS Microbiol. Lett.* **205**, 17–23 (2001).
205. Galperin, M. Y., Natale, D. a, Aravind, L. & Koonin, E. V. A specialized version of the HD hydrolase domain implicated in signal transduction. *J. Mol. Microbiol. Biotechnol.* **1**, 303–5 (1999).
206. Plate, L. & Marletta, M. A. Nitric oxide modulates bacterial biofilm formation through a multicomponent cyclic-di-GMP signaling network. *Mol. Cell* **46**, 449–460 (2012).

Bibliography

207. Miner, K. D., Klose, K. E. & Kurtz, D. M. An HD-GYP cyclic di-guanosine monophosphate phosphodiesterase with a non-heme diiron-carboxylate active site. *Biochemistry* **52**, 5329–5331 (2013).
208. Bellini, D. *et al.* Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol. Microbiol.* **91**, 26–38 (2013).
209. Stelitano, V. *et al.* C-di-GMP hydrolysis by *Pseudomonas aeruginosa* HD-GYP phosphodiesterases: analysis of the reaction mechanism and novel roles for pGpG. *PLoS One* **8**, e74920 (2013).
210. Lovering, A. L., Capeness, M. J., Lambert, C., Hobley, L. & Sockett, R. E. The structure of an unconventional HD-GYP protein from *Bdellovibrio* reveals the roles of conserved residues in this class of Cyclic-di-GMP phosphodiesterases. *MBio* **2**, 1–8 (2011).
211. Rinaldo, S. *et al.* Structural basis of functional diversification of the HD-GYP domain revealed by the *Pseudomonas aeruginosa* PA4781 protein, which displays an unselective bimetallic binding site. *J. Bacteriol.* **197**, 1525–1535 (2015).
212. Miner, K. D. & Kurtz, D. M. Active site metal occupancy and cyclic di-GMP phosphodiesterase activity of *Thermotoga maritima* HD-GYP. *Biochemistry* **55**, 970–979 (2016).
213. Datta, A. K. & Niyogi, S. K. A novel oligoribonuclease of *Escherichia coli*. *J. Biol. Chem.* **250**, 7313–7319 (1975).
214. Niyogi, S. K. & Datta, A. K. A novel oligoribonuclease of *Escherichia coli*. *J. Biol. Chem.* **250**, 7307–7312 (1975).
215. Zuo, Y. & Deutscher, M. P. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* **29**, 1017–1026 (2001).
216. Rezácová, P., Borek, D., Moy, S. F., Joachimiak, A. & Otwinowski, Z. Crystal structure and putative function of small Toprim domain-containing protein from *Bacillus stearothermophilus*. *Proteins* **70**, 311–319 (2008).
217. Zuo, Y. & Deutscher, M. P. Mechanism of action of RNase T: I. Identification of residues required for catalysis, substrate binding, and dimerization. *J. Biol. Chem.* **277**, 50155–50159 (2002).

218. Young Park, A. *et al.* Hydrolysis of the 5'-p-nitrophenyl ester of TMP by oligoribonucleases (ORN) from *Escherichia coli*, *Mycobacterium smegmatis*, and human. *Protein Expr. Purif.* **57**, 180–187 (2008).
219. Hamdan, S. *et al.* Hydrolysis of the 5'-p-nitrophenyl ester of TMP by the proofreading exonuclease (epsilon) subunit of *Escherichia coli* DNA polymerase III. *Biochemistry* **41**, 5266–5275 (2002).
220. Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L. & Dixon, N. E. Structural basis for proofreading during replication of the *Escherichia coli* chromosome. *Structure* **10**, 535–546 (2002).
221. Cheng, Y. & Patel, D. J. Crystallographic structure of the nuclease domain of 3'hExo, a DEDDh family member, bound to rAMP. *J. Mol. Biol.* **343**, 305–312 (2004).
222. Zuo, Y. *et al.* Crystal Structure of RNase T, an exoribonuclease involved in tRNA maturation and end-turnover. *Structure* **15**, 417–428 (2007).
223. Liu, N. *et al.* Nitric oxide regulation of cyclic di-GMP synthesis and hydrolysis in *Shewanella woodyi*. *Biochemistry* **51**, 2087–99 (2012).
224. Boon, E. M. & Marletta, M. a. Ligand discrimination in soluble guanylate cyclase and the H-NOX family of heme sensor proteins. *Curr. Opin. Chem. Biol.* **9**, 441–6 (2005).
225. Boon, E. M., Huang, S. H. & Marletta, M. a. A molecular basis for NO selectivity in soluble guanylate cyclase. *Nat. Chem. Biol.* **1**, 53–9 (2005).
226. Boon, E. M. & Marletta, M. a. Sensitive and selective detection of nitric oxide using an H-NOX domain. *J. Am. Chem. Soc.* **128**, 10022–3 (2006).
227. Huang, Z., Edery, I. & Rosbash, M. PAS is a dimerization domain common to Drosophila period and several transcription factors. *Nature* **364**, 259–262 (1993).
228. Möglich, A., Ayers, R. A. & Moffat, K. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17**, 1282–1294 (2009).
229. Ortmayer, M. *et al.* An oxidative N-demethylase reveals PAS transition from ubiquitous sensor to enzyme. *Nature* **539**, 593–597 (2016).
230. Hefti, M. H., François, K.-J., De Vries, S. C., Dixon, R. & Vervoort, J. The PAS fold. *Eur. J. Biochem.* **271**, 1198–1208 (2004).

Bibliography

231. Zhulin, I., Taylor, B. & Dixon, R. PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox. *Trends Biochem. Sci.* **22**, 331–333 (1997).
232. Ponting, C. & Aravind, L. PAS: a multifunctional domain family comes to light. *Curr. Biol.* **7**, 674–677 (1997).
233. Pellequer, J. L., Wager-Smith, K. a, Kay, S. a & Getzoff, E. D. Photoactive yellow protein: a structural prototype for the three-dimensional fold of the PAS domain superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5884–90 (1998).
234. Cheung, J. & Hendrickson, W. A. Crystal structures of C4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB. *J. Biol. Chem.* **283**, 30256–30265 (2008).
235. Shah, N. *et al.* Reductive evolution and the loss of PDC/PAS domains from the genus *Staphylococcus*. *BMC Genomics* **14**, 524 (2013).
236. Henry, J. T. & Crosson, S. Ligand-binding PAS domains in a genomic, cellular, and structural context. *Annu. Rev. Microbiol.* **65**, 261–86 (2011).
237. Galperin, M. Y. Bacterial signal transduction in a genomic perspective. *Environ. Microbiol.* **6**, 552–567 (2004).
238. Pongratz, I., Antonsson, C., Whitelaw, M. L. & Poellinger, L. Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol. Cell. Biol.* **18**, 4079–88 (1998).
239. Taylor, B. L. & Zhulin, I. B. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63**, 479–506 (1999).
240. Fedorov, R. *et al.* Crystal structures and molecular mechanism of a light-induced signaling switch: The Phot-LOV1 domain from *Chlamydomonas reinhardtii*. *Biophys. J.* **84**, 2474–2482 (2003).
241. Card, P. B., Erbel, P. J. A. & Gardner, K. H. Structural basis of ARNT PAS-B dimerization: Use of a common beta-sheet interface for hetero- and homodimerization. *J. Mol. Biol.* **353**, 664–677 (2005).
242. Key, J., Hefti, M., Purcell, E. B. & Moffat, K. Structure of the redox sensor domain of *Azotobacter vinelandii* NifL at atomic resolution: Signaling, dimerization, and mechanism. *Biochemistry* **46**, 3614–3623 (2007).

243. Ayers, R. A. & Moffat, K. Changes in quaternary structure in the signalling mechanisms of PAS domains. *Biochemistry* **47**, 12078–12086 (2008).
244. Lee, J. *et al.* Changes at the KinA PAS-A dimerization interface influence histidine kinase function. *Biochemistry* **47**, 4051–4064 (2008).
245. Ma, X., Sayed, N., Baskaran, P., Beuve, A. & Van Den Akker, F. PAS-mediated dimerization of soluble guanylyl cyclase revealed by signal transduction histidine kinase domain crystal structure. *J. Biol. Chem.* **283**, 1167–1178 (2008).
246. Gong, W. *et al.* Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 15177–15182 (1998).
247. Pokkuluri, P. R. *et al.* Structures and solution properties of two novel periplasmic sensor domains with c-type heme from chemotaxis proteins of *Geobacter sulfurreducens*: implications for signal transduction. *J. Mol. Biol.* **377**, 1498–517 (2008).
248. Gong, W., Hao, B. & Chan, M. K. New mechanistic insights from structural studies of the oxygen-sensing domain of *Bradyrhizobium japonicum* FixL. *Biochemistry* **39**, 3955–62 (2000).
249. Hao, B., Isaza, C., Arndt, J., Soltis, M. & Chan, M. K. Structure-based mechanism of O₂ sensing and ligand discrimination by the FixL heme domain of *Bradyrhizobium japonicum*. *Biochemistry* **41**, 12952–12958 (2002).
250. Miyatake, H. *et al.* Sensory mechanism of oxygen sensor FixL from *Rhizobium meliloti*: crystallographic, mutagenesis and resonance Raman spectroscopic studies. *J. Mol. Biol.* **301**, 415–31 (2000).
251. Airola, M. V., Watts, K. J., Bilwes, A. M. & Crane, B. R. Structure of concatenated HAMP domains provides a mechanism for signal transduction. *Structure* **18**, 436–448 (2010).
252. Airola, M. V. *et al.* Architecture of the soluble receptor aer2 indicates an in-line mechanism for PAS and HAMP domain signaling. *J. Mol. Biol.* **425**, 886–901 (2013).
253. Gilles-Gonzalez, M. A., Ditta, G. S. & Helinski, D. R. A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*. *Nature* **350**, 170–172 (1991).
254. Sciotti, M. A., Chanfon, A., Hennecke, H. & Fischer, H. M. Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in *Bradyrhizobium japonicum*. *J. Bacteriol.* **185**, 5639–5642 (2003).

Bibliography

255. Fischer, H. M. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58**, 352–386 (1994).
256. David, M. *et al.* Cascade regulation of nif gene expression in *Rhizobium meliloti*. *Cell* **54**, 671–683 (1988).
257. Dunham, C. M. *et al.* A distal arginine in oxygen-sensing heme-PAS domains is essential to ligand binding, signal transduction, and structure. *Biochemistry* **42**, 7701–7708 (2003).
258. Yamawaki, T. *et al.* Regulatory implications of structural changes in Tyr201 of the oxygen sensor protein FixL. *Biochemistry* **55**, 4027–4035 (2016).
259. Akimoto, S., Tanaka, A., Nakamura, K., Shiro, Y. & Nakamura, H. O₂-specific regulation of the ferrous heme-based sensor kinase FixL from *Sinorhizobium meliloti* and its aberrant inactivation in the ferric form. *Biochem. Biophys. Res. Commun.* **304**, 136–142 (2003).
260. Hiruma, Y., Kikuchi, A., Tanaka, A., Shiro, Y. & Mizutani, Y. Resonance Raman observation of the structural dynamics of FixL on signal transduction and ligand discrimination. *Biochemistry* **46**, 6086–6096 (2007).
261. Yamada, S. *et al.* Structure of PAS-linked histidine kinase and the response regulator complex. *Structure* **17**, 1333–1344 (2009).
262. Yano, S. *et al.* Ultraviolet resonance raman observations of the structural dynamics of rhizobial oxygen sensor FixL on ligand recognition. *J. Phys. Chem. B* **117**, 15786–15791 (2013).
263. Key, J., Šrajer, V., Pahl, R. & Moffat, K. Time-resolved crystallographic studies of the heme domain of the oxygen sensor FixL: Structural dynamics of ligand rebinding and their relation to signal transduction. *Biochemistry* **46**, 4706–4715 (2007).
264. Rivera-Cancel, G., Ko, W., Tomchick, D. R., Correa, F. & Gardner, K. H. Full-length structure of a monomeric histidine kinase reveals basis for sensory regulation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 17839–44 (2014).
265. Watts, K. J., Taylor, B. L. & Johnson, M. S. PAS/poly-HAMP signalling in Aer-2, a soluble haem-based sensor. *Mol. Microbiol.* **79**, 686–699 (2011).
266. Sawai, H. *et al.* Structural basis for oxygen sensing and signal transduction of the heme-based sensor protein Aer2 from *Pseudomonas aeruginosa*. *Chem. Commun.* **48**, 6523–6525 (2012).

267. Güvener, Z. T., Tifrea, D. F. & Harwood, C. S. Two different *Pseudomonas aeruginosa* chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. *Mol. Microbiol.* **61**, 106–118 (2006).
268. Ferrández, A. *et al.* Cluster II che genes from *Pseudomonas aeruginosa* are required for an optimal chemotactic response. *J. Bacteriol.* **184**, 4374–4383 (2002).
269. Hong, C. S. *et al.* Chemotaxis proteins and transducers for aerotaxis in *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* **231**, 247–252 (2004).
270. Hong, C. S., Kuroda, A., Takiguchi, N., Ohtake, H. & Kato, J. Expression of *Pseudomonas aeruginosa* aer-2, one of two aerotaxis transducer genes, is controlled by RpoS. *J. Bacteriol.* **187**, 1533–1535 (2005).
271. Aravind, L. & Ponting, C. P. The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol. Lett.* **176**, 111–116 (1999).
272. Borkovich, K. A., Kaplan, N., Hess, J. F. & Simon, M. I. Transmembrane signal transduction in bacterial chemotaxis involves ligand-dependent activation of phosphate group transfer. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 1208–1212 (1989).
273. McNally, D. F. & Matsumura, P. Bacterial chemotaxis signaling complexes: Formation of a CheA/CheW complex enhances autophosphorylation and affinity for CheY. *Biochemistry* **88**, 6269–6273 (1991).
274. Gegner, J. A., Graham, D. R., Roth, A. F. & Dahlquist, F. W. Assembly of an MCP receptor, CheW, and kinase CheA complex in the bacterial chemotaxis signal transduction pathway. *Cell* **70**, 975–982 (1992).
275. Erbse, A. H. & Falke, J. J. The core signaling proteins of bacterial chemotaxis assemble to form an ultrastable complex. *Biochemistry* **48**, 6975–6987 (2009).
276. Borkovich, K. A. & Simon, M. I. The dynamics of protein phosphorylation in bacterial chemotaxis. *Cell* **63**, 1339–1348 (1990).
277. Falke, J. J. & Hazelbauer, G. L. Transmembrane signaling in bacterial chemoreceptors. *Trends Biochem. Sci.* **26**, 257–265 (2001).

Bibliography

278. Swanson, R. V, Schuster, S. C. & Simon, M. I. Expression of CheA fragments which define domains encoding kinase, phosphotransfer, and CheY binding activities. *Biochemistry* **32**, 7623–7629 (1993).
279. Welch, M., Oosawa, K., Aizawa, S. & Eisenbach, M. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8787-8791 (1993).
280. Bren, A. & Eisenbach, M. The N terminus of the flagellar switch protein, FlIM, is the binding domain for the chemotactic response regulator, CheY. *J. Mol. Biol.* **278**, 507–514 (1998).
281. Kato, J., Nakamura, T., Kuroda, A. & Ohtake, H. Cloning and characterization of chemotaxis genes in *Pseudomonas aeruginosa*. *Biosci Biotechnol Biochem* **63**, 155–161 (1999).
282. Rao, F., Ji, Q., Soehano, I. & Liang, Z.-X. Unusual heme-binding PAS domain from YybT family proteins. *J. Bacteriol.* **193**, 1543–51 (2011).
283. Ceulemans, A., Oldenhof, W., Gorller-Walrand, C. & Vanquickenborne, L. G. Gouterman's 'four-orbital' model and the MCD spectra of high-symmetry metalloporphyrins. *J. Am. Chem. Soc.* **108**, 1155–1163 (1986).
284. Gouterman, M. Study of the effects of substitution on the absorption spectra of porphin. *J. Chem. Phys.* **30**, 1139 (1959).
285. Nappa, M. & Valentine, J. The influence of axial ligands on metalloporphyrin visible absorption spectra. Complexes of tetraphenylporphinatozinc. *J. Am. Chem. Soc.* **100**, 5075–5080 (1978).
286. Morgan, R., Kohn, S., Hwang, S.-H., Hassett, D. J. & Sauer, K. BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. *J. Bacteriol.* **188**, 7335–43 (2006).
287. Petrova, O. E. & Sauer, K. Dispersion by *Pseudomonas aeruginosa* requires an unusual posttranslational modification of BdlA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16690–5 (2012).
288. Petrova, O. E. & Sauer, K. PAS domain residues and prosthetic group involved in BdlA-dependent dispersion response by *Pseudomonas aeruginosa* biofilms. *J. Bacteriol.* **194**, 5817–28 (2012).
289. Roy, A. B., Petrova, O. E. & Sauer, K. The phosphodiesterase DipA (PA5017) is essential for *Pseudomonas aeruginosa* biofilm dispersion. *J. Bacteriol.* **194**, 2904–15 (2012).

290. An, S., Wu, J. & Zhang, L.-H. Modulation of *Pseudomonas aeruginosa* biofilm dispersal by a cyclic-di-GMP phosphodiesterase with a putative hypoxia-sensing domain. *Appl. Environ. Microbiol.* **76**, 8160–73 (2010).
291. Thompson, M. W. & Maurizi, M. R. Activity and specificity of *Escherichia coli* ClpAP protease in cleaving model peptide substrates. *J. Biol. Chem.* **269**, 18201–18208 (1994).
292. Petrova, O. E., Cherny, K. E. & Sauer, K. The diguanylate cyclase GcbA facilitates *Pseudomonas aeruginosa* biofilm dispersion by activating BdlA. *J. Bacteriol.* **197**, 174–187 (2015).
293. Wood, S. R. *et al.* Nitrosative stress inhibits production of the virulence factor alginate in mucoid *Pseudomonas aeruginosa*. *Free Radic. Res.* **41**, 208–215 (2007).
294. Remminghorst, U. & Rehm, B. H. a. In vitro alginate polymerization and the functional role of Alg8 in alginate production by *Pseudomonas aeruginosa*. **72**, 298–305 (2006).
295. Remminghorst, U., Hay, I. D. & Rehm, B. H. a. Molecular characterization of Alg8, a putative glycosyltransferase, involved in alginate polymerisation. *J. Biotechnol.* **140**, 176–183 (2009).
296. Owens, R. J. <https://www.addgene.org/41125/>. *Addgene*
297. Gasteiger, E. *et al.* in *The proteomics protocols handbook*, Humana Press 571–607 (2005).
298. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Sixth Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).
299. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
300. Sigma Aldrich. Buffer reference center. <http://www.sigmaaldrich.com/life-science/core-bior> Available at: <http://www.sigmaaldrich.com/life-science/core-bioreagents/biological-buffers/learning-center/buffer-reference-center.html>. (Accessed: 6th July 2017)
301. Bellini, D. *et al.* Dimerisation induced formation of the active site and the identification of three metal sites in EAL-phosphodiesterases. *Sci. Rep.* **7**, 42166 (2017).
302. Zhang, X. & Liu, Z. Superlenses to overcome the diffraction limit. *Nat. Mater.* **7**, 435–441 (2008).

Bibliography

- 303. Schropp, A. *et al.* Hard X-ray scanning microscopy with coherent radiation : Beyond the resolution of conventional X-ray microscopes. *Appl. Phys. Lett.* **100**, 253112 (2012).
- 304. Durbin, S. D. & Feher, G. Protein crystallization. *Annu. Rev. Phys. Chem.* **47**, 171–204 (1996).
- 305. Dunlop, K. V & Hazes, B. When less is more: a more efficient vapour- diffusion protocol. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1797–1800 (2003).
- 306. Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24**, 409–411 (2000).
- 307. Carter, C. W. & Carter, C. W. Protein crystallization using incomplete factorial experiments. *J. Biol. Chem.* **254**, 12219–12223 (1979).
- 308. Garman, E. F. & Schneider, T. R. Macromolecular cryocrystallography. *J. Appl. Crystallogr.* **30**, 211–237 (1997).
- 309. Source, D. L. <http://www.diamond.ac.uk/Home/About/How-Diamond-Works.html>.
- 310. Helliwell, J. R. Synchrotron x-radiation protein crystallography: instrumentation , methods and applications. *Reports Prog. Phys.* **47**, 1403–1497 (1984).
- 311. Synchrotron, E. <http://www.esrf.eu/about/synchrotron-science/synchrotron>.
- 312. Bragg, W. H. & Bragg, W. L. The reflection of X-rays by crystals. *Proc. R. Soc. London . Ser. A* **88**, 428–438 (1913).
- 313. Bragg, W. H. Bakerian lecture: X-rays and crystal structure. *Philos. Trans. R. Soc. London. Ser. A* **215**, 253–274 (1915).
- 314. Rhodes, G. *Crystallography made crystal clear*. (2006).
doi:10.1017/CBO9781107415324.004
- 315. Ewald, P. P. X-ray diffraction by finite and imperfect crystal lattices. *Proc. Phys. Soc.* **52**, 167 (1940).
- 316. Rossman, M. G. The molecular replacement method. *Acta Crystallogr. Sect. A* **46**, 73–82 (1990).
- 317. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 22–25 (2010).

318. Okaya, Y., Saito, Y. & Pepinsky, R. New method in X-ray crystal structure determination involving the use of anomalous dispersion. *Phys. Rev.* **98**, 1857 (1955).
319. Bijvoet, J. M. Phase determination in direct Fourier-synthesis of crystal structures. *Prok K Ned Akad Wet B* **52**, 313–314 (1949).
320. Bijvoet, J. M. Structure of optically active compounds in the solid state. *Nature* **173**, 888–891 (1954).
321. Peerdeman, A. F. & Bijvoet, J. M. The indexing of reflexions in investigations involving the use of the anomalous scattering effect. *Acta Crystallogr. Sect. A* **9**, 1012–1015 (1956).
322. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. Sect. A Found. Crystallogr.* **42**, 140–149 (1986).
323. Brünger, A. T. Free R value - a novel statistical quantity for assessing the accuracy of crystal-structures. *Nature* **355**, 472–475 (1992).
324. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18**, 581–6 (2008).
325. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination - the next generation. *Biochim. Biophys. Acta* **1838**, 78–87 (2014).
326. Liu, Y., Engelman, D. M. & Gerstein, M. Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol.* **3**, research0054 (2002).
327. Bird, L. E. *et al.* Green fluorescent protein-based expression screening of membrane proteins in *Escherichia coli*. *J. Vis. Exp* **95**, 1–7 (2015).
328. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–81 (2006).
329. Lewinson, O., Lee, A. T. & Rees, D. C. The funnel approach to the precrystallization production of membrane proteins. *J. Mol. Biol.* **377**, 62–73 (2008).
330. Hammon, J., Palanivelu, D. V., Chen, J., Patel, C. & Minor, D. L. A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. *Protein Sci.* **18**, 121–133 (2009).

Bibliography

331. Barth, H. G., Boyes, B. E. & Jackson, C. Size exclusion chromatography. *Anal. Chem.* **66**, 595R–620R (1994).
332. Sonoda, Y. *et al.* Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* **19**, 17–25 (2011).
333. Mancusso, R., Karpowich, N. K., Czyzewski, B. K. & Wang, D.-N. Simple screening method for improving membrane protein thermostability. *Methods* **55**, 324–9 (2011).
334. Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem. Sci.* **37**, 343–352 (2012).
335. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. a. & Stevens, R. C. Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**, 351–359 (2008).
336. Tsirigos, K. D., Peters, C., Shu, N., Kall, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407 (2015).
337. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
338. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1260–1273 (2013).
339. Kabsch, W. Xds. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 125–132 (2010).
340. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 72–82 (2006).
341. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
342. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
343. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).
344. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 355–367 (2011).

345. Chen, V. B. *et al.* MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
346. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
347. Kuppuraj, G., Dudev, M. & Lim, C. Factors governing metal-ligand distances and coordination geometries of metal complexes. *J. Physocal Chem.* **113**, 2952–2960 (2009).
348. Robert-Paganin, J., Nonin-Lecomte, S. & Réty, S. Crystal structure of an EAL domain in complex with reaction product 5'-pGpG. *PLoS One* **7**, e52424 (2012).
349. Kulasekara, H. D. *et al.* A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol. Microbiol.* **55**, 368–380 (2005).
350. Liao, R. Z., Yu, J. G. & Himo, F. Phosphate mono- and diesterase activities of the trinuclear zinc enzyme nuclease P1 - Insights from quantum chemical calculations. *Inorg. Chem.* **49**, 6883–6888 (2010).
351. Ivanov, I., Tainer, J. a & McCammon, J. A. Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease IV. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1465–1470 (2007).
352. Romier, C., Dominguez, R., Lahm, A., Dahl, O. & Suck, D. Recognition of single-stranded DNA by nuclease P1: High resolution crystal structures of complexes with substrate analogs. *Proteins Struct. Funct. Genet.* **32**, 414–424 (1998).
353. Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell* **98**, 397–408 (1999).
354. Mol, C. D., Hosfield, D. J. & Tainer, J. a. Abasic site recognition by two apurinicapyrimidinic endonuclease families in DNA base excision repair: the 3' ends justify the means. *Mutat. Res.* **460**, 211–229 (2000).
355. Syson, K. *et al.* Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J. Biol. Chem.* **283**, 28741–28746 (2008).

Bibliography

356. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15 (2016).
357. Adler, J. Chemotaxis in bacteria. *Science* **153**, 708–16 (1966).
358. Drake, D. & Montie, T. C. Flagella, motility and invasive virulence of *Pseudomonas aeruginosa*. *Microbiology* **134**, 43–52 (1988).
359. Barken, K. B. *et al.* Roles of type IV pili, flagellum-mediated motility and extracellular DNA in the formation of mature multicellular structures in *Pseudomonas aeruginosa* biofilms. *Environ. Microbiol.* **10**, 2331–2343 (2008).
360. Schmidt, J. *et al.* The *Pseudomonas aeruginosa* chemotaxis methyltransferase CheR1 impacts on bacterial surface sampling. *PLoS One* **6**, e18184 (2011).
361. McDougald, D., Rice, S. a, Barraud, N., Steinberg, P. D. & Kjelleberg, S. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. *Nat. Rev. Microbiol.* **10**, 39–50 (2012).
362. Shitashiro, M. *et al.* Evaluation of bacterial aerotaxis for its potential use in detecting the toxicity of chemicals to microorganisms. *J. Biotechnol.* **101**, 11–18 (2003).
363. Kuroda, A. *et al.* Molecular cloning and characterization of a chemotactic transducer gene in *Pseudomonas aeruginosa*. *J. Bacteriol.* **177**, 7019–7025 (1995).
364. Taguchi, K., Fukutomi, H., Kuroda, A., Kato, J. & Ohtake, H. Genetic identification of chemotactic transducers for amino acids in *Pseudomonas aeruginosa*. *Microbiology* **143**, 3223–3229 (1997).
365. Miller, R. M. *et al.* *Pseudomonas aeruginosa* twitching motility-mediated chemotaxis towards phospholipids and fatty acids: Specificity and metabolic requirements. *J. Bacteriol.* **190**, 4038–4049 (2008).
366. Alexander, R. P. & Zhulin, I. B. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2885–2890 (2007).
367. Kato, J., Kim, H.-E., Takiguchi, N., Kuroda, A. & Ohtake, H. *Pseudomonas aeruginosa* as a model microorganism for investigation of chemotactic behaviors in ecosystem. *J. Biosci. Bioeng.* **106**, 1–7 (2008).

368. Masduki, A. *et al.* Isolation and characterization of chemotaxis mutants and genes of *Pseudomonas aeruginosa*. *J. Bacteriol.* **177**, 948–952 (1995).
369. Darzins, A. The pilG gene product, required for *Pseudomonas aeruginosa* pilus production and twitching motility, is homologous to the enteric, single- domain response regulator CheY. *J. Bacteriol.* **175**, 5934–5944 (1993).
370. Darzins, A. Characterization of a *Pseudomonas aeruginosa* gene cluster involved in pilus biosynthesis and twitching motility: Sequence similarity to the chemotaxis proteins of enterics and the gliding bacterium *Myxococcus xanthus*. *Mol. Microbiol.* **11**, 137–153 (1994).
371. Whitchurch, C. B. *et al.* Characterization of a complex chemosensory signal transduction system which controls twitching motility in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **52**, 873–893 (2004).
372. Caiazza, N. C., Merritt, J. H., Brothers, K. M. & O'Toole, G. A. Inverse regulation of biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J. Bacteriol.* **189**, 3603–3612 (2007).
373. Hamer, R., Chen, P.-Y., Armitage, J. P., Reinert, G. & Deane, C. M. Deciphering chemotaxis pathways using cross species comparisons. *BMC Syst. Biol.* **4**, 3 (2010).
374. Sampedro, I., Parales, R. E., Krell, T. & Hill, J. E. *Pseudomonas* chemotaxis. *FEMS Microbiol. Rev.* 17–46 (2014). doi:10.1111/1574-6976.12081
375. Airola, M. V. *et al.* HAMP domain conformers that propagate opposite signals in bacterial chemoreceptors. *PLoS Biol.* **11**, e1001479 (2013).
376. Briegel, A. *et al.* Universal architecture of bacterial chemoreceptor arrays. *Proc. Natl. Acad. Sci. Natl Acad Sci U S A* **106**, 17181–17186 (2009).
377. Hess, J. F., Oosawa, K., Kaplan, N. & Simon, M. I. Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell* **53**, 79–87 (1988).
378. Garrity, L. F. & Ordal, G. W. Activation of the CheA kinase by asparagine in *Bacillus subtilis* chemotaxis. *Microbiology* **143**, 2945–2951 (1997).
379. Barak, R. & Eisenbach, M. Correlation between phosphorylation of the chemotaxis protein CheY and its activity at the flagellar motor. *Biochemistry* **31**, 1821–1826 (1992).

Bibliography

380. Toker, A. S. & Macnab, R. M. Distinct regions of bacterial flagellar switch protein FliM interact with FliG, FliN and CheY. *J. Mol. Biol.* **273**, 623–634 (1997).
381. McEvoy, M. M., Bren, A., Eisenbach, M. & Dahlquist, F. W. Identification of the binding interfaces on CheY for two of its targets the phosphatase CheZ and the flagellar switch protein FliM. *J. Mol. Biol.* **289**, 1423–1433 (1999).
382. Springer, W. R. & Koshland, D. E. Identification of a protein methyltransferase as the cheR gene product in the bacterial sensing system. *Proc. Natl. Acad. Sci.* **74**, 533–537 (1977).
383. Anand, G. S., Goudreau, P. N. & Stock, A. M. Activation of methylesterase CheB: Evidence of a dual role for the regulatory domain. *Biochemistry* **37**, 14038–14047 (1998).
384. Anand, G. S. & Stock, A. M. Kinetic basis for the stimulatory effect of phosphorylation on the methylesterase activity of CheB. *Biochemistry* **41**, 6752–6760 (2002).
385. Harwood, C. S., Fosnaugh, K. & Dispensa, M. Flagellation of *Pseudomonas putida* and analysis of its motile behavior. *J. Bacteriol.* **171**, 4063–4066 (1989).
386. García-Fontana, C., Corral Lugo, A. & Krell, T. Specificity of the CheR2 methyltransferase in *Pseudomonas aeruginosa* is directed by a C-terminal pentapeptide in the McpB chemoreceptor. *Sci. Signal.* **7**, ra34 (2014).
387. Hong, C. S. *et al.* The aerotaxis transducer gene aer, but not aer-2, is transcriptionally regulated by the anaerobic regulator ANR in *Pseudomonas aeruginosa*. *J. Biosci. Bioeng.* **97**, 184–90 (2004).
388. Mattick, J. S. Type IV pili and twitching motility. *Annu. Rev. Microbiol.* **56**, 289–314 (2002).
389. Bertrand, J. J., West, J. T. & Engel, J. N. Genetic analysis of the regulation of type IV pilus function by the Chp chemosensory system of *Pseudomonas aeruginosa*. *J. Bacteriol.* **192**, 994–1010 (2010).
390. DeLange, P. A., Collins, T. L., Pierce, G. E. & Robinson, J. B. PilJ localizes to cell poles and is required for type IV pilus extension in *Pseudomonas aeruginosa*. *Curr. Microbiol.* **55**, 389–395 (2007).
391. Darzins, A. The *Pseudomonas aeruginosa* pilK gene encodes a chemotactic methyltransferase (CheR) homologue that is translationally regulated. *Mol. Microbiol.* **15**, 703–717 (1995).

392. O'Toole, G. A. & Kolter, R. Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* **30**, 295–304 (1998).
393. O'Connor, J. R., Kuwada, N. J., Huangyutitham, V., Wiggins, P. A. & Harwood, C. S. Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol. Microbiol.* **86**, 720–729 (2012).
394. Güvener, Z. T. & Harwood, C. S. Subcellular location characteristics of the *Pseudomonas aeruginosa* GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. *Mol. Microbiol.* **66**, 1459–1473 (2007).
395. Ramos, H. C., Rumbo, M. & Sirard, J. C. Bacterial flagellins: Mediators of pathogenicity and host immune responses in mucosa. *Trends in Microbiology* **12**, 509–517 (2004).
396. Shitashiro, M. *et al.* Identification of chemosensory proteins for trichloroethylene in *Pseudomonas aeruginosa*. *J. Biosci. Bioeng.* **99**, 396–402 (2005).
397. Kim, H. E. *et al.* Identification and characterization of the chemotactic transducer in *Pseudomonas aeruginosa* PAO1 for positive chemotaxis to trichloroethylene. *J. Bacteriol.* **188**, 6700–6702 (2006).
398. Wu, H. *et al.* Identification and characterization of two chemotactic transducers for inorganic phosphate in *Pseudomonas aeruginosa*. *J. Bacteriol.* **182**, 3400–3404 (2000).
399. Alvarez-Ortega, C. & Harwood, C. S. Identification of a malate chemoreceptor in *Pseudomonas aeruginosa* by screening for chemotaxis defects in an energy taxis-deficient mutant. *Appl. Environ. Microbiol.* **73**, 7793–7795 (2007).
400. Kim, H.-E., Shitashiro, M., Kuroda, A., Takiguchi, N. & Kato, J. Ethylene chemotaxis in *Pseudomonas aeruginosa* and other *Pseudomonas* species. *Microbes Environ.* **22**, 186–189 (2007).
401. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
402. Buchan, D. W. a, Minneci, F., Nugent, T. C. O., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349-57 (2013).

Bibliography

403. Sasakura, Y. *et al.* Characterization of a direct oxygen sensor heme protein from *Escherichia coli*. Effects of the heme redox states and mutations at the heme-binding site on catalysis and structure. *J. Biol. Chem.* **277**, 23821–7 (2002).
404. Saleh, B. E. A. & Teich, M. C. *Fundamentals of Photonics*. (2001).
405. Hough, M. Personal correspondence. (2015).
406. Manole, A. *et al.* Conformational control of the binding of diatomic gases to cytochrome c'. *J. Biol. Inorg. Chem.* **20**, 675–686 (2015).
407. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta - Bioenerg.* **1767**, 1073–1101 (2007).
408. Pinakoulaki, E., Ohta, T., Soulimane, T., Kitagawa, T. & Varotsis, C. Detection of the His-heme Fe²⁺-NO species in the reduction of NO to N₂O by ba₃-oxidase from *Thermus thermophilus*. *J. Am. Chem. Soc.* **127**, 15161–15167 (2005).
409. Skubák, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777 (2013).
410. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A Found. Crystallogr.* **64**, 112–122 (2007).
411. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 470–478 (2010).
412. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
413. Tickle, I. J. *et al.* Staraniso Cambridge, United Kingdom: Global Phasing Ltd. (2016).
414. Bricogne, G. *et al.* BUSTER version 2.10.3 Cambridge, United Kingdom: Global Phasing Ltd. (2016).
415. Key, J. & Moffat, K. Crystal structures of deoxy and CO-bound bjFixLH reveal details of ligand recognition and signaling. *Biochemistry* **44**, 4627–4635 (2005).
416. Schmidt, B., Ho, L. & Hogg, P. J. Allosteric disulfide bonds. *Biochemistry* **45**, 7429–7433 (2006).
417. Wells, J. A. & Powers, D. B. In vivo formation and stability of engineered disulfide bonds in subtilisin. *J. Biol. Chem.* **261**, 6564–6570 (1986).

418. Kuwajima, K., Ikeguchi, M., Sugawara, T., Hiraoka, Y. & Sugai, S. Kinetics of disulfide bond reduction in alpha-lactalbumin by dithiothreitol and molecular basis of superreactivity of the Cys6-Cys120 disulfide bond. *Biochemistry* **29**, 8240–8249 (1990).
419. Katz, B. A. & Kossiakoff, A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem.* **261**, 15480–15485 (1986).
420. Fernhoff, N. B., Derbyshire, E. R. & Marletta, M. a. A nitric oxide/cysteine interaction mediates the activation of soluble guanylate cyclase. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21602–21607 (2009).
421. Ha, D.-G. & O'Toole, G. A. C-di-GMP and its effects on biofilm formation and dispersion: a *Pseudomonas aeruginosa* review. *Microb. Spectr.* **3**, MB-003-2014 (2015).
422. Valentini, M. & Filloux, A. Biofilms and c-di-GMP signaling: Lessons from *Pseudomonas aeruginosa* and other bacteria. *J. Biol. Chem.* **291**, 12547–12555 (2016).
423. Finn, R. *et al.* The Pfam protein families databases. *Nucleic Acids Res.* **38**, D211–D222 (2010).
424. Schultz, J. & Milpetz, F. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 5857–5864 (1998).
425. Letunic, I., Doerks, T. & Bork, P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 10–13 (2014). doi:10.1093/nar/gku949
426. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, 182–185 (2007).
427. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691-699 (2001).
428. Felsenstein, J. Confidence Limits on phylogenies: An approach using the bootstrap. *Evolution (N. Y.)*. **39**, 783–791 (1985).
429. Kumar, S., Tamura, K. & Nei, M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Bioinformatics* **10**, 189–191 (1994).
430. Kumar, S., Stecher, G. & Tamura, K. MEGA7 : Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

Bibliography

431. Reinelt, S., Hofmann, E., Gerharz, T., Bott, M. & Madden, D. R. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J. Biol. Chem.* **278**, 39189–39196 (2003).
432. Zhou, Y. F. *et al.* C4-Dicarboxylates sensing mechanism revealed by the crystal structures of DctB sensor domain. *J. Mol. Biol.* **383**, 49–61 (2008).
433. Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W. A. & Waldburger, C. D. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. *J. Biol. Chem.* **276**, 41182–41190 (2001).
434. Cho, U. S. *et al.* Metal bridges between the PhoQ sensor domain and the membrane regulate transmembrane signaling. *J. Mol. Biol.* **356**, 1193–1206 (2006).
435. Ukaegbu, U. E. & Rosenzweig, A. C. Structure of the redox sensor domain of *Methylococcus capsulatus* (Bath) MmoS. *Biochemistry* **48**, 2207–2215 (2009).
436. Nash, A. I. *et al.* Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl. Acad. Sci.* **108**, 9449–9454 (2011).
437. Rinaldi, J. *et al.* The β -scaffold of the LOV domain of the brucella light-activated histidine kinase is a key element for signal transduction. *J. Mol. Biol.* **420**, 112–127 (2012).
438. Conrad, K. S., Bilwes, A. M. & Crane, B. R. Light-induced subunit dissociation by a light-oxygen-voltage domain photoreceptor from *Rhodobacter sphaeroides*. *Biochemistry* **52**, 378–391 (2013).
439. Röllen, K. *et al.* Signaling states of a short blue-light photoreceptor protein PpSB1-LOV revealed from crystal structures and solution NMR spectroscopy. *J. Mol. Biol.* **428**, 3721–3736 (2016).
440. Möglich, A. & Moffat, K. Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA. *J. Mol. Biol.* **373**, 112–126 (2007).
441. Rajagopal, S. & Moffat, K. Crystal structure of a photoactive yellow protein from a sensor histidine kinase: conformational variability and signal transduction. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1649–1654 (2003).

442. Borgstahl, G. E., Williams, D. R. & Getzoff, E. D. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278–6287 (1995).
443. Valentini, M., Storelli, N. & Lapouge, K. Identification of C4-dicarboxylate transport systems in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* **193**, 4307–4316 (2011).
444. Stock, J. B., Stock, A. M. & Mottonen, J. M. Signal transduction in bacteria. *Nature* **344**, 395–400 (1990).
445. Stock, J. B., Ninfa, A. J. & Stock, A. M. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol. Rev.* **53**, 450–490 (1989).
446. Chang, F.-Y., Lu, C. L. & Peng, H.-L. Evolutionary analysis of the two-component systems in *Pseudomonas aeruginosa* PAO1. *J. Mol. Evol.* **59**, 725–737 (2004).
447. Petrova, O. E. & Sauer, K. A novel signaling network essential for regulating *Pseudomonas aeruginosa* biofilm development. *PLoS Pathog.* **5**, (2009).
448. Lundgren, B. R. *et al.* Genetic analysis of the assimilation of C5-dicarboxylic acids in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* **196**, 2543–2551 (2014).
449. Tatke, G., Kumari, H., Silva-Herzog, E., Ramirez, L. & Mathee, K. *Pseudomonas aeruginosa* MifS-MifR two-component system is specific for α -ketoglutarate utilization. *PLoS One* **10**, 1–31 (2015).
450. Sonawane, A. M., Singh, B. & Röhm, K. H. The AauR-AauS two-component system regulates uptake and metabolism of acidic amino acids in *Pseudomonas putida*. *Appl. Environ. Microbiol.* **72**, 6569–6577 (2006).
451. Avison, M. B., Horton, R. E., Walsh, T. R. & Bennett, P. M. *Escherichia coli* CreBC is a global regulator of gene expression that responds to growth in minimal media. *J. Biol. Chem.* **276**, 26955–26961 (2001).
452. Amemura, M., Makino, K., Shinagawa, H. & Nakata, A. Cross talk to the phosphate regulon of *Escherichia coli* by PhoM protein: PhoM is a histidine protein kinase and catalyzes phosphorylation of PhoB and PhoM-open reading frame 2. *J. Bacteriol.* **172**, 6300–6307 (1990).
453. Luu, R. A. *et al.* Taxis of *Pseudomonas putida* F1 toward phenylacetic acid is mediated by the energy taxis receptor AER2. *Appl. Environ. Microbiol.* **79**, 2416–2423 (2013).

Bibliography

454. Cai, Y. Unpublished data. (2016).
455. Song, J. & Jensen, R. A. PhhR, a divergently transcribed activator of the phenylalanine hydroxylase gene cluster of *Pseudomonas aeruginosa*. *Mol. Microbiol.* **22**, 497–507 (1996).
456. Herrera, M. C. & Ramos, J. L. Catabolism of phenylalanine by *Pseudomonas putida*: The NtrC-family PhhR regulator binds to two sites upstream from the *phhA* gene and stimulates transcription with σ^{70} . *J. Mol. Biol.* **366**, 1374–1386 (2007).
457. Herrera, M. C., Krell, T., Zhang, X. & Ramos, J. L. PhhR binds to target sequences at different distances with respect to RNA polymerase in order to activate transcription. *J. Mol. Biol.* **394**, 576–586 (2009).
458. Palmer, G. C., Palmer, K. L., Jorth, P. A. & Whiteley, M. Characterization of the *Pseudomonas aeruginosa* transcriptional response to phenylalanine and tyrosine. *J. Bacteriol.* **192**, 2722–2728 (2010).
459. Reimann, C. *et al.* The global activator GacA of *Pseudomonas aeruginosa* PAO positively controls the production of the autoinducer N-butyryl-homoserine lactone and the formation of the virulence factors pyocyanin, cyanide, and lipase. *Mol. Microbiol.* **24**, 309–319 (1997).
460. Bordi, C. *et al.* Regulatory RNAs and the HptB/RetS signalling pathways fine-tune *Pseudomonas aeruginosa* pathogenesis. *Mol. Microbiol.* **76**, 1427–1443 (2010).
461. Chambonnier, G. *et al.* The hybrid histidine kinase LadS forms a multicomponent signal transduction system with the GacS/GacA two-component system in *Pseudomonas aeruginosa*. *PLoS Genet.* **12**, 1–30 (2016).
462. Ritchings, B. W., Almira, E. C., Lory, S. & Ramphal, R. Cloning and phenotypic characterization of *fleS* and *fleR*, new response regulators of *Pseudomonas aeruginosa* which regulate motility and adhesion to mucin. *Infect Immun* **63**, 4868–4876 (1995).
463. Dasgupta, N. *et al.* A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **50**, 809–824 (2003).
464. Yeung, A. T. Y. *et al.* Swarming of *Pseudomonas aeruginosa* is controlled by a broad spectrum of transcriptional regulators, including MetR. *J. Bacteriol.* **191**, 5592–5602 (2009).

465. Kong, W. *et al.* Hybrid sensor kinase PA1611 in *Pseudomonas aeruginosa* regulates transitions between acute and chronic infection through direct interaction with RetS. *Mol. Microbiol.* **88**, 784–797 (2013).
466. Balasubramanian, D., Schneper, L., Kumari, H. & Mathee, K. A dynamic and intricate regulatory network determines *Pseudomonas aeruginosa* virulence. *Nucleic Acids Res.* **41**, 1–20 (2013).
467. Cotter, P. a. & Stibitz, S. C-di-GMP-mediated regulation of virulence and biofilm formation. *Curr. Opin. Microbiol.* **10**, 17–23 (2007).
468. Kuchma, S. L., Kuchma, S. L., Connolly, J. P. & Connolly, J. P. A three-component regulatory system regulates biofilm maturation and type III secretion in *Pseudomonas aeruginosa*. *J. Bacteriol.* **187**, 1441–1454 (2005).
469. Lundgren, B. R., Sarwar, Z., Pinto, A., Ganley, J. G. & Nomura, C. T. Ethanolamine catabolism in *Pseudomonas aeruginosa* PAO1 is regulated by the enhancer-binding protein EatR (PA4021) and the alternative sigma factor RpoN. *J. Bacteriol.* **198**, 2318–2329 (2016).
470. Kruger, N. & Steinbuchel, A. Identification of *acoR*, a regulatory gene for the expression of genes essential for acetoin catabolism in *Alcaligenes eutrophus* H16. *J. Bacteriol.* **174**, 4391–4400 (1992).
471. Pedroni, P., Friedrich, T., Breuer, M., McBeth, D. & Hauer, B. Novel esterase activity of dihydrolipoamide acetyltransferase AcoC of *Pseudomonas putida* identified by mutation of the *acoR* regulator. *Can. J. Chem.* **80**, 692–698 (2002).
472. Nishijyo, T., Haas, D. & Itoh, Y. The CbrA-CbrB two-component regulatory system controls the utilization of multiple carbon and nitrogen sources in *Pseudomonas aeruginosa*. *Mol. Microbiol.* **40**, 917–931 (2001).
473. Li, W. & Lu, C. D. Regulation of carbon and nitrogen utilization by CbrAB and NtrBC two-component systems in *Pseudomonas aeruginosa*. *J. Bacteriol.* **189**, 5413–5420 (2007).
474. Li, Y. *et al.* Identification of a new gene PA5017 involved in flagella-mediated motility, chemotaxis and biofilm formation in *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* **272**, 188–195 (2007).
475. Tan, E. *et al.* Solution structure of the PAS domain of a thermophilic YybT protein homolog reveals a potential ligand-binding site. *J. Biol. Chem.* **288**, 11949–11959 (2013).

Bibliography

- 476. Kearns, D. B. A field guide to bacterial swimming motility. *Nat. Rev. Microbiol.* **8**, 634–644 (2010).
- 477. Wadhams, G. H. & Armitage, J. P. Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**, 1024–1037 (2004).
- 478. Jones, C. W. & Armitage, J. P. Positioning of bacterial chemoreceptors. *Trends Microbiol.* **23**, 1–10 (2015).
- 479. Rebbapragada, a *et al.* The Aer protein and the serine chemoreceptor Tsr independently sense intracellular energy levels and transduce oxygen, redox, and energy signals for *Escherichia coli* behavior. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10541–10546 (1997).
- 480. Merritt, J. H., Brothers, K. M., Kuchma, S. L. & O'Toole, G. A. SadC reciprocally influences biofilm formation and swarming motility via modulation of exopolysaccharide production and flagellar function. *J. Bacteriol.* **189**, 8154–8164 (2007).
- 481. Kuchma, S. L. *et al.* BifA, a cyclic-di-GMP phosphodiesterase, inversely regulates biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *J. Bacteriol.* **189**, 8165–8178 (2007).
- 482. Gorrec, F. The MORPHEUS protein crystallization screen. *J. Appl. Crystallogr.* **42**, 1035–1042 (2009).
- 483. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 484. Schrödinger, L. *The PyMOL molecular graphics system, version 1.3r1.* (2010).