University of Southampton

Faculty of Natural and Environmental Sciences

School of Biological Sciences

Identification of NO-sensing protein domains that regulate bacterial pathogenesis and biofilm formation in *Pseudomonas aeruginosa*

by

Jack William Craddock

Thesis for the degree of Doctor of Philosophy

August 2020

University of Southampton

Abstract

Faculty of Natural and Environmental Sciences

School of Biological Sciences

Thesis for the degree of Doctor of Philosophy

Identification of NO-sensing protein domains that regulate bacterial pathogenesis and biofilm formation in *Pseudomonas aeruginosa*

by

Jack William Craddock

Most species of bacteria preferentially grow in sessile communities, known as biofilms, rather than as free-living planktonic cells. Biofilms can be up to 1000 times more tolerant to antimicrobials compared to their planktonic counterparts, making them a growing problem within the medical and industrial settings. Biofilms formed by *Pseudomonas aeruginosa* are involved in chronic infections, such as those affecting the lungs of Cystic Fibrosis patients. High or low intracellular levels of a bacterial secondary messenger, bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP), regulates biofilm formation and dispersal respectively. The synthesis of c-di-GMP is carried out by diguanylate cyclases (DGC) due to catalytic domains known as GGDEF domains. Whereas c-di-GMP degradation is carried out by phosphodiesterases (PDE) due to catalytic domains known as EAL domains. Low (non-toxic) concentrations of nitric oxide are known to induce a biofilm dispersal through a reduction in the c-di-GMP levels and an increase in the PDE activity. However the nitric oxide sensor and the protein responsible for the reduction in c-di-GMP levels is unknown.

Previously found to be involved in an NO-induced biofilm dispersal, the bi-functional enzyme RbdA (regulation of biofilm disposal) contains a GGDEF and an EAL domain in tandem. Using an enzymatic assay measuring PDE activity, we investigate the relationship between the tandem GGDEF and EAL domains. We find that the isolated EAL domain of RbdA has a higher PDE activity, suggesting that the tandem GGDEF domain negatively influences the activity of the EAL domain. We attempt to further investigate this at the molecular level using X-ray crystallography and structure determination. The structure of the EAL domain of RbdA was

determined and appears to be in a primed state for substrate binding, with a single Mg^{2+} ion bound within the active site. After comparisons to other EAL domain structures, we suggest a schematic for substrate binding to EAL domains.

We investigate an RbdA homologue, PA2072, as previous biological data indicates opposing physiological roles. By comparing the primary and secondary structures of RbdA and PA2072 we suggest that their physiological differences are caused by disparities between their periplasmic regions and / or their putative sensory PAS domains. Protein crystallisation of the PA2072 periplasmic region (a putative CHASE4 domain) and the PA2072 PAS domain were attempted but require further optimisation.

The first and second PAS domain of PA0285 were predicted to bind a haem-*b* and FAD (flavin adenine dinucleotide) cofactor respectively. We hypothesise that NO can be sensed by a haem-bound PAS domain. Using ultraviolet-visible spectroscopy we could only identify a very weak haem-*b* cofactor binding to the PAS1 domain of PA0285 and so requires further investigation. However, we identified a PA0285 PAS2 : FAD binding stoichiometry of approximately 2 : 1. Here we put forward models suggesting that, NO-induced changes to the redox potential are sensed by the FAD bound PAS2 domain, leading to changes in the enzymatic output of PA0285 and potentially biofilm dispersal. Restoring the sensitivity of bacterial cells to antimicrobials by inducing a biofilm dispersal, is thought to be a novel treatment strategy. This work lays some of the foundations required to understand the molecular mechanisms that lead to a biofilm dispersal in *P. aeruginosa*.

Contents

List of Figures

List of Tables

Research Thesis: Declaration of Authorship

Print name: JACK WILLIAM CRADDOCK

Title of thesis: Identification of NO-sensing protein domains that regulate bacterial pathogenesis and biofilm formation in *Pseudomonas aeruginosa.*

I declare that this thesis 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:

Cai Y-M., Hutchin A., Craddock J., Walsh M. A., Webb J. S. and Tews I. Differential impact on motility and biofilm dispersal of closely related phosphodiesterases in *Pseudomonas aeruginosa*. *Scientific reports* 2020;10(1):6232.

Signature: Jack William Craddock Date: 24/08/2020

Acknowledgements

Firstly, I would like to give my upmost thanks to Prof. Jeremy Webb and Dr. Ivo Tews for their guidance, encouragement and their continued support throughout. I am extremely grateful for the opportunity to work on this project. I would also like to thank the Griffiths Memorial Trust and University of Southampton Vice Chancellor's Award for their funding.

I would like to thank Dr. Yuming Cai and Dr. Andrew Hutchin for all their support and help teaching me laboratory techniques and answering all of my many questions. I would like to thank all of the members of the Tews lab group; Dr. Christian Orr, Hayden Fisher, Rachel Bolton and Charlotte Cordery for their advice, discussion and assistance with crystallographic software and lab protocols. Thanks are also given to members of the Doyle, Williamson, East, Werner and Crispin groups at the University of Southampton for their friendship, frequent advice and discussion. I would also like to thank Christopher Holes for his assistance in the crystallography facility at the University of Southampton, as well as Neville Wright for his help and instruction with equipment and Marilyn Jenkins for providing access to consumables. Thanks also to project students Ellie Woods, Jade Ashton and Finnian Clark.

I am very thankful for the constant support and encouragement my family have shown me. I would also like to thank my friends, inside and outside of the lab, for their support and encouragement while writing this thesis.

Finally, I cannot express my thanks enough to Jenny Gibbs for all of the support, care and encouragement she has given me throughout my studies. I am forever thankful to her for putting up with me – her patience is without equal.

List of abbreviations

c-di-GMP	Bis-(3'-5')-cyclic dimeric guanosine monophosphate
CF	Cystic Fibrosis
CFTR	Cystic Fibrosis Transmembrane Conductance Regulator
cGMP	Cyclic guanosine monophosphate
CHASE	Cyclases/histidine kinases associated sensory extracellular
CO	Carbon monoxide
DGC	Diguanylate Cyclase
EPS	Extracellular polymeric substances
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FRET	Fluorescence resonance energy transfer
H-NOX domain	Haem-NO/O_2-binding domain
KO	Knockout
MAD	Multiple wavelength anomalous dispersion
MIR	Multiple isomorphous replacement
MR	Molecular replacement
NO	Nitric Oxide
PAS domain	Per-ARNT-Sim domain
PDE	Phosphodiesterase
pGpG	5'-phosphoguanylyl-(3',5')-guanosine
RMSD	Root mean square deviation
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
WT	Wildtype

Chapter 1 – Introduction

1.1 – *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a rod-shaped aerobic and facultative anaerobic Gram-negative opportunistic pathogen ubiquitous in the environment. As an opportunistic pathogen, *P. aeruginosa* commonly infects the lungs of Cystic Fibrosis (CF) patients. CF is a genetic disease, which is the most common lethal disease to be inherited in Caucasians[1,2], affecting approximately 80,000 individuals worldwide[3]. It is caused by the inheritance of two recessive 'faulty' alleles of the cystic fibrosis transmembrane conductance regulator (CFTR)[1]. Due to where CFTR is expressed multisystem complications arise in CF patients, such as endocrine and exocrine pancreatic dysfunction, defective gastrointestinal transit, intestinal malabsorption, liver disease and accelerated bone loss[2]. It is chronic pulmonary disease however that accounts for the highest rate of morbidity (being in a diseased state) and death in CF patients[2,3].

In the lung the CFTR protein normally maintains the volume and ionic composition of the airway surface fluid[1], functioning predominately as a chloride channel in the apical membrane of epithelial cells[3]. In CF patients the faulty or even absent CFTR protein, depending on the CFTR mutation, results in a build-up of a dehydrated thick mucus in the airways[3]. This thick mucus impairs the mucociliary escalator and its clearing of the airways, causing obstructions in the small airways[3] and resulting in CF patients being highly susceptible to bacterial infections, especially *P. aeruginosa*[1]. It has been shown by Worlitzsch et al.[4] that *P. aeruginosa* are predominately found inside the mucus rather than being attached to the lung epithelial cells. Worlitzsch et al.[4] also suggest that bacterial motility allows motile *P. aeruginosa* to penetrate into stationary mucus, by showing that motile *P. aeruginosa* had moved into stationary mucus whereas fluorescent beads did not. The airway mucus is mostly anaerobic, but being a facultative anaerobe, if appropriate terminal electron acceptors are available *P. aeruginosa* can grow under anaerobic conditions[4].

Despite antibiotic therapies or the host's inflammatory response, *P. aeruginosa* can survive within the CF lung which is considered to be primarily due to their ability to form bacterial biofilms[1,4-6]. Biofilms are structured sessile communities of aggregated bacterial cells in which

the bacteria become embedded in a self-produced extracellular matrix. Usually, the lungs of CF patients become chronically infected with *P. aeruginosa,* residing in the biofilm lifestyle, by the time the patients reach adolescence and early adulthood[6,7]. Chronic bacterial infections along with chronic inflammation of the lungs due to the host's immune response, results in tissue damage[8] and progressive deterioration of lung function and, in most cases of CF, eventually leads to the fatality of the patient as a result[6,7,9]. Therefore, novel treatment strategies to clear the biofilm infection in CF patients' lungs are required.

Current treatments for the pulmonary complications affecting CF patients are shown in Table 1.1 and also include physiotherapy[10] and lung transplantation[11]. However, these current treatment strategies can have negative impacts on the continued adherence to treatments due to the large time commitments required and the complexity of the daily treatments[12]. Furthermore, current treatments for the faulty CFTR protein are insufficient or unavailable for a significant number of CF patients[13]. Recently, it has been reported by Park et al.[14] that individuals lacking the TMEM16A protein also have a non-functional CFTR protein, but the CF respiratory phenotype was not observed in these individuals. Consequently, the authors suggest that inhibitors of the TMEM16A protein could prove a good strategy in the future treatment of CF[14].

Nevertheless, in addition to being the major pathogen for CF patients, *P. aeruginosa* also causes complications and chronic infections in other individuals such as those suffering from burn wounds[15] or those with indwelling medical devices[16]. Consequently, it is of clinical importance to find strategies for the treatment of *P. aeruginosa* biofilm infections.

Table 1.1. **Summary of current pulmonary-directed treatments for cystic fibrosis patients.**

Drug Target	Drug	Mode of Action	Indications	Ref.
Mucus	Dornase alfa	Reduces mucus viscosity by cleaving extracellular DNA.	For patients aged 5 years and older.	10,12
	Mannitol	Rehydrates the airway surface leading to the clearance of mucus due to being hyperosmotic.	For patients aged 18 and older taken with other treatments.	10,12
	Hypertonic saline	Increases the ion concentration in the airway surface liquid, drawing water out of the epithelial cells through osmosis and thinning the mucus.	For patients aged 6 years and older. Normally needs to be taken with a bronchodilator.	12
Inflammation	Azithromycin	The mechanism is not fully understood but possibly inhibits the release of pro-inflammatory mediators.	For patients aged 6 years and older. Reduced clinical benefits after 1 year of use. Emergence of macrolide-resistance in *Staphylococcus aureus* after 6 months of use.	17,18
	(High-dose) Ibuprofen	Inhibits cyclo-oxygenase and thus prevents the synthesis of prostaglandins. Also has a possible antimicrobial activity.	For patients aged 6 to 17 years with a forced expiratory volume in one second (FEV_1) ≥60 % predicted – higher efficacy in young CF patients with mild lung disease.	19,20
CFTR defect	Lumacaftor (used with Ivacaftor)	Improves the cellular processing and trafficking of the mutant F508del-CFTR to increase the number of functional CFTR proteins in the cell membrane.	For patients aged 2 years and older with a homozygous genotype for the *F508del* mutation in the *CFTR* gene.	13,21,22
	Tezacaftor (used with Ivacaftor)	Improves the cellular processing and trafficking of the mutant F508del-CFTR to increase the number of functional CFTR proteins in the cell membrane.	For patients aged 12 years and older with a homozygous genotype for the *F508del* mutation in the *CFTR* gene.	13,21
	Ivacaftor	Enhances the transport of chloride by increasing CFTR channel gating.	For patients aged 1 year and older with a specific gating mutation in the *CFTR* gene (G178R, S549N, S549R, G551D, G551S, G1244E, S1251N, S1255P or G1349D). Or for patients aged 18 years and older with the R117H *CFTR* gene mutation.	13,21
Infection	Flucloxacillin	Anti-staphylococcal antibiotic in the penicillin class. Interferes with bacterial cell wall synthesis.	Used in CF patients as a prophylaxis from point of diagnosis until aged 3 years.	10
	Colistin	Interacts with lipopolysaccharides disrupting the outer and inner membranes of Gram-negative bacteria.	For patients aged 2 years and older.	10
	Tobramycin	Inhibits protein synthesis in bacteria by irreversibly binding to the 30S ribosomal subunit.	For patients aged 6 years and older.	23
	Aztreonam	Inhibits bacterial cell wall synthesis by binding to penicillin binding protein 3.	For patients aged 6 years and older.	24

1.1.1 – Biofilms

It is now understood that most bacterial species, whether in an environmental, medical or industrial setting, favour the sessile biofilm lifestyle rather than the free-living planktonic lifestyle[25]. Bacteria within the biofilm lifestyle are physiologically distinct from their counterparts in the planktonic lifestyle[26], as will be described in the following sections. Bacterial biofilms can form at interfaces, attached to all (known) biotic and abiotic surfaces[16,27] or even attached to other microbial cells which may be of different species[28]. However, biofilms are not just formed by bacteria, they can also be formed by fungi[29], archaea and microalgae[30,31]. In fact in nature, biofilms often consist of many different microbial species as well as non-cellular materials (such as mineral crystals, silt or corrosion particles and clay)[32].

Often biofilms develop complex, highly organised three-dimensional structures[25,29] which differ depending on the species forming the biofilm and the environment in which they are formed[33]. Furthermore, the biofilm structure is dynamic, being capable of adapting in response to changing environmental conditions[33-35]. These structures also consist of actively maintained water channels in which waste products, nutrients and oxygen can be transported[29,36,37]. The water channels surround microcolonies which are distinct and often differentiated multicellular structures comprised of bacterial cells embedded within a self-produced matrix of extracellular polymeric substances (EPS)[5,28,35]. The EPS matrix is formed of exopolysaccharides, extracellular DNA, proteins and lipids[27]. It acts as a scaffold important in maintaining the structure of a biofilm[33], where the majority of a biofilm is formed of the EPS matrix, with only a small proportion being made up of cells[30]. The EPS is highly hydrated, formed of 98 % water[38]. However, like the biofilm structure, the exact composition of the EPS matrix is complex and dynamic and varies between different bacterial species[16] and even with different environmental conditions within the same species.

Examples of biofilms in everyday life include the dental plaque on the surface of teeth, the slime that covers submerged rocks, and the biofouling of ship hulls, water and oil pipes[29,31]. Biofilms also cause serious medical problems by growing on indwelling medical devices (such as pacemakers, catheters and prosthetics) and even living tissues and open wounds[29,30]. Many set up persistent infections, often with the only cure being the removal and replacement of

the implant or infected tissue[16,39-41]. Furthermore, biofilms formed by non-pathogenic microbial species may provide an environment for pathogens to reside in. Yet, biofilms can be useful in the treatment of wastewater[30], in promoting remediation of contamination in the environment[31] and in biofuel production[42]. Nevertheless, it is now thought that over 80 % of all bacterial infections involve biofilms[1,43].

Establishing a biofilm offers the constituent microbial cells protection from desiccation, predation and UV radiation as well as a broad spectrum defence to antimicrobials[32,44] (such as antibiotics and the immune system) shown both *in vitro* and *in vivo*[25,39]. Thus the biofilm lifestyle promotes bacterial survival in hostile conditions. Indeed bacterial biofilms can be up to 1000-fold more tolerant to antibiotics and biocides than compared to their planktonic counterparts[45]. This increased tolerance is further confirmed to be due to the biofilm lifestyle as, when biofilms are resuspended, the sensitivity of the bacterial cells to the antimicrobials is restored[46].

1.1.2 – Antibiotic tolerance mechanisms

The increased tolerance of biofilms against antimicrobials can be explained by different mechanisms and it is likely they all play a part, although the extent of each mechanism varies between species and antimicrobial. The age of a biofilm is also an important determinant in how tolerant the biofilm is to an antimicrobial. It has been shown that younger biofilms are more susceptible (to the antibiotic tobramycin) as compared to mature biofilms[47].

Once thought a key mechanism, there is now conflicting evidence as to whether the EPS matrix of a biofilm can prevent complete antimicrobial penetration by acting as a physical and chemical barrier[16,40]. The EPS matrix is hypothesised to act as a diffusion barrier, due to its reaction with or binding of antimicrobials[40], or its immobilisation of antimicrobial inactivating enzymes (such as β-lactamases)[32]. Supporting this, there are studies that have shown the penetration of antimicrobials through a biofilm is reduced[48]. Yet many report that antimicrobials can readily permeate through a biofilm[49]. Moreover, some studies also present contradicting evidence where biofilms have a reduced penetration to some antibiotics but are permeable to others[15,46,50]. Recently Tseng et al.[15] reported that biofilms of a non-mucoid

strain of *P. aeruginosa* were permeable to the neutral antibiotic ciprofloxacin. However, the authors also found that the EPS matrix was able to sequester tobramycin, a positively charged antibiotic, to the periphery of the biofilm through ionic interactions which protected biofilm cells from being killed[15].

Therefore the ability of the EPS matrix to infer antimicrobial tolerance by reducing an antibiotics penetration through a biofilm seems to be species and antibiotic dependent[40]. However, often even with complete penetration of the antibiotic, cells within the biofilm are refractory to the effects of the antimicrobial, indicating that biofilms have further mechanisms of tolerance[16,44,46].

Most antibiotics target cellular pathways associated with actively growing and metabolising cells[44]. However a mature biofilm is heterogeneous, containing differentiated bacteria and many different micro-environments due to gradients in oxygen, pH, waste and nutrients[42,51]. It has been shown that mature biofilms often express anaerobic genes even when grown under aerobic conditions, whereas younger biofilms do not[52]. As a consequence of these gradients, the interior of a biofilm contains many metabolically inactive, less permeable and slow-growing cells[7,28,53]. Therefore, these cells will have an increased tolerance to many of the antimicrobials. Furthermore, a subpopulation of persister cells which, although may only form a small percentage of the biofilm population, are more frequently found in the biofilm population than compared to the planktonic population[39]. Persister cells are dormant, slow growing variants associated with a much higher tolerance to many antimicrobials which, despite continued antimicrobial exposure, can persist[54]. Also, within a biofilm, horizontal gene transfer between bacterial cells of the same or different species is enhanced due to the cells close proximity to each other[16,32], promoting resistance encoding plasmids to spread quickly throughout the biofilm population.

Thus, when treating bacterial biofilm infections with antibiotics, the symptoms caused by planktonic cells, which naturally leave a biofilm during its life cycle, may be alleviated, but the biofilm infection itself is not eradicated[39,41]. Consequently, biofilm infections are often chronic, and are associated with antimicrobial tolerance and recurring symptoms[39,45]. The development of novel methods to tackle biofilms are, therefore, urgently needed[7]. As a result, understanding a biofilm's lifecycle is crucial, as preventing initial biofilm formation or inducing a biofilm dispersal are both possible mechanisms by which biofilms could be controlled.

1.1.3 – The developmental stages of a biofilm

By using microscopy over 12 days to observe *P. aeruginosa* PAO1 biofilm development when grown within a flow cell with continuous culture, five distinct developmental stages (outlined in Figure 1.1) with their own characteristics were detected[26]. This five stage sequence of the biofilm life cycle seems to be conserved among prokaryotes. Each stage of the biofilm life cycle is formed of cells which are thought to be physiologically distinct from the cells in other stages of the life cycle. However, all developmental stages may be present within a mature biofilm[26].

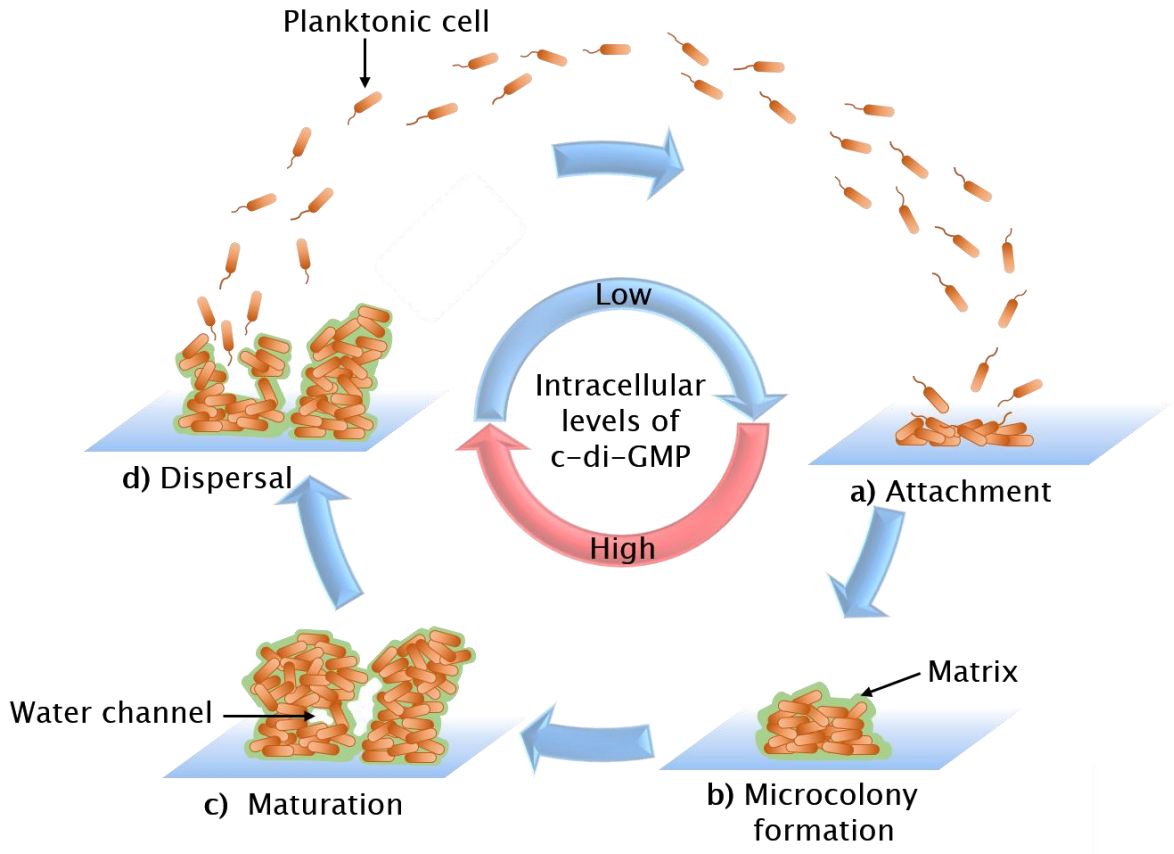

Figure 1.1. **The biofilm life cycle.** (a) Biofilm formation begins with the reversible, followed by irreversible, attachment of planktonic bacterial cells to a surface, an interface or to other bacterial cells. (b) This is followed by microcolony formation due to the cellular proliferation of attached cells which become embedded in a self-produced matrix. (c) The microcolonies will continue to grow to form a mature structured biofilm consisting of different microenvironments and water channels. (d) A subpopulation of bacterial cells will then leave the biofilm by returning to their planktonic state. The life cycle of a biofilm is regulated by the intracellular levels of the secondary messenger Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP).

1.1.3.1 – Stages 1 and 2 – Attachment

Initial attachment (Figure 1.1 A) is reversible, where it is observed that some cells detach, or attach only transiently, before returning to the planktonic state[26]. A conditioning film, formed of proteins and polysaccharides, becomes rapidly adsorbed to exposed surfaces within aqueous environments[38], influencing microbial attachment[32]. At first, when a bacterial cell comes into close proximity with a surface due to fluid flow, gravity, Brownian motion or active motility, there are reversible, transient, non-specific interactions[16], such as electrostatic and van der Waal forces[28]. Microbial cell surface appendages (for example pili (involved in surface sensing[55])) are important in overcoming repulsive forces between the surface and the cell, enabling the cell to remain attached to the surface[28,32,38]. Initial cellular attachment is enhanced by active motility, where it was shown that a flagellar mutant (non-motile) was present at lower numbers at a glass surface of a flow cell than compared to a motile flagellated wild type[26]. When a microbial cell initially interacts with a surface a number of genes are up- and down-regulated within the cell[32].

Irreversible attachment occurs when the motility of surface attached cells ceases, and is associated with the attachment of the long axis of the cell to the surface[26]. Microcolonies then begin to develop through clonal growth while the cells also become embedded in a self-produced EPS matrix, which firmly attaches the bacterial cells together[16,28] (Figure 1.1 B). Still, bacterial cells are well separated from one another with EPS being found between cells in the interstices[26].

1.1.3.2 – Stages 3 and 4 – Maturation

Maturation-1 occurs with the thickening of the microcolonies and the progressive layering of the bacterial cells[26]. Maturation-2 follows when the microcolonies reach their maximum thickness (approximately 100 µm) with most cells of the biofilm being segregated within microcolonies and some microcolonies no longer being in contact with the initial attachment surface[26] (Figure 1.1 C).

1.1.3.3 – Stage 5 – Dispersal

In the final stage of the biofilm life cycle, dispersal events allow bacterial cells to leave a sessile biofilm, returning to the planktonic state to spread and colonise new, possibly more favourable, environments[51] (Figure 1.1 D). Due to cell differentiation (and possible genetic diversification as a result of mutations that may take place within a biofilm) subpopulations of dispersed cells often have a high level of heterogeneity[51]. This variation among dispersed cells is hypothesized to increase the fitness of the population[51]. Different bacterial species utilise different dispersal strategies[28]. Dispersal may be an active process in which cells exhibit a distinct phenotype and require energy and motility to leave the biofilm[51] in response to environmental factors or quorum sensing. Or dispersal may be passive, where cells are lost from the biofilm due to shearing forces[28,51]. After a natural dispersal event (called seeding) in which many cells are rapidly released from the centre of a mature microcolony, non-motile bacteria remain to form a hollow microcolony 'shell' within the biofilm[25,26,28]. Further modes of biofilm dispersal include erosion, in which small numbers of cells are continuously released from a biofilm, and sloughing, where large portions of the biofilm suddenly detach[16,28].

It has been observed in several species of bacteria that, during biofilm development, a subpopulation of cells within a mature microcolony undergo programmed cell death in a spatially and temporally organised manner[5,25,51]. This is thought to benefit the surviving cells within the microcolony which promotes them to differentiate and disperse from the biofilm[5,25]. In biofilms of *P. aeruginosa* PAO1 this cell death has been suggested to be caused by mutations in a filamentous Pf1-like prophage encoded within the genome, causing the normally non-lytic Pf1-like prophage to become lytic and super-infectious[5]. The mutations in the Pf1-like phage have been suggested to be induced by reactive oxygen species, which were found to be present within the centre of mature, but not young, *P. aeruginosa* microcolonies[5]. However, this phage-induced cell death in *P. aeruginosa* biofilms was suggested to occur only when the phage could infect the bacterial cells. This was evidenced by *P. aeruginosa* mutants deficient in both flagella and type 4 pili (both of which are required for and act as receptors for bacteriophage infection), which were resistant to infection and cell death but which still produced a phage that killed the PAO1 WT[5]. Additionally, it has been observed that mutant biofilms that do not undergo a cell death also do not have a dispersal phenotype[25].

Upon biofilm dispersal, the antibiotic sensitivity of the bacterial cells that return to their planktonic state is restored[44,46,47]. Consequently, inducing biofilm dispersal with novel agents and then clearing the more susceptible planktonic bacterial cells with traditional antibiotics is a putative anti-biofilm strategy[45].

1.2 – C-di-GMP – a secondary messenger

The secondary messenger, bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP), is wide-spread in bacteria, being implicated in many of their processes such as, motility (swimming, swarming and twitching), virulence, the cell cycle, differentiation, biofilm formation and dispersal[56-58] and polysaccharide synthesis[59]. Within bacterial cells, high intracellular levels of c-di-GMP are associated with attachment, reduced motility and biofilm formation[33,60], whereas low intracellular levels of c-di-GMP are associated with the motile planktonic phenotype and biofilm dispersal[58]. To bring about these affects, c-di-GMP can act at the level of transcription, translation and post-translation[61,62].

This nucleotide secondary messenger is not naturally produced in archaea or higher eukaryotes and seems to be exclusive to bacteria and lower eukaryotes (such as amoebae)[63]. The level of c-di-GMP within a bacterium is controlled by its synthesis - diguanylate cyclase's (DGCs), and degradation - phosphodiesterase's (PDEs)[33,59-61] (see Figure 1.2).

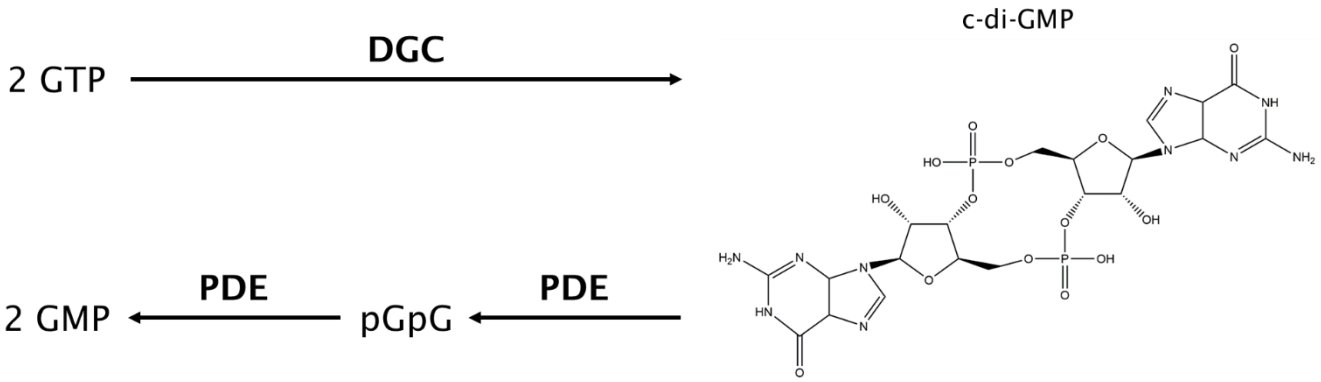

Figure 1.2. **The synthesis and degradation of cyclic dimeric guanosine monophosphate (c-di-GMP).** Catalytic GGDEF domains of diguanylate cyclase's (DGCs) synthesise c-di-GMP from two molecules of GTP. Whereas phosphodiesterase's (PDEs) hydrolyse c-di-GMP to either linear 5'-phosphoguanylyl-(3',5')-guanosine (pGpG) catalysed by EAL domains or to two molecules of GMP catalysed by HD-GYP domains. Here, the enzymatic domains are named due to conserved amino acid sequence motifs. The chemical structure of c-di-GMP was drawn using ChemDraw Professional 16.0.

1.2.1 – C-di-GMP-specific DGCs – GGDEF domains

The enzymatically active domain of c-di-GMP-specific DGC enzymes is the GGDEF domain (approximately 170 amino acids in length[57,64]), named due to the highly conserved amino acid sequence motif of Gly-Gly-Asp-Glu-Phe found in the active site[56,64]. While the first two and last two amino acids of the GGDEF motif are strongly conserved, the third catalytic amino acid of the GGDEF motif can be either Aspartate (D) or Glutamate (E)[56,62] (thus can be written as GG(D/E)EF).

The GGDEF domain was reported by Ryjenkov et al.[64] to be responsible for c-di-GMP-specific DGC activity in which enzymatic assays were carried out using full length proteins (from different branches of the bacterial phylum) containing GGDEF domains as well as isolated GGDEF domains, with only GTP being utilised as a substrate. GGDEF domains synthesise, from two molecules of GTP in a two-step reaction *via* the intermediate 5'-pppGpG, c-di-GMP[57] and the by-products of two pyrophosphate molecules[56]. This particular condensation reaction has only been identified to be carried out by GGDEF domains[65]. It is understood that for this c-di-GMP synthesis to occur, two monomeric GGDEF domains, each bound to a GTP molecule, come together to form an antiparallel[66] homodimer[57,67], with the active site being located at

the dimer interface[56,68]. This would allow the 3'-hydroxyl group of each GTP molecule to be in proximity with the α-phosphate from the opposing GTP molecule, so that the opposing 3'-hydroxyl groups, when deprotonated, can perform an in-line nucleophilic attack on the α-phosphate, leading to the eventual cleavage of the α-β-phosphate phosphodiester bond[67,68]. This would suggest that the mechanism for phosphodiester bond formation in DGCs is the same as that of adenylate cyclases and DNA polymerases[67,68]. Therefore regulatory mechanisms that prevent GGDEF monomers coming together would inhibit DGC activity[56,68]. With adjacent sensory domains being able to regulate the enzymatic activity of GGDEF domains[64].

The GGDEF domain is formed of five α-helices which surround a central five-stranded β-sheet, with an order of $(\alpha_0)\beta_1\alpha_1\alpha_2\beta_2\beta_3\alpha_3\beta_4\alpha_4\beta_5$[69-73] (shown in Figure 1.3). This GGDEF domain fold is similar to that of the adenylate cyclase catalytic core and a DNA polymerase domain and have identical topologies[69]. Furthermore, the mode of nucleotide binding is also similar[68].

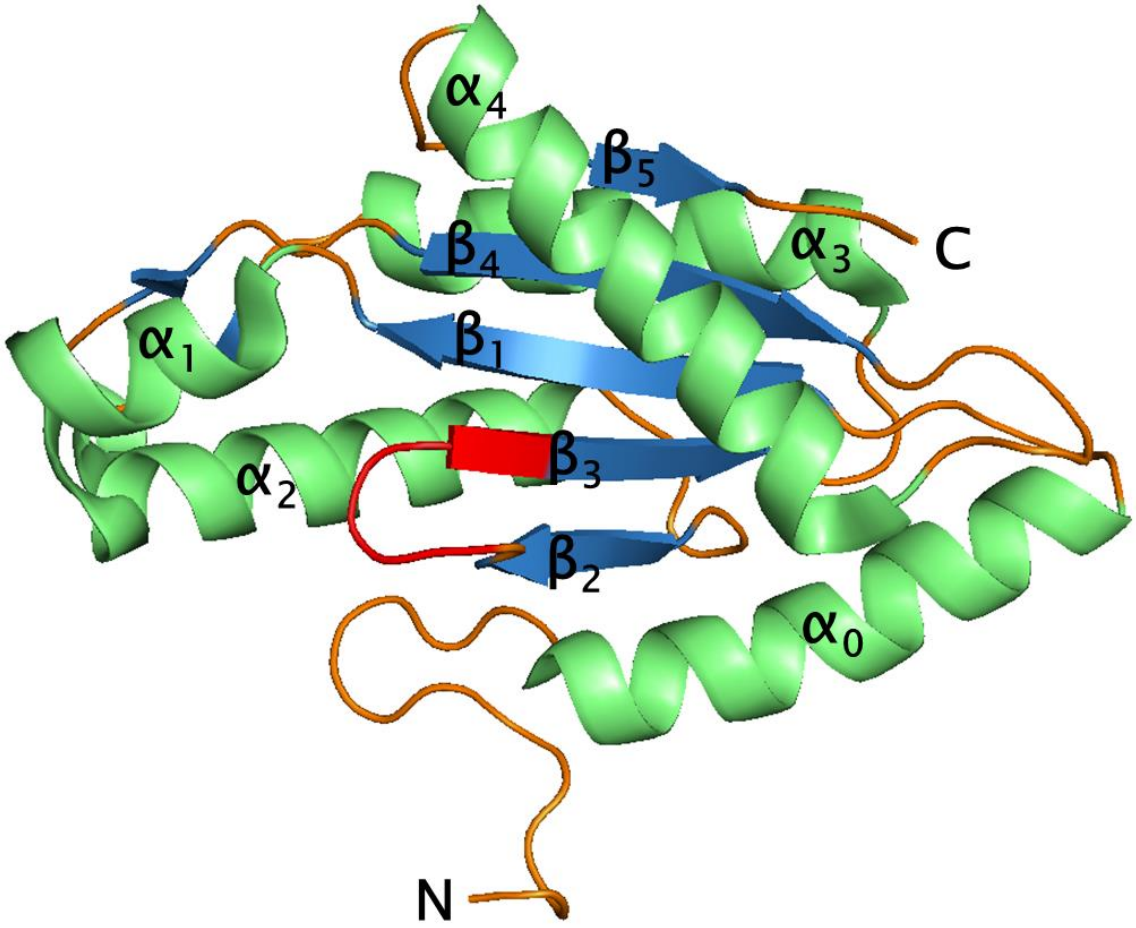

Figure 1.3. **The structure of the GGDEF domain of PleD**. The GGDEF domain of PleD (residues 286-454) from *Caulobacter crescentus* in a cartoon depiction shows a typical GGDEF domain fold, with the GGEEF motif of PleD shown in red, the α-helices in green, β-strands in blue and loops in orange (PDB ID: 1W25 and PyMOL were used to make this image).

The GG(D/E)EF motif, which forms part of the active site, is located on the central β-hairpin (the loop between the β2 and β3 strands, also called β2-β3 hairpin)[69,72]. By using a GTPαS analogue in PleD, the binding of a substrate within the active site of the GGDEF domain could be identified[68]. The α1 and α2 helices of each GGDEF domain form a pocket which the guanyl base binds into with hydrogen bonds formed between the base and a conserved asparagine and aspartate residue within the pocket[70] accounting for the GTP specificity over ATP[69,72].

The β- and γ-phosphates of GTPαS are bound within the active site by a lysine and arginine residue, a Mg^{2+} ion, and the (P) loop between β1 and α1[68]. This Mg^{2+} ion, as well as a second Mg^{2+} ion, is also coordinated by an aspartate (from β_1) and the third residue of the GG(D/E)EF

motif, with these two residues being essential for enzymatic activity[68]. The fourth residue of the GG(D/E)EF motif is also essential for enzymatic activity and may coordinate the second Mg^{2+} ion[67,68,72]. Within the GTPαS bound GGDEF domain of PleD, the α-phosphate does not bind with specific interactions to the protein, and is not within a coordination distance to the second Mg^{2+} ion, which may be due to the thiol modification. A model indicates that the α-phosphate is within coordinating distance to the second Mg^{2+} ion and is in a position that would allow for an attack on the α-β-phosphodiester bond[68]. Furthermore, the active site contains the side chain amino group of a conserved lysine which would be able to stabilise the transition state pentavalent phosphoryl as well as the leaving group[68].

The DGC c-di-GMP synthesis is also inhibited through a negative feedback mechanism in which c-di-GMP binds in a non-competitive manner to an allosteric inhibitory (I) site of GGDEF domains[66,67,69]. The I site is formed of a primary site (I_p) and a secondary site (I_s)[68]. The I_p site consists of an RXXD motif (when 'X' can be any residue) and is positioned upstream of the GGDEF motif by five amino acids on the α2-β2 loop[72], but is located on the opposite side of the structure to the GGDEF motif[56,67,70]. Within the I site c-di-GMP binds as a dimer in which the four guanine bases are intercalated with a closed, U-shaped, conformation[56,69] (Figure 1.4).

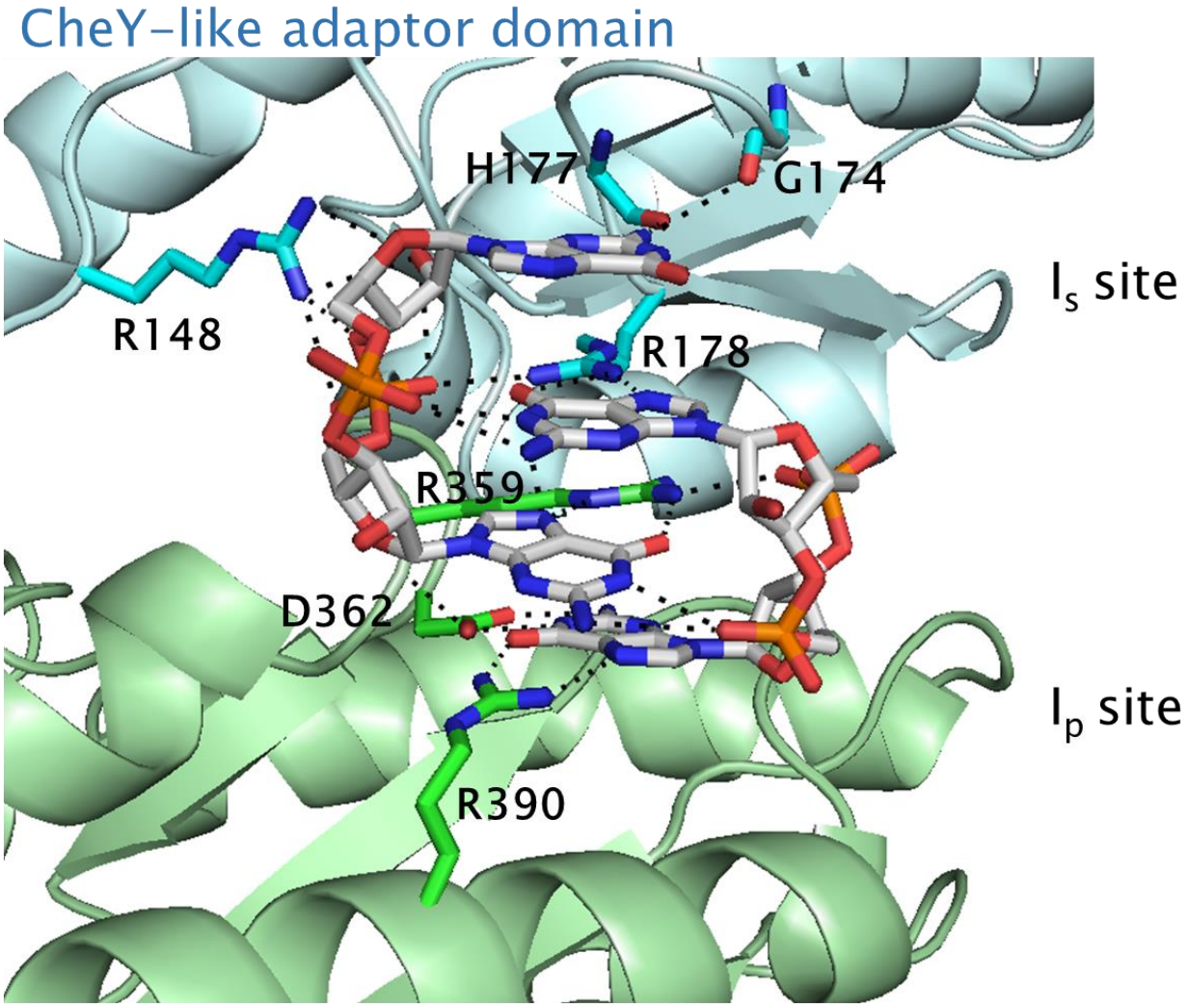

Figure 1.4. **Dimeric c-di-GMP bound within the I site of non-activated PleD.** Dimeric c-di-GMP (shown as sticks in elemental colours) binds in an inhibitory (I) site formed at the interface between the GGDEF domain (green; with a primary I site (I_p)) and CheY-like adaptor domain (cyan; with a secondary I site (I_s)) of non-activated PleD. Residue side-chains that specifically interact with the dimeric c-di-GMP within the I site are labelled and shown as sticks, except for histidine 177 and glycine 174 in which the main chain (shown as sticks) interact with dimeric c-di-GMP rather than their side chains (not shown). Arginine 359 and aspartate 362 form part of the RXXD motif in the I_p site of the GGDEF domain. Figure was made in PyMOL with PDB ID: 1W25, based on 'Figure 3' from Chan et al., 2004[69].

In crystal structures of activated and non-activated PleD, c-di-GMP has been found to bind to the allosteric I site in two different modes, either crosslinking a GGDEF domain (with the I_p site) with its neighbouring tandem CheY-like adaptor domain (with an I_s site) within a monomer[69] (Figure 1.4). Or crosslinking the two GGDEF domains within the dimer (by binding

to the I_p site of one GGDEF domain and an I_s site of the adjacent GGDEF domain (I_{sDGC}))[68]. Both of which cause product inhibition by immobilizing the GGDEF domains[68,69].

The inhibition constants of approximately 1 µM shown for PleD[68,69] and DgcA[67] are within the cellular c-di-GMP concentration range. This product inhibition would impose an upper limit on the (local) intracellular levels of c-di-GMP[69], keeping levels within concentration windows which could stabilise the c-di-GMP circuitry[67] and regulate c-di-GMP's actions in a temporal and/or spatial manner[56]. It has been suggested that high intracellular levels of c-di-GMP can cause toxicity *in vivo*[64,67]. Furthermore, a role for the I site in determining the c-di-GMP levels *in vivo* has also been demonstrated using a DgcA mutant unable to undergo a feedback inhibition[67].

However not all DGC enzymes are inhibited by c-di-GMP binding to an I site, with approximately 40% of GGDEF domains, which contain the conserved GG(D/E)EF motif, not containing the I_p site RXXD motif[67]. Yet the GGDEF domain of XCC4471 from *Xanthomonas campestris,* which does not contain an I site, had a reduced DGC activity when c-di-GMP was present[71]. Interestingly the crystal structure showed two c-di-GMP molecules as a partially intercalated dimer within the active site[71], which may demonstrate that some GGDEF domains lacking I sites are still inhibited by the c-di-GMP product but in a competitive manner[56].

Therefore GGDEF domains could potentially be inhibited *via* three mechanisms which are not mutually exclusive; one in which an inactive conformation is stabilised by c-di-GMP binding to an allosteric I site; another in which dimerization of the active half sites in each of the two monomers is prevented, possibly by regulatory or sensory domains; and one in which c-di-GMP dimers competitively inhibit the active site. Further, the I site may present a drug target for the inhibition of DCG activity which could reduce the c-di-GMP levels and prevent biofilm formation and perhaps even lead to a biofilm dispersal.

1.2.2 – C-di-GMP-specific PDEs

The c-di-GMP-specific PDE activity is due to the EAL or HD-GYP enzymatic domains, named due to the consensus amino acid sequence motif found in their respective active sites; Glu-Ala-Leu in the EAL domain[57] and the two sequence motifs His-Asp and Gly-Tyr-Pro in the HD-GYP domain[66]. PDEs containing an EAL domain are more abundant than PDEs which contain an HD-GYP domain[60]. Moreover HD-GYP domains are not ubiquitously found in bacteria[60], whereas EAL domains are[57]. Catalytically active EAL domains (approximately 250 amino acids long) hydrolyse c-di-GMP to linear 5'-phosphoguanylyl-(3',5')-guanosine (pGpG)[57,60]. Whereas catalytically active HD-GYP domains hydrolyse c-di-GMP, *via* pGpG, to two molecules of GMP[60], and it has also been suggested that HD-GYP domains can use pGpG directly as a substrate.

The PDE activity of dispersed cells was found to be 3-fold higher than the PDE-activity of biofilms and 6-fold higher than that of planktonic cells[74]. Further, Christensen et al.[75] have shown *in vitro* that inducing the overexpression of an *E. coli* c-di-GMP PDE in *P. aeruginosa* resulted in a reduced biofilm formation and the dispersal of established biofilms. The authors also report a proof of concept whereby an increase in c-di-GMP specific PDE activity and reduction in c-di-GMP levels can induce a biofilm dispersal *in vivo*[75]. Therefore understanding the regulatory mechanisms of PDE activity could highlight potential drug targets for the therapeutic induction of biofilm dispersal.

1.2.2.1 – EAL domains

The EAL domain was shown to be responsible for c-di-GMP specific PDE activity by Schmidt et al.[57] using an isolated EAL domain from the *E. coli* protein YahA. It was found that the PDE activity of the EAL domain was specific for c-di-GMP, with the EAL domain being unable to hydrolyse other cyclic nucleotides (cGMP, cAMP and c-di-AMP)[57,76]. However, recently the *Vibrio cholerae* protein, *Vc*EAL (formed only of a single EAL domain), was found to hydrolyse both c-di-GMP and cyclic-GMP-AMP (cGAMP)[77]. EAL domain hydrolysis of c-di-GMP is strongly inhibited by Ca^{2+} but is dependent on the presence of Mn^{2+} or Mg^{2+} ions[56,57,76-79]. Additionally, the EAL domain PDE activity increases with an increasing pH[78].

EAL domains have a highly conserved triosephosphate isomerase (TIM)-barrel-like fold with the topology $\alpha\beta(\beta\alpha)_7$[80,81] (Figure 1.5). The EAL motif is located on the second β strand, which runs antiparallel to the first β strand but parallel to the other β strands. The β-barrel at its C-terminal end has an extended groove to which c-di-GMP binds in its extended conformation[81], with each monomeric EAL domain capable of binding c-di-GMP.

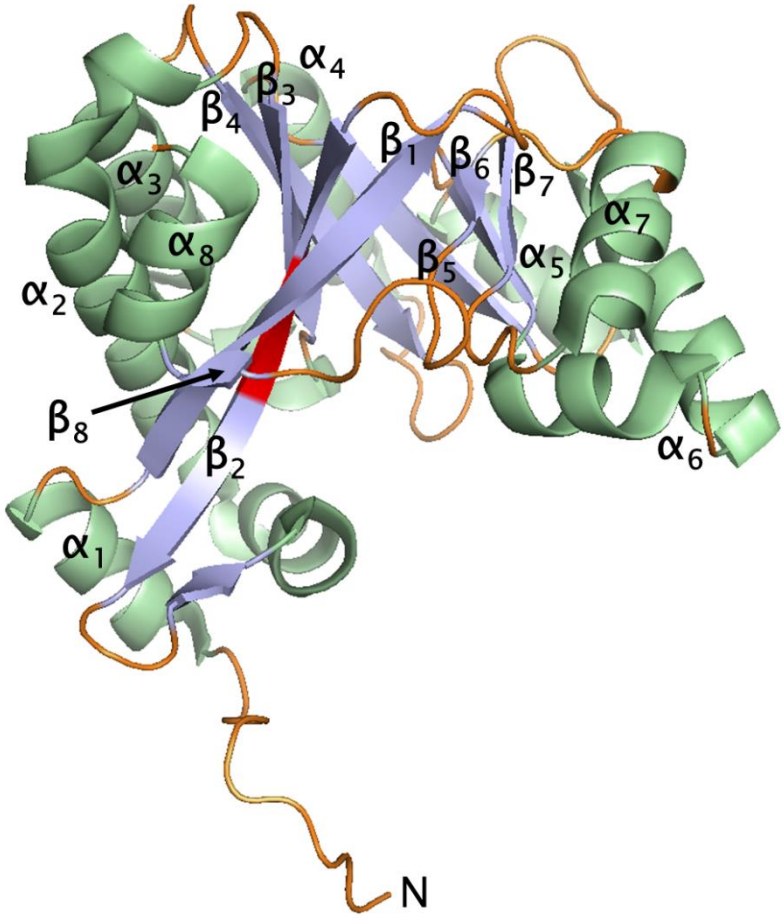

Figure 1.5. **The conserved fold of an EAL domain.** The structure of the EAL domain of YahA from *E. coli* with its non-consensus EVL motif highlighted in red. All α-helices (green) and β-strands (blue) are labelled, loops are shown in orange. Figure was made using PyMOL and PDB ID 4KIE.

Monomeric EAL domains were originally suggested to be catalytically active c-di-GMP specific PDEs[57], however, addition of c-di-GMP to monomeric EAL domains of YahA induced EAL domain dimerization[81]. Furthermore PDE activity was reduced when dimerization of the EAL domain was inhibited by site directed mutagenesis, while the binding affinity of c-di-GMP was not affected[81]. Therefore, it is now realised that EAL domain c-di-GMP specific PDEs are

catalytically active as dimers[81]. The enzymatic activity of the EAL domain is regulated by conformational changes that are induced by dimerization[73,80]. In many crystal structures of the EAL domain, dimers have been observed, however, there are variations in the interaction details and the relative orientation of the two monomers[80,81].

The canonical or 'classic' dimerisation interface of the EAL domains[80] was first shown to interact *via* three helices by Barends et al.[78] using crystal structures of the BLUF-EAL domain protein, BlrP1 from *Klebsiella pneumoniae*. Each EAL domain monomer contributes the α6-helix which run antiparallel to each other and an α5-helical segment which forms a single compound helix, where the amino termini of the helical segments meet end to end[78]. Dimer formation induces structural rearrangements that infer catalytic activity to the enzyme due to the repositioning of active site residues[73,80]. For example, when the EAL domains of MorA from *P. aeruginosa*, undergo dimerization, the α5 helix shortens at the dimerization interface so that the two Asp residues of the catalytic motif DDFGTG, located on the extended β5-α5 loop, can enter the active site to coordinate metal ions important for catalysis[73].

It has previously been proposed that EAL domains hydrolyse c-di-GMP *via* a two-metal catalysis[78]. In catalytically active complexes, the two metal ions (M1 and M2) are bridged by a water molecule (W1), which is likely responsible for giving rise to a hydroxide that then performs the nucleophilic attack on the phosphorus atom of c-di-GMP, to form pGpG[77,78]. In inactive complexes, M1 and M2 are further apart[78] , with the main differences being observed in the M2 binding site between Mg^{2+} and Ca^{2+} c-di-GMP-bound complexes[80]. Though W1 is still positioned in-line with the phosphorus atom it is only being coordinated by M1[81] , which prevents the activation of W1 and thus prevents the hydrolysis of c-di-GMP[78]. Inhibition of EAL domain PDE activity by Ca^{2+} is explained by the differences in the coordination geometry of Ca^{2+} within the M2 binding site compared to Mg^{2+} (or Mn^{2+}) ions and the differential positioning of the first Asp residue of the catalytic sequence motif DDFGTG[80].

In some structures of dimeric catalytically active EAL domains, metal ions can be observed bound in both the M1 and M2 sites with the substrate, c-di-GMP, still bound (i.e. not hydrolysed). The pGpG-bound structure of the EAL domain of PA3825, from *P. aeruginosa*, highlighted a third possible metal binding site[80]. The M1 metal from the pGpG-bound complex can be superimposed with the M1 metal in Mg^{2+} and Ca^{2+} c-di-GMP bound complexes. The M2 site was unoccupied in the pGpG-bound complex but there was a third metal binding site

that directly interacted with the pGpG, coordinated by a phosphate non-bridging oxygen and ribose O2' oxygen (as well as a catalytic Asp residue and water molecule)[80]. Furthermore, in the pGpG-bound EAL domains of CC3396 (from *Caulobacter crescentus*) and *Vc*EAL, which both have all three metal binding sites occupied, the third metal binding site (M3) aligns with the M3 binding site in PA3825[77]. Moreover, a comparison between the pGpG-bound EAL domains of PA3825 and CC3396 suggested that this third metal binding site in PA3825 is genuine, being a hallmark of the pGpG bound state, likely stabilising the transition state during hydrolysis[80].

More recently, the presence of a fourth metal ion binding site (M4) has been identified from the structure of the pGpG-bound EAL domain of VcEAL, which the authors suggest represents a second stage to the product state[77]. The M4 site is suggested by Yadav et al.[77] to become occupied after c-di-GMP is hydrolysed to pGpG, with a conformational change in the β5-α5 loop along with a reorientation of the conserved Glu131 side chain being required to form the M4 binding site. The M4-induced conformational changes results in changes to the coordination of the M2 metal ion which, when compared to the VcEAL-c-di-GMP bound structure, is shifted with the P1 phosphate group of pGpG. This increases the distance between the two hydrolysed ends of pGpG which may prevent the re-ligation of pGpG to c-di-GMP and/ or facilitate product release form the active site[77]. However, as this M4 site has been identified in only one pGpG-bound EAL domain structure and is occupied by the non-catalytic Ca^{2+} ion, it is currently unclear whether this M4 binding site is authentic and required for PDE activity or if it is an artefact.

1.2.2.2 – HD-GYP domains

The *P. aeruginosa* genome encodes three HD-GYP domain-containing proteins, two of which, PA4108 and PA4781, are enzymatically active and have been shown *in vivo* to influence the levels of c-di-GMP while the third, PA2572, is degenerate[59,60].

So far only three crystal structures of HD-GYP domains have been revealed[82,83], one of which is PA4781 from *P. aeruginosa*[60]. The other two are Bd1817 from *Bdellovibrio bacteriovorus* (which was the first HD-GYP domain to be crystallised)[82] and *Pm*GH from *Persephonella*

marina[83]. Of these three HD-GYP domain containing proteins to be crystallised, two (*Pm*GH and PA4781) are catalytically active as c-di-GMP PDEs[60,82,83]. Bd1817 is degenerate which is probably due to the non-consensus active site residues[84] (for example, the tyrosine of the GYP signature is deleted[82]).

The structures of the three HD-GYP domains display an all-alpha fold, being formed of 7 alpha helices[60,82,83] including the 5-helix core, characteristic to the HD domain superfamily. This suggests that this 7 helix all-alpha fold is likely universal for HD-GYP domains, although structures of further HD-GYP domains are needed.

Furthermore, in each of the three HD-GYP domain-containing-proteins to be crystallised so far, the HD-GYP domain is separated from an N-terminal domain by a helical linker region. With Lovering et al.[82] pointing out that as the N-terminal and HD-GYP domains are physically distinct, it would be easy for the N-terminal (sensory) domain to differ allowing for distinct signalling pathways. For example, PA4781 is formed from an N-terminal regulatory domain (a REC domain), a sensing helix and the HD-GYP domain (forming the full length protein REC-helix-HD-GYP, which is activated when the REC domain becomes phosphorylated)[60,84].

The HD motif is located on the second alpha helix of the HD-GYP domain in all three crystallised HD-GYP domains[60,82,83]. The GYP motif is found in similar locations for PA4781, *Pm*GH and Bd1817, that is, on the long loop between the fourth and fifth alpha helix of the HD-GYP domain[60,83]. This GYP loop is the most conserved region within HD-GYP domains (and has the highest degree of structural similarity between the PA4781 and *Pm*GH structures) which suggests its involvement in the function(s) of HD-GYP domains[60] which may not be directly related with catalytic activity. It has been shown by alanine substitutions that the GYP motif of the HD-GYP domain protein, RpfG, interacts with GGDEF domain partner proteins[85]. Interestingly, mutations of RpfGs GYP motif had no effect on its enzymatic activity[83].

For PA4781, only the HD-GYP domain was found to crystallise (after an autoproteolysis of the full-length protein) and was found as a dimer[60]. Although, from using gel filtration, the HD-GYP domain of PA4781 had previously been found to exist mainly as a monomer with full length PA4781 being dimeric[84]. While *Pm*GH was also found to crystallise as a dimer, the authors explained that this was due to the N-terminal GAF domain and helical linker and not to the HD-GYP domain[83]. From four different crystal forms of Bd1817 and analysis of their

protein-protein contacts, the authors suggest that Bd1817 is likely monomeric[82]. Therefore there seems to be variation in the oligomeric states (at least between these three structures) of different HD-GYP domains.

The active sites of the three HD-GYP domains described so far all differ from one another, with the active site of *Pm*GH containing a tri-ion centre, while PA4781 and Bd1817 have bimetallic centres but display their two metal ions in different positions with different binding modes[60]. This led Bellini et al.[83] to propose the division of the HD-GYP family into two evolutionary subgroups (those with a bimetallic and those with a trimetallic centre) based on a phylogenetic comparison which was independent of the type of associated sensory and/ or regulatory domains. Interestingly however, the position and coordination of the first metal is conserved within the three metal binding sites and it is the differential binding of the additional metal ion(s) that gives the variation[60].

The *Pm*GH structure relieved a tri-iron centre with an extended V shape that bound c-di-GMP in a *cis*-conformation. Moreover, the binding site allows for both hydrolysable phosphates of c-di-GMP to interact with the metal centre in turn, allowing for c-di-GMP to be hydrolysed into two molecules of GMP[83].

In PA4781, Glu314 replaces a consensus Alanine residue, so when c-di-GMP (in the *cis* conformation – as observed in *Pm*GH) is modelled into the binding cleft there is a steric clash with c-di-GMPs distal phosphodiester moiety[60,84]. Furthermore, an E314A mutant of PA4781 had a much higher (approximately 17-fold) affinity for c-di-GMP than compared to WT PA4781[84]. However it was suggested that pGpG, due to the absence of a ring constrain (present in c-di-GMP), could be the substrate of PA4781 which was found to have a much higher affinity for pGpG than compared for c-di-GMP[84]. Later, from the structure of PA4781, it could be shown that pGpG could bind within the active site with two possible binding modes, due to being able to enter in two possible orientations[60]. In both cases, the bridging phosphate of pGpG replaces a water ligand and interacts with the bimetallic centre. Also, the phosphate bridging oxygen is close enough for hydrogen bonds to form with Lys224 (thought to protonate the leaving group)[60]. Therefore, the structure of PA4781 appears to be better adapted for binding pGpG instead of c-di-GMP. However, due to a low catalytic activity, the primary function of PA4781 was suggested by Rinaldo et al.[60] not to be the hydrolysis of nucleotides. As a result of the different structural features (which could be due to the

differences in their functions) observed in these three structures, further structures of HD-GYP domains are required.

1.2.3 – Tandem arrangements of c-di-GMP metabolising domains

The GGDEF and EAL or HD-GYP domains are frequently found arranged in tandem within the same protein[56,60], with EAL domains often being found C-terminal to the GGDEF domain[57,86]. Furthermore these enzymatic domains are often found in tandem with regulatory or sensory domains[60]. Thus, while both domains may be capable of enzymatic activity, the activity of these domains could be controlled by a signal so that only one domain is active at any point[56]. Nevertheless, it is common for one of the two tandem domains to be catalytically degenerate[56]. However, in some GGDEF-EAL tandems, when one of the domains is degenerate and catalytically inactive, the GGDEF and EAL domains still have the ability to bind GTP and c-di-GMP respectively, which may have a structural and/or regulatory role[56]. Additionally, HD-GYP domains that are degenerate may also serve as receptors for c-di-GMP or pGpG[60].

Different bacterial genomes contain very varied numbers of proteins with GGDEF, EAL or HD-GYP domains[62]. For example 37 proteins with one or more of these domains are encoded within the *X. campestris* genome[71] and 29 are encoded within the *E. coli* K-12 genome[70]. The *P. aeruginosa* PAO1 genome encodes 17 GGDEF domain containing proteins, 5 EAL domain containing proteins, 16 proteins that contain both GGDEF and EAL domains[87] and 3 proteins that contain a HD-GYP domain.

Within the cell, local pools of c-di-GMP are thought to be regulated by specific DGCs and PDEs to mediate discrete c-di-GMP signalling pathways[88]. On the other hand, as a likely freely diffusible messenger, the levels of c-di-GMP within the cell may change globally, but how this global change in c-di-GMP levels could facilitate specific pathways and not others is unclear[87].

Being able to manipulate and lower the intracellular levels of c-di-GMP, through its degradation (activation of PDEs) and/or inhibition of its synthesis (inhibition of DGCs), or direct manipulation of downstream c-di-GMP signalling pathways, with a drug or small molecule, could represent a novel anti-biofilm therapeutic.

1.3 – Inducing the dispersal of *P. aeruginosa* biofilms

Being a highly diffusible, uncharged, diatomic gas molecule with a single unpaired electron, nitric oxide (NO) at low concentrations is an important signalling molecule in eukaryotes[31,89]. However it is now realised that signalling (non-toxic) concentrations of NO can also be detected by a variety of bacteria, having implications in bacterial biofilm formation and dispersal.

In 2006 Barraud et al.[58] demonstrated that NO is involved in *P. aeruginosa* biofilm dispersal. By growing *P. aeruginosa* mutant biofilms in glass flow cells and comparing them to a wild-type, the authors showed that Δ*nirS* mutants, which lack nitrite reductase and are therefore unable to produce metabolic NO, developed thicker and confluent biofilms with a reduced dispersal. While Δ*norCB* mutants, which lack NO reductase and are therefore unable to remove metabolic NO, contained many hollow voids within the biofilm and displayed a greater amount of cell death. The effluent from Δ*norCB* mutant biofilms also contained an increased number of dispersed viable cells. Furthermore, the addition of sublethal concentrations of exogenous NO to *P. aeruginosa* biofilms induced a biofilm dispersal. Additionally, when this NO induced dispersal was combined with antibiotic treatment, there was a greater biofilm removal from the surface than compared to antibiotic treatment on its own[58]. Sublethal concentrations of NO, when NO donors were added to agar, also increased the swimming and swarming motility of *P. aeruginosa*[52].

Barraud et al.[58] in 2009 demonstrated that NO induced dispersal of *P. aeruginosa* biofilms occurred *via* the stimulation of PDE activity and decreases in the overall intracellular levels of c-di-GMP. This has since been confirmed by other groups[90]. After confirming that GTP could inhibit PDE activity in *P. aeruginosa* cells, GTP was shown to reduce the NO-mediated dispersal of *P. aeruginosa* biofilms in a dose-dependent manner, where GTP alone did not increase the biofilm surface coverage[58]. Using bis(*p*-nitrophenyl) phosphate, a substrate specific to PDEs, low concentrations (500 nM to 50 μM) of the NO-donor sodium nitroprusside (SNP) was found to increase the overall PDE activity in cell extracts of *P. aeruginosa*, whereas SNP at higher concentrations (5 mM and 50 mM) inhibited PDE activity[58]. This corresponds to the earlier finding that low non-toxic concentrations of SNP (nano- and micromolar) induced a biofilm dispersal of *P. aeruginosa* whereas higher concentrations (millimolar) stimulated

biofilm formation[52]. When quantifying intracellular c-di-GMP levels with liquid chromatography-tandem mass spectrometry, the authors showed that the level of c-di-GMP in 5-day-old biofilms and in planktonic cultures of *P. aeruginosa* decreased when exposed to SNP[58].

Although there are contradictory studies which suggest that NO has either no effect or promotes biofilm formation[91], other studies support the finding that *P. aeruginosa* biofilm dispersal can be induced by NO[74,90,92]. Furthermore this NO-induced dispersal has now also been observed in a variety of other bacterial species[45,93].

Although the delivery of gaseous NO alongside antibiotic usage shows promise in the treatment of pulmonary *P. aeruginosa* biofilm infection in CF patients[94], NOs short half-life within a host prevents it reaching other sites of infection (such as indwelling medical devices)[89]. In a further study by Barraud et al.[45] the authors have shown that an NO-donor prodrug, a cephalosporin-3'-diazeniumdiolate, which releases NO upon reaction with the bacteria specific enzyme β-lactamase, is effective in triggering *P. aeruginosa* biofilm dispersal *in vitro*. Interestingly, cell lysates of non-β-lactamase-producing *E. coli* cells also stimulated NO release from the cephalosporin-3'-diazeniumdiolate (albeit at lower levels than compared to β-lactamase containing *P. aeruginosa* cell lysates), which the authors suggested was due to transpeptidases[45]. However, biofilm dispersal is not only induced by NO[31] and it has also been shown that biofilms may disperse in response to various environmental factors, such as a rapid increase in nutrients[95], or a reduction in nutrient or oxygen levels[96].

1.3.1 – A mammalian NO sensor – soluble guanylyl cyclase

In mammals (and also insects and fish) NO is involved in many physiological processes such as neurotransmission, relaxing smooth muscle[97] and platelet aggregation[98,99]. Understanding how NO is sensed within eukaryotic systems could provide a paradigm for how NO is sensed within prokaryotic systems. Guanylyl cyclase is an enzyme that synthesizes cyclic guanosine monophosphate (cGMP) from guanosine triphosphate (GTP). The secondary messenger, cGMP, then triggers an array of downstream signalling cascades that ultimately leads to a

physiological output[99]. There are two major classes of guanylyl cyclase, a membrane-bound and a soluble form[99].

Soluble guanylyl cyclases increase the synthesis of cGMP in response to an activating NO signal[98]. These soluble enzymes can be either homo- or heterodimeric, being formed of α and β subunits[99]. There are several isoforms of the α and β subunits (in vertebrates the α1, α2, β1 and β2 subunits are known[100]), which are distributed in distinct tissues and cells. Each subunit is formed of; a haem-NO/O_2-binding (H-NOX) domain, a Per-Arnt-Sim (PAS) domain, a coiled-coil (CC) domain and a catalytic domain[98,100]. In the best studied soluble guanylyl cyclase, the α1β1 heterodimer, only the H-NOX domain of the β1 subunit is associated with a haem prosthetic group[98,100].

The enzymatic activity of effector proteins which are associated with H-NOX proteins, are regulated by the binding of NO to the H-NOX domain, as shown by biochemical studies[101]. When NO (reversibly) binds to the ferrous *b*-type haem cofactor in the H-NOX domain of the β subunit, this forms a penta-coordinated Fe^{II}-NO complex[99] and displaces the proximal Histidine residue which ligates the haem cofactor[98,100]. This induces a conformational change that is transduced from the H-NOX domain through to the catalytic domain, but the molecular detail of this mechanism is still unclear[98]. However the catalytic domain undergoes a conformational change from an inactive 'open' to an active 'closed' state, as shown by fluorescence resonance energy transfer (FRET) *in vitro* and fluorescence lifetime imaging based FRET *in vivo*[98]. The active site is formed at the interface between two catalytic domains (one catalytic domain from each subunit)[100]. In this way, the catalytic activity of soluble guanylyl cyclases is increased by up to 200-fold due to NO binding to its haem cofactor[100]. H-NOX forms a conserved family of proteins which can be found in both eukaryotes and prokaryotes. Proteins belonging to the H-NOX family are able to discriminate NO from molecular oxygen which is vital to the function of H-NOX proteins as NO sensors[31].

1.3.2 – Bacterial NO sensory domains

1.3.2.1 – The H-NOX domain

It has been found that a H-NOX protein, *Sw*H-NOX, exists within *Shewanella woodyi*, that when knocked out results in a reduced biofilm formation compared to a WT[102]. Furthermore, when WT biofilms were grown in the presence of non-toxic concentrations of NO, there was a decrease in the level of biofilm formation compared to when WT biofilms were grown in the absence of NO[102]. In contrast, the reduced biofilm formation of Δ*hnox* mutants was unaffected by the presence or absence of NO. However when mutants were complimented with the *hnox* gene, biofilm formation and NO sensitivity was restored to WT levels. Moreover, NO exposure was found to cause a decrease in the levels of c-di-GMP in the WT. With c-di-GMP levels of the Δ*hnox* mutant being lower than the WT, whilst being unaffected by the presence or absence of NO[102].

Pull down assays have indicated that *Sw*H-NOX interacts with *Sw*DGC, which has been shown to have both DGC and PDE activities, due to a GGDEF and an EAL domain respectively[102]. Using *Sw*DGC variants where either the GGDEF or EAL domains were inactive, it was found that NO-bound *Sw*H-NOX stimulates the PDE activity of *Sw*DGC, whereas unligated *Sw*H-NOX was found to stimulate the DGC activity of *Sw*DGC[102]. However the enzymatic activity of native *Sw*DGC with its GGDEF and EAL domains both active together, in the presence of unligated or NO-bound *Sw*H-NOX, was not shown in this study.

Nevertheless, this study suggests that in the absence of NO, unligated *Sw*H-NOX interacts with *Sw*DGC to increase the levels of c-di-GMP. But upon *Sw*H-NOX sensing NO, the intracellular levels of c-di-GMP are reduced by *Sw*DGC due to an increase in its PDE activity and a decrease in its DGC activity[102]. In this way *S. woodyi* can couple an NO signal to the c-di-GMP level to regulate biofilm formation.

It has been predicted that H-NOX domains are found in over 300 species of bacteria[31,99]. Although in *P. aeruginosa* no (known) H-NOX protein exists. Nevertheless, *P. aeruginosa* encodes for putative NO sensory domains that are also often found in tandem with GGDEF and/or EAL or HD-GYP domains.

1.3.3 – Putative NO sensory domains in *P. aeruginosa*

1.3.3.1 – FIST domains

FIST (F-box and intracellular signal transduction proteins) domains are found largely in bacterial genomes, often encoded for within operons along with signalling proteins such as histidine kinases, DGCs and PDEs[103]. But FIST domains are also found in a few eukaryotic and archaeal genomes. FIST domains mainly comprise single-domain proteins, but some are also found in combination with other domains such as GGDEF, EAL and PAS domains[104]. The FIST domain is divided into N- and C-terminal subdomains[104]. With the full-length FIST domain being predicted to be formed of 20 β-strands and 7 α-helices[104].

Recently, Hossain and Boon[103] suggested that a FIST domain of a *P. aeruginosa* protein, PA1975, which the authors later termed NosP (NO-sensing protein), was able to bind haem and sense NO. Haem binding of NosP was shown through pulldown assays using haem-agarose, and UV/vis spectra of purified NosP also gave rise to a Soret band which is consistent with haem binding. Furthermore, the authors showed that the N-terminal half of the FIST domain was sufficient for haem binding[103]. This study reported NO dissociation rates of NosP to be similar to the NO dissociation rates of H-NOX domains (including soluble guanylyl cyclase and *Sw*H-NOX)[102,103].

However these authors could not show that NosP was involved in *P. aeruginosa* biofilm regulation, as NosP KO mutants were unable to be produced, although this was not believed to be due to *nosP* being an essential gene[103]. The authors did show however that a putative interacting protein NahK, standing for NosP-associated histidine kinase, is involved in an NO-mediated biofilm dispersal[103].

However if NosP is an NO sensor involved in biofilm dispersal, it does not suggest that it is the only NO sensor within *P. aeruginosa*. In fact there is likely a redundancy in the system.

1.3.3.2 – MHYT domains

MHYT domains, so called due to a conserved pattern of amino acids, are predicted by sequence analysis to be comprised of six transmembrane helices, located in the inner membrane, connected *via* short cytoplasmic loops rich in arginine residues and periplasmic loops rich in charged residues[105] (Figure 1.6). MHYT domains can form single domain proteins or they can be found in tandem with signalling domains, for example MHYT domains are sometimes found in tandem with histidine kinase domains but are more usually found in tandem with GGDEF and EAL domains[105]. Moreover, Galperin et al.[105] suggested that this domain may be involved in sensing oxygen, CO or NO through the coordination of one or two copper ions by the conserved Met and His residues located towards the periplasmic face in the second, fourth and sixth helices of the MHYT domain.

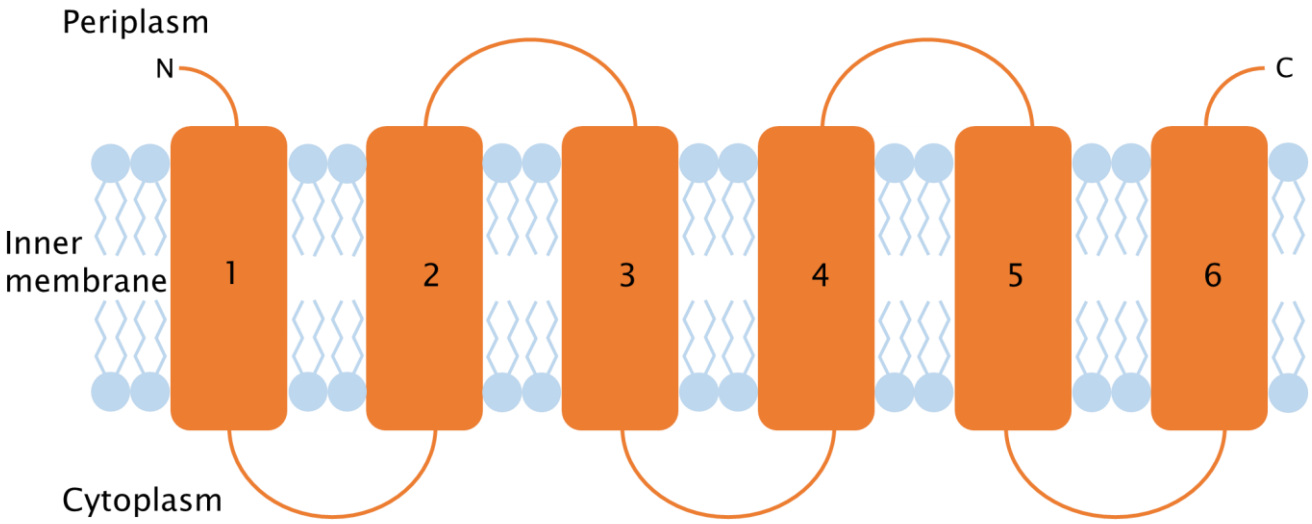

Figure 1.6. **The orientation of a MHYT domain**. Six transmembrane segments predicted to form a MHYT domain are shown as orange rectangles with the periplasmic and cytoplasmic loops shown as orange semi-circles. A highly conserved MHYTXM motif is found on the second, fourth and sixth transmembrane segments located towards the periplasm [106].

Although not all bacterial species encode for MHYT domains, they can be found in phylogenetically distant bacteria[105]. The genome of *P. aeruginosa* PAO1 encodes for two proteins that contain MHYT domains, namely MucR (for mucoid alginate regulator; PA1727)

and NbdA (for NO-induced biofilm dispersion locus A; PA3311) which also contain cytoplasmic GGDEF and EAL domains[90,105].

MucR has seven N-terminal transmembrane regions where the MHYT domain is predicted to be formed of the second to the seventh transmembrane helices[106]. The use of reporter enzymes confirmed that MucR is located at the cytoplasmic membrane, with the C-terminal GGDEF and EAL domains located in the cytosol[106]. Previously it has been shown that MucR is involved in regulating the production of the exopolysaccharide alginate, where an alginate overproducing *P. aeruginosa* strain (PDO300) had a non-mucoid phenotype and approximately a 38 fold reduction in alginate production when the *mucR* gene was deleted[106]. Moreover, when MucR was re-introduced into the Δ*mucR* mutant, the production of alginate increased to levels that exceeded those of the WT and this was found to be dependent on the C-terminus containing the GGDEF and EAL domains[106].

It has been shown that NbdA has PDE activity and that MucR has both PDE and DGC activity, evidenced both *in vitro* and *in vivo*[90]. Moreover the PDE and DGC activity of MucR seem to be inversely regulated *in vivo* depending on the bacterial lifestyle[90]. Furthermore although the GGDEF domain of NbdA (which has the non-canonical AGDEF motif) does not possess DGC activity, it inhibits the PDE activity of NbdA in the absence of GTP[90].

Biofilms of *P. aeruginosa* mutants lacking either *nbdA* or *mucR* did not disperse when exposed to NO, with biofilm architecture also being unaffected by the NO exposure as shown by confocal microscopy[90]. The addition of NO had no significant effect on the PDE activity or the c-di-GMP levels in Δ*nbdA* mutant biofilms in contrast to NO addition to a WT biofilm. However as MucR is present in the Δ*nbdA* mutant, this led the authors to suggest that the c-di-GMP level and PDE activity alterations were NbdA (but not MucR) dependent[90]. Quantitative real-time PCR has suggested that NO regulates the transcription of *nbdA*, where the *nbdA* transcript levels were significantly increased within NO induced dispersed cells compared to untreated biofilms and planktonic cells treated with NO[90]. However if this NO induced biofilm dispersal relies on *nbdA* being transcribed, this would suggest that NbdA is not acting as the primary NO sensor, although NO may still act to stimulate the PDE activity of NbdA.

1.3.3.3 – Per-ARNT-Sim (PAS) domains

Per-ARNT-Sim (PAS) domains are named due to being first recognised in the proteins period (Per) and single-minded (Sim) from *Drosophila* and the aryl hydrocarbon receptor nuclear transporter (ARNT) found in vertebrates[107-109]. PAS domains were originally divided into a PAS and PAC motif (also known as S1 and S2 boxes), however as these two motifs make up a single globular fold, together they are known as a PAS domain[108,110].

All kingdoms of life contain PAS domains[108,109,111], but PAS domains are not found in all species[112]. PAS domains can be found in the cytoplasm[108,112] or the periplasm[113,114] and consist of approximately 100 to 120 amino acids[107]. The canonical PAS fold (shown in Figure 1.7) comprises of a central conserved antiparallel five-stranded β sheet with a strand order of 2-1-5-4-3 and several, more variable, intervening α helices that flank the sheet, with the secondary structure of Aβ Bβ Cα Dα Eα Fα Gβ Hβ Iβ[110,111]. The length, orientation and number of α helices between the β strands can differ, which is in contrast to the conserved β strands[107]. Although the primary sequence of different PAS domains are highly variable, without any residues being conserved universally, their structural homology is often remarkably similar[111,113].

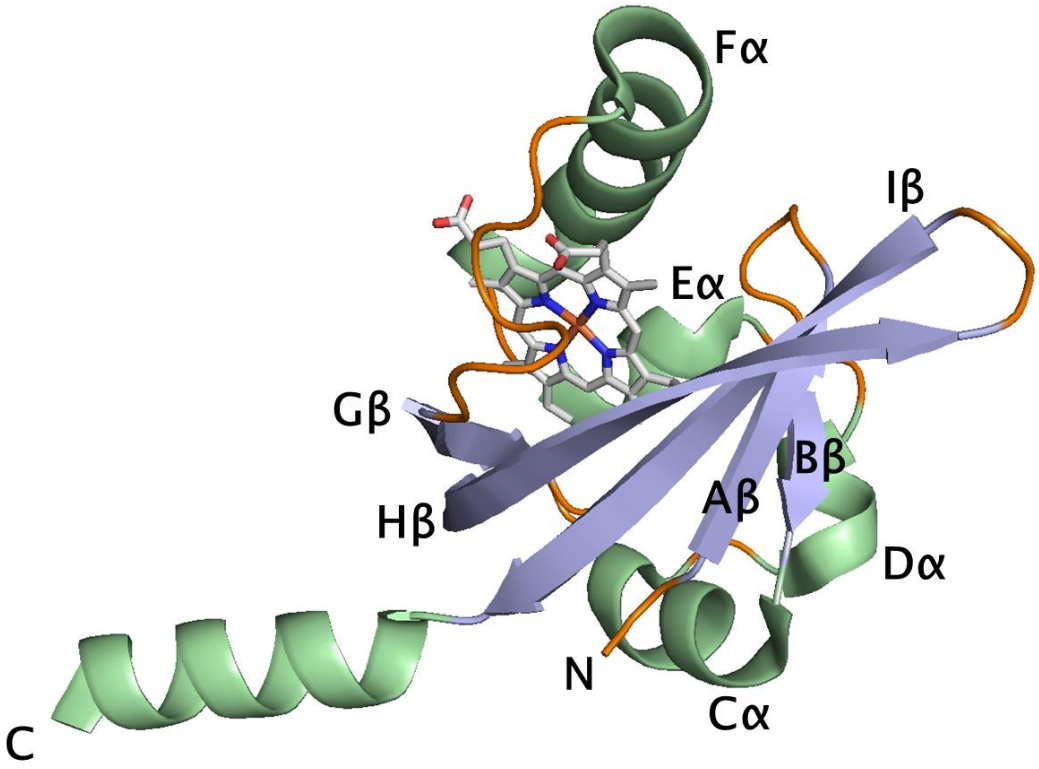

Figure 1.7. **Canonical fold of a PAS domain.** The PAS domain of FixL from *Bradyrhizobium japonicum* with α-helices (green) and β-strands (blue) labelled and loops in orange. A haem *b*-type cofactor, without a ligand, is also bound within the PAS domain and is shown in element colours. Figure was made with PyMOL and the PDB ID: 1XJ3.

Many PAS domains function as sensory domains and can covalently or non-covalently bind an array of molecules and ions which act either as a direct chemical ligand (such as carboxylic acids[113,114], glucose and divalent metal ions) or as a cofactor[115]. Stimuli such as redox potential, voltage, light and oxygen can be sensed by PAS domains with bound cofactors flavin adenine dinucleotide (FAD), flavin mononucleotide (FMN), *p*-coumaric acid and haem[87,116,117] (in its *b* form when cytosolic or *c* form when periplasmic) respectively[107,118]. In PAS domains, most cofactors are bound in a cleft, which is spatially conserved, between the β sheets inner surface and the E and F α helices[107]. It is the variation in the Fα helix that allows the range of different cofactors to be accommodated in different PAS domains[111].

Typically, one or more PAS domains are found N-terminally within multi-domain proteins, being linked to and able to regulate many different effector domains[107,112] (such as histidine kinases, methyl-accepting domains, the GGDEF domain and the EAL domain[117]). For example, in the bifunctional histidine kinase/phosphatase, CckA, the switch between its phosphatase

and kinase activities is mediated by an N-terminal PAS domain and its interaction with PAS domains of DivL[119]. This switch in the activities of CckA was evidenced to be due to the interactions between the PAS domains of CckA and DivL, PAS B and PAS B-D respectively, as variants which lacked these PAS domains were unable to influence the enzymatic activities of CckA[119]. Moreover, PAS domains can mediate the oligomerisation state of a protein and can hetero- or homo-dimerise through several different arrangements involving an interface formed by the PAS core and an amphipathic N-terminal α-helical extension (the PAS N-cap)[115,118]. In multi-domain proteins that contain more than one PAS domain, different signals may be integrated[108] to fine tune the output[119]. It has been suggested that PAS domains also have a role in modulating the affinity of proteins for other proteins and this modulation could be signal dependent[107]. Recently, Ortmayer et al. have reported a haem-bound PAS domain to be enzymatically active as an oxidative *N*-demethylase[120]. Therefore, sensing stimuli, protein-protein interactions, oligomer formation[115] and even enzymatic activity[120] can be mediated and regulated by PAS domains, thus enabling the PAS domain to regulate diverse cellular processes.

The oxygen sensing protein, FixL from *Bradyrhizobium japonicum*, contains an N-terminal haem-binding PAS domain (Figure 1.7) which regulates a C-terminal histidine kinase domain that belongs to a two-component regulatory system[111,121]. Oxygen binds to the distal side of the haem cofactor, which is located in a predominantly hydrophobic pocket where its iron is coordinated by a proximal histidine residue which serves as its axial ligand[111,112]. The axial histidine residue is rigid but the binding of oxygen induces the puckered porphyrin ring to flatten[111]. This is then transduced, due to a shift of the haem propionate side chains and a rotation of a conserved arginine into the haem pocket, to induce conformational changes in the PAS domain, the largest taking place at the FG-loop region (FG loop switch) which is proposed to interact with the kinase domain[111,122,123]. Thus upon oxygen binding to the haem of the FixL PAS domain, the induced conformational changes inhibits the histidine kinase domain[124]. Without a ligand bound to the haem of the FixL PAS domain, the histidine kinase domain is active and undergoes auto-phosphorylation using the γ-phosphate from ATP[111].

The haem of FixLs PAS domain also binds NO. However NO binding was unable to induce the FG-loop shift[122], as the conserved arginine residue, which forms a hydrogen bond to the oxygen ligand in the oxygen bound structure causing the arginine to rotate into the haem

pocket, cannot hydrogen bond to the NO ligand[123]. Thus the arginine does not rotate into the haem pocket in the NO bound structure, and so cannot form a steric clash that would normally result in the conformational change in the FG loop[123]. However, NO binding to FixL partially inhibits kinase activity, suggesting the involvement of other factors and that the FG loop switch is not the only form of regulation that influences the kinase domain[124].

Out of the 41 proteins in *P. aeruginosa* PAO1 that contain a DGC motif or a PDE motif or both, 13 of these proteins also contain one or more PAS domain. Therefore, these 13 *P. aeruginosa* c-di-GMP metabolising proteins may be regulated by their PAS domain(s). Further, as haem is able to bind NO and PAS domains can bind haem, it is reasonable to hypothesize that the DGC and/or PDE activities of one or more of these 13 proteins are regulated by NO to induce the *P. aeruginosa* biofilm dispersal. As in the *Sw*H-NOX – *Sw*DGC system, the activities of tandem GGDEF and EAL domains may be inversely regulated by NO binding to an N-terminal PAS domain(s).

1.3.3.4 – CHASE domains

The first identification of the Cyclases/histidine kinases associated sensory extracellular (CHASE) domains were evidenced in two back-to-back publications by Anantharaman and Aravind[125] and by Mougel and Zhulin[126]. CHASE domains are predicted to act as sensory domains, being found N-terminally to signalling domains[125,126]. The CHASE domains are predicted to bind a variety of low molecular weight ligands (such as short peptides[126] or cytokinin-like adenine derivatives), which are putatively involved in an organisms developmental program[125,127].

CHASE domains were identified in prokaryotes, plants and lower eukaryotes but found to be exclusive to signal transduction proteins, such as histidine kinases, methyl-accepting chemotaxis proteins and adenylyl or diguanylate cyclases with or without a phosphodiesterase domain[125-127]. CHASE domains are always found in the periplasmic or extracellular regions of proteins, flanked by transmembrane regions[126]. There have been eight classes of CHASE domains identified so far, designated CHASE2 through to CHASE9[128], with the *P. aeruginosa* PAS-GGDEF-EAL protein, PA2072, predicted to contain the CHASE4

domain[127]. The amino acid length and predicted secondary structure of the CHASE domains varies between classes. For example, the CHASE2 domains are predicted to be formed of six α-helices and eight β-strands and are 250 to 300 residues long, while the CHASE3 domains are predicted to be formed entirely of four to six α-helices and are 130 to 150 residues long[127]. The CHASE4 domains are comprised of 150 to 160 residues and have an αβ fold, consisting of six β-strands which are encompassed by extended α helices on each boundary and a central loop of 20 to 45 residues[127]. The boundaries of the CHASE domain exhibit a low amount of conservation compared to the strongly conserved motifs found in the central part of the domain[125].

1.4 – Targets potentially implicated in the NO-induced biofilm dispersal pathway

1.4.1 – PA0861 (RbdA)

Through transposon mutagenesis of *P. aeruginosa*, An et al.[129] found that the gene *pa0861* was involved in biofilm development. It was shown that *P. aeruginosa* mutants of *pa0861* display an increased biofilm formation compared to a WT[87,129], which our lab has also shown using *pa0861* knockout mutants in PAO1[130]. Furthermore, while growing WT PAO1 biofilms and *pa0861* KO biofilms, after 6 hrs of incubation the biofilm mass of the WT started to decline whereas the biofilm of the *pa0861* mutant continued to accumulate for 48 hrs[129]. This suggests that normally *pa0861* negatively influences *P. aeruginosa* biofilm formation[131]. An et al.[129] designated the *pa0861* gene *rbdA* (regulation of biofilm disposal). The gene, *rbdA*, encodes an 818 amino acid long protein (see Figure 1.8) formed of two transmembrane domains at its N-terminus, anchoring RbdA to the inner membrane, followed by a putative sensory PAS domain, a GGDEF and an EAL domain[129,131].

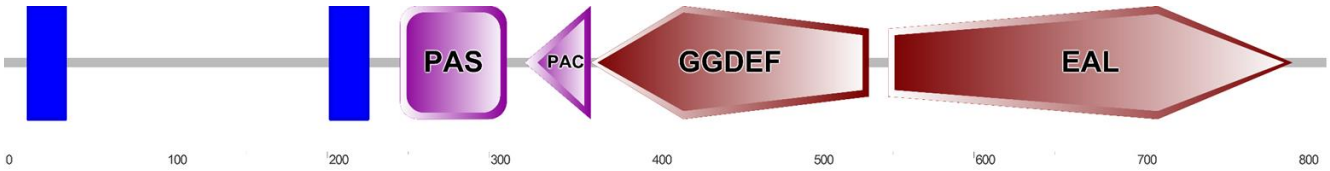

Figure 1.8. **Domain organisation of (full length) RbdA**. RbdA is 818 amino acids in length, formed of two transmembrane domains (indicated by vertical blue rectangles), a single PAS domain (formed of both the PAS (pink square) and PAC (pink triangle)) and a GGDEF and EAL domain (indicated by the red pentagons). Grey lines indicate regions of unknown. Domain organisation was predicted by and figure acquired from SMART[132].

It has been shown that biofilms of the *P. aeruginosa* Δ*rbdA* mutant could not disperse in response to glutamate, mercury chloride, ammonium chloride and NO[74]. It has also been found in our own lab that biofilms of *P. aeruginosa* Δ*rbdA* KO mutants could not undergo a biofilm dispersal in response to NO, whereas WT *P. aeruginosa* biofilms could[130]. This indicated that RbdA may be involved in the NO induced biofilm dispersal mechanism and therefore required further investigation. RbdA contains a GGDEF and EAL domain arranged in tandem with an upstream putative sensory PAS domain. When the EAL motif was substituted for alanine residues in the chromosomal copy of *rbdA* in *P. aeruginosa* PAO1 there was a three-fold increase in biofilm formation[131]. This suggests that the c-di-GMP specific PDE activity of the EAL domain in RbdA down-regulates biofilm formation under normal circumstances[131]. However the possibility of the RbdA EAL mutant being less stable than the WT RbdA protein cannot be ruled out and thus the phenotype observed may be due to loss of the entire protein[131]. The GGDEF domain of RbdA contains the consensus GGDEF signature motif along with the I site RXXD amino acids found upstream of the GGDEF motif by the conserved 5 amino acids[129]. The EAL domain of RbdA contains the signature ELL, which is also found in the functional phosphodiesterase YhjH from *Salmonella enterica*[129]. Furthermore, the EAL domain also contains residues important for substrate-binding (such as Gln569 and Arg589), coordination of the divalent cation and the active site (Lys730)[129].

Therefore, this would suggest that RbdA could have both opposing DGC and PDE activities. Consistent with this, An et al.[129] found that RbdA progressively degraded c-di-GMP to pGpG as shown by high-performance liquid chromatography (HPLC) analysis (i.e. RbdA has c-di-GMP PDE activity). Originally it was suggested that RbdA did not poses DGC activity as c-di-GMP

production could not be directly observed[129]. Recently however, Liu et al.[131] measured DGC activity for RbdA using a mutant which lacked PDE activity. Moreover the DGC activity of an RbdA mutant lacking PDE activity was further increased when the I-site of the GGDEF domain was also inactivated[131].

Additionally, the presence of GTP enhanced the PDE activity of RbdA which was dependent on the GGDEF motif, as when this motif was substituted with alanine residues, the mutant degraded c-di-GMP at rates which were similar when GTP was present or absent[129]. To ensure an enhancement of PDE activity by GTP could not be attributed to an increase in the concentration of c-di-GMP (due to the intrinsic DGC activity of RbdA), the non-hydrolysable GTP analogue Guanosine 5'-β-γ-imido triphosphate (GMPPNP) has also recently been tested[131]. It was found that the PDE activity of RbdA increased as the GMPPNP concentrations were increased, which again was found to be dependent on the A site of the GGDEF domain[131].

Recently the dimeric structure of the cytosolic region of RbdA (cRbdA) consisting of the PAS-GGDEF-EAL domains has been determined by Liu et al.[131] in the apo, GTP bound and c-di-GMP bound states (the apo-sate, PDB code 5XGB, is shown in Figure 1.9). The fold of the PAS domain (residues 255 - 360) of RbdA varies slightly from convention with an antiparallel six-stranded β sheet with the extra 3 residue long β strand being found between Eα and Fα helices[131]. Within the interior of the PAS domain is a hydrophobic pocket in which the side chains of aliphatic residues form a cavity with the potential to bind small organic ligands[131]. However, in the PAS domain sequence of RbdA a haem-coordinating Histidine residue is absent and homologous 3D structures, revealed by an automated database search, also lacked bound haem[131]. Yet, when the PAS domain of RbdA was expressed in isolation a weak haem binding was observed, although there was not spectroscopic evidence for haem binding in an RbdA PAS-GGDEF-EAL construct[131].

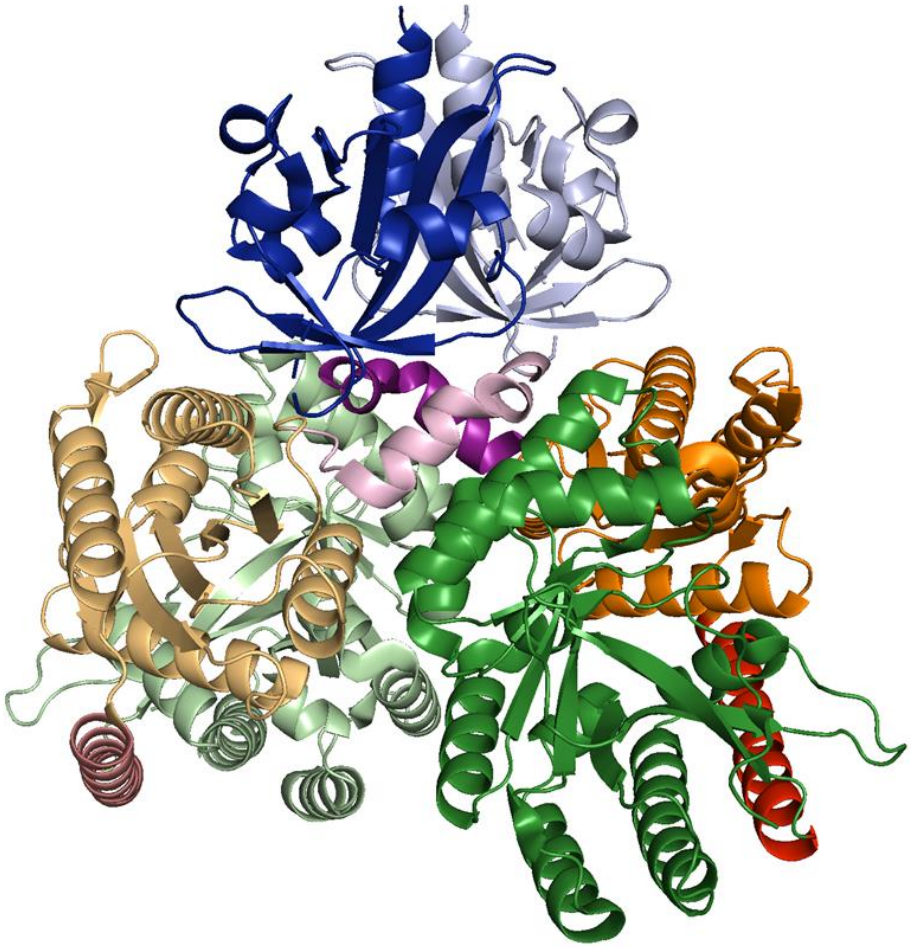

Figure 1.9. **The dimeric arrangement of the cytosolic domains of RbdA**. The cytosolic PAS-GGDEF-EAL domains of RbdA (PDB code 5XGB), depicted as a cartoon, are held together in a dimeric arrangement. The PAS domain is coloured in blue, Signalling (S)-helix in magenta, GGDEF domain in orange, Hinge (H)-helix in red and the EAL domain in green. One monomer is shown in darker colourings with the other monomer of the dimer shown in lighter colourings.

Two helical segments, which together form the signalling helix (S-helix), connect the PAS domain to the GGDEF domain (resides 380 – 536). The GTP bound GGDEF domain of RbdA was elucidated (PDB code 5XGD) and closely resembles the GGDEF domain of PleD bound to GTP[68,131]. In RbdA, the negative charges from the triphosphate group of GTP are neutralised by one Mg^{2+} ion that is coordinated by residues Asp455 of the GGDEF motif and Asp412[131]. In addition the carbonyl oxygen of His425 and the amide group of the side chain of Asn420 form hydrogen bonds with the N2 atom of the guanine[131]. Compared to the apo-structure, the γ-phosphate group of GTP is neutralized by the repositioning of the side chains of Lys525 (which

in the native structure forms a salt-bridge with Asp412) and Arg529. Additionally to accommodate the triphosphate group of GTP, a significant conformational change occurs in the N-terminal region of the α2 helix in the GGDEF domain[131]. However in the present GTP bound dimeric structure of cRbdA, the two active half sites of the GGDEF domains are separated by a distance of approximately 31 Å and so does not represent the active DGC structure[131]. Liu et al.[131] suggest that larger conformational changes, such as the reorientation of whole domains in the RbdA dimer, could result from the propagation of the local conformational changes triggered upon GTP binding. Although these large conformational changes are not observed between the apo- and GTP bound structures, data from small angle X-ray scattering (SAXS) indicates that large conformational changes occur upon GTP binding[131].

In RbdA, the GGDEF domain is connected to an EAL domain (residues 562-795) via an α-helix (residues 537-562) termed the H-helix[131]. This H-helix functions as a hinge which allows the two domains to adopt various orientations with respect to one another[73,131]. In the current structure of cRbdA, the c-di-GMP binding site of the EAL domain is facing the GTP binding site of the GGDEF domain of the same monomer and is suggested to be in a closed or auto-inhibited conformation[131]. The space that would need to be occupied by the EAL partner to form a canonical EAL dimer, is filled by the GGDEF and PAS domains of the partner molecule[131]. Therefore to form the canonical EAL dimer, contacts between the S-helix and the α-6 helix of the EAL partner molecule would need to be broken along with a large rotation of the GGDEF domain and an inward movement of loop-6 of the EAL domain[131]. Although it is suggested that this current cRbdA structure is an auto-inhibited conformation[131], the PDE and DGC activity of RbdA suggests that other alternative dimeric structures must be possible in solution. To identify the variety of conformations that the cytosolic region of RbdA could assume in solution, Liu et al.[131] carried out SAXS experiments. These SAXS experiments demonstrated that cRbdA is extended and flexible but in the presence of GTP, GTP and c-di-GMP or GMP and c-di-GMP the conformation of cRbdA becomes more compact and less flexible. Whereas the conformation of cRbdA when in the presence of c-di-GMP alone showed a mixture of the compact conformation (75%) and a more extended conformation (25%)[131]. However as the molecular shapes determined by SAXS are of relatively low resolution, the

possible dimeric arrangements that could be fit were numerous and thus requires further structural studies[131].

Therefore, RbdA seems to be a promising target that potentially takes part in the NO induced biofilm dispersal pathway. However as many different proteins have been suggested to be implicated in the NO induced biofilm dispersal pathway within *P. aeruginosa*, there is likely to be a redundancy in the system, with more than one mechanism and pathway leading to biofilm dispersal upon NO induction.

1.4.2 – PA2072

The hypothetical *P. aeruginosa* protein PA2072 has a protein domain architecture very similar to RbdA, except, PA2072 is predicted to also contain the periplasmic CHASE4 sensory domain[133]. In our recent publication investigating the phenotypes of gene KO mutants in *P. aeruginosa*, we presented data that evidenced RbdA and PA2072 to have opposing physiological roles[130]. However, we found both RbdA and PA2072 to be involved in the NO-induced biofilm dispersal pathway[130]. We also predict that PA2072 has both DGC and PDE enzymatic activities (as evidenced previously in RbdA[129,131])[130]. Therefore, we were interested how, at the molecular level, PA2072 and RbdA mediate their opposing physiological roles. We hypothesise that the differences between the RbdA and PA2072 putative sensory PAS domains and/ or periplasmic regions account for their physiological differences.

1.4.3 – PA0285

During our lab groups' previous phenotypic investigations with *P. aeruginosa* KO mutants, PA0285 was highlighted as an interesting target due to the increased NO-induced biofilm dispersal in the Δ*pa0285* KO mutant compared to the PAO1 WT. Furthermore, we predict both DGC and PDE enzymatic activities for PA0285[130]. We were interested in how PA0285 could sense NO and how, in the PAO1 WT, PA0285 reduced the extent of NO-induced biofilm dispersal. The Dr I. Tews group (University of Southampton) previously predicted that the first

PAS domain of the PA0285 PAS-PAS-GGDEF-EAL protein, bound a haem-*b* cofactor while the second PAS domain was predicted to bind an FAD cofactor[134]. We hypothesise that one or both of the PA0285 PAS domains can sense NO to modulate the enzymatic outputs of PA0285.

1.5 – Main aims and objectives

NO induces a *P. aeruginosa* biofilm dispersal *via* a decrease in the intracellular c-di-GMP levels. While this could occur with the inhibition of the DGC activity, as NO induced biofilm dispersal is a rapid response it is assumed that this occurs with an increase in c-di-GMP specific PDE activity. The *P. aeruginosa* PAO1 genome contains 16 proteins with tandem GGDEF and EAL domains. In multi-domain proteins the different domains may influence each other's activity. We hypothesize that the c-di-GMP specific PDE activity of the EAL domain is influenced by tandem GGDEF domains. The aim of chapter 3 is to investigate the relationship between the tandem GGDEF and EAL domain of RbdA.

To do this we will:

- Carry out an enzymatic assay, measuring the breakdown of c-di-GMP to pGpG, on the c-di-GMP specific PDE activity of the EAL domain when in the presence and absence of its tandem GGDEF domain when both domains are isolated and when these two domains are kept in tandem.
- Utilise crystallographic structure determination of the EAL, GGDEF and GGDEF-EAL domains of RbdA to highlight structural differences so that a structural mechanism for the GGDEF domains regulation of its tandem EAL domain can be proposed.

Prokaryotic PAS domains are capable of binding cofactors, with cofactor bound PAS domains having the potential to bind or respond to NO. Therefore, we hypothesize that an upstream PAS domain is a putative NO sensor in *P. aeruginosa*. The aim of chapter 4 and 5 is to characterise the PAS domain containing proteins, PA2072 and PA0285, which have been identified as potentially linking NO sensing to a biofilm dispersal phenotype in *P. aeruginosa*.

To do this we will:

- Express and purify constructs to use in crystallographic structure determination in order to investigate conformational changes that may result from NO binding.
- Investigate whether the PAS domains of these proteins bind a cofactor by using ultraviolet-visible light spectroscopy, with a bound cofactor producing a different spectrum to an apo-PAS domain.

Chapter 2 – Materials and Methods

2.1 – Buffers

Lysis buffer: 50 mM Tris, 200 mM NaCl, 5 % (v/v) glycerol, 2 mM β–mercaptoethanol, pH 7.5

Binding buffer: 50 mM Tris, 200 mM NaCl, 5 % (v/v) glycerol, 20 mM imidazole, 2 mM β–mercaptoethanol, pH 7.5

Wash buffer: 50 mM Tris, 300 mM NaCl, 5 % (v/v) glycerol, 80 mM imidazole, 2 mM β–mercaptoethanol, pH 7.5

Elution buffer: 50 mM Tris, 300 mM NaCl, 5 % (v/v) glycerol, 300 mM imidazole, 2 mM β–mercaptoethanol, pH 7.5

Gel filtration buffer 1: 50 mM Tris, 300 mM NaCl, pH 7.5

Gel filtration buffer 2: 50 mM Tris, 300 mM NaCl, 2 mM $MgCl_2$, 2 mM β–mercaptoethanol, pH 7.5

Gel filtration buffer 3: 50 mM Tris, 300 mM NaCl, 2 mM $MgCl_2$, 2 mM β–mercaptoethanol, 50 mM L-glutamic acid, 50 mM L-arginine, pH 7.5

Gel filtration buffer 4: 50 mM Tris, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM $MnCl_2$, 50 mM L-glutamic acid, 50 mM L-arginine, pH 7.5

PDE reaction buffer: 50 mM Bis-Tris Propane, 50 mM NaCl, 5 mM $MgCl_2$, pH 9.35

SDS-PAGE Gel buffer: 3 M Tris, 0.3 % SDS, pH 8.45

SDS-PAGE Anode buffer: 200 mM Tris, pH 8.9

SDS-PAGE Cathode buffer: 100 mM Tris, 100 mM Tricine, 0.1 % SDS, pH 8.2

SDS-PAGE Sample buffer (4x): 242 mM Tris pH 6.8, 7.72 % SDS, 4 mg/mL bromophenol blue, 19.31 % glycerol, 1.187 M β-Mercaptoethanol

2.2 – Bacterial strains and plasmids

All bacterial strains being used in this study are listed in Table 2.1. WT *P. aeruginosa* PAO1 cells were utilized for their genomic DNA for use in construct development. *E. coli* DH5α cells were used for construct DNA amplification and BL21 (DE3) cells were used for protein expression. Cells were grown in Lysogeny broth (LB) medium. In this thesis all constructs were cloned into the pET28a vector so that a hexa-Histidine tag followed by a thrombin cleavage site were included at the N-terminus.

A target gene is cloned into the expression vector, pET28a (Novagen), under the control of a T7 promoter for the expression of a recombinant protein of interest. BL21 (DE3) cells contain the phage T7 RNA polymerase gene under the control of the inducible *lac*UV5 promoter. The addition of isopropyl-β-D-thiogalactopyranoside (IPTG) removes lac repressor proteins from lac operators at the *lac*UV5 promoter and the T7 promoter. Therefore the production of T7 RNA polymerase and thus recombinant protein can be induced within BL21 (DE3) cells with the addition of IPTG. The pET28a vector also contains an N-terminal (and optional C-terminal) hexa-Histidine tag and contains a kanamycin resistance gene. To study the ability of a protein to incorporate a haem cofactor, BL21 (DE3) cells with the haemoprotein expression system plasmid (pHPEX-3[135]; received from University of Kaiserslautern, Germany[136]) was used for the expression of proteins that were potential haem binders. This is because, BL21 (DE3) cells do not already contain the haem receptor gene, *chuA,* so haem uptake into the cells is poor. The pHPEX-3 plasmid encodes for the ChuA haem receptor under the control of the *lac*UV5 promoter and is used in BL21 (DE3) cells alongside the pET28a expression vector. The pHPEX-3 plasmid also contains the *lysS* gene which encodes for a T7 lysozyme. The small amount of T7 RNA polymerase, produced due to the leakage expression under the *lac*UV5 promoter, is inhibited by the T7 lysozyme. Thus preventing target genes under the T7 promoter from being expressed before induction. The pHPEX-3 plasmid contains a tetracycline resistance gene[135].

Table 2.1. **Bacterial strains used in this study**.

Strain	Use	Source
Strains		
E. coli		
DH5α	Production of plasmid of interest	Life Technologies Ltd
BL21 (DE3)	Production of protein of interest	NEB
P. aeruginosa		
PAO1	Wild type control / genomic DNA extraction	University of Washington

2.3 – Construct Development

2.3.1 – Genomic DNA extraction

For all genome extraction procedures overnight cultures of WT *P. aeruginosa* strain PAO1 were used with the Wizard Genomic DNA Purification Kit from Promega. Kit protocol was followed with all centrifugation step carried out at 16,000 x g. However for step 3 incubation was at 37 °C for 30 mins and DNA pellet rehydration was carried out with 200 µL of Rehydration Solution.

2.3.2 – Primers

All primers (non-modified oligonucleotides) used are listed in Table 2.2 and were purchased from Eurofins Genomics. Primers were designed so that restriction enzyme recognition sites were included onto either end of the PCR amplified DNA (NdeI recognition site at 5' and HindIII recognition site at 3') for insertion into the pET28a vector. The NdeI restriction site was used so a hexa-His tag followed by a thrombin cleavage site would be included at the N-terminus of the recombinant protein when expressed from the pET28a vector.

Table 2.2. **The primer sequences used in PCR for producing different constructs**.

Primers	Sequence (5' – 3')*
RbdA 376 FWD	aacttCATATG**CACGATGCGTTGACCG**
RbdA 536 REV	aacttAAGCTTTCA**ATGGAACACCTGGACCCG**
RbdA 549 FWD	aacttCAT**ATGACCTGGGTCCAGCG**
RbdA 797 REV	aacttAAGCTTTCA**GCGACTGAACGGCAGG**
PA2072 45 FWD	aacttCATATG**GACGACATCGCCATCGA**
PA2072 254 REV	aacttAAGCTTTCA**CTCCCGCAGCATGGC**
PA2072 310 FWD	aacttCATATG**GAGGCGGCCTCGGAC**
PA2072 427 REV	aacttAAGCTTTCA**GTGCAGGGACAGATGCTG**
PA0285 88 FWD	aacttCATATG**GAAGCGGTGCGCGAC**
PA0285 213 FWD	aacttCATATG**GACAGCACCCGCGAG**
PA0285 216 REV	aacttAAGCTTTCA**GCGGGTGCTGTCGAATA**
PA0285 333 REV	aacttAAGCTTTCA**GTGGTGGGCGAGGAAG**
PA0285 760 REV	aacttAAGCTT**TCAGTCTTCCGGCAGCG**

* Lower case sequences - assist in the enzymatic identification of the added restriction enzyme recognition site. Upper case sequences – added restriction enzyme recognition site. Upper case underlined sequence - indicates the addition of a stop codon. Bolded upper case sequences - gene specific sequences.

2.3.3 – PCR

Gene sequences of the protein of interest which were to be expressed were amplified from the isolated WT genome of *P. aeruginosa* PAO1 using PCR with a final reaction volume of 50 µL:

1. — µL water
2. 1X Phusion HF/ GC Buffer (NEB) or 1X Q5 Reaction Buffer (NEB)
3. 100 - 500 nM FWD primer
4. 100 - 500 nM REV primer
5. 200 µM of each dNTP (Promega)
6. 20 ng genomic template DNA or 10 ng vector template DNA
7. 0.02 units/µL Phusion High-Fidelity DNA Polymerase (NEB) or 0.02 units/µL Q5 High-Fidelity DNA Polymerase (NEB)

Primer pair specific annealing temperatures and construct specific extension times, as listed in Table 2.3, were used with the Bio-Rad DNA Engine Tetrad 2 Peltier Thermal Cycler to amplify the DNA sequences as follows:

1. Initial denaturation at 98 °C for 3 mins
2. 98 °C for 10 secs
3. Primer pair specific annealing temperature for 15 secs
4. 72 °C for specific extension time (30 secs/1kb for genomic DNA or 15 secs/1kb for vector DNA)
5. Repeat steps 2 through to 4 to give thirty cycles
6. Final extension at 72 °C for 5 mins

Table 2.3. **The specific conditions used in PCR to form the different constructs**.

Construct	FWD Primer	REV Primer	Final conc. (nM) of each primer	Annealing Tm (°C)	Extension time (sec) for genomic DNA
RbdA GGDEF[a]	RbdA 376	RbdA 536	500	64	15
RbdA EAL[a]	RbdA 549	RbdA 797	500	65	23
RbdA GGDEF-EAL[a]	RbdA 376	RbdA 797	500	64	39
PA2072 CHASE[a]	PA2072 45	PA2072 254	500	65	19
PA2072 PAS[a]	PA2072 310	PA2072 427	500	63	11
PA0285 PAS1[b]	PA0285 88	PA0285 216	100	68	12
PA0285 PAS2[b]	PA0285 213	PA0285 333	100	68	12
PA0285 PAS2-GGDEF-EAL[b]	PA0285 213	PA0285 760	250	70	50

[a] Construct produced with Phusion High-Fidelity DNA Polymerase (NEB). [b] Construct produced with Q5 High-Fidelity DNA Polymerase (NEB).

2.3.4 – PCR product purification

The QIAGEN QIAquick PCR Purification Kit was used to purify PCR products. The kit protocol (protocol revised 09/2011) was followed with centrifugation steps carried out for 1 min at 17,900 X g. DNA was eluted with 30 µL of water and a NanoDrop spectrophotometer used to measure DNA concentration.

2.3.5 – Restriction Digestion

The PCR amplified gene sequences and expression vector were digested with restriction enzymes, NdeI (NEB) and HindIII (NEB), to create compatible ends for the insertion of the PCR amplified gene sequence into the expression vector. The total reaction volumes were 20 µL containing either 500 ng insert DNA or 1000 ng vector with 2 µL Cutsmart buffer (NEB), 20 units of NdeI, 20 units of HindIII and water. DNA digestion was carried out at 37 °C for 2 hrs followed by the enzymes being denatured at 80 °C for 25 mins, with 1 unit of Shrimp Alkaline Phosphatase (NEB) being added to the vector digestion only after 1 hr at 37 °C to prevent re-ligation.

2.3.6 – Ligation

The digested DNA insert and vector were ligated to form constructs with 200 cohesive end units of T4 DNA ligase (NEB) at room temperature overnight in a total volume of 10 µL containing 1X T4 ligase buffer and water. The 1:2, 1:1, 2:1, 4:1 and 8:1 insert to vector ligation ratios were carried out.

2.4 – Competent cells

E. coli cells to be made competent were grown in a 2 L baffled flask containing 1 L of LB medium inoculated with 8 mL of an overnight culture. Cells were grown at 37 °C with 180 rpm shaking to an OD_{600} value of 0.4 at which point they were transferred to a 1 L centrifuge tube and chilled on ice for 20 mins. Cells are then centrifuged (Avanti J-20 XPI Beckman Coulter with JLA 8.1000 rotor) at 3,000 x g for 20 mins at 4 °C and the resulting cell pellet was resuspended on ice with 30 mL of cold 100 mM $CaCl_2$. Resuspended cells are transferred to a 50 mL Falcon tube and chilled on ice for 30 mins before centrifuging (Beckman Coulter Allegra X-15R) at 3,000 x g for 10 mins at 4 °C. The resulting cell pellet was resuspended on ice with 10 mL of cold 100 mM $CaCl_2$ containing 15 % (v/v) glycerol. Aliquots of 100 µL were pipetted

in cold 500 µL eppendorf tubes before freezing on dry ice mixed with 100 % ethanol. Competent cells are stored at -80 °C.

2.5 – Transformation

Constructs were transformed into suitable *E. coli* strains (Section 2.2) using 20 µL of $CaCl_2$ competent cells (Section 2.4) and 2 µL of construct DNA. After incubating on ice for 15 mins, cells were heat shocked at 42 °C for 45 secs and returned to ice for 3 mins, before the addition of 200 µL LB and incubation at 37 °C with shaking for 1.5 hrs. Cells (100 µL) were then plated on antibiotic (50 µg/mL kanamycin, 100 µg/mL ampicillin and/ or 10 µg/mL tetracycline) selective LB agar and incubated at 37 °C overnight. Co-transformations were carried out the same way except, 1 µL of construct DNA with 1 µL of pHPEX-3 plasmid were used with 20 µL of suitable $CaCl_2$ competent cells.

2.6 – Extraction of construct DNA

Using overnight cultures of transformed *E. coli* DH5α cells, 5 mL of bacterial culture was pelleted by centrifugation at 17,900 x g for 3 mins. The QIAGEN QIAprep Spin Miniprep Kit was used to extract construct DNA. Kit protocol (protocol revised 02/2015) was followed with all centrifugation steps being carried out for the maximal amount of time (usually 1 min at 17,900 X g), however bacterial pellets were re-suspended in Buffer P2 and the wash step with Buffer PB was omitted. DNA was eluted with 30 µL of water. After producing a new construct, the insert DNA was sequenced with Sanger sequencing (Eurofins Genomics) using appropriate forward and reverse primers, ensuring that vectors contained the DNA of interest.

2.7 – Agarose gel electrophoresis

1 % agarose gels (Melford) were used to separate DNA fragments based on size to check PCR products against a 1 kb DNA ladder (Promega). Gels contained 1 x GelRed (Biotium) to stain the DNA. 5 µL DNA ladder was mixed with 1 µL 6 X Load Dye (Promega) and 5 µL loaded into

a well. 1 µL DNA sample was mixed with 1 µL 6 X Load Dye and 4 µL water, the 6 µL mix was then loaded into consecutive wells. Gel electrophoresis was carried out in 1 X TAE buffer at 120 V for 40 mins. Gels were imaged using the Syngene PXI and Genesys software under UV light.

2.8 – Protein expression and cell harvesting

2.8.1 – Protein expression using BL21 (DE3) cells

Protein expression was carried out in 2 L baffled flasks containing 1 L LB medium with selective antibiotic (50 µg/mL kanamycin) which were inoculated with 8 mL of an overnight culture of transformed *E. coli* BL21 (DE3) cells. Cells were grown at 37 °C with 180 rpm shaking to an OD_{600} value of 0.2 – 0.3. At this point the temperature was either turned down to 18 °C and at an OD_{600} value of 0.6 – 0.8, the bacteria were induced with a final concentration of 1 mM IPTG (Fisher Scientific) and grown at 18 °C with 180 rpm shaking for 18 hrs before harvesting. Or, the cells were further grown at 37 °C, induced with a final concentration of 1 mM IPTG at an OD_{600} value of 0.6 – 0.8, and then grown at 37 °C with 180 rpm shaking for 2 hrs before harvesting.

Initially, for all constructs used in this thesis, protein expression was trialled in BL21 (DE3) cell cultures grown at both 18 °C and 37 °C, except for the PA0285 $PAS1_{88-216}$ and PA0285 PAS2-GGDEF-$EAL_{213-760}$ constructs in which protein expression was only trialled at 18 °C. This was to determine whether the protein preferentially expressed at either temperature. For all of the constructs tested, the protein of interest was expressed regardless of whether the cells were grown at 18 °C or 37 °C. After this initial test, all further protein expressions were carried out by turning the temperature down and growing the cells at 18 °C with 180 rpm shaking for 18 hrs before harvesting. It is thought that expressing recombinant proteins at lower temperatures improves the likelihood that the proteins are folded correctly and reduces the occurrence of inclusion bodies. This is because at lower temperatures the rates of transcription and translation are reduced, preventing the chaperone proteins of the cell from becoming overwhelmed[137]. Lower temperatures also reduce the strength of the hydrophobic interactions that contribute to protein misfolding and aggregation[138].

2.8.2 – Protein expression using BL21 (DE3) cells with pHPEX-3 plasmid

Protein expression was carried out in 2 L baffled flasks containing 1 L LB medium with selective antibiotics (50 µg/mL kanamycin and 10 µg/mL tetracycline) which were inoculated with 8 mL of an overnight culture of co-transformed *E.coli* BL21 (DE3) cells (Section 2.5). Cell cultures were grown at 37 °C with 180 rpm shaking to an OD_{600} value of 0.6 – 0.8 and then induced with a final concentration of 1 mM IPTG with or without a final concentration of 10 µM hemin (Fisher Scientific; hemin dissolved in 1 M NaOH to make a 50 mM stock) added to each 1 L of culture just before induction. The cell cultures were then incubated in the dark at 37 °C with 180 rpm shaking for a further 4 hrs before harvesting.

2.8.3 – Cell harvesting

Cultures of BL21 (DE3) cells, with or without the pHPEX-3 plasmid, were harvested with centrifugation (Avanti J-20 XPI Beckman Coulter with JLA 8.1000 rotor) at 6,238 x g at 4 °C for 20 mins. The resulting cell pellets were re-suspended in a volume of lysis buffer (50 mM Tris, 200 mM NaCl, 5 % (v/v) glycerol, pH 7.5 or pH 8.0) that was approximately three times that of the cell pellets weight. Lysis buffer re-suspended cell pellets were stored at -20 °C.

2.9 – Protein purification

2.9.1 – Immobilised metal affinity chromatography (IMAC)

Lysis buffer re-suspended cell pellets were thawed and lysed by sonication with 10 secs on, 45 secs off and a total sonication time of 5 mins. Lysed cells were centrifuged (Optima XPN-80 Beckman Coulter) at 92,600 x g for 40 mins at 4 °C and the supernatant was used for recombinant protein of interest purification. Recombinant proteins were expressed from the pET28a vectors with an N-terminal hexa-histidine tag allowing for an initial protein purification using immobilised metal affinity chromatography (IMAC). The His-tag of the

recombinant protein can reversibly bind to a Ni^{2+} ion which is chelated by nitrilotriacetic acid (Ni-NTA) which is coupled to a resin bead.

In this thesis, the supernatant was combined with approximately 2 mL of Ni-NTA Superflow resin (Qiagen) pre-equilibrated in approximately 15 mL binding buffer (50 mM Tris, 200 mM NaCl, 5 % (v/v) glycerol, 20 mM imidazole, pH 7.5) and left at 4 °C with constant mixing for 1 hr for the binding of the His-tag to the Ni-NTA. This was then poured into an empty (binding buffer equilibrated) gravity flow column and the flow through collected. Then the column was washed with 20 mL wash buffer (50 mM Tris, 300 mM NaCl, 5 % (v/v) glycerol, 80 mM imidazole, pH 7.5) followed by 10 mL binding buffer. The recombinant protein of interest was then eluted from the column with 10 mL elution buffer (50 mM Tris, 300 mM NaCl, 5 % (v/v) glycerol, 300 mM imidazole, pH 7.5). Samples of this purification process were analysed with SDS-PAGE (Section 2.10).

2.9.2 – Size exclusion chromatography

For all protein samples, the elute was concentrated with an appropriate molecular weight cut off Vivaspin 20 (Sartorius) and centrifugation (Beckman Coulter Allegra X-15R) at 3,540 x g to approximately 1.2 mL. This concentrated elute was then loaded onto a 120 mL column volume HiLoad 16/600 Superdex 75 prep grade size exclusion column (GE Healthcare) pre-equilibrated with at least 1 column volume of gel filtration buffer. The size exclusion column was run, with an AKTA purifier, for 1 column volume using gel filtration buffer at a flow rate of either 1 mL/min or 0.2 mL/min collecting 2 mL fractions. Size exclusion chromatography separates proteins based on their molecular weight, where large molecular weight proteins are eluted from the column quicker than smaller molecular weight proteins. A trace measuring UV absorbance at 280 nm indicated which fractions contained protein and these fractions were analysed for the protein of interest with SDS-PAGE (Section 2.10). Fractions containing the protein of interest were pooled and concentrated for further use.

The HiLoad 16/600 Superdex 75 prep grade column was calibrated using a low molecular weight kit (GE Healthcare). The column was run at 1 mL/min with gel filtration buffer (50 mM Tris, 300 mM NaCl, 50 mM L-glutamic acid, 50 mM L-arginine, pH 7.5) measuring

absorbance at 280 nm. The protein standards (3 mg) were each dissolved in 1 mL gel filtration buffer with 5 % (v/v) glycerol and individually run over the column. PrimeView Evaluation software was used to find the elution volume at the middle of the main absorbance peak. The protein standards known molecular weights were plotted against their elution volumes and a line of best fit was produced in GraphPad Prism (version 7.0).

2.10 – Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis (SDS-PAGE)

A 12 % SDS-PAGE gel 0.75 mm thick was used to ensure protein expression and purification by separating proteins based on size and estimating their sizes by comparing them to a protein ladder consisting of markers of known size (4 µL of PageRuler Prestained Protein Ladder; Thermo Scientific).

12 % SDS-PAGE gels were prepared as follows:

	Stacking gel	*Running gel*
40 % Acrylamide	0.66 mL	3.33 mL
Water	2.9 mL	3.33 mL
SDS-PAGE Gel buffer	1.24 mL	3.33 mL
TEMED	10 µL	10 µL
Ammonium Persulphate (100 mg/mL)	25 µL	50 µL

Protein samples were run on the SDS-PAGE gel by first mixing 6 µL of sample with 3 µL of SDS-PAGE sample buffer (4x), incubating at 100 °C for 10 mins, centrifuging for 5 mins at 17,900 x g and loading 6 µL of each sample into separate wells. SDS-PAGE gels were run at 160 V constantly for 45 mins with SDS-PAGE cathode and SDS-PAGE anode buffers using a Bio-Rad PowerPac 3000. The SDS-PAGE gels were submerged in fixing solution (40 % ethanol, 10 % Acetic acid, 50 % water) for 10 mins, stained with 1 X coomassie blue for 1 hr, and water used to de-stain the gels overnight. SDS-PAGE gels were imaged with the Syngene PXI and Genesys software using white light.

2.11 – Enzymatic assay of c-di-GMP specific PDE activity

After purification and concentration, 1.5 µM of a EAL domain with or without 1.5 µM of a GGDEF domain or 1.5 µM of GGDEF-EAL tandem protein domain were added to the PDE reaction buffer (50 mM Bis-Tris Propane, 50 mM NaCl, 5 mM $MgCl_2$, pH 9.35) and the PDE reaction initiated with the addition of 10 µL 10 mM c-di-GMP (final concentration 100 µM) in a 1 mL total reaction volume. Reactions were incubated at room temperature and monitored by removing 100 µL of sample, and the reaction stopped by addition of 10 µL of 100 mM $CaCl_2$ before being placed on ice until use. Stopped reaction samples were diluted with 890 µL of 5 mM ammonium bicarbonate. 500 µL of this diluted sample was injected onto a 1 mL Resource-Q column, using an AKTA purifier, after pre-equilibrating the Q column with 5 mM ammonium bicarbonate. After a 2 column volume wash step with 5 mM ammonium bicarbonate, the reaction products and substrates were separated by a 25 column volume linear ammonium bicarbonate gradient (5 mM to 1 M), monitoring elution at 253 nm absorbance (mAU). This protocol is adapted from a method described by Sundriyal et al.[81].

Standards at 100 µM of c-di-GMP (Biolog), pGpG (Biolog), GTP (Sigma) and GMP (Sigma), treated exactly the same as the reaction samples, were run separately to determine the formation of reaction products. The concentration of reaction product, pGpG, was determined from a calibration curve (shown in Appendix Figure 7.1) in which known concentrations (0 µM, 5 µM, 10 µM, 20 µM, 40 µM, 60 µM, 80 µM and 100 µM) of pGpG were run over the 1 mL Resource Q column in triplicate. The known pGpG concentrations were plotted against the area under their curves which was found by integration in the UNICORN Evaluation software and a line of best fit produced in GraphPad Prism (version 7.0). Using the equation of the line of best fit the areas under the pGpG peaks obtained from the PDE reaction samples were converted to known pGpG concentrations.

2.12 – Spectroscopic analysis investigating cofactor binding

Ultraviolet-visible light (UV-Vis) spectroscopy was used to determine the presence or absence of cofactor binding within a protein after protein purification (Section 2.9). The ultraviolet (170 nm to 380 nm) and visible (380 nm to 780 nm) light wavelengths are used in UV-Vis spectroscopy. Electrons have discrete energy states, with an electron in the ground state (E0) switching to an excited state (E1) after light with a specific energy (equal to E1 – E0) strikes the electron, this is the absorption of light. The amount of energy associated with the light is specific to each wavelength. Different molecules within a mixed sample can be identified using a UV-Vis absorbance spectrum if they absorb distinct wavelengths of light.

A single wavelength or a range of wavelengths can be used for absorption measurements. As the number of molecules in a solution increases (the concentration increases) the amount of absorption also increases. Thus by measuring the amount of absorbance at a particular wavelength, the concentration of a specific molecule in solution can be determined by using the Beer-Lambert Law ($A = \varepsilon cl$, where A is the absorbance, ε is the molar extinction coefficient (L mol^{-1} cm^{-1}), c is the sample concentration (mol/L) and l is the pathlength (in cm) through the sample).

Initially, using a Shimadzu UV-2600 spectrophotometer, baselines were set measuring absorbance of light between 250 and 700 nm using 1 mL cuvettes (Eppendorf UVette) with a 10 mm optical pathlength, containing 100 µL of gel filtration buffer 1 (50 mM Tris and 300 mM NaCl at pH 7.5 or pH 8.0). Baselines were then measured to produce a UV-Vis spectra between 250 and 700 nm to use as a comparison between the measured baselines and protein samples. Using 100 µL of purified protein with a fresh 1 mL cuvette, UV-Vis spectra were measured between 250 and 700 nm. UV-Vis spectra were analysed with the UVProbe software.

2.13 – X-ray crystallography

2.13.1 – Crystallisation trials

After protein purification (Section 2.9), fractions from size exclusion chromatography containing the protein of interest, as determined from SDS-PAGE gels (Section 2.10), were concentrated with an appropriate molecular weight cut off Vivaspin 2 (Sartorius) and centrifugation (Beckman Coulter Allegra X-15R) at 3,540 x g. Normally, protein was concentrated to approximately 10 mg/mL for initial protein crystallisation trails.

Crystallisation used the vapour diffusion method. This is where a drop of concentrated protein and a drop of precipitant solution are mixed and then incubated within a sealed environment. Initially the protein concentration is too low to favour aggregation into an ordered structure with the droplet being under-saturated. Over time, water diffuses out from the protein/ precipitant droplet into a reservoir of precipitant solution as the solute concentration is higher in the precipitant solution than in the protein/ precipitant droplet. This reduces the drop volume causing the effective protein and precipitant concentrations in the droplet to increase. The droplet becomes super-saturated leading to the nucleation (Labile) phase in which the protein molecules aggregate into crystal nuclei. This then reduces the effective concentration of protein free in solution resulting in the Metastable phase in which crystal growth occurs but nucleation stops (see Figure 2.1).

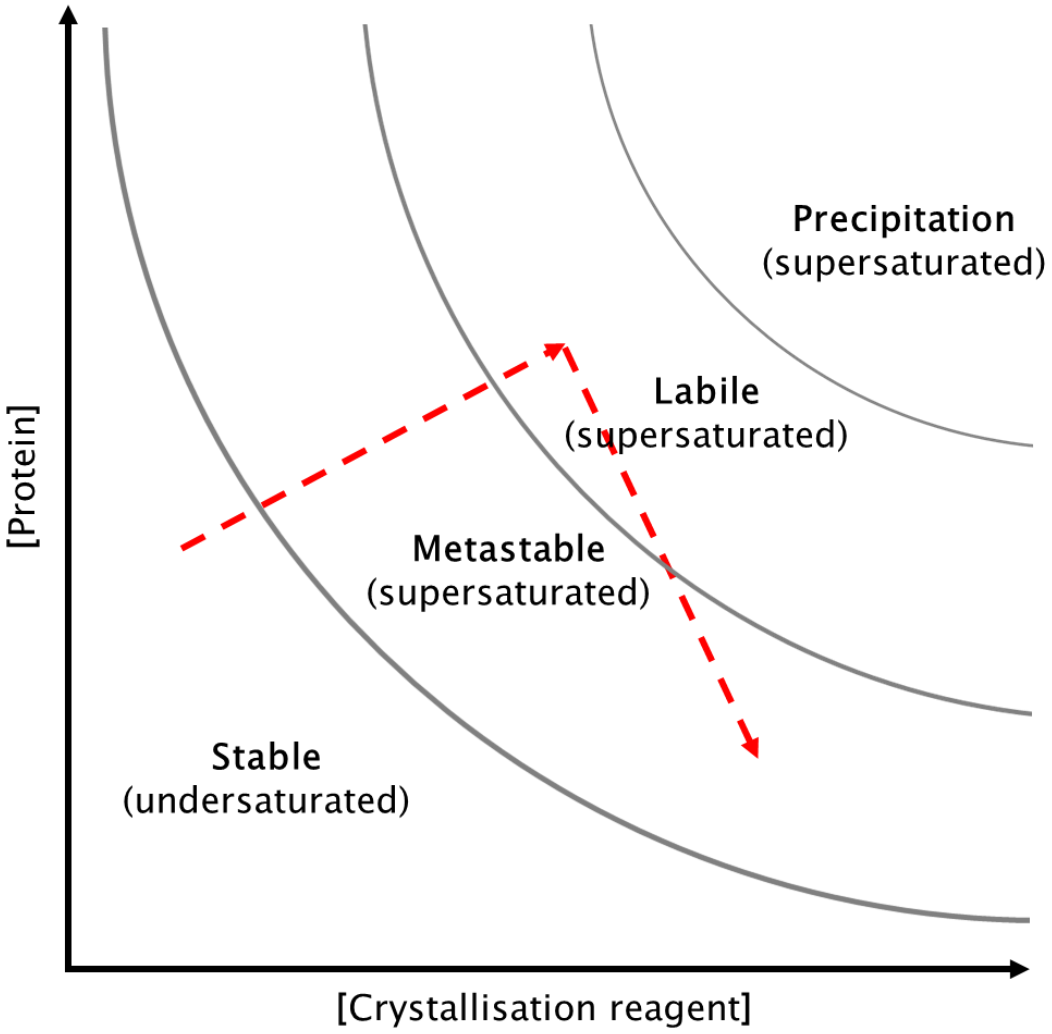

Figure 2.1. **The phase diagram of protein crystallisation.** The path for successful protein crystallisation is shown with red dashed arrows. To begin with the protein solution is undersaturated and stable. As vapour diffusion occurs the concentration of the protein and crystallisation reagent in the crystallisation drop increases. This causes the protein to become supersaturated and at the right protein and crystallisation reagent concentrations the Labile phase is reached in which nucleation of the protein can occur. As the protein comes out of solution to form nuclei, this then reduces the effective protein concentration in the solution resulting in the Metastable phase. In the Metastable phase, protein crystals can grow but cannot nucleate. Therefore if the concentration of the protein and crystallisation reagent does not become supersaturated enough to reach the Labile phase, crystal growth with not occur (unless protein seeds are used which can act as a nuclei for crystal growth in the Metastable phase without the need for the Labile phase). If the solution becomes too supersaturated then the protein will precipitate out of solution.

Initial crystallisation screens used were Morpheus, JCSG$^+$ and PACT *premier* obtained from Molecular dimensions. The Morpheus screen is based on PDB entries, testing a range of pH, PEGs and salt additives and incorporates low molecular weight ligands. JCSG$^+$ is a sparse matrix while PACT *premier* is a systematic screen which, using PEG as the precipitant, tests the effect of pH, cations and anions. Customised screens, prepared with the Alchemist DT Liquid Handling System (Rigaku), were designed based on conditions that gave initial protein crystal hits in order to optimise crystal size and quality.

Protein and screen were set up in 96-condition sitting drop Intelli-plates (Art Robbins Instruments), with each of the 96 different conditions having three wells in which the protein and precipitant solution are combined in different ratios, 2:1, 1:1, 1:2 (top drop to bottom drop) protein to precipitant solution respectively, using a Gryphon nanodrop dispenser (Art Robbins Instruments). Plates were sealed with ClearVue Sheets (Molecular Dimensions) and then incubated at a constant 21 °C. Plates were viewed under a brightfield microscope (Leica) to look for crystal hits. Crystals were then confirmed as protein by viewing under an Ultra-violate fluorescence microscope (JANSi UVEX), with most protein crystals fluorescing bright white due to Tryptophan and Tyrosine residues and most salt crystals being non-fluorescent. If necessary cryo-buffer (75 % (v/v) precipitant solution with 25 % (v/v) glycerol) was used to replace the precipitant solution the crystals had formed in. Protein crystals were then fished with a SPINE standard pin (Molecular dimensions) with an appropriately sized LithoLoop and rapidly cooled to cryogenic temperatures in liquid nitrogen before being stored at cryogenic temperatures and taken to a synchrotron for diffraction experiments.

2.13.2 – Seeding experiments

Seeding experiments can be used to control and improve the number or size of crystals, as well as facilitating a range of crystal forms (polymorphs) that may be achieved. Seeds (sub-microscopic crystals) are produced by crushing protein crystals or crystal fragments. The seeds can then be used in new crystallisation drops to act as a nucleus for crystal growth, avoiding the need for spontaneous nucleation. Seeds therefore allow protein crystals to grow in the Metastable Zone without the need to enter the Labile (Nucleation) Zone (see Figure 2.1).

MicroSeed Bead kits (Molecular Dimensions) were used to produce a seed stock that could then be used in seeding experiments. Seed stocks were made following the protocol supplied with the MicroSeed Bead Kit.

For the PA2072 $PAS_{310-427}$ seeding experiment, PA2072 $PAS_{310-427}$ protein crystals, along with the reservoir solutions, were taken from the PACT *premier* screen wells A1 (100 mM Succinic Acid, sodium phosphate monobasic monohydrate, Glycine (SPG) buffer, pH 4.0 and 25 % w/v PEG 1500) and B1 (100 mM sodium malonate dibasic monohydrate, Imidazole, Boric acid (MIB) buffer, pH 4.0 and 25 % w/v PEG 1500) and placed into the same microcentrifuge tube with MicroSeed Bead. The seed stock was used directly to seed a 96-condition sitting drop Intelli-plate, dispensing 0.2 µL of seed stock into each well (for a seed to protein to precipitant ratio of 2:2:1 (top well), 1:1:1 (middle well) and 2:1:2 (bottom well)). The Gryphon nanodrop dispenser was then used to dispense PA2072 $PAS_{310-427}$ protein at 12 mg/mL and Morpheus precipitant solutions into the seeded Intelli-plate (Section 2.13.1), which was then incubated at 21°C.

For the PA2072 $CHASE4_{45-254}$ seeding experiment, the MicroSeed Bead microcentrifuge tube contained PA2072 $CHASE4_{45-254}$ protein crystals, with reservoir solution, taken from the Morpheus screen well B1 (30 mM Sodium fluoride, 30 mM Sodium bromide, 30 mM Sodium iodide, 100 mM Buffer System 1 (imidazole and MES monohydrate acid), pH 6.5 and 50 % v/v Precipitant Mix 1 (40 % v/v PEG 500 MME and 20 % w/v PEG 20,000)). The seed stock was diluted 1 in 10 (10 µL seed stock into 90 µL water), and then 0.5 µL of diluted seeds were dispensed into the middle wells of a 96-condition sitting drop Intelli-plate (for a seed to protein to precipitant ratio of 5:2:2). The Gryphon nanodrop dispenser was then used to combine PA2072 $CHASE4_{45-254}$ protein at 22 mg/mL and Morpheus precipitant solutions into the seeded plate, before incubating at 21°C.

2.13.3 – Crystallisation of the RbdA $EAL_{549-797}$ domain

Following size exclusion chromatography (Section 2.9.2), 2 mL fractions (usually 28, 29, 30, 31 and 32) containing RbdA $EAL_{549-797}$ were concentrated in a 10,000 MWCO Vivaspin 2 (Sartorius) to approximately 10 mg/mL as measured with a NanoDrop spectrophotometer.

Using the Gryphon micro-disperser (Art Robbins Instruments) as described in Section 2.13.1, crystallisation trials were set up in sitting drop 96-well trays. Crystallisation using the screens Morpheus and JCSG$^+$ incubated at 21 °C gave protein crystal hits in many different conditions after 24 hrs. An optimisation screen (Appendix Table 7.1) based on Morpheus conditions that gave good initial protein crystals was produced. This optimisation screen reproducibly gave protein crystals in nearly all conditions after 24 hrs at 21 °C. Using 0.5 µL of protein and 0.5 µL of precipitant condition in a sitting drop 96-well tray, a single RbdA $EAL_{549-797}$ crystal grown for 72 hrs in condition H11 of the optimisation screen (100 mM Tris (base) and BICINE pH 9, 34 % (w/v) of an ethylene glycol, PEG 8000 mix) was cryo-protected (75 % v/v optimised H11 condition with 25 % v/v glycerol) before cryo-cooling in liquid nitrogen.

2.13.4 – Diffraction data collection for RbdA $EAL_{549-797}$

The resolution limit of all direct imaging techniques is restricted to half the wavelength of the light used[139]. The distance between bonded atoms is approximately 1.5 Å, while the wavelength of visible light is between 4,000 Å to 7,000 Å, meaning that individual atoms are too small and too close together to be independently resolved with the wavelength of visible light. Conversely, X-rays have a wavelength between 0.1 Å to 100 Å, meaning that this form of light has a wavelength short enough to resolve individual atoms[140].

X-rays can be produced by colliding accelerated electrons, emitted from a heated cathode, with a metal (such as copper) anode within X-ray tubes or rotating anode tubes[140]. However, the most intense X-rays are produced at synchrotron radiation sources using storage rings[140,141]. In this thesis, only synchrotron radiation sources (European Synchrotron Radiation Facility (ESRF) and Diamond Light Source (DLS)) were used. At synchrotron radiation sources, electrons, generated by an electron gun, are directed into the linear accelerator and are then further accelerated using electric fields to almost the speed of light within the booster ring. The electrons then enter the storage ring where magnets and insertion devices cause the electrons to bend and undulate. Changing the direction of the electrons causes them to emit electromagnetic radiation, including X-ray radiation. The X-ray beam can then be focused and

filtered by mirrors and monochromators to produce a narrow beam of single wavelength X-rays.

Nevertheless, the highest resolution achieved by direct X-ray microscopy is currently around 7 nm[142] and thus, cannot be used to study protein molecules at atomic detail. Therefore X-ray diffraction patterns, which are produced when X-rays interact with and become diffracted by the electrons of the atoms that make up a molecule, are utilised. The final 'image' is that of a three-dimensional electron density map. To determine the three-dimensional electron density map of a molecule, the diffraction angle, the amplitude and the phase of each of the diffracted X-ray beams (a waves structure factor) must be determined. The mathematical method which converts the diffraction data into a three-dimensional electron density map (and vice versa) is called the Fourier transform.

However, a single molecule will only scatter X-rays weakly, resulting in diffracted X-rays that cannot be detected[140]. A protein crystal, formed of many ordered individual protein molecules aligned in the same orientation, is therefore required in order to produce a diffraction grating that will produce a signal capable of being detected. The crystal can therefore be thought of being formed of many identical repeating 'boxes' which, through translational symmetry, forms the three-dimensional crystal lattice. A single repeating 'box' is called the unit cell[143]. The unit cell can be described, in terms of its dimensions, by the length of three edges **a** (corresponding to the x-axis), **b** (corresponding to the y-axis) and **c** (corresponding to the z-axis); and three angles α, β and γ[140]. Within the unit cell, the smallest group of repeating atoms that can be juxtaposed, by symmetry operations, onto other identical groups is called the asymmetric unit.

The diffraction of X-rays from a protein crystal obeys Bragg's law (Figure 2.2; equation 1), whereby treating the diffracted beams as if they were reflected from sets of equivalent parallel planes within the crystal[144], allows the angles at which the diffracted beams emerge from the crystal to be computed[140]. The various sets of parallel planes that run through the unit cell are each given a three integer Miller index (for the *h*, *k* and *l* axes) which corresponds to the number of sections the given plane set has divided each edge of the unit cell into[140]. A single reflection (or diffraction spot) is therefore produced by each of the sets of parallel planes acting as a source of diffraction.

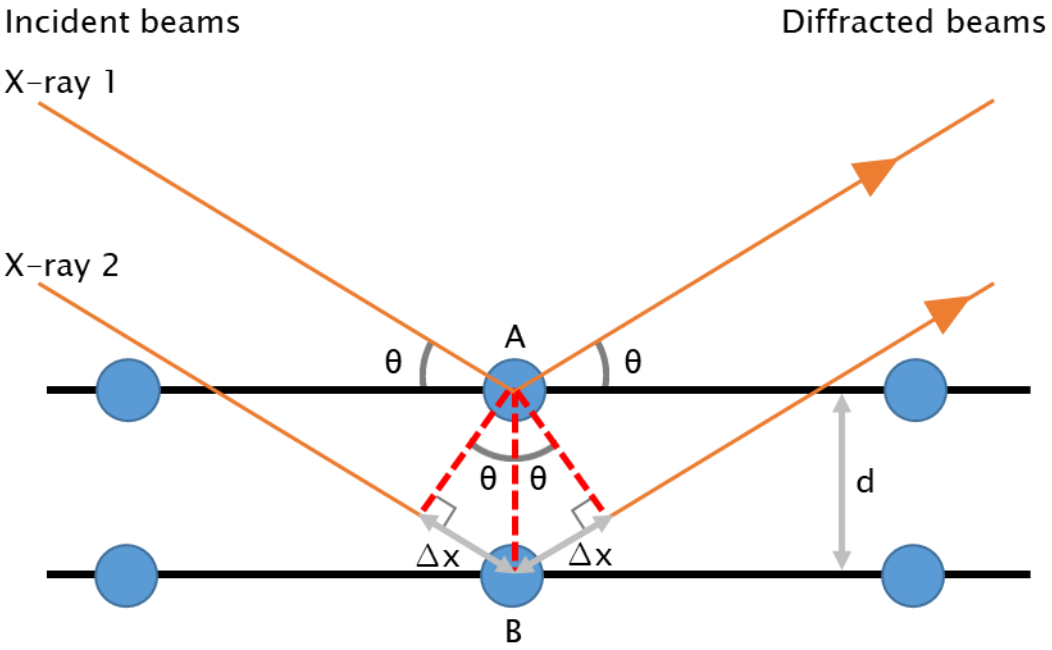

Figure 2.2. **A diagram of Bragg's law.** Prior to the diffraction of X-ray 1 and X-ray 2 from a set of planes within the crystal, both X-ray 1 and X-ray 2 are in phase. X-ray 2 must travel a distance of 2Δx further than X-ray 1, with Δx being equal to $d\sin\vartheta$ (where d is the distance between planes and ϑ is the angle of incidence which is also the angle of reflection). The extra distance travelled by X-ray 2 must be equal to an integer (n) of the X-ray wavelength (λ) for diffracted X-ray 2 to still be in phase with diffracted X-ray 1, producing constructive interference that can be recorded as a diffraction spot. When the extra distance travelled by X-ray 2, or if ϑ is changed so that $2d\sin\vartheta$, equals a fraction of the X-ray wavelength, diffracted X-ray 2 will be out of phase with diffracted X-ray 1 which will result in destructive interference.

$$n\lambda = 2d\sin\vartheta \qquad \text{equation 1}$$

The direction in which a particular X-ray beam was diffracted by the crystal (or, in other words, the beams angle of diffraction), is determined from the position of an individual diffraction spot (or X-ray reflection) which that particular beam made in the diffraction pattern[140]. The direction of the diffracted X-ray beam is dependent upon the distances between the diffracting planes and the orientation of the crystal. Three-dimensional coordinates *h*, *k* and *l* are given to each of the diffraction spots in the diffraction pattern, specifying the direction of each of the X-ray beams[140].

Each of the diffraction spots (or reflections) recorded on a diffraction pattern are the result of a complex wave, produced through the constructive and/ or partly destructive interference of diffracted X-rays from a set of planes within the crystal. Constructive interference will result

in the complex wave having a greater amplitude due to the diffracted X-rays being in phase with one another. Destructive interference results in the diffracted X-rays, due to being in different phases, cancelling one another out and thus the amplitude of the complex wave will be reduced or completely abolished[140,143]. The amplitude of the wave is related to the intensity of the diffraction spot which is measured by the detector, where the square root of the intensity is proportional to the amplitude[145,146].

However, the phases of the X-rays cannot be measured directly by the detector. The phases of the X-ray waves can be determined (estimated) experimentally using multiple isomorphous replacement (MIR), multiple wavelength anomalous dispersion (MAD) or single wavelength anomalous dispersion (SAD) or by using molecular replacement (MR)[140,141,143,146].

The MR method, which was the method used in this thesis for determining the phase information, requires the structure of a homologous protein to already be known. Structural homology is usually implied by a sequence homology, however this is not always the case and many proteins will form multiple distinct structural conformations. Therefore some trial and error or prior knowledge about the proteins structure may be necessary with this method. MR uses Patterson maps of inter-atomic vectors which are generated by squaring the structure factor amplitudes and setting the phase information to zero. A Patterson map is generated for the unknown protein structure, using the diffraction data, as well as for the known homologous protein structure. The two Patterson maps are then compared, with closely matching Patterson maps resulting from the known model having a similar position and orientation to the unknown structure. The homologous model of known structure is therefore rotated and translated within the unit cell, and Patterson maps generated for each alteration, until a Patterson map for the known model, when in a specific orientation and position, closely matches the Patterson map of the unknown structure. The phases for the known model, in this specific orientation and position, can then be calculated and subsequently used to determine the electron density map for the unknown structure.

Briefly, MIR involves measuring the separate diffraction patterns of isomorphous crystals (crystals with the same space group) when heavy atoms are present (usually introduced by soaking the crystal in a heavy metal ion solution) or absent within the protein crystal. By comparing the protein crystals native diffraction pattern with the diffraction pattern of protein crystals which have heavy atoms included, the diffraction data, produced from only

the heavy atoms, can be determined. The heavy atoms have a stronger effect on the diffraction spots (reflections), resulting from having a higher electron density, compared to the other atoms which make up the protein. This then allows the position(s) of the heavy atom(s) within the unit cell to be determined and estimates of the phase information to be deduced.

The anomalous dispersion methods (MAD and SAD) also make use of incorporating heavy atoms into protein crystals. X-rays of a specific wavelength can be absorbed by certain heavy atoms which then reemit the waves with an altered phase. Friedel's law, which states $I_{hkl} = I_{-h-k-l}$ (where *I* is intensity), breaks down when this absorption occurs. At some synchrotron beam-lines, the wavelength of the X-ray can be tuned. Choosing an X-ray wavelength that will be absorbed by the heavy atoms within the protein crystal, allows the position(s) of the heavy atom(s) within the unit cell to be determined. With this information phase estimates can be determined. During an anomalous dispersion experiment, either a single anomalous wavelength is used (SAD) or multiple anomalous wavelengths are used (MAD), with the use of multiple anomalous wavelengths providing a better estimate for the phases.

Once the angles of diffraction, amplitudes and phases for each of the diffracted X-rays are determined, the Fourier transform (equation 2) can be used to calculate the electron density map of the contents of the unit cell[141,146].

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

Where $\rho(x,y,z)$ represents the electron density at a position (x,y,z) in the unit cell of volume *V*. The Miller indices (*hkl*) describes the set of parallel lattice planes which form each reflection. The structure factor amplitude is represented as F_{hkl} with the X-ray phase represented as ϕ.

Before a dataset was collected from a crystal, an initial characterisation was carried out in order to estimate the unit cell and space group of the crystal so that a collection strategy could be determined that would result in a complete dataset. To obtain a complete dataset,

diffraction from all of the lattice planes of the crystal must be collected. Knowing the unit cell dimensions and the space group, and thus the symmetry, of the crystal provides information about the lattice planes that are present. The minimum amount the crystal is required to be rotated by, to ensure that diffraction occurs from each of these sets of lattice planes, is thus dependent upon the unit cell and space group of the crystal. During characterisation, four diffraction images with an oscillation range of 1.0 ° over a 180 ° rotation (one image taken every 45 °) were captured. Data collection was carried out using the oscillation starting angle that was determined by the synchrotron software in the characterisation step. Datasets were collected with fine slicing every 0.1 ° or 0.2 ° collecting at least 180 ° rotation of data, producing either 1800 or 900 diffraction images respectively. This was to ensure that diffraction data from all of the lattice planes was collected. The transmission and exposure time were beamline dependent and were set to values that would provide a good signal whilst minimising the amount of radiation damage that would occur early on during the data collection.

A single dataset was used for the structure determination of RbdA $EAL_{549\text{-}797}$ which was collected on beamline ID23-1 (ESRF), with a wavelength of 0.9789 Å and a Pilatus 6M (Dectris) detector, on 15/09/2017 at 100 K, from a single crystal. The transmission was set to 100 % and exposure time was set to 0.037 secs. A fine slicing of 0.2 ° was used to collect 180 ° of data.

2.13.5 – Processing the RbdA $EAL_{549\text{-}797}$ diffraction data

The following data processing was carried out by Charlotte Cordery.

The X-ray diffraction dataset of RbdA $EAL_{549\text{-}797}$ was indexed and integrated using DIALS[147] in the command line. Indexing determines the unit cell dimensions and suggests possible space groups. During integration, the positions of the diffraction spots (reflections) on each image are determined and given *h*, *k*, *l* coordinates as well as measuring the intensities of each reflection. Here, the data was cut to 2.3 Å using the first 750 images out of the total 900 diffraction images collected, to avoid the possible radiation damage or the rotation of the crystal out of the X-ray beam which could be observed in the later images. In the command

line, DIALS[147] was then used for data reduction and scaling. During this step differences between the intensities of identical reflections (caused by differences in the distance the X-ray has travelled through the crystal as it is rotated, radiation damage, and fluctuations in the incident X-ray beams intensity during the experiment) are compensated for with a scaling algorithm. Partial reflections (single reflections that are split across multiple diffraction images) and reflections related by symmetry were then merged. The reflection intensities are also merged into amplitudes. In the Collaborative Computational Project No. 4[148] (CCP4i2) program suite, POINTLESS[149,150] was used to select the space group $P2_12_12_1$.

The following structure refinement was carried out by Dr Ivo Tews, Charlotte Cordery and Jack Craddock.

The structure of the RbdA $EAL_{549\text{-}797}$ domain was determined using molecular replacement with MOLREP[151] and the *P. aeruginosa* FimX EAL domain (PDB code 3HV8[152]), followed by iterative cycles of refinement using REFMAC[153] and manual model building using COOT[154] in the CCP4i2 software suite. The phase estimates, used to produce an initial electron density map, can be improved by refining the model structure through the application of stereochemical restraints (such as bond angles and bond lengths) and by manual manipulation (such as mutating residues to match the protein sequence and/ or adding in ligands and water molecules). Improvements to the phase information results in improvements to the electron density map[143,146]. Continuous validation of the model was carried out by reviewing various statistics. Calculation of the *Rwork* is a measure of the agreement between the model and the experimental data. The *Rfree* is calculated using the same equation as the *Rwork*, except the *Rfree* uses reflection data (usually 5 %) which has not been used during model refinement. The root mean squared deviation (RMSD) of the bond angles and bond lengths that are present within the model, provides information on the models geometry compared to what would be expected in a protein. The Ramachandran plot shows the percentage of the models residue backbone, using the backbone conformational angles phi (φ) and psi (Ψ), which falls into preferred, allowed or disallowed regions. Structure validation was carried out using the PDB-REDO[155] online server and the worldwide protein data bank (wwPDB)[156] validation server.

Chapter 3 – Interactions between GGDEF and EAL domains regulate c-di-GMP turnover

Pseudomonas aeruginosa PAO1 contains 16 proteins that contain both GGDEF and EAL domains arranged in tandem[87]. These domains have opposing enzymatic functions, so how these domains are regulated so that only one of these two domains has an overriding activity at any one time remains an important question. In some systems, only one of the two domains in the GGDEF-EAL tandem is active with the other being degenerate and/ or acting as a receptor domain[56,57,157]. The GGDEF-EAL tandem domains are often preceded by sensory domains[60]. One idea is that, when both domains are active, the sensory domain regulates the DGC and / or PDE activity by acting as a molecular switch through conformational changes upon a stimuli binding or unbinding to the sensory domain[56]. We were interested in how the presence of opposing enzymatic domains could influence the other domains activity. In this study, the aim was to understand how the presence or absence of the GGDEF domain could influence the PDE activity of the EAL domain.

To study this we used the protein RbdA, which has been shown to possess both DGC and PDE activities[131] and to be involved in the NO induced biofilm dispersal within *P. aeruginosa*[74], with dispersal taking place with an increase in PDE activity[58,75]. Enzymatic assays revealed that the PDE activity of the EAL domain was negatively regulated by the tandem GGDEF domain. This regulation was further probed at the molecular level by attempting to determine the crystal structures of the isolated EAL domain and GGDEF-EAL tandem domain of RbdA.

3.1 – Analysis of the PDE activity of RbdA

We hypothesized that the presence of the GGDEF domain could influence the PDE activity of its tandem EAL domain. To test this hypothesis we use the tandem GGDEF-EAL system of RbdA. It has previously been found that RbdA has PDE activity and that this PDE activity can be increased with the addition of GTP[129] or the GTP analogue GMPPNP[131], which is attributed to GTP or GMPPNP binding to the A site of the GGDEF domain. However it is unclear how the presence of the GGDEF domain alone can influence the PDE activity of its tandem EAL domain.

3.1.1 – The PDE activity of the EAL domain of RbdA is negatively regulated by the tandem GGDEF domain

To determine whether the GGDEF domain alone had any influence over the PDE activity of the EAL domain of RbdA, the isolated $EAL_{549-797}$ and isolated $GGDEF_{376-536}$ domain fragments along with the tandem GGDEF-$EAL_{376-797}$ fragment of RbdA were produced. All RbdA constructs (Table 2.2) used in this thesis were expressed in BL21 (DE3) cells (as detailed in Section 2.8) and the proteins purified with IMAC using Ni-NTA resin beads and size exclusion chromatography (SEC; as detailed in Section 2.9). To help stabilise the three RbdA protein constructs during protein concentration, 50 mM L-glutamic acid and 50 mM L-arginine were added to the gel filtration buffer as described by Golovanov et al.[158]. This was found to reduce the amount of protein that eluted in the SEC void volume (approximately 46 mL), thus reducing the amount of protein that precipitated during SEC.

An example of protein purifications for RbdA $EAL_{549-797}$, RbdA $GGDEF_{376-536}$ and RbdA GGDEF-$EAL_{376-797}$ are shown in Figures 3.1, 3.2, 3.3 respectively.

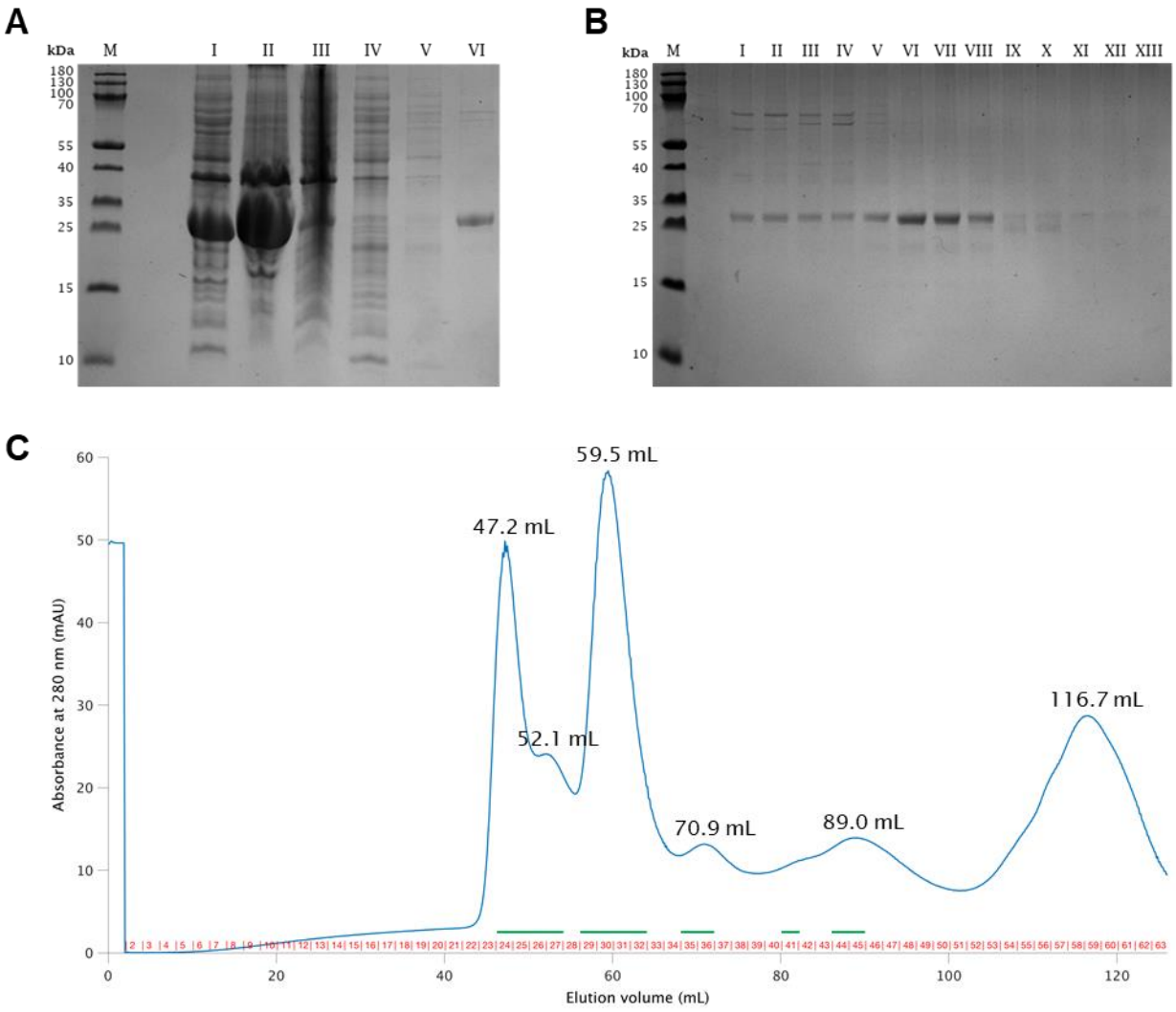

Figure 3.1. **Purification of RbdA EAL$_{549-797}$**. A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of RbdA EAL$_{549-797}$ (approximately 30 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography of the RbdA EAL$_{549-797}$ IMAC eluate showing fractions 24-27, 29, 30-32, 35, 36, 41, 44 and 45 in lanes I –XIII respectively. C) Size exclusion chromatography trace of RbdA EAL$_{549-797}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 prep grade with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

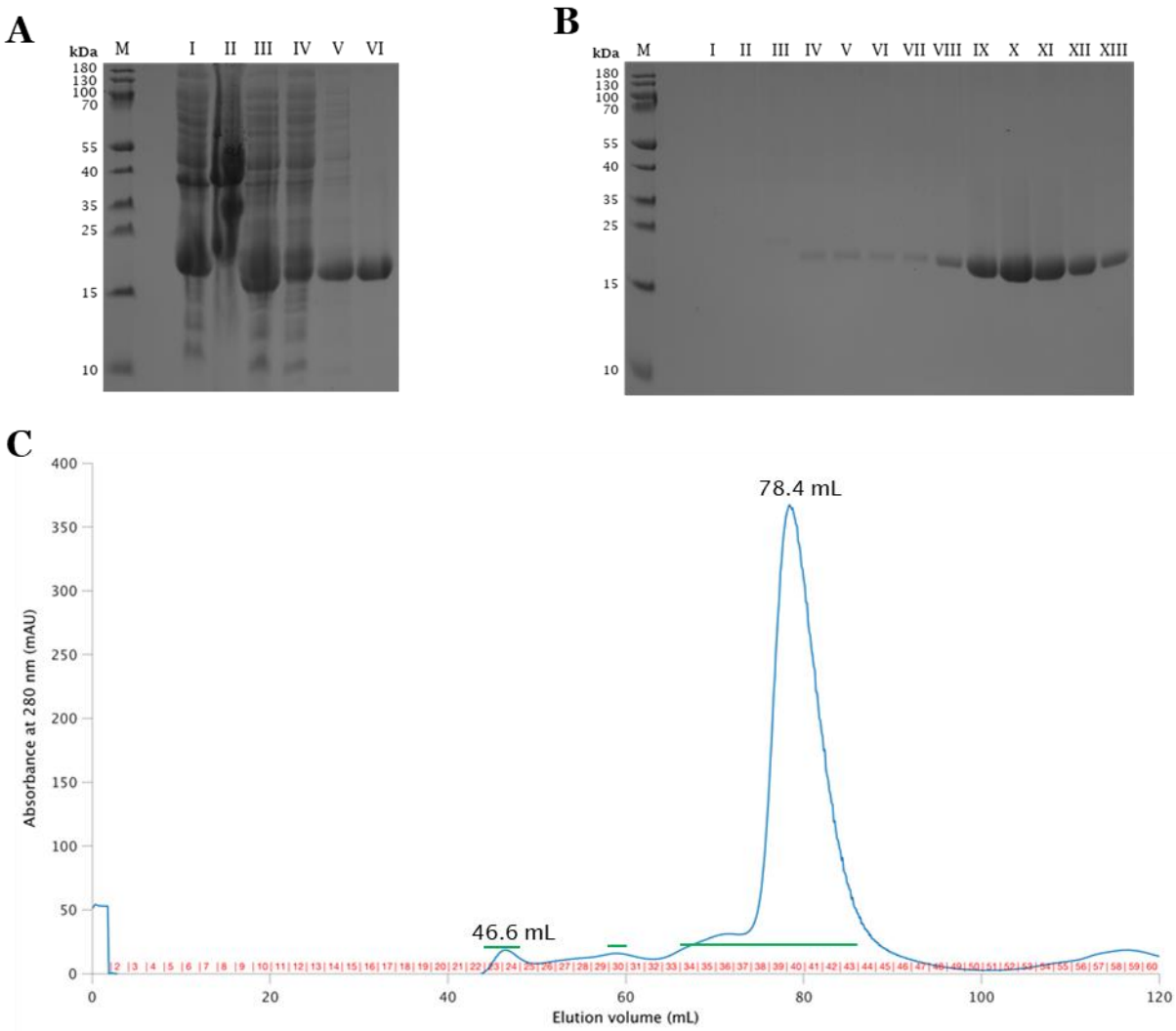

Figure 3.2. **Purification of RbdA GGDEF$_{376-536}$.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of RbdA GGDEF$_{376-536}$ (approximately 20 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography of the RbdA GGDEF$_{376-536}$ IMAC eluate showing fractions 23, 24, 30, 34-43 in lanes I – XIII respectively. C) Size exclusion chromatography trace of RbdA GGDEF$_{376-536}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 prep grade with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

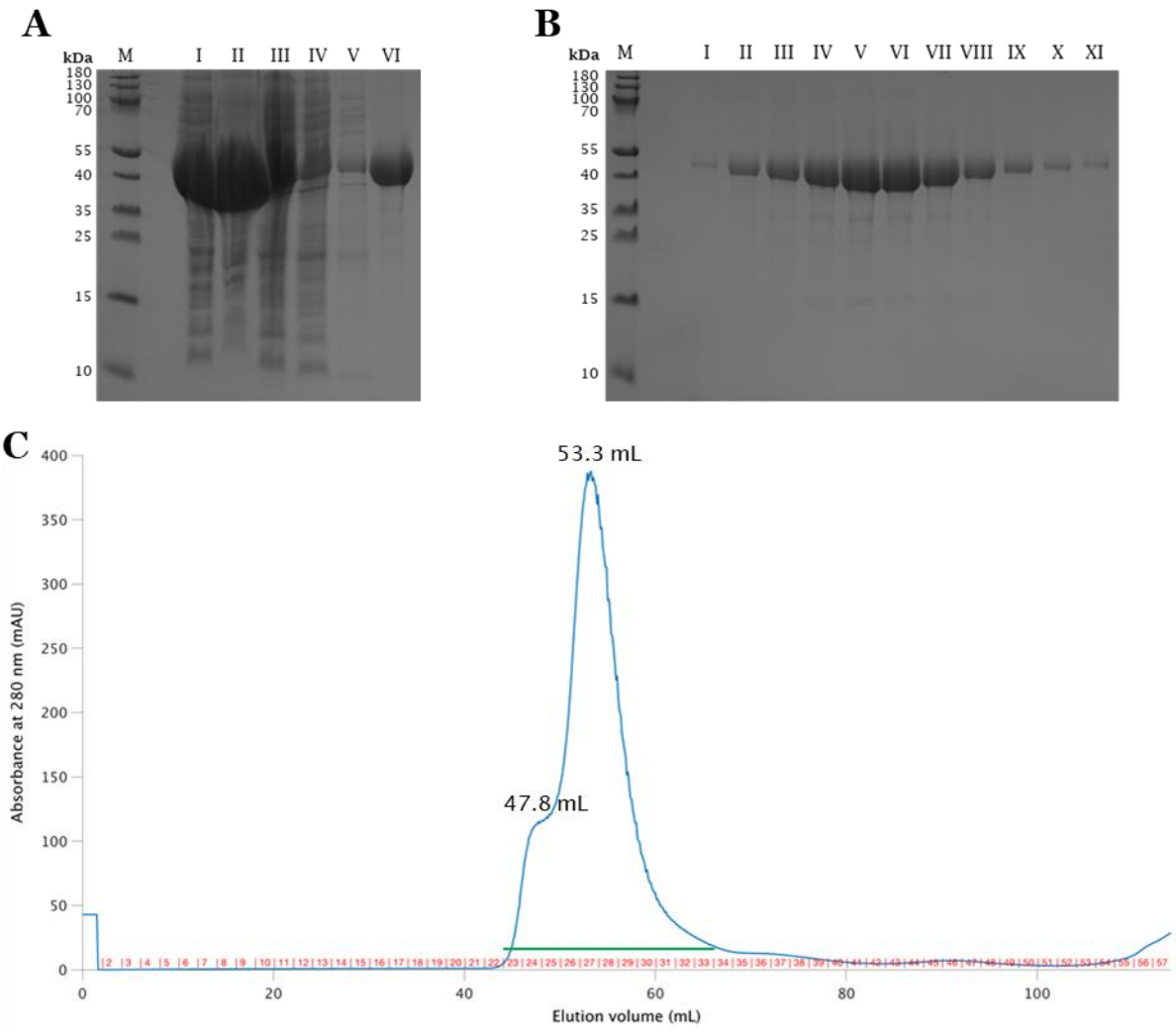

Figure 3.3. **Purification of RbdA GGDEF-EAL$_{376-797}$.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of RbdA GGDEF-EAL$_{376-797}$ (approximately 50 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography of the RbdA GGDEF-EAL$_{376-797}$ IMAC eluate showing fractions 23-33 in lanes I – XI respectively. C) Size exclusion chromatography trace of RbdA GGDEF-EAL$_{376-797}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 prep grade with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

A calibration curve of the SEC column (shown in Appendix Figure 7.2) was used to determine the oligomerisation state of the proteins after purification through SEC. RbdA $EAL_{549-797}$ (molecular weight 30,077.58 Da) is mainly eluted from the SEC column as a dimer at approximately 59.5 mL (as suggested from a calculated molecular weight of 61,787 Da) but also as a monomer at approximately 71 mL (as suggested from a calculated molecular weight of 28,242 Da; Figure 3.1). RbdA $GGDEF_{376-536}$ (molecular weight 20,425.18 Da) is eluted from the SEC column as a monomer at approximately 78.5 mL (as suggested from a calculated molecular weight of 16,950 Da; Figure 3.2) and RbdA GGDEF-$EAL_{376-797}$ (molecular weight 49,728.94 Da) is eluted as a dimer at approximately 53.3 mL (as suggested from a calculated molecular weight of 94,232 Da; Figure 3.3). However, even with the 50 mM L-glutamic acid and 50 mM L-arginine addition in the gel filtration buffer, RbdA $GGDEF_{376-536}$ precipitated soon after elution from the SEC column and so could not be used for further enzymatic analysis. Therefore the initial rates of the c-di-GMP specific PDE activity (when using 100 µM c-di-GMP) were only measured for the RbdA $EAL_{549-797}$ isolated domain and RbdA GGDEF-$EAL_{376-797}$ tandem domain.

To ensure that the proteins were as fresh as possible so as not to lose any enzymatic activity, protein purification and the PDE activity assay (Section 2.11) were carried out on the same day. Using a 1 mL Resource Q column the reaction substrate (c-di-GMP) and reaction product (pGpG) could be separated well using a 25 column volume linear gradient of ammonium bicarbonate (5 mM to 1 M) as shown in Figure 3.4 A. It was also found that c-di-GMP did not break down in the PDE reaction buffer (pH 9.35: 50 mM Bis-Tris Propane, 50 mM NaCl, 5 mM $MgCl_2$) without the addition of protein even after 24 hrs (as shown in Figure 3.4 B). Furthermore it was confirmed that the PDE reaction could be stopped by adding the 100 µL reaction sample to 10 µL 100 mM $CaCl_2$ and placing on ice before the sample was run over the Q column. Running protein alone at 1.5 µM over the Q column also did not produce a significant signal when measuring absorbance at both 280 nm and 253 nm.

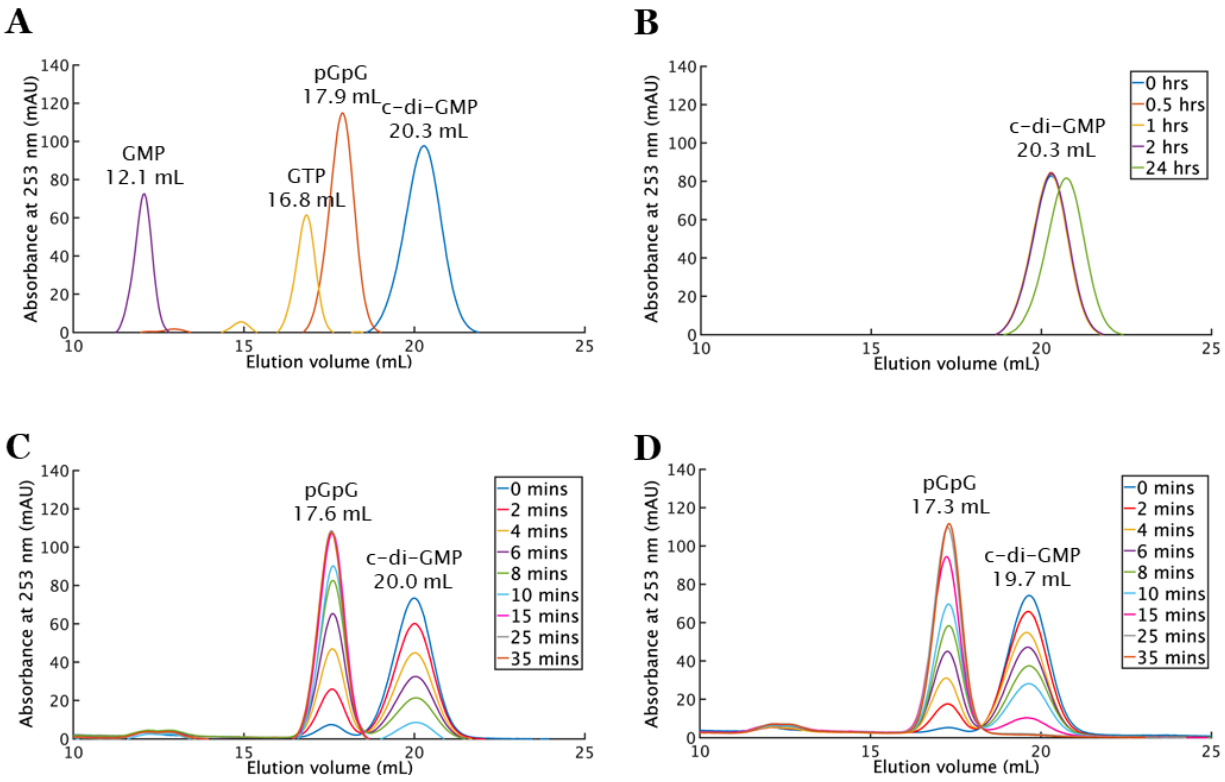

Figure 3.4. **Phosphodiesterase activity of RbdA EAL$_{549-797}$ and RbdA GGDEF-EAL$_{376-797}$.** A) Elution profiles of c-di-GMP, pGpG, GTP and GMP 100 µM standards when run over a Resource Q column measuring absorbance at 253 nm. B) The elution profile of 100 µM c-di-GMP when incubated without enzyme, showing no c-di-GMP breakdown after 24 hours incubation. C) Elution profile of RbdA EAL$_{549-797}$ showing the breakdown of c-di-GMP to pGpG. D) Elution profile of RbdA GGDEF-EAL$_{376-797}$ showing the breakdown of c-di-GMP to pGpG which is slower than compared to that of RbdA EAL (C).

It was found that the isolated EAL$_{549-797}$ domain of RbdA had an increased PDE activity than compared to the RbdA GGDEF-EAL$_{376-797}$ tandem domain (Figure 3.4 C-D). With the RbdA isolated EAL$_{549-797}$ domain having an initial rate of 7.24 ± 0.3975 µM pGpG/min (± standard error) compared to the RbdA GGDEF-EAL$_{376-797}$ tandem initial rate of 4.507 ± 0.1208 µM pGpG/min (shown in Figure 3.5), when in the presence of 100 µM c-di-GMP. These were found to be statically significantly different (P = 0.0002).

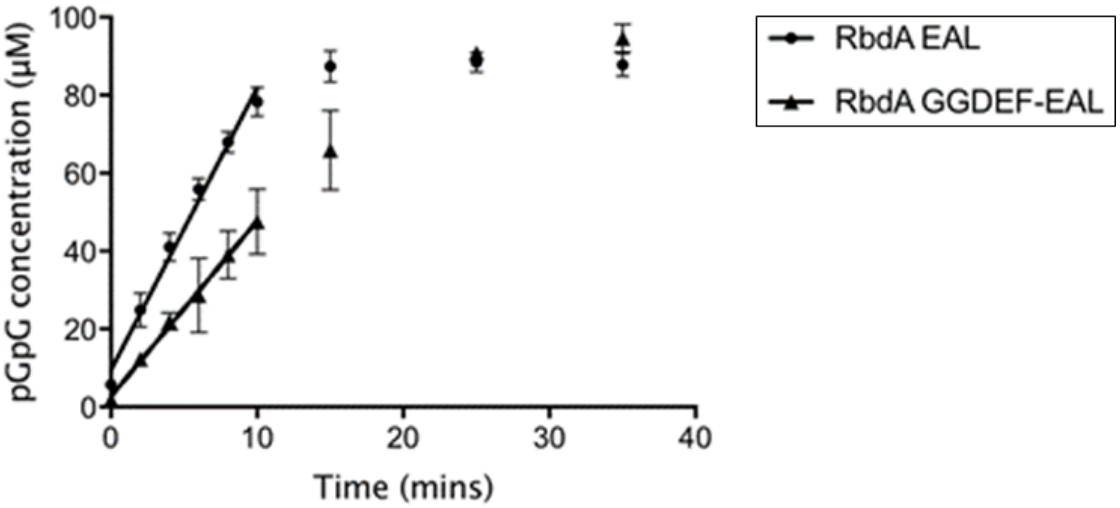

Figure 3.5. **The initial rates of phosphodiesterase activity for RbdA EAL$_{549\text{-}797}$ and RbdA GGDEF-EAL$_{376\text{-}797}$ when using 100 μM c-di-GMP**. Data is plotted in GraphPad Prism (v. 7) and a line of best fit produced to give the equation for RbdA EAL$_{549\text{-}797}$ (circles) of $y = 7.24x + 9.442$ with an $R^2 = 0.9881$ and an equation for RbdA GGDEF-EAL$_{376\text{-}797}$ (triangles) of $y = 4.507x + 2.663$ with an $R^2 = 0.9971$. These were found to be statistically significantly different with $P = 0.0002$.

3.2 – The molecular basis for the GGDEF regulation of the EAL domain in RbdA

To further investigate, at the molecular level, the GGDEF domains negative regulation of its tandem EAL domain in RbdA, we sought to determine the structure of the isolated $EAL_{549-797}$ domain and the tandem GGDEF-$EAL_{376-797}$ domain of RbdA. To this end X-ray crystallography was used. The RbdA $EAL_{549-797}$ isolated domain was successfully crystallised and the structure determined to a resolution of 2.4 Å. While apparent crystallisation of the RbdA GGDEF-$EAL_{376-797}$ tandem domain occurred after approximately 7 months, the diffraction data is still being analysed for structure determination.

3.2.1 – Crystal structure of the isolated $EAL_{549-797}$ domain of RbdA

The isolated RbdA $EAL_{549-797}$ domain was successfully crystallised using 10 mg/mL after 24 hrs at 21 °C with many crystal hits within a screen of Morpheus and a few within JCSG$^+$ (wells A5, A7, B3, B8, C4, C10, F6, G4, G10 and H1). An optimisation screen (shown in Appendix Table 7.1) based on Morpheus was produced which gave crystals in most conditions. A crystal (Figure 3.6) grown in condition H11 of the optimisation screen was diffracted on beamline ID23-1 at the ESRF, wavelength 0.9789 Å (data collected 15/09/2017). Using molecular replacement with the *P. aeruginosa* FimX EAL domain (PDB code 3HV8), the structure for the isolated RbdA $EAL_{549-797}$ domain was determined to 2.4 Å. Full statistics for data collection and data processing are shown in Table 3.1.

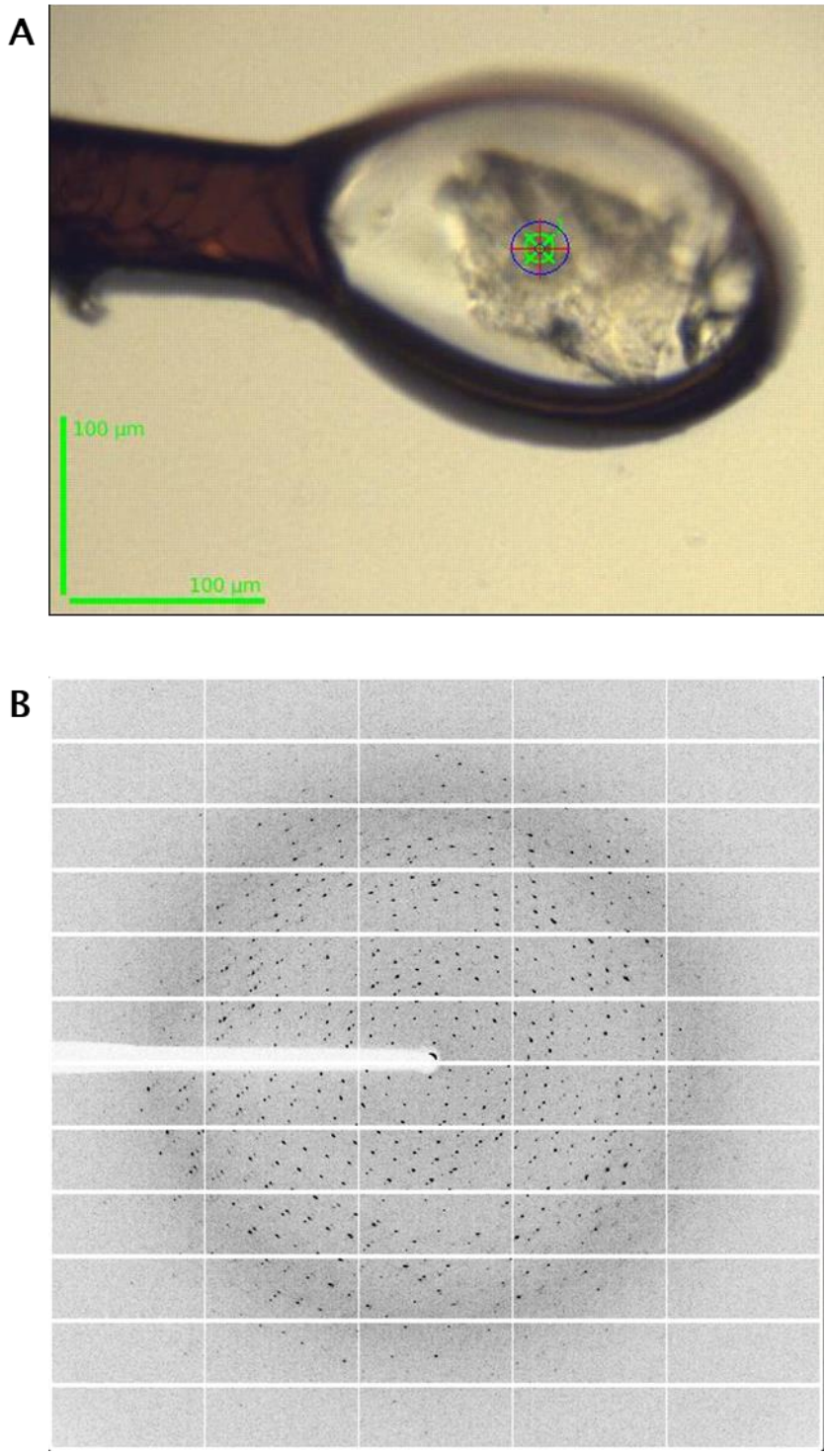

Figure 3.6. **RbdA EAL$_{549-797}$ crystal.** A) The RbdA EAL$_{549-797}$ crystal diffracted on ID23-1 at the ESRF that was used to determine the structure. B) An example of a diffraction pattern from the collected dataset that was used for RbdA EAL$_{549-797}$ structure determination.

Table 3.1. **Data collection and refinement statistics for RbdA $EAL_{549-797}$.**

	RbdA $EAL_{549-797}$
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	64.2, 65.6, 171.9
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	171.87 – 2.30 (2.34 – 2.30)
*R*merge	0.072 (0.689)
CC 1/2	0.978 (0.654)
I / σ*I*	11.4 (0.4)
Completeness (%)	98.8 (89.1)
Redundancy	4.8 (3.0)
Refinement	
Resolution (Å)	2.30
No. reflections	32273
*R*work / *R*free	19.9 / 24.9
No. atoms	
Protein	3907
Ligand/ion	4 / 2
Water	224
B-factors	
Protein	61.9
Ligand/ion	53.77 / 49.31
Water	56.57
RMS Deviations	
Bond lengths (Å)	0.0148
Bond angles (°)	2.13
Ramachandran (%)	
Preferred	97.11
Allowed	2.89
Disallowed	0
PDB Code	

The RbdA $EAL_{549-797}$ domain adopts an $\alpha\beta(\beta\alpha)_6\beta$ – barrel fold, shown in Figure 3.7 A, which is similar to other EAL domain structures. Here, RbdA $EAL_{549-797}$ was found as a 'classic dimer'[80], with the dimer interface formed by the α5 and α6 helices and part of the β5 – α5 loops of each monomer. The dimer interface is held together by four hydrogen bonds and two salt bridges (shown in Figure 3.7 B), with a buried accessible surface area at the dimer interface of 743.2 $Å^2$. In MorA the dimer interface is stabilised by the reciprocal contacts formed

between residues Tyr1316 and Ser1317 of the $_{1310}$DDFGTGYSS$_{1318}$ motif on the β5 – α5 loop of each monomer[73]. The β5 – α5 loop of RbdA contains the non-conserved $_{708}$DDFCAGMSS$_{716}$ motif. However the dimer interface of RbdA EAL$_{549\text{-}797}$ is stabilised by hydrogen bonds formed by the main chain of Gly713 and Met714 of one monomer with the side chain of Ser716 and the main chain of Phe717 respectively of the other monomer. The requirement of the residues DFG(A/S/T)(G/A)(Y/F)(S/A/T)(S/A/G/V/T) in forming the DDFGTGYSS catalytic motif in the β5 – α5 loop has recently been suggested to be too restrictive[159]. However the first two Asp residues of the DDFGTG(YSS) motif are absolutely required for PDE catalytic activity and are involved in the coordination of catalytic metal ions[76] (see below).

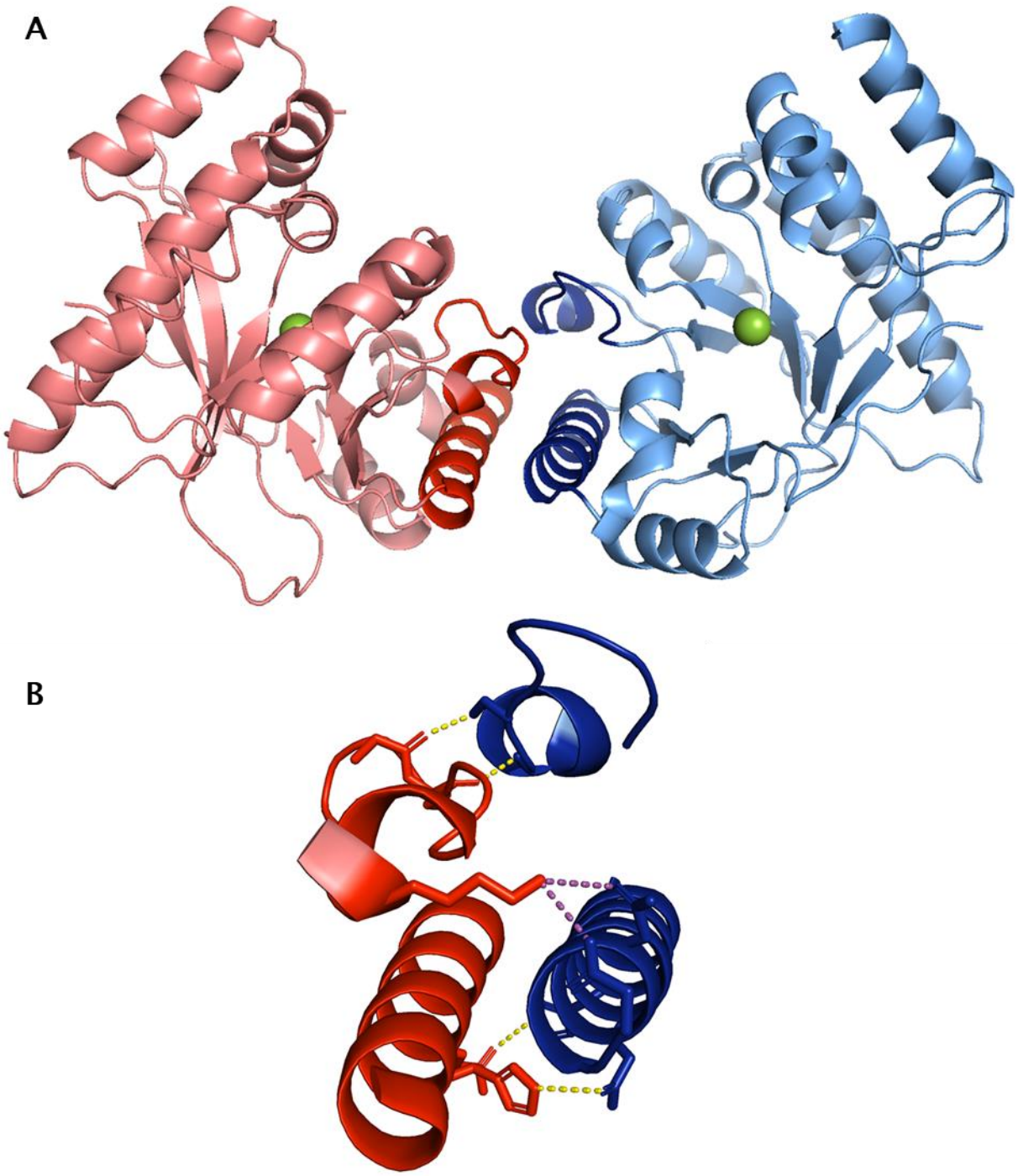

Figure 3.7. **The structure of dimeric RbdA EAL$_{549-797}$.** A) The RbdA EAL$_{549-797}$ dimer is shown in cartoon representation with the monomers coloured in red and blue with the dimer interface highlighted in a darker colouring. Magnesium ions (one in each monomer) are shown as green spheres. B) A zoomed in image of the dimer interface with hydrogen bonds shown as yellow dashed lines and salt bridges shown as purple dashed lines.

This RbdA $EAL_{549-797}$ dimer is asymmetric (an RMSD of 1.297 Å over all atoms) with the major difference between the two monomers occurring at the β5 – α5 loop and the α5 helix (shown in Figure 3.8).

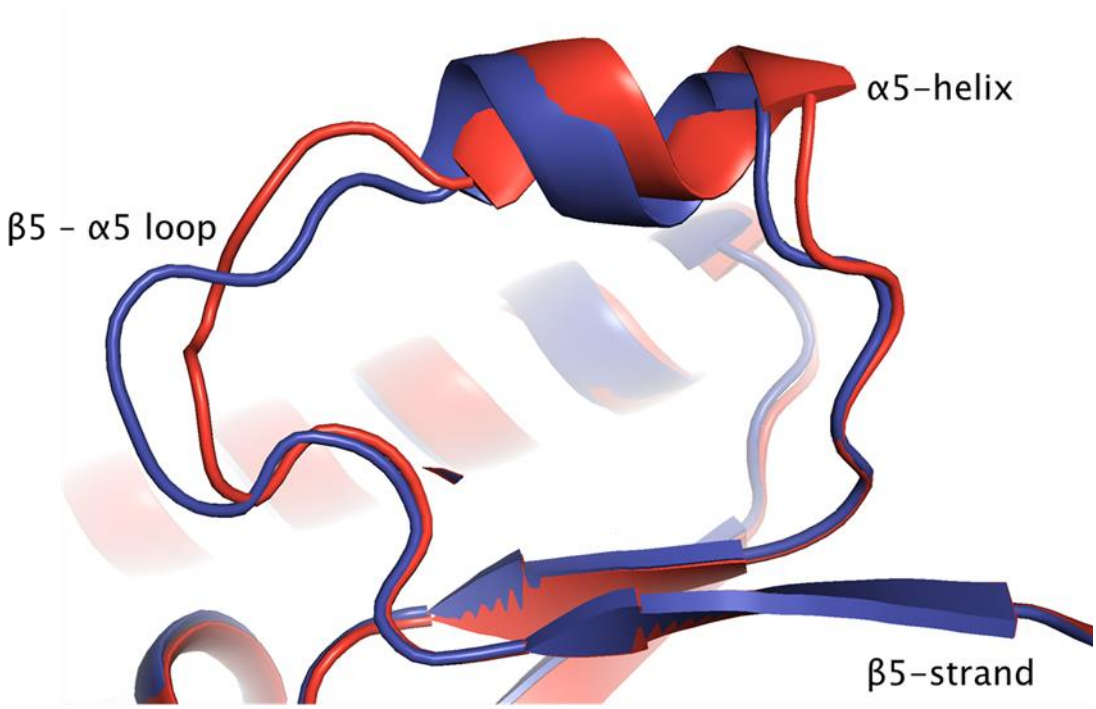

Figure 3.8. **The main differences in the asymmetric RbdA $EAL_{549-797}$ dimer.** When the monomers of the RbdA $EAL_{549-797}$ dimer are superposed the major difference occurs at the β5 – α5 loop with a slight displacement at the α5-helix. The monomers are shown in a cartoon representation.

Each monomer was found with a single metal ion within the active site but without bound substrate (c-di-GMP) or product (pGpG). The metal ions were designated as magnesium ions due to magnesium chloride being present in the size exclusion chromatography purification buffer. When compared to other EAL domain structures, the magnesium ion is found in the M1 position. Although there is a difference between the monomers at the β5 – α5 loop which contains the catalytic DDFGTG motif involved in metal ion coordination[73], the coordination of the magnesium ion is identical. It is coordinated by Glu585, Asn646, Glu678, Asp708 and two water molecules in an octahedral geometry (Figure 3.9). This observed metal ion coordination is consistent with the coordination of an M1 metal ion[80].

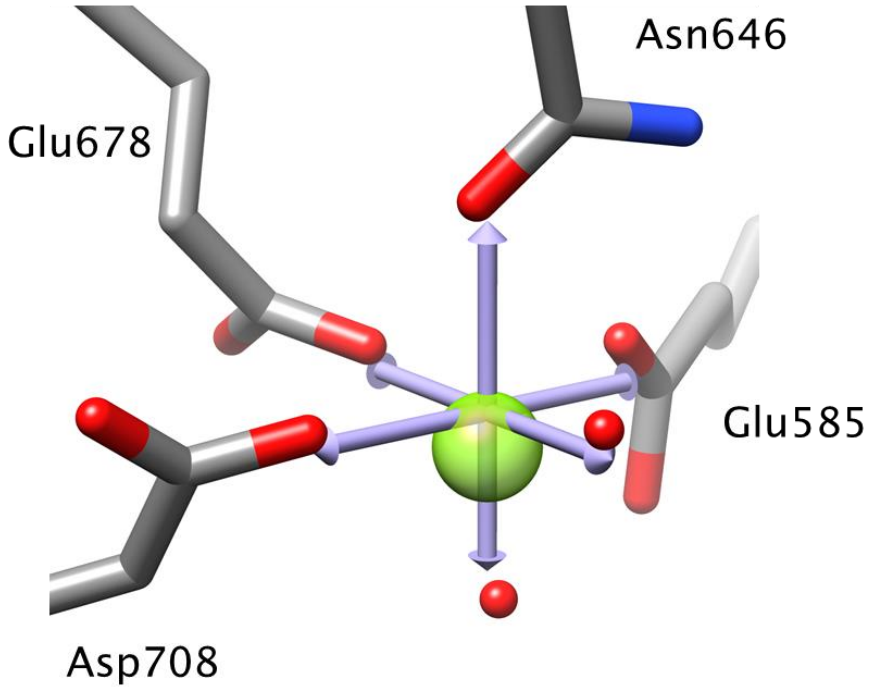

Figure 3.9. **The idealised coordination geometry of the magnesium ion within RbdA EAL$_{549-797}$.** The magnesium ion is depicted as a transparent green sphere, waters as red spheres and the coordinating side chains in grey elemental colours. The idealised octahedral coordination geometry is shown as arrows in purple.

Although this structure of RbdA EAL$_{549-797}$ only contains one metal ion in the M1 position, metal ions (magnesium or manganese) in both the M1 and M2 positions are required for the activation of a hydroxide nucleophile[76,78,160]. While a metal ion in the recently proposed M3 position may be required for stabilising the negative charge of the transition state formed during c-di-GMP hydrolysis to pGpG[80], a further M4 metal ion has been suggested to induce conformational changes which prevents re-ligation and promotes pGpG disassociation[77]. In this RbdA EAL$_{549-797}$ structure the residues, Asp708, Asp709 and Glu765, required for the M2 metal coordination are present, with Asp709 being in a conformation that could be receptive to metal ion coordination (Figure 3.10). However, Asp708 and Glu765 do not have conformations that would be able to coordinate an M2 site metal ion. Glu765 would need to adopt an alternative rotamer, while Asp708, corresponding to an Asp residue that usually coordinates both metal ions in the M1 and M2 positions in other EAL domain structures, cannot achieve the required conformation *via* a rotameric change. Therefore a slight conformational shift, possibly in the β5 – α5 loop, would be required to allow for the Asp708 residue to adopt a conformation receptive to metal ion coordination in the M2 site.

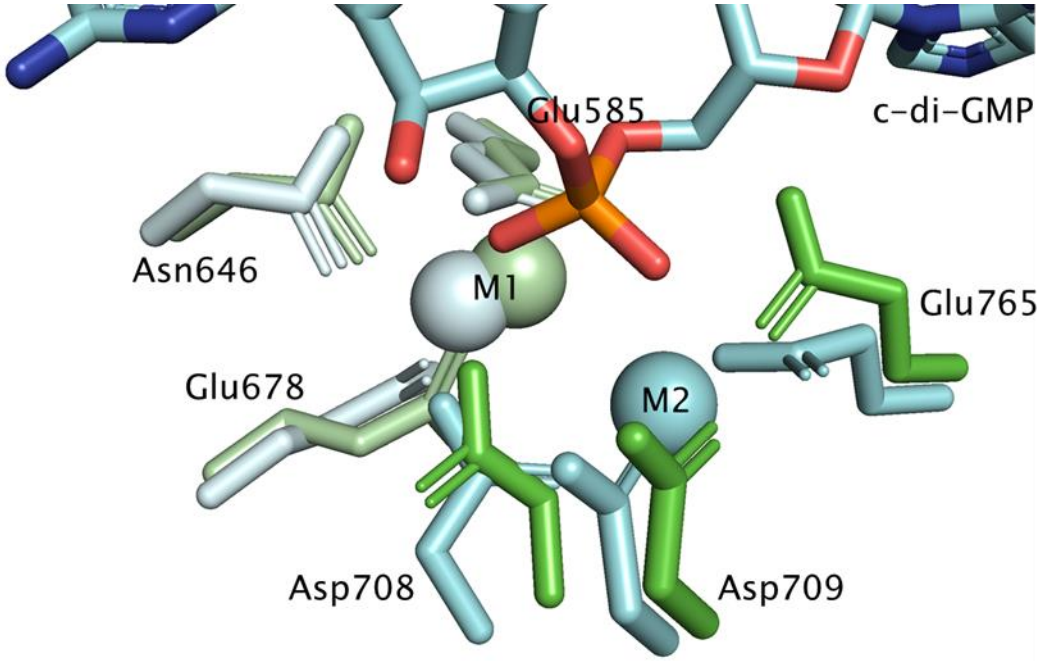

Figure 3.10. **The residues which coordinate a putative M2 site metal ion within RbdA EAL$_{549-797}$.** The RbdA EAL$_{549-797}$ structure (shown in green) is superposed onto MucR EAL (PDB code 5M1T – shown in cyan). The coordinating side chains and c-di-GMP are depicted as sticks with the magnesium ions shown as spheres. Side chains coordinating the magnesium ion in the M1 site are shown in a lighter colour. Residues are labelled according to RbdA numbering.

To date there are three EAL domain structures with metal ions in the proposed M3 site[77,80]. The M3 positioned sodium ion in PA3825 (PDB code 5MFU) is coordinated by a P1 hydroxyl oxygen from pGpG and a ribose O2' oxygen as well as a water molecule and an Asp residue. However in CC3396 (PDB code 3U2E) the magnesium ion in the M3 position is coordinated entirely by the bound pGpG and water molecules, with the ribose O3' oxygen and waters replacing the coordination by the Asp residue found in PA3825[80]. In this RbdA EAL$_{549-797}$ structure, Asp708 would be able to coordinate a sodium ion in the M3 position. However if the M1, M2 and M3 positions were all occupied at once by metal ions, Asp708 as described above, would require a slight conformational change to enable it to coordinate the M2 metal ion. The slight conformational change of the Asp708 residue in RbdA EAL$_{549-797}$ would then enable it to coordinate all three metal ions. In the pGpG bound state of *Vc*EAL (PDB code 6IJ2), the metal sites are occupied by Ca^{2+} ions, with the M3 site Ca^{2+} ion coordinated by a Thr and Asp residue (equivalent to Thr680 and Asp709 in RbdA respectively) as well as a water molecule. It is the water molecule that mediates the M3 site metal ions interaction with the

sugar of pGpG[77]. The Thr680 side chain in our RbdA EAL structure would have to undergo a rotameric change and the α–carbon of the Asp709 residue of our RbdA EAL structure would need to undergo a downwards shift of approximately 1.6 Å to be in a position to coordinate the M3 site metal ion as in *Vc*EAL. As the Asp708 residue of RbdA in our EAL structure would need to undergo a slight conformation change in order to coordinate the M2 metal ion, this would likely also move the Asp709 residue into position to allow for an M3 metal ion coordination. However as Ca^{2+} ions are inhibitory to an EAL domains PDE activity, the differences in the M3 site metal ion coordination in *Vc*EAL compared to CC3396 and PA3825 may be reflective of this inhibitory state. The M3 magnesium ion in CC3396 is coordinated entirely by the bound pGpG and waters, so it is likely that RbdA $EAL_{549-797}$ could accommodate a metal ion in the M3 position. Also a magnesium ion is more likely to be physiologically relevant compared to a sodium ion, thus the observed M3 site metal ion coordination in CC3396 is more likely to be of relevance.

The pGpG bound *Vc*EAL structure (PDB ID: 6IJ2) also reveals the presence of a fourth metal binding site (M4) that is occupied by a Ca^{2+} ion[77]. Compared to the c-di-GMP bound *Vc*EAL structure (PDB ID: 6IH1), the binding of the metal ion to the M4 site takes place after c-di-GMP hydrolysis to pGpG and requires, and presumably induces, conformation changes within the β5-α5 loop. This causes the M2 metal and the hydrolysed phosphate of pGpG to shift, increasing the distance between the hydrolysed phosphate and ribose sugar of pGpG suggesting a mechanism to prevent re-ligation and promote pGpG disassociation[77]. As in *Vc*EAL, conformational changes would be required to take place within our RbdA $EAL_{549-797}$ structure for it to accommodate the M4 metal ion. Within *Vc*EAL the M4 metal ion is coordinated by the peptide carbonyl oxygens of Asp159-Phe160, Ser162-Gly163 and Tyr164-Ser 165 and two water molecules, one of which is stabilised by the side chain of Glu131 (which undergoes a reorientation for M4 metal ion coordination)[77]. However, as Ca^{2+} ions are inhibitory and have different coordination geometries compared to Mg^{2+} or Mn^{2+} ions, the exact coordination of the M4 metal may differ from that observed in the pGpG bound *Vc*EAL structure (PDB ID: 6IJ2) compared to if the M4 site was occupied with catalytic ions. Therefore further investigation into the role and authenticity of this M4 site is required.

To date there are no EAL domain structures that have metals ions within the M1 and M2 positions but without a nucleotide (c-di-GMP or pGpG) bound (Table 3.2). Also, EAL domain

structures that have just one metal bound always occupy the M1 position and not the M2 position. This suggests that c-di-GMP may bring the M2 metal ion into the EAL domains active site as it binds, possibly with c-di-GMP binding to the M2 metal ion outside of the active site beforehand. However, there are EAL domain structures (PDB codes 4RNH, 6PWK, 3GFX and 2W27) which are bound to c-di-GMP with only the M1 site occupied. Interestingly, while not coordinated at the typical M2 site, the c-di-GMP molecule within the EAL active site of MorA (PDB code 4RNH) is coordinating a second metal ion. Although there are also structures with c-di-GMP and pGpG bound without metal ions in the M1 and M2 sites, these EAL domains are inactive as PDEs[152,160-163], except for RbdA (PDB code 5XGE[131]) and BlrP1 (PDB code 3GFY[78]) which are active PDEs. Furthermore, as previously pointed out[80], there are structures of EAL domains in which c-di-GMP is still bound with coordinated M1 and M2 metal ions even though the EAL domains are active PDEs (PDB codes 3N3T[76], 3GFZ and 3GG0[78], 5MF5[80]) (Table 3.2). This may be explained by the requirement of a third metal ion in the M3 position to stabilise the transition state during c-di-GMP hydrolysis[80]. This is further supported by the observation that the M3 metal binding site is not occupied in a c-di-GMP bound EAL-structure but only in pGpG bound structures. It is worth noting that the conformation of the Asp residue (RbdA Asp708) when involved in coordinating the M1 and M2 metal ions, would also be receptive to the coordination of a metal ion in the M3 site (as in PA3825; PDB code 5MFU), although as yet, this has not been directly observed. As shown in this work, RbdA EAL is an active PDE and the structure presented here could represent a state that is ready and primed for hydrolysis or one in which hydrolysis and the leaving of pGpG has just taken place or both. Based on the structural analysis presented here and findings presented by Yadav et al.[77], an updated schematic of c-di-GMP hydrolysis to that proposed by Bellini et al.[80] is presented in Figure 3.11.

Table 3.2. **An updated classification of all EAL domain structures in the Protein Data Base 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 **RbdA-EAL (5XGE)** *FimX-EAL (3HV8)* *FimX-EAL (4FOJ)* *FimX-EAL (4FOU)* *FimX-EAL (4FOK)* *FimX-EAL (4F3H)* *FimX-EAL (4F48)* *LapD-EAL (3PJU)* Dimeric **BlrP1-EAL (3GFY)** *LapD-EAL (3PJT)*	Monomeric DosP-EAL (4HU3) CC3396-EAL (3S83) **YahA-EAL (4KIE)** **PA3825-EAL (4Y9O)** **PA3825-EAL (4Y9N)** **RbdA-EAL (5XGB)** **RbdA-EAL (5XGD)** *FimX-EAL (3HV9)* *LapD-EAL (3PFM)* ***LapD-EAL (3PJW)*** ***LapD-EAL (3PJX)*** ***MorA-EAL (4RNF)*** Dimeric PA3825-EAL (4Y9M) YkuI (2BAS) Imo0131-EAL (4Q6J) MorA-EAL (4RNJ) MorA-EAL (4RNI) **DosP-EAL (4HU4)** *YdiV-EAL (3TLQ)* *FimX-EAL (4AG0)* ***FimX-EAL (4J40)***	Dimeric Mg^{2+} TBD1265-EAL (2R6O) **DcpA (5YRP)** **YahA-EAL (4LYK)** **LapD (6PWJ)** **Bd1971 (6HQ4)** Ca^{2+} PA3825-EAL (4Y8E) **PA0575-EAL (5M3C)** **Lmo0111-EAL (3KZP)** **VcEAL (6IFQ)** Tetrameric Mg^{2+} RocR (3SY8)	Dimeric Mg^{2+} MorA-EAL (4RNH) **LapD (6PWK)** Ca^{2+} BlrP1-EAL (3GFX) YkuI-EAL (2W27)	Dimeric Mn^{2+} BlrP1 (3GFZ) BlrP1 (3GG0) Mg^{2+} PA3825-EAL (5MF5) MucR-EAL (5M1T) TBD1265-EAL (3N3T) Ca^{2+} PA3825-EAL (4Y9P) **PA3825-EAL (5MKG)** CC3396-EAL (4HJF) BlrP1 (3GG1) YahA (4LJ3) **Bd1971 (6HQ5)** **VcEAL (6IH1)** Tetrameric Ca^{2+} **VcEAL (6K6T)**	Dimeric Mn^{2+} (M1 & M3) PA3825-EAL (5MFU) Mg^{2+} (M1, M2 & M3) CC3396-EAL (3U2E) Ca^{2+} (M1, M2 & M3) **VcEAL (6IJ2)** no metal *FimX-EAL (4AFY)*

This table has been taken and modified from 'Table 2' reported by Bellini et al.[80]. Non-catalytic EAL domains are shown in italics, PDB structures that have been added to the table in this thesis are shown in bold.

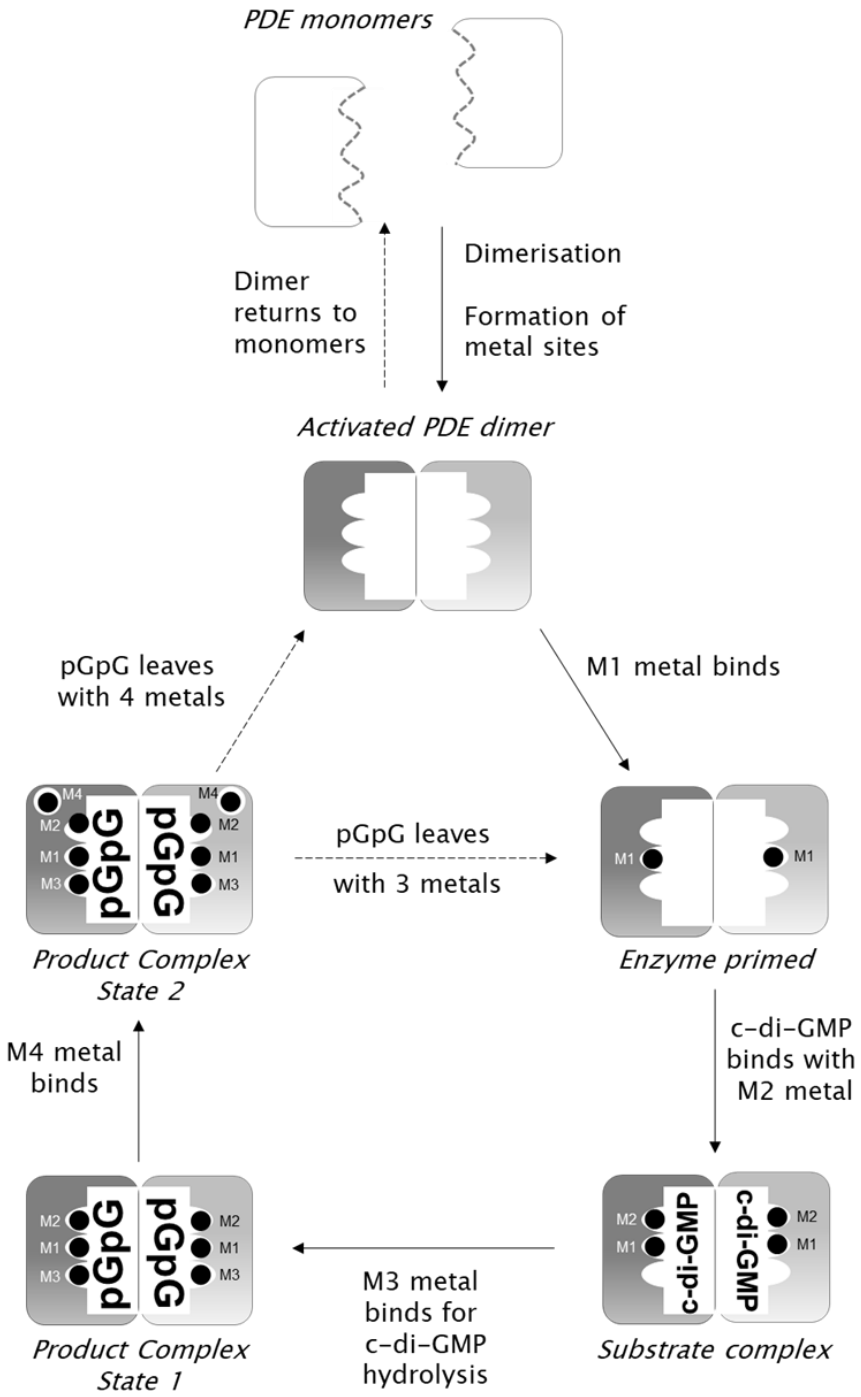

Figure 3.11. **An updated proposed schematic of the EAL domains PDE activity.** The dimerisation of monomeric EAL domains promotes residues involved in metal ion coordination to enter the active site. A metal ion can then be coordinated in the M1 site after which c-di-GMP binds with the M2 metal into the active site. Hydrolysis of c-di-GMP then takes place with an M3 metal required in the pGpG bound complex. The binding of an M4 metal ion induces conformational changes within the β5-α5 loop which causes the M2 metal ion to shift along with the hydrolysed phosphate of pGpG, preventing re-ligation and promoting pGpG disassociation from the active site. It is not known whether pGpG would exit with all four metal ions or with three.

Despite crystal soaks and co-crystallisation attempts with c-di-GMP or pGpG, no datasets collected were observed to contain bound nucleotide. To determine whether this RbdA $EAL_{549-797}$ structure would be receptive to nucleotide (c-di-GMP or pGpG) binding, a structural comparison between this RbdA $EAL_{549-797}$ nucleotide-free structure and a previously determined RbdA structure with c-di-GMP bound (PDB code 5XGE[131], referred to as $RbdA_{c\text{-}di\text{-}GMP}$ here after) was carried out. A structure superposition between the two RbdA structures is displayed in Figure 3.12. Most of the residues which form interactions with the bound c-di-GMP as described by Liu et al.[131] superpose well. The majority of the differences occur between RbdA $EAL_{549-797}$ and $RbdA_{c\text{-}di\text{-}GMP}$ at some of the residue side chains which interact with one of the guanosine bases (G2) of c-di-GMP (Figure 3.12 B and D). In $RbdA_{c\text{-}di\text{-}GMP}$, the G2 guanosine base interacts with; the side chain of Glu768 and the main chain of Leu787 *via* hydrogen bonding; the side chains of Gln569, Glu768 and Leu787 *via* Van der Waals; and the side chain of Phe766 through Pi-Pi stacking. When comparing monomer A of this RbdA $EAL_{549-797}$ dimer with $RbdA_{c\text{-}di\text{-}GMP}$, differences occur at the side chains of Phe766, Glu768 and Leu787 (Figure 3.12 B). Whereas in monomer B of RbdA $EAL_{549-797}$ differences occur at the side chains of Glu768 and Leu787 (Figure 3.12 D). The side chain of Leu787 of RbdA $EAL_{549-797}$ could adopt the same conformation as observed in $RbdA_{c\text{-}di\text{-}GMP}$ through a different rotamer. However the RbdA $EAL_{549-797}$ side chains of Phe766 and Glu768 would require slight conformational adjustments. To note, the conserved residue Arg589, which hydrogen bonds with a non-bridging oxygen of the P2 phosphate of c-di-GMP, is in a different rotameric state in monomer A of RbdA $EAL_{549-797}$ compared to $RbdA_{c\text{-}di\text{-}GMP}$. In monomer B, Arg589 adopts two alternative rotameric states, one of which is the same conformation as that observed in $RbdA_{c\text{-}di\text{-}GMP}$. The different rotameric states and slight conformational adjustments required for c-di-GMP binding in the RbdA $EAL_{549-797}$ structure, may be induced and stabilised as c-di-GMP binds, rather than being required before c-di-GMP could bind. This is supported by slight rotameric and conformational differences observed between an apo-RbdA structure (PDB code 5XGB) and $RbdA_{c\text{-}di\text{-}GMP}$ which was obtained via soaking crystals of apo-RbdA with c-di-GMP[131].

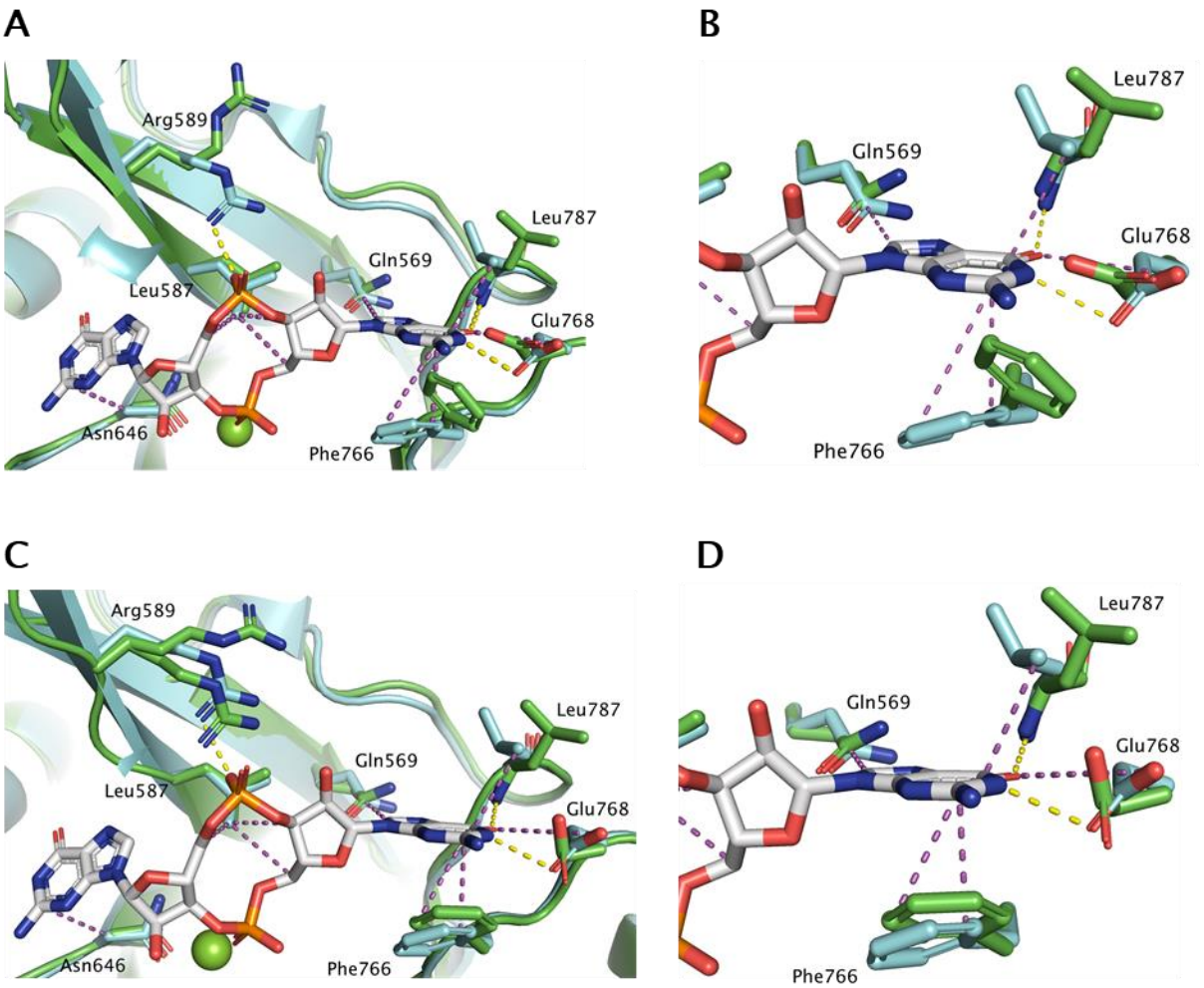

Figure 3.12. **Residues of RbdA EAL which interact with bound c-di-GMP.** A cartoon representation of RbdA $EAL_{549\text{-}797}$ shown in green superposed with an RbdA structure (PDB code 5XGE, referred to here as $RbdA_{c\text{-}di\text{-}GMP}$) shown in cyan bound with c-di-GMP (shown as sticks in grey). Main chains and side chains that interact with c-di-GMP via hydrogen bonds (yellow dashed lines), Van der Waals or Pi-Pi stacking (magenta dashed lines) are shown as sticks. A) Monomer A of the RbdA $EAL_{549\text{-}797}$ dimer superposed with $RbdA_{c\text{-}di\text{-}GMP}$. B) Zoomed in image of the G2 guanosine base of c-di-GMP when monomer A of RbdA $EAL_{549\text{-}797}$ is superposed with $RbdA_{c\text{-}di\text{-}GMP}$. C) Monomer B of the RbdA $EAL_{549\text{-}797}$ dimer superposed with $RbdA_{c\text{-}di\text{-}GMP}$. D) Zoom in of the G2 guanosine base of c-di-GMP when monomer B of RbdA $EAL_{549\text{-}797}$ is superposed with $RbdA_{c\text{-}di\text{-}GMP}$. The M1 site magnesium ion in RbdA $EAL_{549\text{-}797}$ is shown as a green sphere.

3.3 – Conclusion

To investigate whether the PDE activity of an EAL domain could be regulated by a tandem GGDEF domain, the bi-functional enzyme RbdA was studied. Here it was found that the isolated $EAL_{549-797}$ domain of RbdA had a higher PDE activity than compared to when the GGDEF-$EAL_{376-797}$ domains were kept in tandem (Figure 3.5). This would suggest that, in RbdA, the tandem GGDEF domain is negatively regulating the PDE activity of the EAL domain. This finding for RbdA is in contrast to that found for MorA[73] and MucR[164], in which the isolated EAL domains had lower or no PDE activity respectively compared to their GGDEF-EAL tandems which had a higher PDE activity.

To elucidate the structural mechanisms that are responsible for the GGDEF domains negative regulation of the EAL domain in RbdA, crystallographic structure determination was carried out. The structure of the isolated $EAL_{549-797}$ domain of RbdA was successfully determined, with further analysis of the datasets obtained for the RbdA GGDEF-$EAL_{376-797}$ tandem required. The isolated RbdA $EAL_{549-797}$ domain was determined in a canonical dimer with each monomer coordinating a single magnesium ion in the M1 site (Figure 3.7). Despite the RbdA $EAL_{549-797}$ structure containing an empty M2 metal site and no bound nucleotide, this EAL structure appears to be in a primed state, with only subtle conformational changes required for the binding of the substrate and M2 and M3 metal ions (Figures 3.10 and 3.12). Based on this work and previous work[77,80], we have suggested a schematic model for the EAL domains PDE activity (Figure 3.11).

Chapter 4 – Putative nitric oxide sensory domains of PA2072

P. aeruginosa contains the gene *pa2072* which encodes for a hypothetical protein (referred to as PA2072) with a domain organisation almost identical to that of RbdA, as predicted by the protein domain predictors SMART[132] and InterPro[133]. Recently, we published work investigating some of the potential targets involved in the biofilm dispersal pathway[130]. Surprisingly, we found that RbdA and PA2072 had opposing physiological roles within *P. aeruginosa* PAO1. Compared to a PAO1 WT, early stage biofilms of the Δ*rbdA* mutant were thicker with larger microcolonies (5.6 ± 3.2-fold), an increased biomass (2.1 ± 0.6-fold) and an increased surface coverage (1.3 ± 0.5-fold), while Δ*pa2072* mutant biofilms had smaller microcolonies (4.5 ± 1-fold), a reduced biomass (6.6 ± 0.6-fold) and reduced surface coverage (2.7 ± 2.2-fold). Correlating to this, we also found that the intracellular c-di-GMP levels relative to a PAO1 WT were increased in Δ*rbda* mutants (1.5 ± 0.2-fold) but reduced in Δ*pa2072* mutants (1.2 ± 0.1-fold). Furthermore, while the swimming motility of the Δ*rbdA* mutant was reduced, the Δ*pa2072* mutant had an equivalent swimming motility to the WT.

However, as with *rbdA*, the biological data indicated that *pa2072* was involved in the biofilm dispersal pathway. This was evidenced by the reduction in the NO-induced dispersal responses of Δ*rbdA* and Δ*pa2072* mutant biofilms compared to the dispersal response of PAO1 WT biofilms, with biomass reductions of 17 ± 8 %, 13 ± 8 % and 57 ± 5 % after NO treatment respectively[130]. In this published work, I predicted that PA2072 was an active DGC and PDE, based on the conservation of catalytic residues and active sites using sequence analysis and homology modelling of protein structures with the SWISS-MODEL server respectively. RbdA is known to be active as both a DGC and a PDE[129,131].

Here, we wanted to further characterise the PA2072 protein in order to understand, at the structural level, why it has an opposing physiological function to RbdA, even though their protein domain architectures are almost identical. We were interested in whether PA2072 could sense an NO stimulus and hypothesized that this may occur *via* its PAS domain binding to a cofactor. However, PA2072 is also predicted, by InterPro[133], to contain the periplasmic domain CHASE4 (cyclases/histidine kinases associated sensory extracellular[125,126]), which may also act as a sensory domain in PA2072. Of note, RbdA is not predicted by InterPro[133] or SMART[132] to contain a periplasmic domain but a model, proposed by Liu et al.[131], suggests

that a signal may be sensed by a putative periplasmic domain in RbdA to cause local conformational changes that propagate through the protein, resulting in a PDE enzymatic output.

To investigate the differences between RbdA and PA2072 that may give rise to the opposing physiological functions, a pairwise global sequence alignment was carried out. In order to assess whether the PAS domain of PA2072 could bind a cofactor, we expressed and purified the isolated $PAS_{310-427}$ domain of PA2072. Using UV-vis spectroscopy, we found that the PA2072 $PAS_{310-427}$ domain does not appear to bind a cofactor. Protein crystallisation trials were also carried out, but only small, non-diffracting crystals were produced. We also attempted protein crystallisation of the PA2072 CHASE4 domain, with the aim of elucidating the PA2072 CHASE4 structure which may have highlighted a sensory mechanism. While the initial crystallisation trials of the PA2072 $CHASE4_{45-254}$ domain were successful, further optimisation experiments of the $CHASE4_{45-254}$ protein crystals were not.

4.1 – The domain organisation of PA2072 and RbdA are homologous

The online protein domain prediction server, SMART[132], predicted that the domain organisation of RbdA and PA2072 were identical (Figure 4.1). Another online protein domain prediction server, InterPro[133], agreed with the SMART prediction except that the InterPro server predicted an additional periplasmic domain, CHASE4 (see Section 4.3), in PA2072 but not in RbdA. The CHASE4 domain in PA2072 is predicted by InterPro[133] to be formed of residues 57 to 213 (as shown in Figure 4.2). We were interested in whether this extra CHASE4 domain predicted in PA2072 could account for the physiological differences between RbdA and PA2072.

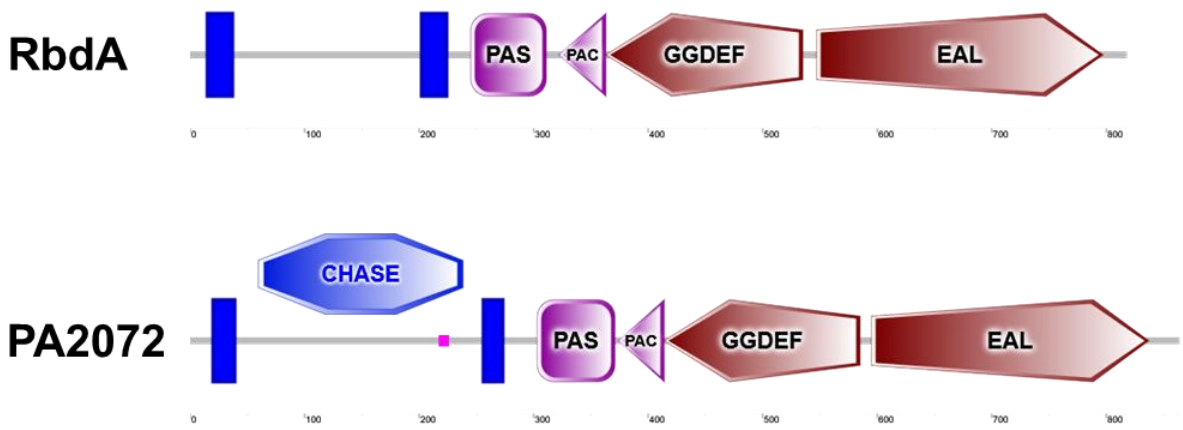

Figure 4.1. **Protein domain organisation of RbdA compared to PA2072**. RbdA is made up of 818 amino acids while PA2072 is formed of 864 amino acids. However the protein domain organisation predicted by SMART[132] is identical, being formed of two transmembrane regions (shown as vertical blue rectangles), a PAS domain (indicated by both the PAS (dark pink square) and PAC (dark pink triangle)), a GGDEF domain and an EAL domain (each shown as a red pentagon). A region of low complexity (shown as a bright pink square) predicted by SMART[132] in PA2072 is shown between residues 217 – 226. PA2072 is predicted by InterPro[133] to also contain a CHASE4 domain (which is indicated by the blue octagon) between residues 57 – 213. Grey lines indicate regions of unknown or unstructured regions. Figures acquired from the online protein domain prediction server SMART[132].

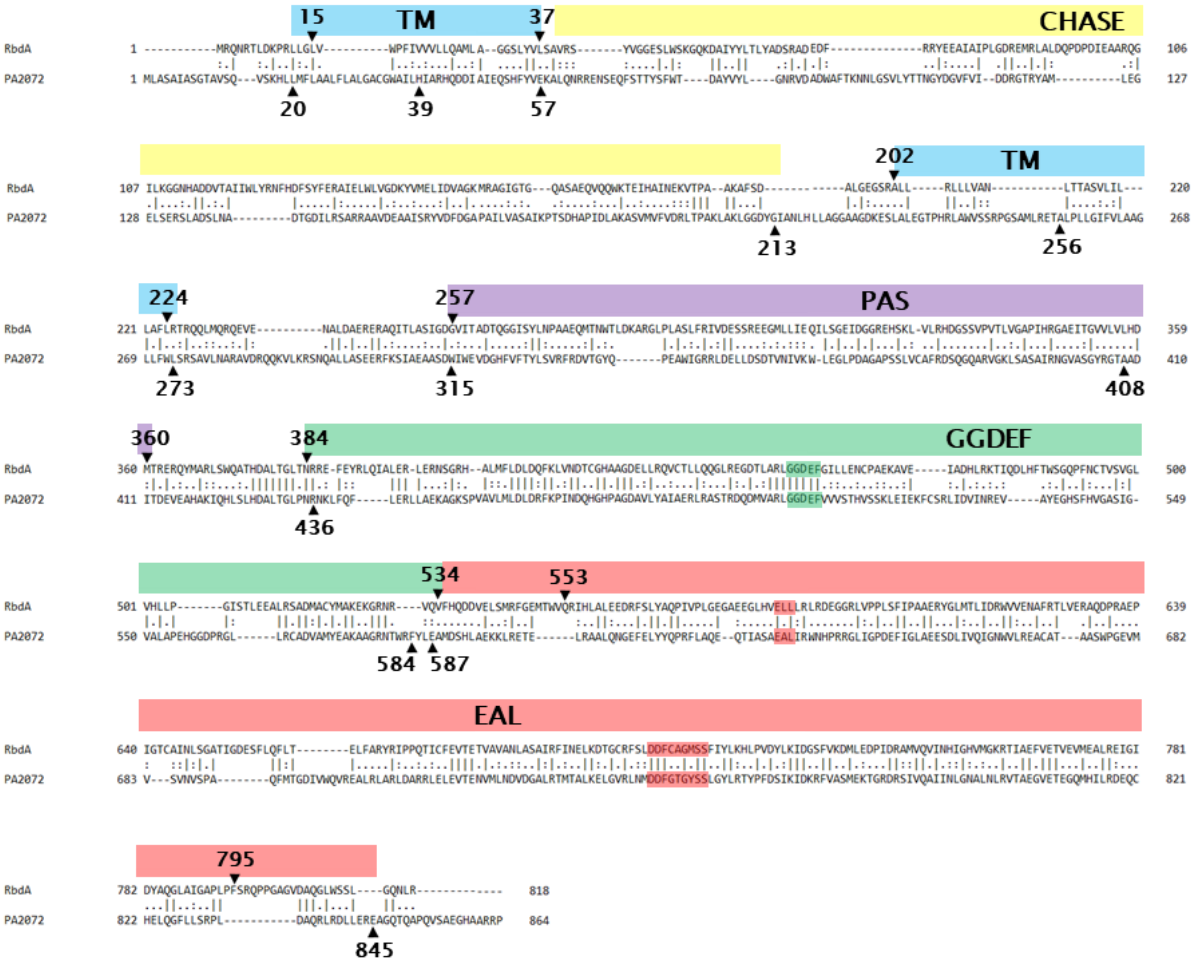

Figure 4.2. **Global protein sequence alignment between RbdA and PA2072.** RbdA and PA2072 global alignment using EMBOSS Needle[165] default parameters (matrix used was BLOSUM62, gap open penalty was 10, gap extend penalty was 0.5 and end gap penalty was false). The '|' between residues indicates that aligned residues are identical, the ':' indicates aligned residues are similar and share physico-chemical properties, the '.' indicates the aligned residues are less well conserved and the '–' in the sequences indicates a gap. The PA2072 CHASE4 domain is shown in yellow, transmembrane regions (TM) in blue, PAS domains in purple, GGDEF domains in green and EAL domains in red. The domain boundaries for RbdA and PA2072 are predicted using the secondary structure prediction server PSIPred[166], and current structural information[131] with the beginning and end of each domain being indicated by black arrows and residue number. Conserved catalytic residues are highlighted in the respective colours.

A global pairwise sequence alignment was carried out between the protein sequences of RbdA and PA2072 using the online server EMBOSS Needle[165] which uses the Needleman-Wunsch algorithm (Figure 4.2). Both protein sequences were aligned over their entire length giving the optimum alignment (including gaps) with an overall sequence identity of 24.5%, a similarity of 37.9% and gaps of 23.7%. The global protein alignment and protein domain

organisation suggests that RbdA and PA2072 are homologous proteins (and therefore paralogous genes).

It can be seen in Figure 4.2 that the differences in the sequence alignment between RbdA and PA2072 mainly occur between approximately the first 370 residues of RbdA and 420 residues of PA2072. Most of these differences between the RbdA and PA2072 protein sequences occur around the TM regions and the region of the predicted CHASE4 domain in PA2072. Currently, there have been eight different classes of CHASE domain identified[127,128] (see Section 1.3.3.4). Upon inspection of the secondary structures forming the periplasmic regions of RbdA and PA2072, the RbdA periplasmic region is predicted by the PSIPRED online server[166,167] to be formed of nine α-helices which vary in length (Figure 4.3 A). In contrast, the periplasmic region of PA2072 is predicted to be formed of both α-helices and β-strands, with a secondary structure order of α1-α2-β1-β2-α3-α4-β3-β4-β5-α5-β6-α6 (Figure 4.3 B) Nevertheless, the periplasmic region of RbdA was suggested by Liu et al. to be a putative sensory domain[131]. It is possible that RbdA also contains a sensory domain within its periplasm which is currently unidentified. However, it is likely that the structural differences predicted between the periplasmic regions of RbdA and PA2072 is a source for the opposing physiological roles of PA2072 and RbdA *in vivo*.

There is a low level of similarity between the PAS domains of PA2072 and RbdA (Figure 4.2). Still, it is well established that the primary sequence identity between PAS domains is low but that their three-dimensional architectures are highly conserved[111,113]. Our lab has previously investigated whether the primary sequence of PAS domains could be used to predict cofactor binding when comparing them to the sequences of PAS domains with known cofactor binding[134]. PAS domains associated with different cofactors can sense a range of different stimuli[107]. Nevertheless, the PAS domains of RbdA and PA2072 were not predicted to bind a cofactor[134]. However, PAS domains can directly sense stimuli, be involved in transducing conformational changes and can mediate protein dimerisation without a bound cofactor[119]. Therefore it is feasible that the differences in the physiological roles of RbdA and PA2072 are due to the differences between their PAS domains.

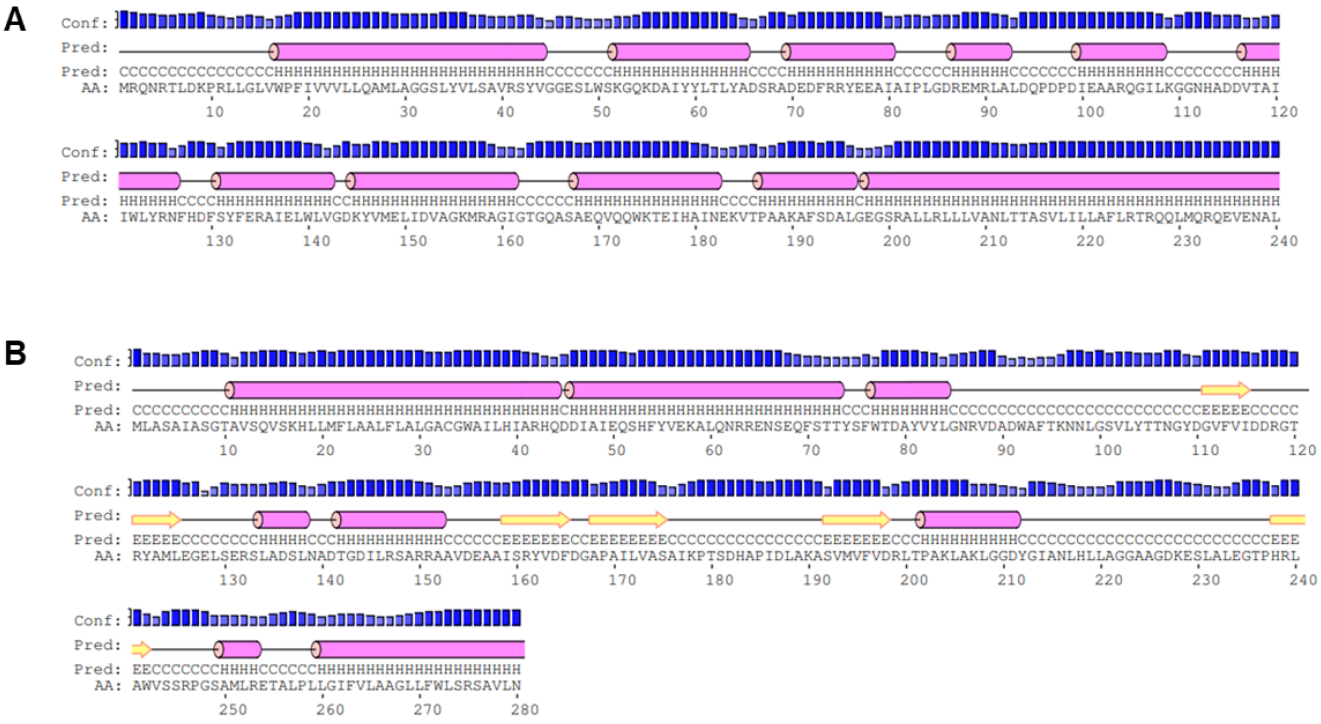

Figure 4.3. **The predicted secondary structures of the putative periplasmic domains of RbdA and PA2072**. A) Shows the predicted secondary structure of the periplasmic region (residues 45 to 197) of RbdA. B) Shows the predicted secondary structure of the putative periplasmic CHASE4 sensory domain (residues 45 to 254) of PA2072. The secondary structures were predicted with the PSIPRED online server[166,167]. Pink cylinders depict α-helices, yellow arrows depict β-strands and grey lines depict coil regions. Conf; Confidence of prediction (level depicted by the height of the blue rectangle), Pred; Prediction (represented in a cartoon and by letter; C – coil, H – α-helix, E – β-strand) and AA; Amino acid (the primary sequence of the protein targets).

The protein sequences of RbdA and PA2072 show a high level of similarity between their GGDEF and EAL domains, with the catalytic GGDEF, EAL and DDFGTGYSS motifs aligning (Figure 4.2). Due to the presence of the consensus GGDEF motif along with the EAL and DDFGTGYSS motifs in the PA2072 protein sequence, and previous homology modelling of the active sites, we predict PA2072 to have both DGC and PDE enzymatic activities respectively[130]. In RbdA, DGC and PDE enzymatic activities have been observed experimentally[129,131], suggesting that the opposing physiological roles between RbdA and PA2072 are not due to differences in enzymatic abilities.

4.2 – The putative sensory PAS domain of PA2072

We hypothesize that the biofilm dispersal-inducing stimulus, NO, is sensed by *P. aeruginosa via* binding to a haem-bound PAS domain. This is based on previous structural evidence that shows NO can bind to a haem cofactor within a PAS domain[122]. However, within *P. aeruginosa*, the NO sensor(s) responsible for the NO-induced biofilm dispersal remains unknown. We investigate the PAS domain of PA2072, due to the reduced NO-induced biofilm dispersal of PAO1 Δ*pa2072* mutant biofilms compared to WT PAO1 biofilms[130].

4.2.1 – The $PAS_{310-427}$ domain of PA2072 does not bind a cofactor

To determine whether or not the PAS domain of PA2072 binds a cofactor, the isolated $PAS_{310-427}$ domain of PA2072 was expressed and the protein purified (as described in Section 2.8 and 2.9). An example of PA2072 $PAS_{310-427}$ protein purification is shown in Figure 4.4. Protein expression was relatively high, 12 mg/mL in 350 µL using 2 mL SEC fractions 38, 39 and 40 (Figure 4.4 C) from a 2 L BL21 (DE3) cell expression. Using the SEC calibration curve (Figure 7.2), we find that the PA2072 $PAS_{310-427}$ domain (molecular weight 15,251.92 Da) is eluted from the SEC column as a monomer at 77.1 mL (as suggested from a calculated molecular weight of 18,518 Da). Upon purifying the PA2072 $PAS_{310-427}$ domain and carrying out UV-Vis spectroscopy, we did not observe any cofactor binding (Appendix B – Figure 7.4).

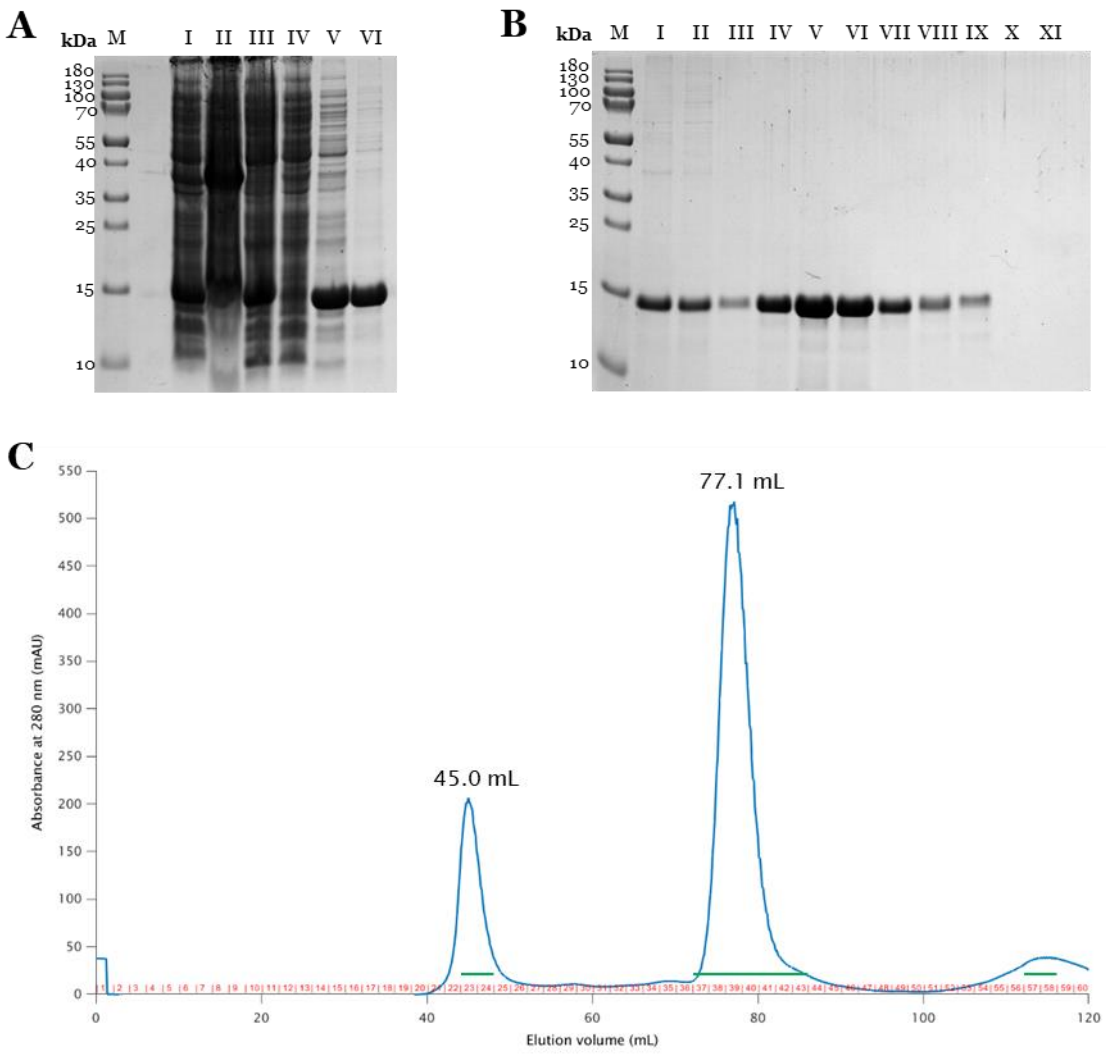

Figure 4.4. **Purification of PA2072 PAS$_{310\text{-}427}$ after expression with BL21 (DE3) cells.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA2072 PAS$_{310\text{-}427}$ (approximately 15 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA2072 PAS$_{310\text{-}427}$ IMAC eluate showing fractions 23, 24, 37-43, 57 and 58 in lanes I – XI respectively. C) SEC trace of PA2072 PAS$_{310\text{-}427}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

Protein crystals of PA2072 $PAS_{310-427}$ (Figure 4.5) were achieved using SEC fractions 38, 39 and 40 (Figure 4.4) when concentrated to 12 mg/mL and incubated at 21°C in the trial screen PACT *premier* (in conditions: 100 mM Succinic Acid, sodium phosphate monobasic monohydrate, Glycine (SPG) buffer, pH 4.0 and 25 % w/v PEG 1500 (well A1), 100 mM sodium malonate dibasic monohydrate, Imidazole, Boric acid (MIB) buffer, pH 4.0 and 25 % w/v PEG 1500 (well B1) and 100 mM DL-Malic acid, MES monohydrate, Tris (MMT) buffer, pH 4.0 and 25 % w/v PEG 1500 (well D1)).

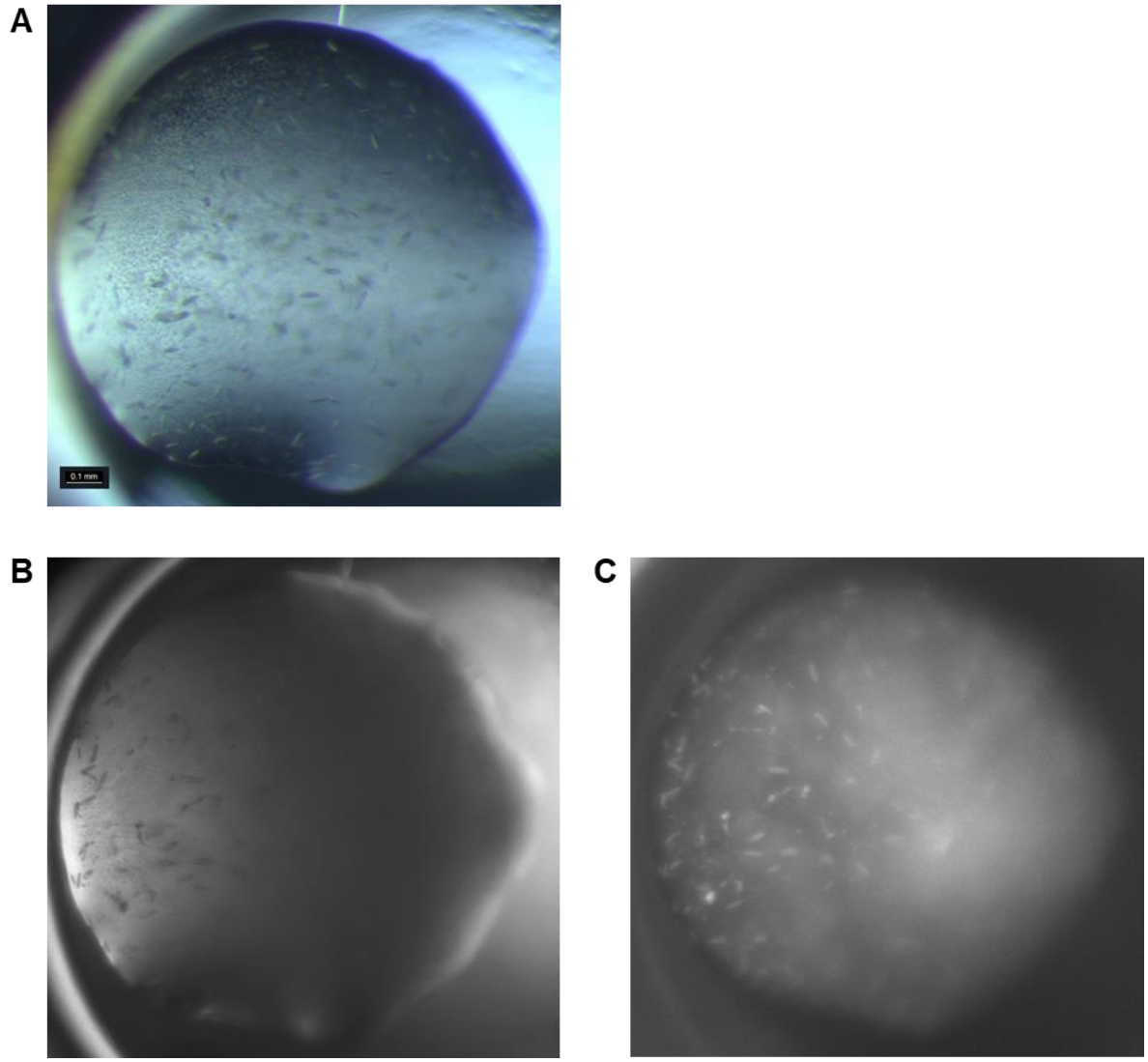

Figure 4.5. **The crystallisation of the PA2072 $PAS_{310-427}$ domain.** A) Protein crystals of the PA2072 $PAS_{310-427}$ domain grown in the trial screen PACT *premier* well B1 (100 mM sodium malonate dibasic monohydrate, Imidazole, Boric acid (MIB) buffer, pH 4.0 and 25 % w/v PEG 1500). The image was taken with a brightfield microscope (Leica) with a 0.1 mm scale bar. B) Brightfield image of the PA2072 $PAS_{310-427}$ protein crystals taken with an ultra-violate fluorescence microscope (JANSi UVEX) at a 5X magnification and with a 0.1 sec exposure. C) Protein crystals of the PA2072 $PAS_{310-427}$ domain (in the same position as in B)) taken under ultra-violet light with the JANSi UVEX microscope at a 5X magnification and with a 1 sec exposure.

In an attempt to obtain larger protein crystals of PA2072 $PAS_{310-427}$, a seeding experiment (as described in Section 2.13.2), using crystals grown in wells A1 and B1 of PACT *premier* with the trial screen Morpheus was carried out. Although there were protein crystals hits (Figure 4.6) in the seeded Morpheus screen, crystals were still small (approximately 50 nm). In order to test for diffraction, a Mesh LithoLoop (Molecular Dimensions) was used to retrieve the small PA2072 $PAS_{310-427}$ crystals from the seeded Morpheus screen, well H9 (100 mM Amino acids (20 mM L-Na-Glutamate; 20 mM Alanine (racemic); 20 mM Glycine; 20 mM Lysine HCl (racemic); 20 mM Serine (racemic)), 100 mM buffer system 3 (Tris (base); Bicine) at pH 8.5, 50 % v/v precipitant mix 1 (40% v/v PEG 500 MME; 20% w/v PEG 20000)). However, these crystals of PA2072 $PAS_{310-427}$ did not produce a diffraction pattern (as tested at the ESRF on beamline ID30A-3 on the 27/02/2018 using a mesh scan as an initial characterisation method of the mesh loop). Therefore the crystallisation, seeding experiment and/ or the diffraction experiment of PA2072 $PAS_{310-427}$ requires further optimisation (discussed in Section 6.2.2).

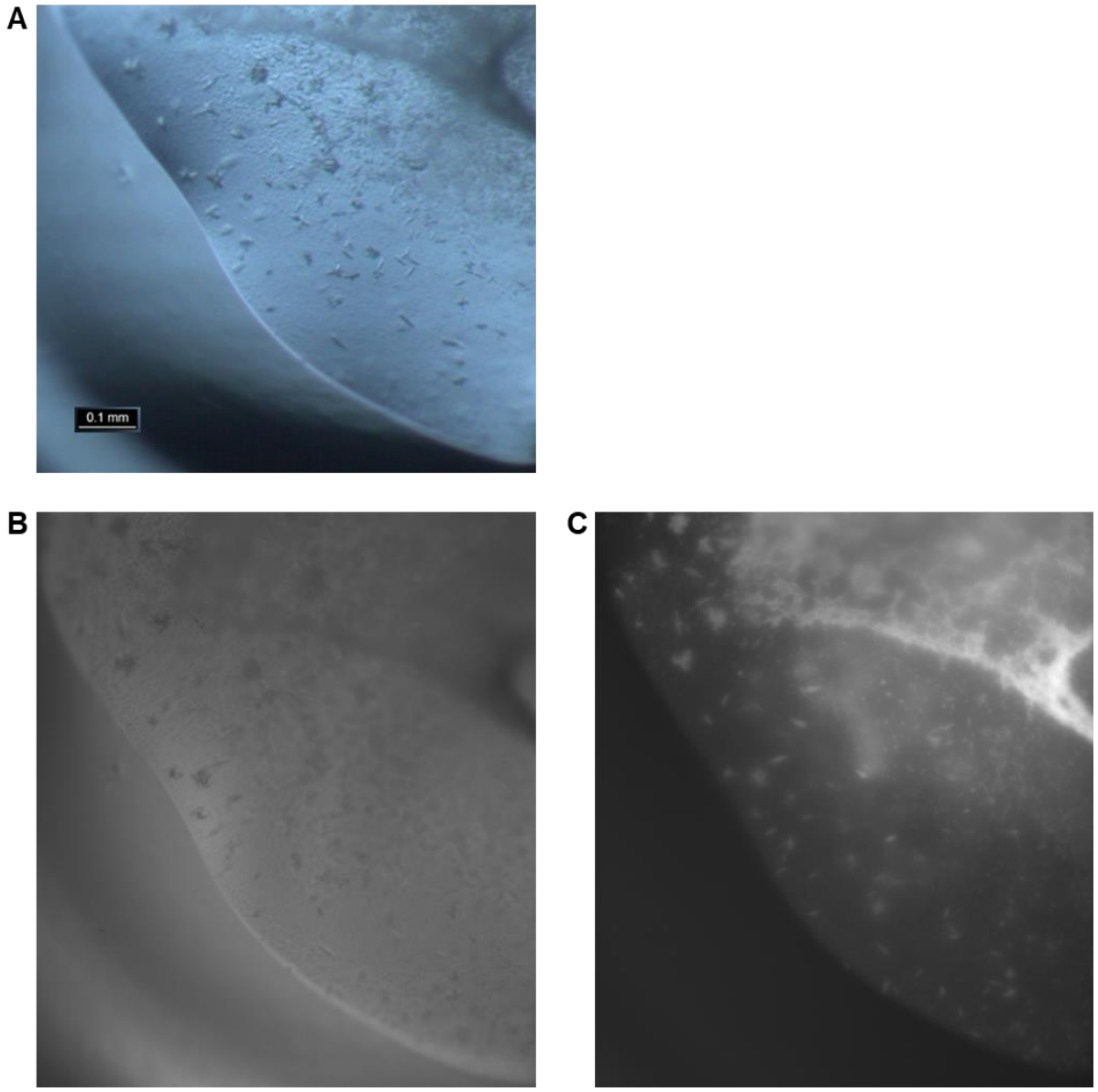

Figure 4.6. **Seeding of the PA2072 $PAS_{310-427}$ domain.** Protein crystals of the PA2072 $PAS_{310-427}$ domain obtained in the Morpheus screen (well H9), after seeding with an undiluted seed stock of PA2072 $PAS_{310-427}$ crystals taken from PACT *premier* wells A1 and B1. A) Image taken with a brightfield microscope (Leica) with an 0.1 mm scale bar. B) Brightfield image of PA2072 $PAS_{310-427}$ protein crystals taken with the ultra-violate fluorescence microscope (JANSi UVEX) at a 15X magnification and with a 0.1 sec exposure. C) Protein crystals of the PA2072 $PAS_{310-427}$ domain (in the same position as in B)) taken under ultra-violet light with the JANSi UVEX microscope at a 15X magnification and with a 1.5 sec exposure.

4.3 – PA2072 contains a putative CHASE4 sensory domain

As well as containing all the same protein domains as RbdA, PA2072 is predicted to also contain the sensory CHASE4 domain within the periplasmic region[133]. We sought to characterise the PA2072 CHASE4 domain due to its potential in being the cause for the physiological differences between RbdA and PA2072. Using the protein secondary structure prediction server, PSIPRED[166,167], and the domain boundaries predicted by InterPro[133,168] and current knowledge in the literature on CHASE4 domains[125-127], residues 45 to 254 of PA2072 were selected as appropriate boundaries to encompass the PA2072 CHASE4 domain. To date there are only two novel CHASE domain structures within the PDB[169,170], limiting our understanding of how these domains function and sense stimuli.

4.3.1 – Crystallisation of the PA2072 CHASE4$_{45\text{-}254}$ domain

To begin characterising the PA2072 CHASE4 domain, we attempted to crystallise the PA2072 CHASE4$_{45\text{-}254}$ domain for further use in X-ray crystallography and structure determination. There was a high level of protein expression (40 mg/mL in 150 µL when concentrated from 2 mL (SEC fraction 36) from a 2L culture) of the PA2072 CHASE4$_{45\text{-}254}$ domain in BL21 (DE3) cells. An example of the PA2072 CHASE4$_{45\text{-}254}$ purification is shown in Figure 4.7.

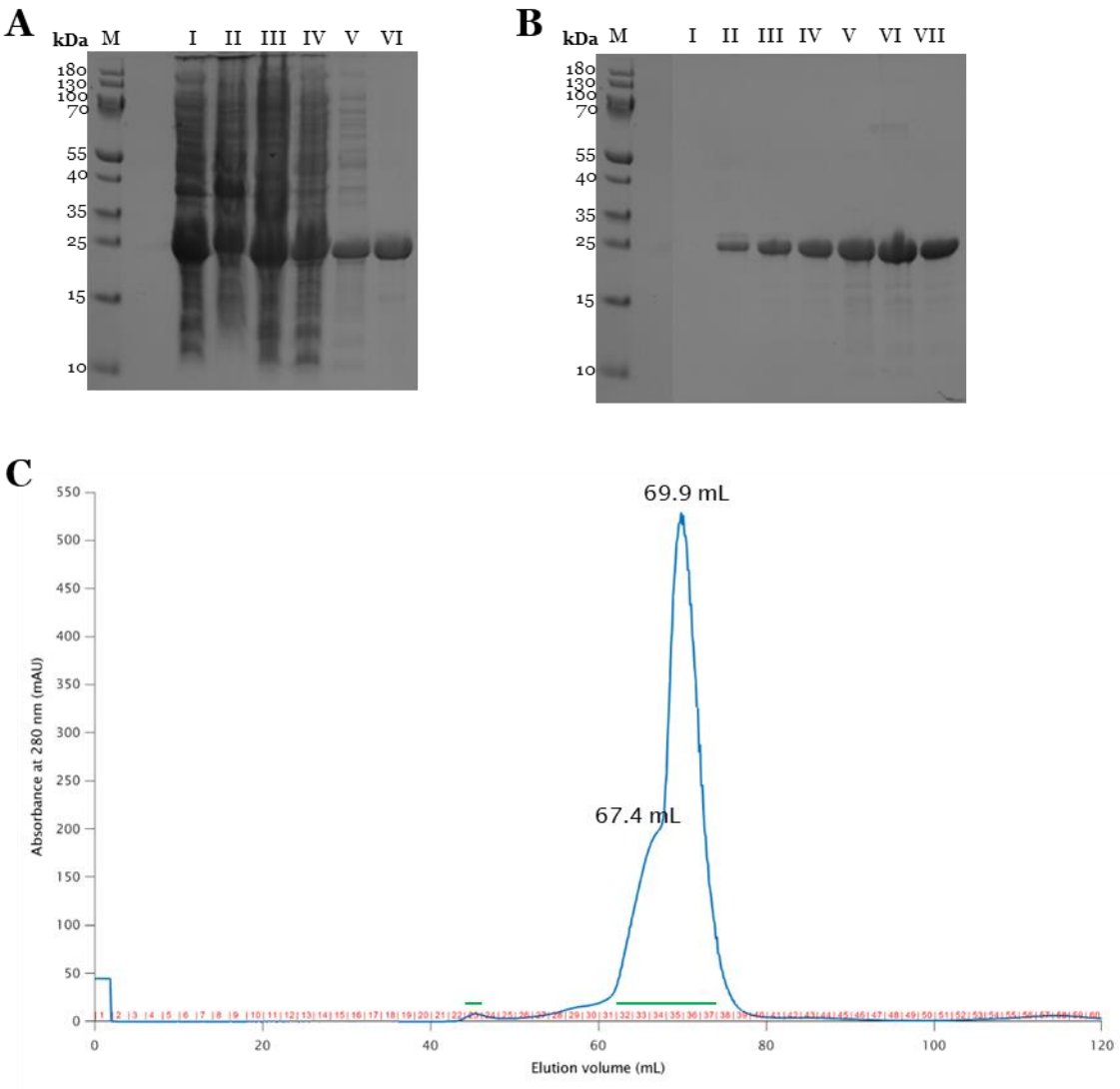

Figure 4.7. **Purification of PA2072 CHASE$_{45-254}$.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA2072 CHASE$_{45-254}$ (approximately 25 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA2072 CHASE$_{45-254}$ IMAC eluate showing fractions 23, 32 – 37 in lanes I – VII respectively. C) SEC trace of PA2072 CHASE$_{45-254}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

From the SEC trace (Figure 4.7 C) and SDS-PAGE gel of the SEC fractions (Figure 4.7 B), it can be observed that PA2072 CHASE4$_{45\text{-}254}$ mainly elutes at approximately 69.9 mL. This suggests that the PA2072 CHASE4$_{45\text{-}254}$ domain (molecular weight of 25,132.96 Da) mainly elutes as a monomer due to the calculated molecular weight of 30,438 Da (using the SEC calibration curve, Figure 7.2). It was also observed in some of the PA2072 CHASE4$_{45\text{-}254}$ SEC traces that there was a second peak (Figure 4.8 A) at approximately 63.3 mL which also produced a protein band on SDS-PAGE gels (Figure 4.8 B) corresponding to the size of PA2072 CHASE4$_{45\text{-}254}$. This suggests that PA2072 CHASE4$_{45\text{-}254}$ can also form a dimer (calculated molecular weight of 47,703 Da), eluted at approximately 63.3 mL.

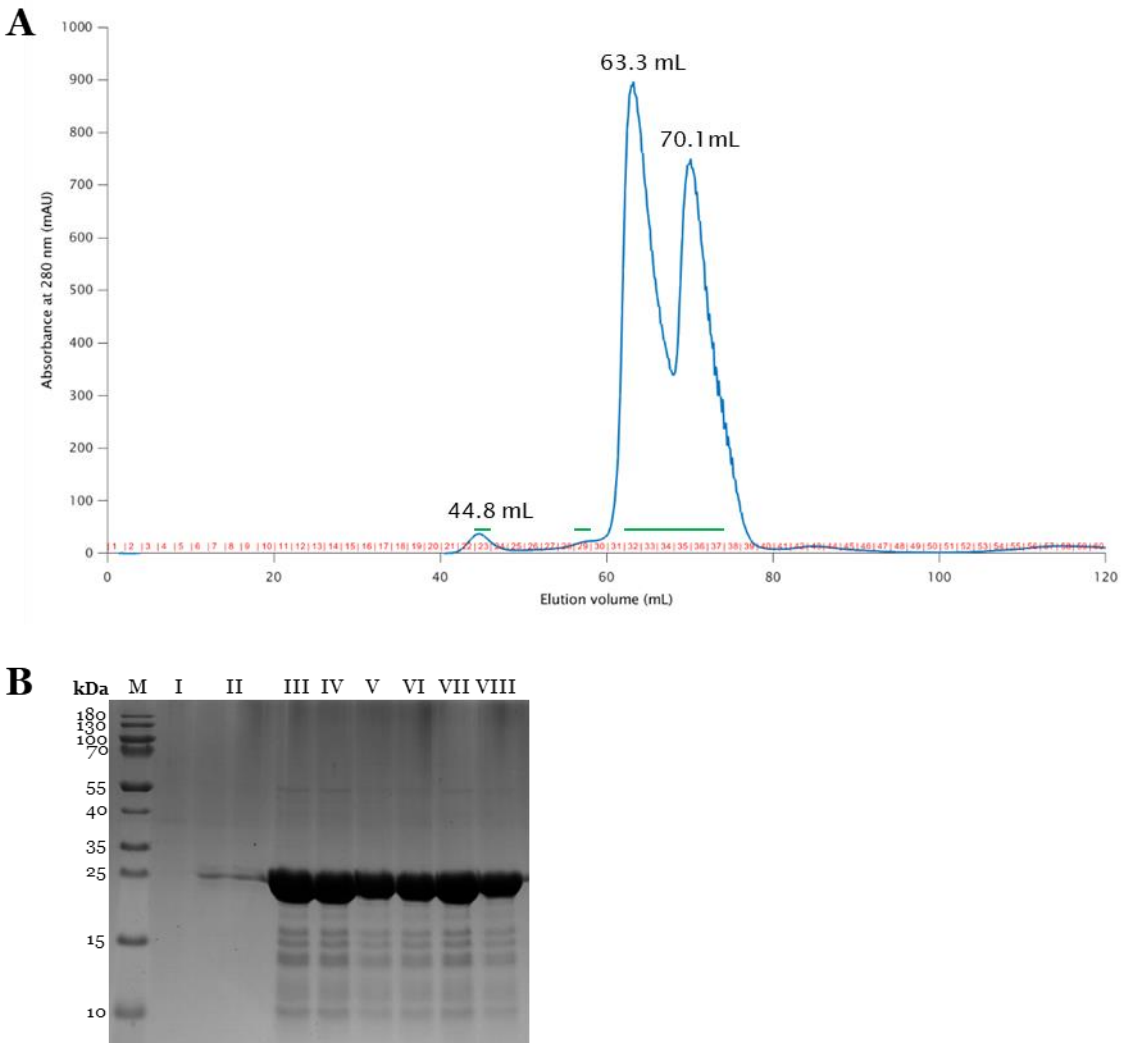

Figure 4.8. **Purification of monomeric and dimeric PA2072 CHASE$_{45-254}$.** A) SEC trace of PA2072 CHASE$_{45-254}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B). B) Sodium dodecyl sulphate polyacrylamide gel electrophoresis of PA2072 CHASE$_{45-254}$ (approximately 25 kDa) after size exclusion chromatography of the PA2072 CHASE$_{45-254}$ immobilised metal affinity chromatography eluate showing fractions 23, 29, 32 – 37 in lanes I – VIII respectively. Lane M contains the molecular weight marker PageRuler prestained protein ladder.

PA2072 CHASE4$_{45-254}$ protein crystal hits were achieved in the trial screen Morpheus (wells A1, B1 and C1), an example is shown in Figure 4.9. However, due to the PA2072 CHASE4$_{45-254}$ protein crystals overlapping and fragile nature, these crystals were unable to be fished for diffraction experiments. Furthermore, a custom screen based on crystallisation conditions that gave crystal hits (Appendix A; Table 7.2) and seeding experiments (described in Section

2.13.2) failed to produce protein crystals. Therefore, the crystallisation and seeding experiments for PA2072 CHASE4$_{45-254}$ require further optimisation (as discussed in Section 6.2.2).

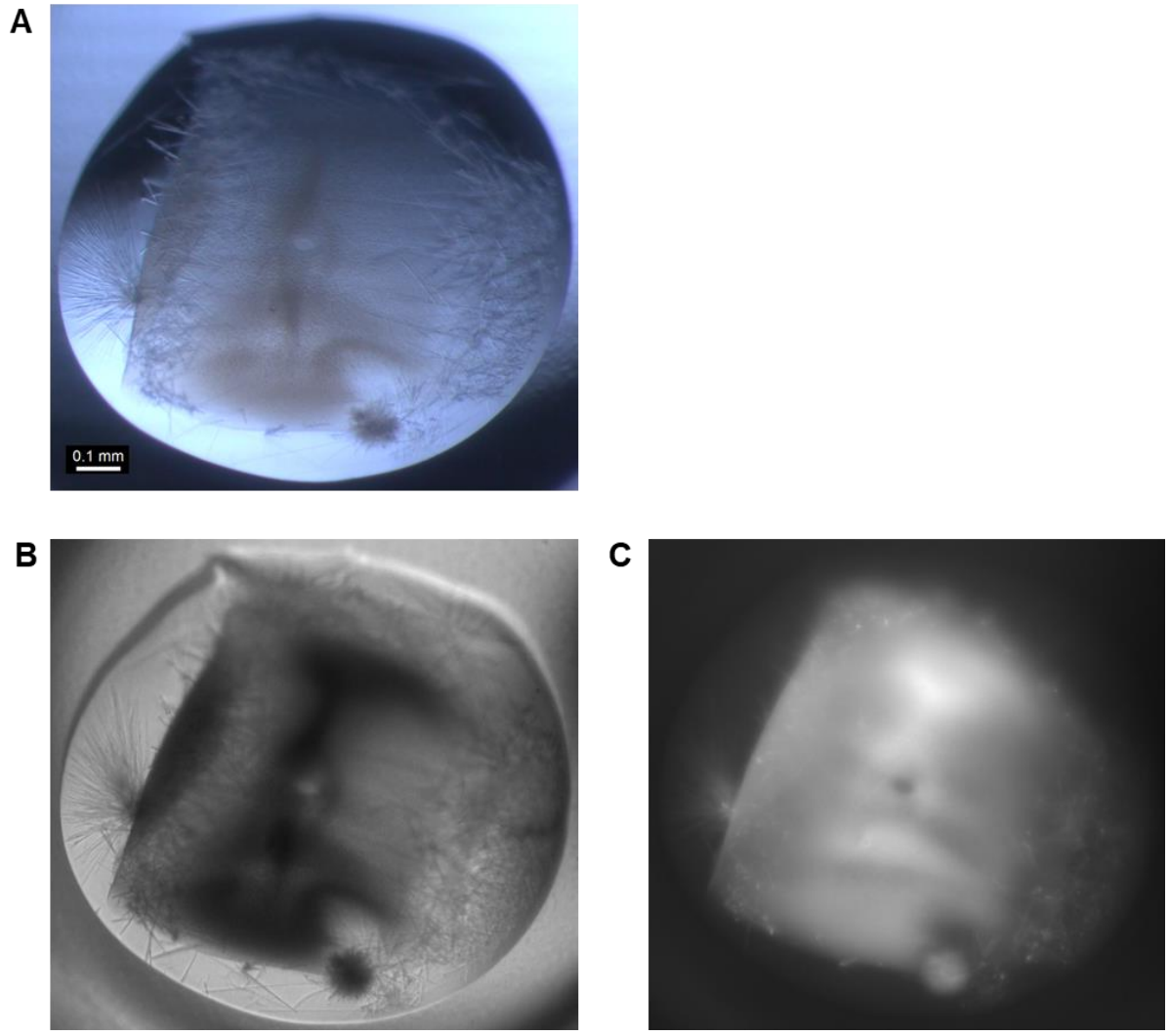

Figure 4.9. **The crystallisation of the PA2072 CHASE$_{45-254}$ domain.** Protein crystals of the PA2072 CHASE$_{45-254}$ domain from the Morpheus screen, well C1. A) Image taken with a brightfield microscope (Leica) with a 0.1 mm scale bar. B) Brightfield image of PA2072 CHASE$_{45-254}$ protein crystals taken with the ultra-violate fluorescence microscope (JANSi UVEX) at a 15X magnification and with a 0.1 sec exposure. C) Protein crystals of the PA2072 CHASE$_{45-254}$ domain (in the same position as in B)) taken under ultra-violet light with the JANSi UVEX microscope at a 15X magnification and with a 1 sec exposure.

4.4 – Conclusion

Here, we investigated PA2072 which, despite having a homologous protein domain architecture, has some opposing physiological roles compared to RbdA[130]. By comparing the primary sequences and the predicted secondary structures of RbdA and PA2072 (Figure 4.2 and Figure 4.3 respectively), we suggest that the physiological differences may be due to the structural differences between the putative periplasmic sensor domains, as well as differences between the PAS putative sensory domains. The periplasmic region of PA2072 has been predicted to contain a CHASE4 sensory domain[127], while the periplasmic region of RbdA has not[133] (Figure 4.1).

Using the PA2072 $PAS_{310-427}$ and $CHASE4_{45-254}$ constructs, we sought to further characterise these two putative sensor domains with crystallographic structure determination. Whilst protein expression and purification resulted in high yields of protein (Figures 4.4 and 4.7), we were not able to produce diffracting protein crystals of the PA2072 $PAS_{310-427}$ domain or the PA2072 $CHASE4_{45-254}$ domain (Figures 4.6 and 4.9 respectively).

Chapter 5 – Cofactor binding in the PAS domains of PA0285

Work carried out leading to our recent publication[130] also identified the *P. aeruginosa* PAO1 gene, *pa0285,* to be involved in the NO stimulated biofilm dispersal pathway. This was highlighted by an increased dispersal of Δ*pa0285* KO mutant biofilms in response to an NO stimulus, compared to the dispersal of PAO1 WT biofilms. It was also found that there were increased levels of intracellular c-di-GMP in planktonic Δ*pa0285* mutants compared to PAO1 WT cells. The *pa0285* gene encodes for a 760 amino acid long hypothetical protein (referred to as PA0285) that is predicted, by SMART[132] and InterPro[133], to include two PAS domains, a GGDEF domain and an EAL domain (depicted in Figure 5.1). We predicted both DGC and PDE enzymatic activities for PA0285 based on the conservation of consensus motifs required for enzyme activity and active site homology modelling[130]. This suggests that the PA0285 protein could regulate the NO induced biofilm dispersal via the intracellular levels of c-di-GMP.

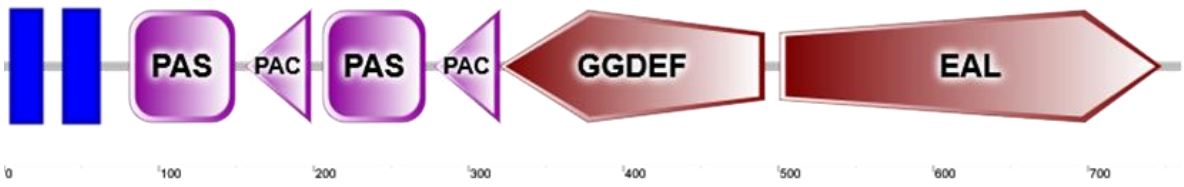

Figure 5.1. **Protein domain organisation of PA0285**. PA0285 is made up of 760 amino acids and is predicted by SMART[132] to be formed of two transmembrane regions (shown as vertical blue rectangles), two PAS domains (with each PAS domain indicated by both the PAS (pink square) and PAC (pink triangle)), a GGDEF domain and an EAL domain (each shown as a red pentagon). Grey lines indicate regions of unknown. Figures acquired from the online protein domain prediction server SMART[132].

Previously in the Dr I. Tews group (University of Southampton), the protein sequences of PAS domains with known cofactor binding were aligned and compared to try and identify sets of conserved coordinating amino acids that would indicate the binding of a specific cofactor[134]. Using this method, PAS domains within *P. aeruginosa* PAO1 with unknown cofactor binding were then predicted to bind certain cofactors, based on the conserved amino acids involved in coordinating a particular cofactor being present within the protein sequence. The first PAS

domain of PA0285 was predicted to bind haem-*b* and the second PAS domain of PA0285 was predicted to bind FAD[134].

As we hypothesize that the NO stimulus, which induces biofilm dispersal, is sensed by a cofactor bound PAS domain, in this chapter our aim was to characterise the cofactor binding of both the PA0285 PAS domains. We found only a very weak haem binding in the first PAS domain of PA0285 ($PAS1_{88-216}$). The second PAS domain of PA0285 ($PAS2_{213-333}$) bound a cofactor that was identified as FAD.

5.1 – Investigating the PAS domains of PA0285

Based on the previous phenotypic data of Δ*pa0285* biofilms[130] and the PAS domain sequence analysis[134], we hypothesized that NO is sensed by the PAS domains of PA0285 *via* bound cofactors, namely haem-*b* (in PAS1) and FAD (in PAS2). In order to test this hypothesis we use UV-Vis spectroscopy (Section 2.12) with PA0285 PAS domain fragments to determine the presence or absence of cofactor binding.

5.1.1 – Potential cofactor binding in the first PAS domain of PA0285

To determine whether the prediction that the PA0285 $PAS1_{88-216}$ domain of PA0285 could bind a haem-*b* cofactor, the isolated PA0285 $PAS1_{88-216}$ domain fragment was produced. The PA0285 $PAS1_{88-216}$ domain was expressed in and purified from BL21 (DE3) cells with an additional pHPEX-3 plasmid (see Section 2.8.2). This additional pHPEX-3 plasmid was used in an attempt to ensure that the cofactor haem would be available to the PA0285 $PAS1_{88-216}$ domain during the expression and folding of the protein. An example of the PA0285 $PAS1_{88-216}$ domain protein purification when expressed in BL21 (DE3) pHPEX-3 cells is shown in Figure 5.2.

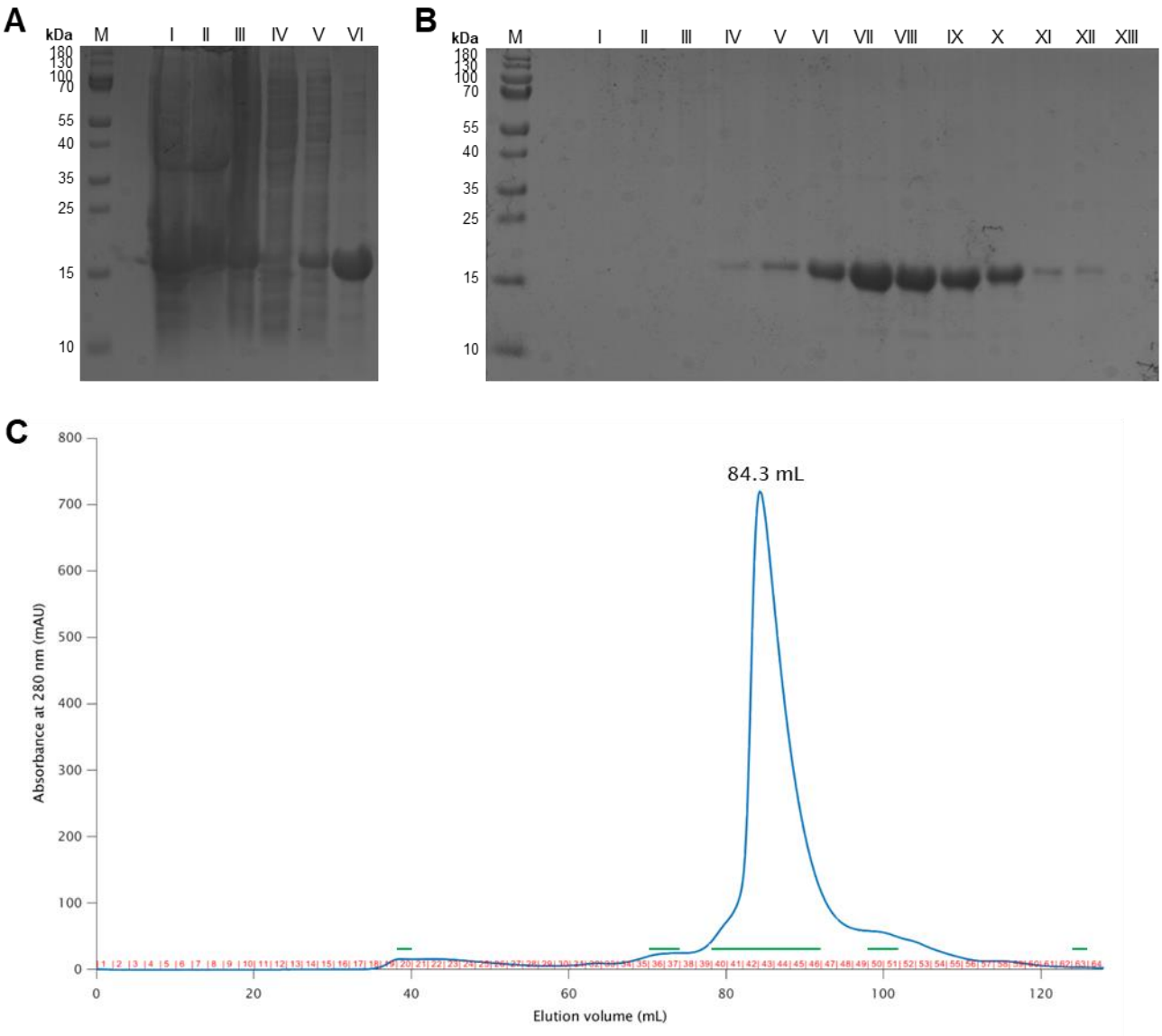

Figure 5.2. **Purification of PA0285 PAS1$_{88-216}$ after expression with BL21 (DE3) pHPEX-3 cells.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA0285 PAS1$_{88-216}$ (approximately 17 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA0285 PAS1$_{88-216}$ IMAC eluate showing fractions 20, 36, 37, 40 – 46, 50, 51, 63 in lanes I – XIII respectively. C) SEC trace of PA0285 PAS1$_{88-216}$ measuring absorbance at 280 nm when run at 0.2 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

Using the SEC column calibration curve (Appendix A; Figure 7.2) and the SEC trace for PA0285 $PAS1_{88-216}$ shown in Figure 5.2, the PA0285 $PAS1_{88-216}$ domain (molecular weight 17,298.43 Da) is suggested to be mainly eluted from the SEC column as a monomer at approximately 84.3 mL with a calculated molecular weight of 11,421 Da.

The UV-Vis spectrum of the purified PA0285 $PAS1_{88-216}$ domain was measured as described in Section 2.12 (Figure 5.3). A UV-Vis spectrum of a fraction from the SEC column that contained no protein (as observed in the SEC trace and SDS-PAGE gel; Figure 5.2) was taken (shown in Figure 5.3 B), to ensure that a haem peak produced in the PA0285 $PAS1_{88-216}$ protein sample was not due to haem contaminant coming from the SEC column from previous experiments. From the UV-Vis spectrum of PA0285 $PAS1_{88-216}$, it can be observed that there is a slight haem peak (insert in Figure 5.3). However this haem peak would account for only 0.048 % of the PA0285 $PAS1_{88-216}$ protein binding to haem (if the protein to haem binding ratio was 1 to 1), using the extinction coefficient values of 29,450 M^{-1} cm^{-1} at 280 nm for PA0285 $PAS1_{88-216}$[171] and 58,400 M^{-1} cm^{-1} at 410 nm for hemin[172]. Therefore, it is unlikely that the PA0285 $PAS1_{88-216}$ domain binds a haem-*b* cofactor and the observed haem peak may represent residual hemin from the purification process (discussed further in Section 6.3.2).

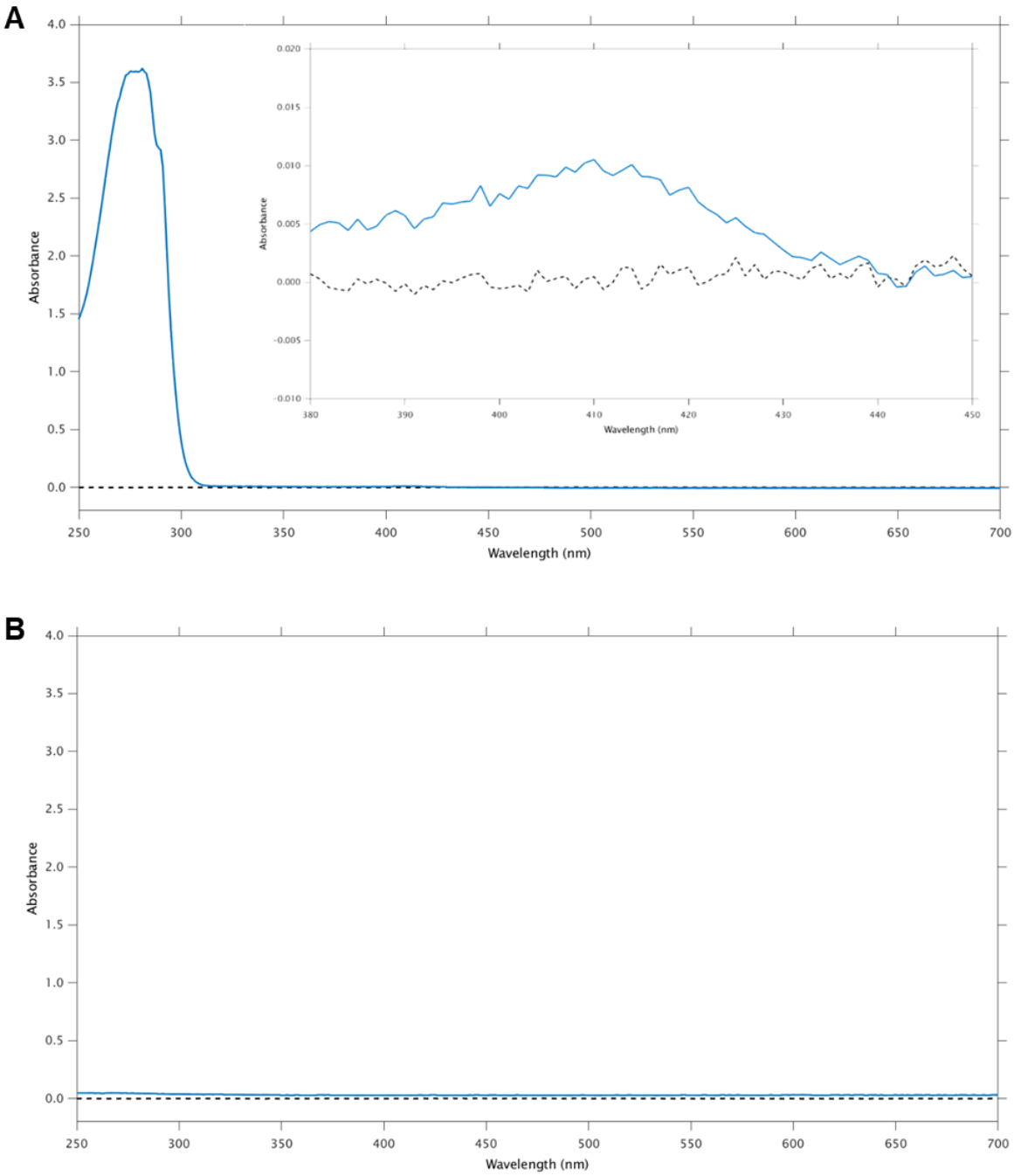

Figure 5.3. **UV-Vis spectrum of the PA0285 PAS1$_{88-216}$ domain after expression in BL21 (DE3) pHPEX-3 cells.** The ultraviolet- visible light spectrum of the isolated PA0285 PAS1$_{88-216}$ domain was measured between 250 - 700 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 8.0) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of PA0285 PAS1$_{88-216}$ was measured in the same way. A) The UV-Vis spectrum of PA0285 PAS1$_{88-216}$ (blue solid trace) with the baseline (black dashed trace). The insert shows a magnified image of the UV-Vis spectrum between 380 – 450 nm. B) The UV-Vis spectrum of size exclusion chromatography fraction 63 (blue solid trace) with the baseline (black dashed trace).

5.1.2 – The second PAS domain of PA0285 binds the cofactor FAD

To determine whether the prediction that the PA0285 PAS2$_{213\text{-}333}$ domain of PA0285 could bind the cofactor FAD, the isolated PA0285 PAS2$_{213\text{-}333}$ domain fragment was produced. An example of the PA0285 PAS2$_{213\text{-}333}$ domain protein purification when expressed in BL21 (DE3) cells is shown in Figure 5.4.

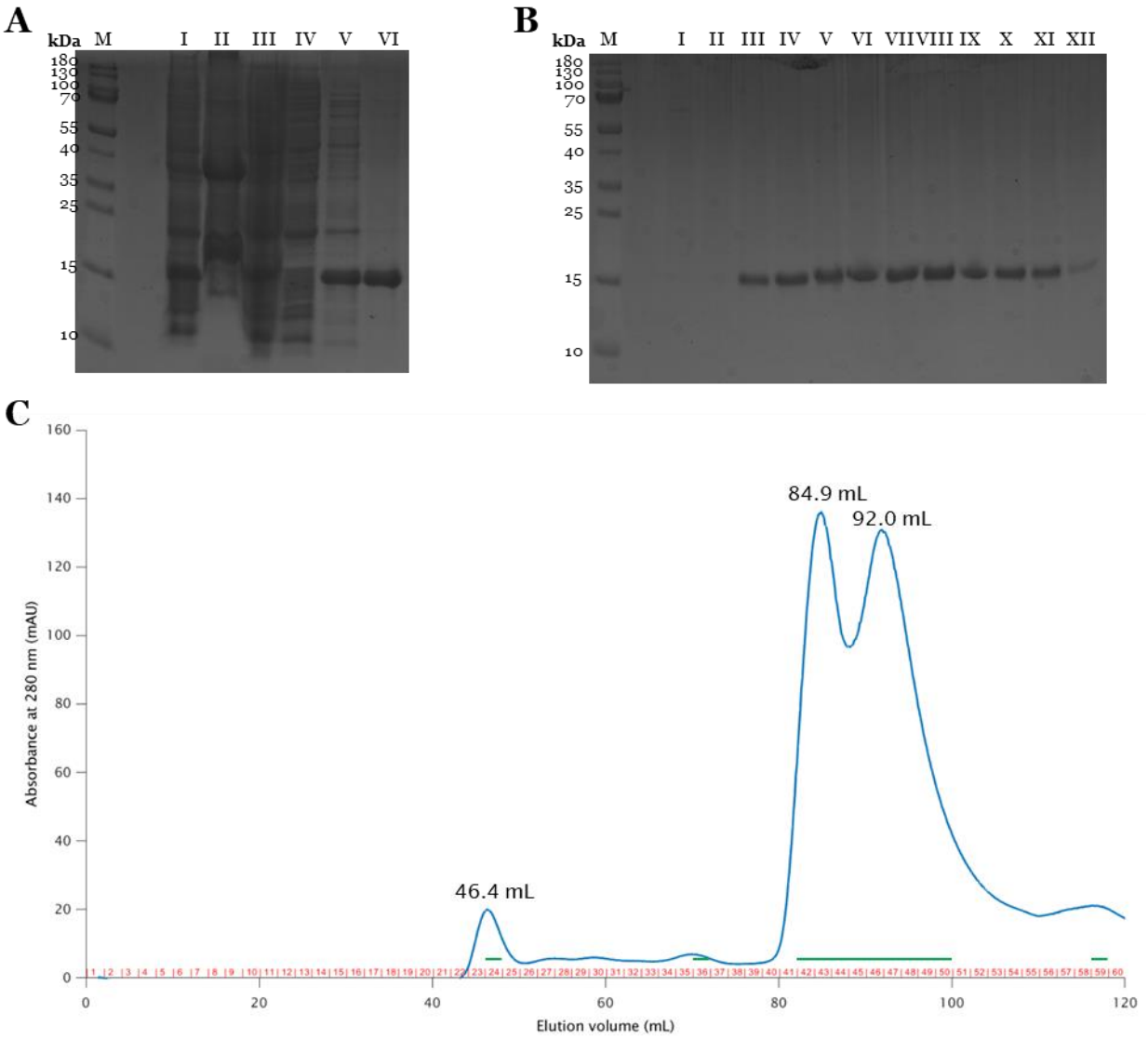

Figure 5.4. **Purification of PA0285 PAS2$_{213-333}$.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA0285 PAS2$_{213-333}$ (approximately 16 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA0285 PAS2$_{213-333}$ IMAC eluate showing fractions 24, 36, 42 – 50, 59 in lanes I – XII respectively. C) SEC trace of PA0285 PAS2$_{213-333}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

Using the SEC column calibration curve (Appendix A; Figure 7.2), the PA0285 $PAS2_{213-333}$ domain (molecular weight 16,328.19 Da) is mainly eluted from the SEC column at approximately 84.9 mL with a calculated molecular weight of 10,963 Da. It was observed that the SEC fractions (42, 43 and 44) under this protein peak were yellow in colour, indicating the presence of a flavin cofactor, FAD or FMN, which are yellow in colour[173]. A second large protein peak at approximately 92.0 mL (Figure 5.4 C), which was also found to contain a protein band indicative of the PA0285 $PAS2_{213-333}$ domain (Figure 5.4 B), has a calculated molecular weight of 6,761 Da. However, the SEC column eluates at this second peak (SEC fractions 46 to 48 inclusive) were colourless. This suggests that the PA0285 $PAS2_{213-333}$ domain elutes with a flavin cofactor as a dimer at 84.9 mL and as a monomer without a bound flavin cofactor at 92.0 mL.

Using the SEC fraction which was yellow in colour and contained the purified PA0285 $PAS2_{213-333}$ domain (SEC fraction 43), the UV-Vis spectrum was measured (Figure 5.5). From the UV-Vis spectrum of PA0285 $PAS2_{213-333}$, the cofactor FAD was identified[174] (Figure 5.5 B). Using Beer-Lambert Law, the extinction coefficient values of 27,960 M^{-1} cm^{-1} at 280 nm for PA0285 $PAS2_{213-333}$[171] with high and low values for the protein bound FAD cofactor of 14,100 M^{-1} cm^{-1} at 452 nm and 10,600 M^{-1} cm^{-1} at 448 nm[175], the PA0285 $PAS2_{213-333}$ to FAD stoichiometry was estimated. It was found that the PA0285 $PAS2_{213-333}$ domain bound an FAD cofactor in a protein : FAD stoichiometry between 1.9 : 1 and 2.7 : 1.

Conversely, from the UV-Vis spectrum of the PA0285 $PAS2_{213-333}$ domain purified in SEC fractions 46 to 48, it was found that this portion of the PA0285 $PAS2_{213-333}$ protein was not bound to the FAD cofactor (Figure 5.6).

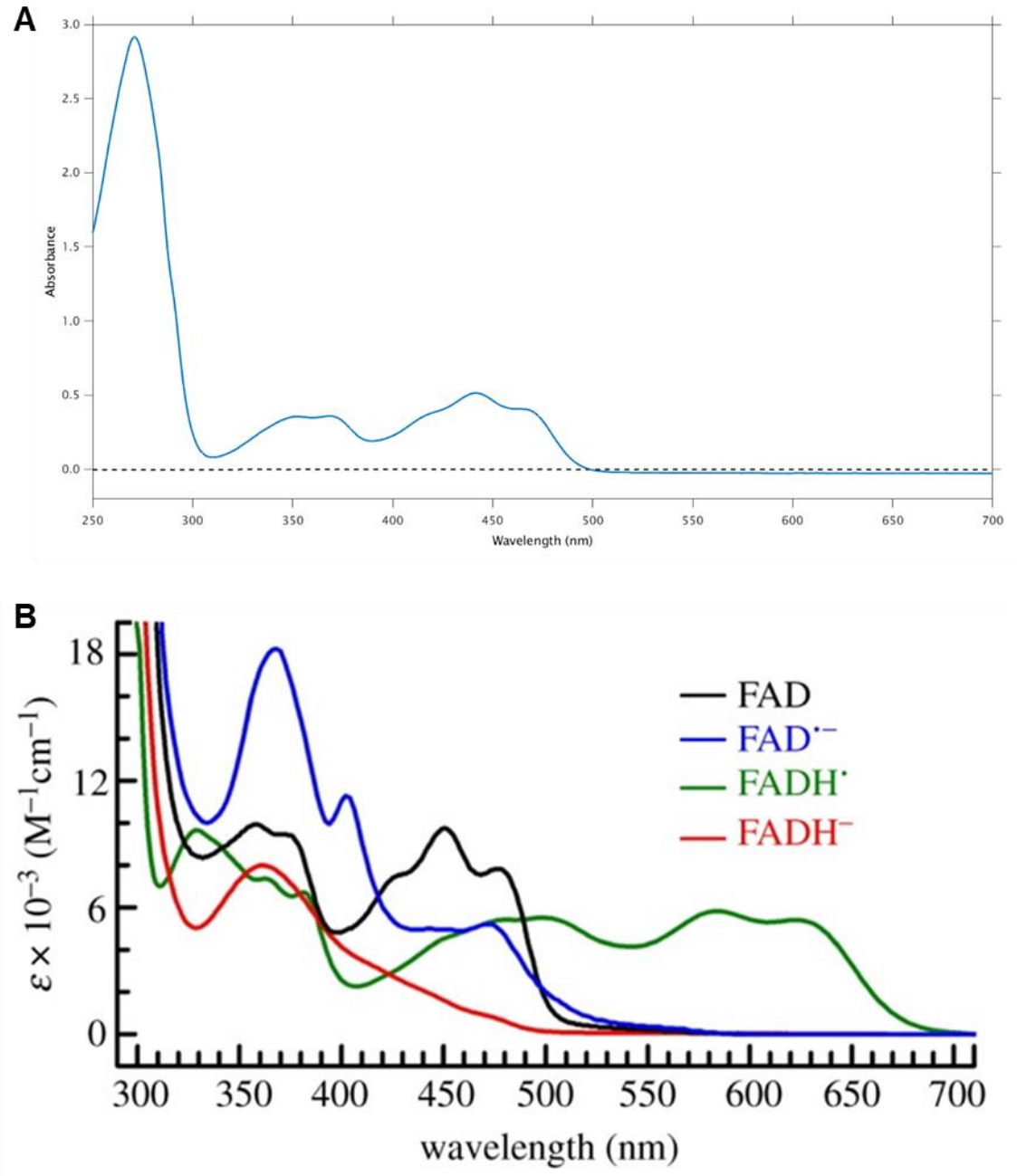

Figure 5.5. **UV-Vis spectrum showing FAD cofactor binding in the PA0285 PAS2$_{213-333}$ domain.** The ultraviolet- visible light spectrum of the isolated PA0285 PAS2$_{213-333}$ domain was measured between 250 - 700 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 8.0) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of PA0285 PAS2$_{213-333}$ was measured in the same way. A) The UV-Vis spectrum of PA0285 PAS2$_{213-333}$ (blue solid trace) with the baseline (black dashed trace). B) The absorption spectrums of the different species of FAD cofactor, taken from E. Evans et al., 2013[174].

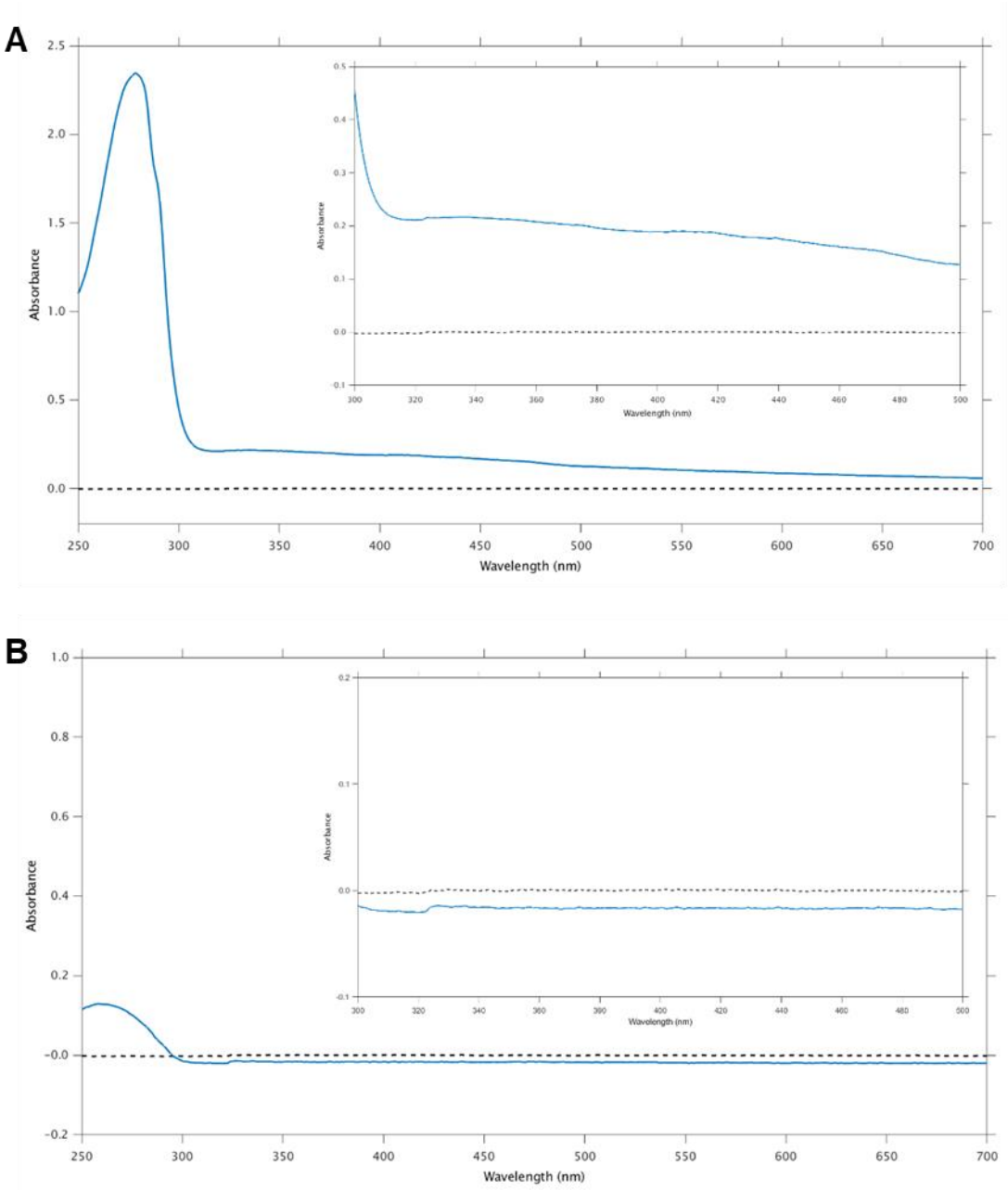

Figure 5.6. **UV-Vis spectrum showing an absence of FAD cofactor binding in the PA0285 PAS2$_{213-333}$ domain**. The ultraviolet- visible light spectrum of the colourless portion of the isolated PA0285 PAS2$_{213-333}$ domain was measured between 250 - 700 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 8.0) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of colourless PA0285 PAS2$_{213-333}$ protein was measured in the same way. A) The UV-Vis spectrum of PA0285 PAS2$_{213-333}$ (blue solid trace) using size exclusion chromatography (SEC) fractions 46 to 48 inclusive, with the baseline (black dashed trace). B) The UV-Vis spectrum of SEC fraction 60 (blue solid trace) with the baseline (black dashed trace). The inserts show a magnified image of the UV-Vis spectrum between 300 – 500 nm.

A PA0285 PAS2-GGDEF-EAL$_{213-760}$ protein fragment was also produced and purified from BL21 (DE3) cells, to determine whether FAD binding to the PAS2 domain could also be achieved in a larger protein construct (Figure 5.7). PA0285 PAS2-GGDEF-EAL$_{213-760}$ (molecular weight 63,955.28 Da) is mainly eluted from the SEC column as a monomer at 59.5 mL (as suggested from the calculated molecular weight of 61,787 Da). The eluted SEC fraction 30 was found to be yellow in colour and the FAD cofactor was found to bind the PA0285 PAS2-GGDEF-EAL$_{213-760}$ fragment (using SEC fraction 30) as evidenced by the UV-Vis spectrum (shown in Figure 5.8 A). Using the same extinction coefficients for protein bound FAD[175] as used above and 63,370 M^{-1} cm^{-1} at 280 nm for PA0285 PAS2-GGDEF-EAL$_{213-760}$[171], the protein : FAD stoichiometry was estimated to between 2.4 : 1 and 3.4 : 1. Another protein peak at 52.5 mL (SEC fraction 27), was also found to contain the PA0285 PAS2-GGDEF-EAL$_{213-760}$ fragment (Figure 5.7 B), however this eluate was colourless and was found not to bind the FAD cofactor (Figure 5.8 B).

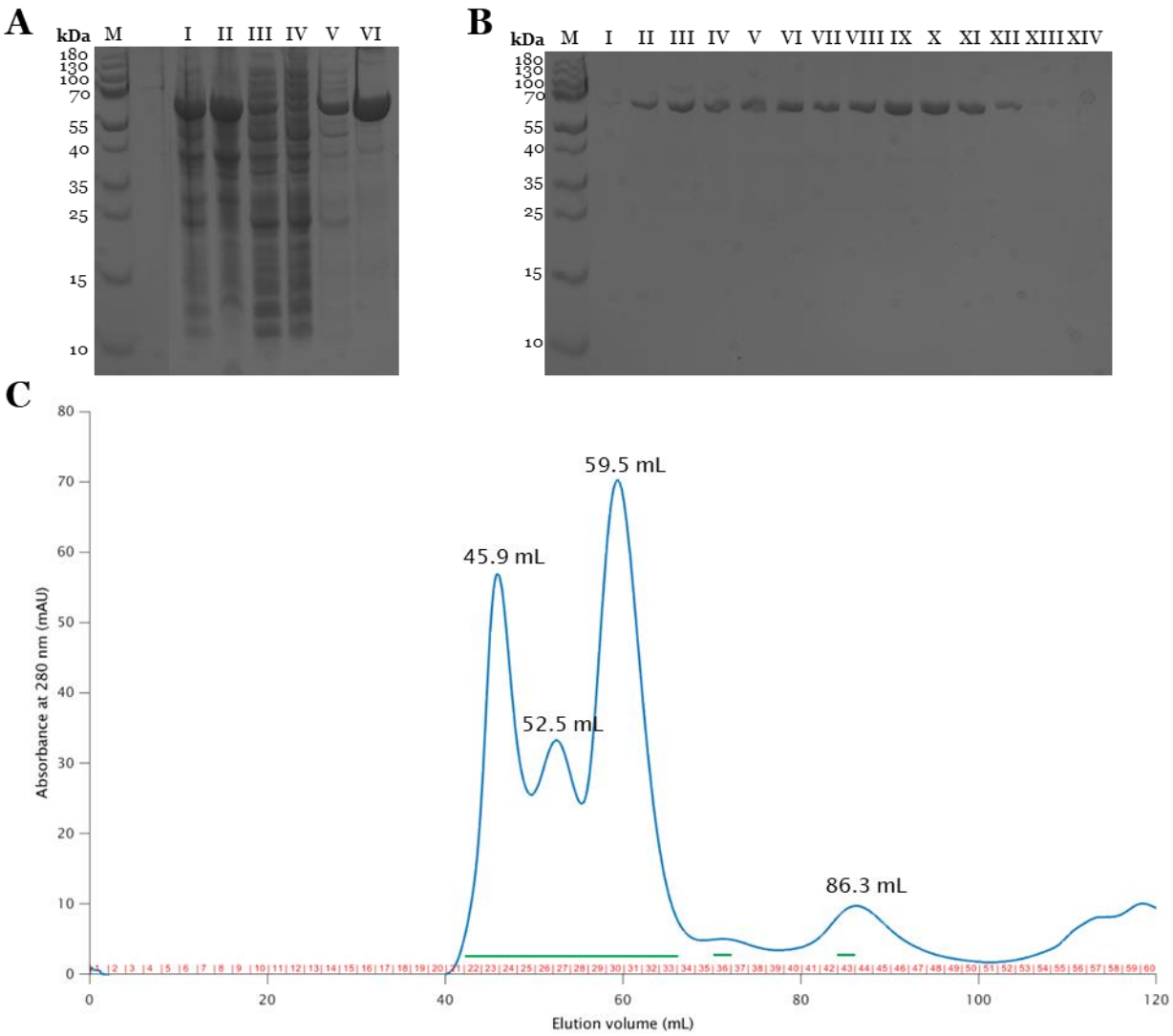

Figure 5.7. **Purification of PA0285 PAS2-GGDEF-EAL$_{213-760}$.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA0285 PAS2-GGDEF-EAL$_{213-760}$ (approximately 64 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA0285 PAS2-GGDEF-EAL$_{213-760}$ IMAC eluate showing fractions 22 – 33, 36, 43 in lanes I – XIV respectively. C) SEC trace of PA0285 PAS2-GGDEF-EAL$_{213-760}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red. Green lines are shown above the fractions that were loaded onto the SDS-PAGE gel in B).

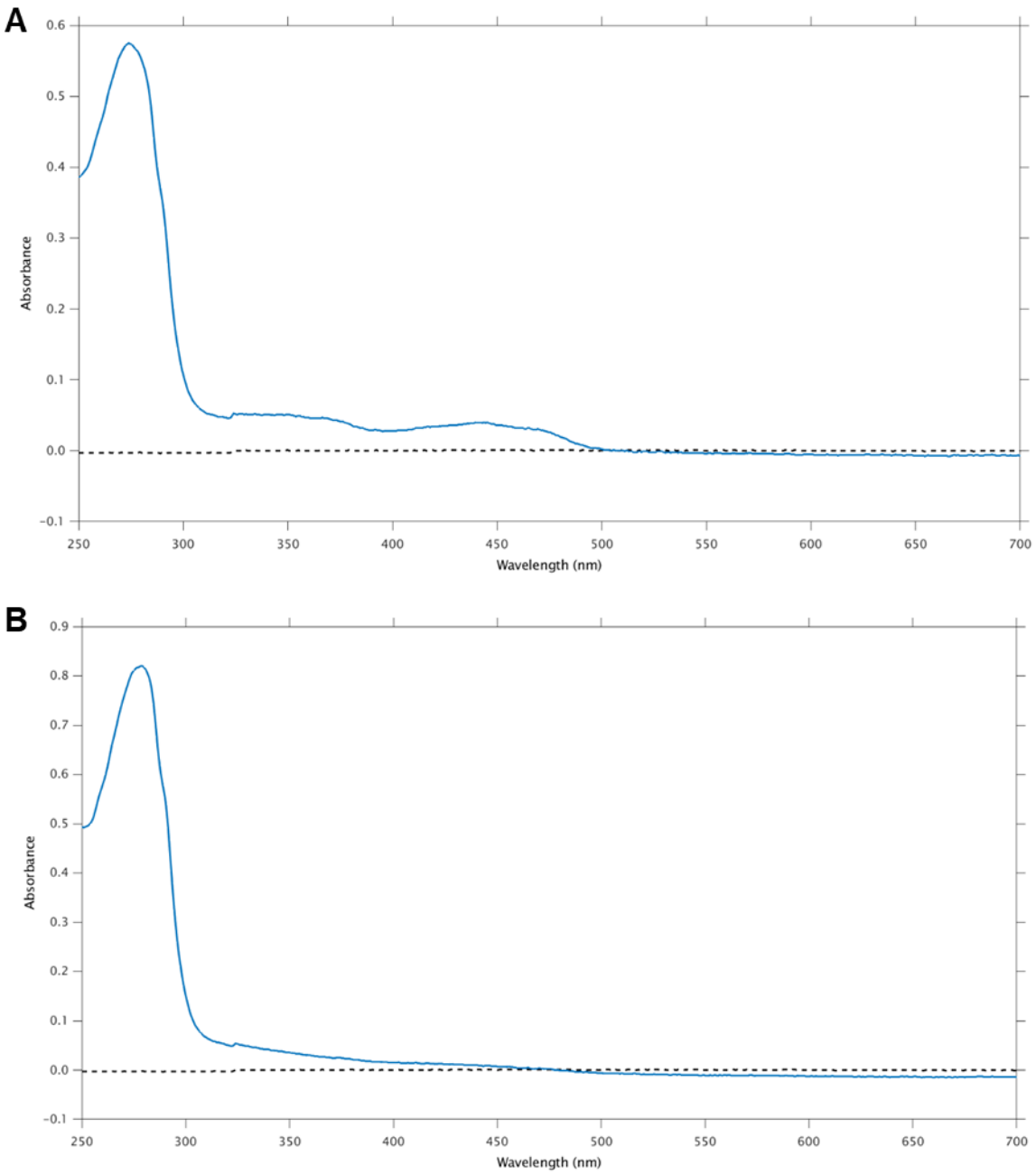

Figure 5.8. **UV-Vis spectrum of the PA0285 PAS2-GGDEF-EAL$_{213-760}$ fragment.** The ultraviolet-visible light spectrum of the PA0285 PAS2-GGDEF-EAL$_{213-760}$ protein fragment was measured between 250 - 700 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 7.5) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of PA0285 PAS2-GGDEF-EAL$_{213-760}$ was measured in the same way. A) The UV-Vis spectrum of PA0285 PAS2-GGDEF-EAL$_{213-760}$ size exclusion chromatography (SEC) fraction 30 (blue solid trace) with the baseline (black dashed trace). B) The UV-Vis spectrum of SEC fraction 27 (blue solid trace) with the baseline (black dashed trace).

5.2 – Conclusion

To investigate whether PA0285 could sense the biofilm dispersal inducing NO stimulus, we sought to characterise the two putative sensory PAS domains. Previous sequence analysis of *P. aeruginosa* PAS domains, investigating conserved sequence motifs that might indicate certain cofactor binding, predicted that the first PAS domain of PA0285 bound a haem-*b* cofactor, while the second PAS domain of PA0285 was predicted to bind an FAD cofactor[134]. Whilst we found a weak haem-*b* binding in the PA0285 $PAS1_{88-216}$ domain, this would only amount to a 0.048 % occupancy if the haem : protein stoichiometry was assumed to be 1 : 1 (Figure 5.3). However we found, in agreement with the PAS-cofactor sequence analysis, the second PAS domain of PA0285 ($PAS2_{213-333}$) bound an FAD cofactor in a stoichiometry between 1.9 : 1 and 2.7 : 1 (Figure 5.5). The FAD cofactor binding could also be achieved in the larger PA0285 PAS2-GGDEF-$EAL_{213-760}$ fragment in a protein : FAD stoichiometry between 2.4 : 1 and 3.4 : 1 (Figure 5.8). In both cases, FAD cofactor binding was obtained without the addition of FAD or FAD derivatives to the growth media or during the protein purification process, validating the FAD cofactor as the genuine cofactor for the PA0285 PAS2 domain.

Chapter 6 – Discussion

In this study we investigated three proteins, RbdA, PA2072 and PA0285, as each were previously highlighted as being involved in a *P. aeruginosa* NO-induced biofilm dispersal[129-131] (Table 6.1). We have shown that the PDE activity of an EAL domain may be negatively regulated by a tandem GGDEF domain, in contrast to previous observations[73,164]. While we were unable to elucidate the molecular mechanism for this negative regulation, we have taken the first step towards its understanding by determining the crystallographic structure of the RbdA $EAL_{549-797}$ domain. Understanding this molecular switch could reveal how two active enzymatic domains with opposing functions are controlled while in tandem so that there is only one output at any given time. We further propose a scheme for an EAL domains PDE activity based on the current structural data available.

PA2072, a protein with a homologous protein domain architecture to RbdA, was investigated to try to determine the reason, on a molecular level, for their opposing physiological roles. While sequence and predicted secondary structure comparisons highlighted the periplasmic regions and / or PAS domains between the two proteins to be a possible cause for the opposing physiological roles, we were not successful in producing diffracting protein crystals for these domains in PA2072. Revealing the molecular mechanisms that lead to the opposing physiological roles of RbdA and PA2072, would aid our understanding of the molecular switches involved in regulating a proteins opposing enzymatic activities.

During this work investigating the two PAS domains of PA0285, that we hypothesise sense NO, we found that the second PAS domain binds a flavin cofactor which we suggest is FAD. Approximately two PA0285 PAS2 domains bind one flavin cofactor. In contrast, a negligible amount of the haem-*b* cofactor was found to be associated with the first PAS domain of PA0285. As putative sensor domains, understanding these PAS domains and the cofactors associated with them may reveal how PA0285 senses NO. Understanding how PA0285 senses NO and the conformational changes that occur within PA0285 upon NO stimulation, could allow for the development of small molecules which inhibit PA0285 and therefore enhance the NO-induced biofilm dispersal using combination therapies. Each of the proteins investigated in this work are discussed in turn, with suggestions of further experiments that could be carried out to improve our understanding.

Table 6.1. **A summary of the phenotypic data previously highlighting RbdA, PA2072 and PA0285 as being involved in an NO-induced biofilm dispersal in *P. aeruginosa*.**

	Phenotype compared to a PAO1 wild type (WT)										
KO mutant	Microcolony size	Biofilm dispersal in response to NO	Biofilm Biomass	Biofilm Surface coverage	Intracellular levels of c-di-GMP	Total protein in EPS	Total poly-sacchrides in EPS	Swimming	Twitching	Swarming	Effect of NO to increase swarming area
Δ*rbdA*	↑	↓	–	–	↑	↑	↑	↓	–	↓	↓
Δ*pa2072*	↓	↓	↓	↓	↓	–	–	–	–	–	↓
Δ*pa0285*	↓	↑	↓	↓	↑	↑	↑	↓	X	–	↑

↓ : a statistically significant reduction in the mutants' phenotypic trait compared to a PAO1 WT.

↑ : a statistically significant increase in the mutants' phenotypic trait compared to a PAO1 WT.

– : no statistically significant difference in the mutants' phenotypic trait compared to a PAO1 WT.

X : a complete absence of the phenotypic trait in the mutant compared to a PAO1 WT.

6.1 – RbdA

6.1.1 – The EAL domain of RbdA

Recently, three structures of the cytosolic region of RbdA (cRbdA) consisting of the PAS-GGDEF-EAL domains were determined, each forming a dimer that was suggested to be held in an auto-inhibited conformation (PDB codes 5XGB, 5XGE and 5XGD)[131]. As PDE and DGC activities were determined for cRbdA this would suggest that cRbdA can adopt different dimeric arrangements to allow for these enzymatic activities, possibly due to the H-helix acting as a hinge[73,131]. Consistent with this, SAXS data of cRbdA suggested at least two different conformations, with cRbdA forming either an extended conformation or a compact conformation which could be induced by the presence of GTP or GMP with c-di-GMP[131]. Based on this SAXS data, the crystallographic structures of cRbdA and the finding that the addition of GTP[129] and the non-hydrolysable GTP analogue, GMPPNP[131], could enhance the PDE activity of RbdA through binding to the GGDEF domains active site, a model for the PDE activity of RbdA was proposed. The model suggests that signal detection, potentially the binding of GTP to the active half-sites of the GGDEF domains, would cause local conformational changes that would propagate through RbdA and cause a release of the auto-inhibited state, resulting in a rearrangement of the EAL domains into a canonical EAL dimer that would poses PDE activity[131].

The dimerisation of EAL domains is known to be required for PDE activity[81], with dimerisation suggested to induce a shortening of the α5 – helix and an extension of the β5 – α5 loop allowing the catalytic Asp residues of the DDFGTG motif to enter the active site for metal ion coordination[73]. Consistent with this, in the RbdA $EAL_{549-797}$ 'classic' dimer presented here, the α5 – helix is formed of 6 residues and the β5 – α5 loop is formed of 9 residues, corresponding to a shortened α5 – helix and an extended β5 – α5 loop. In contrast the cRbdA structures have an elongated α5 – helix formed of 9 residues and a shortened β5 – α5 loop formed of 6 residues. However, unlike in MorA[73] and PA3825[80], even with the elongated α5 – helix and shortened β5 – α5 loop in the cRbdA structures, the catalytic Asp residues of the DDFGTG motif are still located within the active site (Figure 6.1). This observation is inconsistent with the previous model in which an extended α5 – helix and shortened β5 – α5 loop would result

in the removal of the Asp residues of the DDFGTG motif from the active site. It may suggest that the cRbdA structures are not in an auto-inhibited conformation in terms of PDE activity, with the catalytic Asp residues maintained within the active site, or perhaps the cRbdA structures reflect a transition state between an inactive and active PDE state. It should also be noted that, although structural rearrangements would be required for the formation of a canonical EAL dimer in cRbdA, other EAL dimer interfaces may still promote PDE activity due to the repositioning of the Asp residues of the DDFGTG motif, as is observed for PA3825[80]. Furthermore, as cRbdA also poses DGC activity, it is possible that a rearrangement, resulting from local conformation changes potentially induced by a signal, leads to the anti-parallel dimerisation of the GTP binding sites of the two GGDEF domains necessary for DGC activity. However this later possibility requires further investigation, with the GGDEF domain of RbdA suggested to function predominately as a GTP sensor rather than in c-di-GMP synthesis[131].

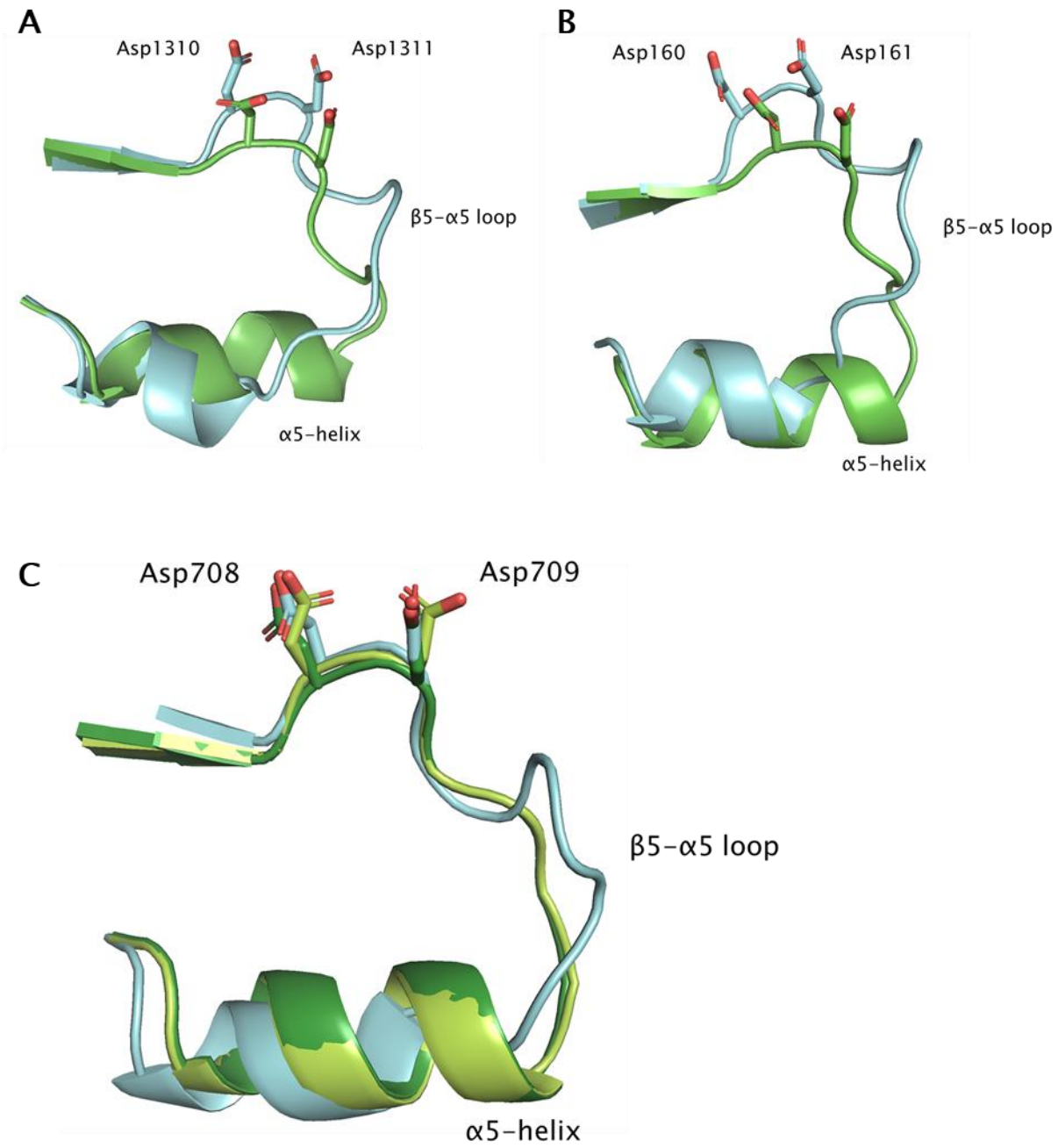

Figure 6.1. **The extended α5 – helix of RbdA does not remove catalytic Asp residues from the active site.** The side chains of catalytic Asp residues of the DDFGTG motif located on the β5 – α5 loop are shown as sticks. A) Overlay of MorA EAL domains with an extended α5 – helix (PDB code 4RNJ; shown in green) and a shortened α5 – helix (PDB code 4RNH; shown in cyan). B) Overlay of PA3825 EAL domains with an extended α5-helix (PDB code 4Y9M; shown in green) and a shortened α5 – helix (PDB code 5MF5; shown in cyan). C) Overlay of RbdA $EAL_{549-797}$ (shown in cyan with a shortened α5 – helix) and the RbdA PAS-GGDEF-EAL (cRbdA) structures in the apo-form (PDB code 5XGB; shown in dark green) and c-di-GMP bound form (PDB code 5XGE; shown in light green). Even with the α5 – helix extended in the cRbdA structures, the catalytic Asp residues are not removed from the active site.

The beginning of the N-terminal helix (α1 – helix) of the RbdA $EAL_{549-797}$ structure forms the C-terminus of the H-helix which connects the GGDEF and EAL domains together. When superposing the RbdA $EAL_{549-797}$ structure with the cRbdA structures, the N-terminal helix of RbdA $EAL_{549-797}$ and the H-helix of the cRbdA structures superpose well. In contrast, when superposing RbdA $EAL_{549-797}$ with the GGDEF-EAL tandem structures of MorA (PDB code 4RNH), LapD (PDB code 3PJX) and PA0575 (PDB code 5M3C) the N-terminal helix of RbdA $EAL_{549-797}$ and the H-helix of these GGDEF-EAL tandem structures do not overlay and do not follow the same direction. This indicates that if the GGDEF domain were present in the RbdA $EAL_{549-797}$ structure (as in the RbdA GGDEF-$EAL_{376-797}$ construct), it may sit in a similar position to the GGDEF domain in the cRbdA structure (i.e. the RbdA GGDEF-$EAL_{376-797}$ construct may form a closed conformation as in the cRbdA structures). However, the N-terminal helix of RbdA $EAL_{549-797}$ would be relatively unconstrained without the tandem GGDEF domain. Therefore the direction of the N-terminal helix in RbdA $EAL_{549-797}$ may not reflect the true position of the GGDEF domain in relation to the EAL domain in the RbdA GGDEF-$EAL_{376-797}$ construct. Furthermore the N-terminal helix of RbdA EAL is shorter than the H-helix which is able to hinge. The hinging of the H-helix in the RbdA GGDEF-$EAL_{376-797}$ construct may allow the GGDEF domain to occupy a different position relative to the EAL domain to that observed for cRbdA. Additionally, in solution proteins are not rigid and the relative positions of the GGDEF and EAL domains are likely to change.

6.1.2 – Further work required to characterise the regulation between the GGDEF and EAL domains of RbdA

The regulation between the GGDEF and EAL domains of RbdA is still not fully characterised or understood. The DGC activities of RbdA $GGDEF_{376-536}$ and RbdA GGDEF-$EAL_{376-797}$ could be measured to determine whether the EAL domain of RbdA can influence the DGC activity of the GGDEF-EAL tandem. However, this may require the production of a different RbdA GGDEF construct to overcome the precipitation of the RbdA $GGDEF_{376-536}$ construct. The use of a different N-terminal tag, such as thioredoxin or maltose-binding protein, instead of the hexa-His tag used here, may increase the solubility of this RbdA $GGDEF_{376-536}$ construct[176]. Alternatively the purification buffers could be amended to stabilise the RbdA $GGDEF_{376-536}$ construct.

It is worth noting that different crystal morphologies for RbdA $EAL_{549-797}$ were observed and these may represent different protein contacts and possibly different nucleotide bound states of RbdA $EAL_{549-797}$. However some of these crystal morphologies (needle morphology) only diffracted to a low resolution. Therefore these low resolution diffracting crystals could be further optimised using seeding experiments. In seeding experiments, protein crystals are crushed to form many small crystals (or seeds), these seeds can then be used to nucleate a crystallisation event when the seeds are added to a solution of protein and crystallisation condition, promoting crystal growth in the metastable zone. Although attempts were also made at co-crystallising RbdA $EAL_{549-797}$ with c-di-GMP or pGpG, as well as crystal nucleotide soaks, these were unsuccessful, indicating that these protocols require further optimisation.

Co-crystallisation of RbdA $EAL_{549-797}$ was attempted by setting up hanging drops with a limited number of precipitant conditions. Although crystals of RbdA $EAL_{549-797}$ were observed, the resulting structures were nucleotide-free. Trakhanov and Quiocho[177] reported that in order to obtain high quality diffracting crystals of leucine/ isoleucine/ valine-binding protein (LIVBP) and leucine-specific binding protein (LBP) in the presence of leucine, the use of divalent cations (cadmium being optimal) was required. Furthermore, the use of different divalent cations produced variations in the crystal morphology and diffraction quality of LIVBP in the presence of leucine[177]. It has been found that divalent metal ions, located on the protein surface, promote protein crystal formation by mediating intermolecular interactions within

the crystal lattice[177,178]. The initial crystallisation screens, Morpheus, JCSG$^+$ and PACT *premier*, each utilise (in varying degrees) different divalent metal ions. Therefore, the RbdA $EAL_{549-797}$ co-crystallisation experiments would benefit from using these broad 96-condition crystallisation screens set up in sitting drops. It may also be necessary to optimise the concentration of the nucleotides being added (the final nucleotide concentrations trialled here for co-crystallisation were 2 mM c-di-GMP or 200 µM pGpG) and the length of time the nucleotides are incubated with the protein before the crystallisation condition is added (in these hanging drop experiments the nucleotides were added after the protein and crystallisation condition had been combined). The use of different metal cations (such as inhibitory calcium ions) may also be necessary to obtain a nucleotide bound state of RbdA $EAL_{549-797}$.

Due to the solvent channels that run throughout the lattice of the crystals and the position of the active site in EAL proteins, a nucleotide (c-di-GMP or pGpG) should be able to access the active site of our RbdA $EAL_{549-797}$ protein during soaking experiments. When nucleotide soaks were carried out, damage and cracking of the crystals was observed. This may represent conformational changes within the protein that are required as the nucleotide binds or it may represent non-optimal soaking conditions (such as soaking time before cryo-cooling or the soaking buffers being used). During protein crystal soaking small conformational changes are usually well tolerated[179]. As only slight conformational or rotameric changes are required for our RbdA $EAL_{549-797}$ structure to be receptive to nucleotide binding, it is likely that the damage to the crystals, observed during the nucleotide soaks, is the result of non-optimal soaking conditions. Interestingly, protein crystals of octopine dehydrogenase with bound NADH were observed to crack within the first few minutes after the crystals were soaked with ligands (L-arginine or pyruvate). However, within a few hours the cracked crystals recovered and damage to the crystals could no longer be observed[180]. Though these soaked crystals diffracted to a lower resolution than crystals which were not soaked, the ligand bound structures were determined[180,181].

Therefore, the RbdA $EAL_{549-797}$ soaking experiment requires further optimisation in terms of the soaking buffer composition and nucleotide concentrations being used (the soaking buffers trialled in our experiment were: 2.5 mM c-di-GMP, 1 mM $MgCl_2$ and 0.5 mM $MnCl_2$; 2.5 mM c-di-GMP, 0.5 mM $CaCl_2$ and 0.5 mM $MgCl_2$; 775 µM pGpG, 1 mM $MgCl_2$ and 0.5 mM

$MnCl_2$) and the soaking time[182] before the cryo-buffer is added and the crystal cryo-cooled (soaking times trialled here were between 1 – 20 mins). As a cryo-buffer was applied after soaking the crystal, it is possible that the nucleotide and/ or metal ions were soaked back out of the crystal. Hence, it may be necessary to also add the nucleotide and/ or the metal ions that are present in the soaking buffer to the cryo-buffer to help prevent this[179]. Further, in order to maintain the protein crystals' diffraction, an increase in the nucleotide concentration in a slow and stepwise manner may be required[179]. In order for Liu et al.[131] to obtain $RbdA_{c\text{-}di\text{-}GMP}$ (PDB code 5XGE), native crystals were transferred to a fresh reservoir solution supplemented with 1 mM c-di-GMP and 2 mM $MgCl_2$, soaking the crystals for 5 mins. The authors observed that longer soak times led to crystal breakage and to a deterioration of the diffraction quality[131]. Therefore, an increase in the metal ion concentration in our soaking buffers may be required in order for the nucleotide to bind into the active sites of the RbdA $EAL_{549\text{-}797}$ proteins within our crystals.

Seeding experiments or co-crystallisation with a nucleotide may also promote the crystallisation of RbdA GGDEF-$EAL_{376\text{-}797}$. Variation in the length of the RbdA GGDEF-$EAL_{376\text{-}797}$ construct at the N- or C-terminus may also be necessary to promote crystallisation. Utilising the cleavage site to remove the N-terminal histidine tag, used for protein purification during IMAC, from the recombinant protein by a protease (thrombin) before protein crystallisation, may also be required[176]. The removal of the histidine tag from a protein can promote protein crystallisation, as the histidine tag is flexible and can therefore prevent the formation of an ordered crystal lattice. Furthermore, due to its flexibility, the histidine tag is often disordered and not observed within the model of the protein structure[183]. However, by comparing the structures of His-tagged and non-tagged proteins of identical sequence (except for a reasonable extension at either the N- or C-terminus), the His-tag was suggested to have little to no effect on the native structure of the protein[183]. In some cases, the His-tag may even promote protein crystallisation[184]. The nucleotide bound states of RbdA $EAL_{549\text{-}797}$ and the structure of RbdA GGDEF-$EAL_{376\text{-}797}$ may provide information on the molecular mechanism for the increased PDE activity of the isolated $EAL_{549\text{-}797}$ domain compared to the GGDEF-$EAL_{376\text{-}797}$ tandem. Further insight and validation of the putative M3 and/ or M4 metal sites may also be provided with pGpG bound EAL domains of RbdA.

6.2 – PA2072

6.2.1 – The homologous proteins RbdA and PA2072

RbdA and PA2072 have homologous protein domain architectures (Figures 4.1 and 4.2), except that PA2072 seems to have an extra periplasmic domain, the CHASE4 domain. Currently, little is understood about CHASE domains. It is suggested that the CHASE domains are involved in the developmental program of an organism[126], and it is unclear whether this would extend to biofilm formation and dispersal. To date (June 2020), there are only two novel structures of CHASE domains available in the PDB (from AH4K (PDB IDs: 3T4J, 3T4K, 3T4L, 3T4O, 3T4Q, 3T4S and 3T4T[169]) and PcrK (PDB ID: 6K62[170])). However, the first CHASE domain structures to be determined, from AH4K, were originally deposited as periplasmic tandem PAS-PAS domains[169], and consequently there may be more CHASE domain structures deposited under the wrong identity. One reason for this is the CHASE domain is suggested to be formed up of two PAS subdomains and another reason is the low sequence identity between CHASE domains[170], making them difficult to identify.

The CHASE4 domain has previously been annotated to be formed of two α-helices which flank six β-strands and a loop, of 20 to 45 residues, in the middle of the CHASE4 domain[127]. Although, in PA2072 the periplasmic region is predicted to consist of α-helices which flank six β-strands, the β-strands are interspersed by three α-helices (Figure 4.3 B). In contrast, the periplasmic region of RbdA is predicted to be formed up of nine α-helices (Figure 4.3 A). This is similar to the CHASE3 domain which is formed up entirely of four to six α-helices[127]. Both CHASE3 and CHASE4 have been found in diguanylate cyclases/phosphodiesterases and have always been found between two transmembrane regions[127]. However the periplasmic region of RbdA does not contain the highly conserved CHASE3 signature motif of Arg-Gly-aromatic-aliphatic-aliphatic-alcohol residues in the α-1 to α-2 loop[127]. Therefore it is unlikely that the periplasmic region of RbdA is a CHASE3 sensory domain. Interestingly, the signature motif of the CHASE4 domain is Trp-Asp-Asp in the α1-helix[127], but this region of PA2072 contains the residues $_{77}$Trp-Thr-Asp$_{79}$. Nevertheless, the differences between the periplasmic regions of RbdA and PA2072 (Figure 4.3), probably result in the opposing physiological outputs we observe *in vivo*.

Despite efforts here, we were unable to produce optimal protein crystals of the PA2072 $CHASE4_{45-254}$ domain for X-ray diffraction studies. However, the structure of the PA2072 CHASE4 domain could hold valuable insight into what and how it senses a stimulus and how the signal could be transduced from the periplasm to the enzymatic effector domain(s) within the protein complex.

PAO1 KO mutants of Δ*rbdA* and Δ*pa2072* have been found to have an impaired NO induced biofilm dispersal (biomass reduction of 17 ± 8% and 13 ± 8% after NO treatment respectively), compared to that of a WT (57 ± 5%)[130]. However, RbdA and PA2072 seem to have opposing roles in *P. aeruginosa* PAO1 biofilm formation. As, when cultured for only 48 hrs for early stage biofilms, Δ*rbdA* formed thicker biofilms with an increased microcolony size, biomass, surface coverage and higher intracellular levels of c-di-GMP compared to a WT. This is in contrast to Δ*pa2072* biofilms which had a reduced microcolony size, biomass, surface coverage and lower intracellular levels of c-di-GMP compared to a WT[130]. This biological data suggests that RbdA functions predominately, during biofilm formation, as a PDE whereas PA2072 seems to function predominantly as a DGC.

The DGC and PDE activities of PA2072 are currently experimentally unknown. However, the GGDEF and EAL domains of PA2072 contain the consensus GGDEF and EAL with the DDFGTGYSS motifs respectively. Along with homology modelling with SWISS-MODEL in our recent Scientific Reports publication[130], we suggest that PA2072 is enzymatically active as both a DGC and PDE. Furthermore, RbdA has been shown to have both DGC and PDE activities experimentally[131], supporting the hypothesis that PA2072 has both DGC and PDE enzymatic activities.

It is interesting to note that, while the protein sequences making up the GGDEF and EAL domains in RbdA and PA2072 are highly similar (Figure 4.2), the sequence alignment between RbdA and PA2072 differs around the H-helix (residues 537-562 of RbdA). These amino acid differences at the H-helix could change the relative orientation of the GGDEF and EAL domain and/ or regulation between the GGDEF and EAL domain in PA2072 compared to in RbdA. This structural and/ or regulatory difference between RbdA and PA2072, at the H-helix, could cause the two proteins to have differing enzymatic outputs when the dispersal stimulus (NO) is absent, thus explaining the observed differences in biofilm formation between the Δ*rbdA* and Δ*pa2072* mutants. The opposing physiological roles observed between Δ*rbdA* and

Δ*pa2072* could also be due to the extra CHASE4 domain within PA2072 which may have a structural and/ or regulatory effect on the downstream effector domains.

In this study we were unable to purify the PA2072 $PAS_{310-427}$ domain with a bound cofactor (Appendix B - Figure 7.4). However, this does not mean that the PA2072 PAS domain does not bind a cofactor *in vivo*. The cofactor could have been unavailable or available only at suboptimal concentrations during the protein expression. Also, the binding affinity of the PA2072 $PAS_{310-427}$ domain for its cofactor could be low and therefore have been unable to bind it or the cofactor could have been lost during the purification process. This may be due to the domain being isolated from the rest of the protein complex. Nevertheless, the proximal Histidine residue required for haem coordination is absent from the PAS domain of PA2072. Therefore it is unlikely that the PAS domain of PA2072 can bind the cofactor haem. Interestingly, when Liu et al. expressed the isolated PAS domain of RbdA weak haem binding was observed but reconstitution experiments with the cRbdA protein did not produce strong spectroscopic evidence for haem binding[131].

However, many PAS domains have been reported to function as sensory domains without a bound cofactor[110], so the PAS domain of PA2072 could function as a sensory domain by directly binding a ligand without requiring a bound cofactor. It is also possible that the sensing of different stimuli by the PA2072 CHASE4 and PAS domain can be integrated, allowing for the fine-tuning of the opposing DGC and PDE enzymatic activities. Furthermore, the CHASE4 domain may act as the sole sensory domain in PA2072 and upon sensing the stimulus transduces the signal to the PAS domain *via* conformational changes. This could lead to a change in the PA2072 PAS domain dimerisation and therefore a change in the dimerisation of the PA2072 GGDEF and EAL domains required for DGC and PDE activity respectively. Further investigation into the CHASE4 and PAS domains, as well as the PA2072 protein complex as a whole, is therefore required.

6.2.2 – Suggested further work in order to characterise PA2072

To further examine which protein domain(s) of RbdA and PA2072 accounts for the differences in their physiological roles, chimeric proteins formed of a mixture of RbdA and PA2072 domains could be constructed (for example, the PAS domain of RbdA could be replaced with the PAS domain of PA2072)[185,186]. The chimera proteins could then be individually expressed in Δ*rbdA* and Δ*pa2072* KO mutants to determine whether the chimera proteins could rescue the mutant phenotype. This could then give us insight into which protein domain(s) of RbdA and PA2072 are responsible for the observed differences in their physiological roles. For example, if a chimera protein, formed of RbdA but with the RbdA PAS domain replaced with the PAS domain of PA2072, could rescue the phenotype of the Δ*pa2072* mutant but not the Δ*rbdA* mutant, then we could suggest that the physiological differences between RbdA and PA2072 resulted from differences in their PAS domains. Further experiments using smaller fragments, amino acid substitutions or deletions could then narrow down the precise region or residue(s) that are responsible for the physiological differences between RbdA and PA2072[185]. It is useful to know the 3D structure of the proteins to be made chimeric in order to aid the design of the chimera(s), although this is not an absolute requirement if a homology model can be generated[185]. Given the structural similarity between RbdA and PA2072 this would be a suitable experiment.

The protein crystals of the PA2072 $CHASE4_{45\text{-}254}$ and $PAS_{310\text{-}427}$ domains require further optimisation for X-ray diffraction and structure determination experiments. This could require the further optimisation of seeding experiments. Undiluted seed stocks were used in the seeding experiment for PA2072 $PAS_{310\text{-}427}$ which led to many small protein crystals instead of fewer larger crystals. Therefore lower seed concentrations (by progressive serial dilutions) should be used in future seeding experiments with PA2072 $PAS_{310\text{-}427}$ in order to try to improve crystal size. Likewise, multiple rounds of seeding may also be required to produce protein crystals that diffract.

Alternatively, we could try using the new Versatile Macromolecular Crystallography micro/ nanofocus (VMXm) beamline at Diamond Light Source[187,188] with the PA2072 $PAS_{310\text{-}427}$ crystals. The beam size of the VMXm beamline is 0.3 – 10 µm vertically and 0.5 – 5 µm horizontally and is therefore used for diffraction experiments with microcrystals as small as

0.5 µm[187,188]. The crystals are prepared on Transmission Electron Microscope grids and aligned to the X-ray beam using a Scanning Electron Microscope, with the samples placed under vacuum for data collection to improve the signal to noise[188]. As the PA2072 $PAS_{310-427}$ crystals produced were small (and could possibly be produced even smaller by using seeding experiments with a concentrated seed stock), the VMXm beamline may be ideal for the diffraction of these crystals.

Furthermore, the Versatile Macromolecular Crystallography *in-situ* (VMXi) beamline at Diamond Light Source[189] could also have been used for diffraction experiments with the crystals of PA2072 $CHASE4_{45-254}$. The VMXi beamline allows for the characterisation and data collection of crystals directly within the crystallisation trays *in situ*[189]. As the VMXi beamline eliminates the need to remove crystals from the crystallisation trays, this beamline would be ideal for the diffraction experiments of PA2072 $CHASE4_{45-254}$, as these crystals were clustered and unable to be fished.

Crystalline material of the PA2072 $CHASE4_{45-254}$ protein, which showed up well under UV-light, were also observed in the trial screens JCSG$^+$ (well H6), Morpheus (well B10 and D6) and PACT *premier* (wells C3 and E3). As this crystalline material was present in a variety of different crystallisation conditions, they may represent different crystal morphologies of PA2072 $CHASE4_{45-254}$. Therefore, if one or more of these crystalline materials were used in seeding experiments, crystals of PA2072 $CHASE4_{45-254}$ that could be fished and diffracted may have been produced. Furthermore, only the PA2072 $CHASE4_{45-254}$ monomers (SEC fraction 36) were used in crystallisation trials in this study. However, the apparent PA2072 $CHASE4_{45-254}$ dimer (Figure 4.8) could have also been used in separate crystallisation trials (using SEC fractions 32 and 33), as dimers can promote the formation of a crystal lattice[190,191].

In this study we were unable to determine a bound cofactor for the PA2072 $PAS_{310-427}$ domain. Simply, this could be because this particular PA2072 $PAS_{310-427}$ domain construct is not compatible with cofactor binding due to the size of the construct itself having an unknown structural impact. Either due to the $PAS_{310-427}$ domain being isolated from the rest of the PA2072 protein complex or because a structural feature necessary for cofactor binding wasn't included in this construct. Therefore it is possible that further PA2072 PAS domain constructs of different sizes and/ or different start and end sites need to be screened for cofactor binding.

Reconstitution experiments of the PA2072 $PAS_{310-427}$ domain with the haem cofactor could also be trialled to investigate the proteins haem binding capabilities. In these experiments the purified apoprotein is incubated for a period of time with an excess of the haem derivative hemin[192]. The unbound hemin is then purified from the rest of the protein[192] and UV-Vis spectroscopy can be carried out[136]. The position of the Soret band produced by hemin in UV-Vis spectroscopy is shifted when bound to a protein (~410 nm) compared to the Soret band of free unbound hemin in solution (~385 nm)[193]. However, for the hemin cofactor to adopt a native haem coordination in the binding cleft of the PAS domain, the apoprotein may have to be first unfolded and then refolded around the cofactor.

To determine whether the predicted enzymatic activities of PA2072, based on the conservation of catalytic residues and homology modelling, is accurate, the DGC and PDE enzymatic activities of the PA2072 GGDEF and EAL domains should be investigated (as in Section 3.1 for RbdA). The *in vitro* enzymatic activities of PA2072 may also shed light on the apparent opposing physiological roles of RbdA and PA2072.

6.3 – PA0285

6.3.1 – The PAS domains of PA0285

We hypothesize that the inducer of biofilm dispersal, NO, is sensed by a PAS domain bound to a haem cofactor due to NO previously being found to bind to haem within a PAS domain[122] and PAS domains being widely found associated with c-di-GMP DGC and/ or PDE domains[116,117]. PA0285 was highlighted as a target due to our previous biological phenotyping of *P. aeruginosa* KO mutants, which found that Δ*pa0285* biofilms had a different structure with a reduced biomass and reduced surface coverage compared to a PAO1 WT biofilm. Also upon induction with NO, the Δ*pa0285* mutant biofilms dispersed more than compared to the PAO1 WT biofilms. Further, we found that Δ*pa0285* mutants also had a reduced swimming motility (mediated by flagella[194]) and completely lacked a twitching motility (mediated by type IV pili[195-197]) compared to a PAO1 WT. Along with an altered swarming phenotype (dependant on type IV pili and flagella[194,198,199] in *P. aeruginosa*), the Δ*pa0285* mutant displays an increased swarming when induced with NO compared to a PAO1 WT. We have previously suggested that swarming motility is required for NO-induced biofilm dispersal[130] and, taken together, this data may suggest that, under normal conditions, PA0285 interacts with and/ or regulates type IV pili and/ or the flagella within *P. aeruginosa*.

In contradiction to the finding that the Δ*pa0285* mutant biofilm dispersed more than the PAO1 WT, we have previously found that the overall intracellular levels of c-di-GMP are significantly higher in planktonic Δ*pa0285* mutants compared to the intracellular c-di-GMP levels of a PAO1 WT. Nevertheless, it is becoming accepted that local pools of c-di-GMP are regulated by spatially distinct proteins within a cell to mediate specific responses[88]. For example, experiments investigating the phenotypes of two *P. aeruginosa* DGC mutants, Δ*roeA* and Δ*sadC*, found that the total cellular c-di-GMP levels and the observed phenotypic outputs were not correlated[200]. Equally, local pools of c-di-GMP could be explained by the formation of protein complexes in which the c-di-GMP turnover protein(s) is in the immediate vicinity of the c-di-GMP receptor/ effector protein(s)[201]. Being associated with the inner cellular membrane, PA0285 could be spatially regulated within the cell and/ or could form a protein complex with other effector/ receptor protein(s). Localised c-di-GMP pools could explain the

increased NO-induced biofilm dispersal of Δ*pa0285* mutants while still maintaining higher global c-di-GMP levels compared to a PAO1 WT.

Here, we found that the PAS2 domain of PA0285 binds to an FAD cofactor in an approximate 2 : 1 stoichiometry. PAS domains bound to a flavin cofactor (FMN or FAD; Figure 6.2) can be used to sense a variety of stimuli such as blue light, oxygen concentrations and redox[118,202]. Fully oxidised FMN/ FAD can be reduced to FMNH/ FADH *via* the acceptance of a hydrogen ion plus an electron at the N5 atom of the tricyclic dimethyl-isoalloxazine ring, and FMNH/ FADH can be further reduced to $FMNH_2$/ $FADH_2$ *via* the N1 atom of the isoalloxazine ring and vice versa. With the FMN/ FAD cofactor in PAS domains being evidenced to undergo a reversible oxidation-reduction process[203]. The flavin may be in an oxidised or reduced form in its redox resting state, depending on the physiological environment and signalling pathway[204]. A stimulus which changes the redox state of a cell can be sensed by a change in the redox state of an FMN or FAD cofactor.

Known as a flavin redox switch, the different redox states of the flavin cofactor are able to mediate conformational changes within a protein, affecting intra- or inter-domain interactions[118,204]. This couples various cellular processes to the redox state of the cell[204]. For example, the redox-mediated regulators, NifL[118], Aer[202] and *Ax*DGC2[203] bind an FAD cofactor *via* PAS domains to sense changes in the redox state to regulate transcription, bacterial motility and cellulose synthesis respectively.

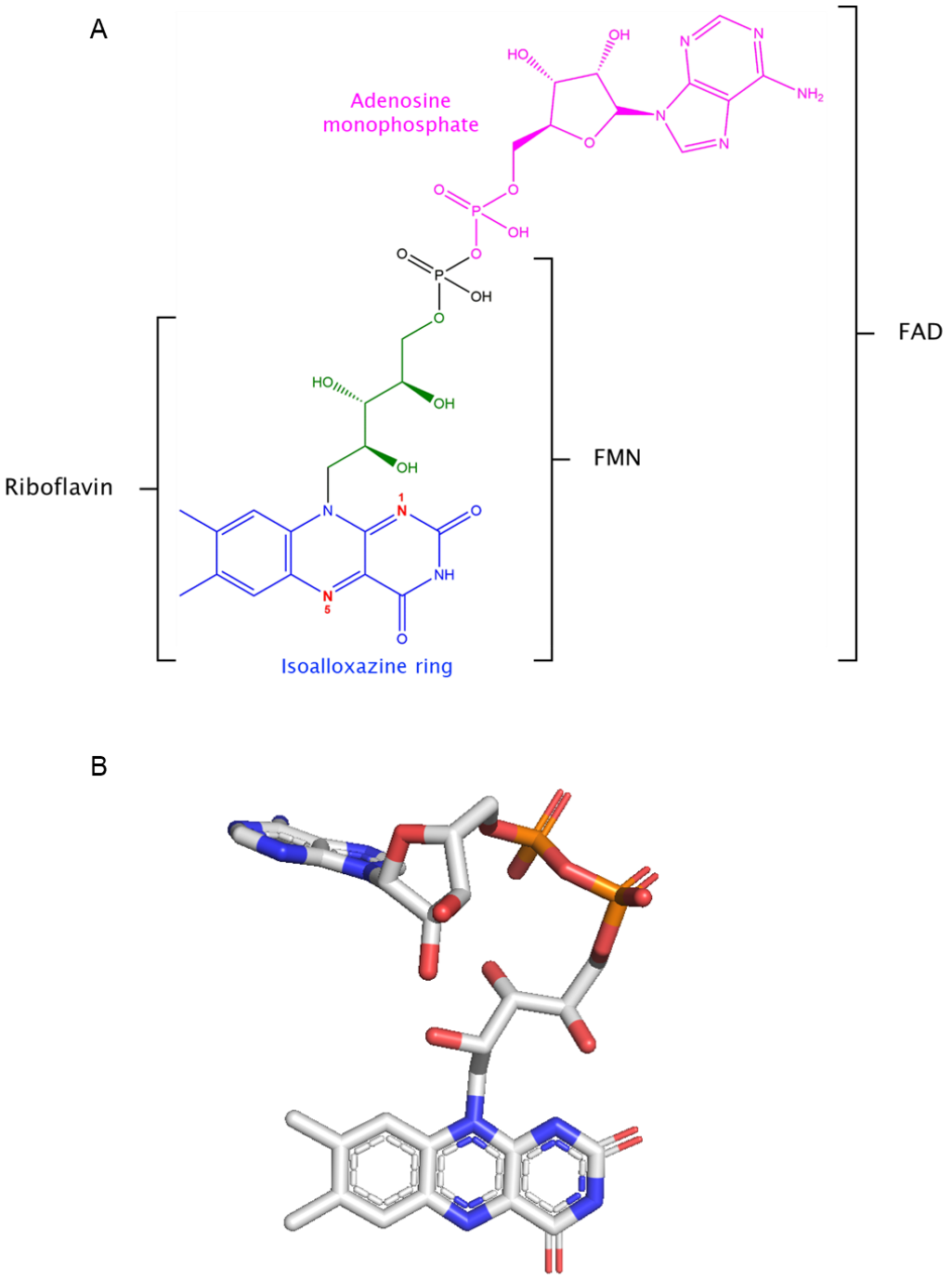

Figure 6.2. **Structure of flavin adenine dinucleotide (FAD)**. A) The chemical structure of FAD drawn with ChemDraw Professional (version 16.0.1.4). The isoalloxazine ring, riboflavin, flavin mononucleotide (FMN) and Adenosine monophosphate components that together form the FAD cofactor are indicated. The N1 and N5 atoms of the isoalloxazine ring that are the sites of protonation / deprotonation in the reduction/ oxidation of flavin cofactors are numbered and coloured red. B) The three-dimensional structure of an FAD cofactor from the first PAS domain of NifL (PDB ID: 2GJ3 chain A). Figure produced with PyMOL with Carbon atoms shown in grey, Oxygen atoms shown in red, Nitrogen atoms shown in blue, Phosphate atoms shown in orange

*Ax*DGC2 from *Acetobacter xylinum* consists of a PAS, GGDEF and EAL domain, with the EAL domain being catalytically inactive towards c-di-GMP, compared to the GGDEF domain which synthesises c-di-GMP and is regulated by the redox state of the FAD-bound PAS domain[203]. It was found that the DGC catalytic rate of *Ax*DGC2 was higher in its oxidised state compared to in its reduced state. This was attributed to the redox state of the FAD cofactor and not to the presence of oxygen due to the DGC activity being unaffected when measuring the kinetic parameters of the oxidised *Ax*DGC2 protein in oxygenated or anaerobic buffer[203]. It is suggested that when the N5 atom in the isoalloxazine ring is protonated or deprotonated, changing the redox state of the flavin cofactor, a reorganisation of the internal hydrogen bonds in the flavin binding pocket ensues, inducing structural changes in the PAS domain[118,203]. These structural changes in the PAS domain then go on to propagate to the C-terminal domains to bring about a functional switch, such as an exposure of a protein binding surface or an increase in enzymatic activity[118,203,204].

The intracellular redox state of a cell can be affected by NO and the reactive NO-derived species[205] which can consequently alter the redox state of an FAD cofactor. Therefore, it is possible that in *P. aeruginosa* the NO stimulus mediates a change in the redox state of the FAD cofactor in the PA0285 PAS2-GGDEF-EAL system, causing conformational changes that lead to a regulation of the GGDEF domain and/ or the EAL domain. As the previous phenotypical data evidenced an increased biofilm dispersal in Δ*pa0285* mutants compared to the PAO1 WT, this may suggest that, under normal conditions, upon NO-induced changes to the cellular redox state, PA0285 acts as a DGC (Figure 6.3). Therefore, PA0285 could have the function in setting the minimum threshold required for inducing a biofilm dispersal.

In Figure 6.3, model 1 shows PA0285 dimerising at the PAS2, GGDEF and EAL domains upon NO induced oxidation of the cytoplasm, resulting in both DGC and PDE activity. This model is not plausible as both opposing enzymatic domains would be working against one another which is not energetically favourable and not logical. Model 2 shows PA0285 dimerising at both the PAS1 and PAS2 domains of PA0285. While the PAS1 domain of PA0285 may participate in the dimerisation of PA0285, it is unlikely that the PAS1 domain is required for dimerisation as the isolated PA0285 $PAS2_{213-333}$ domain was still able to form dimers. Furthermore, the isolated PA0285 $PAS1_{88-216}$ construct was found to only purify as a monomer (Figure 5.2).

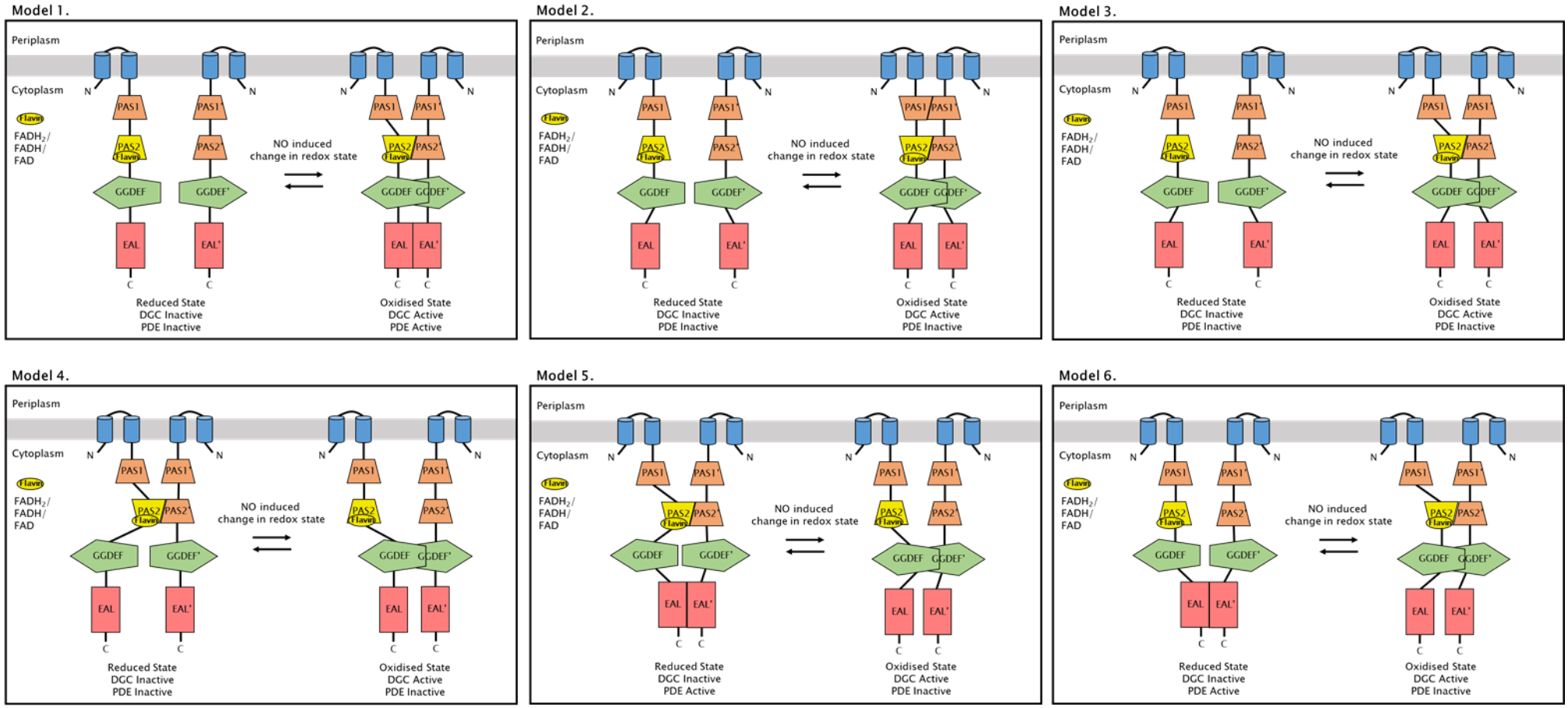

Figure 6.3. **Proposed models of the possible mechanisms for PA0285.** PAS (Per-ARNT-Sim) domains are shown as an orange trapezoid with a flavin cofactor (either FAD, FADH or $FADH_2$ and shown as a yellow oval) bound PAS domain shown as a yellow trapezoid. GGDEF domains (named after the residues that form a conserved catalytic motif) are shown as a green pentagon. EAL domains (named after the residues that form a conserved catalytic motif) are shown as red rectangles. Dimerisation of the GGDEF domains is required for diguanylate cyclase (DGC) activity with dimerisation of the EAL domains being required for phosphodiesterase (PDE) activity. Blue cylinders represent transmembrane domains and the grey rectangle represents the inner cellular membrane.

Although model 3 could be correct, it does not explain the increased levels of c-di-GMP observed in the planktonic Δ*pa0285* mutants compared to the PAO1 WT. Model 4 is unlikely to be correct for the same reason as model 3. In model 4 the PA0285 PAS2 domain is dimeric under reducing conditions but monomeric under oxidising conditions. In this work the PA0285 flavin–bound $PAS2_{213-333}$ domain was purified as a dimer under oxidising conditions, but the flavin-bound PAS2-GGDEF-$EAL_{213-760}$ complex was purified as a monomer (as indicated by the SEC column) under oxidising conditions. However, both the PA0285 $PAS2_{213-333}$ domain and the PAS2-GGDEF-$EAL_{213-760}$ complex were found bound to FAD in a stoichiometry of approximately 2 : 1. Therefore, to determine whether the redox potential has an effect on the oligomeric state of PA0285, purification of PA0285 $PAS2_{213-333}$ and PAS2-GGDEF-$EAL_{213-760}$ under reducing conditions should be carried out and compared to protein purification under oxidising conditions. In model 5 the dimeric PAS2 domains of PA0285 facilitates the dimerisation of the EAL domains and therefore PDE activity under reducing conditions. This is in agreement with the biological data that found increased intracellular levels of c-di-GMP in planktonic cells of Δ*pa0285* mutants compared to a wild type. This would suggest that PA0285 has a PDE activity in the WT under normal conditions. Redox changes induced by NO are sensed by the PAS2-bound flavin cofactor, causing conformational changes which are propagated throughout PA0285 causing the PAS2 and the EAL domains to become monomeric, while allowing the GGDEF domain to bind its partner GGDEF' domain leading to DGC activity. This is in agreement with the phenotypical data which found the Δ*pa0285* biofilm mutant had an increased biofilm dispersal compared to a WT, suggesting that PA0285 is active as a DGC upon NO induced biofilm dispersal in the WT under normal conditions. Model 6 is based on the same phenotypic data as model 5, except that the PAS2 domains are monomeric under reducing conditions and dimeric under oxidising conditions.

It is also possible that the PAS1 domain in PA0285 further modulates the enzymatic activities of the GGDEF and EAL domains and/ or allows for the integration of additional stimuli. In this study we were only able to evidence a very weak haem binding to the PA0285 $PAS1_{88-216}$ domain, with only a 0.048 % occupancy if assuming a 1 : 1 stoichiometry. The slight haem peak observed in these UV-Vis spectrums may be from a non-specific binding of the haem to the protein or some contaminant haem that made it through or came from the purification process. Thus it is possible the PA0285 PAS1 domain does not bind haem under normal

circumstances. Conversely, it is also possible that the binding affinity of the PA0285 PAS1 domain for haem is low and that the haem was lost during the purification process, possibly due a structural alteration resulting from the PA0285 PAS1 domain being isolated from the protein complex. Equally the low occupancy of haem in PA0285 $PAS1_{88\text{-}216}$ could represent a low availability of haem during protein expression and a need for further optimisation of the expression system. Furthermore, the haem peak is shifted to a wavelength of 410 nm (from 385 nm of free hemin) which suggests the haem cofactor is bound within a binding cleft in the PA0285 PAS1 domain[172]. Additionally, the proximal His residue, which serves as the axial ligand to haem, is present within the PAS1 domain of PA0285 (His133). Also present within the PAS1 domain of PA0285 is an aromatic residue, Trp95, corresponding to residues Trp45 in DosP and Phe176 in BjFixL which interact with the haem cofactor within the binding pocket[116]. Therefore the haem binding ability of the PA0285 PAS1 domain requires further investigation.

6.3.2 – Further work required to understand the regulation of PA0285

Here, we suggest that PA0285 could be involved in the regulation of the *P. aeruginosa* type IV pili and/ or flagella. In order to assess whether PA0285 has any effect on the presence of the type IV pili and/ or flagella, bacterial cells taken from the centre or the edge of a swarm could be studied with electron microscopy[198], comparing the swarming cells of a Δ*pa0285* mutant with a PAO1 WT. Further, to determine whether PA0285 localises with the type IV pili or flagellum, the localisation of PA0285 within the *P. aeruginosa* cell could be determined. This could be investigated by attaching a Green Fluorescent Protein to PA0285, expressing the fusion protein in Δ*pa0285* mutants and carrying out fluorescence microscopy[206].

We have previously predicted the GGDEF and EAL domains of PA0285 to be catalytically active, as determined by the presence of the consensus motifs within the protein sequence that infer DGC and PDE enzymatic activity along with active site homology modelling[130]. However the enzymatic activities of the PA0285 GGDEF and EAL domains should be experimentally determined. This could also provide further insight into the regulation between the GGDEF and EAL domains (see Section 3.1). The potential regulation of these enzymatic activities by the N-terminal cofactor bound PAS domain(s) in response to stimuli, such as changes in redox potential, should also be investigated. This could be carried out by measuring the enzymatic activities of the PA0285 protein in oxidising and reducing buffers and/ or in the presence and absence of an NO donor (such as SNP). This would require part of the protein preparation to be carried out within an anaerobic glove box to prevent the oxidation of the cofactor[203].

In this study, we were not able to definitively conclude a haem-*b* cofactor binding to the PA0285 $PAS1_{88\text{-}216}$ domain and possible reasons for this are discussed above. Here, the BL21 (DE3) pHPEX-3 expression system was used when producing the PA0285 $PAS1_{88\text{-}216}$ domain (Section 2.8.2) to try to ensure that the haem cofactor would be available to the PA0285 $PAS1_{88\text{-}216}$ domain during production. Another protein expression system, *E. coli* Nissle 1917 cells, also increases the amount of intracellular haem available to proteins during expression due to the presence of a chromosomal copy of the *chuA* haem receptor gene[136]. However, Nissle 1917 cells do not possess a chromosomal copy of the T7 RNA polymerase and so the use of this system either requires the gene of interest to be cloned into a vector so that is it

not under the control of a T7 promoter[136], or the use of a T7 RNA polymerase expressing vector such as pAR1219. Both the BL21 (DE3) pHPEX-3 and Nissle 1917 protein expression systems require further optimisation. It could be useful to vary the temperature that the cells are grown at after induction, the concentration of the inducer, the cell density at which they are induced, the length of time between induction and cell harvesting, and the growth medium, in order to determine whether the levels of target protein expression and the amount of cofactor incorporation are affected. This is because these variables have been, in some cases, shown to have a significant effect on the amount of target protein that is expressed within some cell lines[207-209] and thus, these variables should be optimised.

Further investigation into the cofactor binding of the PA0285 PAS1 domain using the BL21 (DE3) pHPEX-3 or Nissle 1917 expression system would benefit from the use of a haem-binding positive and negative control protein. This would ensure that any haem binding observed for the PA0285 PAS1 domain is not contaminant haem from the purification process (negative control) and that the availability of haem to the protein during expression is not limiting (positive control). Cofactor reconstitution experiments could also be carried out[210,211], followed by UV-Vis spectroscopy, to determine whether the PA0285 PAS1 domain could bind to the haem cofactor or not. Another experiment would be to construct a H133A and/or a W95A mutant of PA0285 $PAS1_{88-216}$ and measure the haem occupancy with UV-Vis spectroscopy, as these residues would be important for the ligation of the haem cofactor. This would determine whether the amount of haem binding is diminished or not and whether the (small amount of) haem cofactor is binding specifically within the binding pocket of the PAS1 domain of PA0285 or non-specifically to the protein surface[116].

In this study we identified the flavin cofactor bound to the PA0285 $PAS2_{213-333}$ domain to be FAD, due to the comparison of our UV-Vis spectrums with previous FAD-bound protein spectrums[174] and the predicted PA0285 FAD cofactor binding based on the PAS domain sequence comparisons[134]. However, as the UV-Vis spectrums for protein-bound FMN and FAD cofactors are similar, High Performance Liquid Chromatography could be used with a set of cofactor standards to experimentally determine whether the FMN or FAD cofactor is present in the PA0285 PAS2 domain. FMN has an elution time of 18 mins producing a broad peak, whereas FAD has an elution time of 9.5 mins producing a sharper peak[212]. Mass spectrometry is another technique that could have been used to determine the identity of the cofactor

bound to PA0285 $PAS2_{213-333}$, as FAD has a molecular weight of 785.55 Da whereas FMN has a molecular weight of 456.34 Da.

To determine the exact PA0285 $PAS2_{213-333}$ and PA0285 PAS2-GGDEF-$EAL_{213-760}$ to FAD stoichiometry, a UV-Vis spectrometry method that measures the amount of free FAD could be used. In this method, a known concentration of purified protein (with FAD bound in an unknown stoichiometry) is denatured (for example by heating the protein sample) to release the FAD. Centrifugation then removes the denatured protein and after transferring the supernatant to a cuvette, the absorbance of free FAD cofactor (extinction coefficient of 11,300 M^{-1} cm^{-1}) can be measured at 450 nm with a UV-Vis spectrometer[212,213]. The concentration of free FAD in solution can then be solved with the Beer-Lambert Law and compared to the starting protein concentration to determine the protein to FAD cofactor stoichiometry. Furthermore, the protein specific FAD extinction coefficient can then be solved with the Beer-Lambert Law once the concentration of FAD is known[212].

The approximate 2 : 1 PA0285 $PAS2_{213-333}$ and PAS2-GGDEF-$EAL_{213-760}$ to FAD stoichiometry reported here could be of biological significance or it could represent a suboptimal FAD incorporation during protein expression or loss of the FAD cofactor during the purification process. This may be represented by the SEC protein peaks that were observed to contain the PA0285 protein of interest but not the FAD cofactor, but equally these peaks could represent aggregated or miss-folded proteins (Figures 5.4 C and 5.7 C). Therefore, after SEC purification, the PA0285 PAS2 domain fragments could be deflavinated and then reconstituted with the FAD cofactor[214] using an excess of exogenous FAD to ensure a 1 : 1 protein to FAD stoichiometry which could be confirmed with a UV-Vis spectrum. The 1 : 1 protein to FAD stoichiometry sample could then be re-run through the SEC column to observe the effects of this 1 : 1 stoichiometry on the oligomerisation state of the PA0285 protein using its elution volume. The UV-Vis spectrum can then be re-measured to determine whether the sample is still in a 1 : 1 protein to FAD stoichiometry or whether this has changed during the purification process. Investigating whether the 2 : 1 PA0285 protein to FAD stoichiometry observed here is of biological significance or just an experimental artefact would aid the development and understanding of the PA0285 mechanism of action.

Unfortunately, the protein crystallisation attempts of PA0285 $PAS1_{88-216}$, $PAS2_{213-333}$ and PAS2-GGDEF-$EAL_{213-760}$ were unsuccessful. This could be due to a number of factors, including;

the protein concentration, the incubation temperature of the crystallisation trays (only 21°C was trialled), the homogeneity of the protein samples (there was not a 1 to 1 protein to cofactor ratio, with successful protein crystallisation requiring a homogenous protein sample), the amino acids at the beginning or end of the protein fragments or the presence of the affinity tag used for purification (N-terminal hexa-His tag)[215,216]. Furthermore, in this study the cytoplasmic domains of PA0285 were expressed and purified in isolation or as a fragment from the two N-terminal transmembrane domains. The transmembrane domains may mediate a structural restriction to the downstream cytosolic domains and the removal of the transmembrane domains may result in changes to the cytosolic domain(s) conformation or orientation relative to one another.

The change in Gibbs free energy, Δ*G*, governs whether it is feasible for protein molecules in solution to become ordered to form a protein crystal[217]. A process is thermodynamically favourable when Δ*G* is equal to a value less than 0, which depends on the change in enthalpy, Δ*H*, and the change in entropy, Δ*S*, at a constant given temperature, *T*, in an equation written as: $\Delta G = \Delta H - T\Delta S$.

Protein crystallisation is an entropy-driven process[217,218]. It is thermodynamically unfavourable for a system to become more ordered, as occurs when soluble proteins form an ordered crystal lattice, due to the negative change in entropy[217]. Nevertheless, if the ordering of protein molecules into a crystal releases water molecules (and possibly other small solvent molecules) from the protein surface so that the entropy of the system increases to an extent so that Δ*G* is equal to a negative value, protein crystallisation will be favourable. Of course, additional water and solvent molecules may become trapped within the protein crystal, leading to an entropy decrease and so favouring soluble protein molecules[219]. Therefore, a thermodynamic driving force for protein crystallisation is the release of structured water and solvent molecules from the surface of the soluble proteins[217,219,220].

However, the entropy 'cost' of crystallisation (the amount of entropy that is lost in forming a crystal) is also increased by amino acids on the protein surface which have large, hydrophilic and highly flexible side chains (such as Lys, Glu and Gln). This due to the random motions of the flexible side chains in solution (the residues have a high conformational entropy) and the loss of this motion (and thus loss of entropy) when forming crystal contacts[220]. This leads to a higher entropic barrier that must be overcome for these flexible residues to become ordered

within a crystal[221]. In contrast, amino acids on the protein surface with short, less flexible side chains (such as Ala, Thr and Ser), are more likely to order water and solvent molecules in solution and have less conformational entropy to lose when forming crystal contacts. Thus, residues on the protein surface with less flexible side chains, form a smaller entropy 'cost' during crystallisation compared to the larger flexible amino acid side chains[217,220].

Therefore, to promote the crystallisation of proteins (and possibly these PA0285 constructs), substituting flexible residues for less flexible residues, through site-directed mutagenesis, at the protein surface can be beneficial[221]. This technique is called surface-entropy reduction. Several studies have shown that the substitution of protein surface residues can lead to protein crystallisation and/ or new crystal forms[220-222].

Limitations of this technique include the risk of altering the proteins native structure[217], altering the biochemical behaviour of the protein and altering the proteins interactions with binding partners[218]. Furthermore, replacing hydrophilic residues on the protein surface with nonpolar residues can reduce the proteins stability and solubility[218], but this issue can be circumvented by the use of Tyr or Thr[221]. Without some structural information of the protein beforehand, identifying which flexible amino acids are appropriate targets to mutate, without causing a disruption to the overall protein structure, can be challenging[218]. However, as there are already many structures of PAS, GGDEF and EAL domains available in the PDB, these could help in the substitution-target identification process for the PA0285 constructs. Furthermore, the online surface-entropy reduction prediction (SERp) server is designed to identify residues within the protein sequence that could be mutated to help facilitate the proteins crystallisation[223].

The successful crystallisation of the PA0285 protein in part (as a PAS1-PAS2, PAS1-PAS2-GGDEF, PAS2-GGDEF, PAS1-PAS2-GGDEF-EAL or PAS2-GGDEF-EAL fragment) or as a full membrane protein complex would shed light on; 1) how the PAS1 and/ or PAS2 domain regulate the enzymatic domains in response to stimuli, 2) the regulation between the GGDEF and EAL domains and 3) how PA0285 could be inhibited with a small molecule (with a combination therapy of PA0285 inhibition and NO treatment putatively increasing the amount of biofilm dispersal than compared to using NO treatment alone).

6.4 – Conclusion

Within *P. aeruginosa* there are 41 proteins which contain at least one c-di-GMP regulatory domain (either a GGDEF, EAL or HD-GYP domain). There are 16 proteins which contain both GGDEF and EAL domains in tandem, with 9 of these proteins also containing at least one putative PAS sensor domain. Therefore this system is either redundant, is a complex network of proteins working together or a mixture of the two. It is highly unlikely that a single protein will be responsible for the regulation of biofilm formation and biofilm dispersal through increasing and decreasing the global intracellular levels of c-di-GMP *via* the activity of a single GGDEF and EAL domain respectively. Furthermore, the distinct phenotypes observed in related *P. aeruginosa dgc* or *pde* gene mutants suggests that these proteins are not redundant[87]. It is therefore likely, that many proteins sense the NO stimulus and, by each regulating a particular component *via* changing the local c-di-GMP levels, together induce a biofilm dispersal.

7 – Appendix

Appendix A

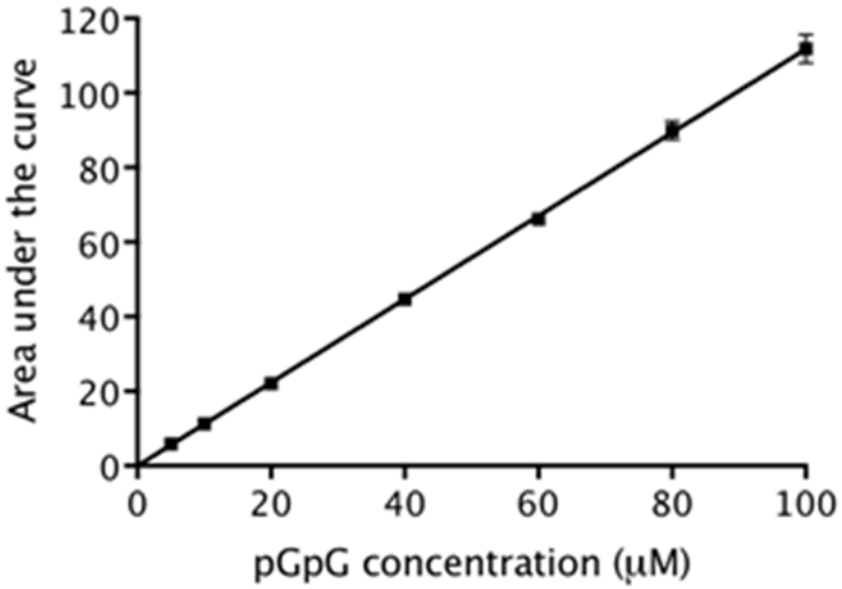

Figure 7.1. **Resource Q pGpG calibration curve**. Standards of pGpG are prepared at a known concentration and then treated in the same way as the reaction samples (i.e. 10 µL 100 mM $CaCl_2$ added and diluted with 890 µL 5 mM ammonium bicarbonate) before being run over the Resource Q column measuring absorbance at 253 nm. Area under the peak is then plotted against the known pGpG concentration in GraphPad Prism (v. 7) and a line of best fit produced giving the equation $y = 1.117x$ with an $R^2 = 0.9999$.

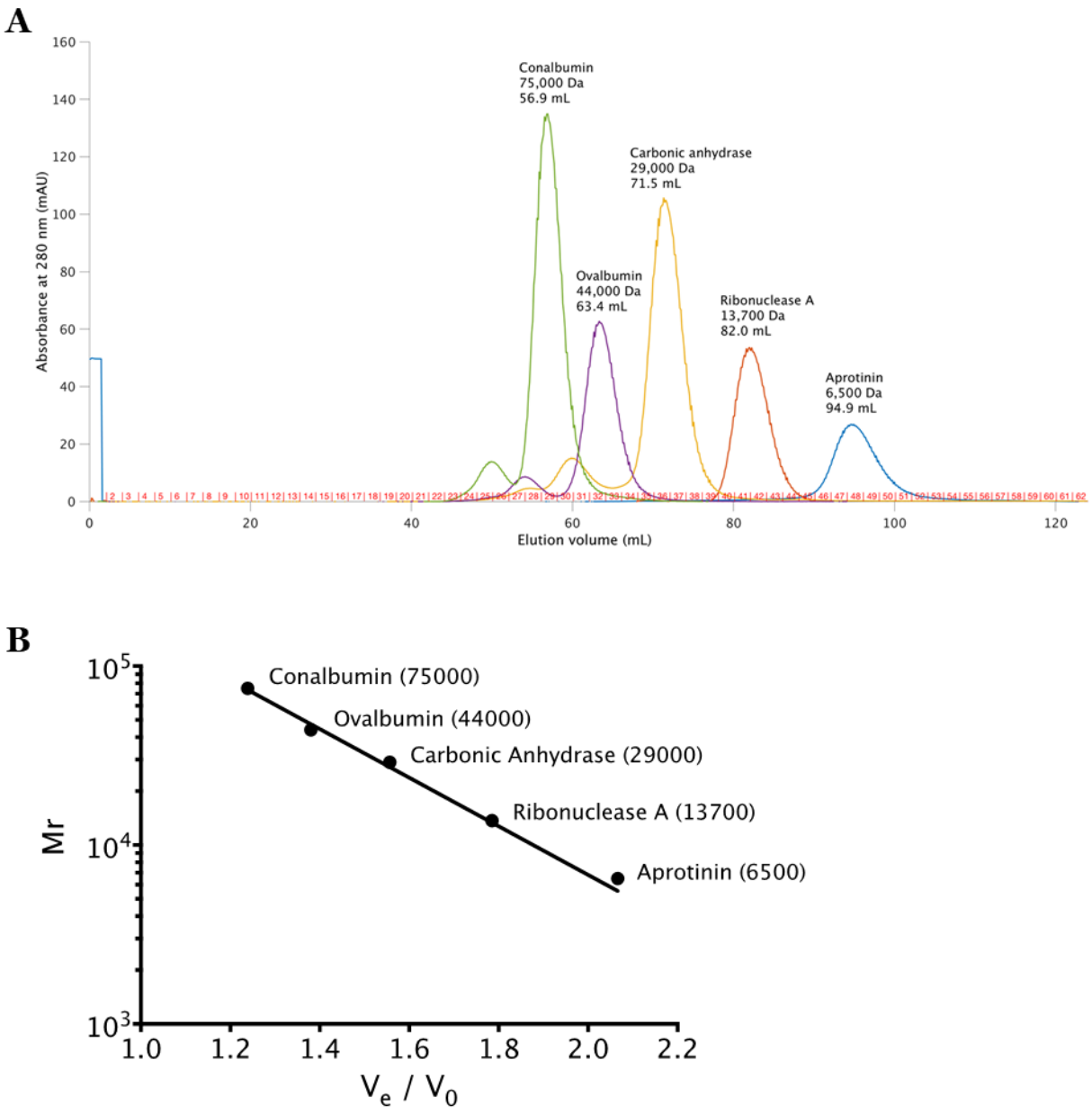

Figure 7.2. **Calibration of the HiLoad 16/600 Superdex 75 prep grade size exclusion chromatography column**. A) Size exclusion chromatography trace measuring absorbance at 280 nm of protein standards of known size (indicated on the trace with their respective molecular weights and elution volumes) when run on the HiLoad 16/600 Superdex 75 prep grade column at 1 mL/min. B) Calibration curve of the size exclusion chromatography column using the elution volumes of the protein standards and their known molecular weights. GraphPad Prism (v. 7) was used to produce a line of best fit with the equation: y = 10^(-1.358x + 6.55) with an R^2 = 0.9943.

Table 7.1. **RbdA EAL$_{549-797}$ crystallisation optimisation screen**.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 27 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 29 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 31 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 27 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 29 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 31 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M monosaccharides 32 % w/v precipitant mix 2
B	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 27 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 29 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 31 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 27 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 29 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 31 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.12 M ethylene glycols 32 % w/v precipitant mix 2
C	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 27 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 29 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 31 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 27 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 29 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 31 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 0.1 M amino acids 32 % w/v precipitant mix 2
D	0.1 M buffer system 3 (pH 8.5) 27 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 29 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 31 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.5) 27 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 29 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 31 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.5) 32 % w/v precipitant mix 2
E	0.1 M buffer system 3 (pH 8.1) 26 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 34 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 36 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.1) 26 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.1) 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.1) 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.1) 32 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.1) 34 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.1) 36 % w/v precipitant mix 2
F	0.1 M buffer system 3 (pH 8.4) 26 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 34 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 36 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.4) 26 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.4) 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.4) 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.4) 32 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.4) 34 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.4) 36 % w/v precipitant mix 2
G	0.1 M buffer system 3 (pH 8.7) 26 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 34 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 36 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 8.7) 26 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.7) 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.7) 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.7) 32 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.7) 34 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 8.7) 36 % w/v precipitant mix 2
H	0.1 M buffer system 3 (pH 9) 26 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 28 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 30 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 32 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 34 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 36 % w/v precipitant mix 1	0.1 M buffer system 3 (pH 9) 26 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 9) 28 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 9) 30 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 9) 32 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 9) 34 % w/v precipitant mix 2	0.1 M buffer system 3 (pH 9) 36 % w/v precipitant mix 2

Table 7.2. **Trialled optimisation screen for PA2072 CHASE4$_{45\text{-}254}$ protein crystallisation.**

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 10 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 10 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 16 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 16 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 16 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 16 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 16 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 16 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 16 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 10 % w/v PEG 6000	0.2 M sodium iodide	0.2 M sodium iodide 16 % w/v PEG 3350
B	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 12 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 12 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 18 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 18 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 18 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 18 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 18 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 18 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 18 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 12 % w/v PEG 6000	0.2 M sodium iodide 2 % w/v PEG 3350	0.2 M sodium iodide 18 % w/v PEG 3350
C	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 14 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 14 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 20 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 20 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 20 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 20 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 20 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 20 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 20 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 14 % w/v PEG 6000	0.2 M sodium iodide 4 % w/v PEG 3350	0.2 M sodium iodide 20 % w/v PEG 3350
D	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 16 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 16 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 22 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 22 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 22 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 22 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 22 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 22 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 22 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 16 % w/v PEG 6000	0.2 M sodium iodide 6 % w/v PEG 3350	0.2 M sodium iodide 22 % w/v PEG 3350
E	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 18 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 18 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 24 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 24 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 24 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 24 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 24 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 24 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 24 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 18 % w/v PEG 6000	0.2 M sodium iodide 8 % w/v PEG 3350	0.2 M sodium iodide 24 % w/v PEG 3350
F	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 20 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 20 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 26 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 26 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 26 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 26 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 26 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 26 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 26 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 20 % w/v PEG 6000	0.2 M sodium iodide 10 % w/v PEG 3350	0.2 M sodium iodide 26 % w/v PEG 3350
G	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 22 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 22 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 28 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 28 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 28 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 28 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 28 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 28 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 28 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 22 % w/v PEG 6000	0.2 M sodium iodide 12 % w/v PEG 3350	0.2 M sodium iodide 28 % w/v PEG 3350
H	0.1 M buffer system 1 (pH 6.5) 0.09 M halogens 24 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.12 M alcohols 24 % v/v precipitant mix 3	0.1 M buffer system 1 (pH 6.5) 0.06 M divalents 30 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.12 M monosaccharides 30 % v/v precipitant mix 1	0.1 M buffer system 1 (pH 6.5) 0.09 M NPS 30 % v/v precipitant mix 1	0.1 M sodium acetate (pH 5.5) 0.09 M NPS 30 % v/v precipitant mix 1	0.1 M citrate (pH 6.0) 0.09 M NPS 30 % v/v precipitant mix 1	0.1 M sodium cacodylate (pH 6.0) 0.09 M NPS 30 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 0.09 M NPS 30 % v/v precipitant mix 1	0.1 M MES (pH 6.0) 24 % w/v PEG 6000	0.2 M sodium iodide 14 % w/v PEG 3350	0.2 M sodium iodide 30 % w/v PEG 3350

Appendix B

Protein expression using Nissle 1917 cells with pAR1219 plasmid

Nissle 1917 cells (received from University of Kaiserslautern, Germany[136]) contain a chromosomal copy of the haem receptor *chuA*, which enables haem uptake into the cell. As Nissle 1917 cells do not already encode for T7 RNA polymerase, the pAR1219 plasmid (Sigma), which contains the T7 RNA polymerase gene under the control of the *lac*UV5 promoter, is used alongside the pET28a expression vector. The pAR1219 plasmid contains an ampicillin resistance gene.

Protein expression was carried out in 2 L baffled flasks containing 1 L LB medium with selective antibiotics (50 µg/mL kanamycin and 100 µg/mL ampicillin) which were inoculated with 8 mL of an overnight culture of co-transformed *E.coli* Nissle 1917 cells (same method as in Section 2.5; 1 µL of construct DNA with 1 µL of pAR1219 plasmid used with 20 µL of Nissle 1917 $CaCl_2$ competent cells). Cell cultures were grown at 37 °C with 180 rpm shaking to an OD_{600} value of 1.2 – 1.5 and then induced with a final concentration of 1 mM IPTG with or without a final concentration of 10 µM hemin (50 mM hemin stock prepared as in Section 2.8.2) added to each 1 L of culture just before induction. After induction the temperature of the shaker was turned down to 17 °C and the cell cultures were incubated in the dark for a further 18 hrs with 180 rpm shaking before harvesting. Cell harvesting was carried out as in Section 2.8.3, except centrifugation at 6,238 x g was carried out for 30 mins.

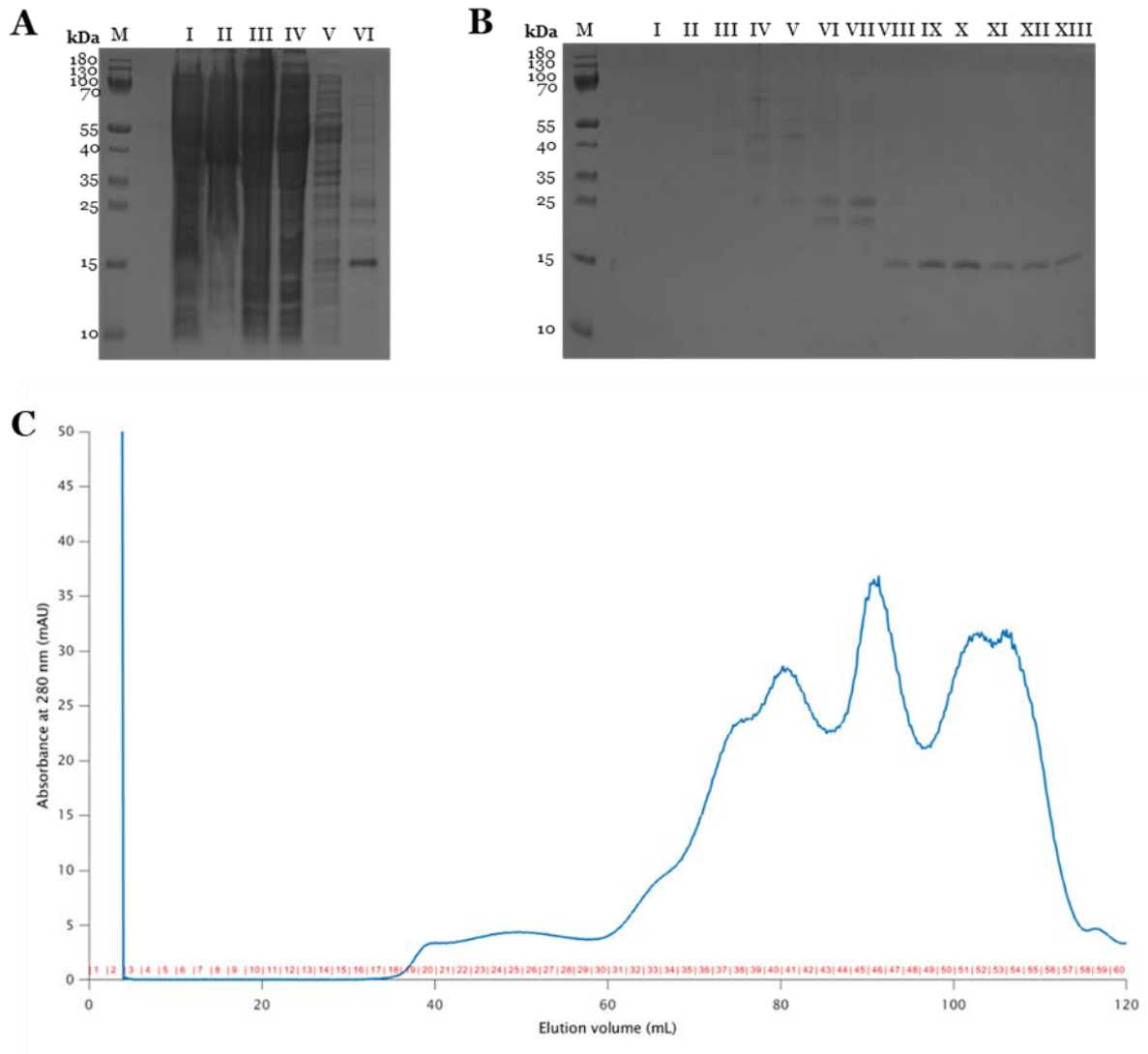

Figure 7.3. **Purification of PA2072 PAS$_{310-427}$ after expression with Nissle 1917 cells.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA2072 PAS$_{310-427}$ (approximately 15 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA2072 PAS$_{310-427}$ IMAC eluate showing fractions 20, 25, 33, 37, 38, 40, 41, 45 – 47, 51, 52, 54 in lanes I – XIII respectively. C) SEC trace of PA2072 PAS$_{310-427}$ measuring absorbance at 280 nm when run at 1 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red.

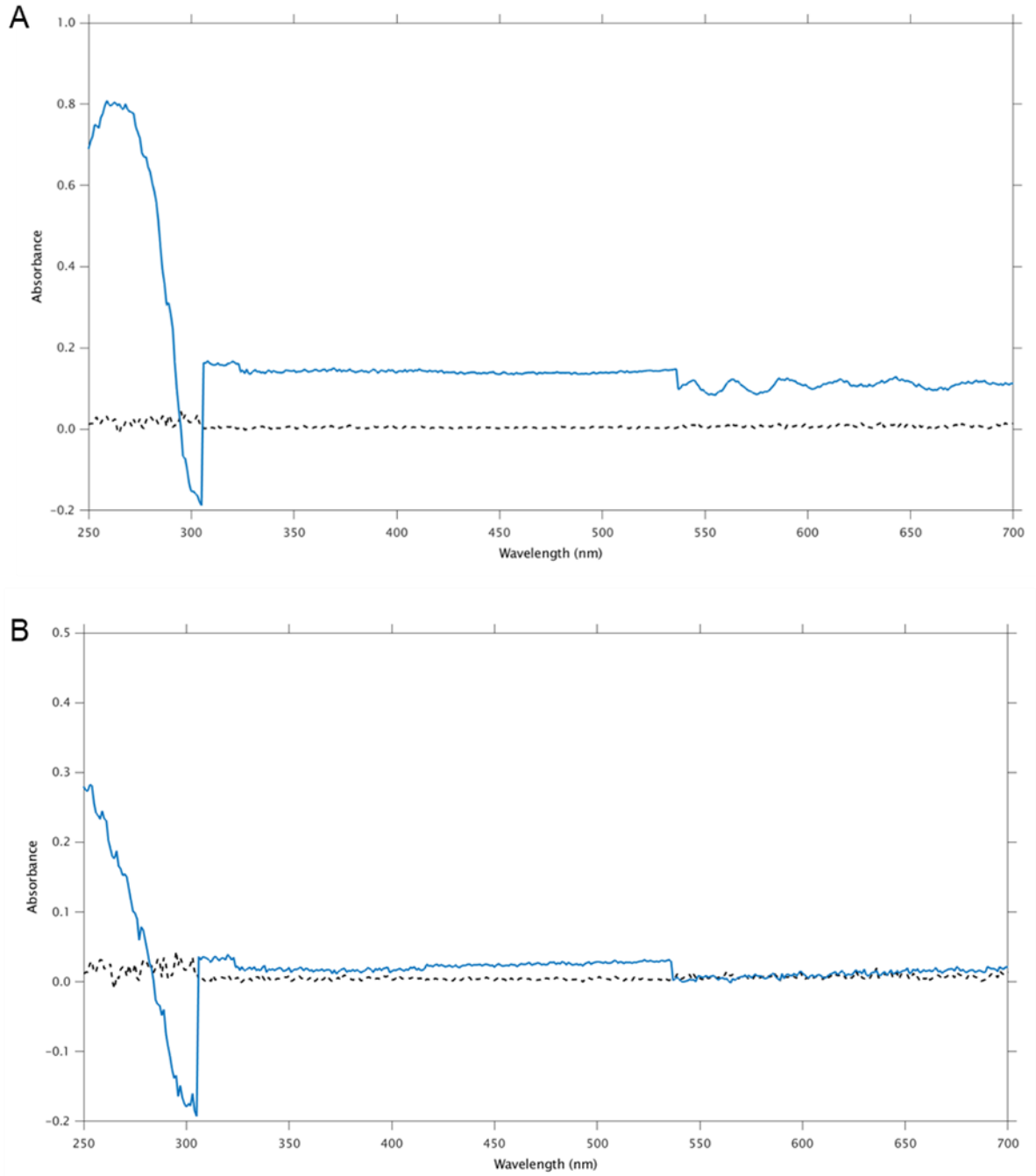

Figure 7.4. **UV-Vis spectrum of the PA2072 $PAS_{310-427}$ domain after expression in Nissle 1917 cells.** The ultraviolet- visible light spectrum of the isolated PA2072 $PAS_{310-427}$ domain was measured between 250 - 700 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 7.5) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of PA2072 $PAS_{310-427}$ was measured in the same way. A) The UV-Vis spectrum of PA2072 $PAS_{310-427}$ size exclusion chromatography (SEC) fractions 45 to 47 (blue solid trace) with the baseline (black dashed trace). B) The UV-Vis spectrum of SEC fractions 51, 52 and 54 (blue solid trace) with the baseline (black dashed trace).

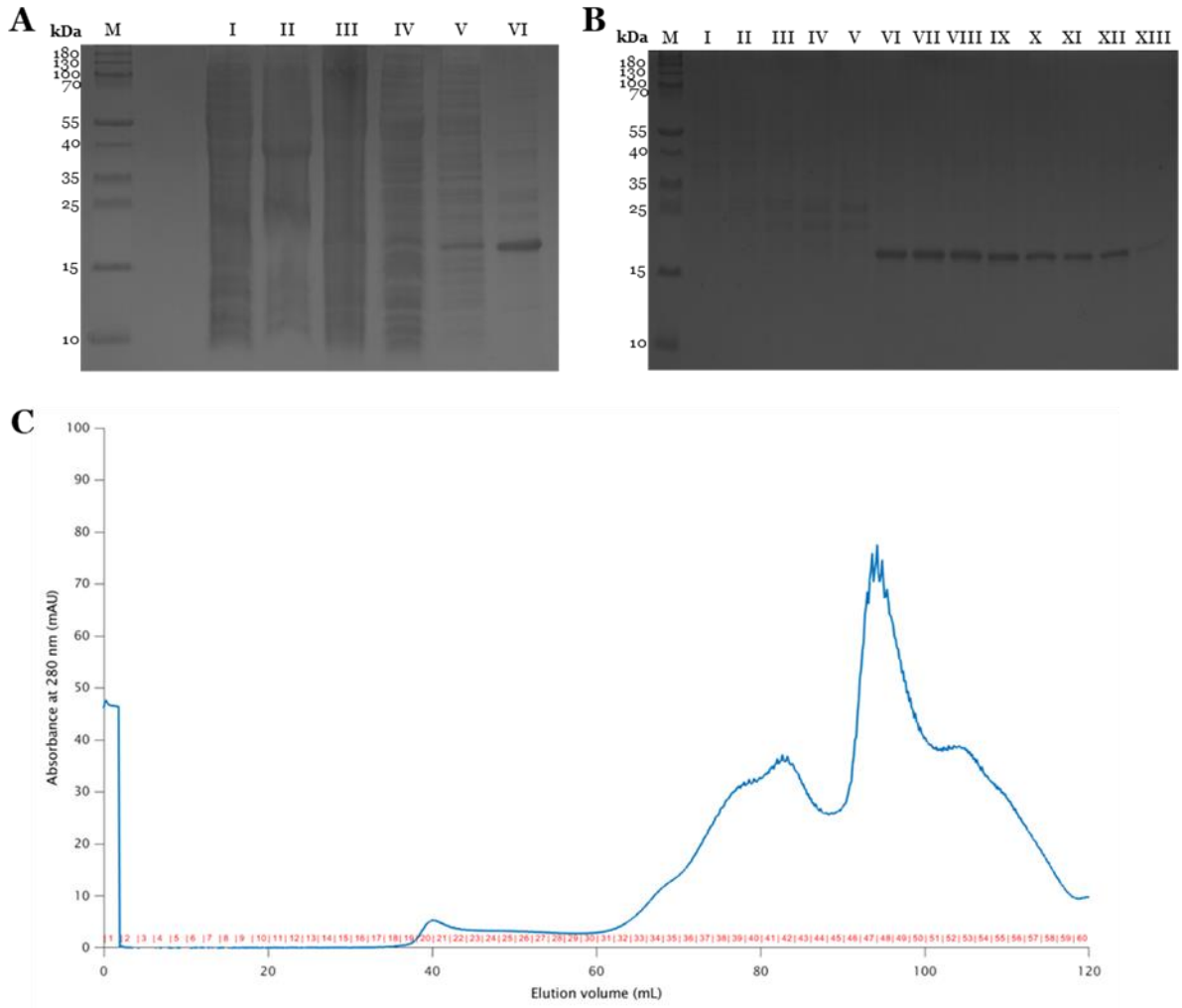

Figure 7.5. **Purification of PA0285 PAS1$_{88-216}$ after expression with Nissle 1917 cells.** A) Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) of PA0285 PAS1$_{88-216}$ (approximately 17 kDa) after purification with immobilised metal affinity chromatography (IMAC) using Ni resin beads. Lane M contains the molecular weight marker PageRuler prestained protein ladder. The sonicated cells, cell pellet after ultracentrifugation, supernatant after ultracentrifugation, flow through from Ni beads, wash from Ni beads and elution from Ni beads are shown in lanes I – VI respectively. B) SDS-PAGE after size exclusion chromatography (SEC) of the PA0285 PAS1$_{88-216}$ IMAC eluate showing fractions 39 – 43, 47 – 50, 53 – 55, 61 in lanes I – XIII respectively. C) SEC trace of PA0285 PAS1$_{88-216}$ measuring absorbance at 280 nm when run at 0.2 mL/min on a HiLoad 16/600 Superdex 75 pg with collected 2 mL fractions indicated in red.

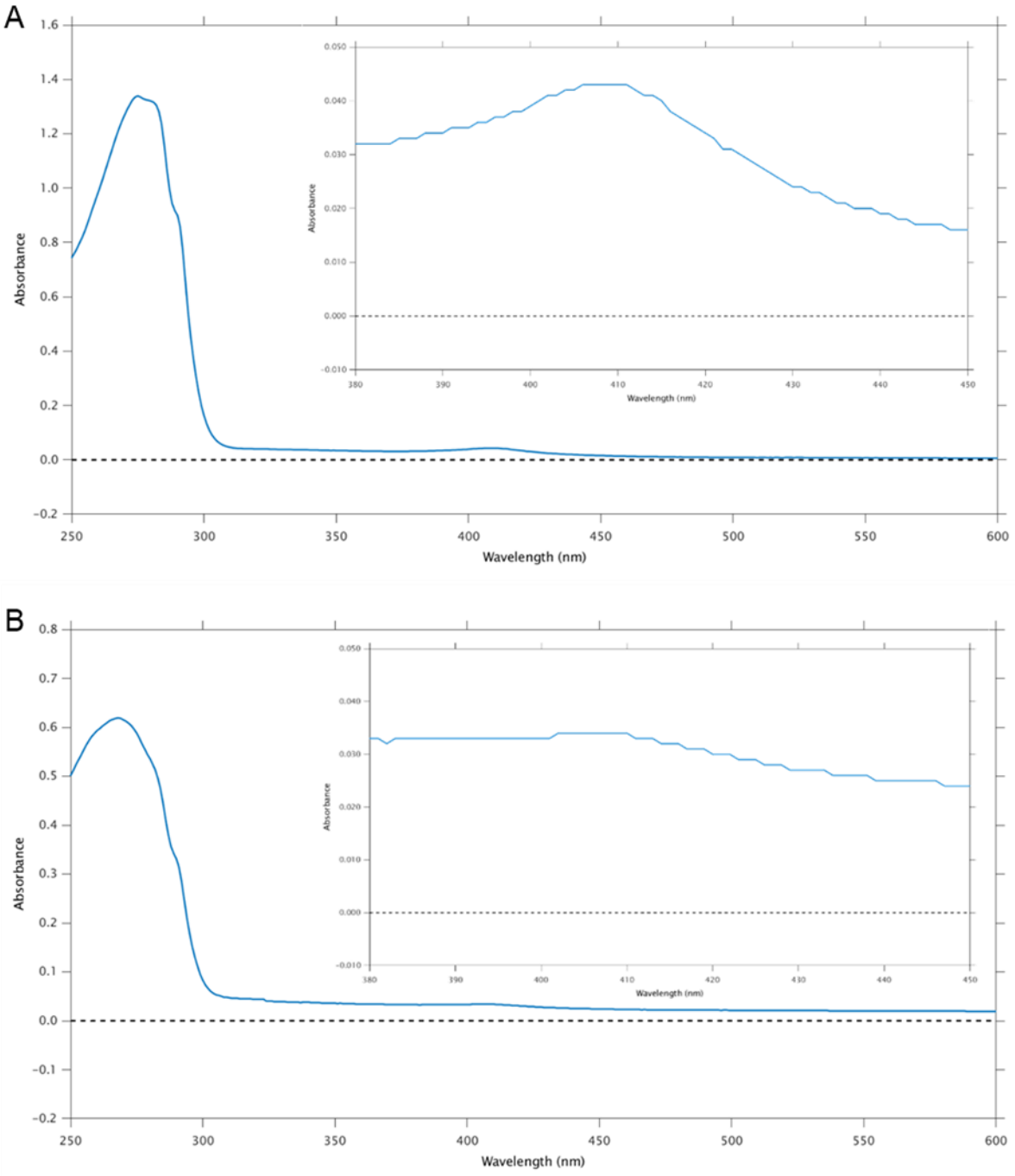

Figure 7.6. **UV-Vis spectrum of the PA0285 PAS1$_{88-216}$ domain after expression in Nissle 1917 cells.** The ultraviolet- visible light spectrum of the isolated PA0285 PAS1$_{88-216}$ domain was measured between 250 - 600 nm using a Shimadzu UV-2600 spectrophotometer. A baseline was first set using 100 µL of buffer the protein sample was purified in (namely: 50 mM Tris and 300 mM NaCl at pH 8.0) with a 1 mL Eppendorf UVette and the 10 mm optical pathlength and measured to record the baseline. Using 100 µL of sample, the UV-Vis spectrum of PA0285 PAS1$_{88-216}$ was measured in the same way. A) The UV-Vis spectrum of PA0285 PAS1$_{88-216}$ size exclusion chromatography (SEC) fractions 47 to 50 (blue solid trace) with the baseline (black dashed trace). B) The UV-Vis spectrum of SEC fractions 52 to 55 (blue solid trace) with the baseline (black dashed trace). The inserts show a magnified image of the UV-Vis spectrums between 380 – 450 nm.

Appendix C

Additional constructs were produced during this work, with the primer sequences and the conditions used during PCR given in Table 7.3 and 7.4 respectively. The additional constructs were produced following the same method as in Section 2.3, except for the two RbdA PAS constructs which were previously produced by Andrew Hutchin using the pOPINF expression vector and a ligation-independent In-Fusion cloning method.

Briefly, primers were designed with specific base extensions (Table 7.3) that are complementary to the pOPINF expression vector. The specific conditions used during PCR to produce these RbdA PAS constructs are given in Table 7.4. The PCR products were purified using AMPure XP magnetic beads (Agencourt/ Beckman), by adding 90 µL of the magnetic bead solution to each 50 µL finished PCR reaction, thoroughly mixing and incubating at room temperature for 5 mins. DNA fragments over 100 bp bind to the magnetic beads which were then pelleted by placing a magnet under the solution for 5 mins before the solution was removed. The DNA-bound magnetic beads were washed twice with the addition of 200 µL 70 % ethanol for 30 secs before removing completely and leaving for 10 - 20 mins to allow for evaporation. DNA fragments were eluted from the magnetic beads with 30 µL of 10 mM Tris at pH 8.0, pelleting the magnetic beads with a magnet and transferring the supernatant. 2 µL of purified PCR product, 1 µL of linearised pOPINF (100 ng) expression vector and 7 µL of water were added to lyophilised In-Fusion enzyme (Clontech) and incubated at 42 °C for 30 mins. The In-Fusion enzyme degrades the specific DNA base extensions, which were included at both termini of the target gene due to the way the primers were designed, into single stranded DNA overhangs that are complementary to the DNA sequences at the ends of the linear expression vector, allowing the vector and target gene to anneal to one another. The reaction was then deactivated with the addition of 40 µL TE buffer and placed on ice. The expression vector with target gene insert becomes circularised within the *E. coli* cells following transformation.

Table 7.3. **The primer sequences used in PCR for producing additional constructs**.

Primers	Sequence (5' – 3')[a]
RbdA 229 FWD[b]	*AAGTTCTGTTTCAGGGCCCG***CAACTGATGCAGCGCCAGG**
RbdA 231 FWD[b]	*AAGTTCTGTTTCAGGGCCCG***ATGCAGCGCCAGGAAGTGG**
RbdA 377 REV[b]	*ATGGTCTAGAAAGCT*TTA**ATCGTGGGTCGCCTGCCAG**
PA2072 427 FWD	aacttCATATG**CACGATGCCCTGACCG**
PA2072 587 REV	aacttAAGCTTTCA**TTCCAAGTAGAAACGCCAGGTA**
PA2072 587 FWD	aacttCATATG**GAAGCCATGGACTCCCATC**
PA2072 845 REV	aacttAAGCTTTCA**CTCGCGCTCCAGCAGA**
PA2072 864 REV2	aacttAAGCTT**TCAAGGCCGGCGCGC**
PA0285 493 FWD	aacttCATATG**CGCGTGCTGACCGCC**
PA5442 284 FWD	aacttCATATG**GAGGCCAGCGAACTGG**
PA5442 408 FWD	aacttCATATG**GAGGCGGCCAGCGAG**
PA5442 412 REV	aacttAAGCTTTCA**CTCGCTGGCCGCCTC**
PA5442 527 REV	aacttAAGCTTTCA**GTGCAGCAGGTGACGC**

[a] Italicised upper case sequences - added for In-Fusion cloning. Lower case sequences - assist in the enzymatic identification of the added restriction enzyme recognition site. Upper case sequences - added restriction enzyme recognition site. Upper case underlined sequence - indicates the addition of a stop codon. Bolded upper case sequences - gene specific sequences. [b] Primers designed by Andrew Hutchin.

Table 7.4. **The specific conditions used in PCR to form the additional constructs**.

Construct	FWD Primer	REV Primer	Final conc. (nM) of each primer	Annealing Tm (°C)	Extension time (sec) for genomic DNA
RbdA PAS[a]	RbdA 229	RbdA 377	600	65	40
RbdA PAS[a]	RbdA 231	RbdA 377	600	67	40
PA2072 PAS-GGDEF-EAL[b]	PA2072 310	PA2072 864	50	72	50
PA2072 GGDEF[c]	PA2072 427	PA2072 587	500	64	15
PA2072 EAL[c]	PA2072 587	PA2072 845	500	65	24
PA2072 EAL[b]	PA2072 587	PA2072 864	100	72	26
PA0285 PAS1-PAS2[b]	PA0285 88	PA0285 333	100	68	23
PA0285 PAS1-PAS2-GGDEF-EAL[b]	PA0285 88	PA0285 760	50	71	61
PA0285 EAL[b]	PA0285 493	PA0285 760	100	71	24
PA5442 PAS1[b]	PA5442 284	PA5442 412	100	69	12
PA5442 PAS2[b]	PA5442 408	PA5442 527	100	69	12
PA5442 PAS1-PAS2[b]	PA5442 284	PA5442 527	500	68	23

[a] Construct designed and produced by Andrew Hutchin using Phusion Flash DNA Polymerase Master Mix (Thermo Scientific). [b] Construct produced with Q5 High-Fidelity DNA Polymerase (NEB). [c] Construct produced with Phusion High-Fidelity DNA Polymerase (NEB).

Except for both of the RbdA PAS constructs, all constructs were cloned into the expression vector pET28a alongside an N-terminal hexa-Histidine tag with a thrombin cleavage site, using the restriction enzymes NdeI and HindIII and then T4 DNA ligase (as described in Section 2.3). Both of the RbdA PAS constructs were cloned by Andrew Hutchin at the Oxford Protein Production Facility - UK into the pOPINF expression vector alongside an N-terminal hexa-Histidine tag with a Rhinovirus 3C protease cleavage site using ligation-independent In-Fusion cloning.

The protein expressions and purifications for all these additional constructs (described in Table 7.5), were carried out as in Section 2.8 and Section 2.9 respectively, except for the RbdA PAS constructs which required 50 µg/mL carbenicillin for pOPINF plasmid selection.

Table 7.5. **The protein expressions and observations made with the additional constructs**.

		SDS-PAGE detectable protein following expression in *E. coli* BL21 (DE3) cells or Nissle 1917 (with pAR1219 vector) cells after 1 mM IPTG induction at an OD_{600} value of 0.6 – 0.8 or an OD_{600} value of 1.2 – 1.5 respectively and growth with 180 rpm at either 17 °C for 18 hrs, 18 °C for 18 hrs or 37 °C for 2 hrs before cell harvesting			Qualitative observations on the protein yields after immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) based on SDS-PAGE gels and SEC traces	
		BL21 (DE3) cells		Nissle 1917 cells (with pAR1219 vector)		
Construct	Range of amino acids	18 °C for 18 hrs	37 °C for 2 hrs	17 °C for 18 hrs	IMAC	SEC
RbdA PAS	229 - 377	No	-	-	-	-
RbdA PAS	231 - 377	Yes	Yes	-	Protein yields were poor. Expression at 18 °C was best - at 37 °C protein was mainly found within the cell pellet.	Protein yields very poor. Most of the protein was found within the precipitant peak.
PA2072 PAS-GGDEF-EAL	310 - 864	Yes	-	-	High yields of protein expression.	Most of the protein was found within the precipitant peak - further optimisation is required.
PA2072 GGDEF	427 - 587	Yes	Yes	-	High protein yields obtained from expression at both 18 °C and 37 °C.	-
PA2072 EAL	587 - 845	Yes	Yes	-	Protein only found within the cell pellet.	-

		SDS-PAGE detectable protein following expression in *E. coli* BL21 (DE3) cells or Nissle 1917 (with pAR1219 vector) cells after 1 mM IPTG induction at an OD_{600} value of 0.6 – 0.8 or an OD_{600} value of 1.2 – 1.5 respectively and growth with 180 rpm at either 17 °C for 18 hrs, 18 °C for 18 hrs or 37 °C for 2 hrs before cell harvesting			Qualitative observations on the protein yields after immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) based on SDS-PAGE gels and SEC traces	
		BL21 (DE3) cells		Nissle 1917 cells (with pAR1219 vector)		
Construct	Range of amino acids	18 °C for 18 hrs	37 °C for 2 hrs	17 °C for 18 hrs	IMAC	SEC
PA2072 EAL	587 - 864	Yes	Yes	-	Expression at 18 °C resulted in good yields – the protein was found only within the cell pellet with 37 °C expression.	Good protein yields – some contaminants still found with the target protein.
PA0285 PAS1-PAS2	88 - 333	Yes	Yes	-	Good protein yields – highest protein yield obtained from expression at 18 °C. Eluted proteins were yellow in colour – probable flavin cofactor binding.	-
PA0285 PAS1-PAS2-GGDEF-EAL	88 - 760	Yes	-	-	Good protein yield. Eluted protein was slightly yellow in colour – probable flavin cofactor binding.	-
PA0285 EAL	493 - 760	Yes	-	-	Very high protein yield.	-
PA5442 PAS1	284 - 412	Yes	Yes	-	High protein yields – expression at 18 °C resulted in higher protein yield.	-

		SDS-PAGE detectable protein following expression in *E. coli* BL21 (DE3) cells or Nissle 1917 (with pAR1219 vector) cells after 1 mM IPTG induction at an OD_{600} value of 0.6 – 0.8 or an OD_{600} value of 1.2 – 1.5 respectively and growth with 180 rpm at either 17 °C for 18 hrs, 18 °C for 18 hrs or 37 °C for 2 hrs before cell harvesting			Qualitative observations on the protein yields after immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) based on SDS-PAGE gels and SEC traces	
		BL21 (DE3) cells		Nissle 1917 cells (with pAR1219 vector)		
Construct	Range of amino acids	18 °C for 18 hrs	37 °C for 2 hrs	17 °C for 18 hrs	IMAC	SEC
PA5442 PAS2	408 - 527	No	No	No	-	-
PA5442 PAS1-PAS2	284 - 527	-	-	No	-	-

SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis; *E. coli*, *Escherichia coli*; IPTG, Isopropyl β-d-1-thiogalactopyranoside.

References

1. Tolker-Nielsen T. Pseudomonas aeruginosa biofilm infections: From molecular biofilm biology to new treatment possibilities. *Apmis* 2014;122:1-51.
2. Lam J, Vaughan S, Parkins MD. Tobramycin Inhalation Powder (TIP): An Efficient Treatment Strategy for the Management of Chronic Pseudomonas aeruginosa Infection in Cystic Fibrosis. *Clinical medicine insights. Circulatory, respiratory and pulmonary medicine* 2013;7:61-77.
3. Lubamba B, Dhooghe B, Noel S, et al. Cystic fibrosis: Insight into CFTR pathophysiology and pharmacotherapy. *Clinical Biochemistry* 2012;45(15):1132-44.
4. Worlitzsch D, Tarran R, Ulrich M, et al. Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *Journal of Clinical Investigation* 2002;109(3):317-25.
5. Webb JS, Thompson LS, James S, et al. Cell death in Pseudomonas aeruginosa biofilm development. *Journal of Bacteriology* 2003;185(15):4585-92.
6. de la Fuente-Nunez C, Reffuveille F, Fernandez L, et al. Bacterial biofilm development as a multicellular adaptation: antibiotic resistance and new therapeutic strategies. *Current Opinion in Microbiology* 2013;16(5):580-89.
7. Rybtke M, Hultqvist LD, Givskov M, et al. Pseudomonas aeruginosa Biofilm Infections: Community Structure, Antimicrobial Tolerance and Immune Response. *Journal of Molecular Biology* 2015;427(23):3628-45.
8. Bjarnsholt T, Jensen PO, Fiandaca MJ, et al. Pseudomonas aeruginosa Biofilms in the Respiratory Tract of Cystic Fibrosis Patients. *Pediatric Pulmonology* 2009;44(6):547-58.
9. Gellatly SL, Hancock REW. Pseudomonas aeruginosa: new insights into pathogenesis and host defenses. *Pathogens and Disease* 2013;67(3):159-73.
10. Edmondson C, Davies JC. Current and future treatment options for cystic fibrosis lung disease: latest evidence and clinical implications. *Therapeutic Advances in Chronic Disease* 2016;7(3):170-83.
11. Schwensen HF, Moser C, Perch M, et al. Pseudomonas aeruginosa antibody response in cystic fibrosis decreases rapidly following lung transplantation. *Journal of Cystic Fibrosis* 2020;19(4):587-94.
12. Southern KW, Clancy JP, Ranganathan S. Aerosolized agents for airway clearance in cystic fibrosis. *Pediatric Pulmonology* 2019;54(6):858-64.
13. Cuyx S, De Boeck K. Treating the Underlying Cystic Fibrosis Transmembrane Conductance Regulator Defect in Patients with Cystic Fibrosis. *Seminars in Respiratory and Critical Care Medicine* 2019;40(6):762-74.
14. Park JH, Ousingsawat J, Cabrita I, et al. TMEM16A deficiency: a potentially fatal neonatal disease resulting from impaired chloride currents. *Journal of medical genetics* 2020.
15. Tseng BS, Zhang W, Harrison JJ, et al. The extracellular matrix protects Pseudomonas aeruginosa biofilms by limiting the penetration of tobramycin. *Environmental Microbiology* 2013;15(10):2865-78.
16. Kierek-Pearscon K, Karatan E. Biofilm development in bacteria. In: Laskin AI, Bennett JW, Gadd GM (eds.) *Advances in Applied Microbiology, Vol 57*. San Diego: Elsevier Academic Press Inc; 2005 p79-111.
17. Klingel M, Stanojevic S, Tullis E, et al. Oral Azithromycin and Response to Pulmonary Exacerbations Treated with Intravenous Tobramycin in Children with Cystic Fibrosis. *Annals of the American Thoracic Society* 2019;16(7):861-67.
18. Samson C, Tamalet A, Thien HV, et al. Long-term effects of azithromycin in patients with cystic fibrosis. *Respiratory Medicine* 2016;117:1-6.

19. Konstan MW, VanDevanter DR, Sawicki GS, et al. Association of High-Dose Ibuprofen Use, Lung Function Decline, and Long-Term Survival in Children with Cystic Fibrosis. *Annals of the American Thoracic Society* 2018;15(4):485-93.
20. Shah PN, Marshall-Batty KR, Smolen JA, et al. Antimicrobial Activity of Ibuprofen against Cystic Fibrosis-Associated Gram-Negative Pathogens. *Antimicrobial Agents and Chemotherapy* 2018;62(3):22.
21. Clancy JP. Rapid therapeutic advances in CFTR modulator science. *Pediatric Pulmonology* 2018;53:4-11.
22. Loukou I, Moustaki M, Plyta M, et al. Longitudinal changes in lung function following initiation of lumacaftor/ivacaftor combination. *Journal of Cystic Fibrosis* 2020;19(4):534-39.
23. Geller DE, Nasr SZ, Piggott S, et al. Tobramycin Inhalation Powder in Cystic Fibrosis Patients: Response by Age Group. *Respiratory Care* 2014;59(3):388-98.
24. Elson EC, Mermis J, Polineni D, et al. Aztreonam Lysine Inhalation Solution in Cystic Fibrosis. *Clinical Medicine Insights-Circulatory Respiratory and Pulmonary Medicine* 2019;13:10.
25. Webb JS, Givskov M, Kjelleberg S. Bacterial biofilms: prokaryotic adventures in multicellularity. *Current Opinion in Microbiology* 2003;6(6):578-85.
26. Sauer K, Camper AK, Ehrlich GD, et al. Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. *Journal of Bacteriology* 2002;184(4):1140-54.
27. Wolska KI, Grudniak AM, Rudnicka Z, et al. Genetic control of bacterial biofilms. *Journal of Applied Genetics* 2016;57(2):225-38.
28. Kaplan JB. Biofilm Dispersal: Mechanisms, Clinical Implications, and Potential Therapeutic Uses. *Journal of Dental Research* 2010;89(3):205-18.
29. Roilides E, Simitsopoulou M, Katragkou A, et al. How Biofilms Evade Host Defenses. *Microbiology Spectrum* 2015;3(3):10.
30. Mazza MG. The physics of biofilms-an introduction. *Journal of Physics D-Applied Physics* 2016;49(20):24.
31. Arora DP, Hossain S, Xu YM, et al. Nitric Oxide Regulation of Bacterial Biofilms. *Biochemistry* 2015;54(24):3717-28.
32. Prakash B, Veeregowda BM, Krishnappa G. Biofilms: A survival strategy of bacteria. *Current Science* 2003;85(9):1299-307.
33. Tolker-Nielsen T. Biofilm Development. *Microbiology Spectrum* 2015;3(2):12.
34. Nielsen AT, Tolker-Nielsen T, Barken KB, et al. Role of commensal relationships on the spatial structure of a surface-attached microbial consortium. *Environmental Microbiology* 2000;2(1):59-68.
35. Tolker-Nielsen T, Brinch UC, Ragas PC, et al. Development and dynamics of Pseudomonas sp biofilms. *Journal of Bacteriology* 2000;182(22):6482-89.
36. Davey ME, Caiazza NC, O'Toole GA. Rhamnolipid surfactant production affects biofilm architecture in Pseudomonas aeruginosa PAO1. *Journal of Bacteriology* 2003;185(3):1027-36.
37. Stoodley P, Debeer D, Lewandowski Z. Liquid Flow In Biofilm Systems. *Applied and Environmental Microbiology* 1994;60(8):2711-16.
38. Donlan RM. Biofilm formation: A clinically relevant microbiological process. *Clinical Infectious Diseases* 2001;33(8):1387-92.
39. Wu H, Moser C, Wang HZ, et al. Strategies for combating bacterial biofilm infections. *International Journal of Oral Science* 2015;7(1):1-7.
40. Mah TFC, O'Toole GA. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology* 2001;9(1):34-39.
41. Stewart PS, Costerton JW. Antibiotic resistance of bacteria in biofilms. *Lancet* 2001;358(9276):135-38.
42. Flemming HC, Wingender J, Szewzyk U, et al. Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* 2016;14(9):563-75.

43. Wagner VE, Iglewski BH. P. aeruginosa Biofilms in CF Infection. *Clinical Reviews in Allergy & Immunology* 2008;35(3):124-34.
44. Stewart PS. Mechanisms of antibiotic resistance in bacterial biofilms. *International Journal of Medical Microbiology* 2002;292(2):107-13.
45. Barraud N, Kardak BG, Yepuri NR, et al. Cephalosporin-3 '-diazeniumdiolates: Targeted NO-Donor Prodrugs for Dispersing Bacterial Biofilms. *Angewandte Chemie-International Edition* 2012;51(36):9057-60.
46. Walters MC, Roe F, Bugnicourt A, et al. Contributions of antibiotic penetration, oxygen limitation, and low metabolic activity to tolerance of Pseudomonas aeruginosa biofilms to ciprofloxacin and tobramycin. *Antimicrobial Agents and Chemotherapy* 2003;47(1):317-23.
47. Anwar H, Vanbiesen T, Dasgupta M, et al. Interaction of Biofilm Bacteria with Antibiotics in a Novel in vitro Chemostat System. *Antimicrobial Agents and Chemotherapy* 1989;33(10):1824-26.
48. Debeer D, Srinivasan R, Stewart PS. Direct Measurement of Chlorine Penetration into Biofilms During Disinfection. *Applied and Environmental Microbiology* 1994;60(12):4339-44.
49. Vrany JD, Stewart PS, Suci PA. Comparison of recalcitrance to ciprofloxacin and levofloxacin exhibited by Pseudomonas aeruginosa biofilms displaying rapid-transport characteristics. *Antimicrobial Agents and Chemotherapy* 1997;41(6):1352-58.
50. Singh R, Ray P, Das A, et al. Penetration of antibiotics through Staphylococcus aureus and Staphylococcus epidermidis biofilms. *Journal of Antimicrobial Chemotherapy* 2010;65(9):1955-58.
51. Barraud N, Kjelleberg S, Rice SA. Dispersal from Microbial Biofilms. *Microbiology Spectrum* 2015;3(6):15.
52. Barraud N, Hassett DJ, Hwang SH, et al. Involvement of nitric oxide in biofilm dispersal of Pseudomonas aeruginosa. *Journal of Bacteriology* 2006;188(21):7344-53.
53. Anderson GG, O'Toole GA. Innate and induced resistance mechanisms of bacterial biofilms. *Bacterial Biofilms* 2008;322:85-105.
54. Lewis K. Persister Cells. In: Gottesman S, Harwood CS (eds.) *Annual Review of Microbiology, Vol 64, 2010*. Palo Alto: Annual Reviews; 2010 p357-72.
55. Ellison CK, Kan JB, Dillard RS, et al. Obstruction of pilus retraction stimulates bacterial surface sensing. *Science* 2017;358(6362):535-38.
56. Romling U, Galperin MY, Gomelsky M. Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiology and Molecular Biology Reviews* 2013;77(1):1-52.
57. Schmidt AJ, Ryjenkov DA, Gomelsky M. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *Journal of Bacteriology* 2005;187(14):4774-81.
58. Barraud N, Schleheck D, Klebensberger J, et al. Nitric Oxide Signaling in Pseudomonas aeruginosa Biofilms Mediates Phosphodiesterase Activity, Decreased Cyclic Di-GMP Levels, and Enhanced Dispersal. *Journal of Bacteriology* 2009;191(23):7333-42.
59. Ryan RP, Lucey J, O'Donovan K, et al. HD-GYP domain proteins regulate biofilm formation and virulence in Pseudomonas aeruginosa. *Environmental Microbiology* 2009;11(5):1126-36.
60. Rinaldo S, Paiardini A, Stelitano V, 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. *Journal of Bacteriology* 2015;197(8):1525-35.
61. Caly DL, Bellini D, Walsh MA, et al. Targeting Cyclic di-GMP Signalling: A Strategy to Control Biofilm Formation? *Current Pharmaceutical Design* 2015;21(1):12-24.
62. Jenal U, Malone J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annual Review of Genetics*. Palo Alto: Annual Reviews; 2006 p385-407.
63. Schaap P. Cyclic Di-nucleotide Signaling Enters the Eukaryote Domain. *Iubmb Life* 2013;65(11):897-903.

64. Ryjenkov DA, Tarutina M, Moskvin OV, et al. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. *Journal of Bacteriology* 2005;187(5):1792-98.
65. Romling U, Liang ZX, Dow JM. Progress in Understanding the Molecular Basis Underlying Functional Diversification of Cyclic Dinucleotide Turnover Proteins. *Journal of Bacteriology* 2017;199(5):16.
66. Jenal U, Reinders A, Lori C. Cyclic di-GMP: second messenger extraordinaire. *Nature reviews. Microbiology* 2017.
67. Christen B, Christen M, Paul R, et al. Allosteric control of cyclic di-GMP signaling. *Journal of Biological Chemistry* 2006;281(42):32015-24.
68. Wassmann P, Chan C, Paul R, et al. Structure of BeF3--modified response regulator PleD: Implications for diguanylate cyclase activation, catalysis, and feedback inhibition. *Structure* 2007;15(8):915-27.
69. Chan C, Paul R, Samoray D, et al. Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101(49):17084-89.
70. Zahringer F, Lacanna E, Jenal U, et al. Structure and Signaling Mechanism of a Zinc-Sensory Diguanylate Cyclase. *Structure* 2013;21(7):1149-57.
71. Yang CY, Chin KH, Chuah MLC, et al. The structure and inhibition of a GGDEF diguanylate cyclase complexed with (c-di-GMP)(2) at the active site. *Acta Crystallographica Section D-Biological Crystallography* 2011;67:997-1008.
72. Tarnawski M, Barends TRM, Schlichting I. Structural analysis of an oxygen-regulated diguanylate cyclase. *Acta Crystallographica Section D-Biological Crystallography* 2015;71:2158-77.
73. Phippen CW, Mikolajek H, Schlaefli HG, et al. Formation and dimerization of the phosphodiesterase active site of the Pseudomonas aeruginosa MorA, a bi-functional c-di-GMP regulator. *Febs Letters* 2014;588(24):4631-36.
74. Roy AB, Petrova OE, Sauer K. The Phosphodiesterase DipA (PA5017) Is Essential for Pseudomonas aeruginosa Biofilm Dispersion. *Journal of Bacteriology* 2012;194(11):2904-15.
75. Christensen LD, van Gennip M, Rybtke MT, et al. Clearance of Pseudomonas aeruginosa Foreign-Body Biofilm Infections through Reduction of the Cyclic Di-GMP Level in the Bacteria. *Infection and Immunity* 2013;81(8):2705-13.
76. Tchigvintsev A, Xu XH, Singer A, et al. Structural Insight into the Mechanism of c-di-GMP Hydrolysis by EAL Domain Phosphodiesterases. *Journal of Molecular Biology* 2010;402(3):524-38.
77. Yadav M, Pal K, Sen U. Structures of c-di-GMP/cGAMP degrading phosphodiesterase VcEAL: identification of a novel conformational switch and its implication. *Biochemical Journal* 2019;476:3333-53.
78. Barends TRM, Hartmann E, Griese JJ, et al. Structure and mechanism of a bacterial light-regulated cyclic nucleotide phosphodiesterase. *Nature* 2009;459(7249):1015-18.
79. Tamayo R, Tischler AD, Camilli A. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. *Journal of Biological Chemistry* 2005;280(39):33324-30.
80. Bellini D, Horrell S, Hutchin A, et al. Dimerisation induced formation of the active site and the identification of three metal sites in EAL-phosphodiesterases. *Scientific Reports* 2017;7:11.
81. Sundriyal A, Massa C, Samoray D, et al. Inherent Regulation of EAL Domain-catalyzed Hydrolysis of Second Messenger Cyclic di-GMP. *Journal of Biological Chemistry* 2014;289(10):6978-90.
82. Lovering AL, Capeness MJ, Lambert C, et al. 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* 2011;2(5):8.
83. Bellini D, Caly DL, McCarthy Y, et al. Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Molecular Microbiology* 2014;91(1):26-38.

84. Stelitano V, Giardina G, Paiardini A, et al. C-di-GMP Hydrolysis by Pseudomonas aeruginosa HD-GYP Phosphodiesterases: Analysis of the Reaction Mechanism and Novel Roles for pGpG. *Plos One* 2013;8(9):13.
85. Ryan RP, McCarthy Y, Andrade M, et al. Cell-cell signal-dependent dynamic interactions between HD-GYP and GGDEF domain proteins mediate virulence in Xanthomonas campestris. *Proceedings of the National Academy of Sciences of the United States of America* 2010;107(13):5989-94.
86. Romling U, Gomelsky M, Galperin MY. C-di-GMP: the dawning of a novel bacterial signalling system. *Molecular Microbiology* 2005;57(3):629-39.
87. Kulesekara H, Lee V, Brencic A, et al. Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3 '-5 ')-cyclic-GMP in virulence. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(8):2839-44.
88. Massie JP, Reynolds EL, Koestler BJ, et al. Quantification of high-specificity cyclic diguanylate signaling. *Proceedings of the National Academy of Sciences of the United States of America* 2012;109(31):12746-51.
89. Privett BJ, Broadnax AD, Bauman SJ, et al. Examination of bacterial resistance to exogenous nitric oxide. *Nitric Oxide-Biology and Chemistry* 2012;26(3):169-73.
90. Li Y, Heine S, Entian M, et al. NO-Induced Biofilm Dispersion in Pseudomonas aeruginosa Is Mediated by an MHYT Domain-Coupled Phosphodiesterase. *Journal of Bacteriology* 2013;195(16):3531-42.
91. Zaitseva J, Granik V, Belik A, et al. Effect of nitrofurans and NO generators on biofilm formation by Pseudomonas aeruginosa PAO1 and Burkholderia cenocepacia 370. *Research in Microbiology* 2009;160(5):353-57.
92. Nguyen TK, Selvanayagam R, Ho KKK, et al. Co-delivery of nitric oxide and antibiotic using polymeric nanoparticles. *Chemical Science* 2016;7(2):1016-27.
93. Craven M, Kasper SH, Canfield MJ, et al. Nitric oxide-releasing polyacrylonitrile disperses biofilms formed by wound-relevant pathogenic bacteria. *Journal of Applied Microbiology* 2016;120(4):1085-99.
94. Howlin RP, Cathie K, Hall-Stoodley L, et al. Low-Dose Nitric Oxide as Targeted Anti-biofilm Adjunctive Therapy to Treat Chronic Pseudomonas aeruginosa Infection in Cystic Fibrosis. *Molecular Therapy* 2017;25(9):2104-16.
95. Sauer K, Cullen MC, Rickard AH, et al. Characterization of nutrient-induced dispersion in Pseudomonas aeruginosa PAO1 biofilm. *Journal of Bacteriology* 2004;186(21):7312-26.
96. Schleheck D, Barraud N, Klebensberger J, et al. Pseudomonas aeruginosa PAO1 Preferentially Grows as Aggregates in Liquid Batch Cultures and Disperses upon Starvation. *Plos One* 2009;4(5):15.
97. Denninger JW, Marletta MA. Guanylate cyclase and the (NO)-N-./cGMP signaling pathway. *Biochimica Et Biophysica Acta-Bioenergetics* 1999;1411(2-3):334-50.
98. Pan J, Yuan H, Zhang X, et al. Probing the Molecular Mechanism of Human Soluble Guanylate Cyclase Activation by NO in vitro and in vivo. *Scientific Reports* 2017;7:14.
99. Derbyshire ER, Marletta MA. Structure and Regulation of Soluble Guanylate Cyclase. In: Kornberg RD (ed.) *Annual Review of Biochemistry, Vol 81*. Palo Alto: Annual Reviews; 2012 p533-59.
100. Montfort WR, Wales JA, Weichsel A. Structure and Activation of Soluble Guanylyl Cyclase, the Nitric Oxide Sensor. *Antioxidants & Redox Signaling* 2017;26(3):107-21.
101. Carlson HK, Vance RE, Marletta MA. H-NOX regulation of c-di-GMP metabolism and biofilm formation in Legionella pneumophila. *Molecular Microbiology* 2010;77(4):930-42.
102. Liu N, Xu YM, Hossain S, et al. Nitric Oxide Regulation of Cyclic di-GMP Synthesis and Hydrolysis in Shewanella woodyi. *Biochemistry* 2012;51(10):2087-99.
103. Hossain S, Boon EM. Discovery of a Novel Nitric Oxide Binding Protein and Nitric-Oxide-Responsive Signaling Pathway in Pseudomonas aeruginosa. *ACS infectious diseases* 2017;3(6):454-61.

104. Borziak K, Zhulin IB. FIST: a sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes. *Bioinformatics* 2007;23(19):2518-21.
105. Galperin MY, Gaidenko TA, Mulkidjanian AY, et al. MHYT, a new integral membrane sensor domain. *Fems Microbiology Letters* 2001;205(1):17-23.
106. Hay ID, Remminghorst U, Rehm BHA. MucR, a Novel Membrane-Associated Regulator of Alginate Biosynthesis in Pseudomonas aeruginosa. *Applied and Environmental Microbiology* 2009;75(4):1110-20.
107. Moglich A, Ayers RA, Moffat K. Structure and Signaling Mechanism of Per-ARNT-Sim Domains. *Structure* 2009;17(10):1282-94.
108. Hefti MH, Francoijs KJ, de Vries SC, et al. The PAS fold - A redefinition of the PAS domain based upon structural prediction. *European Journal of Biochemistry* 2004;271(6):1198-208.
109. Rojas-Pirela M, Rigden DJ, Michels PA, et al. Structure and function of Per-ARNT-Sim domains and their possible role in the life-cycle biology of Trypanosoma cruzi. *Molecular and Biochemical Parasitology* 2018;219:52-66.
110. Henry JT, Crosson S. Ligand-Binding PAS Domains in a Genomic, Cellular, and Structural Context. In: Gottesman S, Harwood CS (eds.) *Annual Review of Microbiology, Vol 65*. Palo Alto: Annual Reviews; 2011 p261-86.
111. Gong WM, Hao B, Mansy SS, et al. Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction. *Proceedings of the National Academy of Sciences of the United States of America* 1998;95(26):15177-82.
112. Taylor BL, Zhulin IB. PAS domains: Internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews* 1999;63(2):479-506.
113. Liu YC, Machuca MA, Beckham SA, et al. Structural basis for amino-acid recognition and transmembrane signalling by tandem Per-Arnt-Sim (tandem PAS) chemoreceptor sensory domains. *Acta Crystallographica Section D-Structural Biology* 2015;71:2127-36.
114. Reinelt S, Hofmann E, Gerharz T, et al. The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *Journal of Biological Chemistry* 2003;278(40):39189-96.
115. Heintz U, Meinhart A, Winkler A. Multi-PAS domain-mediated protein oligomerization of PpsR from Rhodobacter sphaeroides. *Acta Crystallographica Section D-Biological Crystallography* 2014;70:863-76.
116. Rao F, Ji QA, Soehano I, et al. Unusual Heme-Binding PAS Domain from YybT Family Proteins. *Journal of Bacteriology* 2011;193(7):1543-51.
117. Anantharaman V, Koonin EV, Aravind L. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *Journal of Molecular Biology* 2001;307(5):1271-92.
118. Key J, Hefti M, Purcell EB, et al. Structure of the redox sensor domain of Azotobacter vinelandii NifL at atomic resolution: Signaling, dimerization, and mechanism. *Biochemistry* 2007;46(12):3614-23.
119. Mann TH, Shapiro L. Integration of cell cycle signals by multi-PAS domain kinases. *Proceedings of the National Academy of Sciences of the United States of America* 2018;115(30):7166-73.
120. Ortmayer M, Lafite P, Menon BRK, et al. An oxidative N-demethylase reveals PAS transition from ubiquitous sensor to enzyme. *Nature* 2016;539(7630):593-97.
121. Key J, Moffat K. Crystal structures of deoxy and CO-bound bjFixLH reveal details of ligand recognition and signaling. *Biochemistry* 2005;44(12):4627-35.
122. Gong WM, Hao B, Chan MK. New mechanistic insights from structural studies of the oxygen-sensing domain of Bradyrhizobium japonicum FixL. *Biochemistry* 2000;39(14):3955-62.
123. Hao B, Isaza C, Arndt J, et al. Structure-based mechanism of O-2 sensing and ligand discrimination by the FixL heme domain of Bradyrhizobium japonicum. *Biochemistry* 2002;41(43):12952-58.

124. Yamawaki T, Ishikawa H, Mizuno M, et al. Regulatory Implications of Structural Changes in Tyr201 of the Oxygen Sensor Protein FixL. *Biochemistry* 2016;55(29):4027-35.
125. Anantharaman V, Aravind L. The CHASE domain: a predicted ligand-binding module in plant cytokinin receptors and other eukaryotic and bacterial receptors. *Trends in Biochemical Sciences* 2001;26(10):579-82.
126. Mougel C, Zhulin IB. CHASE: an extracellular sensing domain common to transmembrane receptors from prokaryotes, lower eukaryotes and plants. *Trends in Biochemical Sciences* 2001;26(10):582-84.
127. Zhulin IB, Nikolskaya AN, Galperin MY. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. *Journal of Bacteriology* 2003;185(1):285-94.
128. Hengge R, Galperin MY, Ghigo JM, et al. Systematic Nomenclature for GGDEF and EAL Domain-Containing Cyclic Di-GMP Turnover Proteins of Escherichia coli. *Journal of Bacteriology* 2016;198(1):7-11.
129. An SW, Wu JE, Zhang LH. Modulation of Pseudomonas aeruginosa Biofilm Dispersal by a Cyclic-Di-GMP Phosphodiesterase with a Putative Hypoxia-Sensing Domain. *Applied and Environmental Microbiology* 2010;76(24):8160-73.
130. Cai Y-M, Hutchin A, Craddock J, et al. Differential impact on motility and biofilm dispersal of closely related phosphodiesterases in Pseudomonas aeruginosa. *Scientific Reports* 2020;10(1):6232.
131. Liu C, Liew CW, Wong YH, et al. Insights into Biofilm Dispersal Regulation from the Crystal Structure of the PAS-GGDEF-EAL Region of RbdA from Pseudomonas aeruginosa. *Journal of Bacteriology* 2018;200(3):19.
132. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research* 2018;46(D1):493-96.
133. Mitchell AL, Attwood TK, Babbitt PC, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* 2019;47(1):351-60.
134. Hutchin AJ, (2017). *Molecular mechanisms of nitric oxide induced Pseudomonas aeruginosa biofilm dispersal*. University of Southampton, School of Biological Sciences, PhD Thesis, 143-146.
135. Varnado CL, Goodwin DC. System for the expression of recombinant hemoproteins in Escherichia coli. *Protein Expression and Purification* 2004;35(1):76-83.
136. Fiege K, Querebillo CJ, Hildebrandt P, et al. Improved Method for the Incorporation of Heme Cofactors into Recombinant Proteins Using Escherichia coli Nissle 1917. *Biochemistry* 2018;57(19):2747-55.
137. Kiefhaber T, Rudolph R, Kohler HH, et al. Protein Aggregation *in vitro* and *in vivo:* A Quantitative Model Of The Kinetic Competition Between Folding and Aggregation. *Bio-Technology* 1991;9(9):825-29.
138. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology* 2014;5:17.
139. Heintzmann R, Ficz G. Breaking the resolution limit in light microscopy. *Briefings in Functional Genomics & Proteomics* 2006;5(4):289-301.
140. Rhodes G. *Crystallography Made Crystal Clear*. Third Edition ed: Academic Press - Elsevier; 2006.
141. Papiz MZ, Winter G. X-Ray Crystallography, Biomolecular Structure Determination Methods. *Encyclopedia of Spectroscopy and Spectrometry, 3rd Edition, Vol 4: S-Z* 2017:640-47.
142. Rosner B, Finizio S, Koch F, et al. Soft x-ray microscopy with 7 nm resolution. *Optica* 2020;7(11):1602-08.
143. Cowtan K. Phase Problem in X-ray Crystallography, and Its Solution. In: Encyclopedia of life sciences: Wiley, Chichester, 2003.

144. Bragg WH, Bragg WL. The reflection of X-rays by crystals. *Proc. R. Soc. Lond. A* 1913;88(605):428-38.
145. Smyth MS, Martin JHJ. x Ray crystallography. *Journal of Clinical Pathology-Molecular Pathology* 2000;53(1):8-14.
146. Taylor GL. Experimental phasing and radiation damage Introduction to phasing. *Acta Crystallographica Section D-Structural Biology* 2010;66:325-38.
147. Winter G, Waterman DG, Parkhurst JM, et al. DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D-Structural Biology* 2018;74:85-97.
148. Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D-Structural Biology* 2011;67:235-42.
149. Evans P. Scaling and assessment of data quality. *Acta Crystallographica Section D-Structural Biology* 2006;62:72-82.
150. Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D-Biological Crystallography* 2011;67:282-92.
151. Vagin A, Teplyakov A. MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* 1997;30:1022-25.
152. Navarro M, De N, Bae N, et al. Structural Analysis of the GGDEF-EAL Domain-Containing c-di-GMP Receptor FimX. *Structure* 2009;17(8):1104-16.
153. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Structural Biology* 1997;53:240-55.
154. Emsley P, Lohkamp B, Scott WG, et al. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* 2010;66:486-501.
155. Joosten RP, Long F, Murshudov GN, et al. The PDB_REDO server for macromolecular structure model optimization. *Iucrj* 2014;1:213-20.
156. Burley SK, Berman HM, Bhikadiya C, et al. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research* 2019;47(D1):D520-D28.
157. Christen M, Christen B, Folcher M, et al. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *Journal of Biological Chemistry* 2005;280(35):30829-37.
158. Golovanov AP, Hautbergue GM, Wilson SA, et al. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society* 2004;126(29):8933-39.
159. El Mouali Y, Kim H, Ahmad I, et al. Stand-Alone EAL Domain Proteins Form a Distinct Subclass of EAL Proteins Involved in Regulation of Cell Motility and Biofilm Formation in Enterobacteria. *Journal of Bacteriology* 2017;199(18):17.
160. Rao F, Yang Y, Qi YN, et al. Catalytic mechanism of cyclic Di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from Pseudomonas aeruginosa. *Journal of Bacteriology* 2008;190(10):3622-31.
161. Navarro M, Newell PD, Krasteva PV, et al. Structural Basis for c-di-GMP-Mediated Inside-Out Signaling Controlling Periplasmic Proteolysis. *Plos Biology* 2011;9(2):21.
162. Newell PD, Monds RD, O'Toole GA. LapD is a bis-(3 ',5 ')-cyclic dimeric GMP-binding protein that regulates surface attachment by Pseudomonas fluorescens Pf0-1. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106(9):3461-66.
163. Chin KH, Kuo WT, Yu YJ, et al. Structural polymorphism of c-di-GMP bound to an EAL domain and in complex with a type II PilZ-domain protein. *Acta Crystallographica Section D-Biological Crystallography* 2012;68:1380-92.
164. Hutchin AJ, (2017). *Molecular mechanisms of nitric oxide induced Pseudomonas aeruginosa biofilm dispersal*. University of Southampton, School of Biological Sciences, PhD Thesis, 89-91.

165. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 2019;47(1):636-41.
166. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* 2019;47(1):402-07.
167. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology* 1999;292(2):195-202.
168. Finn RD, Attwood TK, Babbitt PC, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Research* 2017;45(1):190-99.
169. Hothorn M, Dabi T, Chory J. Structural basis for cytokinin recognition by Arabidopsis thaliana histidine kinase 4. *Nature Chemical Biology* 2011;7(11):766-68.
170. Chen P, Jiao X, Zhang YK, et al. The crystal structure of the phytopathogenic bacterial sensor PcrK reveals different cytokinin recognition mechanism from the plant sensor AHK4. *Journal of Structural Biology* 2019;208(1):69-76.
171. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM (ed.) *The Proteomics Protocols Handbook*: Humana Press; 2005 p571 - 607.
172. Yang F, Xia XA, Lei HY, et al. Hemin Binds to Human Cytoplasmic Arginyl-tRNA Synthetase and Inhibits Its Catalytic Activity. *Journal of Biological Chemistry* 2010;285(50):39437-46.
173. Piano V, Palfey BA, Mattevi A. Flavins as Covalent Catalysts: New Mechanisms Emerge. *Trends in Biochemical Sciences* 2017;42(6):457-69.
174. Evans EW, Dodson CA, Maeda K, et al. Magnetic field effects in flavoproteins and related systems. *Interface Focus* 2013;3(5):17.
175. Ghisla S, Massey V, Lhoste JM, et al. Fluorescence and Optical Characteristics of Reduced Flavines and Flavoproteins. *Biochemistry* 1974;13(3):589-97.
176. Derewenda ZS. The use of recombinant methods and molecular engineering in protein crystallization. *Methods* 2004;34(3):354-63.
177. Trakhanov S, Quiocho FA. Influence of divalent cations in protein crystallization. *Protein Science* 1995;4(9):1914-19.
178. Rajavel M, Gopal B. Analysis of multiple crystal forms of Bacillus subtilis BacB suggests a role for a metal ion as a nucleant for crystallization. *Acta Crystallographica Section D-Structural Biology* 2010;66:635-39.
179. Mueller I. Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallographica Section D-Structural Biology* 2017;73:79-92.
180. Hoeppner A, Schmitt L, Smits SHJ. Proteins and Their Ligands: Their Importance and How to Crystallize Them. In: Ferreira S (ed.) *Advanced Topics on Crystal Growth*: IntechOpen; 2013 p3-42.
181. Smits SHJ, Mueller A, Schmitt L, et al. A structural basis for substrate selectivity and stereoselectivity in octopine dehydrogenase from Pecten maximus. *Journal of Molecular Biology* 2008;381(1):200-11.
182. Hassell AM, An G, Bledsoe RK, et al. Crystallization of protein-ligand complexes. *Acta Crystallographica Section D-Biological Crystallography* 2007;63:72-79.
183. Carson M, Johnson DH, McDonald H, et al. His-tag impact on structure. *Acta Crystallographica Section D-Biological Crystallography* 2007;63:295-301.
184. Tajika Y, Sakai N, Tamura T, et al. Crystal structure of hypothetical protein PH0828 from Pyrococcus horikoshii. *Proteins-Structure Function and Bioinformatics* 2004;57(4):862-65.
185. Adrian-Segarra JM, Lorchner H, Braun T, et al. Identification of Functional Protein Regions Through Chimeric Protein Construction. *Jove-Journal of Visualized Experiments* 2019(143):10.
186. Shanmugaratnam S, Eisenbeis S, Hocker B. A highly stable protein chimera built from fragments of different folds. *Protein Engineering Design & Selection* 2012;25(11):699-703.

187. Trincao J, Warren A, Aller P, et al. VMXm: a new sub-micron beamline for macromolecular crystallography at Diamond Light Source. *Acta Crystallographica a-Foundation and Advances* 2015;71:191-91.
188. Warren A, Evans G, Trincao J, et al. VMXm: A new micro/nanofocus protein crystallography beamline at Diamond. *Acta Crystallographica a-Foundation and Advances* 2018;74:10-10.
189. Sanchez-Weatherby J, Sandy J, Mikolajek H, et al. VMXi: a fully automated, fully remote, high-flux insitu macromolecular crystallography beamline. *Journal of Synchrotron Radiation* 2019;26:291-301.
190. Forse GJ, Ram N, Banatao DR, et al. Synthetic symmetrization in the crystallization and structure determination of CelA from Thermotoga maritima. *Protein Science* 2011;20(1):168-78.
191. Banatao DR, Cascio D, Crowley CS, et al. An approach to crystallizing. proteins by synthetic symmetrization. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103(44):16230-35.
192. Dalton DA, delCastillo LD, Kahn ML, et al. Heterologous expression and characterization of soybean cytosolic ascorbate peroxidase. *Archives of Biochemistry and Biophysics* 1996;328(1):1-8.
193. Wissbrock A, George AAP, Brewitz HH, et al. The molecular basis of transient heme-protein interactions: analysis, concept and implementation. *Bioscience Reports* 2019;39:11.
194. Kearns DB. A field guide to bacterial swarming motility. *Nature Reviews Microbiology* 2010;8(9):634-44.
195. Huang BX, Whitchurch CB, Mattick JS. FimX, a multidomain protein connecting environmental signals to twitching motility in Pseudomonas aeruginosa. *Journal of Bacteriology* 2003;185(24):7068-76.
196. Semmler ABT, Whitchurch CB, Mattick JS. A re-examination of twitching motility in Pseudomonas aeruginosa. *Microbiology-Sgm* 1999;145:2863-73.
197. Tala L, Fineberg A, Kukura P, et al. Pseudomonas aeruginosa orchestrates twitching motility by sequential control of type IV pili movements. *Nature Microbiology* 2019;4(5):774-80.
198. Kohler T, Curty LK, Barja F, et al. Swarming of Pseudomonas aeruginosa is dependent on cell-to-cell signaling and requires flagella and pili. *Journal of Bacteriology* 2000;182(21):5990-96.
199. Murray TS, Kazmierczak BI. Pseudomonas aeruginosa exhibits sliding motility in the absence of type IV pili and flagella. *Journal of Bacteriology* 2008;190(8):2700-08.
200. Merritt JH, Ha DG, Cowles KN, et al. Specific Control of Pseudomonas aeruginosa Surface-Associated Behaviors by Two c-di-GMP Diguanylate Cyclases. *Mbio* 2010;1(4):9.
201. Richter AM, Possling A, Malysheva N, et al. Local c-di-GMP Signaling in the Control of Synthesis of the E. coli Biofilm Exopolysaccharide pEtN-Cellulose. *Journal of Molecular Biology* 2020.
202. Repik A, Rebbapragada A, Johnson MS, et al. PAS domain residues involved in signal transduction by the Aer redox sensor of Escherichia coli. *Molecular Microbiology* 2000;36(4):806-16.
203. Qi YN, Rao F, Luo Z, et al. A Flavin Cofactor-Binding PAS Domain Regulates c-di-GMP Synthesis in AxDGC2 from Acetobacter xylinum. *Biochemistry* 2009;48(43):10275-85.
204. Becker DF, Zhu WD, Moxley MA. Flavin Redox Switching of Protein Functions. *Antioxidants & Redox Signaling* 2011;14(6):1079-91.
205. Basudhar D, Ridnour LA, Cheng R, et al. Biological signaling by small inorganic molecules. *Coordination Chemistry Reviews* 2016;306:708-23.
206. Nicastro GG, Kaihami GH, Pulschen AA, et al. c-di-GMP-related phenotypes are modulated by the interaction between a diguanylate cyclase and a polar hub protein. *Scientific Reports* 2020;10(1):3077.
207. Larentis AL, Nicolau JFMQ, Esteves GdS, et al. Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor. *BMC research notes* 2014;7:671-71.

208. Boshtam M, Shahreza HK, Feizollahzadeh S, et al. Expression and purification of biologically active recombinant rabbit monocyte chemoattractant protein1 in Escherichia coli. *Fems Microbiology Letters* 2018;365(9).
209. Fazaeli A, Golestani A, Lakzaei M, et al. Expression optimization, purification, and functional characterization of cholesterol oxidase from Chromobacterium sp. DS1. *Plos One* 2019;14(2):15.
210. Tan E, Rao F, Pasunooti S, et al. Solution Structure of the PAS Domain of a Thermophilic YybT Protein Homolog Reveals a Potential Ligand-binding Site. *Journal of Biological Chemistry* 2013;288(17):11949-59.
211. Fruk L, Kuo CH, Torres E, et al. Apoenzyme Reconstitution as a Chemical Tool for Structural Enzymology and Biotechnology. *Angewandte Chemie-International Edition* 2009;48(9):1550-74.
212. Aliverti A, Curti B, Vanoni MA. Identifying and quantitating FAD and FMN in simple and in iron-sulfur-containing flavoproteins. *Flavoprotein Protocols* 1999;131:9-23.
213. Lewis JA, Escalante-Semerena JC. The FAD-dependent tricarballylate dehydrogenase (TcuA) enzyme of Salmonella enterica converts tricarballylate into cis-aconitate. *Journal of Bacteriology* 2006;188(15):5479-86.
214. Hefti MH, Vervoort J, van Berkel WJH. Deflavination and reconstitution of flavoproteins - Tackling fold and function. *European Journal of Biochemistry* 2003;270(21):4227-42.
215. Li XX, Xu XD, Dan YY, et al. The Factors during Protein Crystallization: A Review. *Crystallography Reports* 2008;53(7):1261-66.
216. Russo Krauss I, Merlino A, Vergara A, et al. An Overview of Biological Macromolecule Crystallization. *International Journal of Molecular Sciences* 2013;14(6):11643-91.
217. Derewenda ZS, Vekilov PG. Entropy and surface engineering in protein crystallization. *Acta Crystallographica Section D-Structural Biology* 2006;62:116-24.
218. Moon AF, Mueller GA, Zhong XJ, et al. A synergistic approach to protein crystallization: Combination of a fixed-arm carrier with surface entropy reduction. *Protein Science* 2010;19(5):901-13.
219. Vekilov PG, Feeling-Taylor AR, Yau ST, et al. Solvent entropy contribution to the free energy of protein crystallization. *Acta Crystallographica Section D-Biological Crystallography* 2002;58:1611-16.
220. Derewenda ZS. Rational protein crystallization by mutational surface engineering. *Structure* 2004;12(4):529-35.
221. Cooper DR, Boczek T, Grelewska K, et al. Protein crystallization by surface entropy reduction: optimization of the SER strategy. *Acta Crystallographica Section D-Structural Biology* 2007;63:636-45.
222. Loll PJ, Xu P, Schmidt JT, et al. Enhancing ubiquitin crystallization through surface-entropy reduction. *Acta Crystallographica Section F-Structural Biology Communications* 2014;70:1434-42.
223. Goldschmidt L, Cooper DR, Derewenda ZS, et al. Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Science* 2007;16(8):1569-76.