Characterisation of 'degenerate' cyclic di-GMP signalling proteins of *Escherichia coli*

Nicola Sarah Whiting, B. Sc. Hons.

A thesis submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Molecular Biology and Biotechnology

The University of Sheffield

Sheffield

September 2014

Abstract

Since its discovery 25 years ago, the cyclic di-GMP signalling pathway has been an ever-expanding area of scientific research. This pathway, ubiquitous in Bacteria, regulates a range of bacterial phenotypes in response to various environmental cues. Two groups of enzymes make and break the secondary messenger c-di-GMP, these being the diguanylate cyclase (DGC) and phoshodiesterase (PDE) enzymes respectively. Both classes of enzymes contain a consensus motif of amino acid residues, necessary for the catalytic activity of the proteins; specifically a GGDEF motif for DGC proteins and an EAL or HD-GYP motif for PDE enzymes. However proteins exist which contain non-conserved sequences at these sites and are therefore deemed likely to be catalytically inactive. The roles of these so called 'degenerate' proteins are gradually being elucidated and have been found to be regulatory proteins, initiating a variety of cellular effects. This thesis focuses on the degenerate DGC and PDE enzymes in *E. coli* K-12, of which there are four. At the start of this project, two of these (YcgF and YdhA) had been characterised, whilst YdiV and YeaI were largely uncharacterised.

Analysis of the YeaI protein confirmed the protein to be a catalytically inactive diguanylate cyclase, however the poor solubility of the protein vastly restricted characterisation of the protein. YdiV was shown to be a catalytically inactive phosphodiesterase. Chacterisation of the protein showed YdiV to bind to the *E. coli* transcriptional factor FlhDC, where it acted as an anti-FlhDC factor. The FlhDC factor is the master regulator of flagellar gene transcription, shown to bind to DNA and cause gene transcription. However, upon YdiV binding to FlhDC, DNA was shown to be dissociated thereby halting gene expression. Kinetic and thermodynamic analysis of the YdiV:FlhDC binding interaction revealed the proteins to have a nM binding affinity. *In vitro* transcription and EMSA analysis revealed the existence of a transitional or intermediate YdiV:FlhDC:DNA species which retains the FlhDC transcriptional activation ability. Phenotypically, overproduction of the *ydiV* gene caused a vast decrease in flagella production and FliC detection. The regulation of this YdiV:FlhDC interaction is currently unknown. Nucleotide binding assays have revealed a possible interaction between YdiV and c-di-GMP, which potentially could regulate the activity of YdiV and thereby the YdiV:FlhDC interaction. Furthermore, *in vitro* analysis has identified a YdiV:FlhDC:FliT species, which if produced *in vivo*, would add another level of regulation and complexity to the flagellar gene expression. Thus, YdiV has a gene regulatory role in *E. coli*, repressing flagella biosynthesis via its interactions with FlhDC.

Acknowledgements

I would first and foremost like to thank my supervisors, Professor Jeff Green and the late Professor Pete Artymiuk, both of whom were sources of constant support and encouragement. Without their project management and scientific advice I most certainly would not be where I am today and I am so grateful for everything they have both done to help me.

Huge thanks go to everyone in F10, past and present, for teaching me everything I know and for all your help and advice. A special thanks to Dr Melissa Lacey for her patience and support with anything and everything c-di-GMP related. Thanks also go to Dr Arthur Moir for N-terminal sequencing, Bioserve for antibody production, Thomas Curran for providing chemostat samples and Dr Sveta Sedelnikova for gel filtration. I would also like to thank Dr Andrea Hounslow for carrying out the NMR experiments and all her advice, Chris Hill for EM and Dr Srdjan Vitovski for all his expertise planning the BLItz experiments. Thanks to my project student Rachel Hickie for all her efforts with partial proteolysis assays.

Mum and Dad, thank you for your endless support and love throughout this time, you have been so understanding and I can't thank you enough.

I have been so lucky to work in a place where my colleagues are some of my best friends. To my old housemates, Tom, Ash and Arthur: we have got through this all together! Four years, four Ph.D's and hundreds of memories. A special mention also goes to Serina, Laura and Katie, who have been brilliant friends over the past few years. Thank you all for your friendships, the encouragement you have given me and for making work a very sociable and fun place to be.

A final thank you goes to Adam who has been at the forefront of my Ph.D highs and lows over the past few years. You have picked up the pieces and cheered me up more times that I can remember. Thank you for all the support, love and encouragement you have given me, you have been my absolute rock.

Contents

Abstract.....**i**
Acknowledgments.....**ii**
List of figures.....**viii**
List of tables.....**xii**
List of Abbreviations.....**xiii**
1.0 Introduction.....**1**
1.1 Mechanisms of signal transduction in *Bacteria*.....**1**
1.2 Cyclic di-GMP signalling.....**3**
1.3 Cellular effects of c-di-GMP signalling.........**4**
1.3.1 Bacterial motility and its regulation by c-di-GMP signalling.....5
1.3.2 Biofilm formation and the c-di-GMP regulation of this phenotype..9
1.4 Cyclic di-GMP effector proteins.....**12**
1.4.1 The PilZ domain class of effector proteins.....12
1.4.2 The role of degenerate GGDEF/EAL domain proteins as c-di-GMP effector proteins.....13
1.4.3 Riboswitches as c-di-GMP effector proteins.....14
1.4.4 The detection of unrelated proteins as effector proteins.....14
1.5 C-di-GMP sensory domains.....**15**
1.5.1 PAS domains as sensory elements.....15
1.5.2 The presence of GAF as a sensory domain.....16
1.5.3 REC as a commonly found sensory domain.....16
1.5.4 Less common sensory domains.....17
1.6 The biosynthesis and degradation of c-di-GMP.....**17**
1.6.1 Diguanylate cyclase proteins.....19
1.6.2 Cyclic di-GMP-specific phosphodiesterase proteins.....22
1.7 The multiplicity of GGDEF/ EAL domains.....**23**
1.8 The presence of composite and degenerate domains.....**25**
1.8.1 The roles of degenerate GGDEF/EAL domains.....26
1.9 The cyclic di-GMP pathway in *E. coli*.....**27**
1.9.1 *Escherichia coli* strain K-12.....27
1.9.2 Cyclic di-GMP proteins in *E. coli* K-12.....28
1.10 Aims of the project.....**31**
2.0 Materials and Methods.....**33**
2.1 Strains and Plasmids.....**33**
2.2 Growth of *E.coli*.....**34**
2.2.1 Growth media.....34
2.2.2 *Escherichia coli* growth.....35
2.2.3 Measurement of bacterial growth.....36
2.2.4 Supplementation of media.....36
2.2.5 Storage of strains.....36
2.2.6 Making electrically competent cells.....36
2.2.7 Transformation of electrically competent cells.....37
2.3 DNA techniques.....**37**

2.3.1 Genomic DNA isolation....37
2.3.2 Primer design....37
2.3.3 Polymerase chain reaction....39
2.3.4 Plasmid purification....39
2.3.5 Agarose gels....39
2.3.6 Restriction digestion of DNA....40
2.3.7 DNA isolation and purification from an agarose gel....40
2.3.8 Quantification of DNA concentration....41
2.3.9 DNA ligation reactions....41
2.3.10 DNA purification....41
2.3.11 Screening of transformants....41
2.4 Gene expression analysis....42
2.4.1 β-galactosidase assays....42
2.5 Microscopy techniques....43
2.5.1 Transmission electron microscopy....43
2.6 Protein methods....44
2.6.1 Protein overproduction....44
2.6.2 Cell breakage....44
2.6.3 Affinity purification....45
2.6.4 Anion exchange chromatography....46
2.6.5 Resolubilisation of the insoluble pellet....47
2.6.6 Thrombin cleavage....47
2.6.7 Buffer exchange and concentrating of protein samples....48
2.6.8 Quantification of protein concentrations....49
2.6.9 SDS-PAGE....50
2.6.10 Western blot....52
2.6.11 Antibody production....53
2.6.12 Glutaraldehyde crosslinking assays....54
2.6.13 Gel filtration....54
2.6.14 N-terminal sequencing....55
2.6.15 Crystallisation trials....56
2.7 Protein:nucleotide interactions....56
2.7.1 Enzymatic reactions....56
2.7.2 Partial proteolysis....57
2.7.3 Nuclear Magnetic Resonance....58
2.7.4 Fluorescence spectroscopy....59
2.7.5 Pull-down assays with nucleotides....60
2.8 Protein:protein interactions....61
2.8.1 Pull-down assays....61
2.8.2 Competition assay pull-down....62
2.8.3 Bio-Layer interferometry for YdiV:FlhDC interaction....63
2.8.4 Isothermal titration calorimetry....64
2.9 Protein:DNA interactions....65
2.9.1 Radioactive labelling of DNA....65
2.9.2 Electromobility shift assays....65

2.9.3 *In vitro* transcription assays....67
2.9.4 Autoradiography....68
2.9.5 DNA:protein kinetics....68
2.10 Analysis software....70
2.10.1 Image J quantification....70
2.10.2 DNA and protein alignments....70
2.10.3 Phylogenetic analysis....70
3.0 Characterisation of the 'degenerate' *Escherichia coli* diguanylate cyclase protein YeaI....71
3.1 Introduction....71
3.2 Production of *yeaI*$_{GGDEF}$ overproduction plasmids....72
3.3 Optimisation of YeaI$_{GGDEF}$ overproduction....76
3.4 Solubility trials of YeaI....79
3.5 Purification of His$_6$-YeaI$_{GGDEF}$....82
3.6 Analysis of YeaI$_{GGDEF}$ enzymatic activity....82
3.7 Nucleotide binding by YeaI$_{GGDEF}$....84
3.8 Oligomeric analysis of YeaI$_{GGDEF}$....93
3.9 Optimising the yield of soluble YeaI$_{GGDEF}$....95
3.9.1 Ligation of YeaI$_{GGDEF}$ into pCOLD TF vector....97
3.9.2 Overproduction and solubility trials....99
3.9.3 Solubility trials to optimise overproduction of TF-YeaI$_{GGDEF}$....100
3.9.4 Optimisation of YeaI$_{GGDEF}$ purification....100
3.9.5 Large scale purification....106
3.10 Discussion....111
4.0 Characterisation of the *Escherichia coli* 'degenerate' phosphodiesterase protein YdiV....115
4.1 Introduction....115
4.2 Cloning, over-expression and purification of YdiV....118
4.2.1 Cloning of the *ydiV* gene into an over-expression plasmid....118
4.2.2 Optimisation of YdiV overproduction....120
4.2.3 Purification of YdiV –His$_6$ fusion protein....122
4.3 Oligomeric analysis of YdiV....125
4.4 Confirmation of the 'degenerate nature' of the YdiV EAL domain....128
4.5 Determination of YdiV and 'ligand' binding....129
4.5.1 Partial proteolysis analysis....129
4.5.2 Fluorescence spectroscopy of YdiV....135
4.5.3 Nuclear Magnetic Resonance analysis of YdiV in the presence of nucleotides....139
4.5.4 Cloning and over-production of the fusion protein Clp-His$_6$....144
4.5.5 Nuclear Magnetic Resonance analysis of Clp in the presence of c-di-GMP....148
4.6 Characterisation of the effects of *ydiV* over-expression and mutations on bacterial motility....150
4.6.1 Determination of FliC expression levels according to *ydiV* expression....150

4.6.2 Flagella changes as a result of *ydiV* over-expression or mutations....152
4.7 Identification of *ydiV* induction conditions....154
4.7.1 YdiV protein production analysis....154
4.7.2 Gene expression analysis of *ydiV* in the presence of quorum sensing molecules....155
4.7.3 Promoter alignment....160
4.7.4 Phylogenetic analysis....167
4.8 Discussion....171
5.0 Characterisation and functional investigation of the *Escherichia coli* YdiV and FlhDC interaction....177
5.1 Introduction....177
5.2 Construction and over-production of the YdiV and FlhDC proteins....179
5.2.1 Cloning of FlhDC into over-expression plasmids....179
5.2.2 Over production of the FlhDC protein....182
5.2.3 Purification of His-tagged proteins....184
5.3 Determination of YdiV:FlhDC binding....186
5.4 Characterisation of the YdiV and FlhDC binding interaction....188
5.4.1 Characterisation of YdiV binding to the immobilised FlhDC protein....189
5.4.2 Characterisation of FlhDC binding to the immobilised YdiV protein....192
5.4.3 Characterisation of YdiV and FlhDC interactions in solution....194
5.5 Production of YdiV and FlhDC antibodies....195
5.5.1 Bleed specificity tests....198
5.5.2 Terminal bleed antibody sensitivity tests....201
5.6 Determination of YdiV and FlhDC concentrations at different growth rates....201
5.6.1 Analysis of YdiV expression levels....201
5.6.2 Analysis of FlhDC expression levels....201
5.7 DNA-binding capacity of FlhDC in the presence of YdiV....206
5.8 Characterisation of binding kinetics of DNA:FlhDC interaction....210
5.8.1 Control experiments....211
5.8.2 Kinetics of the FlhDC:DNA interaction....213
5.8.3 Kinetics of the YdiV:DNA interaction....217
5.9 Effect of YdiV concentration on *in vitro* transcription of *fliD*....219
5.9.1 Deciphering the role of YdiV in the FlhDC:DNA complex....220
5.10 Investigation of ligand interaction to the YdiV:FlhDC complex....222
5.10.1 Protein and nucleotide pull-down assays....225
5.10.2 Fluorescence spectroscopy....227
5.10.3 Determination of the DNA-binding capacity of FlhDC in the presence of nucleotides....232
5.10.4 Analysis of the effect, if any, of nucleotides upon FlhDC mediated *in vitro* transcription....234

5.10.5 Concluding remarks regarding nucleotide binding to the FlhDC or FlhDC:YdiV complex....236
5.11 Competition assays with FliT/FliZ....236
5.11.1 Cloning of FliT and FliZ overproduction plasmids....236
5.11.2 Overproduction of GST-FliT and GST-FliZ....239
5.11.3 Control assays....239
5.11.4 Competition assays....241
5.12 Discussion....243
6.0 Final discussion....250

References....259

List of figures

Figure 1.1: Schematic diagram of bacterial flagella components 6
Figure 1.2: Schematic diagram of the cascade and regulation of flagella gene synthesis in *E. coli* 7
Figure 1.3: Schematic diagram showing the stages of biofilm formation 10
Figure 1.4: The mechanism of cyclic di-GMP turnover 18
Figure 1.5: Characterisation of active and inactive GGDEF and EAL domain proteins 20
Figure 1.6: Schematic diagram of the 29 GGDEF/EAL domain proteins present in *E. coli* K-12 29
Figure 2.1: SDS-PAGE protein markers 51
Figure 3.1: Structure and degeneracy of the YeaI protein 73
Figure 3.2: Schematic plasmid maps for the insertion of *yeaI*$_{GGDEF}$ into pET32a and pET21a 75
Figure 3.3: Optimisation of overproduction conditions for the YeaI$_{GGDEF}$-His$_6$ fusion protein in the *E. coli* BL21 (λDE3)/pGS2341 strain 77
Figure 3.4: Optimisation of overproduction conditions for the His$_6$-YeaI$_{GGDEF}$ fusion protein in the *E. coli* BL21 (λDE3)/pGS2263 strain 78
Figure 3.5: Optimisation of solubility conditions for the YeaI$_{GGDEF}$-His$_6$ fusion protein in the *E. coli* BL21 (λDE3)/pGS2341 strain 80
Figure 3.6: Optimisation of solubility conditions for the His$_6$-YeaI$_{GGDEF}$ fusion protein in the *E. coli* BL21 (λDE3)/ pGS2263 strain 81
Figure 3.7: HiTrap chelating chromatography to purify His$_6$-YeaI$_{GGDEF}$ 83
Figure 3.8: DGC/PDE assay controls, using active DGC and PDE enzymes 85
Figure 3.9: Confirmation of YeaI as a 'degenerate' or inactive DGC by HPLC analysis 86
Figure 3.10: Trypsin cleavage optimisation assay for the cleavage of YeaI$_{GGDEF}$ 88
Figure 3.11: Partial proteolysis of YeaI$_{GGDEF}$ in the presence of nucleotides 89
Figure 3.12: Quantification of YeaI$_{GGDEF}$ partial proteolysis products upon nucleotide addition 91
Figure 3.13: Oligomerisation analysis of YeaI$_{GGDEF}$ 94
Figure 3.14: Gel filtration trace for YeaI$_{GGDEF}$ 96
Figure 3.15: Schematic plasmid maps for the insertion of *yeaI*$_{GGDEF}$ into the pCOLD TF vector 98
Figure 3.16: Optimisation of overproduction conditions for the TF-YeaI$_{GGDEf}$ fusion protein in *E. coli* BL21 (λDE3) cells 101
Figure 3.17: Solubility optimisation for TF-YeaI$_{GGDEF}$ 102
Figure 3.18: Purification of TF-YeaI$_{GGDEF}$, with thrombin incubation at 4^{o}C 104
Figure 3.19: Purification of TF-YeaI$_{GGDEF}$, with thrombin incubation at 20^{o}C 104
Figure 3.20: Purification of TF-YeaI$_{GGDEF}$, with calcium addition to buffers 105
Figure 3.21: HiTrap chelating chromatography as an initial step to purify TF-YeaI$_{GGDEF}$ 107
Figure 3.22: Thrombin cleavage as a secondary step for YeaI$_{GGDEF}$ purification from TF-YeaI$_{GGDEF}$ 108

Figure 4.1: Primary and secondary structure of YdiV 116
Figure 4.2: Schematic plasmid map for the insertion of *ydiV* into pET21a 119
Figure 4.3: Optimisation of overproduction conditions for the YdiV protein in *E. coli* BL21 (λDE3)/pGS2354 cells 121
Figure 4.4: Identification of soluble YdiV following overproduction 123
Figure 4.5: HiTrap chelating chromatography as a one-step purification of YdiV-His_6 fusion protein 124
Figure 4.6: Oligomeric analysis of YdiV 126
Figure 4.7: Gel filtration analysis of YdiV 127
Figure 4.8: Control reactions for determining YdiV PDE activity, using known active DGC and PDE enzymes 130
Figure 4.9: Confirmation of YdiV as a 'degenerate' or inactive PDE by HPLC analysis 131
Figure 4.10: Trypsin cleavage optimisation assay for the cleavage of YdiV protein 133
Figure 4.11: Partial proteolysis of YdiV in the presence of nucleotides 134
Figure 4.12: Fluorescence spectroscopic analysis of YdiV in the presence of various nucleotides at 26^{o}C 137
Figure 4.13: Fluorescence spectroscopic analysis of YdiV and c-di-GMP binding at various constant temperatures 138
Figure 4.14: Proton NMR spectra focused on the nucleotide chemical shift in the presence and absence of YdiV 141
Figure 4.15: Proton NMR spectra in the 6.3-9.1 ppm region, for YdiV in the presence and absence of nucleotides 143
Figure 4.16: Schematic plasmid map for the insertion of *clp* into pET21a 145
Figure 4.17: HiTrap chelating chromatography as a one-step purification of Clp-His_6 fusion protein 147
Figure 4.18: Proton NMR spectrum in the 5.71-6.15 ppm region for c-di-GMP, in the presence and absence of Clp 149
Figure 4.19: Proton NMR spectrum in the 6.3-9.1 ppm region, for Clp in the presence and absence of c-di-GMP 149
Figure 4.20: Detection of FliC production in various strains, including Δ*ydiV* and *ydiV* over-expression strains 151
Figure 4.21: Electron Microscopy visualisation of flagella for various strains, including Δ*ydiV* and *ydiV* over-expression strains 153
Figure 4.22: Comparison of YdiV protein expression levels at various growth rates 156
Figure 4.23: Schematic of SdiA induced YdiV expression 157
Figure 4.24: β-Galactosidase assay results using the *E. coli ydiV-lacZ* strain 159
Figure 4.25: β-Galactosidase assay results using the *E. coli ydiV-lacZ*/pET28a:*slyA* strain 161
Figure 4.26: Promoter alignments for the *ydiV* upstream sequences in *E. coli* and *S. enterica* strains 162
Figure 4.27: Phylogram of *E. coli* and *S. enterica* strains based on similarities in *ydiV* promoter sequences 168
Figure 4.28: Cladogram to reveal the relatedness of *E. coli* and *S. enterica* strains based on similarities in *ydiV* promoter sequences 169

Figure 5.1: Schematic plasmid maps for the insertion of *flhDC* into pET21a constructs 180
Figure 5.2: Optimisation of overproduction conditions for the $(FLAG)_3$-FlhDC protein in *E. coli* BL21 (λDE3)/pGS2382 cells 183
Figure 5.3: Optimisation of overproduction conditions for the FlhDC-His_6 protein in *E. coli* BL21 (λDE3)/pGS2441 cells 183
Figure 5.4: HiTrap chelating chromatography as a one-step purification of FlhDC-His_6 protein 185
Figure 5.5: Binding interaction between the *E. coli* proteins FlhDC and YdiV 187
Figure 5.6: Assessment of FlhDC and YdiV protein purity for kinetic and thermodynamic interaction analysis 190
Figure 5.7: BLItz binding spectra for YdiV binding to immobilized FlhDC 191
Figure 5.8: BLItz binding spectra for FlhDC binding to immobilized YdiV 193
Figure 5.9: Isothermal Titration Calorimetry curve for the FlhDC YdiV titration 196
Figure 5.10: Schematic of the Li *et al.* (2012) model for YdiV:$FlhD_4C_2$ interaction 197
Figure 5.11: Primary immune response antibody specificity test 199
Figure 5.12: Secondary immune response antibody specificity test 200
Figure 5.13: Anti-YdiV terminal bleed sensitivity test 202
Figure 5.14: Anti-FlhDC terminal bleed sensitivity test 202
Figure 5.15: FlhDC expression changes at various growth rates 204
Figure 5.16: Relative intensities of FlhDC expression, as quantified by Image-J software 205
Figure 5.17: EMSA of radiolabelled P*fliD* DNA with varying quantities of FlhDC and YdiV proteins 208
Figure 5.18: BLItz binding spectra for DNA-protein control experiments 212
Figure 5.19: BLItz binding spectra for “target” DNA immobilisation and FlhDC addition 214
Figure 5.20: BLItz binding spectra for “control” DNA immobilisation and FlhDC addition 214
Figure 5.21: BLItz binding spectra for “control” and “target” DNA immobilisation and YdiV addition 218
Figure 5.22: *In vitro* transcription of *fliD* in the presence of the transcription factor FlhDC and anti-FlhDC factor YdiV 221
Figure 5.23: EMSA to determine the role of YdiV in the P*fliD* , FlhDC and YdiV reaction mix 223
Figure 5.24: Control EMSA for FlhDC and anti-YdiV 224
Figure 5.25: YdiV:FlhDC pull-down assays in the presence of nucleotides 226
Figure 5.26: Quantification of the relative amounts of eluted FlhDC and YdiV proteins resulting from nucleotide pull-down assays 228
Figure 5.27: Fluorescence spectroscopic analysis of the YdiV:FlhDC complex in the presence of various nucleotides 230
Figure 5.28: Fluorescence spectroscopic analysis of the YdiV:FlhDC complex and c-di-GMP addition for three biological replicates 231
Figure 5.29: EMSA to show the P*fliD*:FlhDC binding capacity in the presence of various nucleotides 233
Figure 5.30: *In vitro* transcription levels of *fliD* in the presence of FlhDC and excess 235

nucleotides
Figure 5.31: Schematic plasmid maps for the insertion of *fliT and fliZ* into pGEX-KG vectors 237
Figure 5.32: Control pull-down assay, testing potential binding partners of the transcription factor FlhDC 240
Figure 5.33: Competition pull-down assay, testing potential competitors of the FlhDC:YdiV interaction 242
Figure 5.34: The structure of models developed to simulate inhibition of $FlhD_4C_2$-mediated gene expression by YdiV 247
Figure 5.35: Graphical representation of the effect of YdiV on DNA:FlhDC dissociation 248

List of tables

Table 2.1: Strains of *Escherichia coli* used in this study 33
Table 2.2: Plasmids used in this study 34
Table 2.3: List of primers used in this study 38
Table 2.4: Cell breakage buffers 45
Table 2.5: Affinity chromatography purification buffers 46
Table 2.6: Anion exchange buffers 46
Table 2.7: Thrombin cleavage buffers for TF:Yeal$_{GGDEF}$ purifications 48
Table 2.8: Components of SDS-PAGE gels and buffers 52
Table 2.9: Western blot buffer components 53
Table 2.10: Components of electroblotting buffers 56
Table 2.11: Nucleotide:protein pull-down assay buffers 60
Table 2.12: Protein:protein pull-down assay buffers 61
Table 2.13: BLItz programme details for AR2G biosensor protocols 63
Table 2.14: EMSA gel and buffer components 66
Table 2.15: List of constituents in IVT gels, dyes, markers and buffers 67
Table 2.16: BLItz programme details for SA biosensor protocols 69
Table 3.1: Primers used in amplification of the *yeaI$_{GGDEF}$* gene for insertion into the plasmid pET21a 74
Table 3.2: Sequencing primers used for the plasmid pET21a 74
Table 3.3: Primers used in amplification of the *yeaI$_{GGDEF}$* gene for insertion into the plasmid pCOLD TF 97
Table 3.4: Sequencing primers for pCOLD TF vector 99
Table 4.1: Primers used in amplification of the *ydiV* gene for ligation into the plasmid pET21a 118
Table 4.2: Sequencing primers used for the plasmid pET21a. 120
Table 5.1: Primers for the production of two overproduction plasmids producing the fusion proteins FlhDC-His$_6$ and (FLAG)$_3$-FlhDC 181
Table 5.2: Primers for amplification of P*fliD* for EMSA assays 206
Table 5.3: Primers for synthesis of target and control dsDNA for BLItz analysis 210
Table 5.4: Primers for IVT P*fliD- fliD* DNA synthesis 219
Table 5.5: Primers used for amplification of the *fliT and fliZ* genes for insertion into the pGEX-KG plasmid 238
Table 5.6: Sequencing primers for pGEX-KG 238
Table 5.7: Model reactions and parameter values 247

Abbreviations

ADP: Adenosine diphosphate
AMP: Adenosine monophosphate
APS: Ammonium persulphate
ATP: Adenosine triphosphate
BLItz: Bio-layer inferometry
cAMP: Cyclic adenosine monophosphate
cGMP: Cyclic guanosine monophosphate
Cyclic di-AMP/ c-di-AMP: Bis-(3'5')-cyclic dimeric adenosine monophosphate
Cyclic di-GMP/ c-di-GMP: Bis-(3'5')-cyclic dimeric guanosine monophosphate
DGC: Diguanylate cycase
dH_2O: Deionized water
DNA: Deoxyribonucleic acid
EMSA: Electromobility shift assay
EPS: Exopolysaccharide
g: Gravitational force
GDP: Guanosine diphosphate
GMP: Guanosine monophosphate
GTP: Guanosine triphosphate
HPLC: High-performance liquid chromatography
IPTG: Isopropyl β-D-1-thiogalactopyranoside
I-site: Inhibition site
ITC: Isothermal titration calorimetry
IVT: *In vitro* transcription
Kb: Kilobases
KDa: Kilodaltons
LB: Lennox broth
NMR: Nuclear magnetic resonance
NO: Nitrogen oxide
OD: Optical density
ONPG: O-Nitrophenyl-β-D-galactoside
PAGE: Polyacrylamide gel electrophoresis
PDE: Phosphodiesterase
P*fliD*: Promoter region for the gene *fliD*
pGpG: 5'-phosphoguanylyl-(3'->5')- guanosine
ppGpp: Guanosine pentaphosphate
SDS: Sodium dodecyl sulphate
TF: Trigger factor

Chapter 1: Introduction

1.1 Mechanisms of signal transduction in *Bacteria*

The survival of every organism is dependent on cellular signalling mechanisms, to detect changes in external and internal conditions and respond accordingly. There are several mechanisms in which bacteria respond to these changes of environment; these being certain transcription factors, two component systems and secondary messenger systems.

Transcription factors are the simplest mechanism in which a protein detects a signal and causes a direct response intracellularly. One example in *E. coli* is arabinose catabolism, which is regulated by the AraC protein. AraC binds to arabinose and responds by inducing the expression of five operons encoding proteins for arabinose transport, catabolism and autoregulation (Schleif, 2000).

Another frequently used mechanism for signal transduction is the two-component system. This involves a histidine kinase (HK) protein (often membrane-bound), which acts as a sensory component and a response regulator (RR) protein, which induces a cellular response to the signal. In *E. coli* 30 HK proteins have been identified alongside a similar number of response proteins, indicating the diversity of this system to both detect signals and respond. Often, upon detection of an exogenous signal, the HK domain undergoes auto-phosphorylation, which subsquently leads to phosphorylation of the cognate RR. The activated response regulator then induces a cellular change, by mediating changes in gene transcription (Yamamoto *et al.,* 2005).

Other systems involve small molecules known as secondary messengers, which work in the same basic manner, detecting a stimulus and responding by altering the concentration of an intracellular secondary messenger. The messenger molecule(s) subsequently interacts with various effector proteins, resulting in a variety of cellular effects. In this way, the small molecules are responsible for signal transduction, producing a cellular response to a signal (Kalia *et al.,* 2013; Botsford & Harman, 1992).

One of the best characterised secondary messenger pathways is the cyclic AMP (cAMP) pathway, which acts as a four component system and results in changes in gene expression (Botsford & Harman, 1992; Borrelli *et al.,* 1992).

In *E. coli*, the activity of the only adenylate cyclase (CyaA) (responsible for cAMP synthesis) is controlled by interaction with the glucose phosphotransferase system (PTS). Phosphoenolpyruvate initiates a phosphorelay in which the components of the glucose PTS (PtsI, PtsH, Crr and PtsG) are phosphorylated. PtsG mediates the uptake of glucose across the inner membrane with concomitant conversion to glucose-6-phosphate. Thus, when glucose is abundant phosphate flows through the relay and the phosphorylation state of the proteins that constitute the glucose PTS is minimal. When the bacteria are glucose-starved the PTS is fully phosphorylated and Crr~P interacts with CyaA to promote adenylate cyclase activity. Hence, the bacteria respond to glucose-starvation by increasing the intracellular concentration of cAMP. This is perceived by the cAMP receptor protein (CRP), a transcription factor that exhibits enhanced site-specific DNA-binding when in complex with cAMP. Binding of CRP-cAMP to target promoters' results in the regulation of several hundred genes, many of which are associated with the utilization of alternative carbon sources. A single cAMP phosphodiesterase (CpdA) could act to lower the concentration of cAMP in *E. coli*, but its activity with cAMP is low and perhaps a more likely route to lower intracellular cAMP concentrations is via TolC-mediated export (Botsford & Harman, 1992; Borrelli *et al.,* 1992; Hantke *et al*., 2011).

Other secondary messengers include (p)ppGpp, which acts in the stringent response to enable bacterial adaptation to nutrient starvation conditions (Chatterji & Ojha, 2001). Cyclic di-GMP and cyclic di-AMP are cyclic dinucleotides which have been identified as important signalling molecules (Hengge, 2009; Corrigan & Gründling, 2013). A poorly understood and recently identified secondary messenger is cyclic GMP (An *et al.,* 2013).

Whilst these signalling pathways regulate very different processes, they operate in a similar fashion and represent a significant proportion of the bacterial signalling portfolio (Kalia *et al.,* 2013). The widespread use of nucleotides as secondary messengers opens the possibility of other nucleotides being involved in currently unknown/uncharacterised regulatory pathways.

1.2 Cyclic di-GMP signalling

Bis-(3'-5')-cyclic dimeric guanosine monophosphate (cyclic di-GMP/ c-di-GMP) was first discovered in 1987 to be an activator of a cellulose synthase enzyme in *Gluconacetobacter xylinus,* previously named *Acetobacter xylinium* (Ross *et al.,* 1987). Since this discovery, research activity has increased dramatically and it has been found that c-di-GMP is a ubiquitous secondary messenger in bacteria (Hengge, 2009; Römling *et al.,* 2005; Römling *et al.,* 2013).

The c-di-GMP pathway is similar to cAMP signalling (Section 1.1) but with a greater level of multiplicity, consisting of multiple sensory domains, multiple enzymes responsible for c- di-GMP turnover and numerous effector proteins activated by c-di-GMP, producing diverse cellular effects (Hengge, 2009; Römling *et al.,* 2013).

In the c-di-GMP pathway, external stimuli such as light or the binding of DNA, carbohydrates or proteins are stimulants for the activation of c-di-GMP turnover enzymes. The formation of c-di-GMP occurs by the activity of diguanylate cyclase (DGC; Section 1.6.1) proteins whilst phosphodiesterase (PDE; Section 1.6.2) enzymes break down this secondary messenger molecule. The number of these enzymes varies from species to species, however it is common for an organism to contain multiple proteins (10's-100's) for both enzymatic activities. The intracellular c-di-GMP acts as a secondary messenger molecule, activating various effector proteins which lead to numerous phenotypic responses in the cell. The cellular responses depend on the precise levels of intracellular c-di-GMP, but effect major bacterial phenotypes such as cell motility, virulence, biofilm formation and cell morphology (Römling *et al.,* 2005; Galperin, 2004; Hengge, 2009).

The widespread presence of this signalling pathway within the bacterial kingdom and the diversity of the cellular phenotypes related to c-di-GMP, have cemented the importance of this secondary messenger in bacterial adaptation and survival. Moreover, c-di-GMP is a bacterial-specific molecule, providing a whole array of potential targets for novel antibiotic design (Hengge, 2009; Römling *et al.,* 2013).

1.3 Cellular effects of c-di-GMP signalling

Bis-(3'-5')-cyclic dimeric guanosine monophosphate signalling pathways result in an array of cellular characteristics, achieved by modulating the activities of various effector proteins (Hengge, 2009). These signalling outputs include regulating biofilm formation, exopolysaccharide production, gene transcription, flagellum activity and adhesin secretion (Newell *et al.,* 2009; Hickman *et al.,* 2005; Kirillina *et al.,* 2004; Tischler & Camilli, 2004; Simm *et al.,* 2004; Hickman & Harwood, 2008).

Controlling gene regulation is one of the major cellular effects of c-di-GMP signalling. One example of this comes from the *E. coli* YddV protein, which upon overexpression, caused the up-regulation of 109 genes and repression of 87 genes (Méndez-Ortiz *et al.,* 2006). In conditions of high c-di-GMP (created by over-production of the c-di-GMP producing enzyme YddV), genes encoding cell wall proteins were induced, whilst sugar metabolism genes were repressed (Méndez-Ortiz *et al.,* 2006). Another example is VieA (a c-di-GMP turnover enzyme) from *Vibrio cholerae* which represses transcription of the *vps* genes encoding exopolysaccharide proteins (Tischler & Camilli, 2004).

Bis-(3'-5')-cyclic dimeric guanosine monophosphate is also associated with controlling the virulence of pathogens, antibiotic production and regulating cell cycle progression (Tischler & Camilli, 2005; Aldridge *et al.,* 2003; Hengge, 2009; Duerig *et al.,* 2009). For example, in *V. cholerae*, PilZ domain proteins (Section 1.4.1) have been shown to bind c-di-GMP specifically and regulate the expression of virulence genes as well as biofilm and motility (Pratt *et al.,* 2007).

However, the major role of c-di-GMP signalling is in regulating the predominant lifestyle choice in bacteria; between single motile organisms and biofilm community-based lifestyles. Generally high levels of c-di-GMP cause biofilm formation via the synthesis of extracellular polymeric substances (EPS). A decrease in the intracellular c-di-GMP levels supports a more motile phenotype, increasing flagella synthesis and repressing biofilm favouring genes (Hengge, 2009; Kirillina *et al.,* 2004; Simm *et al.,* 2004).

1.3.1 Bacterial motility and its regulation by c-di-GMP signalling

The motility of bacteria is often dependent on the presence or absence of flagella on the bacterial surface (Macnab, 1977 and references within). A single flagellum consists of a basal membrane-bound structure which is attached to a rotating hook, leading to a long flagellum filament (Figure 1.1). The assembly of a flagellum is a sequential process, starting with the MS-ring, the C-ring and the type III secretion apparatus. Next to be formed, is the rod which spans the periplasmic space, before the P-ring and L-ring associate. At the cell surface, the hook structure is assembled before finally the filament subunits assemble, creating the external appendage (Mcnab, 2003; 1996).

In *E. coli* there are ~ 50 genes which are involved in flagella production and export, all of which need to be tightly regulated in order to produce flagella under the appropriate conditions (Figure 1.2) (Chilcott & Hughes, 2000). These 50 genes are divided into a hierarchy of 3 classes, named Class I, II and III or early, middle and late genes (Liu & Matsumura, 1994; Chilcott & Hughes, 2000). The Class I genes consist of the *flhDC* operon only, which encodes the proteins FlhD and FlhC, forming the transcription factor $FlhD_4C_2$. This protein complex is known as the 'master regulator' of flagellum gene transcription, binding to the promoter regions of all Class II genes and enhancing their transcription. The Class II genes include basal-body rod proteins proteins such as FlgB, chaperone proteins such as FliT and the σ^{28} factor FliA. FliA is then necessary to bind upstream of all Class III genes and induce their transcription, encoding proteins such as the flagellum filament protein FliC (Chilcott & Hughes, 2000; Macnab, 2003; Liu & Matsumura, 1994).

This transcriptional cascade is tightly regulated in a multi-faceted manner. The transcription of *flhDC* is regulated by systems responding to environmental cues such as high osmolarity, quorum sensing and heat shock (Shin & Park, 1995; Sperandio *et al.,* 2002; Wei *et al.,* 2001). The production of FlhDC regulates flagellum synthesis, by up-regulating the Class II genes. Some of the products of Class II transcription then act to both activate (FliZ) and repress (FliT) the activity of FlhDC in a feedback loop, to ensure precise control of FlhDC activity (Kutsukake *et al.,* 1999; Chilcott & Hughes, 2000). Furthermore, other Class II transcription products (FliA and FlgM) are responsible for the transcription of Class III genes, with FliA necessary for the activation of these genes directly and FlgM down-regulating the activity of FliA (Kutsukake & Iino, 1994; Barembruch & Hengge, 2007). In this manner, the production of flagella is tightly regulated.

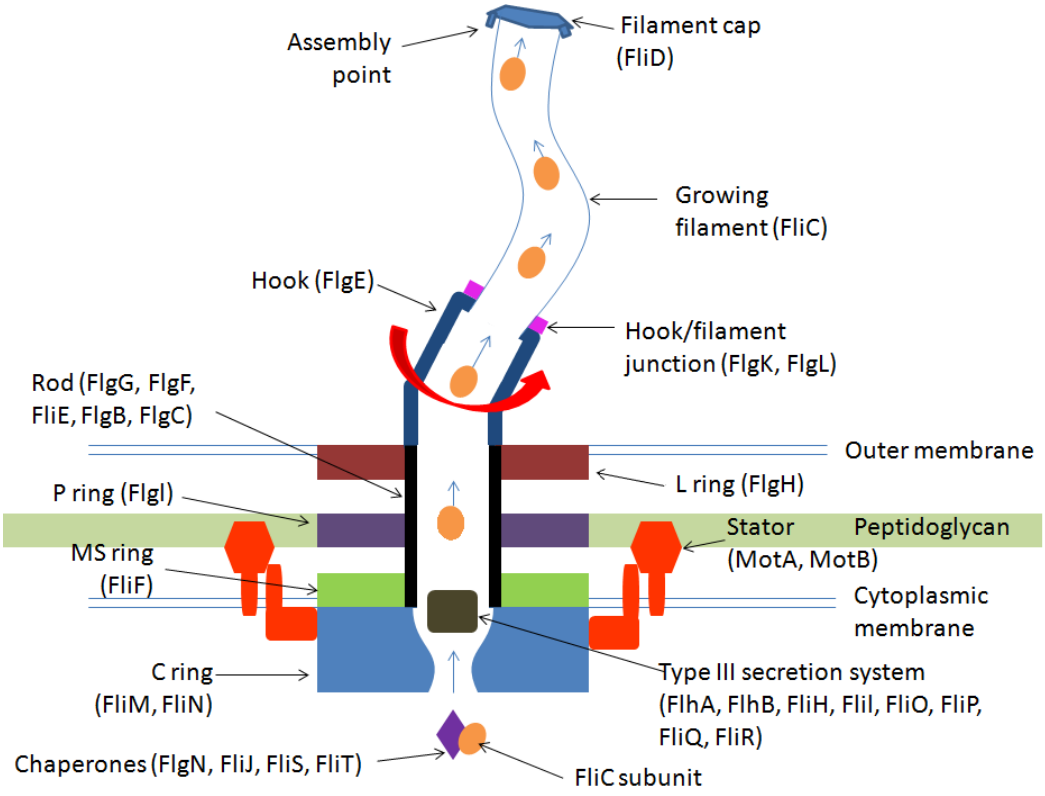


Figure 1.1: Schematic diagram of bacterial flagellar components

Cross-sectional diagram of the flagellar appendage in bacteria, showing the location and role of the flagellar proteins. The flagellum is a hollow tube, as clearly shown in the diagram, enabling proteins to be secreted from the cell outside, in order to allow the formation and growth of the filament.

The assembly of this structure is tightly controlled and sequential. The MS-ring forms first, closely followed by the C-ring and the type III secretion system. Following this, the rod structure is assembled, prior to the P-ring and L-ring being produced. The hook is then assembled, before finally the filament subunits are added, producing the long tail-like appendage known as a flagellum.

(Macnab, 1996; 2003)

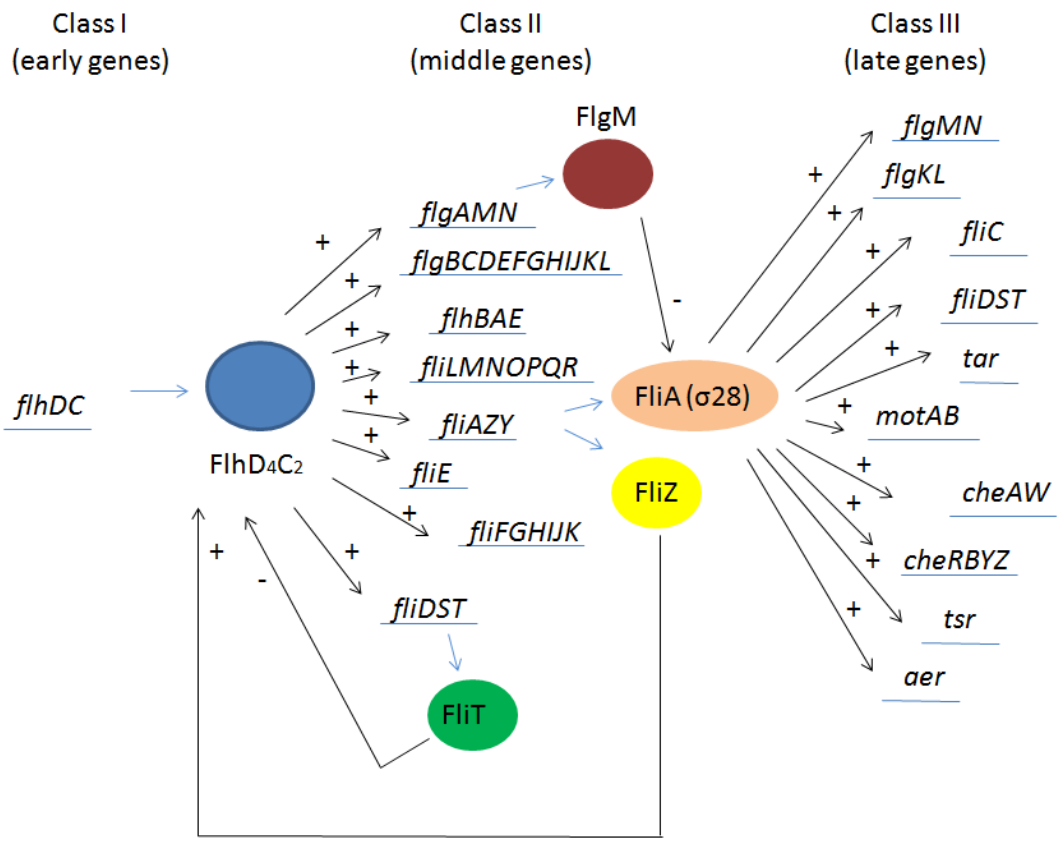


Figure 1.2: Schematic diagram of the cascade and regulation of flagellar gene synthesis in *E. coli*

The *E. coli* flagellar genes are arranged in a three-tier system, named Class I, II and III (or early, middle and late) genes. This system is highly regulated by a number of activators and repressors, fine-tuning the expression of the genes.

The regulatory transcription system is shown here, with genes and operons indicated by italicised writing, whilst coloured shapes indicate proteins. Activators (or up-regulators) are denoted by + labels and repressors (or down-regulators) denoted by – labels.

(Chilcott & Hughes, 2000; Macnab, 2003; Liu & Matsumura, 1994)

The link between intracellular c-di-GMP levels and motility was first noted by Simm *et al.* (2004) and since then has been identified in species across the bacterial kingdom. Low intracellular levels of c-di-GMP characteristically produce motile bacteria, by inducing flagella synthesis and repressing biofilm production (Simm *et al.,* 2004; Krasteva *et al.,* 2010; Ryjenkov *et al.,* 2006; Römling *et al.,* 2013).

In *V. cholerae*, a transcriptional regulator VpsT has been shown to bind to c-di-GMP and inversely control extracellular matrix production and motility. Upon deletion of the *vpsT* gene, cells displayed an increased motility, indicating a direct role of this protein in regulating motility. Upon c-di-GMP binding, VpsT undergoes dimerisation to produce an active form which strongly decreases the expression of various flagellum genes. Additionally, the transcriptional regulator up-regulates the expression of polysaccharide biosynthesis (*vsp*) genes. Therefore the protein has a dual role, both inhibiting motility and increasing biofilm formation in high c-di-GMP conditions (Krasteva *et al.,* 2010).

FleQ is a c-di-GMP-dependent flagella regulator in *Pseudomonas aeruginosa*. FleQ has been shown to bind c-di-GMP and upon binding, controls the expression of the *pel* genes (encoding polysaccharide biosynthesis proteins) and the flagella regulon (Hickman & Harwood, 2008).

In *E. coli*, YcgR and YhjH were identified as putative motility-inducing proteins. Both proteins were found to restore the motility of a non-motile *hns* mutant (Ko & Park, 2000). YhjH is an active phosphodiesterase and a member of the EAL domain family (Pesavento *et al.,* 2008; Section 1.6.2). YcgR is a PilZ domain protein, which has been shown to bind to c-di-GMP (Ryjenkov *et al.,* 2006; Section 1.4.1).

Upon interaction with c-di-GMP, YcgR has been shown to affect motility, causing flagellar rotation direction to change (inducing a counterclockwise bias) and the rate of flagellar rotation to decrease. Two distinct models have been hypothesised to account for the roles of YcgR, however there are distinct differences between models (i.e. whether Yc gR interacts with the FliM/FliG proteins, thereby inducing a counterclockwise bias, or the MotA protein and thereby reduces the efficiency of energy transfer and slows flagellar rotation) which need further investigation before a clarified mechanism of action for YcgR is known (Paul et al., 2010; Boehm *et al.,* 2010) These dual roles of YcgR have led to the

protein being labelled as having a “brake or clutch” effect on flagellar motility (Römling *et al.,* 2013).

1.3.2 Biofilm formation and c-di-GMP regulation of this phenotype

The c-di-GMP signalling pathway has been linked to biofilm production from the initial discovery of c-di-GMP in 1987, where the secondary messenger was found to activate the cellulose synthase enzyme in *G. xylinus* (Ross *et al.,* 1987). Since then, this apparent link has been confirmed across the bacterial kingdom, identifying many DGC and PDE enzymes that regulate cellulose production and biofilm formation (Simm *et al.,* 2004; Tischler & Camilli, 2004; GarcÍa *et al.,* 2004; Bomchil *et al.,* 2003).

Bacterial biofilms are of great significance, with biofilms present in over 80% of all (acute and chronic) human infections. Common infections such as urinary tract infections, catheter infections, chronic wounds, pneumonia and even dental plaque are all produced by bacterial biofilms (Costerton *et al.,* 1999; Bjarnsholt, 2013).

The term biofilm refers to a group of microorganisms which adhere to one another on a surface, aided by an array of EPS such as DNA, proteins and polysaccharides (O’Toole *et al.,* 2000; Pratt & Kolter, 1998). The molecular mechanism of biofilm formation is characterised by five distinct steps (Figure 1.3). Firstly bacteria make an initial contact with a surface. The vast majority of cells are released from the surface and remain motile; however some bacterial cells anchor themselves to the surfaces via pili. Once the cells are firmly attached, the biofilm matures by both cell division and recruitment of bacteria to the site. Biofilm matrices are produced, which enclose the cells and forms a protective barrier. This matrix is composed of both internal (i.e. polysaccharides synthesised within the cell and exported out) and external (i.e. minerals and soil particles) materials. This biofilm can then last for an indefinite time, depending on environmental conditions. Under certain conditions, the biofilm is dispersed, releasing the mass of cells into single motile cells once more (Pratt & Kolter, 1998; O’Toole *et al.,* 2000; Stoodley *et al.,* 2002)

In order for biofilms to be produced, cells rely on quorum sensing (QS) methods to communicate with other cells and inform the cells about the surrounding environment. Quorum sensing is the process in which self-generated signalling molecules called autoinducers (AIs), are produced by cells and secreted into the environment. Bacteria have

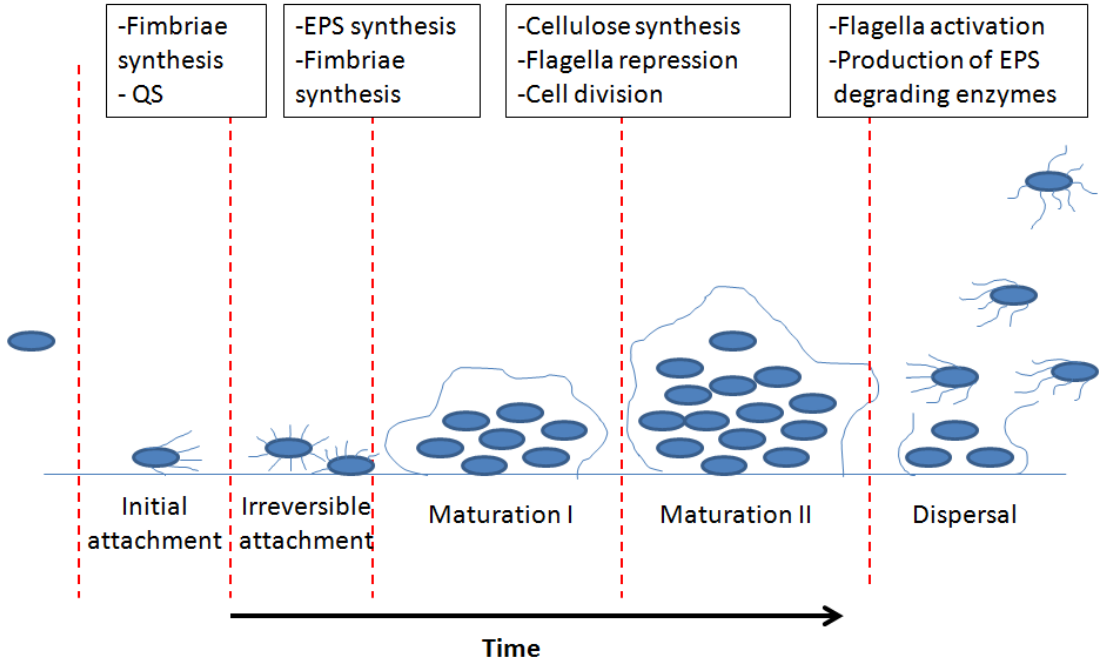


Figure 1.3: Schematic diagram showing the stages of biofilm formation

Diagram to show how single bacterial cells (shown as blue ovals) aggregate to form a biofilm structure with subsequent dispersal into individual cells. The red dashed lines indicate the transition to a new step (of which there are five) and the boxes indicate the key processes occurring between each transition.

The cells initially attach to a surface through weak bonds, but can form strong irreversible attachments by the production of fimbriae structures which adhere the cells to the surface. The decision to irreversibly attach to a surface is largely due to QS, enabling cells to determine whether to remain motile and independent or form an aggregated biofilm.

Once bacteria have become firmly attached to a surface, they start producing cellulose and other extracellular polymeric substances (EPS) which form the biofilm matrix. At this maturation stage, bacteria stop producing flagellae. The cells can remain in this biofilm for extended periods of time, continuing to divide and produce EPS components, all of which give the aggregated cells an increased survival rate.

Once the environmental conditions have changed, and there is no advantage to remain in a biofilm, the cells begin to disperse from the biofilm. To break apart the EPS matrix, cells synthesise EPS-degrading enzymes, releasing the aggregated cells. Flagella synthesis is re-activated by the cells, enabling swimming to new environments.

(O'Toole *et al.,* 2000; Pratt & Kolter, 1998; Stoodley *et al.,* 2002)

a threshold level for AI's, meaning that the bacteria only detect these molecules when multiple nearby bacteria are secreting the molecule. In this manner, bacteria are informed of the surrounding population density and this induces transcriptional changes within the cells, such as production of exopolysaccharides or pili formatIon, both of which are crucial to biofilm formation (Bassler & Losick, 2006; Hammer & Bassler, 2003; Parsek & Greenberg, 2005).

Quorum sensing thereby enables bacterial cells to act as a population rather than individual cells, which can be advantageous for survival, with biofilms offering protection from UV, water-limiting conditions, predation, disinfectants and antibiotics (Chang *et al.,* 2007; Fux *et al.,* 2005; Matz *et al.,* 2005; Elasri & Miller, 1999). The precise environmental conditions which cause biofilm formation (by QS) vary from organism to organism. For example, the microaerobic bacterium *Campylobacter jejuni* forms biofilms in stressful high O_2 levels (20%) as a survival strategy (Reuter *et al.,* 2010). Gram-negative bacteria such as *E. coli* and *Salmonella* survive low nutrient starvation conditions by the formation of biofilms (Kolter *et al.,* 1993). Furthermore, low levels of β-lactam antibiotics induce the production of extracellular DNA and biofilm formation in *Staphylococcus aureus* (Kaplan *et al.,* 2012).

One of the key components necessary for biofilm formation in *E. coli* and *Salmonella* strains is the synthesis of cellulose, which acts as the major exopolysaccharide in the biofilm matrix (Zogaj *et al.,* 2001). Additionally, bacterial cells require adherence proteins and motility appendages, such as curli fibres, flagella and pili, in order to make the initial attachment to surfaces for biofilm formation (Pratt & Kolter, 1998; Olsén *et al.,* 1989). Cyclic di-GMP signaling proteins have been implicated in biofilm formation by activation of these key biofilm-inducing systems, across a range of bacterial species.

In *P. aeruginosa,* FleQ has been identified as a regulator of bacterial biofilm in a c-di-GMP-dependent manner. Under low c-di-GMP conditions, FleQ forms a complex with an accessory protein FleN and the complex binds upstream of biofilm associated genes and represses transcription by bending the DNA. However, in high c-di-GMP environments, DNA bending does not occur, thereby inducing the expression of biofilm associated genes, specifically *pel* and *psl* (Baraquet *et al.,* 2012; Hickman & Harwood, 2008).

VieA is a *V. cholerae* EAL domain protein (Section 1.6.2) which hydrolyses intracellular c-di-GMP. Upon activation of the protein (and c-di-GMP hydrolysis), the protein represses transcription of exopolysaccharide synthesis (*vps*) genes, thereby inhibiting biofilm formation (Tischler & Camilli, 2004).

In *E. coli*, the transcription factor CsgD activates transcription of the *csgBA* operon which encodes curli fibre structures (Römling *et al.,* 1998; Prigent-Combaret *et al.,* 2001). The transcription of *csgDEFG* and therefore levels of CsgD in the cell, are regulated by pairs of c-di-GMP signalling proteins. In particular, YdaM and YciR (a DGC and PDE respectively; Section 1.6) and YegE and YhjH (a DGC and PDE respectively) are two pairs of proteins which regulate CsgD levels (Pesavento *et al.,* 2008; Weber *et al.,* 2006). These enzymes fine-tune the intracellular levels of c-di-GMP, which at high concentrations stimulates the transcription of CsgD, thereby controlling the synthesis of curli fibre (Pesavento *et al.,* 2008; Weber *et al.,* 2006).

Futhermore, CsgD activates the transcription of proteins YaiC and YoaD (DGC and PDE activities respectively) which again control the concentration of cellular c-di-GMP (Brombacher *et al.,* 2006). In turn, this c-di-GMP pool regulates the activity of the BcsA protein and cellulose formation (Zogaj *et al.,* 2001; Amikam & Galperin, 2006; Ryjenkov *et al.,* 2006; Brombacher *et al.,* 2006).

1.4 Cyclic di-GMP effector proteins

Whole arrays of c-di-GMP binding proteins exist in bacterial cells, and it is these effectors which convert intracellular fluctuations of c-di-GMP concentrations into phenotypic cellular effects. Whilst a vast number of effector proteins remain unknown, an ever-growing number of c-di-GMP binding proteins have been characterised, and can be categorised into multiple classes (Römling *et al.,* 2013; Corrigan & Gründling, 2013). Here four classes; PilZ domain proteins, degenerate GGDEF/EAL proteins, riboswitches and unrelated proteins, are discussed.

1.4.1 The PilZ domain class of effector proteins

One of the most common effector proteins is the c-di-GMP binding PilZ domain. This domain was first identified in cellulose synthase enzymes from *G. xylinus*, but was soon

revealed as a wide-spread domain (Amikam & Galperin, 2006). The PilZ domain was identified in *V. cholerae* proteins PlzC and PlzD, shown to regulate motility, biofilm formation and virulence of the bacteria in a c-di-GMP-dependent manner (Pratt *et al.,* 2007). The *E. coli* protein YcgR and *C. crescentus* protein DgrA are just two further examples of PilZ proteins, binding c-di-GMP and inducing a downstream signalling cascade resulting in the regulation of cell motility (Ryjenkov *et al.,* 2006; Christen *et al.,* 2007).

PilZ domain effectors have been shown to regulate a number of cellular functions, the most common of which are motility, exopolysaccharide synthesis and DNA binding functions (Ryjenkov *et al.,* 2006; Christen *et al.,* 2007; Merighi *et al.,* 2007; Johnson *et al.,* 2011).

1.4.2 The role of degenerate GGDEF/EAL domain proteins as c-di-GMP effector proteins

The aforementioned DGC and PDE enzymes are responsible for c-di-GMP turnover in bacteria (Section 1.6). Proteins with these catalytic activities are characterised by a highly conserved motif of residues necessary for the enzymatic function. These motifs are GGDEF (for DGC activity) and EAL or HD-GYP (for PDE activity). The protein domains containing these motifs are designated as GGDEF, EAL and HD-GYP domain proteins (Simm *et al.,* 2004; Paul *et al.,* 2004; Hengge, 2009).

A significant proportion of the GGDEF/EAL domain proteins lack this highly conserved motif and are presumed to be catalytically inactive (Section 1.8) (Schmidt *et al.,* 2005; Suzuki *et al.,* 2006). However, several of these proteins have been identified as c-di-GMP receptor-effector proteins, activated by binding to c-di-GMP and initiating a cellular response (Römling *et al.,* 2013; Duerig *et al.,* 2009; Lee *et al.,* 2007; Newell *et al.,* 2009). At present, no degenerate HD-GYP domains have been identified as effector proteins.

In *C. crescentus*, PopA is a degenerate GGDEF domain protein but binds c-di-GMP at its intact I-site (Section 1.8.1). Upon c-di-GMP binding, PopA is activated and induces a downstream cascade of protein recruitment, culminating in the degradation of the replication inhibitor protein CtrA. Ultimately this leads to progression of the *C. crescentus* cell cycle from G1 to S phase (Duerig *et al,* 2009). Another example of a degenerate GGDEF effector protein is PelD in *P. aeruginosa*. PelD is a highly diverged GGDEF domain protein, lacking most structural features of DGCs, retaining only an intact I-site. This protein binds

c-di-GMP at its I-site which activates the protein and induces post-translational regulation of PEL polysaccharide synthesis (Lee *et al.,* 2007; Christen *et al.,* 2006).

Degenerate EAL domain proteins have also been shown to be c-di-GMP effector proteins. In *Pseudomonas fluorescens*, LapD was identified as a degenerate GGDEF-EAL domain protein. Neither domain is catalytically active, however LapD becomes activated by c-di-GMP binding to the EAL domain of the protein. Upon activation, the protein induces cell adhesion and biofilm formation (Newell *et al.,* 2009).

1.4.3 Riboswitches as c-di-GMP effector proteins

Alongside protein effector systems, RNA effector systems have also been identified, named as riboswitches. Riboswitches are non-coding RNA molecules which exist in a characteristic secondary structure. These RNA segments bind c-di-GMP at high affinity and regulate gene transcription according to c-di-GMP concentrations. Virulence genes, flagellum and pilus biosynthesis genes have been identified as genes regulated by riboswitches. Whilst this effector class is still relatively uncharacterised, it may represent a highly important class, given the already high numbers of riboswitches identified in some species (Sudarsan *et al.,* 2008; Barrick & Breaker, 2007).

1.4.4 The detection of unrelated proteins as effector proteins

Protein sequence alignments have been useful in identifying effector proteins of class 1 and 2 (PilZ domain proteins and degenerate GGDEF/EAL domain proteins). However, several proteins have been revealed as c-di-GMP effector proteins which do not fit into these classes and appear to have no common motif or domain unifying them.

One example is FleQ which is a transcriptional regulator in *P. aeruginosa*. FleQ is a DNA-binding protein, known to regulate expression of various genes involved in exopolysaccharide production and flagellum-based motility. FleQ acts in a c-di-GMP-dependent manner, repressing exopolysaccharide synthesis but up-regulating flagellum assembly genes in low c-di-GMP environments. Under high c-di-GMP conditions, FleQ has the opposite effect, initiating exopolysaccharide synthesis and repressing motility genes (Hickman & Harwood, 2008).

Several other transcription factors have been identified as c-di-GMP binding proteins, such as Clp from *Xanthomonas campestris*, a member of the CRP/FNR superfamily, which controls virulence gene expression (Chin *et al.,* 2010). In *Burkholderia cenocepacia*, the Bcam1349 protein is a CRP/FNR family transcriptional regulator. This protein regulates biofilm formation and virulence genes in a c-di-GMP-dependent manner (Fazli *et al.,* 2011).

1.5 C-di-GMP sensory domains

Further expansion of the c-di-GMP signaling pathway comes from the multitude of sensory domains, which detect external stimuli and cause an activation cascade down the c-di-GMP pathway. The vast majority of GGDEF and EAL domains are linked to N-terminal transmembrane sensory domains that detect the external signal and activate the coupled DGC or PDE activity. Some GGDEF/EAL domains are coupled to multiple N-terminal sensory domains, demonstrating the complex signal integration which occurs (Hengge, 2009; Römling *et al.,* 2013). A small proportion of GGDEF/EAL domains have an attached cytosolic domain which acts as a self-contained effector protein, and here the GGDEF/EAL domains simply mediate a phosphoryl-transfer reaction between the sensory and effector domains (Jenal & Malone, 2006; Schmidt *et al.,* 2005). The most frequently identified sensory domains are PAS, GAF and REC domains (Henry & Crosson, 2012; Römling *et al.,* 2013).

1.5.1 PAS domains as sensory elements

PAS domains, named after the first proteins in which it was recognised (Per, ARNT and Sim) are widespread sensory domains, with each of the Per, ARNT and Sim proteins identified as regulatory proteins (Ponting & Aravind, 1997; Nambu *et al.,* 1991). PAS domains have been found that detect light, oxygen and other gases and redox potential (Ponting & Aravind, 1997; Zhulin *et al.,* 1997; Taylor & Zhulin, 1999).

PAS domains have two regions of conserved amino acid residue sequences, but otherwise have very high sequence deviation, often only with 20% sequence identity to other PAS domains (Zhulin *et al.,*1997; Taylor & Zhulin, 1999; Henry & Crosson, 2012). The sensory specificity of PAS domains is partly dependent on an associated cofactor, with haem or flavin commonly found (Taylor & Zhulin, 1999).

In *E. coli* the *dosCP* operon encodes the proteins DosC and DosP, both of which control intracellular c-di-GMP concentration in an oxygen-dependent manner (Tuckerman *et al.,* 2009). The DosP protein is an oxygen-sensing c-di-GMP PDE, with an N-terminal PAS domain which acts as a haem-binding oxygen sensor. The DosC protein is a DGC protein with an N-terminal GCS (globin-coupled sensor) domain (Tuckerman et al., 2009; Hou *et al.,* 2001). This GCS domain is a globin domain which binds oxygen at its haem site. Both of these sensory domains thereby enable intracellular oxygen sensing, and maintain controlled intracellular c-di-GMP metabolism via the C-terminal coupled DGC and PDE activities of DosC and DosP respectively (Tuckerman *et al.,* 2009).

1.5.2 The presence of GAF as a sensory domain

GAF domains are another type of sensory domain found in the c-di-GMP pathway. Protein domains containing strong secondary structure homology were identified in cGMP-specific and –stimulated phosphodiesterases, *Anabaena* adenylate cyclases and *E. coli* FlhA and were named GAF according to these proteins (Aravind & Ponting, 1997). These domains bind to cyclic nucleotides cAMP and cGMP and in doing so cause the activation of DGCs and PDEs (Heikaus *et al.,* 2009). A GAF domain containing a non-haem iron centre has also been shown to detect NO (Tucker *et al.,* 2008).

In *Myxococcus xanthus,* SgmT is crucial for regulating the composition of the extracellular matrix. This protein consists of an N-terminal GAF domain and a C-terminal GGDEF domain. Despite binding to an unknown ligand, the GAF domain is essential for SgmT activity, suggesting that ligand binding to GAF induces a conformational change that enables c-di-GMP binding to the GGDEF domain and subsequent downstream cellular effects (Petters *et al.,* 2012).

1.5.3 REC as a commonly found sensory domain

REC (standing for receiver) domains are so named due to their role as a phosphate acceptor or receiver protein, specifically at a conserved Asp residue (Galperin, 2006 and references within). These REC domains, also called CheY domains, are frequently used sensory domains in the c-di-GMP pathway, often coupled to GGDEF/EAL/HD-GYP domains. Upon

phosphorylation, the REC domains are activated and initiate downstream signalling events, activating neighbouring or co-localised proteins (Römling *et al.,* 2013; Paul *et al.,* 2007).

The *C. crescentus* protein PleD is a REC-containing protein, consisting of 2 N-terminal REC domains and a C-terminal GGDEF DGC domain. Both REC-domains are required for the activity of PleD, which is responsible for pole development during the *C. crescentus* cell cycle (Paul *et al.,* 2004). Upon activation of the REC domains by phosphorylation, the REC structure is repackaged, which enables the GGDEF domains to form active DGC dimers (Paul *et al.,* 2007).

1.5.4 Less common sensory domains

There is also a multitude of less common sensory domains that have been characterised. For example, the *E. coli* protein YcgF is involved in regulation gene transcription and is activated by the detection of blue light. YcgF consists of an N-terminal BLUF domain and a C-terminal EAL domain, with the BLUF domain being activated upon detection of blue light, causing a cascade of changes resulting in gene regulation (Tschowri *et al.,* 2009). MASE1 and MASE2 domains, have yet to be completely characterised but have been found to sense oxygen in *Desulfovibrio vulgaris* (Nikolskaya *et al.,* 2003). Additionally the MASE1 domain from the *E. coli* protein YfgF has been found to regulate bacterial responses to aspartate (Lacey *et al.,* 2013).

Numerous sensory domains have been identified, detecting a multitude of signalling inputs. The sheer variety of stimuli gives a clear indication of the breath of environments in which the c-di-GMP pathway is responsive to. Additonally, sensory domains can be both cytoplasmic (such as PAS or GAF) or can be periplasmic, again maximising the environments and signals which cause c-di-GMP fluctuations (Jenal & Malone, 2006; Römling *et al.,* 2005).

1.6 The biosynthesis and degradation of c-di-GMP

The turnover of the secondary messenger c-di-GMP is dependent on two families of enzymes; the diguanylate cyclases (DGCs) and the phosphodiesterases (PDEs). These two families of enzymes control the cellular level of c-di-GMP, with DGCs and PDEs synthesising and degrading the molecule respectively (Figure 1.4). Both enzyme classes have conserved

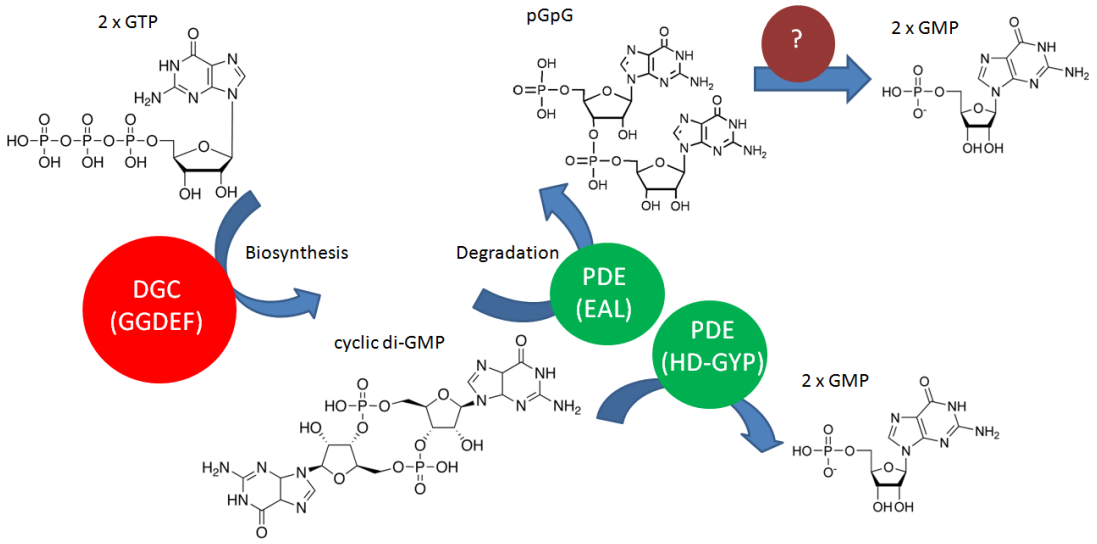


Figure 1.4: The mechanism of cyclic di-GMP turnover

The level of intracellular c-di-GMP is determined by the relative activities of DGC and PDE enzymes, characterized by consensus GGDEF and EAL/HD-GYP motifs respectively.

The DGC proteins act to synthesize c-di-GMP from two molecules of GTP and therefore the activity of these enzymes increases the cellular c-di-GMP levels. PDE enzymes have an opposing role to the DGC proteins, catalyzing the hydrolysis of c-di-GMP either into pGpG or GMP. The HD-GYP domains degrade c-di-GMP directly into 2 molecules of GMP, whilst EAL domain proteins degrade c-di-GMP into pGpG, which is further degraded into GMP molecules by a separate enzyme.

The identity of the enzyme which converts pGpG into GMP, often referred to as having PDE-B activity, is currently unknown but is believed to be ubiquitous in all c-di-GMP signalling bacteria.

(Ryjenkov *et al.,* 2005; Simm *et al.,* 2004; Schmidt *et al.,* 2005; Tamayo *et al.,* 2005; Chang *et al.,* 2001; Ryan *et al.,* 2006, Povolotsky & Hengge, 2012)

domains, which are named after the amino acid residues located in the active sites. DGCs contain the conserved GGDEF motif and are thereby named GGDEF domains, whilst PDEs contain a conserved EAL or HD-GYP domain and are named accordingly (Paul *et al.,* 2004; Hengge, 2009).

Initially GGDEF and EAL domains were the only identified c-di-GMP turnover enzymes, being the most numerous. However, several bacterial species contained GGDEF proteins but lacked EAL proteins, such as *Thermotoga maritima* which contains nine GGDEF proteins but zero EAL domains (Galperin *et al.,* 2001). This therefore raised the question whether alternative PDEs were present in these species. Investigations revealed the HD-GYP domain as a possible alternative PDE, which was confirmed to be enzymatically active and responsible for the hydrolysis of c-di-GMP into GMP molecules (Galperin *et al.,* 1999; 2001; Ryan *et al.,* 2006; Povolotsky & Hengge, 2012).

1.6.1 Diguanylate cyclase proteins

Jenal and co-workers made the initial link between GGDEF-motif proteins and the synthesis of c-di-GMP and proposed the idea of GGDEF proteins being universal regulatory proteins (Paul *et al.,* 2004). This hypothesis was found to be true, with GGDEF domains found to be characteristic of DGC activity and the synthesis of c-di-GMP (Ryjenkov *et al.,* 2005). These domains are characterized by a GGDEF or GG[D/E][D/E]F residue motif, which is a consistent feature of all active DGC enzymes in any bacterial species (Galperin *et al.,* 2001; Ryjenkov *et al.,* 2005).

Diguanylate cyclase proteins are catalytic proteins which synthesize one molecule of c-di-GMP from two molecules of GTP in the cell (Figure 1.4) (Simm *et al.,* 2004; Ryjenkov *et al.,* 2005). These proteins are active as dimers, formed of two GGDEF domains (Figure 1.5) (Chan *et al.,* 2004; Paul *et al.,* 2007). The A-site (active site) is located in the cleft between the two GGDEF domains, with each GGDEF motif at this dimerisation interface. The GGDEF motif residues and other conserved residues in the domains make contacts with the two GTP molecules and are required for the catalysis of c-di-GMP synthesis (Chan *et al.,* 2004; Navarro *et al.,* 2009).

The dimerisation of GGDEF domains is a crucial stage for the activation of DGC activity. The majority of GGDEF domains are coupled to N-terminal domains which are thought to

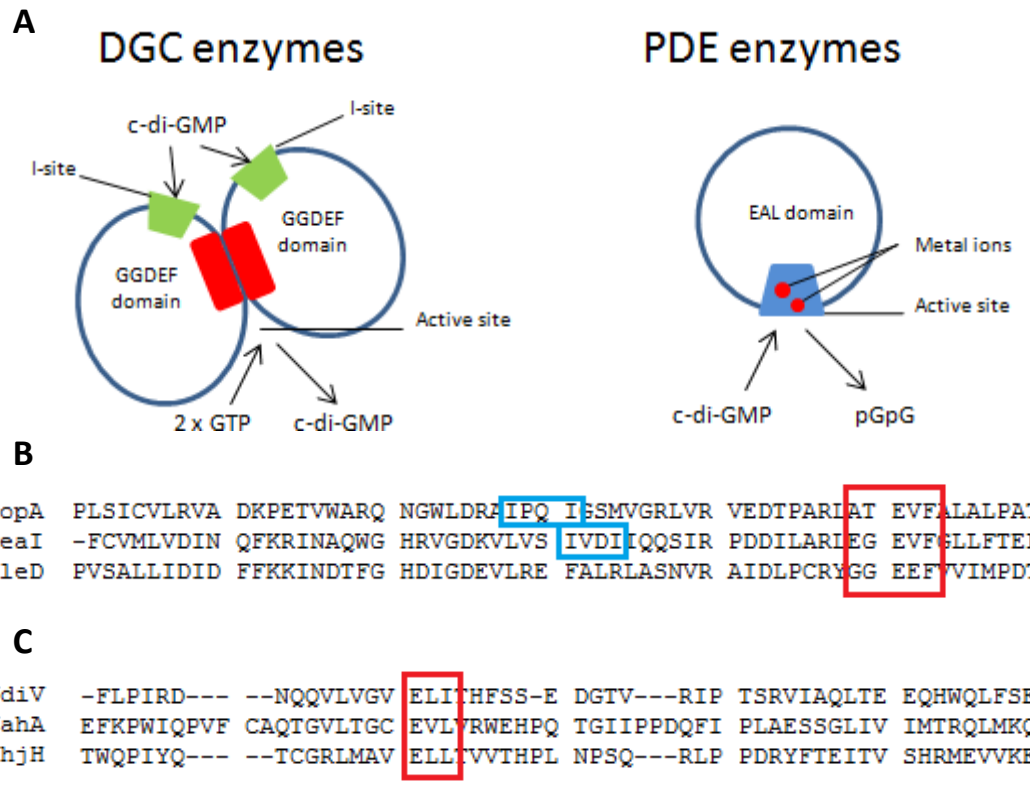


Figure 1.5: Characterisation of active and inactive GGDEF and EAL domain proteins

A) Schematic diagrams of active DGC and PDE domains. Active DGCs are produced upon the dimerisation of two GGDEF containing subunits, bringing two GGDEF motifs together (shown as red boxes) to form an active site at this interface. Inhibition sites (shown as green boxes) act to bind c-di-GMP and therefore set an upper limit on the activity of DGCs. Active PDEs are characterized as monomeric EAL domain proteins, with Mg^{2+} or Mn^{2+} (shown as red circles) bound at the active site (shown as a blue region), in order to bind c-di-GMP.

B) Protein sequence alignment of GGDEF domains, with the GGDEF consensus sequence boxed in red and the inhibition site shown in blue. PleD is a known active GGDEF domain, whilst PopA is a well-characterised 'degenerate' GGDEF domain (both from *C. crescentus*). YeaI (from *E. coli*) was chosen as a putative inactive DGC protein. Clearly, both PopA and YeaI lack retention of the GGDEF motif, but have an I-site.

C) Protein sequence alignment of EAL domains, with the EXL consensus sequence boxed in red. Proteins YahA and YhjH from *E. coli* were chosen to represent active PDE enzymes, whilst *E. coli* YdiV is a putative inactive PDE protein. Clearly, both YahA and YhjH contain the EXL motif, whilst this is lost in YdiV. For both B and C, protein sequences were obtained from Uniprot and analysed by Mega6 software using ClustalW parameters (Section 2.10.2).

(Paul *et al.,* 2004; Ryjenkov *et al.,* 2005; Chan *et al.,* 2004; Paul *et al.,* 2007; Christen *et al.,* 2006; Simm *et al.,* 2004; Schmidt *et al.,* 2005; Chang *et al.,* 2001; Tamayo *et al.,* 2005; Minasov *et al.,* 2009, Sommerfeldt *et al.,* 2009)

initiate dimerisation of GGDEF domains. In the case of PleD, phosphorylation of the REC N-terminal domains induces dimerisation of the GGDEF domains, forming a catalytically active DGC (Chan *et al.,* 2004; Paul *et al.,* 2007; Sommerfeldt *et al.,* 2009).

In addition to the catalytic active-site in DGCs, another binding site is present in some GGDEF domains which binds c-di-GMP and is responsible for allosteric control of DGC activity (Paul *et al.,* 2004; Chan *et al.,* 2004). This I-site, characterised by a conserved RXXD motif is located within the GGDEF domain, upstream from the A-site and is present within single- and multi-domain DGCs (Christen *et al.,* 2006). This non-competitive product inhibition is vital for c-di-GMP homeostasis in the cell by setting an upper limit for the levels of c-di-GMP in the cell (Christen *et al.,* 2006).

The characterisation of GGDEF domains as an active DGC depends on the conservation of the GG[D/E][D/E]F motif as well as the RXXD I-site upstream of this (Ryjenkov *et al.,* 2005; Sommerfeldt *et al.,* 2009).

X-ray crystallography of several DGC proteins has revealed a common secondary structure between proteins, consisting of a 5 β-strand core surrounded by 5 or 6 α helices, with 2 short parallel β-strands at one end of the protein (Chan *et al.,* 2004; De *et al.,* 2009; Navarro *et al.,* 2009; Forouhar *et al.,* unpublished). Dimerisation of GGDEF domains occurs by the formation of various ionic and hydrophobic interactions between α-helices α4 and α5 of each domain, forming the dimerisation interface (Chan *et al.,* 2004). The active sites from both GGDEF domains (located at a β-hairpin) are inward facing, towards the other GGDEF, whilst both I-sites are on the outer faces of the domains (Chan *et al.,* 2004).

The mechanism of DGC catalytic activity was serendipitously determined by the crystallization of PleD in complex with c-di-GMP (Chan *et al.,* 2004). In the crystal, a c-di-GMP molecule was found bound to tetrameric PleD. This GGDEF tetramer was composed of two dimers, associated head-to head, with c-di-GMP interacting with two GGDEF active sites, one from each of the two dimers (Chan *et al.,* 2004). This tetrameric species was also identified in WspR, which was also crystallized in complex with c-di-GMP (De *et al.,* 2009). Therefore it is believed that this transitional tetrameric formation is the general mechanism by which DGC proteins catalyse c-di-GMP synthesis.

The crystal structure of PleD was also instrumental in determining the manner in which c-di-GMP feedback inhibition occurs. PleD consists of two REC domains and a GGDEF domain, which then dimerises and produces two interacting GGDEF domains and a stem region consisting of four REC domains. Two intercalated c-di-GMP molecules were identified bound to the DGC- stem interface, with each c-di-GMP forming interactions with the other c-di-GMP molecule as well as to the REC domain and the GGDEF domain (Chan *et al.,* 2004). This binding explains the mechanism of c-di-GMP inhibition, with domain immobilization preventing the continued catalytic activity of PleD (Chan *et al.,* 2004). This domain immobilization mechanism of c-di-GMP feedback was also identified in WspR, suggesting a probable DGC-wide mechanism (De *et al.,* 2009).

1.6.2 Cyclic di-GMP-specific phosphodiesterase proteins

Phosphodiesterase enzymes are required for c-di-GMP catabolism and constitute the enzymatic activity characteristic of EAL or HD-GYP domain proteins, with *E. coli* containing only EAL proteins (Chang *et al.,* 2001; Galperin, 2004; Simm *et al.,* 2004; Schmidt *et al.,* 2005). The EAL domains are characterized by an EAL consensus motif at the active site, which is the location of c-di-GMP hydrolysis (Schmidt *et al.,* 2005). There are 10 consensus residues which are necessary for PDE catalytic activity (Tchigvintsev *et al.,* 2010). These conserved residues include the EAL motif (actually consensus EXL) and seven other residues, essential for general acid-base catalysis and for c-di-GMP binding (Rao *et al.,* 2008; Tchigvintsev *et al.,* 2010).

EAL domain proteins are widely distributed amongst *Bacteria* and are enzymatically specific to c-di-GMP. The EAL domains hydrolyse c-di-GMP into an intermediary pGpG (5'-phosphoguanylyl-(3'-5')-guanosine) molecule, which is later broken down to 2 molecules of GMP by an unknown PDE-B enzyme (Figure 1.4) (Chang *et al.,* 2001; Simm *et al.,* 2004; Schmidt *et al.,* 2005; Ross *et al.,* 1990; Tamayo *et al.,* 2005).

EAL-derived PDE activity functions in a two-metal ion catalytic mechanism, requiring two cations, either Mg^{2+} or Mn^{2+}, to be bound to the EAL domain (Figure 1.5) (Schmidt *et al.,* 2005;Tamayo *et al.,* 2005; Tchigvintsev *et al.,* 2010). X-ray crystallography revealed the role of the divalent ions in c-di-GMP hydrolysis. The crystal structure of TBD1265 in *T. denitrificans* in complex with c-di-GMP revealed that two metal ions were bound to the

active site via six conserved residues. The metal ions, along with the conserved residues of the protein, were responsible for the coordination of c-di-GMP as well as a potentially catalytic water molecule (Tchigvintsev *et al.,* 2010).

Whilst Mg^{2+} and Mn^{2+} have repeatedly been shown to activate PDE activity, it has been found that other cations (Ca^{2+} and Zn^{2+}) had the opposite effect, inhibiting the PDE activity of VieA in *V. cholerae* (Tamayo *et al.,* 2005).

The oligomeric state of active EAL domains has been investigated following the discovery of the dimerisation of GGDEF domains for functional DGC activity. The *E. coli* PDE YahA was found to be monomeric in solution, indicating the catalytic activity of monomeric EAL domains (Schmidt *et al.,* 2005). However crystallographic studies of EAL proteins revealed a dimeric structure. YkuI (from *Bacillus subtilis*) is an EAL- protein, identified as a catalytically inactive monomer in solution, but a dimeric oligomer in a crystal lattice (Minasov *et al.,* 2009). This raises the question whether dimerisation of EAL domains may have a regulatory role in controlling their activity.

As with the DGC family, X-ray crystallography has provided the protein structures of EAL proteins and revealed structural similarities within the family. The EAL domains have been shown to have a central 8 β-barrel structure surrounded by 9 or 10 α-helices (Navarro *et al.,* 2009; 2011).

Whilst the hydrolysis of c-di-GMP into the linear pGpG molecule is well characterised, the manner in which pGpG is further degraded into GMP is relatively unknown. EAL domains have been identified to hydrolyse pGpG into GMP, but many orders of magnitude slower than the hydrolysis of c-di-GMP into pGpG and thereby have been deemed to be physiologically irrelevant (Ross, 1987; Tal *et al.,* 1998). Therefore, it is widely believed that a specific protein, currently of unknown identity, is present in all c-di-GMP signalling bacteria, which has PDE-B activity.

1.7 The multiplicity of GGDEF/EAL domains

The GGDEF and EAL domains are almost ubiquitous in *Bacteria*, however they are absent in *Archaea* and *Eukarya*, indicating that these domains are specific to the bacterial genomes

(Galperin, 2004; Römling *et al.,* 2005). The number of GGDEF and EAL/HD-GYP domain proteins is species dependent, with some bacterial species having 100's (*Vibrio vulnificus*) whilst others lack any (*Helicobacter pylori*) (Galperin, 2004).

This multiplicity of GGDEF/EAL proteins was initially puzzling, given that the proteins carried out the same two catalytic reactions. A partial explanation for this multiplicity was found by studying the expression conditions of the 29 GGDEF/EAL domain proteins in *E. coli* (Sommerfeldt *et al.,* 2009). Gene profiling revealed differential expression of the GGDEF/EAL domain genes, dependent on growth conditions such as temperature, solid or liquid medium and the bacterial growth phase. Additionally, some genes were under the control of the general stress response sigma factor, RpoS (Sommerfeldt *et al.,* 2009). It was therefore clear that the GGDEF/EAL proteins could have independent and specialised functions in the cell, presumably able to detect specific stimuli and elicit specific effector functions.

Clearly under certain growth conditions, only a limited number of the 29 GGDEF/EAL proteins are expressed and therefore present in the cell. However, the question remains how these limited GGDEF/EAL domains produce independent and specific effects and how cross-talk is avoided. Various hypotheses have been proposed, although it is highly probable that a variety of these regulatory systems work in parallel to fine-tune this network.

Micro-compartmentalisation is one theory in which c-di-GMP is made in various pools in the cell and only acts on co-localised effector proteins (Weber *et al.,* 2006). An example of this is in *C. crescentus*, where various GGDEF/EAL domain proteins are localised to the cell poles, thereby regulating c-di-GMP pools directly at the cell poles and thus activating cell pole effector proteins. In this way, GGDEF/EAL proteins in *C. crescentus* regulate cell cycle progression (Duerig *et al.,* 2009). The co-localisation of DGCs and PDEs with their targets has been experimentally demonstrated in several cases. In *G. xylinus*, proteins DgcA and PdeA co-purify with the cellulose synthase enzyme, indicating their specific function of regulating cellulose production (Ross *et al.,* 1987).

Whilst it is known that differential expression of the GGDEF/EAL proteins occurs, the levels and activity of the proteins may also be regulated. For example, the rate of protein

proteolysis may be controlled, in order to degrade certain GGDEF/EAL proteins upon a change in condition (Hengge, 2009). Additionally environmental or intracellular stimuli may be necessary to regulate the activity of DGCs and PDEs. This has been shown to be true for several proteins, including PleD, in which phosphorylation of the N-terminal REC domain is required for PleD enzymatic activity (Paul *et al.,* 2004; Chan *et al.,* 2004; Ryjenkov *et al.,* 2005).

An alternative to the micro-compartmentalisation hypothesis is the existence of c-di-GMP gradients in the cell, with enzymes only becoming active at a threshold level (Römling *et al.,* 2013).

1.8 The presence of composite and degenerate domains

Sequence alignments have predicted that both active and inactive classes of GGDEF and EAL domain proteins exist (Schmidt *et al*., 2005; Suzuki *et al.,* 2006). The catalytically active classes are predicted to retain the active site motif, whilst inactive domains are identified as such by a lack of sequence conservation in the consensus GG[D/E][D/E]F and EAL (actually EXL consensus) sites, as well as the supplementary conserved residues discussed previously (Sections 1.6.1 and 1.6.2) (Schmidt *et al.,* 2005; Suzuki *et al.,* 2006; Rao *et al.,* 2008; Ryjenkov *et al.,* 2005). Over 20% of all known EAL domain proteins lack consensus residues, suggesting that a large number of catalytically inactive proteins exist (Rao *et al.,* 2008). Accordingly, experimental analysis of these proteins deviating from the consensus motifs confirms a much reduced or non-existent enzymatic activity, thereby identifying these sets of proteins as catalytically inactive or degenerate (Christen *et al.,* 2005; Tal *et al.,* 1998; Rao *et al.,* 2008; Suzuki *et al.,* 2006; Kazmierczak *et al.,* 2006).

A significant proportion of GGDEF/EAL domain proteins are in fact coupled to a second GGDEF/EAL domain, forming GGDEF-EAL domain fusion proteins. This dual domain composition is counter-intuitive, due to the catalytically opposing roles of these domains. However, the vast majority of these dual domain proteins are composed of one catalytically active and one inactive domain, with conserved and non-conserved motif sequences respectively. In these cases, the active domain carries out DGC or PDE activity whilst the catalytically inactive domain may have a regulatory or secondary function (Christen *et al.,* 2005; Lacey *et al.,* 2010).

In general, active GGDEF and EAL domains do not exist in the same protein due to their opposing roles. However, one exception exists in the form of protein MSDGC-1 in *Mycobacterium smegmatis*, which contains both GGDEF and EAL domains and exhibits DGC and PDE activity in the organism (Kumar & Chatterji, 2008). This dual enzymatic activity is yet to be seen in any other species.

In addition, some proteins exist in which both domains deviate from the consensus motifs, producing a catalytically inactive protein consisting of both GGDEF and EAL domains. An example of this is CsrD from *E. coli*. This protein acts in a GGDEF-EAL dependent manner to impart its function as a gene regulatory protein (Suzuki *et al.,* 2006).

Degenerate GGDEF/EAL proteins also exist as singular proteins, either linked to an N-terminal sensory or effector domain or present as a single domain protein. These degenerate proteins have been identified to have an ever-expanding number of diverse functions in bacteria.

1.8.1 The roles of degenerate GGDEF/EAL domains

Understanding the roles of these degenerate GGDEF/EAL domains is crucial to fully comprehending both the individual roles of these proteins and the diversity of these GGDEF/EAL domains and the c-di-GMP signaling pathway. Some of the catalytically inactive or 'degenerate' GGDEF and EAL domain proteins have been characterized and reveal the breath of functions that these proteins impart.

FimX is a GGDEF-EAL domain protein in *P. aeruginosa* in which the degenerate EAL domain is crucial for the function of the protein (Qi *et al.,* 2011). FimX is composed of a two N-terminal sensory domains (REC and PAS), a GGDEF domain and a C-terminal EAL domain. FimX localizes to a single pole of the bacterial cell, encoded by a localization motif in the REC domain, which results in type IV pili formation (Kazmierczak *et al.,* 2006; Qi *et al.,* 2011). The EAL domain lacks the consensus EAL motif and lacks efficient c-di-GMP PDE activity (Rao *et al.,* 2008). However, it has been shown that the degenerate EAL domain binds to c-di-GMP and as a result causes long-range conformational changes in the REC domain of FimX that are necessary for FimX single pole localization (Navarro *et al.,* 2009; Qi *et al.,* 2011). Upon deletion of the EAL domain, or mutation of the EVL sequence, c-di-GMP

binding affinity was lost in the EAL domain, resulting in no conformational change in REC and a loss of FimX unipolar localization, preventing correct pili assembly (Qi *et al.,* 2011).

PopA is a *C. crescentus* protein, consisting of two N-terminal REC domains and a GGDEF domain. This GGDEF domain lacks the conserved residues and is therefore an inactive DGC. Despite this, the GGDEF domain can bind c-di-GMP at its conserved I-site and in this way regulates cell cycle progression. In *C. crescentus* cells, CtrA acts as the replication initiation inhibitor and is responsible for controlling the transition of cells from G1 phase to S phase, at which point the DNA will be replicated and cell division can subsequently occur. At low intracellular c-di-GMP concentrations, CtrA is present in the cells and therefore blocks DNA replication initiation. When c-di-GMP levels are elevated, c-di-GMP binds to PopA and in doing so, activates the function of this protein. The activated PopA localizes itself and CtrA to the cell poles, where the protease ClpXP is present and degrades CtrA. Hence, PopA initiates the transition of cells from G1 phase into S phase and does so in a c-di-GMP dependent manner (Duerig *et al,* 2009).

In composite proteins, degenerate domains can affect the enzymatic activity of their neighbouring GGDEF/EAL domain. The *C. crescentus* protein CC3396 consists of a degenerate GGDEF domain and an intact and catalytically active EAL domain. The GGDEF domain, whilst lacking DGC activity, is able to bind GTP and in doing so increases the PDE activity of its partner EAL domain (Christen *et al.,* 2005).

1.9 The cyclic di-GMP pathway in *E. coli*

The *E. coli* K-12 strain was chosen as an experimental model in which to study the cyclic di-GMP pathway, due to its well characterised genome and loss of pathogenicity in the human intestine.

1.9.1 *Escherichia coli* strain K-12

Escherichia coli strain K-12 was isolated in 1922 from a patient's stool sample who was recovering from diphtheria and since then has become a widely-used prokaryotic model (Bachmann, 1972). The *E. coli* K-12 strain contains a genome of 4.6 million bases encoding 4288 protein-coding genes (Blattner *et al.,* 1997). *Escherichia coli* K-12 strain contains many copies of DGCs and PDEs (totalling 29), which are largely thought to have been produced by

transduction during evolution. In addition, small scale DNA mutations are often caused by UV radiation and certain chemicals and therefore feasibly accounts for some of the DGC and PDE sequence variations and even variations in the GGDEF/EAL motif sequence (Serres et al., 2009).

The genome size of *E. coli* K-12 is smaller than the larger genomes of pathogenic *E. coli* strains, such as enteropathogenic *Escherichia coli* (EPEC), consisting of 4.9 million bases and encoding 4703 proteins (Iguchi et al., 2009). Whilst *E. coli* K-12 encodes 29 cyclic di-GMP turnover proteins, pathogenic *E. coli* strain O104:H4(*E. coli* EPEC has not been sequenced yet) contains 2 extra DGCs, one of which (DgcX) is by far the most strongly expressed DGC observed in *E. coli* (Richter et al., 2014).

1.9.2 Cyclic di-GMP proteins in *E. coli* K-12

Escherichia coli K-12 has 29 GGDEF/EAL proteins, including the YahA protein present in some but not all *E. coli* strains (Sommerfeldt *et al.,* 2009). Of these 29 proteins, 12 are GGDEF domains, 10 are EAL domain proteins and 7 are composite GGDEF-EAL proteins (Figure 1.6). All of the GGDEF domains are in tandem with various sensory domains such as PAS, GAF, HAMP, MASE1 and MASE2 domains. Equally the majority of EAL domain proteins are in tandem with N-terminal domains; however the identity and function of these remains largely unknown.

Analysis of the 29 GGDEF/EAL proteins in *E. coli* K-12 has identified 12 of these to be active DGCs, 13 active PDEs and 4 enzymatically inactive proteins, hypothesised to be potentially regulatory proteins (Figure 1.6) (Sommerfeldt *et al.,* 2009). These potential regulators (YeaI, YdhA, YcgF and YdiV) lack the conserved GGDEF/EAL motifs and therefore were predicted to be catalytically inactive (Schmidt *et al*., 2005, Suzuki *et al.,* 2006). Due to their unlikely catalytic role, the proteins are considered probable regulatory proteins, retained in *E. coli* after years of evolutionary pressures. YcgF and YdhA have been particularly well characterised, and clearly have an important role in *E. coli* gene regulation.

YcgF is a blue light sensor consisting of an N-terminal BLUF domain and a C-terminal EAL domain (with a degenerate EAL motif). The EAL-domain has been confirmed to lack PDE activity but is essential to the role of YcgF. Following detection of blue light by the BLUF

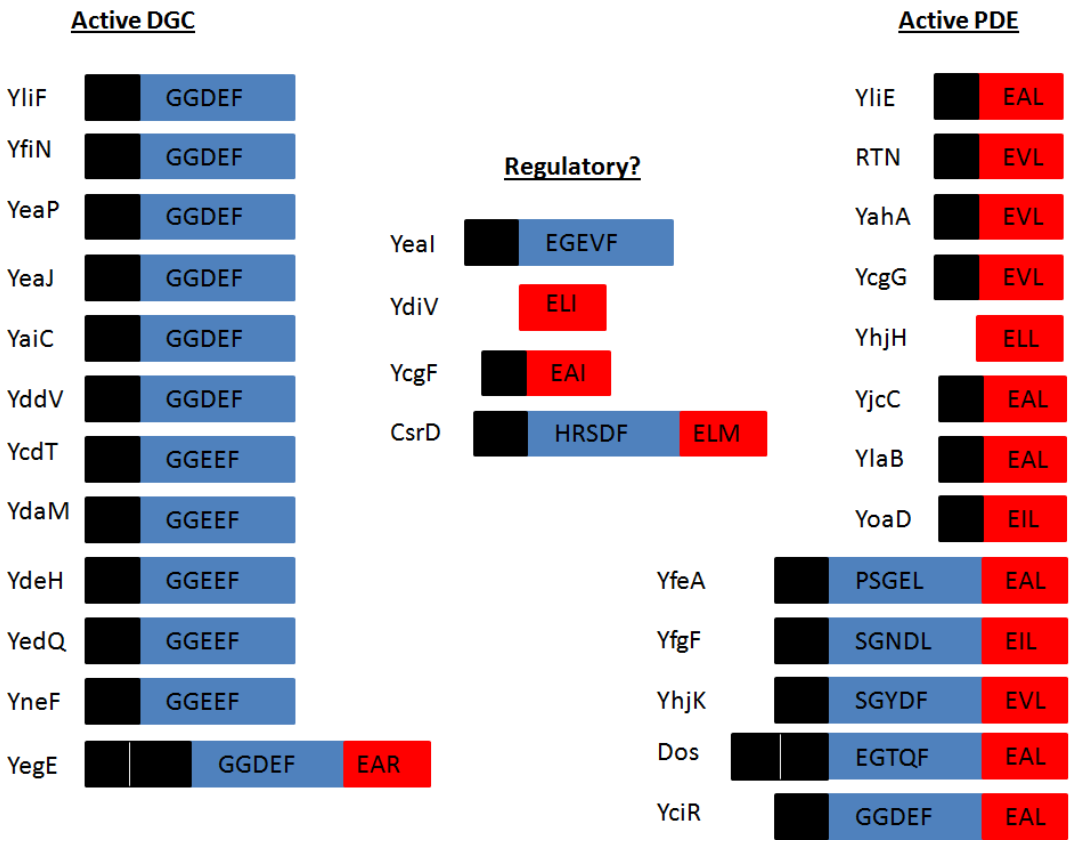


Figure 1.6: Schematic diagram of the 29 GGDEF/EAL domain proteins present in *E. coli* K-12

Each protein is named and its domain structure indicated, with all proteins being categorized into one of three categories: active DGCs, active PDEs or potential regulatory proteins.

GGDEF domains are shown in blue, with the motif sequence specified. EAL domain proteins are shown in red, again with the motif sequence specified. Black unlabelled domains represent N-terminal domains of various types.

(Sommerfeldt *et al.,* 2009; Schmidt *et al*., 2005; Suzuki *et al.,* 2006)

domain, YcgF interacts with the transcription factor YcgE in an EAL-dependent manner. This produces an active YcgE protein, which induces the expression of eight genes, including acid-resistance genes and production of biofilm-associated substances (Tschowri *et al.,* 2009).

YhdA, more commonly referred to as CsrD (Carbon storage regulator D) is a membrane bound protein containing both degenerate GGDEF and EAL motif domains. This protein is a gene expression regulator, working in the CsrA/B/C system. CsrA (Carbon storage regulator A) is an RNA-binding protein, responsible for controlling the translation and altering mRNA stability of target mRNA's. CsrA activity is tightly regulated by CsrB and CsrC non-coding RNA molecules, which bind to CsrA and prevent CsrA binding to target RNAs. These regulatory CsrB/C RNA molecules also require tight control and it is CsrD that carries out this role by altering rates of CsrB/C RNA degradation. Whilst the TM domains of CsrD are not required, the degenerate GGDEF and EAL domains are essential for this function (Suzuki *et al.,* 2006).

YeaI is less well-characterised, identified as a two-domain protein composed of an N-terminal transmembrane domain and a cytosolic GGDEF domain (Sommerfeldt *et al.,* 2009; Daley *et al.,* 2005). The GGDEF motif deviates from the consensus, indicating a putative degenerate DGC (Figure 1.6). Phenotypic analysis of the protein revealed that YeaI over-production caused cells to display increased levels of biofilm formation and sedimentation under anaerobic conditions at 28^{o}C, indicating a possible function of the protein (Lacey, M. *et al.,* unpublished).

YdiV is a cytosolic EAL domain protein. Whilst the consensus EXL motif and other conserved residues are lacking in YdiV, there have been contradictory classifications of the protein as both a catalytically active and inactive PDE (Rao *et al.,* 2008; Hisert *et al.,* 2005; Tchigvintsev *et al.,* 2010). Phenotypic analysis of strains overproducing YdiV indicated a possible role of this protein in regulating biofilm production, motility and sedimentation, all of which are c-di-GMP related phenotypes. Upon over-production of YdiV, cells exhibited decreased motility under anaerobic and aerobic conditions at 28 and 37^{o}C. YdiV over-production also caused cells to exhibit decreased biofilm formation at 28^{o}C and increased sedimentation phenotypes. Despite these clear phenotypes upon YdiV over-production,

Δ*ydiV* exhibited no detectable differences compared to wild type cells (Lacey, M. *et al.,* unpublished).

1.10 Aims of the project

The research summarized in this Introduction has shown that c-di-GMP is a crucial second messenger in bacteria. Modulation of c-di-GMP concentrations by the action of multiple opposing DGC and PDE enzymes, whose activities respond to a range of environmental cues, control fundamental 'lifestyle choices'.

Diguanylate cyclase activity is associated with GGDEF domains; phosphodiesterase activity is associated with EAL domains. *Escherichia coli* possess 12 GGDEF proteins, 10 EAL proteins and 7 composite GGDEF/EAL proteins. Intringingly, 4 (YcgR, CsrD, YeaI and YdiV) of these 29 proteins are predicted to lack either DGC or PDE activity because of the replacement of key active site amino acids. The predicted lack of catalytic activity suggested a regulatory role for these 'degenerate' proteins in c-di-GMP signaling pathways. Characterisation of YcgR and CsrD provide some support for this view. Preliminary work with *E. coli* strains over-producing YeaI or YdiV suggested a role in controlling phenotypes typically associated with c-di-GMP signaling. However, at the outset of this project how YeaI and YdiV contributed to c-di-GMP signaling in *E. coli* was unknown. The hypothesis tested was that YeaI and YdiV are catalytically inactive GGDEF and EAL domain protens, respectively, that act as effector proteins in c-di-GMP signaling pathways. Therefore the aim of the research described in this thesis was to isolate and characterize the YeaI and YdiV proteins.

Due to the likely catalytic inactivity of these proteins as DGCs and PDEs, comprehending their roles, potentially as c-di-GMP effector proteins, is crucial to understanding how c-di-GMP cellular effects are achieved. Currently, the characterisation of degenerate proteins has revealed numerous cellular effects, ranging from pili assembly, targeted protein degradation and gene regulation. In some cases, the activation of degenerate proteins correlates to signalling molecules binding to the protein (i.e. c-di-GMP or GTP), however some proteins are activated independently of signal molecule binding. It is hoped that understanding these degenerate GGDEF/EAL proteins will increase our knowledge and

understanding of both the breadth of c-di-GMP signaling pathways and the contributions made by these proteins.

Chapter 2: Materials and methods

2.1 Strains and plasmids

Escherichia coli strains and plasmids used are listed in Tables 2.1 and 2.2 respectively.

Table 2.1: Strains of *Escherichia coli* used in this study

Strain	Relevant characteristics	Source of strain
MG1655	F^-,λ^- ,*ilvG*$^-$*rfb*$^-$*50rph*$^-$*1*	Lab collection
BL21 λ(DE3)	Lysogen of λDE3 carrying a copy of T7 RNA polymerase gene under the control of the IPTG-inducible *lacUV5* promoter	Novagen
DH5α	*SupE44, lacU169 (*ϕ80 *lacZ* m15), *hsdR17, recA1, endA1, gyrA96, thi-1, relA1*	Novagen
JRG6578	MC1000 Δ*fliC*, FRT::Km::FRT	Dr Graham Stafford, University of Sheffield
JRG6577	MC1000 *flhB*::*lacZ* (Gm), *flhDC*::Km (Kn^R)	Dr Graham Stafford, University of Sheffield
JRG6097	W3110 *ΔlacU169*::Tn10, F^-, lambda$^-$, *thyA36, deoC2,* IN(*rrnD-rrnE*)I, *Δ(argF-lac)U169*, *ydiV*::*lacZ* (Tc^R)	Prof. R. Hengge, Institut für Biologie, Berlin
JRG6130	MG1000Δ*ydiV*	Lab collection
JRG6058	MG1655Δ*fnr* pBR322	Lab collection
JRG6134	MG1655 pBR322:*ydiV*	Lab collection
JRG5512	DH5α/pGS2263	Lab collection
JRG5715	DH5α/pGS2341	This work
JRG6551	DHSα/pGS2399	This work
JRG6450	DH5α/pGS2354	This work
JRG6497	DH5α/pGS2368	This work
JRG6601	DH5α/pGS2441	This work
JRG6513	DH5α/pGS2381	This work
JRG6514	DH5α/pGS2382	This work
JRG6656	DH5α/pGS2467	This work
JRG6664	DH5α/pGS2472	This work

Tc^R indicates strains containing tetracycline resistance, Kn^R indicates kanamycin resistant strains.

Table 2.2: Plasmids used in this study

Plasmid	Relevant characteristics	Source
pET21a	Expression vector encoding C-terminal His tag (Ap^R)	Novagen
pGEX-KG	Expression vector encoding GST tag (Ap^R)	Lab stock
pCOLD TF	Cold-shock expression vector encoding N-terminal TF chaperone tag (Ap^R)	Takara
pGS2468	pET28a:*slyA* (Kn^R)	Lab stock
pGS2263	pET32a: $yeaI_{GGDEF}$ (Ap^R)	Lab stock
pGS2341	pET21a:$yeaI_{GGDEF}$ (Ap^R)	This work
pGS2399	pCOLD TF: $yeaI_{GGDEF}$ (Ap^R)	This work
pGS2354	pET21a: *ydiV* (Ap^R)	This work
pGS2368	pET21a:*clp* (Ap^R)	This work
pGS2441	pET21a:*flhDC* (Ap^R)	This work
pGS2381	pET21a:$FLAG_3$ (Ap^R)	This work
pGS2382	pET21a:$FLAG_3$:*flhDC* (Ap^R)	This work
pGS2467	pGEX-KG:*fliT*(Ap^R)	This work
pGS2472	pGEX-KG:*fliZ* (Ap^R)	This work

Ap^R indicates plasmids containing ampicillin resistance and Kn^R indicates kanamycin resistance plasmids.

2.2 Growth of *E. coli*

2.2.1 Growth media

Escherichia coli was grown in Lennox broth (LB) medium, the constituents of which are listed below. Infrequently, *E. coli* was grown in motility broth in order to analyse motility of any type, the components of this medium are listed below. The ingredients were placed in a beaker and made up to 1 litre with dH_2O before transferring to a 1 L Duran bottle and then autoclaved at 121^oC, 15 psi in order to sterilise. Media constituents were supplied from Oxoid Ltd.

LB (litre): 10 g NaCl, 10 g Tryptone, 5 g yeast extract

Motility broth (litre): 5 g NaCl, 5 g Tryptone

LB agar (litre): 10 g NaCl, 10 g Tryptone, 5 g yeast extract, 15 g bacteriological agar

Motility broth agar (litre): 5 g NaCl, 5 g Tryptone, 3 g bacteriological agar

Agar was autoclaved and cooled to 50°C before antibiotics were added and pouring into petri dishes, approximately 25 ml per dish. Once the agar plates were set, they were stored at 4°C. When antibiotic resistance was present, the relevant antibiotic was added to the media at the appropriate concentration (Section 2.2.4).

Chemostat samples of *E. coli* MG1655 cells were provided by Thomas Curran. These bacteria were grown in Evans minimal medium. The recipe for Evans minimal medium is listed below.

> Evans minimal medium (20 litres): 100 ml 2 M $NaH_2PO_4.2H_2O$, 100 ml 2 M KCl, 100 ml 0.25 M $MgCl_2$, 500 ml 4 M NH_4Cl, 100 ml 0.4 M Na_2SO_4, 100 ml 0.004 M $CaCl_2.2H_2O$, 100 ml of trace elements and 7.6 g of nitrilotriacetic acid ($C_6H_6NNa_3O_6$).
>
> Trace elements (2.5 litres): 1.03 g ZnO, 13.5 g $FeCl_3.6H_2O$, 5 g $MnCl_2.4H_2O$, 0.43 g $CuCL_2.2H_2O$, 1.19 g $CoCl_2.6H_2O$, 0.16 gH_3BO_3, 0.01 g $Na_2MoO_4.H_2O$ and 20 ml of (37%) HCl.

Evans minimal medium was adjusted to pH 6.95 prior to autoclaving, after which the medium was supplemented with 20 mM sterile glucose and 30 µg/ml of $Na_2SeO_3.5H_2O$ before use.

2.2.2 *Escherichia coli* growth

Escherichia coli cultures were grown in antibiotic supplemented (Section 2.2.4) growth media at 37°C with 250 rpm shaking until the required optical density OD_{600} was reached (Section 2.2.3).

Overnight cultures were produced by inoculating 5 ml of antibiotic-supplemented growth medium (Section 2.2.4) with a single colony from an agar plate, and were then subjected to incubation at 37°C and 250 rpm shaking.

When studying bacterial motility, motility broth was used to culture the cells (Section 2.2.1). In this thesis, bacteria were grown in motility broth at 37°C with 250 rpm shaking in conical flasks. Cells were subsequently cultured by low-speed centrifugation (~4000*xg*) and gentle resuspension of cell pellets in order to reduce flagellar shearing.

Samples were provided by Thomas Curran of *E. coli* MG1655 cells cultured in a glucose-limited chemostat in Evans minimal medium (Section 2.2.1). The cells were grown at pH 6.95 at 37^{o}C with 400 rpm stirring. The bacterial strain was cultured at varying growth rates, maintained by altering the dilution rate. Four dilution rates (0.05, 0.1, 0.2 and 0.5 /h) were analysed, producing samples of bacteria grown at four different growth rates. A 0.05 /h dilution rate is equivalent to a 13.9 h doubling time, a 0.1 /h dilution rate equivalent to a 6.9 h doubling time, 0.2 /h dilution rate equivalent to a 3.5 h doubling time and a 0.5 /h dilution rate equivalent to a 1.4 h doubling time.

2.2.3 Measurement of bacterial growth

Escherichia coli growth was determined by measuring the OD_{600} of an aliquot (1 ml) of culture against a growth medium blank using a Unicam UV/Vis spectrophotometer.

2.2.4 Supplementation of media

Growth media were supplemented with antibiotics to give final concentrations of 100 µg/ml ampicillin, 20 µg/ml kanamycin or 10 µg/ml tetracycline, sourced from Melford and Sigma respectively. Isopropyl β-D-1-thiogalactopyranoside (Melford) was added to a final concentration between 100 µg/ml and 20 µg/ml (to induce protein over-expression).

2.2.5 Storage of strains

Strains were stored on solid media at 4^{o}C for one month. Glycerol stocks of strains were maintained for longer storage at -20^{o}C. Glycerol stocks were made from cell pellets of 5 ml overnight cultures resuspended in 1.25 ml LB. To the cell suspension, 1 ml of 80% sterile glycerol and the required antibiotics were added.

2.2.6 Making electrically competent cells

A primary culture of *E. coli* DH5α or BL21 λ(DE3) was made in which 5 ml LB was inoculated with the strain and incubated overnight at 37°C (Section 2.2.2). An aliquot (500 µl) of primary culture was used to inoculate 50 ml of LB and this was incubated for 1.5-2 h at 37°C. After this time, (OD_{600} of 0.3-0.6) the bacteria were collected by centrifugation at

4500 x*g* for 10 min using a desk-top microcentrifuge (Sigma 1-15). The supernatant was removed and the cells were washed with 50 ml ice-cold water and re-pelleted. The cell washing was carried out a total of three times. The final time, the cells were re-suspended in the 500 µl water. Aliquots (100 µl) of cell suspensions were frozen at -80^{o}C until required.

2.2.7 Transformation of electrically competent cells

Escherichia coli cells were transformed by the addition of 1 µl plasmid DNA (up to 1 µg) to 100 µl of electrically competent cells (Section 2.2.6). This mixture was transferred to an ice cold 1 mm electroporation cuvette (Gene Flow) and cell-shocked (1800 V, 1 mm path length) in an electroporator (Hybaid CS-100). Lennox broth (1 ml) was added to the shocked cells and the mixture was incubated at 37^{o}C for 1 h. Cells were then pelleted at 3500 x*g* for 5 min before being resuspended in 50 µl LB and plated onto LB agar plates containing the relevant antibiotic and incubated at 37°C overnight.

2.3 DNA techniques

2.3.1 Genomic DNA isolation

Genomic DNA of *E. coli* MG1665 was isolated using a Genomic DNeasy kit (QIAGEN) from an aliquot (700 µl) of an overnight culture in LB broth. By following manufacturer's guidelines, this method produced DNA of approximately 50 ng/ml.

2.3.2 Primer design

Primers were designed to be approximately 15-35 nucleotides long, have a GC content of 50-60% and a melting temperature of between 55 and 70^{o}C. Restriction sites were incorporated as needed and preceeded by the sequence TTTT.

Occasionally primers were designed to contain modifications such as 5' Biotin tags for DNA immobilisation to streptavidin surfaces. When sequences were required to be inserted into vectors, for example a $(FLAG)_3$ tag, oligonucleotides were designed for the entire inserted region and therefore were significantly longer than the usual oligonucleotides designed for gene amplification. Primers used in this work are listed in Table 2.3.

Table 2.3: List of primers used in this study

Primer	Description of use	Nucleotide sequence
NSW9	pET21a:*ydiV* forward	TTTTCATATGAAGATTTTTTTGGAGAATCTTTATC
NSW10	pET21a:*ydiV* reverse	TTTTCTCGAGTCGCTGAACCAACGTCGTTAT
NSW11	pET21a:*yeaI* forward	TTTTCATATGAAGGACTTCTGTGTCATGCTG
NSW12	pET21a:*yeaI* reverse	TTTTCTCGAGAATATGATGAAGAATCACTTTGTT
NSW13	pET21a:*clp* forward	TTTTCATATGAGCCCAGGAAATACGAC
NSW14	pET21a:*clp* reverse	TTTTCTCGAGGCGGGTGCCGTACAGCAC
NSW15	pET21a: FLAG forward	TATGGACTACAAGGACGATGACGACAAGGACTACAAGGA CGATGACGACAAGGACTACAAGGACGATGACGACAAGG
NSW16	pET21a: FLAG reverse	GATCCCTTGTCGTCATCGTCCTTGTAGTCCTTGTCGTCATCG TCCTTGTAGTCCTTGTCGTCATCGTCCTTGTAGTCCA
NSW17	*flhDC* into pET21a: FLAG forward	TTTTGGATCCCATACCTCCGAGTTGCTGAAA
NSW18	*flhDC* into pET21a:FLAG reverse	TTTTAAGCTTTTAAACAGCCTGTACTCTCTGTTC
NSW19	pCOLD:*yeaI* forward	TTTTCATATGATTCAGTCAACACGTATTTCCA
NSW20	pCOLD:*yeaI* reverse	TTTTGGATCCTTAAATATGATGAATAATCACTTTG
NSW21	pCOLD sequencing forward	GCGAAAGTGACTGAAAAAGAA
NSW22	pCOLD sequencing reverse	ATCGATTATTTATTTCCTGAAAAC
NSW24	P*ydiV* reverse	TTTTGGTAACCGTCCCTCTCGAA
NSW27	P*fliD* forward	TTTTTCTAGAGATTCGTTATCCTATATTGCAAGT
NSW28	P*fliD* reverse	TTTTGCGATTTCCTTTTATCTTTCG
NSW35	pET21a:*flhDC* forward	TTTTGGATCC ATGCAT ACC TCC GAG TTG CTG
NSW36	pET21a:*flhDC* reverse	TTTT CTCGAG AAC AGC CTG TAC TCT CTG TTC
NSW39	IVT FliD forward	TTTTGGATCCGATTCGTTATCCTATATTGCAAGT
NSW40	IVT FliD reverse	TTTTGAATTCGCGCGCTTTTCAGCGTACC
NSW59	pGEX-KG: *fliT* forward	TTTTGGATCCAACCATGCACCGCATTTATATTTC
NSW60	pGEX-KG:*fliT* reverse	TTTTAAGCTTTCAAAAGAGGTTATCCTGCGGAG
NSW61	pGEX-KG:*fliZ* forward	TTTTGGATCCATGGTGCAGCACCTGAAAAGAC
NSW62	pGEX-KG:*fliZ* reverse	TTTTAAGCTTTTAATATATATCAGAAGAAGGCAGGC
NSW65	pGEX-KG sequencing forward	GACCATCCTCCAAAATCGGA
NSW66	pGEX-KG sequencing reverse	GAGGTTTTCACCGTCATCAC

NSW84	Blitz target forward	Biotin-AGCCATTTTTTGTTAGTCGCCGAAATACTCTTTTCTC TGCCCCTTATTCCCGCTATTAAAAAAAACAATT
NSW85	Blitz control forward	Biotin- TGCGTCATCCTTCGCGCTGT
NSW86	Blitz control reverse	TTTTGACTTGCAATATAGGATAACG

2.3.3 Polymerase Chain Reactions

Polymerase Chain Reactions (PCR's) were carried out to amplify specific genes or regions of DNA from *E. coli* MG1655 genomic DNA (Section 2.3.1). Reactions (total volume 20 µl) contained 10 µl 2x Extensor long range PCR master mix (Thermo Scientific), 1 µl each of forward and reverse primers (100 pmol/µl) (Section 2.3.2), 1 µl of *E. coli* MG1655 template plasmid DNA (~200 ng/µl) (Section 2.3.1) and 7 µl sterile water.

The PCR reactions were run using a cycling programme on a Techne® TC-3000 machine. A typical programme is shown below, which was altered depending on the primer and DNA characteristics, following Extensor PCR master mix manufacturer's guidelines.

Initial denaturation	94°C	5 min	
Denaturing	94°C	30 s	x 30 cycles
Annealing	50°C	30 s	
Extension	72°C	1 min	
Final extension	72°C	5 min	

Samples were then separated by agarose gel electrophoresis for analysis (Section 2.3.5).

2.3.4 Plasmid purification

Plasmid DNA was purified from an aliquot (5 ml) of *E. coli* primary culture using a QIAprep® Spin Miniprep Kit (QIAGEN) and following manufacturer's instructions.

2.3.5 Agarose gels

Agarose gels (1%) were used as standard, consisting of 1 g of agarose (Melford) dissolved in 100 ml 1 x TAE buffer. To visualise smaller DNA fragments, 1.5% agarose gels were used (1.5 g agarose per 100 ml). Gel red (Biotium) was added to the solution at a 1/1000

dilution to allow UV detection of DNA, and the gel poured into a sealed tray fitted with a comb. Once the agarose was set, the comb was removed and 1 x TAE buffer added to the tank. DNA samples were mixed with 5 x loading dye (Bioline) and then loaded into the wells of the gel, alongside a DNA marker. Depending on the nature of the DNA to be analysed, samples varied between 5-20 µl (including loading dye and water if necessary). To check PCR reactions, 2-3 µl volumes of DNA were analysed, whilst gel extraction techniques required loading of 15-20 µl DNA. The gel was then run at 100 V for 60 min (1% gels) or 90 min (1.5% gels) using a Power Pac 300 (Bio-Rad) and DNA visualised using a UV transilluminator (UVITEC).

The recipe for 50 x TAE buffer (1 litre) was 242 g Tris, 57.1 ml glacial acetic acid, 100 ml 0.5 M EDTA, pH 8. Hyperladder I (Bioline) was used as a standard DNA marker, added to each agarose gel in order to determine the size of analysed DNA.

2.3.6 Restriction digestion of DNA

Restriction digests were carried out to cleave DNA at specific points and enzymes were purchased from MBI Fermentas. In all cases, 20 µl double digests were carried out using 14 µl DNA (0.5-1 µg), 1 µl (10 units) of each enzyme, and 2 µl of appropriate 10 x buffer (unless stated by manufacturer otherwise). The buffer was chosen to allow both enzymes to have full, or almost full enzymatic activity, according to the manufacturer's instructions.

Digests were incubated for 90 min at 37^{o}C before analysis on an agarose gel to check that cleavage had occurred (Section 2.3.5).

2.3.7 DNA isolation and purification from an agarose gel

Aliquots (2 µl) of PCR's (Section 2.3.3) were separated on an agarose gel (Section 2.3.5) to check for correct DNA amplification, before the remainder of the sample was loaded on a second agarose gel for purification. Using the QIAquick Gel Extraction Kit (QIAGEN), the amplified DNA was excised from the gel and purified.

2.3.8 Quantification of DNA concentration

Purified DNA was quantified using an Eppendorf Biophotometer. A sample of 100 µl dH_2O was used as a blank, before measuring the absorbance of a 1/100 dilution of DNA (adding 1 µl to 99 µl dH_2O) at OD_{260}. This was repeated in triplicate and an average calculated.

2.3.9 DNA Ligation reactions

Ligation reactions were carried out overnight at 4°C as 20 µl reactions. The equation below allows the calculation of the appropriate amount of insert and vector for each reaction, including 2 µl 10 x T4 DNA Ligase buffer and 1 µl T4 DNA Ligase enzyme (10 units) (Thermo scientific).

$$\frac{\text{ng vector} \times \text{Kb insert}}{\text{Kb vector}} \times 3/1 = \text{ng insert}$$

The ligation reaction was used to transform competent *E. coli* DH5α cells and the transformants were selected by plating on antibiotic supplemented LB plates (Section 2.2.7). The plates were then incubated overnight at 37°C.

2.3.10 DNA purification

DNA purification was carried out using a "QIAquick PCR Purification Kit" (QIAGEN) and following the manufacturer's instructions.

2.3.11 Screening of transformants

To check whether transformed bacteria (Section 2.2.7) contained the insert or just re-ligated plasmid, they were screened using PCR reactions. Transformants were picked from relevant agar plates using a sterile tip and stabbed onto another LB plate (supplemented with the relevant antibiotic). Using the same tip, the same transformant was inoculated into a 20 µl PCR reaction, consisting of 18 µl 1.1 X ReddyMix PCR master mix (Thermo scientific) and 1 µl of each forward and reverse primer (100 pmol/µl). This was repeated for 5-15 colonies per transformation.

A typical PCR programme used was:

- Initial denaturation 94°C 5 min
- Denaturing 94°C 30 s
- Annealing 50°C 30 s
- Extension 68°C 2 min
- Final extension 68°C 7 min

} x 30 cycles

PCR products were then separated on agarose gels alongside a control of undigested plasmid (Section 2.3.5). Candidate plasmids were purified using a QIAprep Mini-prep kit (QIAGEN) and sent to GeneService for sequencing at Source Bioscience LifeSciences plc (Section 2.3.4).

Once confirmed that the ligation was successful, the plasmid was stored and if for protein overproduction work, the plasmid was used to transform electrically competent *E. coli* BL21 (λDE3) cells (Section 2.2.7).

2.4 Gene expression analysis

2.4.1 β-Galactosidase assays

β-Galactosidase assays were used to analyse *ydiV* promoter activity in the presence of various AHL chemicals. The strains of interest were grown aerobically overnight in LB medium before sub-culturing until mid-log phase was achieved (Sections 2.2.1 and 2.2.2). Initially all assays used strain *E. coli* JRG6097, before a *slyA* over-production plasmid was used to transform this strain for additional studies (JRG6097/pGS2468) (Sections 2.3.4 and 2.2.7) (Table 2.1 and 2.2).

To carry out the β-galactosidase assays, the OD_{600} of the cells for each condition was recorded using an Unicam UV/Vis spectrometer, using the corresponding medium as a blank (Section 2.2.3). For each overnight/growth condition, a reaction was produced consisting of 550 µl Z-buffer, 20 µl chloroform, 10 µl 0.1% SDS and 250 µl cell sample. The composition of Z-buffer was 16.1 g $Na_2HPO_4.7H_2O$, 5.5 g of $NaH_2PO_4.H_2O$, 0.75 g of KCl and 0.246 g $MgSO_4.7H_2O$ per litre of dH_2O, with β-mercaptoethanol added prior to use (2.7 µl/ml).

The samples were vortexed for 10 s before incubation at 28°C for 5 min. Following equilibration at 28°C of both samples and ONPG, the reactions were initiated by the addition of 200 µl 4 mg/ml ONPG. The samples were closely monitored at 28°C, recording the time taken for each sample to turn a pale yellow colour. When the colour change occurred the reactions were terminated by the addition of 500 µl 1 M Na_2CO_3. The samples were subsequently centrifuged at 17,000 x*g* for 5 min and the OD_{420} measured (Unicam UV/Vis spectrometer).

If the solutions failed to change colour within the usual 30 min period, the reactions were left for additional incubation and stopped after a suitable time.

For each condition, three biological replicates were made, and the assays were performed in triplicate. From the raw data, the average MU (Miller units) were determined, and the standard deviation calculated (shown as error bars on the graph) for each condition. Miller Units were calculated by the following equation:

$$\mathrm{MU} = \frac{\mathrm{A420}}{\mathrm{V\ x\ T\ x\ A600}}\ \text{x } 1000$$

V= volume (µl) of cells used to find OD_{420}
T= time (min) from reaction initiation to termination

N-(β-ketocaproyl)-L-homoserine lactone and *N*-hexanoyl-DL-homoserine lactone (Sigma) were added at various stages throughout the experiments. The concentration of AHL used also varied, either at a final concentration of 1 µM or 50 µM as indicated.

2.5 Microscopy techniques

2.5.1 Transmission Electron Microscopy (TEM)

Overnight bacterial cultures (5 ml) were grown in motility broth at 37°C, before an aliquot (2 ml) was added to 50 ml motility broth in 200 ml conical flasks (Sections 2.2.1 and 2.2.2). The cells were grown until exponential phase was reached, with an OD_{600} of ~0.18 (Section 2.2.3). At this point, the bacteria were harvested at 4000 x*g* for 10 min and resuspended in

50 µl motility broth. To preserve the cells, they were added to fixing solution (3% glutaraldehyde in 0.1 M sodium cacodylate) in a 1:1 ratio, before visualisation by TEM.

For TEM visualization, the fixed samples were placed onto 400 mesh copper grids (Agar Scientific Stansted Essex) coated with a formvar film. The fixed samples were thoroughly mixed by vortexing before a small volume (6-10 µl) was added to the grid and incubated for 30 s before excess solution was drawn off by blotting paper. To stain cells, the same volume of 1% phosphotungstic acid (6-10 µl) was added for 30 s before excess was drawn off again by blotting paper.

The copper grids were then ready to be observed under a FEI G2 Bio-twin tecnai 120 kV TEM (Hillsboro) (Mr Chris Hill, University of Sheffield).

2.6 Protein methods

2.6.1 Protein overproduction

Primary cultures were set up by inoculating 5 ml aliquots of growth media (with appropriate antibiotic) with a single colony (of the strain of interest) from an agar plate, and growing overnight at 37^{o}C (Sections 2.2.1; 2.2.2 and 2.2.4). The overnight cultures were then added to 500 ml of antibiotic supplemented LB in 2 L conical flasks for incubation at 37^{o}C until an OD_{600} ~0.6 was achieved (Section 2.2.3).

IPTG was added to the cells (in all cases when the plasmid contained an inducible *lac* promoter) at a concentration of between 100 µg/ml to 20 µg/ml (Section 2.2.4) and the cells were grown for between 2 or 24 h at either 15 or 25^{o}C, depending when maximal expression and solubilisation had been shown for that specific protein. Once cultured, the bacterial cells were harvested at 18,500 x*g* in an ultracentrifuge for 30 min and the pellets were stored at -20^{o}C.

2.6.2 Cell breakage

Cell pellets were defrosted on ice and resuspended in 1-10 ml of appropriate breakage buffer dependent on the specific strain utilised (Table 2.4). Cells were then lysed by 3

rounds of 30 s sonication at 16 microns. Following this, the cell-free extract was separated into soluble and insoluble fractions by centrifugation at 27,000 x*g* for 15 min.

Table 2.4: Cell breakage buffers

YdiV-His_6, FlhDC-His_6, GST-FliT, GST-FliZ, FlhDC-$(FLAG)_3$ (For ITC, BLItz, IVT, EMSA, Antibody stimulation, pull-down assays and YdiV crosslinking, partial proteolysis and nucleotide reactivity)	20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, pH 7.5
Clp-His_6, YdiV-His_6 (NMR)	20 mM Tris, 500 mM NaCl, 5% glycerol, pH 7.4
YeaI-His_6 (nucleotide reactivity assays)	25mM HEPES, 500 mM NaCl, 5% glycerol, pH 7.5
YeaI-His_6 (partial proteolysis, glutaraldehyde)	20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, pH 7.5
TF-YeaI (purification), YeaI-His_6 (gel filtration)	20 mM Tris, 500 mM NaCl, 5% glycerol, pH 7.4

Cell pellets of over-produced proteins were resuspended into one of these various breakage buffers (unless otherwise specified). The breakage buffer chosen was dependent on both the protein overproduced and the subsequent assays the protein would be used for, as specified.

2.6.3 Affinity purification

All His_6-tagged proteins were purified using affinity chromatography methods. A 1 ml HiTrap chelating column (GE healthcare) was attached to the AKTA Prime system and initialised according to manufacturer's instructions.

Bacterial cell pellets were resuspended in the relevant breakage buffer (Table 2.4), the cells lysed by sonication and the phases separated (Section 2.6.2). The soluble fraction was filtered using a 0.2 µm filter before applying the fraction to a 10 ml loop attached to the AKTA™ Prime equipment. The His-chelating programme was used, using the relevant binding and elution buffers (Table 2.5), which separated the loaded protein sample into a number of 1 ml fractions.

Table 2.5: Affinity chromatography purification buffers

Binding buffer (frequently used for YdiV and FlhDC, YeaI for some assays)	20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.5
Elution buffer (frequently used for YdiV and FlhDC, YeaI for some assays)	20 mM sodium phosphate, 0.5 M NaCl, 0.5 M imidazole, pH 7.5
Binding buffer (NMR- YdiV, Clp, YeaI)	20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.4
Elution buffer (NMR- YdiV, Clp, YeaI)	20 mM Tris, 500 mM NaCl, 500 mM imidazole, pH 7.4
Binding buffer (His_6-YeaI)	25 mM HEPES, 0.2 M NaCl, pH 7.5
Elution buffer (His_6-YeaI)	25 mM HEPES, 0.125 M NaCl, 0.5 M imidazole, pH 7.5
Ni-loading buffer	100 mM $NiSO_4$

Three different pairs of binding and elution buffers were used for affinity chromatography, depending on the breakage buffer utilised to resolubilise the cell pellet. Whilst some buffers were chosen as optimal purification buffers for specific proteins, others were chosen based on the assay in which the protein would be used.

A trace reader showed the A_{280} levels, thereby indicating the quantity of protein in various fractions. The eluted fractions were electrophoresed on SDS-PAGE gels to identify which fractions contained the protein of interest (Section 2.6.9).

2.6.4 Anion exchange chromatography

Anion exchange chromatography was carried out as a secondary purification step for $YeaI_{GGDEF}$. Following His chelating chromatography (Section 2.6.3), the $YeaI_{GGDEF}$ protein was dialysed into binding buffer (Table 2.6) (Section 2.6.7), before anion exchange chromatography.

Table 2.6: Anion exchange buffers

Binding buffer	25 mM HEPES, pH 8.0
Elution buffer	25 mM HEPES, 1 M NaCl, pH 8.0

The protein was applied using a 10 ml loop to a 1 ml HiTrap Q HP column (GE healthcare) which was attached to the AKTA Prime system. The AKTA system was initialised according to manufacturer's instructions and the Anion Exchange programme used. The programme used buffers recommended in the manufacturer's instructions (Table 2.6). The loaded

protein was separated into a number of 1 ml fractions, recording the A_{280} levels of the eluted fractions for protein detection.

2.6.5 Resolubilisation of insoluble YeaI protein

In an attempt to re-solubilise the insoluble cell pellet containing $YeaI_{GGDEF}$ protein, a denaturing and refolding protocol was undertaken. The cell pellet was resuspended in 10 ml of binding buffer (6 M guanidine hydrochloride, 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole, 1 mM 2-mercaptoethanol, pH 8.0) and agitated at 4^{o}C for several hours. The resulting mixture was separated into soluble and insoluble fractions by centrifugation at 27,000 x*g* for 15 min. The soluble fraction was then filtered through a 0.2 µm filter and applied to a 10 ml loop attached to a 1 ml Hi-Trap Chelating column on the AKTA Prime system. The refolding programme was followed, according to the manufacturer's guidelines and recommended buffer compositions. The protein was separated into a number of 1 ml fractions, recording the A_{280} levels of these eluted fractions. The eluted fractions containing high A_{280} levels signified the presence of protein in the samples and indicated which samples to analyse by SDS-PAGE (Section 2.6.9).

2.6.6 Thrombin cleavage

In some cases, thrombin cleavage was an important purification step, required to cleave the protein of interest ($YeaI_{GGDEF}$) from the large fused tag. For all variations of this assay, samples were taken throughout each protocol, enabling SDS-PAGE analysis to determine the level of success (Section 2.6.7).

Optimisation of this protocol utilised Ni-NTA resin for small scale reactions. Cell pellets of the pCOLD TF:$YeaI_{GGDEF}$ overproduction strain (*E. coli* BL21 λ(DE3)/pGS2399) were resuspended in breakage buffer (Table 2.7), the cells broken by sonication, clarified by centrifugation and the soluble fraction retained for use (Section 2.6.2). The Ni-NTA resin was prewashed in dH_2O and then wash buffer/TCB (Table 2.7) (twice each) in a 1:1 ratio before addition of soluble protein to the resin, incubating at 4^{o}C for 1 h. Thrombin (20 units) was then added to the protein and incubated overnight at either 4 or 20^{o}C. The supernatant was removed, the beads washed once again before an imidazole rich elution buffer was added.

Large scale purifications of $YeaI_{GGDEF}$, used a 1 ml HiTrap column to purify the fusion TF-$YeaI_{GGDEF}$ protein from the crude soluble fraction, using the His-chelating programme as detailed previously (Section 2.6.3). Following affinity chromatography, the elution fractions were combined and dialysed into TCB buffer (Section 2.6.7). Following dialysis, thrombin (20 units) was added to the protein and incubated for 16 h at 20^{o}C to cleave the fusion protein into TF and $YeaI_{GGDEF}$ components. The cleaved mix was applied to a 1 ml HiTrap column either manually, or via the AKTA Prime system. The manual purification step involved the application of the protein to the column, manually washing the column with TCB and then addition of an elution buffer. The automatic purification involved protein loading to the column and an elution gradient applied using the His-chelating programme, as detailed previously (Section 2.6.3).

The buffers used for the thrombin purification step are detailed in Table 2.7.

Table 2.7: Thrombin cleavage buffers for TF:$YeaI_{GGDEF}$ purifications

Breakage buffer	20 mM Tris, 500 mM NaCl, 5% glycerol, pH 7.4
Wash buffer	20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.4
TCB	20 mM Tris, 500 mM NaCl, 2.5 mM Ca^{2+}, pH 7.4
Elution buffer	20 mM Tris, 500 mM NaCl, 500 mM imidazole, pH 7.4

2.6.7 Buffer exchange and concentrating of protein samples

Following purification, the fractions containing the protein of interest were either used directly or were buffer exchanged. Vivaspin 6 ml or 20 ml concentrator columns (Vivascience) were used to concentrate the protein and to remove small protein contaminants from the sample. Columns were carefully chosen to elute proteins no more than half the size of the desired protein i.e. a 10 KDa cut off was used for a 28 KDa protein. Samples were concentrated by centrifugation at 4000 x*g* for 5 min intervals, with Bradford assays (Section 2.6.8) carried out regularly, until the desired concentration of protein was achieved.

Vivaspin columns were used for all FlhDC and YdiV fusion proteins, both to buffer exchange and to concentrate the proteins. For buffer exchange purposes, a 20 ml concentrator would be used of the appropriate protein retention size, adding 5 ml of protein into the column. The column would be made up to 20 ml by addition of dialysis buffer and the

samples buffer exchanged via centrifugation until the column volume reached 5 ml. This buffer exchange was repeated three times, ensuring minimal unwanted buffer remnants remained.

The dialysis buffer varied depending on the protein and elution buffer conditions, but the dialysis buffer was usually identical to the elution buffer without imidazole.

For ITC and BLItz assays, as well as antibody stimulation, YdiV and FlhDC were dialysed into 20 mM sodium phosphate, 0.5 M NaCl, pH 7.5. In addition YdiV was dialysed into the same buffer for partial proteolysis, glutaraldehyde and nucleotide reactivity assays. NMR experiments used Clp and YdiV dialysed into 20 mM Tris, 500 mM NaCl, 20 mM arginine, pH 7.4.

$YeaI_{GGDEF}$ was buffer exchanged into 20 mM sodium phosphate, 0.5 M NaCl, pH 7.5 for partial proteolysis assays and glutaraldehyde assays. Gel filtration assays took place with $YeaI_{GGDEF}$ protein dialysed into 50 mM Tris HCl, 50 mM NaCl, pH 8.0. HPLC assays were carried out following dialysis of $YeaI_{GGDEF}$ into 10 mM HEPES, 0.125 M NaCl, pH 7.5.

All remaining assays did not require this buffer exchange step, utilising protein either in its crude soluble form or purified form directly in elution buffer.

2.6.8 Quantification of protein concentration

Protein concentrations were determined using Bradford assays. A blank of 200 µl Bio-Rad reagent (Bradford, 1976) and 800 µl MilliQ water was initially used, before protein samples of 1-20 µl were added and the absorbance at A_{595} determined using a Unicam HEλIOS spectrophotometer. Protein concentration was then calculated by using a standard BSA curve.

Additionally, when highly accurate protein concentrations were required, spectroscopic techniques were used. An aliquot (10-20 µl) of purified protein was diluted into the chosen protein dialysis buffer to give a 1 ml sample, which was added to a quartz cuvette. The spectrophotometric absorbencies of the protein were recorded at 280 and 320 nm. The absorbance values at these two wavelengths were then used alongside the proteins'

extinction coefficient (ExPaSy ProtParam tool) to calculate the molar concentration using the following equation:

$$\text{Concentration (M)} = (A_{280}\text{-}A_{320})/ \text{ extinction coefficient}$$

2.6.9 SDS-PAGE

SDS-PAGE gels were used to analyse proteins present throughout the over-expression and purification process, with 12% resolving gels used to visualise proteins 20 KDa upwards and 18% for smaller proteins.

SDS-PAGE was carried out using the Bio-Rad Mini Protean 3 system following the manufacturer's instructions. Glass plates (Bio-Rad) were cleaned with ethanol and clamped into position on a gel casting stand. The appropriate resolving gel was added to the plates until it reached 1 cm from the top and water-saturated isopropanol was layered on top of the gel to enable it to set with a level surface. Once fully set, the isopropanol and unpolymerised acrylamide were poured off the resolving gel and the surface of the gel washed thoroughly with dH_2O. The stacking gel was then added above the set resolving gel and the comb inserted. Once set, the gel was removed from the stand and clamped into the gel running box, where the comb was removed and 1 x running buffer added.

Samples were prepared for SDS-PAGE analysis by mixing the protein sample in a 1:1 ratio with SDS-PAGE loading dye and heating at 95^{o}C for 5 min. The samples were then added to the wells (5-25 µl), alongside 10 µl of a protein marker and electrophoresis carried out at 200 V for up to 70 min. Once complete, the gels were removed from their plates and placed into Coomassie stain for 1-2 h, and then into destain until stained polypeptides were clearly visible. Gels were dried using Invitrogen gel-drying kits, using a solution of 30% methanol and 5% glycerol.

The SDS-PAGE gel components, running buffer, sample preparation and gel staining buffers are listed in Table 2.8. The protein markers used for SDS-PAGE gels are shown in Figure 2.1.

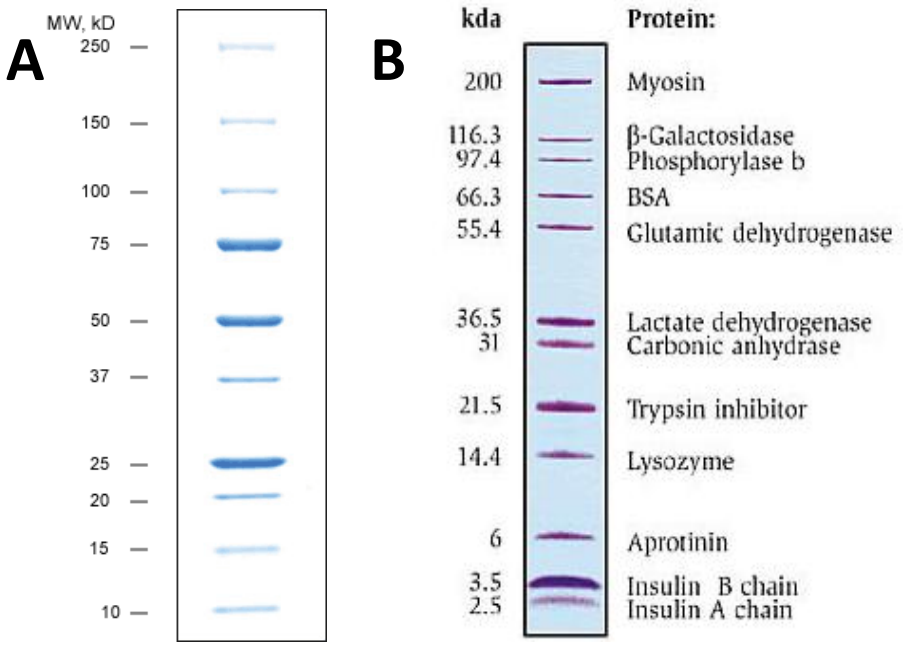


Figure 2.1: SDS-PAGE protein markers

Protein markers were used on every SDS-PAGE gel in order to determine the molecular weight of the analysed proteins. Additionally, these markers enabled the assessment of successful transfers by Western blot.

A: Precision plus protein marker used as both unstained and prestained (All Blue) forms (Bio-Rad). B: Invitrogen Mark 12 unstained marker (Invitrogen).

Table 2.8: Components of SDS-PAGE gels and buffers

Solution	Constituents
12% resolving gel	1.28 ml dH_2O, 2.5 ml 30% acrylamide (Bio-Rad), 2.35 ml 1.5 M Tris pH 8.8, 62.5 µl 10% SDS, 62.5 µl 10% APS and 6.25 µl TEMED (N,N,N',N'-tetramethyl-ethane-1,2-diamine)
18% resolving gel	3.75 ml 30% acrylamide, 2.35 ml 1.5 M Tris pH 8.8, 62.5 µl 10% SDS, 62.5 µl 10% APS and 6.25 µl TEMED
6% stacking gel	2.46 ml dH_2O, 0.75 ml 30% acrylamide, 0.47 ml 1 M Tris pH 6.8, 37.5 µl 10% SDS, 37.5 µl 10% APS, 3.75 µl TEMED
1X Running buffer	250 mM Tris-base, 0.1 M glycine, 0.1% SDS
SDS-PAGE loading dye	0.35 ml 1 M Tris HCl pH 6.8, 0.6 ml 50% glycerol, 100 mg SDS, 93 mg DTT, 50 µl 1% bromophenol blue
Coomassie stain	10% methanol, 10% acetic acid, 0.05% Coomassie brilliant blue
Destain	10% acetic acid, 10% methanol

2.6.10 Western blot

Samples to be analysed by Western blot were electrophoresed on the relevant SDS-PAGE gels, using Bio-Rad Prestained protein marker (Figure 2.1) as a visible marker (Section 2.6.9). The SDS-PAGE gel was transferred to Hybond-C Extra nitrocellulose (Amersham Biosciences) membrane using a semi-dry method. In preparation for the blot, the blotting paper and SDS-PAGE gel were soaked in transfer buffer for 20 min. Meanwhile, a piece of membrane was soaked in dH_2O for 10 min before being transferred to transfer buffer for the remaining 10 min. Keeping all components submerged in transfer buffer, the blot was then assembled, following manufacturer's instructions and placed into the Transblot Electrophoretic Transfer Cell (Bio-Rad). Once fully assembled, the transfer was carried out at 100 mA for 1 h, with gentle buffer stirring.

Once complete, the membrane was removed from the transfer cell and blocked in 50 ml of blocking buffer for 30 min. The blocking buffer was removed and the membrane washed several times (3 x) with PBS-Tween. Subsequently, primary antibody (50 ml) was added to the membrane, prepared at a suitable antibody dilution in blocking buffer, which was incubated for 1 h. Following this, the membrane was further washed in PBS-Tween (3 x) before the secondary antibody was added (50 ml) for 1 h. Anti-rabbit labelled with horse radish peroxidase (Amersham Biosciences) was the chosen secondary antibody for all experiments, used at a 1/10,000 dilution throughout. The secondary antibody was discarded and the membrane washed with PBS-Tween (3 x) prior to a final wash in PBS. All wash and incubation steps were carried out at room temperature with gentle agitation.

The membrane was then removed and drip dried before placing protein-side up on clingfilm. Western ECL reagents (Pierce) were then made up following manufacturer's instructions (6 ml) and were poured onto the membrane and incubated in the dark for 5 min. The ECL reagents were then thoroughly drained off the membrane, before the membrane was wrapped in cling film and placed in an X-ray film cassette. In a dark room, a piece of Hyperfilm ECL (GE healthcare) film was placed onto the membrane and sealed in the cassette to expose for a suitable time (20 s-10 min) before the film was removed from the cassette and developed.

The buffers used for the gel-membrane transfer and for the Western blot are listed in Table 2.9.

Table 2.9: Western blot buffer components

10 x PBS (1 litre)	80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4, 2.4 g KH_2PO_4, pH 7.4
1 x PBS-Tween	10 % 10 X PBS, 0.05 % Tween 20
Transfer buffer (1 litre)	5.8 g Tris base, 2.9 g glycine, 200 ml methanol, 3.7 ml 10% SDS
Blocking buffer	5% Marvel, 0.05% Tween 20 in 1 x PBS

2.6.11 Antibody production

Polyclonal antibodies were raised against the protein complex $FlhD_4C_2$ and the monomeric protein YdiV in rabbits, carried out by BioServe, University of Sheffield. Both proteins were purified in their His-tagged fusion form (using strains BL21 λ(DE3)/pGS2354 and BL21 λ(DE3)/pGS2441), and subsequently dialysed into sodium phosphate buffers (Sections 2.6.1; 2.6.2; 2.6.3 and 2.6.7).

Each rabbit was injected with up to 1 ml of 1 mg/ml protein by BioServe, taking test bleeds after a primary injection, and then after a second round of injections. After each test bleed, the rabbit serum was tested by Western blot analysis (Section 2.6.10) to assess antibody production, specificity and potency. The serum was used at dilutions of 1/100, 1/500, 1/1000 and 1/5000, with non-immunised rabbit sera being used as a negative control. This was carried out for each protein, for both the first and second round of rabbit injections, before the final bleed was tested, to give a starting dilution for Western blots.

The final bleed antibody serum was aliquoted into 500 µl tubes, with the majority stored at -80^{o}C for long-term use, and aliquots for current use kept at 4^{o}C.

2.6.12 Glutaraldehyde crosslinking assays

Glutaraldehye samples (20 µl) were produced, consisting of 16 µl purified protein (0.5 -1 mg/ml) (Section 2.6.8) and between 0 and 1% glutaraldehyde, making up to the final volume with the relevant protein binding buffer, as described in Section 2.6.3. The reactions were incubated at room temperature for 15 min before being quenched by 10 µl of 1 M TRIS pH 6.8. Reactions were frozen until use, when the samples were denatured by boiling at 95°C for 5 min with 30 µl SDS-PAGE loading buffer and electrophoresed by SDS-PAGE analysis (Section 2.6.9).

Crosslinking assays were carried out for both YdiV and YeaI proteins, analysing both as purified His-tagged fusion proteins (produced from overproduction strains *E. coli* BL21 λ(DE3)/pGS2354 and *E.coli* BL21 λ(DE3)/pGS2263) (Section 2.6.3).

2.6.13 Gel filtration

A Hi-Load Superdex 200GL column (GE Healthcare) was pre-equilibriated with buffer (identical to the protein preparation buffer, Sections 2.6.3 and 2.6.7) prior to the application of purified protein to the column. $YeaI_{GGDEF}$ was added (1 ml of 1.5 mg/ml) in 50 mM Tris HCl, 50 mM NaCl, pH 8.0 and YdiV was added (100 µl of 3 mg/ml) in 20 mM sodium phosphate, 500 mM NaCl, pH 7.5 buffer, with both proteins analysed individually. The column was developed at 0.7 ml/min flow rate, using the identical buffer as that used to purify the protein. An attached computer recorded the absorbance at 280 nm of the eluant and therefore identified the elution volume of each protein. The Kav (gel phase distribution coefficient) was calculated using the following equation:

$$\mathrm{Kav} = \frac{\mathrm{Ve} - \mathrm{Vo}}{\mathrm{Vt} - \mathrm{Vo}}$$

In the equation above, Ve = elution volume of protein, Vo = void volume of column (7.8 ml) and Vt = total volume of column (24 ml).

The protein molecular weight could then be estimated from the elution times, using a standard curve based on proteins of known size. This calibration curve was produced by the application of a four protein mix (Ferritin, Aldolase, Ovalbumin and Ribonuclease) of known molecular weight to the Superdex column, and determination of their respective elution volumes. The logMW of these four proteins was plotted against the Kav value for each protein and gave a calibration curve for molecular weight calculation of tested proteins.

2.6.14 N-terminal sequencing

N-terminal sequencing was used to identify co-purifying proteins produced upon YdiV and $YeaI_{GGDEF}$ purifications. YdiV and $YeaI_{GGDEF}$ were over-expressed individually and the proteins purified (detailed in Sections 2.6.1-2.6.3). In each case, the protein mixture (total concentration of 0.8 mg/ml) was initially separated on an SDS-PAGE gel (Section 2.6.9) before the gel was blotted onto an Immuno-Blot™ PVDF membrane (Bio-Rad). After a brief soaking in methanol, the membrane was transferred to an electroblotting buffer, consisting of 10 mM CAPS, 10% (v/v) methanol, pH 11. The gel was also soaked in this buffer for 5 min, before the transblotting sandwich was constructed and the electroblotted at room temperature for 30 min at 50 V using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad).

Once electroblotted, the membrane was removed from the transblotting sandwich and rinsed with dH_2O before a brief soaking in 100% methanol. The Immuno-Blot PVDF membrane was subsequently Coomassie stained (Table 2.10) for 1 min. The membrane was destained and thoroughly rinsed in dH_2O prior to the membrane being dried at room temperature. The polypeptides of interest were then excised from the membrane and amino acid sequences ascertained by Dr A.J.G. Moir (Krebs sequencing and synthesis facility) using the Applied Biosystems protein sequencer.

The electroblotting buffers are detailed in Table 2.10.

Table 2.10: Components of electroblotting buffers

Electroblotting buffer	10 mM CAPS, 10 % (v/v) methanol, pH 11
Coomassie Stain	0.1% (w/v) Coomassie brilliant blue R25O, 40 % (v/v) methanol, 1 % (v/v) acetic acid
Destain	50 % (v/v) methanol

2.6.15 Crystallisation trials

Crystallisation trials were carried out for proteins using a crystallisation robot (Matrix Hydra II PlusOne) and commercially available screens (QIAGEN). The robot was cleaned and utilised following manufacturer's instructions. Robotic trials were carried out for YdiV protein and for the YdiV:FlhDC complex. YdiV was purified and dialysed into a 20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4 buffer and concentrated to 10-15 mg/ml (Sections 2.6.3; 2.6.7).

The YdiV:FlhDC complex was produced by carrying out a pull-down assay (Section 2.8.1) using overproduction strains for theYdiV-His_6 and $(FLAG)_3$-FlhDC proteins. The pull-down was carried out as described in Section 2.8.1, using the buffers specified in Section 2.8.1. The protein complex was analysed by SDS-PAGE (Section 2.6.9) before dialysis of the protein into a 20 mM sodium phosphate, 500 mM NaCl, pH 7.5 buffer (Section 2.6.7). The YdiV:FlhDC complex was then concentrated to 9 mg/ml before robotic trials were carried out again using the Matrix Hydra II PlusOne robot and commercial screens (QIAGEN).

2.7 Protein:nucleotide interactions

Nucleotides were tested for modulation of protein behaviour in various ways, using the following range of nucleotides: GTP, GDP, GMP, cAMP, AMP, ADP, ATP (Sigma), ppGpp (TriLink BioTechnologies), c-di-GMP, c-di-AMP, pGpG (BioLog).

2.7.1 Enzymatic reactions

Reactions (15 µl) were composed of 1 µM purified protein, 10 mM Mn^{2+}/Mg^{2+} and 100 µM substrate, made up to 15 µl with dialysis buffer. The reactions were incubated for 1 h at

37°C before stopping the reaction by heating at 95°C for 5 min. This assay was carried out for both $YeaI_{GGDEF}$ and YdiV proteins, overproduced from strains *E. coli* BL21 λ(DE3)/pGS2263 and *E. coli* BL21 λ(DE3)/pGS2354 (Section 2.6.1) and purified by affinity chromatography (Sections 2.6.2; 2.6.3 and 2.6.7)

To prepare samples for HPLC analysis, the reaction samples were made up to 100 µl with the addition of dH_2O and the samples were transferred to HPLC tubes. The ELITE La Chrom HPLC machine was then initiated following manufacturer's instructions, and systematically took up programmed samples and loaded them onto a C18 LiChroCART 125-3 column. The samples were then subject to a 30 min programme of:

0 min = 100% buffer A
20 min = 20% buffer A
25 min = 20% buffer A
26 min = 100% buffer A
30 min = 100% buffer A

The buffers used in the assay were filter sterilized and consisted of:
Buffer A = 100 mM KH_2PO_4, 4 mM $(butyl)_4$ NH_4HSO_4P pH 5.9
Buffer B= 66% methanol

The flow rate was set at 0.7 ml/min with the sample size being 45 µl for all samples. To standardise the column, a mixture of GTP/GDP/GMP (each at 1 mM) were loaded onto the column.

2.7.2 Partial proteolysis

Partial proteolysis methods were carried out using recombinant purified protein, creating a characteristic cleavage pattern for each protein by the addition of trypsin to samples. This method was used to analyse nucleotide:protein interaction by incubating proteins with various nucleotides and comparing the trypsin cleavage patterns. Both $YeaI_{GGDEF}$ and YdiV were analysed by partial proteolysis, as His-tagged fusion proteins overproduced from strains *E. coli* BL21 λ(DE3)/pGS2263 and *E. coli* BL21 λ(DE3)/pGS2354 (Section 2.6.1) and purified by affinity chromatography (Sections 2.6.2; 2.6.3 and 2.6.7).

Optimal conditions for trypsin digests of protein were determined by using various trypsin concentrations (stocks of 1 mg/ml and 0.2 mg/ml, made in 0.1 M HCl), incubation temperatures (37^{o}C or RT) and incubation times (1-10 min). Trypsin stocks were used, adding 3 µl of stock solution into each 30 µl reaction sample, producing samples containing trypsin at 100 µg/ml (from 1 mg/ml stock) and 20 µg/ml (from 0.2 mg/ml stock). Samples were set up in tubes, containing 10 µl YeaI (1 mg/ml) or 5 µl YdiV (3 mg/ml) purified protein, 3 µl of trypsin at required concentration, and made up to 30 µl with 10 mM Tris-HCl, pH 8. These samples were set up, before incubation at the optimal temperature and duration, before the reactions were stopped by the addition of 4 µl 10% SDS and heating at 100^{o}C for 10 min. SDS-loading dye was added to the samples before electrophoresis on a high percentage SDS-PAGE gel (Section 2.6.9) along with a control sample lacking any trypsin.

Once optimal conditions had been determined, the reactions were repeated with the addition of nucleotides (final concentration of 2 mM), ensuring to produce protein and nucleotide control sample lacking trypsin, as well as the same reaction mix in the presence of trypsin. Once more these samples were separated by SDS-PAGE methods, enabling comparisons between samples.

2.7.3 Nuclear Magnetic Resonance

Proton NMR was carried out for YdiV in the presence of various nucleotides (pGpG, c-di-GMP, ppGpp, GMP, GDP and GTP) as well as analysing the YdiV protein and the individual nucleotides independently.

Conditions for NMR were optimized to get the best signal to noise, with the best detection of nucleotides/proteins being achieved when using a Tris buffer (20 mM Tris, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM $Mg^{2+}$and 0.5 mM EDTA. Samples were prepared to give a final concentration of 30 µM protein, nucleotide, or of both protein and nucleotide in the optimal conditions. Purified YdiV-His_6 protein was used after dialysing the protein into the Tris buffer detailed above (Sections 2.6.3; 2.6.7).

Samples were prepared to give a final volume of 450 µl, before 50 µl D_2O was added and the solutions mixed. The samples were then analysed by Dr. Andrea Hounslow, University

of Sheffield. The samples analysed using on a Bruker 800MHz Avance I NMR spectrometer and Bruker Topspin2.1 software.

2.7.4 Fluoresence spectroscopy

The fluorescence spectra of proteins in the presence and absence of nucleotides were obtained using a Cary Elipse fluorescence spectrophotometer. An excitation wavelength of 280 nm was used, scanning the emission between 300 nm and 400 nm. The excitation slit was set to 5 nm, and emission slit set to 20 nm.

For all experiments, a water bath was utilised, enabling the temperature to be controlled (26, 30 or 35^{o}C) throughout the series of experiments.

For YdiV samples, 0.4 µM purified YdiV-His_6 protein (in 20 mM Tris, 500 mM NaCl, pH 7.5) (Section 2.6.3 and 2.6.7) and 12 µM Mg^{2+} (giving a 30-fold excess) were added giving a final volume of 3 ml. Nucleotides were added to the samples to give a 1:1 molar ratio to the protein, i.e. at a final concentration of 0.4 µM. In all cases, the sample buffer was 10 mM Tris-HCl, pH 7.4. The reaction samples were analysed in 3 ml quartz cuvettes, at a specified constant temperature. Reference samples using N-acetyl-L-tryptophanamide (NATA) (Sigma) were used as a tryptophan analogue, adding NATA at a concentration to give a similar fluorescence intensity as YdiV. In this case, 3 ml samples were produced consisting of 2.3 µM NATA, again with nucleotides at 0.4 µM, Mg^{2+} at 12 µM, all in a 10 mM Tris-HCl, pH 7.4 buffer.

The YdiV:FlhDC fluorescence detection was carried out a similar manner, with each 3 ml sample consisting of 50 µM FlhDC:YdiV complex, 1.5 mM Mg^{2+} (30-fold excess), with nucleotides added in a 1:1 ratio to the protein, again at 50 µM. The sample buffer in this case was 20 mM sodium phosphate, pH 7.5, analysing the samples at a specified constant temperature. The FlhDC:YdiV complex used in this assay was formed by a pull-down method (Section 2.8.1), using the buffers specified in Section 2.8.1. The complex was eluted in 20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.5. Complex formation and purification was confirmed by SDS-PAGE (Section 2.6.9).

2.7.5 Pull-down assays with nucleotides

Ni-NTA beads (QIAGEN) were used as a small-scale protein immobilisation tool, enabling immobilization of the YdiV:FlhDC complex for testing the interaction with individual nucleotides.

The fusion proteins YdiV-His_6 and $(FLAG)_3$-FlhDC (from strains *E. coli* BL21 λ(DE3)/pGS2354 and *E. coli* BL21 λ(DE3)/pGS2382) were produced by overproduction methods (Section 2.6.1), resuspending the harvested cell pellets in a breakage buffer (Table 2.11). The cells were lysed by sonication and the phases separated by centrifugation (Section 2.6.2). The soluble fraction was clarified using a 0.2 µm filter and retained for use, at ~ 1 mg/ml (Section 2.6.8).

Table 2.11: Nucleotide:protein pull-down assay buffers

Breakage buffer	20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, pH 8
Wash buffer	20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 8
Elution buffer	20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 8

Prior to use, the Ni-NTA resin was washed in a 1:1 ratio with dH_2O and wash buffer, with two washes of each (Table 2.11). Both solutions were added and mixed with the resin and then centrifuged at 100 x*g* for 1 min. The supernatant was then removed and the pre-washed resin aliquoted into reaction tubes.

Each reaction sample contained pre-washed resin (100 µl) and an equal volume of crude soluble YdiV-His_6. The mixture was agitated for 30 min at room temperature before the sample was separated by centrifugation and unbound protein removed. The resin was washed three times before the addition of crude soluble $(FLAG)_3$-FlhDC (100 µl) and incubation and agitation for 10 min at room temperature. Following centrifugation, the supernantant was removed, and three wash steps carried out. Finally a specified nucleotide (30 µl of 50 mM) was added to the samples, again in breakage buffer, incubating the samples for 10 min at room temperature. Each nucleotide was added individually, thereby having one reaction sample per nucleotide tested. The eppendorfs were centrifuged again in order to remove the supernatant, before two further wash steps occurred. Finally elution buffer (50 µl) was added to the resin, the samples agitated and centrifuged before the final soluble sample was collected (Table 2.11).

Centrifugation steps always involved 1 min of centrifugation at 100 x*g*, with wash steps always involving the addition of 1 ml of wash buffer, inverting the eppendorfs to fully mix the sample, followed by centrifugation.

Supernatant fractions were taken at each stage of the pull-down process for analysis by electrophoresis (Section 2.6.9).

2.8 Protein:protein interactions

2.8.1 Pull-down assays

Pull-down assays were used to study the possible interaction between YdiV and FlhDC. Following overproduction of the YdiV-His_6 and $(FLAG)_3$-FlhDC proteins (from strains *E. coli* BL21 λ(DE3)/pGS2354 and *E. coli* BL21 λ(DE3)/pGS2382), the cell pellets were resuspended in breakage buffer and lysed by sonication (Sections 2.6.1 and 2.6.2). The cell phases were separated by centrifugation and the soluble extracts filtered using a 0.2 µm filter, retaining these extracts for use.

A 1 ml Hi-Trap column (GE healthcare) was manually loaded with dH_2O (10 ml) before charging the column with 100 mM $NiSO_4$ (10 ml). The column was then loaded once more with dH_2O (10 ml) prior to column equilibration with the addition of wash buffer (10 ml). Soluble YdiV extract was added to the column (7 ml at ~ 1 mg/ml) and the column further washed with the addition of wash buffer (10 ml). The soluble cell-free extract FlhDC was added to the column (10 ml at ~ 1 mg/ml) and again the column washed with wash buffer (10 ml). Finally an imidazole rich elution buffer (10 ml) was added to the column, to elute all His-tagged proteins. To clean the Hi-Trap column, dH_2O (10 ml) and 25% ethanol (10 ml) was passed through the column, prior to storage of the column at 4^{o}C. The buffers utilized in this assay are detailed in Table 2.12.

Table 2.12: Protein:protein pull-down assay buffers

Breakage buffer	20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, pH 7.5
Wash buffer	20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.5
Elution buffer	20 mM Sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.5

Fractions were collected at each stage of the pull-down assay, for subsequent analysis by SDS-PAGE (Section 2.6.9) or Western blot (Section 2.6.10).

2.8.2 Competition assay pull-down

Pull-down methods were utilized for competition assays, producing soluble extracts of YdiV-His_6, FlhDC-His_6, $(FLAG)_3$-FlhDC, GST-FliT and GST-FliZ proteins, all of which were encoded by plasmids (Table 2.2) which were used to transform *E. coli* BL21 λ(DE3) cells (Section 2.2.7) for protein overproduction. The relevant proteins were overproduced and the cells harvested (Section 2.6.1). The cell pellets were resuspended in breakage buffer, prior to sonication, centrifugation and clarification of the soluble extracts (Section 2.6.2).

A 1 ml Hi-Trap column (GE healthcare) was washed with dH_2O (10 ml) before charging it with 100 mM $NiSO_4$ (5 ml) and then washing the column with dH_2O (10 ml) and wash buffer (10 ml).

For the control assays, FlhDC was immobilised on the column, by the addition of cell-free extract FlhDC-His_6 (5 ml of 1.0 mg/ml) following column equilibration. The column was thoroughly washed with wash buffer (10 ml) before addition of GST-FliT or GST-FliZ as cell-free extracts (5 ml of 1.5 mg/ml). Unbound proteins were eluted from the column by the addition of wash buffer (10 ml) before the elution of His-tagged proteins and any binding partners by the addition of elution buffer (10 ml).

The competition assays were a slight variation on the control assays, with three proteins (all as cell-free extracts) passaged down the column rather than two. The column was charged and equilibrated as before. Soluble crude YdiV-His_6 (5 ml at 1.0 mg/ml) was added to the column first, before wash buffer (10 ml) was added. The second protein loaded onto the column was FlhDC-$(FLAG)_3$ (5 ml at 1.0 mg/ml) before another wash buffer step (10 ml). Finally GST-FliT or GST-FliZ (5 ml at 1.5 mg/ml) was added, before a final wash step (10 ml) and elution step using elution buffer (5 ml). All proteins were added as cell-free extracts at concentrations of ~1.0- 1.5 mg/ml (Section 2.6.8).

To clean the Hi-Trap column, dH_2O (10 ml) and 25% ethanol (10 ml) was passed through the column, prior to storage of the column at 4°C. The buffers utilized in this assay are listed in Table 2.12.

Fractions (0.5 ml) were collected at each stage of the pull-down assay, for subsequent analysis by SDS-PAGE (Section 2.6.9) or Western blot (Section 2.6.10).

2.8.3 Bio-Layer interferometry for YdiV: FlhDC interaction

The BLItz detection system (fortéBIO) was utilised to determine the binding kinetics of the YdiV and FlhDC interaction. This system used amine reactive (AR2G) biosensors (fortéBIO) to covalently immobilize one of the proteins to the probe surface.

The probes were hydrated in buffer for 10 min prior to use, before the probes were loaded onto the BLItz machine and the reaction programme selected (Table 2.13).

Table 2.13: BLItz programme details for AR2G biosensor protocols

Step Number	Step type	Duration (s)	Position
1	Initial baseline	60	Tube
2	Chemical activation	300	Tube
3	Load sample	600	Tube
4	Quenching	300	Tube
5	Baseline	120	Tube
6	Association	300	Drop
7	Dissociation	600	Tube

These experiments consisted of seven distinct steps, with the biosensor physically moved to the appropriate solution vessel at each stage. Tube steps required 250 µl of solution to be present in a black eppendorf, whilst drop stages required 5 µl of a solution to be present in a concave drop. The concave surface was thoroughly washed between samples, to ensure no contamination.

Both the YdiV and FlhDC proteins were purified with C-terminal His-tags (Section 2.6.3), overproduced using plasmids pGS2354 and pGS2382 respectively (Table 2.2) in *E. coli* BL21 λ(DE3) cells and dialysed into sodium phosphate buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.5) (Section 2.6.7). This dialysis buffer was used throughout the experiments for baseline and dissociation stages (steps 1, 5 and 7). The AR2G probes were chemically

activated (step 2) by addition of the 'amine coupling second generation reagent kits' following the manufacturer's guidelines. Protein immobilisation (step 3) was subsequently achieved by a 1/10 dilution of concentrated protein (1-3 mg/ml stock) into 10 mM sodium acetate pH 5. Unbound protein was removed by addition of 1 M ethanolamine pH 8.5 (step 4) which quenched the reaction. Following step 5, one of the proteins was immobilised to the AR2G probe, with excess chemicals and proteins removed. Detection of the binding interaction was achieved by addition of the partner protein in solution at concentrations varying from 9 nM to 10 µM (step 6). Following experiment completion, the biosensor was removed from the BLItz machine, and discarded.

The concentration of the immobilised protein was fixed throughout the series of experiments, varying the concentration of soluble protein added in step 6 (Section 2.6.8). This produced a series of data which could be analysed using the BLItz Pro software (fortéBIO) in order to determine reaction constants for each experiment.

2.8.4 Isothermal titration calorimetry

The binding interaction YdiV:FlhDC was studied in solution, in an ITC detector (TA instruments). Both proteins were purified (Section 2.6.3) as His-tagged fusion proteins (using plasmids pGS2382 and pGS2354) and were dialysed into sodium phosphate dialysis buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.5) (Section 2.6.7) and degassed before use.

System cleaning and equilibration was carried out following manufacturer's instructions. The FlhDC protein complex (44 µM) (Section 2.6.8) was loaded into the sample chamber and YdiV (235 µM) loaded into the syringe. The syringe was loaded into the chamber, and the system equilibrated to 25^{o}C with 250 rpm agitation in the chamber. Once the system was fully equilibrated and the heat changes in the chamber minimal, the titrations were initiated. YdiV was injected into the reaction chamber from the loaded syringe, with 25 injections taking place every 180 s, each of volume 2 µl.

Analysis of the detected heat changes was carried out using the NanoAnalyse software (TA instruments). Following manufacturer's instructions, the baseline heat changes were determined and subtracted from the detected changes to produce a corrected heat change trace. The molar ratio of the chamber protein (FlhDC) and syringe protein (YdiV) were

determined by the NanoAnalyse software (TA instruments), calculating the ratio after each injection. These molar ratios could then be plotted against the corrected heat changes for the duration of the experiment. At this point, various binding models could be added to the graphs and the model optimally fitted to the data. Several binding models were tested against the data and variables altered to produce the best-fit curve. Once the binding model was determined, the thermodynamic parameters of the interaction were calculated by the software, giving interaction stoichiometry and binding constants values.

The binding interaction was repeated in triplicate, yielding multiple data sets to fit binding models to, thereby increasing confidence in the binding data. Between each analysis, the reaction chamber and syringe were cleaned following manufacturer's instructions.

2.9 Protein:DNA interactions

2.9.1 Radioactive labelling of DNA

DNA fragments were amplified by PCR reactions (Section 2.3.3) and checked for appropriate amplification on agarose gels (Section 2.3.5). Following this, DNA was purified by a Qiagen PCR purification kit (Section 2.3.10), and then was digested with appropriate restriction enzyme (XbaI) (Section 2.3.6) before repeating the Qiagen PCR purification step (Section 2.3.10). The DNA was then quantified to find the appropriate amount to radioactively label (Section 2.3.8).

A 30 µl reaction mixture was then made, including a suitable volume of DNA to give 150 ng DNA, 3 µl Klenow buffer, 0.5 µl Klenow (5 units) (Fermentas) and 0.5 µl [α-^{32}P] - dCTP (10 mCi/ml) (Perkin Elmer), making up the volume with dH_2O. The reaction was then left for 1 h on the radiation bench before a final PCR purification was carried out (Section 2.3.10) and the labelled DNA solution was frozen.

2.9.2 Electromobility shift assays (EMSA)

EMSA gels were made, using Bio-Rad plates and tanks, as described for SDS-PAGE gels. The gels were made using the components listed in Table 2.14, and the gels clamped into the Bio-Rad tanks before adding 0.5 x TBE buffer to the tank (Table 2.14).

Table 2.14: EMSA gel and buffer components

10 x TBE buffer (1 litre)	108 g Tris, 55 g boric acid, 40 ml EDTA (0.5 M, pH 8)
Gel shift gels	5.63 ml dH_2O, 1.5 ml 30% acrylamide (37:5:1), 0.38 ml 10 x TBE, 75 µl APS, 7.5 µl TEMED
EMSA binding buffer (10 x)	400 mM Tris pH 8, 10 mM EDTA, 10 mM $MgCl_2$, 5% glycerol
EMSA binding buffer (1 x) (20µl)	17.5 µl 10 x EMSA binding buffer, 2 µl 100 mM DTT, 0.5 µl 100 mg/ml BSA
EMSA loading dye	50 % glycerol, bromophenol blue

After removing the comb from the gel, each well was scraped using a needle to ensure all the wells were clean and free of polymerized acrylamide, before pre-running the gel.

Samples were prepared, containing 1 µl of 1 x EMSA binding buffer, 1 µl of radio-labelled DNA (at 5 ng/µl), 1 µl of 2 mg/ml Calf Thymus DNA, 2 µl of 500 mM NaCl, making up to 10 µl with dH_2O (Table 2.14). In samples containing protein, the volume of 500 mM NaCl was reduced to compensate for protein addition (in 500 mM NaCl), ensuring a final volume of 10 µl in all cases. All samples had a final salt concentration of 100 mM. The samples were prepared and mixed, before being transferred to a radiation area where the radiolabelled DNA was added (Section 2.9.1). From here the samples were incubated for 30 min at 37^oC, before 2 µl of EMSA loading dye was added to each sample, mixed and 10 µl loaded to the gel (Table 2.14).

In cases where antibodies were added (Section 2.6.11), the antibodies were diluted 1/10 into salt buffer, and 1 µl added after the incubation period just prior to EMSA loading dye. For nucleotide addition (Section 2.7), nucleotides were diluted into dH_2O and added prior to DNA addition and incubation, to give a final concentration of 100 µM in the samples.

The EMSA gel was pre-run at 30 mA per gel for 30 min using a power pac 300 (Bio-Rad) before it was slowed to 10 mA per gel, whilst loading of samples occurred. Once loaded, the gel was run at 30 mA per gel for approximately 25 min, or until the dye front had run off the gel.

Once complete, the tank was drained and the gel was dried and visualized by autoradiography (Section 2.9.4).

2.9.3 *In vitro* transcription (IVT) assays

Large sequencing gel plates were cleaned with 70% ethanol, before the "ears" plate was prepared further by rubbing in 500 µl Gel Slick Solution (Lonza) to ease gel separation later. The plates were assembled with spacers between the back and ear plates, before being sealed tightly with tape.

The gel mix was then made and added using a syringe, ensuring no air bubbles formed (Table 2.15). A comb was then added and the gel was left to set.

Table 2.15: List of constituents in IVT gels, dyes, markers and buffers

IVT gel	60 ml of 6% sequencing gel (19:1 acrylamide, 1 x TBE, 7 M urea), 600 µl 10% APS, 60 µl TEMED
10 x TBE buffer (1 litre)	108 g Tris, 55 g boric acid, 40 ml EDTA (0.5 M, pH 8)
Perfect RNA Marker (0.1-1 Kb)	0.75 µg RNA template mix, 80 mM HEPES pH 7.5, 12 mM $MgCl_2$, 10 mM NaCl, 10 mM DTT, 2 mM ATP, 2 mM GTP, 2 mM CTP, 0.1 mM UTP, 50 U T7 RNAP (Novagen), 20 units RiboLock RNase inhibitor (Fermentas), 5 µCi [α-^{32}P]- UTP (800 Ci mmol^{-1}) (Perkin Elmer)
Stop/Loading dye	95% formamide, 20 mM EDTA, pH 8, 0.05% bromophenol blue, 0.05% xylene cyanol

The prepared IVT gel was clamped into position into a Flowgen gel running tank, and 1 x TBE buffer added at the top and bottom of the gel (Table 2.15). Urea and acrylamide were removed from the gel wells using a syringe, prior to pre-running the gel at 55 W for approximately 1 h using a Pharmica power supply. During this time the samples were prepared to run on the gel. Once ready, the samples were loaded into the wells, ensuring a lane on either side of the samples contained just stop/loading dye and the gel was run at 50 W for approximately 2 h, or until the first tracking dye had run off the gel. Once complete, buffer was drained from the tank, and the gel was dried and visualized by autoradiography (Section 2.9.4).

Reactions for IVT analysis (10.5 µl) contained 1 pmole *E. coli* RNA polymerase, 0.1 pmole linear DNA, 40 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 1 mM DTT, 75 mM KCl, 0.1 mM EDTA, 5% glycerol, 250 µg/ml bovine serum albumin (BSA) and relevant concentrations of the chosen

transcriptional regulators. In cases of ligand addition, the nucleotides (Section 2.7) were added to give a final concentration of 1 mM at the same point as the protein regulators.

These reactions were incubated at 37^{o}C for 15 min before transcription was initiated by the addition of 2 µl of NTP mix to the tubes. The NTP mix contained 50 µM UTP, 1 mM of each ATP, CTP and GTP and 1 µCi [α-^{32}P]- UTP (Perkin Elmer). Transcription was terminated after a 15 min 37^{o}C incubation by the addition of 12.5 µl stop/loading dye, before the samples and marker were analysed on an IVT gel (Table 2.15).

A 0.1-1 Kb marker was made using Perfect RNA Marker Template Mix (Novagen). The reaction mixture was made and was incubated for 1 h at 37^{o}C, before the reaction was terminated by a 1:20 addition of Stop/Loading dye (Table 2.15). The marker was then stored at -20^{o}C before being used at ~2 ng for gel calibration.

2.9.4 Autoradiography

Following electrophoresis, EMSA and IVT gels were transferred onto Whatman 3MM filter paper and the gels were dried in a Scie-Plas vacuum drier at 80^{o}C. Once dry, the gels were exposed to Kodak medical X-ray film at -70^{o}C for a required period of time. Once complete, the films were developed in Kodak Industrex developer and then Kodak Industrex fixer before being washed in water and air-dried.

2.9.5 DNA:protein kinetics

The BLItz (fortéBIO) system was utilised to measure the kinetics of DNA: protein interactions. This system utilized streptavidin (SA) biosensors (fortéBIO), enabling immobilization of biotinylated DNA.

Double-stranded DNA was produced for the region of interest (P*fliD*) and a control region, using primer pairs NSW84 and 28 and NSW85 and 86 respectively, with both forward primers containing a 5' Biotin tag (Table 2.3). Following amplification of these regions (Section 2.3.3) the DNA was purified by gel extraction from an agarose gel (Section 2.3.7) and concentrated in a spin-vac to give concentrated samples. For all BLItz assays, both the target and control DNA were used at a concentration of 35 µg/ml (Section 2.3.8).

The biosensors were first hydrated in buffer for 10 min before being loaded onto the BLItz machine. These experiments consisted of 5 distinct steps (Table 2.16).

Table 2.16: BLItz programme details for SA biosensor protocols

Step number	Step type	Duration (s)	Biosensor position
1	Initial baseline	60	Tube
2	Loading	300	Drop
3	Baseline	60	Tube
4	Association	300	Drop
5	Dissociation	600	Tube

All stages in which the biosensor was submerged into a tube required 250 µl of solution to be present in the tube, whilst drop stages required 5 µl of a solution to be present in a concave drop. The concave surface was thoroughly washed between samples, to ensure no contamination.

Both the YdiV and FlhDC proteins were purified in their His-tagged fusion forms (produced from plasmids pGS2354 and pGS2441 in *E. coli* BL21 λ(DE3) cells) (Section 2.6.3) and dialysed into sodium phosphate buffer (20 mM sodium phosphate, 500 mM NaCl, pH 7.5), the same buffer used throughout the experiments for baseline and dissociation stages (steps 1,3 and 5). DNA was immobilized in step 2, with proteins added in the association step 4. For each interaction, the experiments were repeated several times, with a fixed DNA concentration and a range of protein concentrations (between 5000 nM and 10 nM) (Section 2.6.8).

Following completion of an experiment, the biosensor was removed from the BLItz machine, and discarded. Analysis of the experiment was carried out using the BLItz Pro software (fortéBIO), fitting a curve to the detected biosensor thickness changes. Using this fitted line, the reaction constants could be determined as well as the level of error and average values across several experiments.

2. 10 Analysis software

2.10.1 Image J

Image J software (http://imagej.nih.gov/ij/) was used to quantify the intensity of proteins or DNA from qualitative assays such as SDS-PAGE gels (Section 2.6.9) and Western blot analysis (Section 2.6.10). This software was used following manufacturer's instructions.

2.10.2 DNA and protein alignments

Promoter alignment was carried out for the *ydiV* promoter region in a range of *E. coli* and *S. enterica* strains. The DNA sequences were retrieved from xBASE (http://www.xbase.ac.uk/) and aligned computationally using the Mega6 software (http://www.megasoftware.net) and the ClustalW parameters. The promoter motifs were detected using Softberry promoter prediction software (www.softberry.com) which were subsequently transferred manually to the *ydiV* alignment.

Protein alignments were carried out to highlight the degeneracy of the YeaI and YdiV proteins alongside known active DGC's and PDE's respectively. The protein sequences were retrieved from UniProt (http://www.uniprot.org/) and aligned computationally using the Mega6 software and ClustalW parameters (http://www.megasoftware.net).

2.10.3 Phylogenetic analysis

Phylogenetic analysis of the *ydiV* promoter regions was carried out by Mega6 software (http://www.megasoftware.net) which produced phylogenetic trees of the strains based purely on the degree of nucleotide similarity.

Chapter 3: Characterisation of the 'degenerate' *Escherichia coli* diguanylate cyclase protein YeaI

3.1 Introduction

The diguanylate cyclase (DGC) family of bacterial proteins are involved in cyclic di-GMP signalling by synthesising the secondary messenger molecule c-di-GMP (Simm *et al.,* 2004). These DGCs have a conserved active site motif, composed of the residues GG[D/E]EF and an sporadic inhibitory (I) site for c-di-GMP feedback. The conserved GG[D/E]EF motif is necessary for the catalytic activity of DGCs to convert GTP molecules into c-di-GMP (Simm *et al.,* 2004; Chan *et al.,* 2004).

YeaI is a two domain protein found in *E. coli*, composed of an N-terminal membrane-bound domain and a C-terminal cytosolic GGDEF domain. The N-terminal of the protein consists of multiple predicted transmembrane regions, of unknown function (Sommerfeldt *et al.,* 2009; Daley *et al.,* 2005). The C-terminal domain is a member of the GGDEF family, having sequence similarities to the family, and a conserved I-site, but lacking the consensus GGDEF motif, with an EGEVF sequence found instead (Sommerfeldt *et al.,* 2009; Figure 1.6). There have been conflicting views on YeaI and its role as a DGC. Whilst Sommerfeldt *et al.* (2009) hypothesized that the protein was catalytically inactive, Sanchez-Torres *et al.* (2011) found evidence that YeaI was an active DGC despite its altered active site motif. Whilst functional roles of the protein have been determined, deleting the *yeaI* gene resulted in no change in the intracellular c-di-GMP levels. However this has been interpreted to be insignificant, with any changes in c-di-GMP only acting at the YeaI protein locality (Sanchez-Torres *et al.,* 2011).

The phenotypic roles of YeaI have been characterised, and show that the protein is involved in regulating both motility and biofilm lifestyles of bacteria (Sanchez-Torres *et al.,* 2011). Analysis of a *yeaI* deletion strain revealed a 4-fold increase in cell motility and a 30-fold increase in biofilm formation. This indicated a phenotypic role of YeaI in repressing both biofilm formation and motility to some extent, having functional roles in the cell despite possessing an EGEVF rather than GGDEF motif (Sanchez-Torres *et al.,* 2011). Active

DGC proteins synthesize c-di-GMP in the cell, high levels of which characteristically cause biofilm formation and repress motility (Simm *et al.,* 2004). The phenotypic analysis of YeaI therefore contradicts to some extent the characteristic roles of an active DGC, with YeaI repressing rather than inducing biofilm synthesis (Sanchez-Torres *et al.,* 2011).
Work in this chapter has determined the following:

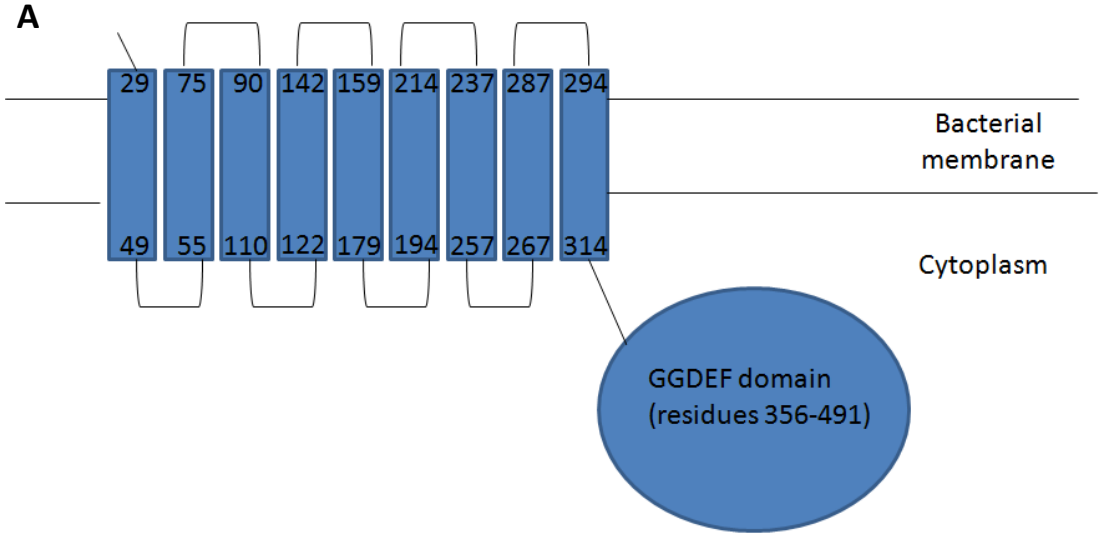
- The YeaI GGDEF domain is an inactive DGC protein (Section 3.6)
- The YeaI GGDEF domain has no clear nucleotide binding capabilities (Section 3. 7)
- There is support to suggest that the YeaI GGDEF domain has, like active DGCs, retained the ability to form dimers (Section 3.8)
- Overproduction of the protein is especially poor, producing low levels and poor purity samples (Section 3.3)
- Cloning DNA encoding the YeaI GGDEF domain into a pCOLD TF vector has largely overcome the overproduction problem, however purification still remains problematic and the main priority for future work (Section 3. 9)

3.2 Production of YeaI GGDEF domain overproduction plasmids

As previously stated the YeaI protein consists of an N-terminal membrane bound domain and a cytosolic GGDEF domain (Figure 3.1) (Daley *et al.,* 2005). The N-terminal region of the protein consists of several transmembrane helices, with linker regions separating these helices. Nine possible transmembrane regions have been identified computationally by UniProtKB (www.uniprot.org) using various prediction tools, with each transmembrane helix of ~20 amino acid residues in length (Figure 3.1). The precise location of the cytosolic GGDEF domain was determined by the domain annotations found again at UniProtKB, identifying residues 356-491 of YeaI to be homologous to other proteins belonging to the DGC family. The secondary structure of this DGC domain in YeaI is unknown, however X-ray crystallography of other proteins in the DGC family have shown a common structure between DGC proteins, consisting of a 5 β-strand core surrounded by 5 or 6 α-helices (De *et al.,* 2009; Navarro *et al.,* 2009).

Membrane bound proteins are notoriously difficult to study and purification of such proteins is very complex. It was therefore decided to study only the C-terminal GGDEF domain of YeaI (referred to from here on in as $YeaI_{GGDEF}$), anticipating that this would simplify purification and still provide useful information on the function of the protein. The

A



B

```
PopA    ETAMRVMELA RSFRRGESIR GALEKARSSG LMDAATGLFT RDLFAAHLAR LASAARERSR
YeaI    IEVSSKLFVV SFLIYN--IF QELQLSSKLA VHDVLTNIYN RRYFFNSVES LLSRPVVKD-
PleD    ELSARVKTQI QRKRYTDYLR NNLDHSLELA VTDQLTGLHN RRYMTGQLDS LVKRATLGGD

PopA    PLSICVLRVA DKPETVWARQ NGWLDRAIPQ IGSMVGRLVR VEDTPARLAT EVFALALPAT
YeaI    -FCVMLVDIN QFKRINAQWG HRVGDKVLVS IVDIIQQSIR PDDILARLEG EVFGLLFTEL
PleD    PVSALLIDID FFKKINDTFG HDIGDEVLRE FALRLASNVR AIDLPCRYGG EEFVVIMPDT

PopA    NQNAACAAAE RIAAVIGCTA FDAGEDRAPF VCEFDIGVAE VQ-PGEGAVK ALERAAAAAL
YeaI    NSAQAKIIAE RMRKNVELLT GFSNRYDVPE QMTISIGTVF STGDTRNISL VMTEADKALR
PleD    ALADALRIAE RIRMHVSGSP FTVAHGREML NVTISIGVSA TAGEGDTPEA LLKRADEGVY

PopA    KREAG----- -----
YeaI    EAKSEGGNKV IIHHI
PleD    QAKASGRNAV VGKAA
```

Figure 3.1: Structure and degeneracy of the YeaI protein

A) Schematic of YeaI protein, indicating two clear domains; an N-terminal transmembrane (TM) region and a C-terminal cytoplasmic domain. Computational prediction tools have identified nine possible TM helices (UniProtKB), the identity of which is indicated by the amino acid residue numbers at either end of the helices. The cytoplasmic C-terminal domain of the protein shows amino acid residue homology to the DGC family of proteins.

B) Alignment of GGDEF domains of YeaI (from *E. coli*) alongside the active DGC protein PleD (from *C. crescentus*) and the degenerate GGDEF domain protein PopA (from *C. crescentus*) (Section 1.8.1). The sequences were retrieved from Uniprot and the alignment carried out by Mega6 software (Section 2.10.2). The final residues of the N-terminal domains of the proteins are shown in grey, to highlight the start of the GGDEF domains of each protein. The GGDEF active site region is boxed in red, highlighting the degenerate nature of YeaI. The I-site of the proteins is boxed in blue, showing the lack of an I-site for PleD, but the presence of an I-site in both the PopA and YeaI proteins.

DNA corresponding to YeaI residues 356-491 was therefore used for production of $YeaI_{GGDEF}$ overproduction plasmids.

A pET32a:$yeaI_{GGDEF}$ plasmid (pGS2263) had previously been made, producing an N-terminal His_6-$YeaI_{GGDEF}$ fusion protein. To investigate whether the position of the His_6 tag affected the solubility of the protein, the $yeaI_{GGDEF}$ gene was also cloned into a pET21a vector, to give a C-terminal tagged fusion protein rather than an N-terminal tag (Figure 3.2). The $yeaI_{GGDEF}$ gene was amplified from genomic *E. coli* MG1655 DNA by PCR using primers NSW11 and NSW12, giving NdeI and XhoI restriction sites at either end of the PCR product (Table 3.1) (Section 2.3.3).

Table 3.1: Primers used in amplification of the $yeaI_{GGDEF}$ gene for insertion into the plasmid pET21a

Primer	Sequence	Function
NSW11	TTTTCATATGAAGGACTTCTGTGTCATG CTG	Forward primer for $yeaI_{GGDEF}$ into pET21a vector
NSW12	TTTTCTCGAGAATATGATGAAGAATCAC TTTGTT	Reverse primer for $yeaI_{GGDEF}$ into pET21a vector

The grey shaded sequence indicates an NdeI restriction site and the boxed sequence shows an XhoI restriction site. No stop codon was encoded by the reverse primer, in order to allow transcription to continue into the vector and transcribe the C-terminal His_6 tag, encoding a fusion protein.

The purified PCR product and pET21a vector (Novagen) were digested with NdeI and XhoI enzymes, prior to ligation (Sections 2.3.4; 2.3.6; 2.3.9 and 2.3.10). The ligated DNA was then used to transform electrically competent *E. coli* DH5α cells and transformants were selected for on ampicillin supplemented agar (Sections 2.2.4 and 2.2.7). Following colony PCR screening, DNA sequencing confirmed the identity of the insert, using T7 promoter and T7 terminator primers (Table 3.2) (Section 2.3.11).

Table 3.2: Sequencing primers used for the plasmid pET21a

Primer	Sequence	Function
T7 promoter primer	TAATACGACTCACTATAGGG	Forward primer for pET21a sequencing
T7 terminator primer	GCTAGTTATTGCTCAGCGG	Reverse primer for pET21a sequencing

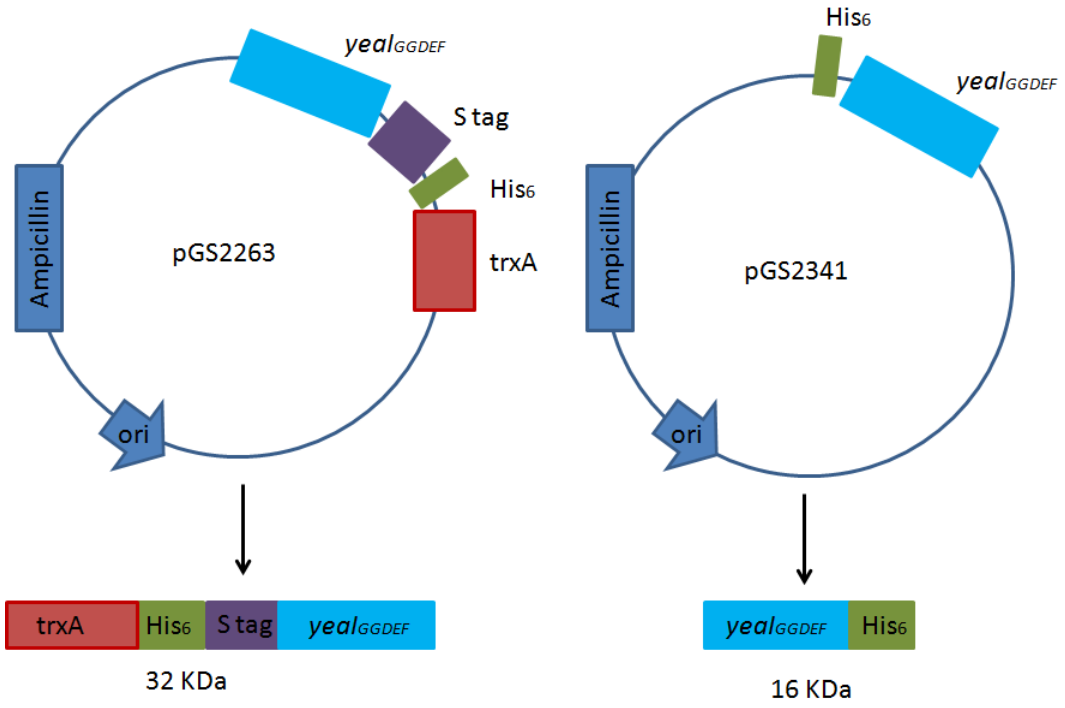


Figure 3.2: Schematic plasmid maps for the insertion of *yeal*$_{GGDEF}$ into pET32a and pET21a

Maps for the pGS2263 and pGS2341 plasmids, derived from pET32a and pET21a vectors respectively. The positions of the various tags, antibiotic resistance cassette and *yeal*$_{GGDEF}$ insertion site are indicated, as well as the fusion proteins produced in each case.

In both cases, the origin of replication is indicated, labelled as ori. The differing sizes of the fusion proteins are accounted for by the various tags used, with His$_6$ tag (0.8 KDa), S tag (1.7 KDa) and TrxA (11.8 KDa) fused to the *yeal*$_{GGDEF}$ (15.3 KDa). Alongside the defined tags, linker regions exist between the tags and insert and thereby contribute to the fusion protein molecular weight.

The pGS2263 construct encodes an N-terminal TrxA, His$_6$ and S tag, each of which is linked together by up to 11 amino acid residues. This entire N-terminal region contributes 16.8 KDa, giving a total molecular weight of 32.2 KDa.

The pGS2341 construct gives a C-terminal His$_6$ fusion protein, with 2 amino acid residues linking *yeal*$_{GGDEF}$ and the His$_6$ tag, giving a total molecular weight of 16.4 KDa.

Plasmids containing no mutations in the *yeaI*$_{GGDEF}$ sequence were named pGS2341 (pET21a::*yeaI*$_{GGDEF}$).

3.3 Optimisation of YeaI$_{GGDEF}$ overproduction

Both plasmids pGS2341 (pET21a::*yeaI*$_{GGDEF}$) and pGS2263 (pET32a:*yeaI*$_{GGDEF}$) were used to transform electrically competent *E. coli* BL21 λ(DE3) cells (Section 2.2.7) for overproduction trials, altering variables such as IPTG concentration and induction time. Trials were carried out using 100 ml cultures, growing the cells at 37^{o}C before inducing the cells at lower temperatures, where the cells remained for subsequent growth. An aliquot (1 ml) of an overnight culture was added to 100 ml of LB containing ampicillin, and the cells were grown at 37^{o}C with shaking at 250 rpm until an OD_{600} of ~0.6 was achieved (Section 2.6.1). IPTG was added, to a final concentration of either 100 μg/ml or 20 μg/ml and the cultures incubated at 25^{o}C or 4^{o}C (to reduce production of inclusion bodies) with shaking. Cell samples were taken before induction (0 h), and at regular hourly intervals. The samples were prepared and analysed by SDS-PAGE (Section 2.6.9).

A clear overproduced polypeptide was present in the post-induction fractions, which was believed to correspond to YeaI$_{GGDEF}$, at ~13 KDa in the pET21a construct (migrating slightly faster than its expected molecular weight of 16 KDa) and at ~30 KDa in the pET32a construct (of expected molecular weight 32 KDa) (Figure 3.3 and 3.4). This significant size variation of YeaI$_{GGDEF}$ in the two constructs can be explained by the contrasting nature of the tags, with a <1 KDa His_6 tag fusion to YeaI$_{GGDEF}$ in the pET21a construct, but >16 KDa fused to the protein in the pET32a construct (Figure 3.2).

Analysing the pET21a construct first, the YeaI$_{GGDEF}$-His_6 protein was overproduced to a similar extent when using either IPTG concentration (Figure 3.3A and B). Induction duration appeared to have a more significant effect on overproduction, with 4 h or more producing the highest yield of YeaI$_{GGDEF}$-His_6 (Figure 3.3). The pET32a construct, in which the tag was N-terminal, again yielded high levels of overproduced His_6-YeaI$_{GGDEF}$, especially at conditions of 25^{o}C, either with 100 or 20 μg/ml IPTG for 3 or more h induction (Figure 3.4A). Whilst induction temperatures of 4^{o}C produced lower levels of overproduced protein, reasonable yields were seen using 100 μg/ml IPTG for 5 or 6 h (Figure 3.4B).

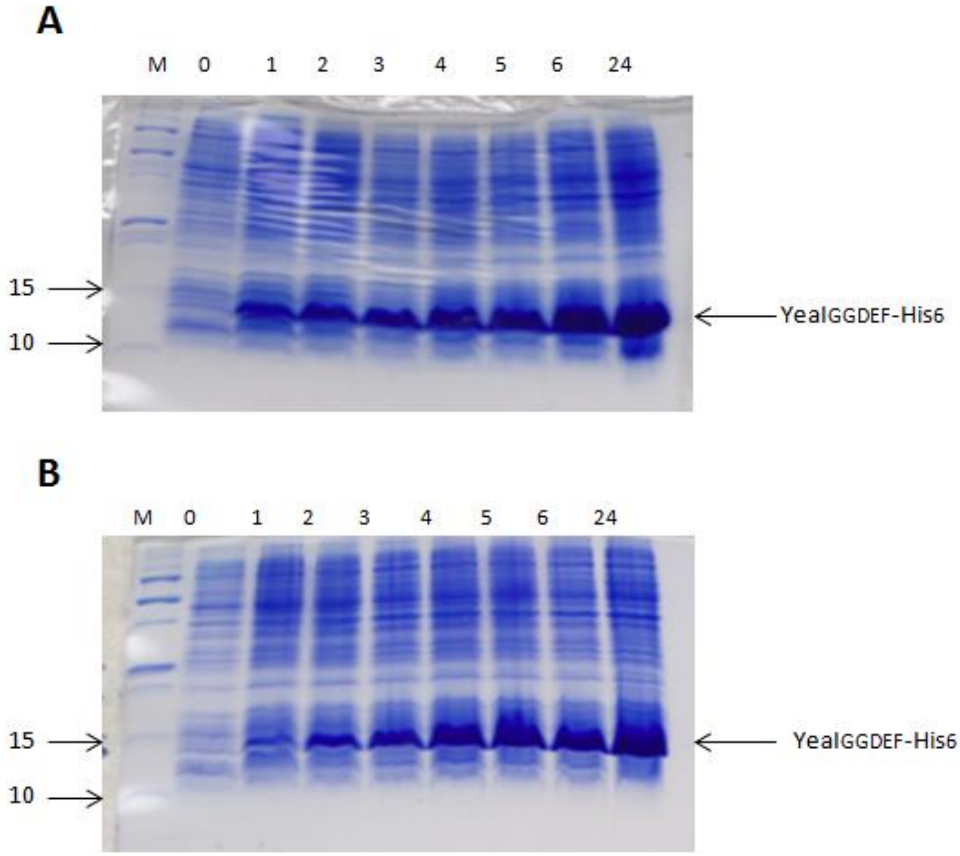


Figure 3.3: Optimisation of overproduction conditions for the Yeal$_{GGDEF}$-His$_6$ fusion protein in the *E. coli* BL21 (λDE3)/pGS2341 strain

Coomassie-stained SDS-PAGE gel of samples taken pre- and post-induction of *yeal$_{GGDEF}$* expression in the pET21a vector.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-24 indicate induction times in h, with Yeal$_{GGDEF}$ production being induced using either a final concentration of 100 µg/ml (A) or 20 µg/ml (B) IPTG, as specified.

The position of Yeal$_{GGDEF}$-His$_6$ is indicated (expected size 16 KDa).

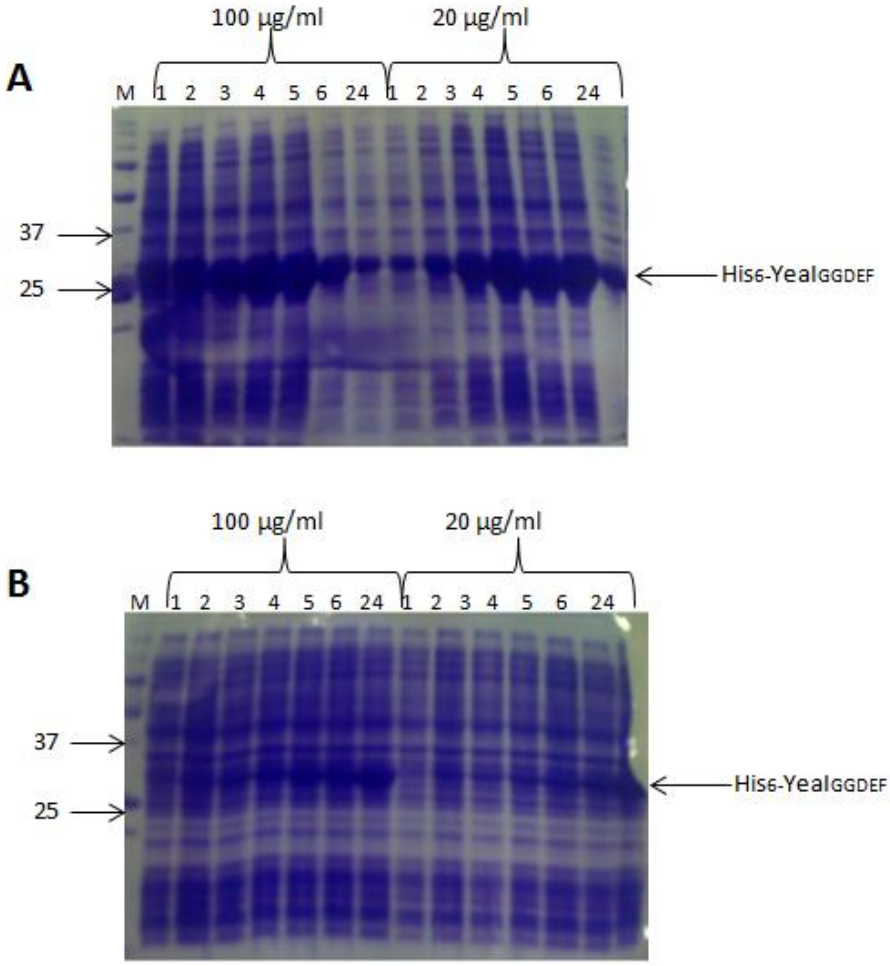


Figure 3.4: Optimisation of overproduction conditions for the His_6-$Yeal_{GGDEF}$ fusion protein in the *E. coli* BL21 (λDE3)/ pGS2263 strain

Coomassie-stained SDS-PAGE gel of samples taken pre- and post-induction of *$yeaI_{GGDEF}$* expression in the pET32a vector.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-24 indicate induction times in h, with $YeaI_{GGDEF}$ production being induced using either a final concentration of 100 µg/ml or 20 µg/ml IPTG, as specified. Induction temperatures were either 25°C (A) or 4°C (B).

The position of His_6- $YeaI_{GGDEF}$ is indicated (expected size 32 KDa), in reality indicating the larger fusion protein TrxA-His_6-S tag-$YeaI_{GGDEF}$.

For ease of differentiation between both fusion proteins, theTrxA-His_6-S tag-$YeaI_{GGDEF}$ fusion protein produced by overproduction from plasmid pGS2263 is labelled as His_6-$YeaI_{GGDEF}$ throughout this chapter, to clearly differentiate between this protein and the C-terminally tagged $YeaI_{GGDEF}$-His_6 protein.

3.4 Solubility trials of YeaI

Whilst the overproduction trials gave an indication of the optimal conditions for $YeaI_{GGDEF}$ production, conditions that would give the highest level of soluble $YeaI_{GGDEF}$ needed to be established. Solubility trials were therefore carried out for both the pET32a:$YeaI_{GGDEF}$ strain (giving the His_6-$YeaI_{GGDEF}$ protein) and the pET21a:$YeaI_{GGDEF}$ strain (producing $YeaI_{GGDEF}$-His_6 protein) at the various overproduction conditions. Cell pellets of both $YeaI_{GGDEF}$ overproduction strains (*E. coli* BL21 λ(DE3)/pGS2262 and *E. coli* BL21 λ(DE3)/pGS2341), grown under the various conditions, were prepared for solubility trials (Tables 2.1 and 2.2). The cell pellets were resuspended in breakage buffer, the cells broken by sonication, prior to clarification into distinct soluble and insoluble phases (Section 2.6.2). The insoluble fractions were resuspended in breakage buffer, and both components analysed by SDS-PAGE (Section 2.6.9). This showed that the proteins were mostly insoluble, with generally >90% of the protein existing in the insoluble fraction (Figures 3.5 and 3.6). Small but potentially useable levels of $YeaI_{GGDEF}$ were however found in the soluble fractions in some conditions, for example, when using the pET21a construct with 2 or 3 h induction with 20 µg/ml IPTG (Figure 3.5B). Equally although faint, the same induction conditions were optimal for the solubility of $YeaI_{GGDEF}$ in the pET32a construct, with cell growth at either 25 or 4^{o}C (Figure 3.6B and 3.6D).

It was concluded that the optimal conditions for maximum protein solubility were obtained using the pET32a construct (His_6-*$YeaI_{GGDEF}$*) with induction at 25^{o}C for 3 h by 20 µg/ml IPTG. These conditions were then used for all future culturing of the His_6-$YeaI_{GGDEF}$ strain.

Given the low solubility levels of $YeaI_{GGDEF}$ even after optimisation, attempts to solubilise the insoluble pellet were carried out. The insoluble cell pellet resulting from the overproduction protocol was resuspended in 6 M guanidine hydrochloride and refolding was attempted on a 1 ml HiTrap chelating column using the refolding programme on the

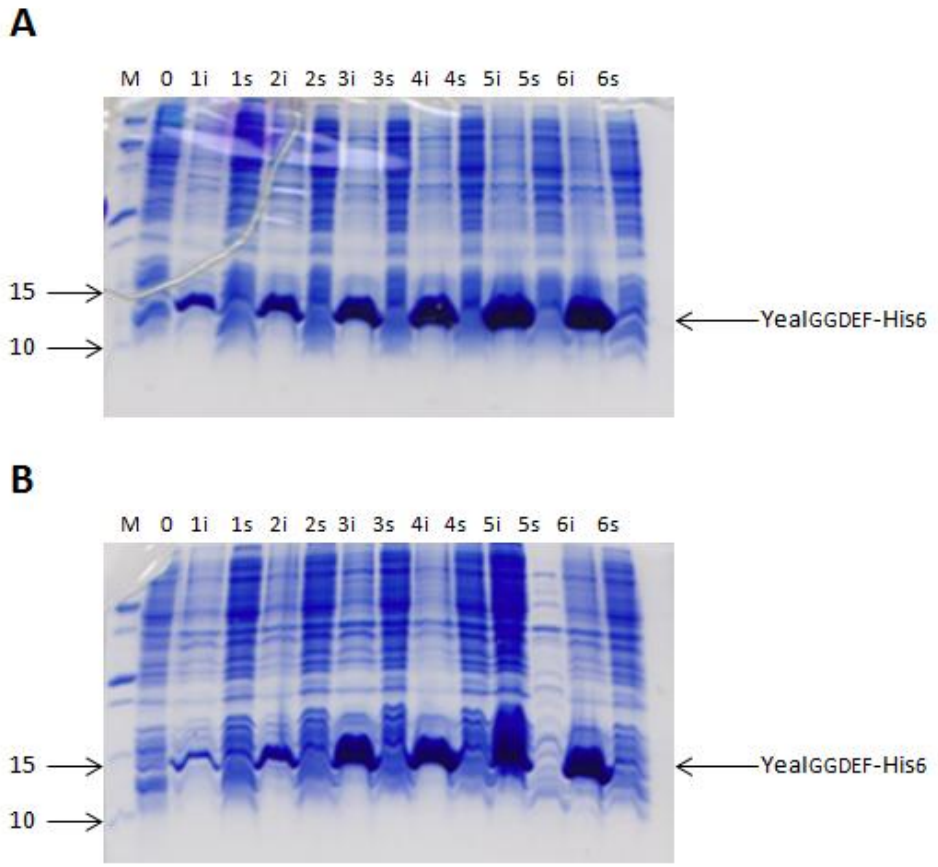


Figure 3.5: Optimisation of solubility conditions for the $Yeal_{GGDEF}$-His_6 fusion protein in the *E. coli* BL21 (λDE3)/pGS2341 strain

Coomassie-stained SDS-PAGE gel of insoluble and soluble cell samples, analysing $Yeal_{GGDEF}$ solubility in the pET21a vector.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-6 indicate induction times in h, with "i" indicative of insoluble fractions and "s" indicative of soluble fractions. $Yeal_{GGDEF}$ solubility was analysed for overproduction using either a final concentration of 100 µg/ml (A) or 20 µg/ml (B) IPTG.

The position of $Yeal_{GGDEF}$-His_6 is indicated (expected size 16 KDa).

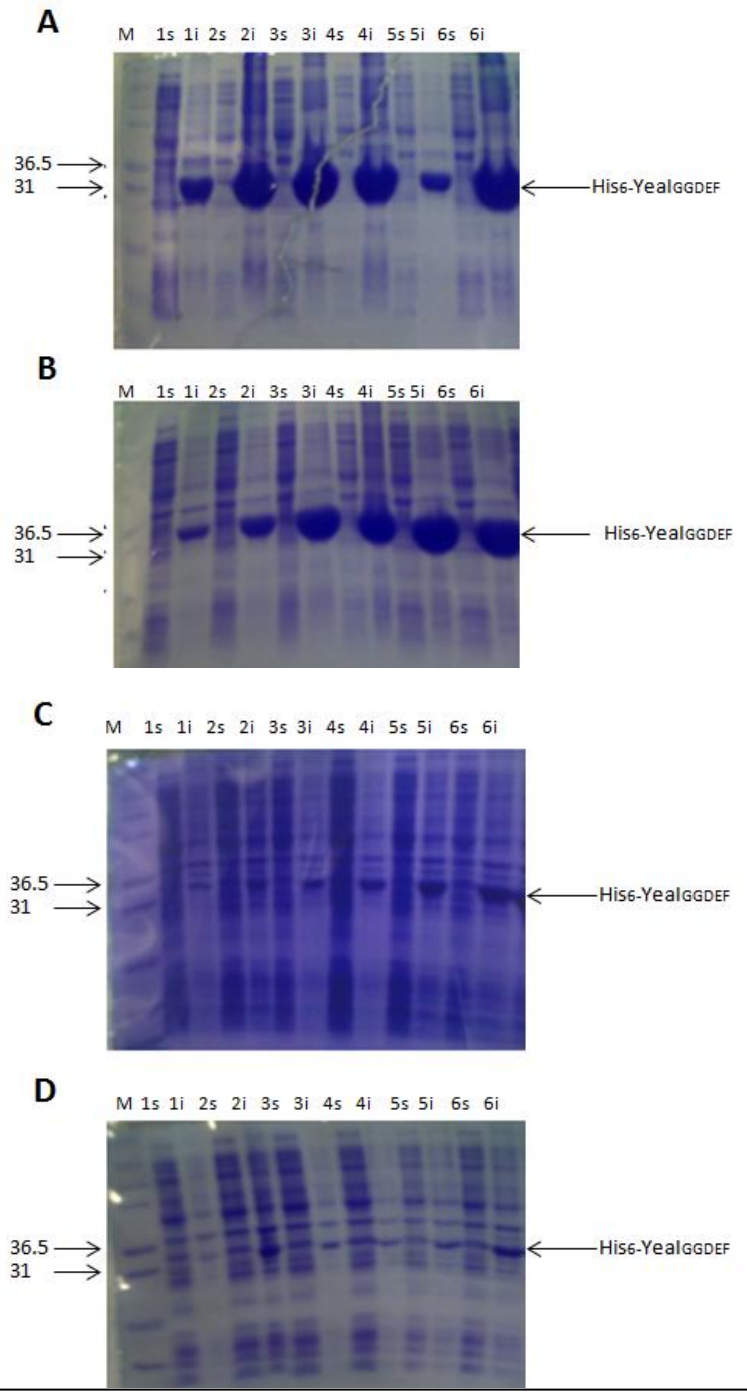


Figure 3.6: Optimisation of solubility conditions for the His_6-$Yeal_{GGDEF}$ fusion protein in the *E. coli* BL21 (λDE3)/ pGS2263 strain

Coomassie-stained SDS-PAGE gels analysing the composition of soluble and insoluble cell fractions, identifying $Yeal_{GGDEF}$ solubility in the pET32a vector. Different IPTG concentrations and growth temperature combinations were used to induce gene expression, with A) 100 µg/ml IPTG at 25°C, B) 20 µg/ml IPTG at 25°C, C) 100 µg/ml IPTG at 4°C and D) 20 µg/ml IPTG at 4°C.

For all gels M: Mark 12 protein marker (molecular weights in KDa). Numbers 0-6 indicate induction times in h, with "i" indicative of insoluble fractions and "s" indicative of soluble fractions. The position of His_6-$Yeal_{GGDEF}$ is indicated (expected size 32 KDa).

AKTA prime (Section 2.6.5). However this was unsuccessful, producing no improved protein levels, and was therefore not further investigated.

3.5 Purification of His_6-$YeaI_{GGDEF}$

Despite solubility levels of $YeaI_{GGDEF}$ being poor, purification of the protein was undertaken, using the N-terminal His-tagged construct (pET32a: *$YeaI_{GGDEF}$*). Frozen cell pellets of the $YeaI_{GGDEF}$ overproduction strain (1-5 litres of BL21 λ(DE3)/pGS2263) were thawed and resuspended in 25 mM HEPES breakage buffer, before the cells were broken by sonication and clarified by centrifugation (Section 2.6.2). The soluble fraction was retained and filtered before application onto a 1 ml HiTrap column for affinity chromatography (Section 2.6.3). The His-chelating programme was run on an ATKA prime, following manufacturer's instructions, and gave a multi-peak trace, indicating various stages of protein elution (Figure 3.7A). The initial peak is characteristic of non-specific protein elution, whilst the later peak usually indicates elution of the stronger binding His-tagged protein. Elution fractions spanning this later peak were analysed by SDS-PAGE and showed relatively pure protein at approximately 32 KDa, the expected molecular weight for the His_6-$YeaI_{GGDEF}$ protein (Figure 3.7B). The fractions containing the most pure His_6-$YeaI_{GGDEF}$were then combined and concentrated using a 10 KDa cut-off Vivaspin column, using HEPES or sodium phosphate dialysis buffer to buffer exchange the protein (Section 2.6.7). The protein concentration was assessed by Bio-Rad assays and if necessary, further concentrating the protein until a sufficiently high concentration was achieved (Section 2.6.8).

Anion exchange was investigated as a second purification step, hoping to remove some contaminating proteins (Section 2.6.4). However, whilst some contaminating polypeptides were lost, the yield of $YeaI_{GGDEF}$ was compromised (data not shown); hence this step was omitted in subsequent purification procedures.

3.6 Analysis of $YeaI_{GGDEF}$ enzymatic activity

To assess whether $YeaI_{GGDEF}$ was an active or inactive DGC protein, enzymatic assays were carried out with GTP (the substrate for DGCs) as well as various other nucleotides. The assays consisted of 1 µM purified protein (Section 3.5) and 10 mM Mg^{2+} or Mn^{2+} incubated with various nucleotides (100 µM) at 37°C for 1 h. Following incubation, reactions were heat inactivated at 95°C for 5 min, and then analysed by HPLC analysis, with absorption at

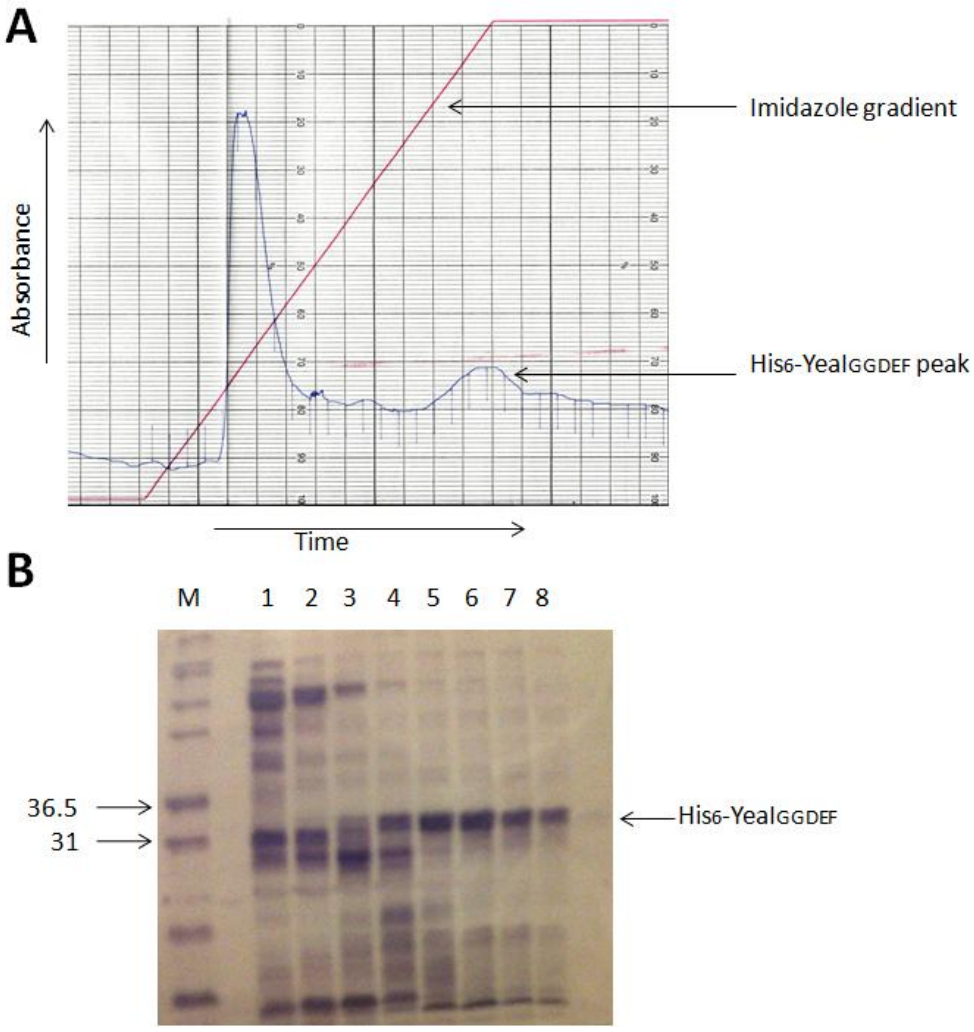


Figure 3.7: HiTrap chelating chromatography to purify His_6-$Yeal_{GGDEF}$

A) Elution profile of His_6-$Yeal_{GGDEF}$ from HiTrap chelating chromatography. Cell-free extract containing His_6-$Yeal_{GGDEF}$ was applied to a 1 ml HiTrap chelating column and fractionated by an imidazole gradient (linear gradient from 0-0.5 M). The blue trace shows the detected A_{280} levels, indicative of protein levels, with the red trace showing the imidazole gradient. The expected His_6-$Yeal_{GGDEF}$ location is indicated.

B) Coomassie-stained SDS-PAGE gel of His_6-$Yeal_{GGDEF}$ purification.
M= Mark 12 protein marker (molecular weights indicated in KDa), lanes 1-8 are elution fractions from the HiTrap column spanning the His_6-$Yeal_{GGDEF}$ peak seen in the elution profile (A).

The position of His_6-$Yeal_{GGDEF}$ is indicated (of expected molecular weight 32 KDa).

254 nm (Section 2.7.1).

Control reactions were carried out, using catalytically active DGC and PDE proteins. Purified protein samples of PleD (a known DGC) and YfgF (a known PDE) were kindly donated by Dr. Melissa Lacey (University of Sheffield) and incubated with known substrates (GTP and c-di-GMP respectively), before separating these samples by HPLC (Figure 3.8). For both proteins, a distinct profile shift was detected when the nucleotides alone and when the nucleotide and relevant protein were mixed, showing that the assay was able to detect the enzymatic activity of the proteins by the ability to differentiate between nucleotides.

The assay was then used to study $YeaI_{GGDEF}$ enzymatic activity. This showed identical profiles when the black (nucleotide only) and pink (nucleotide and protein) traces were overlayed (Figure 3.9). Identical traces were seen when using Mg^{2+} or Mn^{2+} (data not shown for Mn^{2+} series). This confirms that $YeaI_{GGDEF}$ does not act under any of the tested nucleotides. Of particular importance is of course GTP, the substrate regularly converted by DGCs into c-di-GMP. This too, had an identical elution trace when GTP individually was loaded, as when the GTP and YeaI mix was analysed, showing that GTP remained in the sample. The elution times for GTP, c-di-GMP and pGpG were 12.0, 13.3 and 11.6 min respectively, with control reaction carried out at a later date, with minor elution differences here being accounted for by buffer pH fluctuations.

This experiment therefore confirmed that the TrxA-His_6-S tag-$YeaI_{GGDEF}$ fusion protein lacks DGC activity, with no reactivity with GTP, or the other tested nucleotides. This supports the 'degenerate' nature of $YeaI_{GGDEF}$ reported by Sommerfeldt *et al.* (2009) and contradicts the conclusion of Sanchez-Torres *et al.* (2011).

3.7 Nucleotide binding by $YeaI_{GGDEF}$

After confirming that $YeaI_{GGDEF}$ was an inactive DGC, we considered the possibility that $YeaI_{GGDEF}$ might bind a nucleotide, such as GTP or c-di-GMP. It is known that the GGDEF catalytic motif is altered to EGEVF in YeaI, however that did not necessarily imply that the binding cleft was lost entirely (Sommerfeldt *et al.,* 2009). In addition, GGDEF domains are known to have I-sites, which act by binding c-di-GMP to produce a negative feedback loop, slowing the catalytic activity of the DGCs. These I-sites are characterised by an RXXD motif, which is maintained in YeaI (Sommerfeldt *et al.,* 2009). To investigate the nucleotide

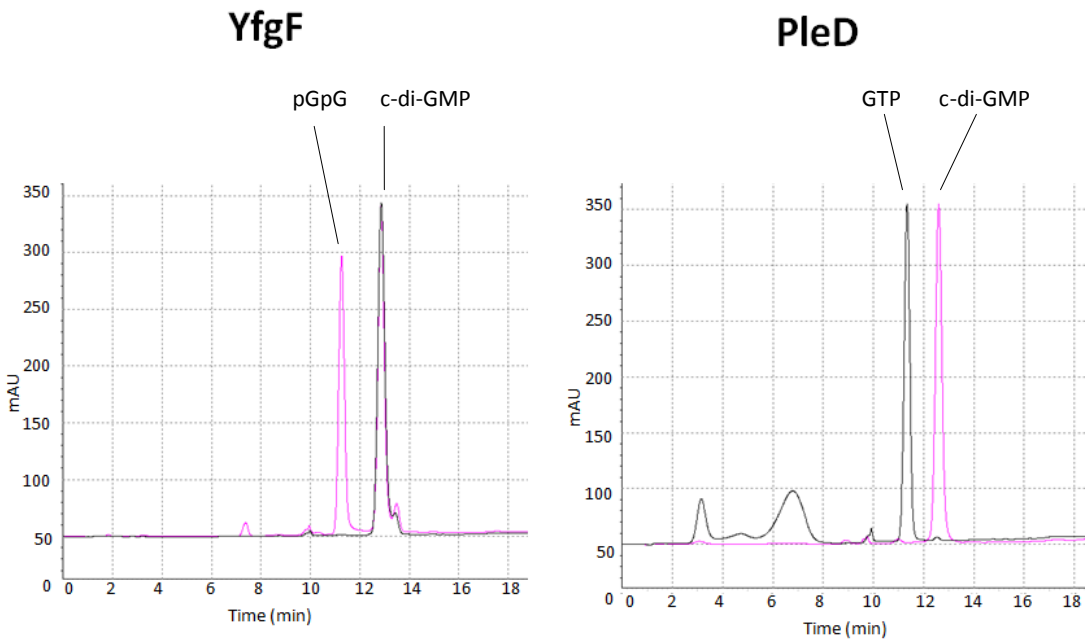


Figure 3.8: DGC/PDE assay controls, using active DGC and PDE enzymes

HPLC elution traces, plotting retention time (min) against intensity (mAU).

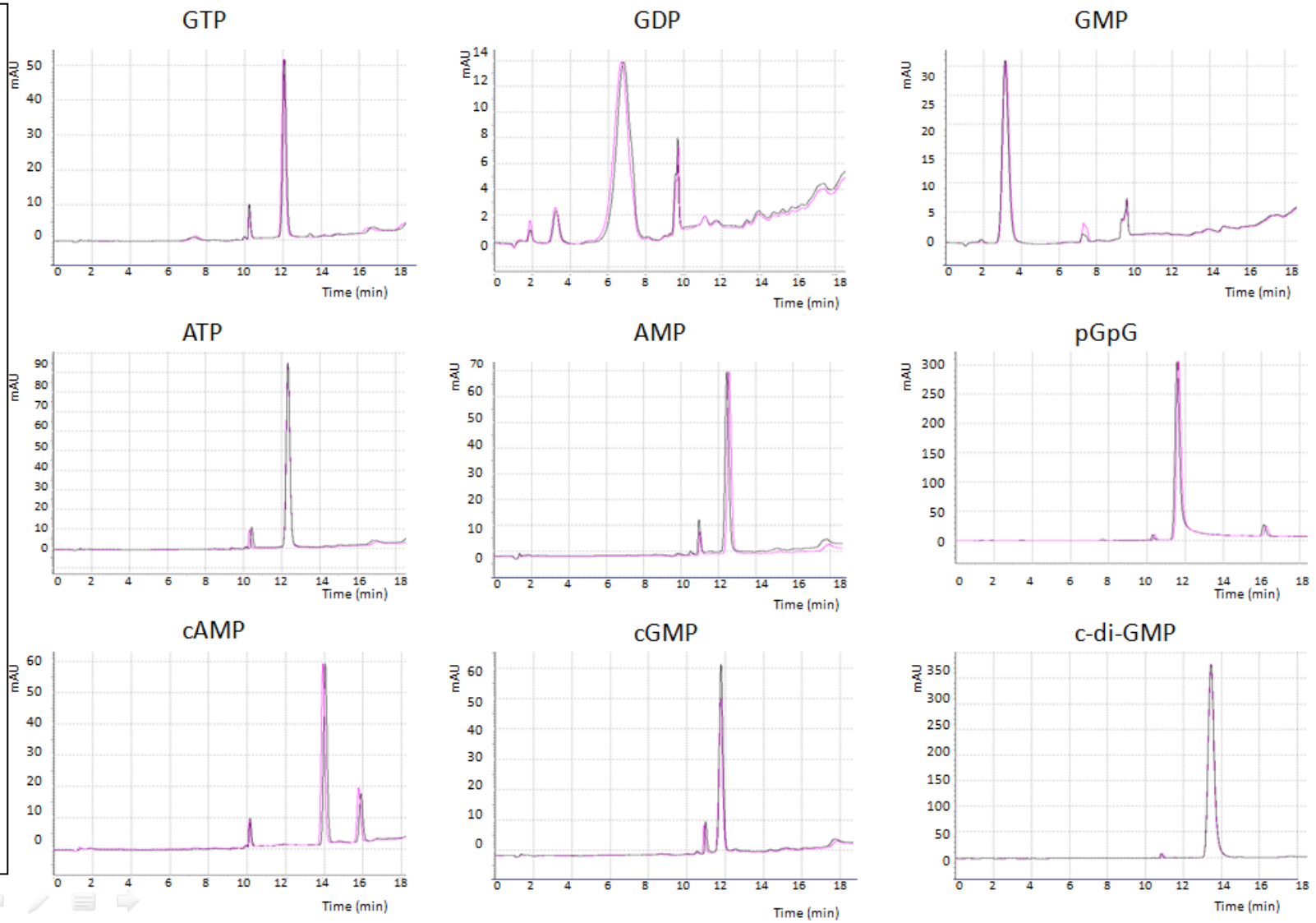
On the left hand side is the overlayed trace for YfgF, an active PDE. This protein is therefore expected to convert c-di-GMP into pGpG, which was tested by analysing a c-di-GMP only sample with a c-di-GMP and YfgF sample. The black trace (c-di-GMP only) gives a peak corresponding to c-di-GMP, whilst the pink trace (YfgF and c-di-GMP) indicates the presence of a major new peak, confirming the activity of the protein.

On the right hand side, is the overlayed trace for PleD, an active DGC. DGCs are known to convert GTP into c-di-GMP, so this activity was tested by analysing GTP both in the presence and absence of PleD and comparing the elution traces. Here the black (GTP only) trace is completely separate from the pink (PleD and GTP) trace, confirming turnover of GTP by the protein.

Figure 3.9: Confirmation of YeaI as a 'degenerate' or inactive DGC by HPLC analysis

HPLC elution traces for each indicated nucleotide, with and without YeaI addition, in the presence of 10 mM Mg^{2+}. The various nucleotides (as specified) were assayed independently (black traces), as well as in a mixture with the YeaI protein (pink traces).

In all cases, the traces are of nucleotide retention time (min) against the 254 nm intensity (mAU).



binding ability of $YeaI_{GGDEF}$, partial proteolysis assays were carried out.

Partial proteolysis was used to identify nucleotide binding by analysing trypsin cleavage profiles of the protein in the presence of various nucleotides. The highly specific enzyme trypsin cleaves at the C-terminal of Lys and Arg residues in a protein, but only cleaving at the accessible surface residues rather than buried ones. For each cleavage the enzyme makes, an additional polypeptide fragment is produced, appearing as a new species on an SDS-PAGE gel. Theoretically ligand binding can conceal or reveal (by inducing a conformational change) cleavage sites, either reducing or decreasing the number of fragments seen on a gel. Of course, nucleotide binding may occur without affecting the environment of Lys or Arg residues, which would be undetectable and is therefore one of the major limitations of the assay.

Optimisation of trypsin cleavage was achieved using 10 µg purified $YeaI_{GGDEF}$ (Section 3.5), adding trypsin enzyme at different stock concentrations (1 mg/ml or 0.2 mg/ml), different times (1, 5 or 10 min) and incubating at different temperatures (RT or 37^{o}C) (Section 2.7.2). The samples were then analysed by SDS-PAGE (Section 2.6.9) which identified the optimal trypsin cleavage conditions to be 100 µg/ml trypsin (using a 1 mg/ml trypsin stock) incubated for 5 min with $YeaI_{GGDEF}$ at room temperature (Figure 3.10). Under these conditions the $YeaI_{GGDEF}$ was clearly cleaved into smaller fragments with some intact protein remaining, and the contaminating polypeptide(s) remaining mainly intact.

Following optimisation, partial proteolysis was carried out, incubating purified $YeaI_{GGDEF}$, trypsin and excess nucleotide (Section 2.7) (2 mM final concentration) under the optimised cleavage conditions, before stopping the reactions and electrophoresing the samples on an SDS-PAGE gel (Sections 2.7.2 and 2.6.9). The ligand free samples acted as control reactions, to compare the nucleotide addition reactions with, and indicated production of two cleavage polypeptides upon trypsin addition (Figure 3.11- indicated by arrows). Nucleotide addition produced only minor profile differences, specifically upon ppGpp, GDP and ATP addition. The presence of these three nucleotides caused the major cleavage product of ~ 15 KDa (as seen in the no ligand sample) to become much more prominent. However, whilst these differences might be indicative of ligand binding, any slight gel loading inaccuracies could cause the same intensity variations seen. The major 15 KDa cleavage product could possibly correspond to the TrxA protein of the TrxA-His_6-S tag-

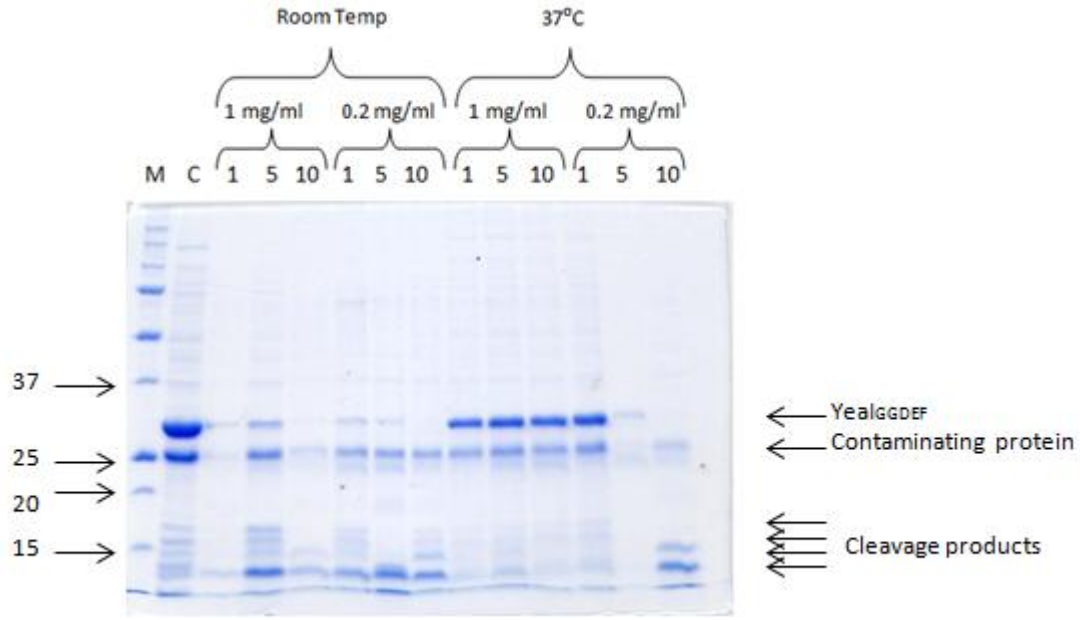


Figure 3.10: Trypsin cleavage optimisation assay for the cleavage of $Yeal_{GGDEF}$

Coomassie- stained SDS-PAGE gel showing trypsin cleavage optimisation for YeaI.

M: Precision plus protein standards, with molecular weights shown in KDa. C: Concentrated YeaI at 1 mg/ml (15 µg loaded) with no trypsin addition.
The remaining samples contain 10 µg YeaI with trypsin addition under the conditions indicated. Trypsin concentrations (1 mg/ml or 0.2 mg/ml stocks giving 100 µg/ml and 20 µg/ml trypsin per sample respectively), incubation temperatures (room temperature or 37°C) and times (1, 5 and 10 min) were varied, using all possible combinations.

Position of the intact YeaI (expected molecular weight 32 KDa) is indicated, along with the major contaminating protein, and the cleavage products produced by trypsin digestion.

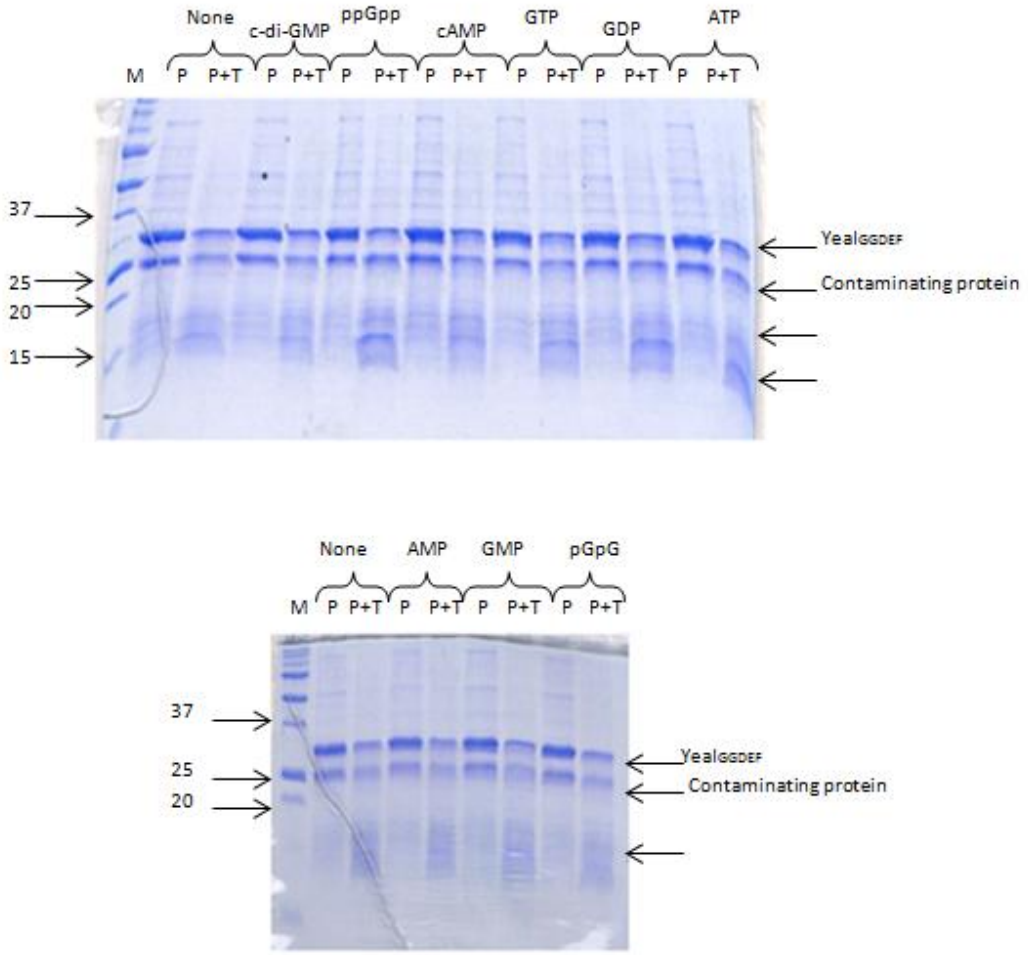


Figure 3.11: Partial proteolysis of Yeal$_{GGDEF}$ in the presence of nucleotides

Coomassie-stained SDS-PAGE gel showing partial proteolysis assays of YeaI and various nucleotides.

M: Precision plus protein standards marker (with molecular weights shown in KDa). For each lane, P indicates a protein only lane, with YeaI protein (10 µg) and no trypsin addition. Lanes labelled P+T indicates the same amount of YeaI with the addition of trypsin, using the 1 mg/ml stock as determined by optimisation trials. Ligands were added to the reaction mix as indicated at 2 mM final concentration, and incubated with the protein and trypsin mix.

The position of the intact Yeal$_{GGDEF}$ (of expected molecular weight 32 KDa) is indicated, along with the major contaminating protein and the visible cleavage products (identified by the unlabelled arrows).

$YeaI_{GGDEF}$ fusion. TrxA has an expected molecular weight of 12 KDa, with the linker region between TrxA and $YeaI_{GGDEF}$ providing a likely relatively unfolded region, which dependent on the nature of the residues, may well provide sites for trypsin cleavage.

To account for gel loading variations, quantification of the protein levels visible on the SDS-PAGE gels (Figure 3.11) were carried out using Image-J software, enabling comparisons between samples (Section 2.10.1). The levels of protein following trypsin cleavage were analysed; specifically the levels of intact protein and the predominant cleavage product (~ 15 KDa), to analyse the effect of nucleotide addition on the extent of protein cleavage (Figure 3.12). There are clear fluctuations in the levels of intact protein (blue) for the various samples, averaging at 4500 ±500 arbitrary units. This variation is likely to be predominately caused by gel loading inaccuracies. However, the more drastic variation is in the levels of the major cleavage product (red). For most samples this averages at ~ 841 arbitrary units, approximately $1/5^{th}$ of the level of intact protein. However, the ppGpp, GDP and ATP supplemented samples have much higher levels of this cleavage product, with ppGpp addition producing the 15 KDa protein fragment at a level of 2500 arbitrary units, altering the previous ratio between intact protein and cleavage product from 1:5 to a 1:2 ratio. The addition of GDP and ATP to protein samples also increased the level of this cleavage product, quantified at levels of ~ 2000 and ~1500 arbitrary units respectively (giving a 1:2 or 1:3 ratio of intact protein: cleavage product). Of importance, for all three nucleotides (ppGpp, GDP and ATP), the levels of intact protein were approximately the average level (quantified to 4500 arbitrary units), and therefore the increase in cleavage fragment levels cannot be accounted for by SDS-PAGE gel overloading.

The presence of these three nucleotides appears to have altered the profile of $YeaI_{GGDEF}$ protein cleavage, suggesting a possible interaction with $YeaI_{GGDEF}$. One explanation is that nucleotide binding to the protein has induced a conformational change to the protein which has increased the exposure of a trypsin cleavage site, thereby causing a higher level of cleavage at this site. However, whilst this is a feasible hypothesis, the low purity protein sample casts doubts to this explanation, with a contaminated and smeary image seen in the SDS-PAGE gels (Figure 3.11). Not only does the smeary image make quantification difficult, and more error-prone, but the presence of the contaminating protein also raises the possibility that this protein is being digested by trypsin too, and therefore contributing to the cleavage pattern. Whilst there is some indication of nucleotide binding to $YeaI_{GGDEF}$,

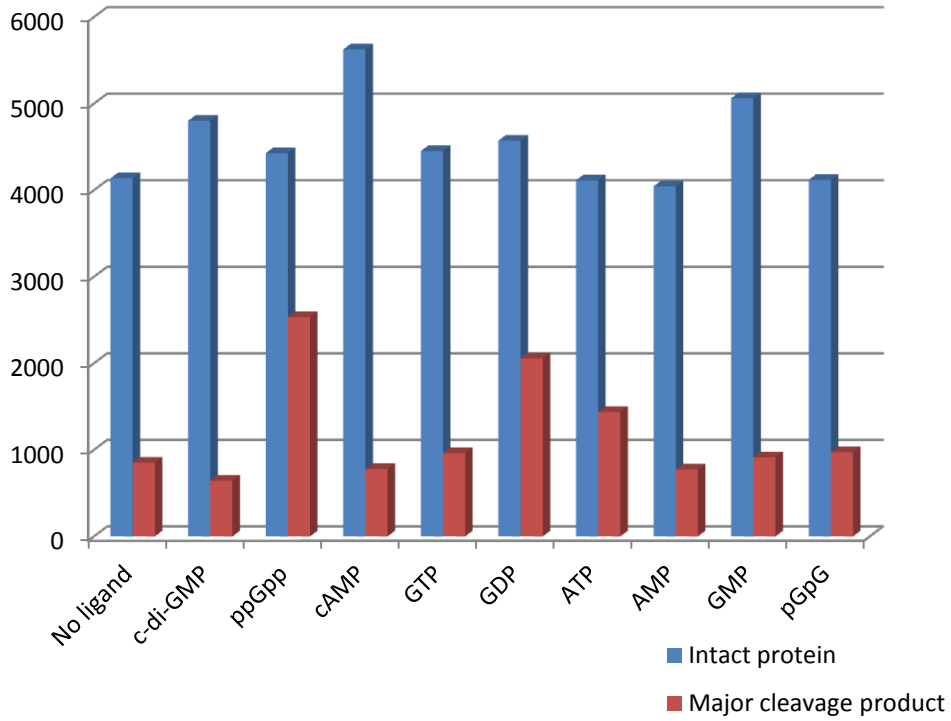


Figure 3.12: Quantification of $Yeal_{GGDEF}$ partial proteolysis products upon nucleotide addition

Image-J quantification of protein levels following trypsin partial proteolysis for $Yeal_{GGDEF}$ protein samples containing various nucleotides.

The levels of intact $Yeal_{GGDEF}$ protein (blue) and levels of the major cleavage product at ~ 15 KDa (red) following trypsin cleavage of $Yeal_{GGDEF}$ were quantified from the SDS-PAGE gels in Figure 3.11. The protein levels were analysed in the presence of various nucleotides (as specified).

Values on the y-axis show arbitrary units referring to the intensities of the protein species.

a more concentrated and purer protein sample is essential to enable further assays to be carried out, such as NMR, fluorescence spectroscopy and to repeat the partial proteolysis assays. Combining the results from various interaction-probing techniques and simply using a cleaner protein sample, would clarify these possible protein and nucleotide interactions, giving a more conclusive result.

Of note, the identity of the major contaminating species at approximately 25 KDa was investigated, as it could be an interaction partner of $YeaI_{GGDEF}$. The contaminating protein was isolated from an SDS-PAGE gel in which the $YeaI_{GGDEF}$ protein preparation was electrophoresed (Section 2.6.9). This contaminating protein was then analysed by N-terminal sequencing (Dr. Arthur Moir- University of Sheffield) (Section 2.6.14). N-terminal sequencing identified this protein to have the amino acid residue sequence SRVCQVT, belonging to the N-terminus of the protein. This protein sequence was submitted to the FASTA database (http://www.ebi.ac.uk/Tools/sss/fasta/) and identified the sequence to certainly belong to the 50S ribosomal protein RpmB (synonym L28). RpmB is a component of the 50S ribosome subunit and is required for ribosomal assembly (Maguire & Wild, 1997). However, RpmB has an expected molecular weight of only 9 KDa, compared to the ~ 25 KDa seen by protein electrophoresis, indicating the possibility of a protein complex involving RpmB to account for this higher molecular weight species. Several 50S ribosomal subunits have been identified to artificially crosslink with other 50S ribosomal subunits, including RpmB (L28) with Rpl1 (L9) (Walleczek *et al.,* 1989; Redl *et al.,* 1989). Together this RpmB-Rpl1 complex gives a molecular weight of 25 KDa, however this complex has only been identified in artificial crosslinking conditions rather than *in vivo* (Walleczek *et al.,* 1989; Redl *et al.,* 1989). Therefore, the visible contaminating species can reasonably confidently be identified as the ribosomal subunit RpmB in complex with an additionally protein, possibly L9, but equally possibly another cellular protein. Further investigations are thereby crucial here in order to determine the exact identity of the co-purifying 25 KDa species.

Despite identification, there is no obvious reason why RpmB would be co-purifying with $YeaI_{GGDEF}$, with no reported involvement of $YeaI_{GGDEF}$ in ribosomal assembly or function. To investigate further, pull-down assays could be carried out, using either purified proteins or overproduced soluble extracts, immobilizing $YeaI_{GGDEF}$ to a HiTrap column via its His_6 tag, and adding untagged RpmB to the column. Following wash and elution steps, analysis of

the protein fractions collected throughout the experiment would identify whether or not an interaction was occurring between RpmB and $YeaI_{GGDEF}$ or whether the appearance of RpmB in the gels was purely an artefact from the protein preparation.

3.8 Oligomeric analysis of $YeaI_{GGDEF}$

The oligomeric state of $YeaI_{GGDEF}$ was investigated with glutaraldehyde crosslinking assays and gel filtration analysis. DGC proteins are characteristically active as dimers, with an active site formed in the cleft between two GGDEF units (Chan *et al.,* 2004). We therefore wished to determine whether $YeaI_{GGDEF}$ maintained this characteristic, or whether this was lost upon motif degeneration. Of note, TrxA is a monomer, meaning that any oligomerisation detected for the TrxA-His_6-S-tag-$YeaI_{GGDEF}$ protein would be dependent on $YeaI_{GGDEF}$ interactions only.

Cross-linking assays were carried out, adding glutaraldehyde to protein samples, which acted as a low-specificity crosslinker for any two amino groups. Samples were made consisting of purified $YeaI_{GGDEF}$ (Section 3.5) and varying quantities of glutaraldehyde (0-1%), which were incubated at room temperature before quenching the reactions with 1M Tris pH 6.8 (Section 2.6.12). Samples were then analysed by SDS-PAGE gel (Section 2.6.9).

The Coomassie-stained SDS-PAGE gel clearly showed the monomeric protein (of expected molecular weight 32 KDa) in the low glutaraldehyde samples, the proportion of which decreases as the % of glutaraldehyde increases (Figure 3.13). At 0.01 and 0.02% glutaraldehyde, two additional species are present; one between 50 and 100 KDa, and also an extremely large species at >250 KDa (Figure 3.13). The large >250 KDa species is most logically explained as being a product of non-specific glutaraldehyde crosslinking, simply forming a very large artificial aggregate. Using the monomeric size of 32 KDa, a dimer should theoretically be 64 KDa and a trimer 96 KDa, meaning that the species between 50-100 KDa is either a dimer or a trimer with aberrant motility. Since active DGCs form dimers (Chan *et al.,* 2004), it is more likely that this oligomeric form is a $YeaI_{GGDEF}$ dimer rather than a trimer.

Gel filtration analysis was carried out by loading a purified sample of $YeaI_{GGDEF}$ protein (20 mM sodium phosphate, 100 mM NaCl pH 7.5) (Section 3.5) onto an equilibrated Superdex

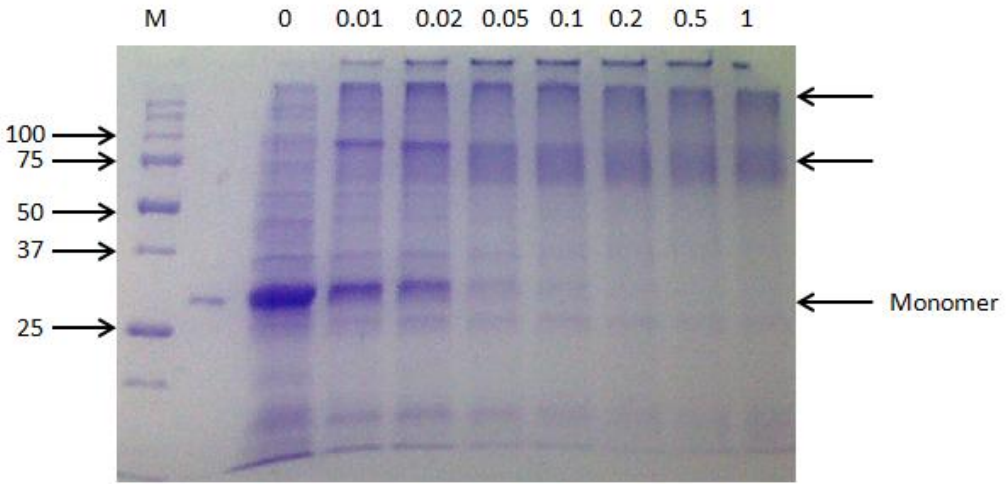


Figure 3.13: Oligomerisation analysis of $YeaI_{GGDEF}$

Coomassie-stained SDS-PAGE gel of $YeaI_{GGDEF}$ in the presence of various % glutaraldehyde.

M: Precision plus protein standards (molecular weights shown in KDa). The remaining lanes show YeaI samples, with varying % of glutaraldehyde added, as specified.

Monomeric $YeaI_{GGDEF}$ of expected size 32 KDa, is indicated.

Of note, a contaminating species at ~ 25 KDa is faint but present, as previously seen in partial proteolysis (Figures 3.8 and 3.9).

200 gel filtration column (Section 2.6.13). Gel filtration acts to separate proteins based purely on size, eluting larger proteins early, retaining smaller proteins in the column for longer. The A_{280} value (indicative of protein levels) was recorded as the sample passed through the gel filtration column, quantifying the level of protein eluted at various points. The gel filtration trace showed multiple protein peaks (Figure 3.14). Standard traces, using reference proteins of known molecular weight enabled conversion of the retention volumes into molecular weights. The predominant peak eluted at a retention volume of 14.52 ml is indicative of an 82 KDa species, which is similar to the dimeric or trimeric oligomer detected by crosslinking assays (Figure 3.13, 3.14). Of note, a monomeric peak is absent in this trace, with the only other significant peaks at retention volumes of 15.49 ml (equivalent to 44 KDa) and 16.45 ml (equivalent to 22 KDa). The smaller 22 KDa is a contaminating species, apparent in the aforementioned crosslinking and partial proteolysis assays and identified by N-terminal sequencing as being a complex containing L28 (Section 3.7). The 44 KDa species can therefore be accounted for in two ways; as a contaminating protein (meaning that all the $YeaI_{GGDEF}$ is present as this 82 KDa species), or as a $YeaI_{GGDEF}$ monomer bound to cellular cofactors that have been co-purified. Either way, there is a limit to the certainty of the conclusions, given the poor purification and yield of the protein. In this case, it was only possible to produce 0.18 mg of low purity protein to analyse by gel filtration, making analysis difficult. Ideally, cleaner protein sample at a much higher purity would be used for both gel filtration and cross-linking assays in order to achieve more definitive results.

3.9 Optimising the yield of soluble $YeaI_{GGDEF}$

One of the recurring problems experienced throughout the work with $YeaI_{GGDEF}$ was the high level of protein insolubility, with only a very small proportion of the overproduced protein being soluble and available for future use.

Attempts to alleviate this problem, by cloning $YeaI_{GGDEF}$ into different vectors, altering the fused tag position, by additional purification steps and by attempting to solubilise the insoluble pellet, did not significantly alter the solubility and yield of purified protein. Whilst some assays were possible with the limited protein produced, many demanded more protein than it was possible to produce. This meant that further binding assays such as NMR or fluorescence and structure solving techniques such as X-ray crystallography were

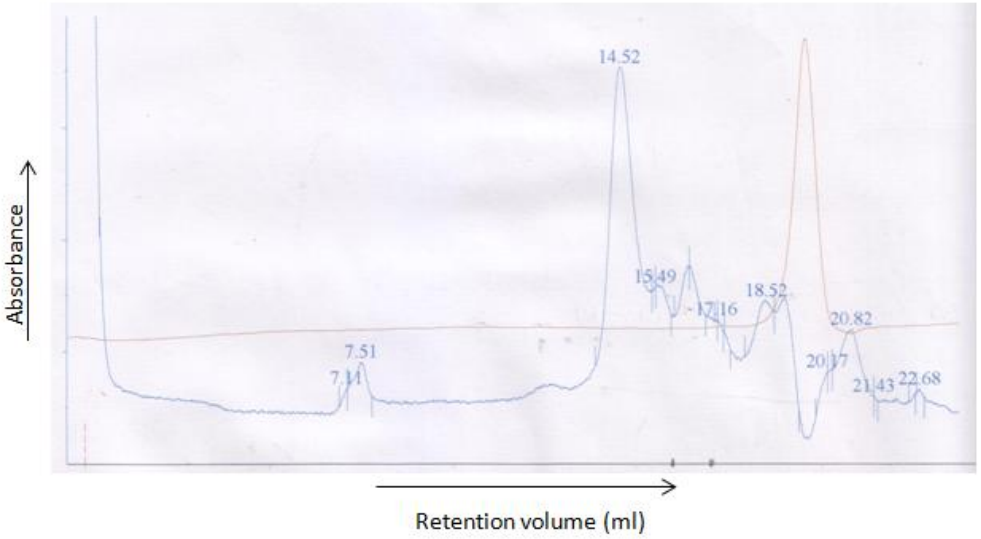


Figure 3.14: Gel filtration trace for $Yeal_{GGDEF}$

Elution profile of His_6-$Yeal_{GGDEF}$ from gel filtration chromatography.

Purified $Yeal_{GGDEF}$ (concentration 0.18 mg/ml in 20 mM sodium phosphate, 100 mM NaCl, pH 7.5 buffer) was applied to an equilibrated Superdex 200 gel filtration column.

The absorbance of eluant at 280 nm (blue trace) was plotted against retention volume, with the numbers at the crest of each peak specifying the retention volume of each eluted protein. The conductivity applied to the column (red line) and the point at which the protein was added to the gel filtration column (pink dotted line) are shown.

simply impossible. Therefore, increasing the solubility of the protein was a major hurdle and demanded an alternative method for $YeaI_{GGDEF}$ production before further experiments could be undertaken.

3.9.1 Ligation of $YeaI_{GGDEF}$ into pCOLD TF vector

In an attempt to increase solubility of the overproduced YeaI protein, the *yeaI* $_{GGDEF}$ sub-gene was cloned into a pCOLD TF vector (Takara), to form a fusion protein with an N-terminal trigger factor (TF) soluble chaperone, a His_6 tag and a cleavage site between these N-terminal tags and our protein of interest $YeaI_{GGDEF}$ (Figure 3.15). The TF protein is a chaperone, anticipated to assist protein folding and to reduce insoluble protein production (Hoffmann *et al.,* 2010). Moreover, the vector is a cold-shock expression vector, inducing protein expression at lower temperatures. This is dually beneficial, suppressing the expression of other cellular proteins with the decreasing growth temperature and theoretically increasing protein solubility too, reducing inclusion body formation (Takara Bio manufacturer's instructions).

The *yeaI* $_{GGDEF}$ sub-gene was amplified from genomic *E. coli* MG1655 DNA and ligated into pCOLD TF by use of primers NSW11 and NSW20. These primers were designed to contain an NdeI and a BamHI restriction site at either end of the PCR fragment respectively (Table 3.3).

Table 3.3: Primers used in amplification of the *yeaI*$_{GGDEF}$ gene for insertion into the plasmid pCOLD TF

Primer	Sequence	Function
NSW11	TTTTCATATGAAGGACTTCTGTGTCAT GCTG	Forward primer for *yeaI*$_{GGDEF}$ into pCOLD TF vector
NSW20	TTTTGGATCCTTAAATATGATGAATAA TCACTTTG	Reverse primer for *yeaI*$_{GGDEF}$ into pCOLD TF vector

The grey shaded sequence indicates an NdeI restriction site and the boxed sequence shows a BamHI restriction site. A stop codon was included in the reverse primer (NSW20) to terminate transcription at the end of the *yeaI*$_{GGDEF}$ gene.

PCR reactions were carried out using high fidelity polymerase, *E. coli* MG1655 genomic DNA and primers NSW 11 and NSW20 to amplify *yeaI*$_{GGDEF}$ (Section 2.3.1; 2.3.3). The resultant

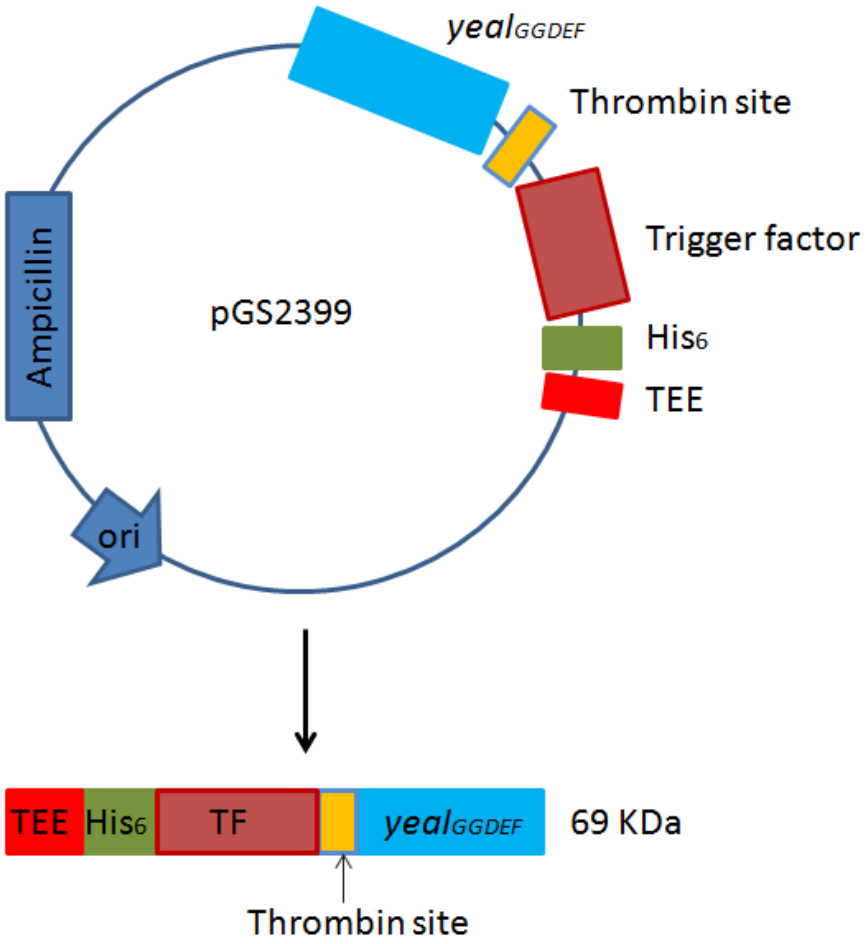


Figure 3.15: Schematic plasmid maps for the insertion of *yeal*$_{GGDEF}$ into the pCOLD TF vector

Map for the pGS2399 plasmid, derived from pCOLD TF vector, indicating locations of the tags, antibiotic resistance cassette and *yeal*$_{GGDEF}$ insertion site, and the fusion protein produced.

The origin of replication is indicated and labelled as ori.

This N-terminal region consists of a translation enhancing element (TEE), a His_6 tag and trigger factor (TF) tag adjacent to a thrombin cleavage site which separates the N-terminal region to the *yeal*$_{GGDEF}$ insertion site. Alongside the defined tags, linker regions exist between the various tags and insert and therefore contribute to the fusion protein molecular weight.

The fusion protein (expected molecular weight 69 KDa) is cleavable at three sites, one of which is the chosen thrombin site. Cleavage at this site should produce two fragments: Yeal$_{GGDEF}$ (expected molecular weight 16 KDa) and the TF tag (expected molecular weight 53 KDa).

PCR product was cleaned by PCR purification (Section 2.3.10) and digested using the NdeI and BamHI restriction enzymes (Sections 2.3.6). The purified pCOLD TF plasmid (Takara) was digested with the same restriction enzymes (NdeI and BamHI) (Sections 2.3.4 and 2.3.6). The ligated product consisting of the *yeaI*$_{GGDEF}$ fragment and pCOLD TF vector were used to transform electrically competent *E. coli* DH5α cells with selection on ampicillin supplemented agar (Sections 2.3.9 and 2.2.7). The resultant colonies were selected and screened using colony PCR with the sequencing primers NSW21 and NSW22 to assess which plasmids contained the correctly sized insert (Table 3.4) (Section 2.3.11).

Table 3.4: Sequencing primers for pCOLD TF vector

Primer	Sequence	Function
NSW21	GCGAAAGTGACTGAAAAAGAA	Forward sequencing primer for pCOLD TF vector
NSW22	ATCGATTATTTATTTCCTGAAAAC	Reverse sequencing primer for pCOLD TF vector

Plasmids containing an appropriate insert were checked by DNA sequencing (Source Bioscience) using the same sequencing primers (Table 3.4).

Plasmids containing no mutations in the *yeaI*$_{GGDEF}$ sequence were named pGS2399 (pCOLD TF:*yeaI*$_{GGDEF}$). The plasmid was then used to transform electrically competent *E. coli* BL21 λ(DE3) cells (Sections 2.2.7).

3.9.2 Overproduction and solubility trials

Trials were carried out to determine the optimal overproduction conditions for the TF-YeaI$_{GGDEF}$ fusion protein using 500 ml cultures of the *E. coli* BL21 (λDE3)/pGS2399 strain and inducing the cells at 15^{o}C (specified in the Takara manufacturer's instructions) by adding either 100 μg/ml or 20 μg/ml final concentrations of IPTG inducer (Section 2.2.4). In practise an aliquot (5 ml) of overnight culture was added to 500 ml of ampicillin supplemented LB. The cells were then grown at 37^{o}C with shaking at 250 rpm until an OD_{600} of ~-0.6 was reached (Sections 2.2.2, 2.2.3 and 2.6.1). Following this IPTG was added, at the two different concentrations, and cultures grown at 15^{o}C with shaking at 250 rpm for induction. Samples were taken pre-induction (0 h), and then at 2, 4, 6 and 24 h after induction and were analysed by SDS-PAGE (Section 2.6.9).

The SDS-PAGE gel showed overproduction of the fusion protein at the expected size of 69 KDa in all of the induction samples, confirming successful overproduction of the protein (Figure 3.16). Regarding optimal induction conditions, all four induction times produced a similar level of protein. Following this, all eight conditions were analysed to assess the optimal conditions for soluble protein production.

3.9.3 Solubility trials to optimise overproduction of TF-YeaI$_{GGDEF}$

Following the overproduction trials, solubility trials were carried out to determine which induction conditions produced the highest proportion of soluble TF-YeaI$_{GGDEF}$ protein.

The cultures used for overproduction trials were also used for solubility trials (*E. coli* BL21 λ(DE3)/ pGS2399), simply taking a 20 ml sample at the various induction times (2, 4, 6 and 24 h) and pelleting the samples (Section 2.6.1). The cell pellets were resuspended in Tris breakage buffer (pH 7.4) before breaking the cells open by sonication and clarification of the sample (Section 2.6.2). The clarified samples were then separated into soluble and insoluble fractions, resuspending the insoluble fractions in breakage buffer, prior to analysis of both fractions by SDS-PAGE (Sections 2.6.2 and 2.6.9).

The solubility gels demonstrated more fluctuations in protein levels, with the highest proportion of soluble TF-YeaI$_{GGDEF}$ appearing to be after 24 h of induction, using a final concentration of 20 μg/ml IPTG, which was therefore chosen as the optimal overproduction condition (Figure 3.17).

3.9.4 Optimisation of YeaI$_{GGDEF}$ purification

In order to purify YeaI$_{GGDEF}$, the large chaperone trigger factor (TF) tag was cleaved using the thrombin cleavage site in the pCOLD TF vector adjacent to the inserted YeaI$_{GGDEF}$ site (Figure 3.15).

Prior to large scale purification of YeaI$_{GGDEF}$, the conditions for small scale preparation were optimised. Cell pellets of the pCOLD TF:YeaI$_{GGDEF}$ overproduction strain (*E. coli* BL21 λ(DE3)/pGS2399) were resuspended in 20 mM Tris breakage buffer (pH 7.4), the cells were

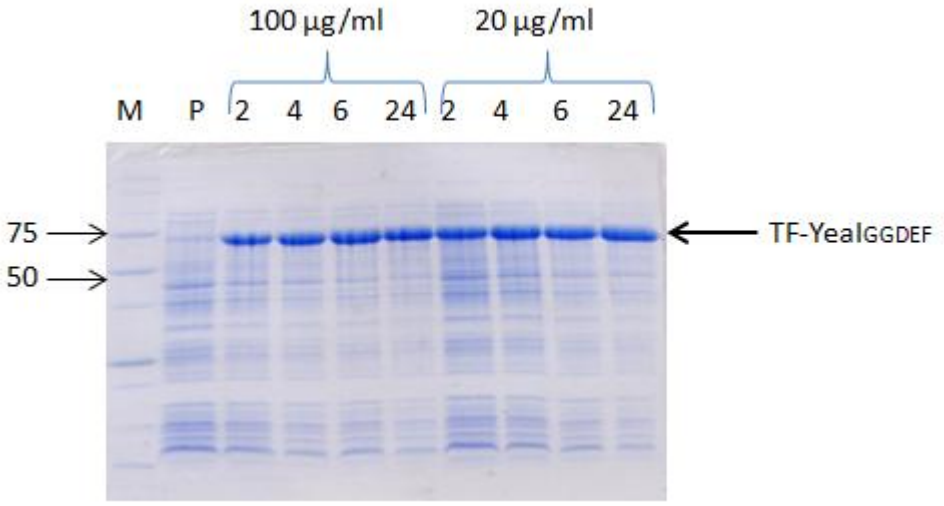


Figure 3.16: Optimisation of overproduction conditions for the TF-Yeal$_{GGDEf}$ fusion protein in *E. coli* BL21 (λDE3) cells

M: Precision plus protein standards (molecular weights shown in KDa).

P: Pre-induction samples, with all other samples showing overproduction samples. The numbers 2, 4, 6 and 24 refer to the time (in h) of induction, with IPTG final concentrations added as specified.

The fusion protein is present at its expected size (69 KDa) as indicated.

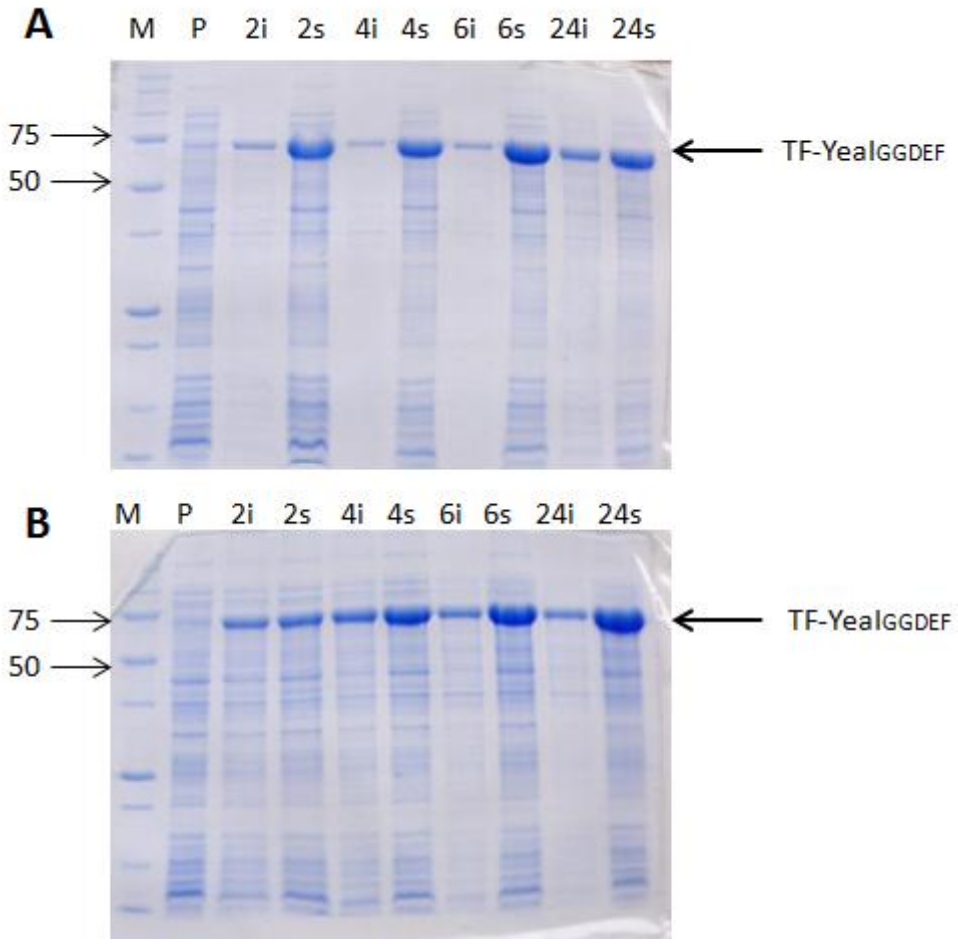


Figure 3.17: Solubility optimisation for TF-Yeal$_{GGDEF}$

Coomassie-stained SDS-PAGE gels, analysing the soluble and insoluble fractions for cells induced with a final concentration of IPTG at (A) 100 µg/ml and (B) 20 µg/ml.

In both cases, M: Precision plus protein standards (molecular weights indicated in KDa). P: Pre-induction cellular sample. All other lanes are soluble and insoluble fractions, with the numbers 2, 4, 6 and 24 referring to induction duration in h. The insoluble and soluble fractions are indicated by "i" and "s" respectively.

In both cases, the location of TF-Yeal $_{GGDEF}$ (expected molecular weight of 69 KDa) is indicated.

broken by sonication, clarified by centrifugation and the soluble fraction retained for use (Section 2.6.2). The soluble fraction was purified by affinity chromatography, adding a small volume to pre-washed Ni-NTA beads, incubating at 4^{o}C for 1 h to immobilise the fusion protein via the His_6-tag. The beads were then thoroughly washed in 20 mM Tris buffer, before thrombin (20 units) was added to the beads and incubated overnight at 4^{o}C (Section 2.6.6). Following this, the supernatant was removed, which should contain the released $YeaI_{GGDEF}$ protein, before washing the beads once more, and eluting the bound His_6-TF protein by an imidazole Tris buffer.

Samples were taken at various stages throughout the purification, enabling analysis by SDS-PAGE (Figure 3.18) (Section 2.6.9). The initial purification appeared not hugely successful, with a significant quantity of the fusion protein present in the unbound sample (lane 3) indicating the Ni-NTA beads were saturated. The thrombin cleavage step was successful to an extent, cleaving the 69 KDa fusion protein into two components: TF and $YeaI_{GGDEF}$ (lane 6). However thrombin cleavage (lane 6) was expected to release $YeaI_{GGDEF}$, whilst retaining the His_6-TF on the resin until the imidazole addition step. However, this was not the case, with both the His_6-TF and $YeaI_{GGDEF}$ being released after addition of thrombin (lane 6) and predominately uncleaved protein being eluted with imidazole addition (lane 9).
Whilst, separation of TF and $YeaI_{GGDEF}$ was a problem which needed to be addressed, the efficiency of thrombin cleavage also required optimisation. The purification was therefore repeated incubating the thrombin at 20^{o}C rather than 4^{o}C, and analysing samples once more by SDS-PAGE gel (Section 2.6.9). Thrombin cleavage was improved at 20^{o}C (Figure 3.19), increasing the proportion of protein cleaved by the enzyme, compared to that at 4^{o}C (Figure 3.18). However, once again, both the $YeaI_{GGDEF}$ and TF proteins were released from the resin directly following thrombin addition (Figure 3.19- lane 4) with no significant quantities of TF remaining bound to the resin until imidazole addition (lane 7).

Further optimisation was achieved by the addition of calcium (at a final concentration of 2.5 mM) and removal of imidazole (previously at 20 mM) to the wash buffers. The assay was carried out using the adjusted buffers and upon SDS-PAGE analysis of the assay samples, showed an increased efficiency of thrombin cleavage, with ~ 1:1 ratio of un-cleaved and cleaved protein (Figure 3.20). In addition, the intact fusion protein (TF:$YeaI_{GGDEF}$) remained bound upon thrombin addition whilst previously being released at this point (Figure 3.18 and Figure 3.19).

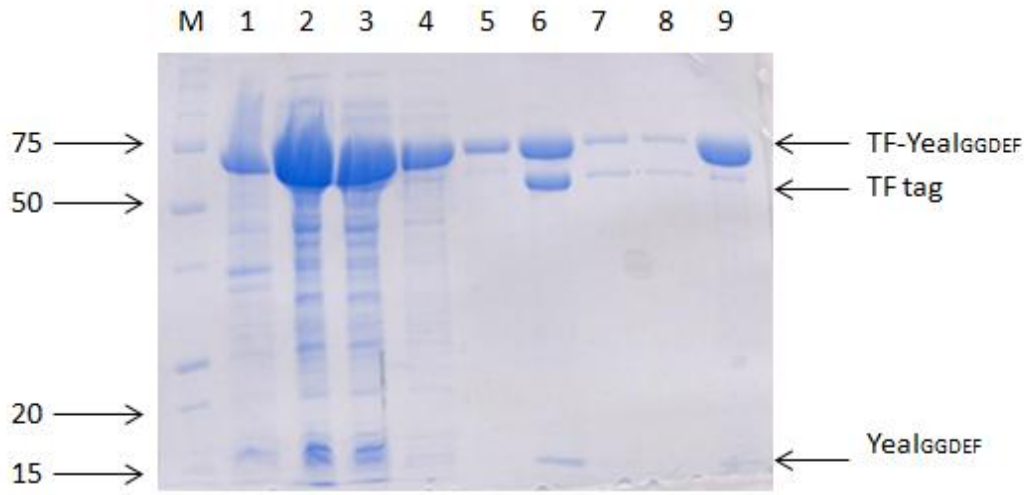


Figure 3.18: Purification of TF-Yeal$_{GGDEF}$, with thrombin incubation at 4^{o}C

Coomassie-stained SDS-PAGE gel showing the initial attempt to purify Yeal$_{GGDEF}$.

M: Precision plus protein standards marker (molecular weights shown in KDa), Lane 1: Insoluble fraction, 2: Soluble fraction, 3: Unbound proteins, 4: Wash steps 1+2, 5: Wash steps 3 + 4, 6: Thrombin step, 7: Wash 5+6, 8: Wash 7+8, 9: Imidazole addition.

The fusion protein (expected MW: 69 KDa), TF tag (expected MW: 53 KDa) and Yeal$_{GGDEF}$ (expected MW: 16 KDa) components are indicated by their respective arrows.

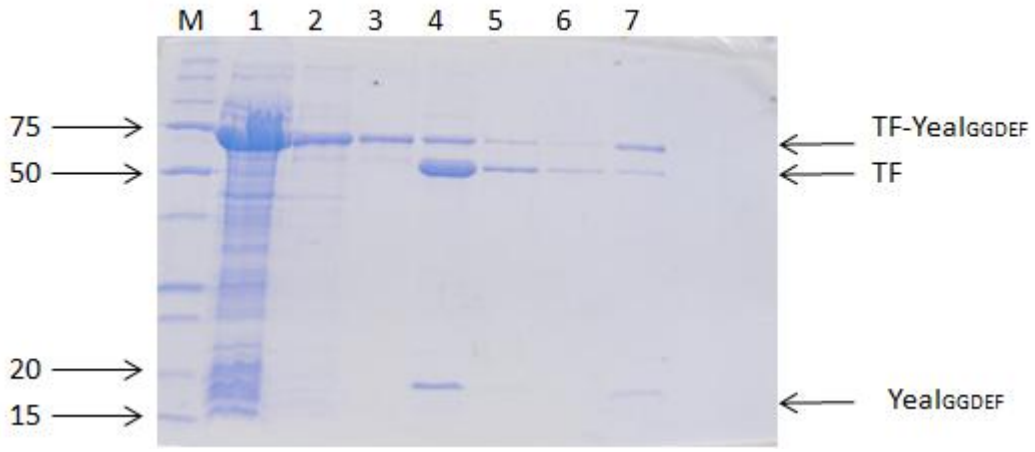


Figure 3.19: Purification of TF-Yeal$_{GGDEF}$, with thrombin incubation at 20^{o}C

Coomassie-stained SDS-PAGE gel, analysing samples in the purification procedure.

M: Precision plus protein standards marker (molecular weights shown in KDa), Lane 1: Unbound proteins, 2: Wash steps 1+2, 3: Wash steps 3 + 4, 4: Thrombin addition, 5: Wash 5+6, 6: Wash 7+8, 7: Imidazole addition.

The fusion protein (expected MW: 69 KDa), TF tag (expected MW: 53 KDa) and Yeal$_{GGDEF}$ (expected MW: 16 KDa) components are indicated by arrows.

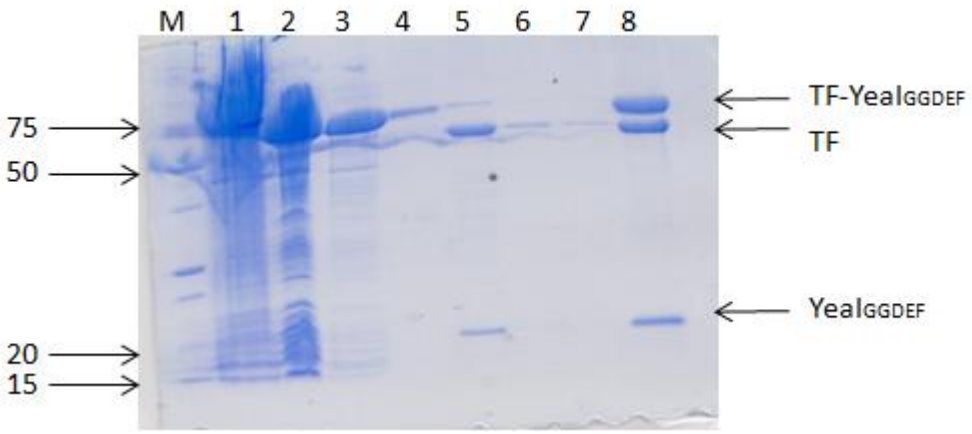


Figure 3.20: Purification of TF-Yeal$_{GGDEF}$, with calcium addition to buffers

Coomassie-stained SDS-PAGE gel, analysing samples throughout the purification steps.

M: Precision plus protein standards marker (molecular weights shown in KDa), Lane 1: Soluble YeaI, 2: Unbound proteins, 3: Wash steps 1+2, 4: Wash steps 3 + 4, 5: Thrombin step, 6: Wash 5+6, 7: Wash 7+8, 8: Imidazole addition.

The fusion protein (expected MW: 69 KDa), TF tag (expected MW: 53 KDa) and Yeal$_{GGDEF}$ (expected MW: 16 KDa) components are indicated by arrows.

The major hurdle to be overcome which became apparent in these trials was the lack of protein (TF and $YeaI_{GGDEF}$) separation. This appears to be due to insufficient TF-His_6 immobilisation to the resin, which may suggest the formation of a potential interaction between $YeaI_{GGDEF}$ and TF. In an attempt to enhance the immobilisation of TF-His_6, a 1 ml HiTrap column was used for the large-scale purification, to provide a larger surface area for immobilisation compared to the low volume Ni-NTA resin used for optimisation trials.

3.9.5 Large scale purification

The purification was then scaled up, using a 1 ml HiTrap column for purification rather than Ni-NTA resin beads. The use of a HiTrap column was dual purpose, firstly to limit the resin reaching saturation for the fusion protein TF:$YeaI_{GGDEF}$ in order to maximise the yield of $YeaI_{GGDEF}$. Secondly, the column was used in an attempt to maximise the immobilisation of TF-His_6 with the column (due to its increased surface area for interactions), ideally releasing only $YeaI_{GGDEF}$ upon thrombin addition.

Cell pellets of the pCOLD TF:YeaI overproduction strain were prepared once more (Section 3.9.2), retaining the soluble fraction for use (Sections 2.6.1 and 2.6.2). The filtered soluble fraction was then applied to a 1 ml HiTrap column for affinity chromatography, mirroring the small-scale purification. The His-chelating programme was run on the ATKA prime, giving a single peak absorbance trace for the eluted fractions (Figure 3.21A) (Section 2.6.3). Elution fractions spanning this peak from the trace were then analysed by SDS-PAGE, and identified the fractions containing TF-$YeaI_{GGDEF}$ (Figure 3.21B).

The TF-$YeaI_{GGDEF}$ fractions were dialysed into TCB buffer (Table 2.7), acting to remove imidazole and to add Ca^{2+} to the buffer, which improved thrombin cleavage in the initial trials (Figure 3.20) (Section 2.6.6). Thrombin (20 units) was then added to these dialysed protein samples and left for 16 h at 20^{o}C, again using the preferential incubation temperature determined previously. Following successful thrombin cleavage, cleaving the vast majority of the fusion TF-$YeaI_{GGDEF}$ into the two consistuent protein components (Figure 3.22A-lane 2), the mixture was manually applied to a 1 ml HiTrap column. Once again, it was proposed that the TF tag would bind to the HiTrap column via the His_6 tag, whilst untagged $YeaI_{GGDEF}$ would flow through the column. Following this, upon imidazole addition (at a single concentration) the bound TF tag and any uncleaved fusion protein would theoretically be eluted. Despite a lack of success in optimisation trials, it was hoped

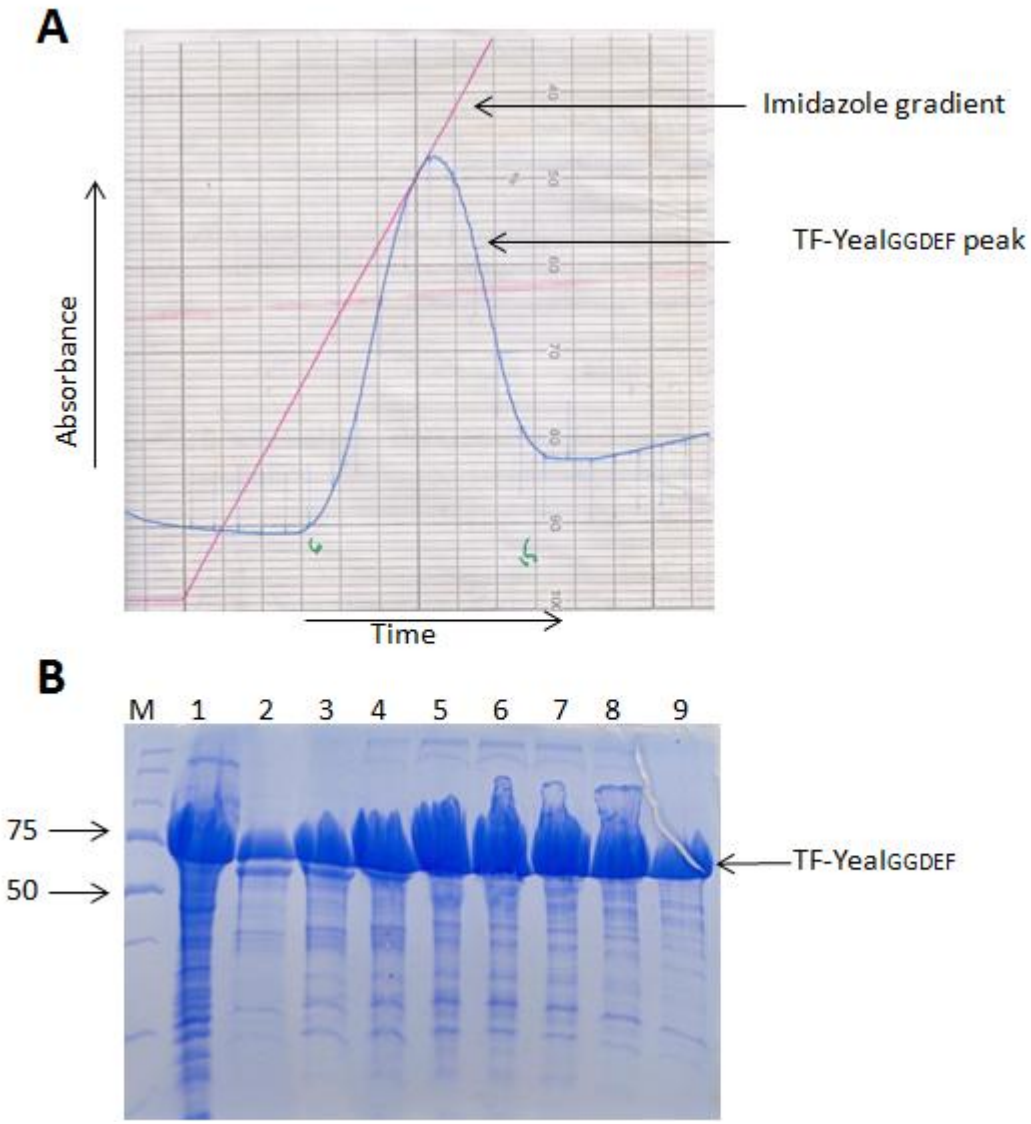


Figure 3.21: HiTrap chelating chromatography as an initial step to purify TF-Yeal$_{GGDEF}$

A) Elution profile of TF-Yeal$_{GGDEF}$ from HiTrap chelating chromatography. Cell-free extract containing TF-Yeal$_{GGDEF}$ was applied to a 1 ml HiTrap chelating column and fractionated by an imidazole gradient (linear gradient from 0-0.5 M). The blue trace shows the detected OD_{280} levels, indicative of protein levels, with the red trace showing the imidazole gradient. The expected TF-Yeal$_{GGDEF}$ location is indicated.

B) Coomassie-stained SDS-PAGE gel of TF-Yeal$_{GGDEF}$ purification. M= Precision plus protein standards (molecular weights indicated in KDa), lanes 1 shows the crude soluble fraction, lanes 2-9 show elution fractions 7-14 collected after imidazole gradient application, and represent fractions corresponding to the labelled peak in A.

The position of TF-Yeal$_{GGDEF}$ is indicated (of expected molecular weight 69 KDa).

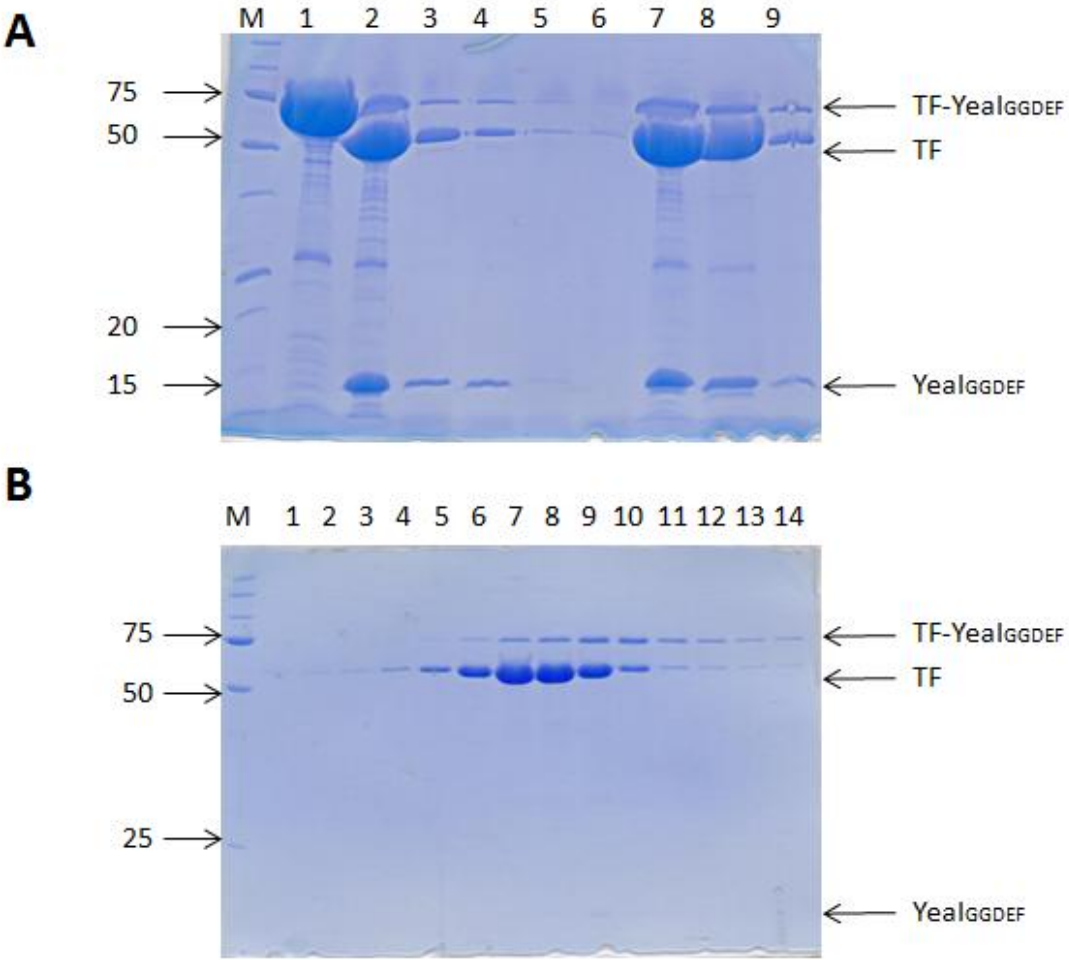


Figure 3.22: Thrombin cleavage as a secondary step for $Yeal_{GGDEF}$ purification from TF-$Yeal_{GGDEF}$

Coomassie-stained SDS-PAGE gels for large scale cleavage and purification of $Yeal_{GGDEF}$ from the TF-$Yeal_{GGDEF}$ fusion protein.

Thrombin was used to cleave the TF-$Yeal_{GGDEF}$ fusion protein into the two components. Isolating $Yeal_{GGDEF}$ was then attempted by applying the cleaved mixture to a HiTrap chelating column and eluting bound protein with imidazole buffer in (A) a single step, (B) an elution gradient by running the His-chelating programme.

In both cases, M: Precision plus protein standards marker (molecular weights shown in KDa).

A) Lane 1: Soluble TF-$Yeal_{GGDEF}$ following initial HiTrap, 2: Post thrombin cleavage, 3: Flow-through following application of mixture to second HiTrap column, 4-6 Wash steps, 7-10: elution fractions following imidazole addition to the column.
B) Lane 1-14: Elution fractions following application of a linear imidazole gradient (0- 0.5 M), using the His-chelating programme on the AKTA.

The fusion protein (expected MW: 69 KDa), TF tag (expected MW: 53 KDa) and $Yeal_{GGDEF}$ (expected MW: 16 KDa) components are indicated by arrows.

that the protocol would be improved on a larger scale, with a much increased accessibility for TF to bind to the resin in the 1 ml HiTrap column than in the small volume of beads used in optimisation. However, as was the case for the optimisation assays, this did not occur, with both the TF tag and $YeaI_{GGDEF}$ present at a low level in the wash steps following addition to the column (Figure 3.22- lane 3) and at a greater level in the imidazole elution fractions (Figure 3.22A- lanes 7-9).

Endeavouring to solve the problem of co-elution, the cleaved protein mix was also applied to a HiTrap column and the His-chelating programme run on the AKTA (following manufacturer's instructions) (Section 2.6.3). This provided a linear imidazole gradient of 0-0.5 M to the column, enabling sequential elution of proteins dependant on the protein's affinity. Elution fractions were analysed by SDS-PAGE (Section 2.6.9) but again failed to produce isolate $YeaI_{GGDEF}$, with very faint quantities of $YeaI_{GGDEF}$ visible in lanes 8-10, in the same elution fractions as the TF fragment (Figure 3.22B).

Clearly, a vast improvement in the overproduction and solubility of $YeaI_{GGDEF}$ was achieved upon fusion with the TF chaperone, compared to the previous overproduction systems used. Whilst this is significant progress, there is still difficulty isolating the $YeaI_{GGDEF}$ protein. The consistent co-elution of $YeaI_{GGDEF}$ and TF suggests that an interaction is occurring between them. This interaction hypothesis would explain why $YeaI_{GGDEF}$ was present in imidazole elution fractions, with the TF bound being immobilized to the HiTrap column, and therefore retaining $YeaI_{GGDEF}$ in the column until imidazole addition. Futhermore, a TF:$YeaI_{GGDEF}$ interaction may cause the His_6 tag to be buried within the protein, explaining why the affinity of TF for the HiTrap column was weaker than expected, present in pre-imidazole wash steps rather than being entirely immobilised (Figures 3.18, 3.19, 3.20 and 3.22).

In some cases, a tag is not of great concern, such as a small His_6 tag which would be unlikely to cause much disruption to the native structure and function of the protein. However, the TF tag is more than three times the size of the protein of interest $YeaI_{GGDEF}$, and therefore may be having a dramatic effect on the protein characteristics. It was therefore essential to remove the TF tag. Given more time, gel filtration or Vivaspin size exclusion columns could be used to solve this problem, theoretically separating the TF and $YeaI_{GGDEF}$ proteins on size. Equally, the pCOLD TF vector contains three cleavage sites

between the TF and the cloned $YeaI_{GGDEF}$, enabling the efficiency of other enzymes to be investigated. However, if an interaction between TF and $YeaI_{GGDEF}$ has been formed, these methods may not alleviate the problem, potentially requiring denaturing conditions such as guanidine hydrochloride addition to unfold the proteins. Assuming this treatment disrupted the TF:$YeaI_{GGDEF}$ interaction, $YeaI_{GGDEF}$ could be isolated by gel filtration but would subsequently require refolding prior to use in assays.

Searching the literature, it appears that the inability to separate TF with the fused protein of interest may be a common phenomenon. The Zhou *et al.* (2014) group reported successful purification of the recombinant protein pCOLD TF-SsSAD (*Sapium sebiferum* roxb stearoyl-acyl carrier protein desaturase). Whilst this group do not report unsuccessful cleavage of the TF from the SsSAD, it is interesting that this group used only the recombinant protein for further work (Zhou *et al.,* 2014). Another group used the same pCOLD TF vector to overproduce silicatein protein (Schröder *et al.,* 2012). In this case, the purified fusion protein was cleaved by thrombin into the TF and the pro-silicatein protein, which was subsequently cleaved from pro-silicatein into the mature silicatein. Therefore, whilst silicatein was successfully isolated from the TF protein, there was no report that the N-terminal pro-peptide region (present in the pro-silicatein but not mature silicatein protein) was itself separated from the TF (Schröder *et al.,* 2012).

The TF protein is a protein folding chaperone, existing at the surface of ribosomes or in the cytosol. Whilst the mechanism of action is well-characterised for ribosomally-associated TF, the activity of cytosolic TF is less known. The crystal structure of an interaction between TF and a ribosomal protein S7 (from *Thermotoga maritima*) has been solved and shows two S7 molecules encapsulated in a cage of two TF molecules. The two TF proteins in this structure were aligned and formed a hydrophilic cage, suggested to have a cellular role to promote folding of the S7 protein. Whilst this encapsulation has only been detected for T7 in complex with S7, this interaction is predicted to be non-specific due to a very large interaction interface between TF and S7 and S7 being poorly packed in the TF cage (Martinez-Hackert & Hendrickson, 2009).

Therefore, the hypothesis of an interaction between TF and $YeaI_{GGDEF}$ appears a likely explanation for the poor cleavage (Figure 3.22), with $YeaI_{GGDEF}$ feasibly being trapped inside TF.

3.10 Discussion

Work described in this chapter aimed to characterise the *E. coli* protein YeaI, and has yielded a system to improve the solubility of the protein and thereby aid future work.

It was previously suspected that the $YeaI_{GGDEF}$ was fairly insoluble, with low yields reported using the plasmid pGS2263, producing an N-terminal TrxA-His_6-S tag-$YeaI_{GGDEF}$ fusion protein. We therefore changed the location and nature of the fused tag, to determine whether this improved protein solubility. Successful cloning of $YeaI_{GGDEF}$ into a pET21a vector yielded a minimal fusion protein of $YeaI_{GGDEF}$-His_6, with a C-terminal tag (Figure 3.2). Both plasmids (pGS2263 and pGS2341) were used to transform electrically competent *E. coli* BL21 (λDE3) cells and overproduction and solubility optimisation trials were carried out. In both cases, successful overproduction of the fusion proteins occurred. However upon fractionation, both strains yielded extremely low (<5%) soluble protein, with varied induction times, temperatures and IPTG concentrations having little effect on solubility. The optimal conditions for solubility were chosen, and the protein purified by a one-step affinity chromatography method, which produced low but sufficient levels of TrxA-His_6-S tag-$YeaI_{GGDEF}$ protein for use in a limited range of assays.

Contradictory views on the enzymatic activity of YeaI have been reported in the literature. The GGDEF motif is known to deviate in YeaI, which in itself indicates a possible lack of catalytic function. However there have been opposing views on YeaI's enzymatic activity, with Sommerfeldt *et al.* (2009) finding the protein to be catalytically inactive, whilst Sanchez-Torres *et al.* (2011) deduced that YeaI is enzymatically active. Nucleotide reactivity assays were carried out in this work, and identified that $YeaI_{GGDEF}$ did not react with GTP (the substrate of DGC domains) or various other nucleotides (Figure 3.9). This identified $YeaI_{GGDEF}$ as an inactive DGC and supported the findings of Sommerfeldt *et al.* (2009).

Resulting from the assignment of YeaI as a degenerate DGC, the ability to interact with nucleotides and the oligomeric state of the protein were investigated.

Despite the lack of reactivity, it was suggested that the protein may have retained the ability to bind to nucleotides. Other degenerate DGC proteins have been found to bind nucleotides and were functionally regulated by this binding. Whether the same was true

for YeaI was investigated by partial proteolysis, allowing visual identification of any contaminating protein, and thereby reducing the risk of producing false positive results. Minor variations in the digestion profiles were seen, with species intensity altering as nucleotides were added to the reaction mix (Figure 3.11). However at no point did the profile change dramatically, with no polypeptide fragment being lost or emerging in the samples upon nucleotide addition. The major polypeptide cleavage fragment appeared at an increased intensity upon addition of the nucleotides ppGpp, GDP and ATP, which could indicate enhanced exposure of a cleavage site when these nucleotides bind to the protein. Quantification of the proteolysis protein species revealed that the intensity changes of the predominant cleavage fragment (upon ppGpp, GDP and ATP) were not an artefact of sample loading errors, with no change in the levels of intact protein in these samples (Figure 3.12). Of note, GTP addition did not alter the proteolysis profile at all, suggesting no binding of GTP to $YeaI_{GGDEF}$.

Despite three nucleotides causing an apparent change in protein cleavage profiles, interpretation of these results was difficult due to poor protein purity, with no certainty that the contaminating proteins did not contribute to these changes (Figure 3.11). In addition, nucleotide binding is only detectable by this technique if the binding alters the exposure of trypsin cleavage sites in the protein. Therefore, whilst this technique suggests no binding of $YeaI_{GGDEF}$ to GTP, other techniques such as NMR or fluorescence could be used to supplement this data, and confirm this lack of binding. The same techniques could also support or disprove the possible binding to ppGpp, GDP and ATP. Both NMR and fluorescence assays required increased protein purity, thereby increasing the necessity for an improved protein overproduction method.

The DGC protein family form catalytically active dimers, in which an active site is formed in the cleft between two GGDEF units (Chan *et al.,* 2004). Gel filtration and glutaraldehyde assays were carried out to determine the oligomeric structure of $YeaI_{GGDEF}$. In both cases, a potential dimeric species was formed, indicating that the protein may have retained this feature of active DGC proteins (Figures 3.13 and 3.14). However, the low levels of protein purity made interpretation difficult, as the oligomeric species formed is of a molecular weight between that expected of a dimer and a trimer. One hypothesis for the aberrant migration of the protein is from non-linear crosslinking in the glutaraldehyde assays, which is likely to alter the mass/charge ratio of proteins and therefore could explain the

inconclusive results, producing a species between the expected dimeric and trimeric sizes. Thus, it is suggested that the oligomeric species produced is more likely to represent a dimer, with perhaps additional contaminating or co-purified proteins becoming attached to the dimeric GGDEF increasing the molecular weight in gel filtration and with non-linear crosslinking affecting the migration of proteins in glutaraldehyde assays. To increase confidence in these results, purer protein samples would be used for these assays, with a much reduced level of contaminating species, thereby giving the oligomeric state formed by $YeaI_{GGDEF}$ only. Nevertheless, it appears that TrxA-His_6-S tag-$YeaI_{GGDEF}$ is not a monomer.

Following on from these assays, increasing the yield of $YeaI_{GGDEF}$ became a priority, because the amount and poor quality of the protein limited the range of experiments that could be carried out, and impaired our confidence in those we could carry out. At this point, the gene was cloned into the cold shock expression vector pCOLD TF. Successful cloning and overproduction of the $YeaI_{GGDEF}$ protein was confirmed. The major success though, was the hugely increased solubility levels of the fusion protein, with more than 50% of the protein being present in the soluble fraction of the cell extract (Figure 3.17). However purification of this TF-$YeaI_{GGDEF}$ was challenging, with problems in cleaving the TF-$YeaI_{GGDEF}$ protein with thrombin. Once overcome, separating the TF chaperone and $YeaI_{GGDEF}$ was problematic, with both proteins eluting from HiTrap chelating columns at the same imidazole gradient suggesting the occurrence of a possible TF:$YeaI_{GGDEF}$ interaction (Figure 3.22). Whilst the fusion protein TF-$YeaI_{GGDEF}$ could be purified to a high concentration and purity, it was decided not to carry out assays with the fusion protein. This was due to the large size difference between the 53 KDa TF and 16 KDa $YeaI_{GGDEF}$ and the possible interactions between TF and $YeaI_{GGDEF}$. These $YeaI_{GGDEF}$:TF interactions could potentially shield binding sites or cleavage sites of $YeaI_{GGDEF}$, or cause conformational changes to $YeaI_{GGDEF}$, all of which could alter the results from any assays carried out. Future work to purify $YeaI_{GGDEF}$ could investigate the use of detergents, urea or guanidine hydrochloride to disrupt the $YeaI_{GGDEF}$:TF interaction, in an attempt to isolate $YeaI_{GGDEF}$ for use in subsequent assays.

To conclude, this work supports the notion that YeaI is an inactive DGC protein, as reported by Sommerfeldt *et al.* (2009). In addition, $YeaI_{GGDEF}$ exists as an oligomer (probably dimeric species), and there is no evidence of nucleotide binding. The most significant progress has been increasing the solubility of $YeaI_{GGDEF}$, with the TF-$YeaI_{GGDEF}$ protein having a much improved level of solubility than the previously used His-tagged constructs. In addition, a

very high level of purification of the fusion protein was achieved. Once the issue of separating the TF and $YeaI_{GGDEF}$ components has been overcome, the high yield and purity of the $YeaI_{GGDEF}$ will enable a much broader repertoire of assays to be carried out, with increased confidence in all results.

One possible focus for future work could be analysing the contaminating L28 species, seen in the glutaralahyde and partial proteolysis assays, to determine the exact identity of the L28 species (whether it is L28-L9) and whether an interaction between $YeaI_{GGDEF}$ and L28 is produced. If a genuine interaction was formed, studying the function of this interaction would be of paramount importance.

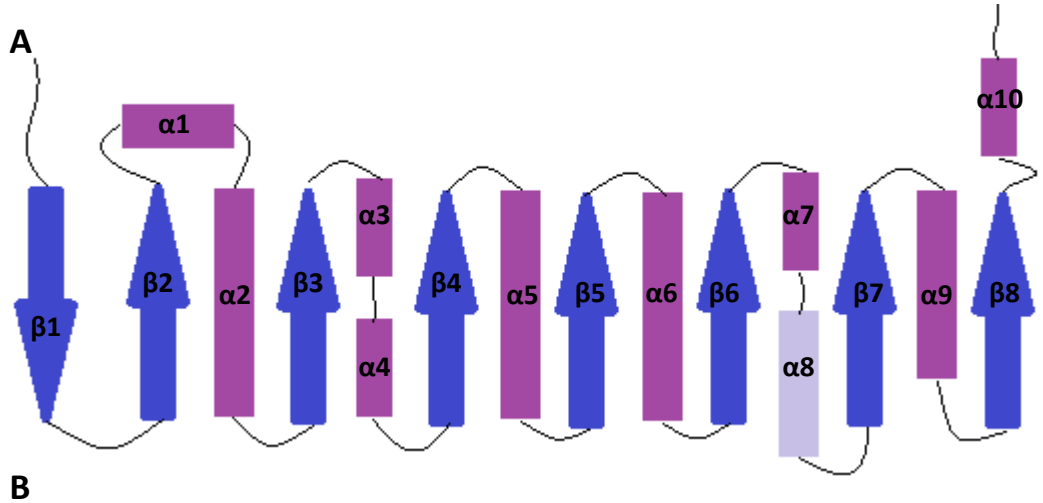
Chapter 4: Characterisation of the *Escherichia coli* 'degenerate' phosphodiesterase protein YdiV

4.1 Introduction

The secondary messenger molecule cyclic di-GMP (c-di-GMP) is controlled by two groups of enzymes in *E. coli*; the diguanylate cyclases and phosphodiesterases, which are responsible for the anabolism and catabolism of c-di-GMP respectively (Simm *et al.,* 2004). The phosphodiesterase (PDE) family of bacterial proteins have a conserved motif in the active site, consisting of the amino acid residues EAL. This site is required for the catalytic activity of these proteins, converting cyclic di-GMP molecules into linear pGpG molecules, prior to the breakdown of pGpG into 2 GMP (Simm *et al.,* 2004; Schmidt *et al.,* 2005).

YdiV is a cytosolic protein, composed of a single PDE domain. However unlike the majority of PDEs, YdiV deviates from the consensus motif EAL, possessing instead an ELI motif and also lacking 7 out of 10 other conserved catalytic residues (Figure 4.1)(Sections 1.6.2 and 1.9) (Tchigvintsev *et al.,* 2010). Despite this lack of conservation, there has been some controversy regarding it's catalytically activity, with Hisert *et al.* (2005) finding YdiV to be catalytically active (and renaming the protein CdgR for c-di-GMP regulator) whilst other groups have found the protein to be inactive (Tchigvintsev *et al.,* 2010; Wada *et al.* 2011).

An X-ray crystallography structure has been obtained to 1.9 Å for the YdiV protein. Whilst the sequence similarities to other EAL domains is low, with <20% amino acid identity, the 3D structure shows the characteristic EAL topology of a TIM-like barrel fold (Figure 4.1). Of note, a groove is present in the YdiV structure, which acts to bind c-di-GMP in other EAL domain proteins (Li *et al.,* 2012). One of the focal points of this investigation was the nucleotide binding potential of the protein. Various other 'degenerate' c-di-GMP signalling proteins have been shown to be regulated by nucleotide interactions. For example, PopA in *Caulobacter crescentus* and PelD in *Pseudomoas aeruginosa* are both degenerate GGDEF proteins whose functions are activated by c-di-GMP binding (Duerig *et al.,* 2009; Lee *et al.,* 2007). Of particular note is FimX from *P. aeruginosa*, which consists of degenerate GGDEF and EAL domains and is responsible for twitching motility. This protein is activated by high-affinity binding of c-di-GMP to its degenerate EAL domain (Navarro *et al.,* 2009). Therefore



B

```
YdiV    KIFLENL--- ----YHSDCY ---------- ---------- ---------- ----------
YahA    KQLFEKLGIQ SDITFWRDIF FQYNPEIISA TGSNSHRYIN DNHYHHIVTP EAISLALENH
YhjH    ESLQERR--- ----FWLQCE RAY------- ---------- ---------- ----------

YdiV    -FLPIRD--- --NQQVLVGV ELITHFSS-E DGTV---RIP TSRVIAQLIE EQHWQLFSEQ
YahA    EFKPWIQPVF CAQTGVLTGC EVLVRWEHPQ TGIIPPDQFI PLAESSGLIV IMTRQLMKQT
YhjH    TWQPIYQ--- --TCGRLMAV ELLTVVTHPL NPSQ---RLP PDRYFTEITV SHRMEVVKEQ

YdiV    LELLKSCQHF FIQHKLFAWL NLTPQVATLL LERDNYAGEL LKYP----FI ELLINENYPH
YahA    ADILMPVKHL LPDN-FHIGI NVSAGCFLAA GFEKECLNLV NKLGNDKIKL VLELTERNPI
YhjH    IDLLAQKADF FIEHGLLASV NIDGPTLIAL RQQPKILRQI ERLP----WL RFELVE---H

YdiV    LNEGKDNRGL LSLSQVYP-L VLGNLGAGNS TMKAVFDGLF TRVMLDKSFI QQQITHRSFE
YahA    PVTPEARAIF DSLHQHNITF ALDDFGTGYA TYRYLQAFPV DFIKIDKSFV QMASVDEISG
YhjH    IRLPKDS-TF ASMCEFGP-L WLDDFGTGMA NFSALSEVRY DYIKIARELF VMLRQSPEGR

YdiV    PFIRAIQAQI SPCCNCIIAG GIDTAEILAQ ITPFDFHALQ GCLWP-AVPI NQITTLVQR-
YahA    HIVDNIVELA RKPGLSIVAE GVETQEQADL MIGKGVHFLQ GYLYSPPVPG NKFISEWVMK
YhjH    TLFSQLLHLM NRYCRGVIVE GVETPEEWRD VQNSPAFAAQ GWFLSRPAPI ETLNTAVLAL
```

Figure 4.1: Primary and secondary structure of YdiV

A) Topology of YdiV, as determined by X-ray crystallography (Li *et al.,*2012) YdiV consists of 10 α-helices and 8 β-strands, forming a modified TIM-like barrel structure consistent with other EAL domain proteins (Li *et al.,* 2012).

B) Protein sequence alignment of EAL domains of YdiV, YahA and YhjH (all from *E. coli*). The sequences were retrieved from Uniprot and the alignment carried out by Mega6 software (Section 2.10.2).

The EAL domain active site region, of consensus sequence EXL is boxed in red, highlighting the degenerate nature of YdiV. Tchigvintev *et al.* (2010) identified 10 conserved residues that were required for the catalytic activity of EAL domains. These 10 residues are boxed in blue, with YahA having all 10 of the conserved residues. YhjH has a conserved EXL motif and retains 9/10 conserved residues. On the other hand, the predicted 'degenerate' protein YdiV both lacks the EXL consensus motif and lacks 7/10 conserved residues (Tchigvintsev *et al.,* 2010).

it was hypothesised that the degenerate EAL protein YdiV, interacts with c-di-GMP or other nucleotides.

When this study began, the role of YdiV was still largely unknown, with phenotypic analysis suggesting that YdiV affected biofilm, sedimentation and motility levels in *E. coli* cells, with *ydiV* overexpression causing a decrease in cell motility (Dr Melissa Lacey, unpublished work). During the project, the role of YdiV was determined, revealing that the protein bound to a transcription factor $FlhD_4C_2$ to inhibit activation of flagella genes (Wada *et al.,* 2012).

Whilst functional roles of YdiV have been elucidated, the mechanism of modulating YdiV activity is not understood, with no known trigger of protein activity. An inducer of *ydiV* gene expression has been reported: SdiA causes the upregulation of *ydiV* in the presence of extracellular autoinducer 1 (AI-1). This is a quorum sensing mechanism in which extracellular AI-1 produced by other microbial species causes the activation of SdiA and subsequent expression of *ydiV* in *E. coli* (Zhou *et al.,* 2008). Equally inhibition of *ydiV* expression is thought to occur under high glucose conditions (Zhou *et al.,* 2008). Part of this investigation therefore included studies of *ydiV* expression, aimed at finding conditions in which the gene was expressed, analysing potential quorum-sensing signalling molecules, and different growth rates of bacteria.

The investigations carried out in this chapter determined the following:

- YdiV is a catalytically inactive EAL domain (PDE) protein (Section 4.4)
- Preliminary evidence of c-di-GMP binding to YdiV (Section 4.5)
- YdiV is a monomer in solution, unable to retain the PDE dimeric form (Section 4.3)
- Upon overproduction of *ydiV*, there was a significant reduction in the number of flagella per cell accompanied by the absence of FliC (Section 4.6)
- Conditions for YdiV overproduction from its own promoter could not be determined, indicating very low levels *in vivo* (Section 4.7)

4.2 Cloning, over-expression and purification of YdiV

4.2.1 Cloning of the *ydiV* gene into an over-expression plasmid

To enable over-expression of the YdiV protein, the *ydiV* gene was ligated into the over-expression plasmid pET21a to create a C-terminal His_6-tag fusion with *ydiV* (Figure 4.2). The *ydiV* gene was amplified from genomic *E. coli* MG1655 DNA by PCR using primers NSW9 and NSW10 (Section 2.3.2). These primers were designed to contain NdeI and XhoI restriction sites at either end of the PCR fragment (Table 4.1). The stop codon of the *ydiV* gene was not included in the reverse primer, in order to allow a C-terminal His_6 tag to be fused to the gene, creating a fusion protein for easy purification.

Table 4.1: Primers used in amplification of the *ydiV* gene for ligation into the plasmid pET21a

Primer	Sequence	Function
NSW9	TTTTCATATGAAGATTTTTTTGGAGA ATCTTTATC	Forward primer for *ydiV* gene
NSW10	TTTTCTCGAGTCGCTGAACCAACGTC GTTAT	Reverse primer for *ydiV* gene (no stop codon in order to form YdiV-His_6 fusion)

The grey shaded sequence indicates an NdeI restriction site (including the start codon ATG) and the boxed sequence shows an XhoI restriction site.

PCR reactions were carried out using high fidelity polymerase, *E. coli* MG1655 genomic DNA and primers NSW9 and NSW10 to amplify the *ydiV* gene (Sections 2.3.1 and 2.3.3). The resultant PCR product was isolated by PCR purification (Section 2.3.10) and digested using the NdeI and XhoI restriction enzymes (Sections 2.3.6). The pET21a plasmid (Novagen) was purified and digested with the same restriction enzymes (NdeI and XhoI) (Sections 2.3.4 and 2.3.6). The *ydiV* fragment and linearised pET21a vector were ligated via their sticky ends and used to transform electrically competent *E. coli* DH5α cells (Sections 2.3.9 and 2.2.7). Transformants were selected on ampicillin supplemented agar (Section 2.2.4). The resultant colonies were screened using colony PCR methods with the appropriate T7 promoter and terminator primers to assess which plasmids contained the correctly sized insert (Table 4.2) (Section 2.3.11). Plasmids containing an appropriately sized insert were checked by DNA sequencing (Source Bioscience) using the same sequencing primers (Table 4.2).

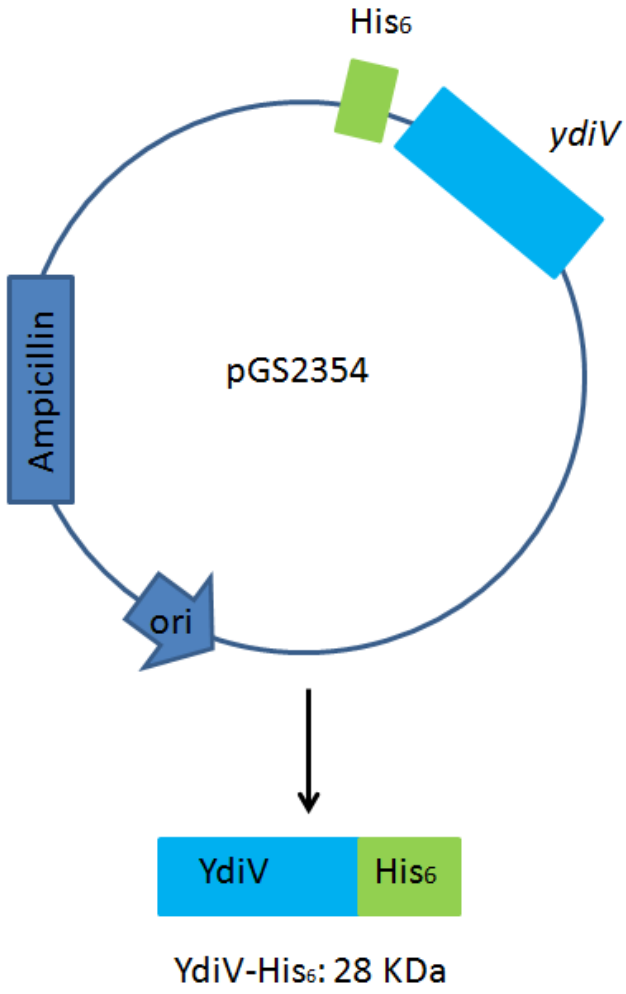


Figure 4.2: Schematic plasmid map for the insertion of *ydiV* into pET21a

Plasmid map for the pGS2354 vector, indicating the position of the His_6-tag, antibiotic resistance cassette and the fusion protein produced.

The origin of replication is indicated and labelled as ori.

The construct encodes a C-terminal His_6 fusion protein, with two amino acid residues linking *ydiV* and the His_6 tag, giving a total molecular weight of 28.1 KDa.

Table 4.2: Sequencing primers used for the plasmid pET21a

Primer	Sequence	Function
T7 promoter primer	TAATACGACTCACTATAGGG	Forward primer for pET21a sequencing
T7 terminator primer	GCTAGTTATTGCTCAGCGG	Reverse primer for pET21a sequencing

Plasmids containing no mutations in the *ydiV* sequence were named pGS2354 (pET21a::*ydiV*). The plasmid was then used to transform electrically competent *E. coli* BL21 λ(DE3) cells (Section 2.2.7).

4.2.2 Optimisation of YdiV overproduction

Trials were carried out in order to determine the optimal conditions for YdiV-His_6 overproduction, altering variables such as IPTG concentration and induction time, using 100 ml cultures and growing the cells at 37^{o}C before induction at 25^{o}C, where the cells remained for subsequent growth. An aliquot (1 ml) of an overnight culture (of strain *E. coli* BL21 λ(DE3)/pGS2354) was added to 100 ml of LB containing ampicillin, and the cells were grown at 37^{o}C with shaking at 250 rpm until an OD_{600} of ~0.6 was achieved. IPTG was added, to a final concentration of either 100 μg/ml or 20 μg/ml and the cultures incubated at 25^{o}C with shaking (Sections 2.2.2, 2.2.3 and 2.6.1). Samples were taken before induction (0 h), and every hour after induction until 5 h was reached. The samples were prepared and analysed by SDS-PAGE, as described in Section 2.6.9 (Figure 4.3).

A polypeptide at ~25 KDa was present in the post-induction fractions, which was believed to correspond to YdiV, although it migrated slightly further than the expected 28.12 KDa molecular mass (including His_6 tag and linker). The protein abundance appeared much greater in the lower IPTG samples (20 μg/ml compared to 100 μg/ml), so the optimal IPTG concentration deemed to be 20 μg/ml. Induction duration also had some impact on protein overproduction level, with a 2 h induction period chosen as the optimal duration for maximum protein levels. Following optimisation of YdiV overproduction, the protein was purified in a one-step affinity chromatography method.

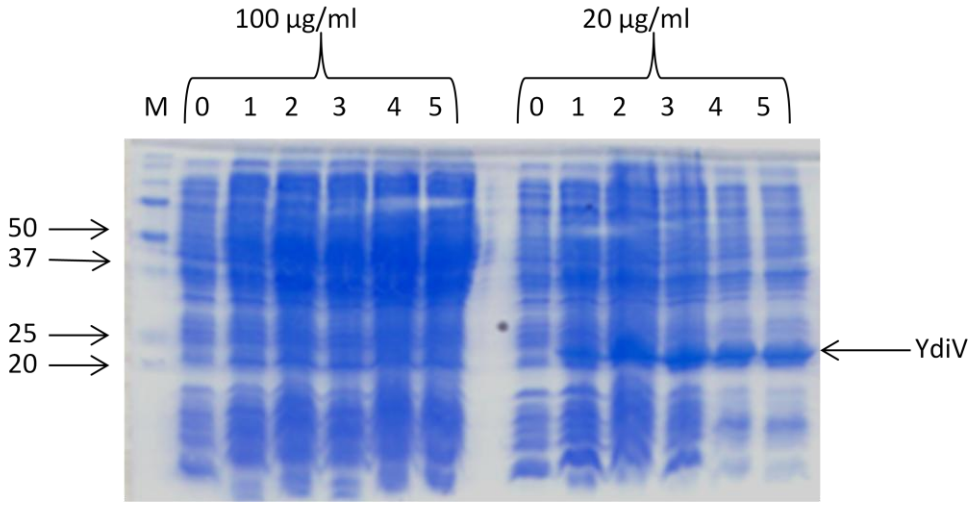


Figure 4.3: Optimisation of overproduction conditions for the YdiV protein in *E. coli* BL21 λ(DE3)/pGS2354 cells

Coomassie-stained SDS-PAGE gel of samples taken pre- and post-induction of *ydiV* expression.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-5 indicate induction times in h, with YdiV production being induced using either a final concentration of 100 µg/ml or 20 µg/ml IPTG, as specified.

The position of YdiV is indicated.

4.2.3 Purification of YdiV–His_6 fusion protein

Purification of His-tagged YdiV was achieved by a single-step Ni-NTA affinity chromatography method. Frozen cell pellets of the YdiV overproduction strain (*E. coli* BL21 λ(DE3)/ pGS2354) from 1-5 litres of culture (Section 4.2.2) were thawed and resuspended in breakage buffer, the cells were then broken by sonication, clarified by centrifugation and the soluble fraction retained for future use (Section 2.6.2). Following separation of the insoluble and soluble cell fractions, the location of YdiV was determined by analysing the insoluble and soluble fractions on an SDS-PAGE gel (Section 2.6.9) and showed the presence of YdiV in the soluble fraction (Figure 4.4). The filtered soluble fraction was applied to a 1ml HiTrap column for affinity chromatography (Section 2.6.3). The His-chelating programme was run on an ATKA, following the manufacturer's instructions, and gave a multi-peak A_{280} trace (Figure 4.5A). The initial peaks in the elution trace are characteristic of the elution of weakly-bound proteins; i.e. those with multiple histidine residues which make some level of interaction with the Ni-NTA column, whilst the larger later eluting peak represented the His_6-tagged protein. The fractions spanning this major peak were separated on an SDS-PAGE gel (Figure 4.5B). From the SDS-PAGE gel it is clear that YdiV was purified to a high level using this single purification step, with lanes 5-8 (fractions 13-16) showing very clean and pure protein, with all major contaminating bands removed. These fractions were therefore used for all further assays and experiments, calculating the protein concentration of YdiV prior to use (Section 2.6.8). The protein migrated further than its expected molecular weight of 28.12 KDa, at ~25 KDa, however N-terminal amino acid sequencing (Dr Arthur Moir, University of Sheffield) gave a protein amino acid sequence of N' MKIF thereby confirmed the identity of the protein to be YdiV (Section 2.6.14).

In some cases, the protein was dialysed into a different buffer in order to remove imidazole, whilst other times the protein was used as prepared, depending on the nature of the assay (Section 2.6.7).

Crystallisation trials were carried out for YdiV in an attempt to elucidate the structure of the protein (Section 2.6.15). YdiV was purified and dialysed into 20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4 buffer (Sections 2.6.3; 2.6.7). The protein was concentrated to 10-15 mg/ml and tested for crystallisation with a variety of commercial screens (specifically pH clear, PEG, PACT, JCSG+, Classics and MPD) (QIAGEN) (Section 2.6.15). No

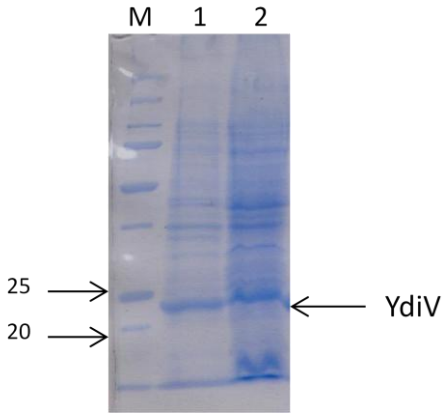


Figure 4.4: Identification of soluble YdiV following overproduction

SDS-PAGE gel of the insoluble and soluble cell fractions following protein overproduction. The cell pellet from cultures of the YdiV overproduction strain was resuspended in breakage buffer, broken by sonication and the debris was pelleted. Samples of the insoluble and soluble fractions were analysed on an SDS-PAGE gel to identify the location of YdiV.

M: Precision Plus protein standards (molecular weights shown in KDa). Lane 1: Insoluble cell sample, 2: Soluble cell sample. The position of YdiV is indicated, showing that some of the protein does remain in the insoluble fraction, but is also present in the soluble fraction.

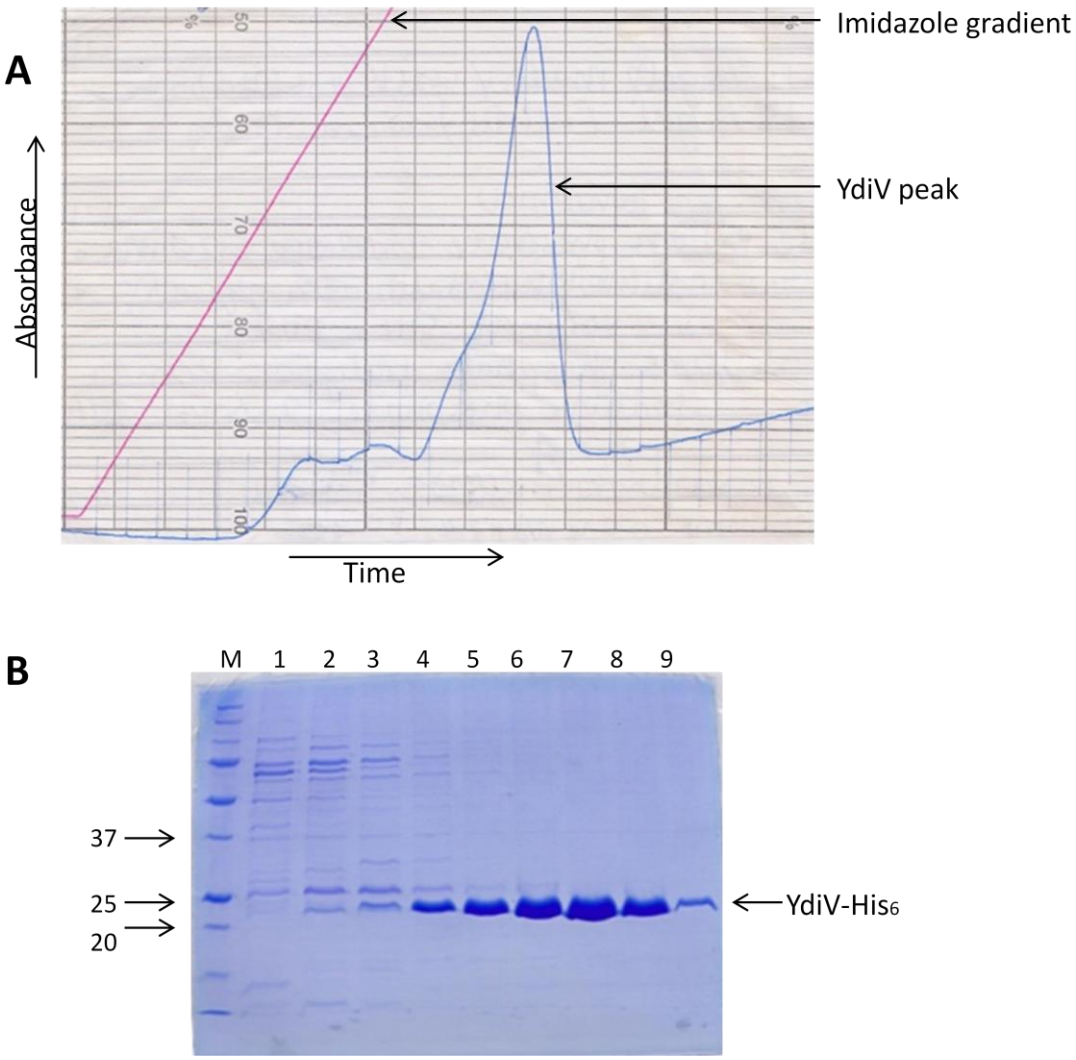


Figure 4.5: HiTrap chelating chromatography as a one-step purification of YdiV-His_6 fusion protein

A) Elution profile of YdiV-His_6 from HiTrap chelating chromatography. Cell-free extract containing YdiV-His_6 was applied to a 1 ml HiTrap chelating column and fractionated by application of an imidazole gradient (linear gradient from 0-0.5 M). The blue trace shows the A_{280} values, indicative of protein levels, with the red trace representing the imidazole gradient. The expected YdiV-His_6 location is indicated.

B) Coomassie-stained SDS-PAGE gel of YdiV-His_6 purification. HiTrap elution fractions spanning the major peak from the Ni-NTA trace. M= Precision plus protein standards (molecular weights indicated in KDa), lanes 1-9 show elution fractions 9-17 collected during the imidazole gradient application, and represent fractions corresponding to the labelled peak in A.

protein crystals were identified using these screens and therefore this was not further investigation.

4.3 Oligomeric analysis of YdiV

To investigate the oligomeric state of the YdiV protein, both glutaraldehyde crosslinking assays and gel filtration techniques were carried out.

Glutaraldehyde acts as a low-specificity crosslinker between amino groups in proteins, enabling oligomeric states of a protein to be fixed and therefore identified. Samples consisting of purified YdiV protein (Section 4.2.3) and various quantities of glutaraldehyde were incubated at room temperature before quenching the reactions with 1 M Tris-HCl pH 6.8 (Section 2.6.12). The samples were then analysed on an SDS-PAGE gel (Section 2.6.9). The SDS-PAGE gels showed the monomeric protein at ~25 KDa (expected size of 28 KDa) in the 0% glutaraldehyde sample, with a weak contaminating species at ~50 KDa (Figure 4.6). Upon addition of 0.01% or 0.02% glutaraldehyde, the pattern of polypeptides was similar to that without glutaraldehyde, with YdiV at 25 KDa and a contaminating species at 50 KDa. The main difference was the presence of protein at higher molecular weights with a possible species at ~100 KDa and smearing at the very top of the PAGE gel. The 100 KDa is most likely to be an artefact, resulting from the non-specific nature of the glutaraldehyde rather than the formation of a tetrameric YdiV species. Therefore, crosslinking assays do not give any indication of YdiV oligomer formation.

Gel filtration was used as a size-exclusion method in which to study the native oligomeric state of YdiV. The Superdex 200 column was calibrated by applying a mixture of four reference proteins of known MW (Ferritin, Aldolase, Ovalbumin and Ribonuclease A) (Section 2.6.13). A plot of the Kav (dependent on elution volume of the protein) and logMW was produced (Figure 4.7A). This plot provided a method to determine the MW of a protein simply from its elution volume from the column.

To analyse the molecular weight of YdiV, a purified sample (3 mg/ml in 20 mM sodium phosphate, 500 mM NaCl, pH 7.5) (Section 4.2.3) was applied to a pre-equilibrated Superdex200 gel filtration column (Section 2.6.3 and 2.6.13). The A_{280} value (indicative of protein quantity) was recorded as the sample passed down the column, and showed a clear peak at an elution volume of 16.76 ml (Figure 4.7B). The minor peaks seen following the

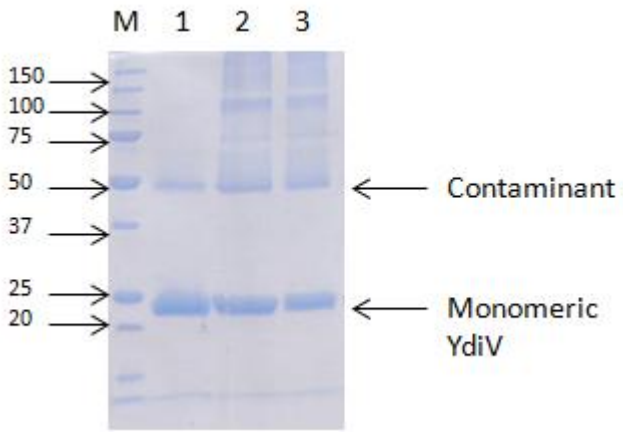


Figure 4.6: Oligomeric analysis of YdiV

SDS-PAGE gel of YdiV treated with gluataraldehyde assays.

M: Precision Plus protein standards (molecular weights indicated in KDa). Lanes 1: YdiV with 0% gluataraldehyde, 2: YdiV with 0.01% glutaraldehyde 3: YdiV with 0.02% glutaraldehyde.

The monomeric YdiV and contaminating species positions are indicated.

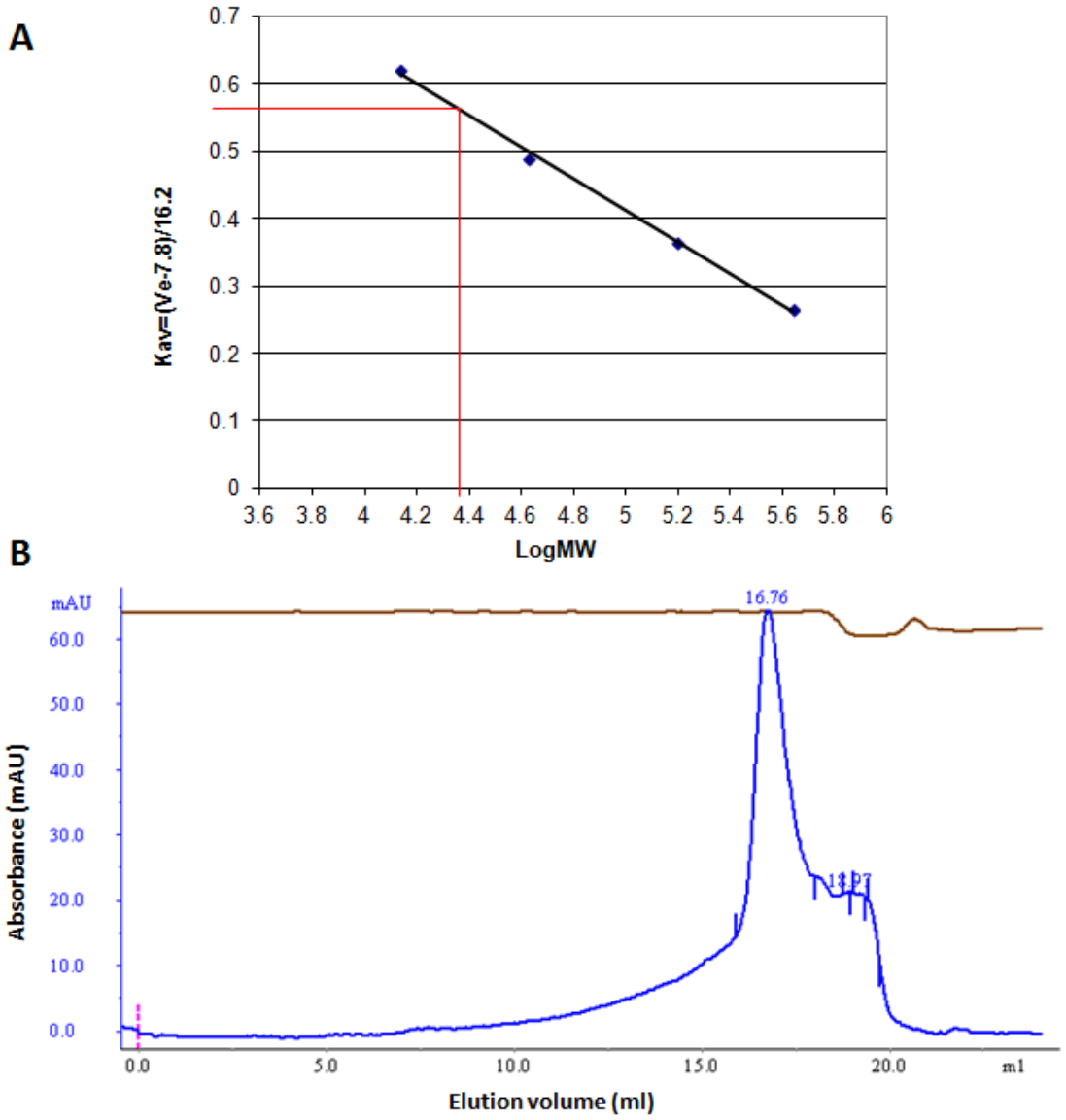


Figure 4.7: Gel filtration analysis of YdiV

A) Calibration plot for Superdex 200 gel filtration column
B) Elution profile of YdiV from gel filtration chromatography

A calibration plot (A) was carried out using a mixture of four proteins of known MW which were applied to the column. A plot of the Kav (dependent on elution volume of the protein) and logMW was produced. This enabled calculation of the MW of YdiV, based on the elution volume detected in B.

Purified YdiV (concentration 3 mg/ml in 20 mM sodium phosphate, 500 mM NaCl pH 7.5) was applied to an equilibrated Superdex200 gel filtration column.

The elution profile (B) shows the absorbance of eluant at 280 nm (blue trace) against elution volume, with the numbers at the crest of each peak specifying the retention volume of each eluted protein. The conductivity applied to the column (brown line) and the point at which the protein was added to the gel filtration column (pink dotted line) are shown.

predominant peak were eluted from the column by a change in the applied conductivity to the column (brown trace) and are smaller MW proteins (Figure 4.7B).

To calculate the MW of YdiV, the elution volume of 16.76 ml was used to determine the proteins Kav (gel phase distribution coefficient) value, using the equation below.

$$\mathrm{Kav} = \frac{\mathrm{Ve} - \mathrm{Vo}}{\mathrm{Vt} - \mathrm{Vo}}$$

In the equation above, Ve = elution volume (ml) of protein, Vo =void volume of column (7.8 ml) and Vt= total volume of column (24 ml). Using this equation, the major peak in the elution profile (Figure 4.6B) gave a Kav value of 0.553, which was then applied to the calibration curve (red line on Figure 4.6A). The corresponding logMW determined by the best-fit line was between 4.35 and 4.40 which when converted to a MW gave a value of 22.38 to 25.12 KDa.

Therefore taken together, cross-linking and gel filtration indicate that YdiV exists as a monomer. Whilst unknown at the time of undertaking this study, the X-ray crystallography structure of YdiV was resolved consisting of two YdiV molecules per asymmetric unit (Li *et al.,* 2012). This was analysed to consist of two YdiV monomers packed tightly together rather than an YdiV dimer. The crystal structure revealed key differences between the two YdiV molecules compared to other dimerised EAL domains. The dimerization helices of the YdiV monomers were much further apart from one another than in other EAL domains and the interaction interface between monomers was much smaller than expected for EAL domain proteins (Li *et al.,* 2012). It was therefore suggested that YdiV could not form stable dimers in solution, which was confirmed by gel-filtration analysis which detected YdiV as a monomeric species (Li *et al.,* 2012). The independent data provided here confirms this.

4.4 Confirmation of the 'degenerate nature' of the YdiV EAL domain

To investigate whether YdiV was an active or inactive PDE, enzymatic assays were carried out for YdiV with c-di-GMP (the substrate for EAL domain proteins) as well as various other nucleotides. The assays consisted of 1 µM protein and 10 mM Mg^{2+} or Mn^{2+} being incubated with various nucleotides at 37^{o}C for 1 h before the reactions were halted by rapid heating to 95^{o}C for 5 min. The samples were then analyzed by HPLC (Section 2.7.1).

Control reactions were carried out using known DGC and PDE proteins, to confirm the ability of the assay to differentiate between nucleotides, as detailed in Section 3.6 (Figure 4.8). After demonstrating the effectiveness of the technique, it was used to study YdiV enzymatic activity, using purified YdiV-His_6 protein (Section 4.2.3). The YdiV traces show identical traces for all pairs of samples, with the black (nucleotide only) and pink (protein and nucleotide) traces overlaying in all cases (Figure 4.9). This indicated that YdiV did not have enzymatic activity for any of the tested nucleotides. Identical traces were produced when using Mn^{2+} or Mg^{2+} (data not shown for Mn^{2+} series). Of particular importance was c-di-GMP, the substrate regularly converted by EAL domains into linear pGpG. This too, had an identical elution trace when c-di-GMP alone was loaded and when the c-di-GMP and YdiV mix was analysed, confirming that c-di-GMP remained unaltered in the sample and was not degraded by YdiV.

The elution times for GTP, c-di-GMP and pGpG were 22.8, 8.6 and 6.8 min respectively. The control reactions were carried out at an earlier date, with elution differences here being accounted for by buffer pH fluctuations and a well-preserved column which had undergone less use.

This experiment therefore confirmed the hypothesis that YdiV is an inactive PDE lacking activity with c-di-GMP or other nucleotides, in agreement with the data from Tchigvintsev *et al.* (2010) and Wada *et al.* (2011), but in contrast to Hisert *et al*. (2005).

4.5 Determination of YdiV 'ligand' binding

After showing that YdiV was an inactive or 'degenerate' PDE the possibility that YdiV may bind a nucleotide, such as c-di-GMP was investigated (Section 2.7). It was known that the EAL catalytic motif was lost in the YdiV protein, however that did not imply that the substrate binding cleft was lost entirely. To investigate this, several assays were used including partial proteolysis and intrinsic fluorescence before a subset of nucleotides, deemed most likely for binding, were analysed by proton NMR.

4.5.1 Partial proteolysis analysis

Partial proteolysis was used as a technique to potentially identify nucleotide binding to

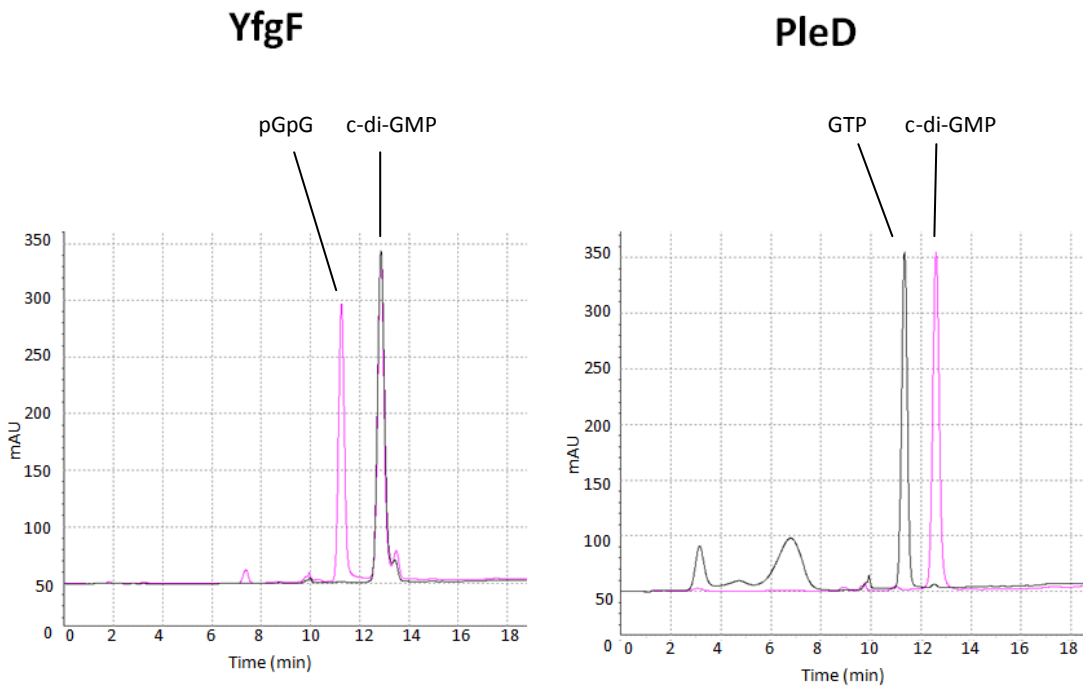


Figure 4.8: Control reactions for determining YdiV PDE activity, using known active DGC and PDE enzymes

HPLC elution traces, with retention time (min) on the x axis and intensity on the y axis.

On the left hand side is the overlayed trace for YfgF, an active PDE. This protein is able to convert c-di-GMP into pGpG, and this was established by comparing a c-di-GMP only trace with a c-di-GMP plus YfgF sample. The black trace shows the nucleotide only sample, corresponding to c-di-GMP. The pink trace is the nucleotide and protein trace, in this case YfgF and c-di-GMP. The trace indicates the absence of c-di-GMP and the presence of a new peak, confirming the catalytic activity of the protein.

On the right hand side, is the overlayed trace for PleD, an active DGC. DGCs are known to convert GTP into c-di-GMP, so this activity was tested by analysing GTP both in the presence and absence of PleD and comparing the elution traces. The pink trace shows nucleotide and protein, whilst the black trace is the nucleotide only. Here the black GTP trace is completely lost and a new profile (pink) trace emerged upon PleD addition, confirming catalytic activity.

These control assays confirmed the suitability of these experiments as a means to differentiate nucleotides.

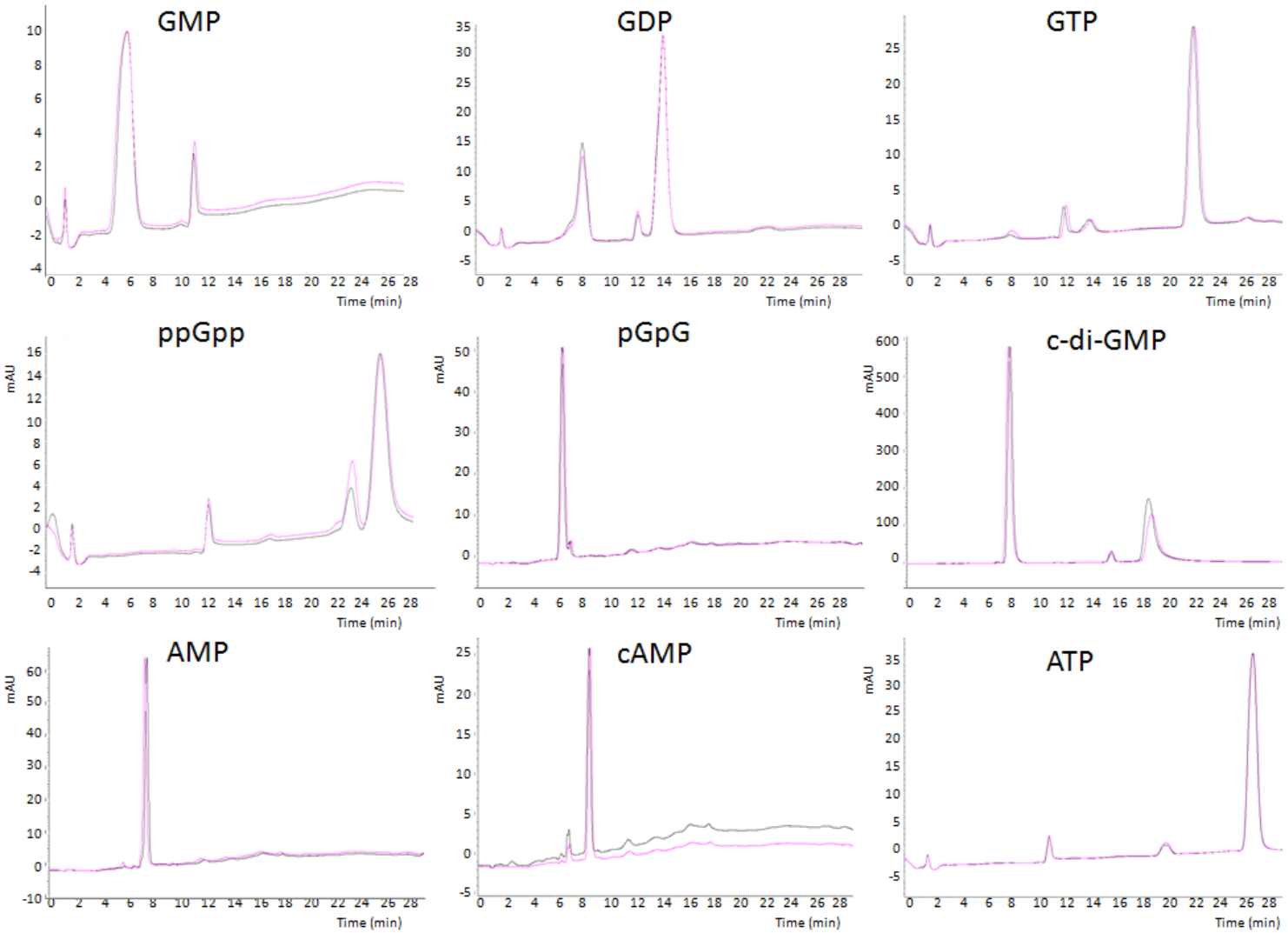


Figure 4.9: Confirmation of YdiV as a 'degenerate' or inactive PDE by HPLC analysis

HPLC elution traces from reactions containing the indicated nucleotide, with and without YdiV addition, in the presence of 10 mM Mg^{2+}. The various nucleotides (as specified) were assayed independently (black traces), as well as in a mixture with the YdiV protein (pink traces).

In all cases, the traces have nucleotide retention time (min) on the x axis, and intensity (at 254nm) on the y axis.

YdiV as indicated by different patterns of trypsin cleavage (Section 3.7). Optimisation of trypsin cleavage was initially achieved using 15 µg purified YdiV (Section 4.2.3), adding trypsin enzyme at different stock concentrations (1 mg/ml or 0.2 mg/ml), for different times (1, 5 or 10 min) and incubating at different temperatures (RT or 37^{o}C). Following the optimisation assays, 1 mg/ml stock trypsin (producing samples containing 100 µg/ml trypsin) and a reaction time of 10 min at RT was chosen for the partial proteolysis assays (Figure 4.10). This condition produced clear protein cleavage, with the intact YdiV being cleaved into three products, whilst some uncleaved protein was still visible.

Partial proteolysis was carried out, as indicated in Section 2.7.2, incubating purified YdiV, trypsin and excess nucleotide (2 mM final concentration) under the optimised cleavage conditions, before stopping the reactions and separating the samples by SDS-PAGE (Figure 4.11).

The resulting SDS-PAGE gels revealed some minor changes upon nucleotide addition (Figure 4.11). Firstly, the no nucleotide control samples of protein and trypsin yielded the same polypeptide profile as observed in the optimisation reactions, with some intact protein remaining and the three breakdown products appearing. Of note, cleavage product 1 actually consisted of two distinct species, a more intense major species which was noted previously, and a smaller product, which had previously been missed. This provided a profile to compare all the other samples to. When c-di-GMP was present, cleavage product 1 (highlighted by a black box) appeared to migrate more slowly and was weaker and less well-defined than the no nucleotide sample, despite the track appearing to be equally loaded compared to the control. This could potentially mean a binding interaction had occurred and was altering the ability of trypsin to cleave YdiV. Equally when ppGpp was present, cleavage product 1 appeared to be more defined, apparently lacking the smaller MW species seen in the no nucleotide sample and instead having one single species at this site. This could be significant, as it could mean that the ppGpp was binding to YdiV and shielding a cleavage site. In the presence of cAMP, a similar situation was seen to c-di-GMP with a weak polypeptide species at cleavage site 1. The remaining nucleotides do not appear to alter the trypsin cleavage pattern in any obvious manner, with very similar fragment patterns seen for all other nucleotides.

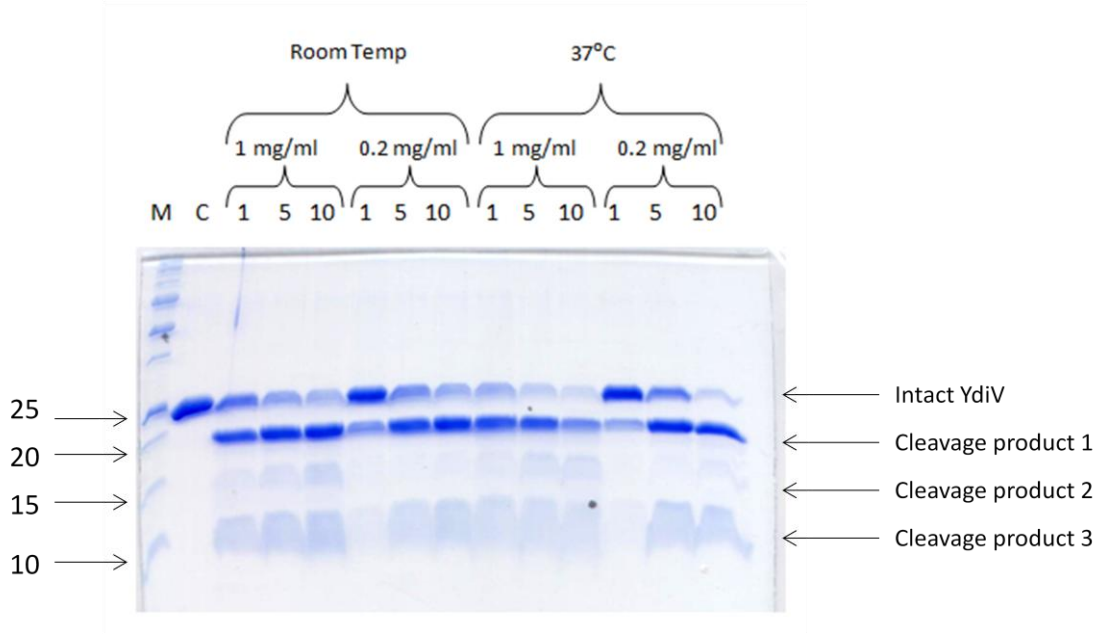


Figure 4.10: Trypsin cleavage optimisation assay for the cleavage of YdiV protein

Coomassie-stained SDS-PAGE gel showing trypsin cleavage optimisation for the YdiV protein.

M: Precision plus protein standards marker, with molecular weights shown in KDa.
C: YdiV (5 µl of 3 mg/ml) with no trypsin addition.
The remaining samples had the same quantity of YdiV (5 µl of 3mg/ml) but with the addition of trypsin at the various concentrations and incubation temperatures and times. The numbers 1, 5, 10 refer to the time (min) over which the incubation occurred prior to stopping the reactions. Trypsin stocks of 1 mg/ml and 0.2 mg/ml were used, adding 3 µl of stock solution to final sample volumes of 30 µl (generating samples containing trypsin at 100 µg/ml and 20 µg/ml respectively). The reaction incubation temperatures were either at room temperature or 37°C, as stated.

Position of the intact YdiV (of expected molecular weight 28 KDa) is indicated, along with the three cleavage products.

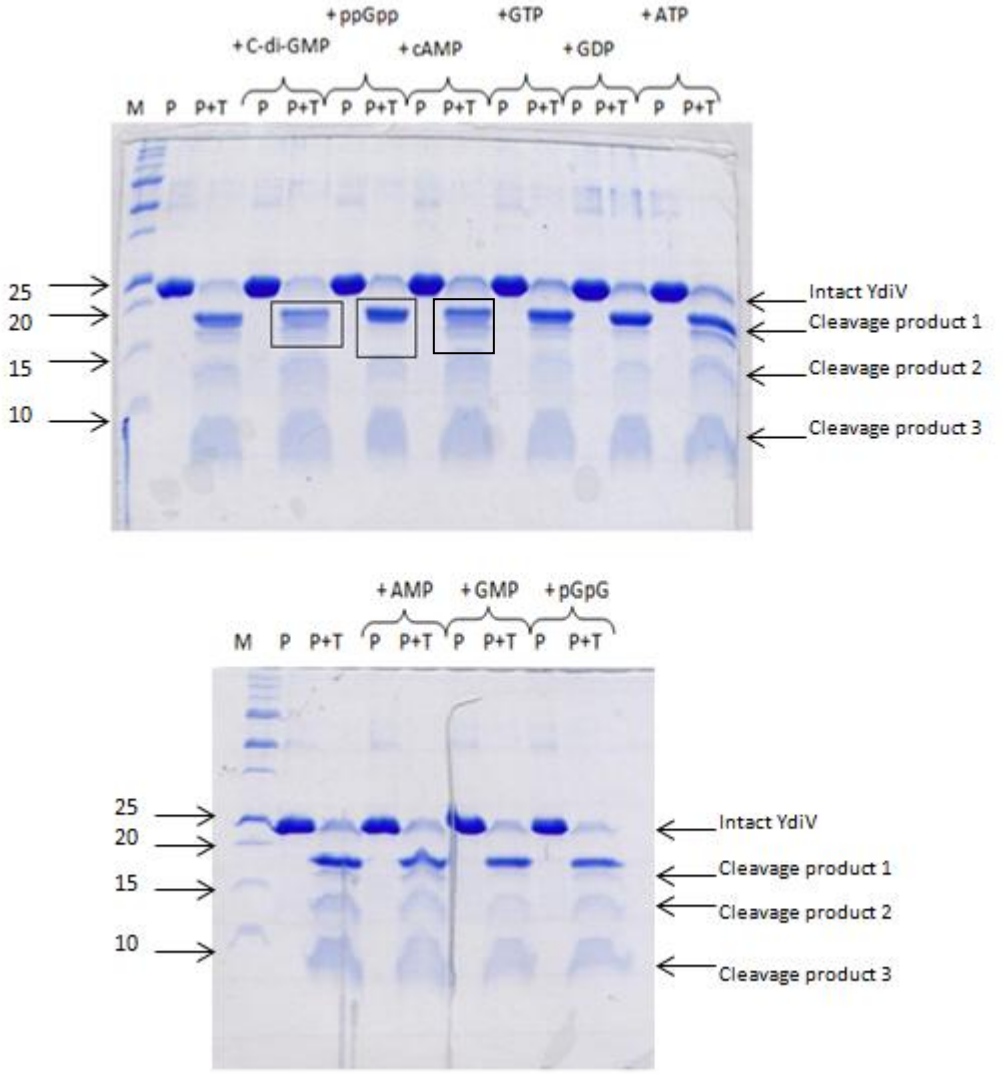


Figure 4.11: Partial proteolysis of YdiV in the presence of nucleotides

Coomassie-stained SDS-PAGE gel showing partial proteolysis assays with YdiV and various nucleotides.

M: Precision plus protein standards marker, with molecular weights shown in KDa. For each lane, P indicates a protein only lane, with YdiV protein (5 µl of 3mg/ml) and no trypsin addition. Lanes labelled P+T indicate the same amount of YdiV with the addition of trypsin, using the 1 mg/ml stock as determined by optimisation trials. Ligands were added to the reaction mix as indicated at 2 mM final concentration, and incubated with the protein and trypsin mix.

Position of the intact YdiV (of expected molecular weight 28 KDa) is indicated, along with the previously identified well-defined cleavage products from optimisation.

Black boxes are used to highlight regions of interest.

The major limitation with this assay is in determining how significant these digest differences are, which is why other experimental techniques have been used. From these assays, it was tentatively suggested that there were reasons to investigate nucleotide binding further, especially YdiV binding to ppGpp and c-di-GMP.

4.5.2 Fluorescence spectroscopy of YdiV

To test the rather ambiguous results obtained from the partial proteolysis assays, intrinsic fluorescence spectroscopy was carried out, to assess the potential binding of nucleotides by YdiV.

Intrinsic tryptophan fluorescence was measured for YdiV before and after nucleotide addition, to determine whether any enhancement or quenching of fluorescence occurred. Intrinsic fluorescence comes from the proteins fluorescent residues; tryptophan, and to a lesser extent phenylalanine and tyrosine residues, of which YdiV has 3W, 6Y and 15F residues. Theoretically, the technique can therefore detect nucleotide binding to a protein, if the binding results in changes in conformation that alter the fluorescent properties of these residues. This method again has limitations, with only 24 residues in YdiV (accounting for 10% of the protein) having significant effect on fluorescence, potentially limiting its ability to report on nucleotide interactions.

Samples were made, consisting of 0.4 µM of purified YdiV (Section 4.2.3) dialysed into Tris-HCl buffer pH 7.4, 30-fold excess Mg^{2+} (final concentration of 12 µM) in a 10 mM Tris-HCl pH 7.4 buffer (Section 2.7.4). Nucleotides were added such that a 1:1 ratio to the protein was achieved, i.e. 0.4 µM.

Reference samples using N-acetyl-L-tryptophanamide (NATA) were used as a tryptophan analogue, adding the various nucleotides to the NATA samples to determine any changes when no interaction was occurring. The level of variation in the NATA samples could be used to assess whether any variation seen in the YdiV samples was real or simply an artefact. For NATA samples, NATA was added to give similar fluorescence intensities as YdiV (at 2.3 µM) and ligands were again added at 0.4 µM, Mg^{2+} at 12 µM in a 10 mM Tris-HCl pH 7.4 buffer.

Samples were made, mixed thoroughly in a 3 ml cuvette. Fluorescence readings were taken at 26^{o}C, with an excitation wavelength of 280 nm, and emission readings recorded between 300-400 nm (Figure 4.12). The most obvious difference in the resultant spectra was that the YdiV emission peak was at ~330 nm, compared to the NATA emission peak at ~360 nm. This is simply due to the contrasting nature of the fluorescence emitters, with NATA having a purely Trp emission, which exists freely in solution, whilst the YdiV emission is due to the protein tryptophan residues, which are therefore influenced by the protein environment.

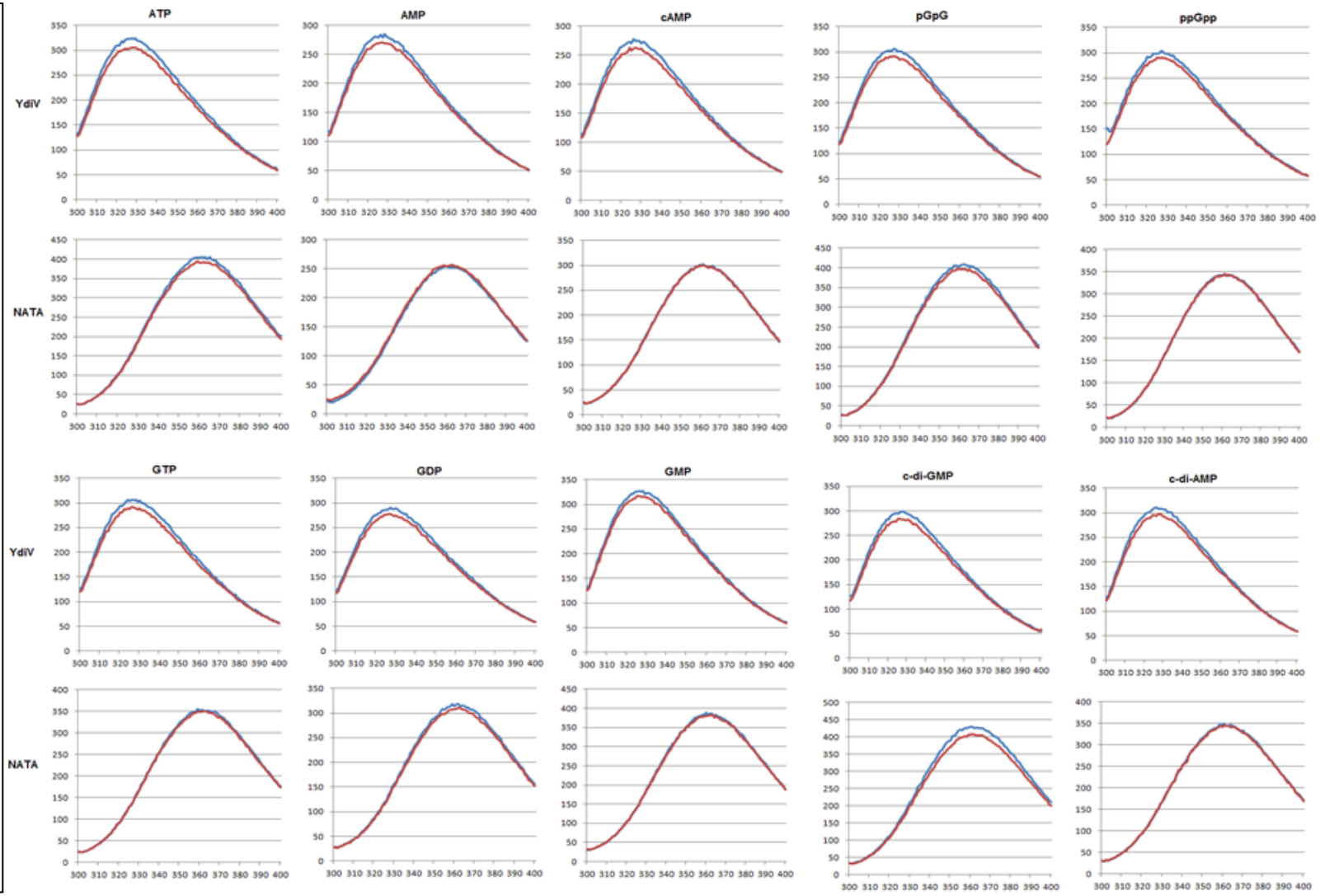
When comparing the YdiV traces in the presence of nucleotides to the NATA and nucleotide traces, a few minor differences were seen (Figure 4.12). In the presence of pGpG and c-di-GMP, the minor variations observed for YdiV with and without these nucleotides was approximately the same as the NATA variation, indicating no apparent binding to YdiV. The remaining nucleotides caused a slight decrease in YdiV fluorescence (red traces compared to the protein only blue traces in Figure 4.12), which in many cases (GTP, GMP, cAMP) was a larger change than the reference NATA samples with and without these nucleotides. However, even the biggest change observed, for YdiV with and without GTP addition, resulted in a less than a 10% decrease in fluorescence intensity, which is a relatively small change. Of note, there was no significant red/blue shift in the wavelength of maximum fluorescence of the samples upon nucleotide addition (Figure 4.12). Any changes in the wavelength maxima were of ~ 1 nm difference, within the same variation range as for the multiple YdiV only samples and therefore irrelevant.

To investigate how reliable these changes were, the c-di-GMP assays were repeated at various constant temperatures (26, 30 and 35^{o}C) using both YdiV and the reference NATA compound. The traces obtained showed quite a significant variation in response to temperature (Figure 4.13). At 26^{o}C and 35^{o}C, both YdiV and NATA underwent a similar fluorescence quenching upon c-di-GMP addition, suggesting that no interaction had taken place between YdiV and c-di-GMP. However at 30^{o}C, the result was different, with the YdiV sample undergoing a significantly larger quench when c-di-GMP was present compared to the NATA sample. Taken on its own, this 30^{o}C assay might suggest that binding was occurring between c-di-GMP and YdiV, however it is unlikely that the binding would be so temperature specific, and therefore this experiment is more likely to show the level of error and variability in this technique. Since the quenching seen at 30^{o}C for YdiV and

Figure 4.12: Fluorescence spectroscopic analysis of YdiV in the presence of various nucleotides at 26°C

For all the graphs, the blue traces indicate the YdiV(0.4 µM) or NATA analogue only, as specified. The red traces indicate the same sample (YdiV or NATA) with the addition of specified nucleotide (0.4 µM).

In all cases the fluorescence intensity (y axis) is shown against wavelength in nm (x axis).



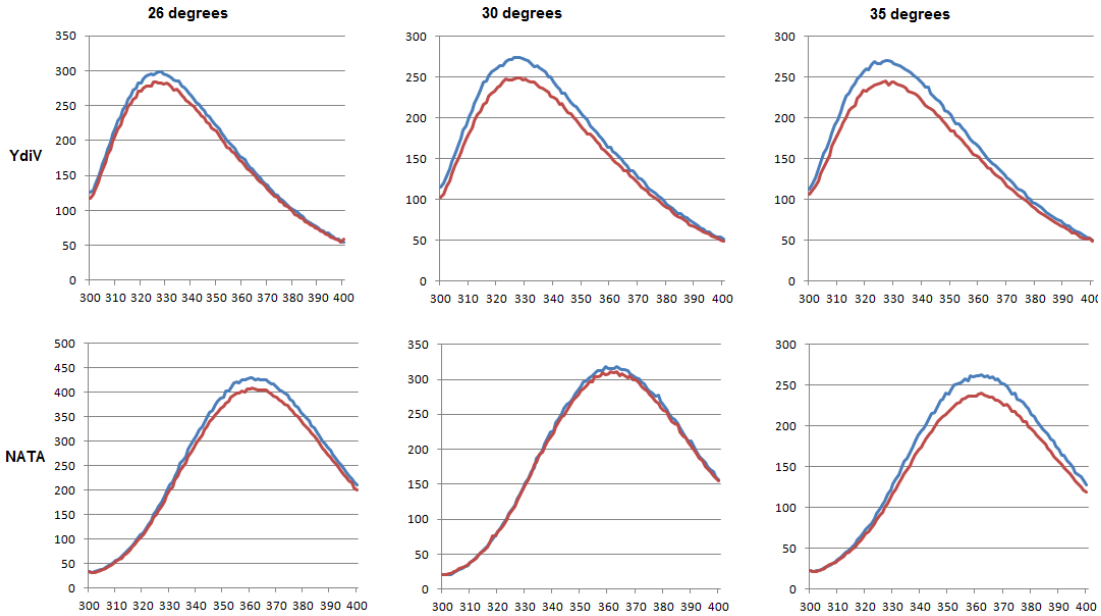


Figure 4.13: Fluorescence spectroscopic analysis of YdiV and c-di-GMP binding at various constant temperatures

The c-di-GMP series of assays were carried out at different temperatures (26, 30 and 35^{o}C), with YdiV and NATA as specified. In all cases, blue traces show YdiV (0.4 µM) or NATA only, with red traces showing YdiV or NATA in the presence of c-di-GMP (0.4 µM)

The emission intensity (y-axis) is shown against emission wavelength (x-axis) for all graphs.

c-di-GMP is a similar decrease (~10%) to several of the YdiV-nucleotide differences in Figure 4.12, this questions the validity of these changes, and the likelihood is that they are simply variations within experimental error.

This technique therefore did not rule out nucleotide binding by YdiV, but equally it did not suggest that ligand binding occurred or provide any information on the identity of the potential ligand.

4.5.3 Nuclear Magnetic Resonance analysis of YdiV in the presence of nucleotides

Following the initial two techniques used to analyse nucleotide binding, proton NMR was carried out for a smaller subset of nucleotides, as a final attempt to reach a conclusion on this matter. Proton NMR studies the environment of H nuclei in a protein or nucleotide with proton nuclei having characteristic chemical shifts depending on the molecular structure and environment in which the proton resides.

The technique of NMR exploits the natural magnetic field associated with all atoms, placing samples into a large magnet and thereby lining up the nuclear spin with the NMR magnet, resulting in net magnetisation in the direction of B_0 (the static magnetic field). The spin is then perturbed by applying rf (radio-frequency) energy to the sample, causing the magnetisation to flip 90^o, no longer aligned with the applied magnetic field. The magnetisation precesses in the magnetic field, inducing an electric current in a detection coil, at a resonance frequency of each nucleus. The excited nuclei decay with time back to the equilibrium position, resulting in an FID (free-induction decay) which is recorded. The recorded signal has a low signal-to-noise ratio, therefore signal averaging is applied, enabling lots of individual decay curves to be added together and thereby increasing this ratio. This is then Fourier transformed into a spectrum showing the chemical shift of the sample, and indicating the environments of all hydrogen nuclei. From these spectra, the environments of simple molecules can be analysed and protein and ligand binding detected (Wüthrich, 1986)

Proton NMR was carried out for YdiV with and without the addition of various nucleotides (Section 2.7.3). A smaller subset of nucleotides was chosen for this analysis, due to the increased workload of this technique, choosing c-di-GMP and ppGpp, which had shown

potential as ligands in the partial proteolysis analysis in Section 4.5.1. Alongside these, pGpG was chosen, simply due to its similarity to ppGpp, and the GXP group chosen (GMP, GDP and GTP) due to GTP and GMP being the enzymatic substrates and breakdown products in the c-di-GMP signalling pathway. Samples were made consisting of 30 µM purified YdiV (Sections 2.6.3 and 2.6.7), 30 µM nucleotide, or 30 µM of both protein and nucleotide, all of which were analysed in a Tris-HCl buffer (20 mM Tris, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM Mg^{2+} and 0.5 mM EDTA (Section 2.7.3) . The samples were then analysed by proton NMR (Dr. Andrea Hounslow, University of Sheffield) and spectra recorded for samples from 1-11 ppm.

Samples of nucleotides individually and the nucleotides in mixtures with the YdiV protein, as well as an YdiV only sample, were analysed, enabling the nucleotide peaks to be compared when YdiV was both absent and present. Additionally the protein (amide) region of the spectra could be compared between an YdiV only sample and the YdiV plus nucleotides sample.

A representative nucleotide chemical shift was found to be in the ~6 ppm range, which was not overlapping with protein spectra and therefore could be compared for each set of nucleotides, tested singularly and when mixed with YdiV. The spectra were stacked in order to enable comparison between each pair of samples (Figure 4.14). Theoretically if a protein and nucleotide were to interact, the nucleotide would make binding contacts with the protein, therefore changing the environment of the hydrogen nuclei in the nucleotide. This would be detected by NMR as a change in linewidth of the peaks as well as a change in the chemical shifts. The MW of a complex is directly related to the relaxation decay rate of the NMR signal, which is detectable as a change in linewidth. Therefore the small MW nucleotides would have a sharp peak, as the rate of decay of the magnetisation is slow, whilst large proteins which have fast NMR decay rate would produce a broad peak. If a nucleotide were to bind to a protein, it would be enveloped in this protein to some extent and therefore take on the MW of the protein, detectable as a broadened peak in the NMR spectra. In addition, if a nucleotide and protein were to bind, the protein would alter the environment of the nucleotide hydrogen nuclei and therefore cause a change in the chemical shift.

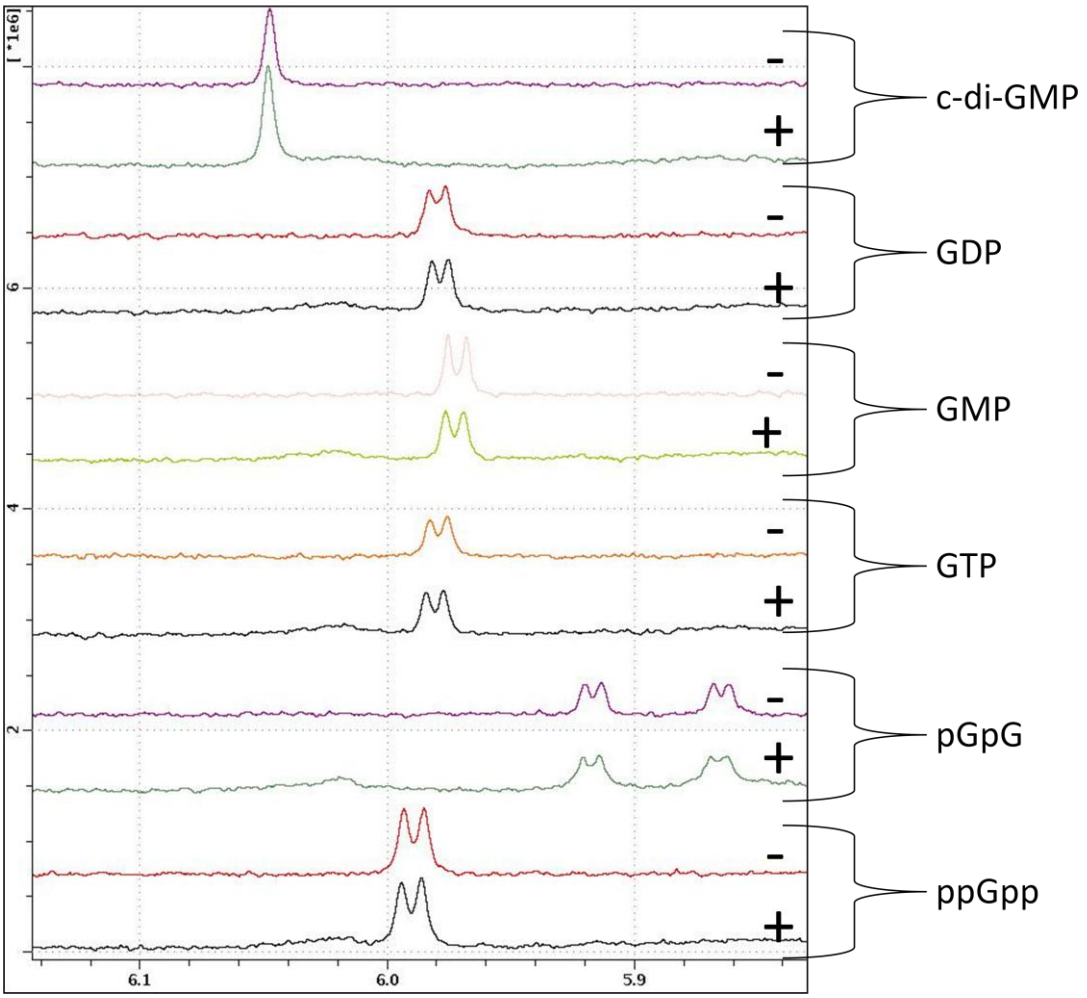


Figure 4.14: Proton NMR spectra focused on the nucleotide chemical shift in the presence and absence of YdiV

Stacked proton NMR spectra, showing the nucleotide chemical shift region of the spectra for each sample.

Samples contained 30 µM nucleotide, or 30 µM of both nucleotide and YdiV in a Tris-HCl buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM Mg^{2+} and 0.5 mM EDTA. In all cases, nucleotides are indicated with +/- referring to presence or absence of YdiV protein.

The spectrum for each sample was recorded between 1-11 ppm chemical shift, and the region shown here between 5.8-6.15 ppm covers the nucleotide-derived portion of the spectra only.

From the spectra obtained, at 5.8-6.15 ppm the stacked pairs of nucleotide and nucleotide plus protein peaks do not seem to change upon protein addition (Figure 4.14). For all six nucleotides the pairs of peaks appear similar with little change in width or in chemical shift. The only minor change is upon pGpG addition when the doublet dips appear smaller than when pGpG alone was present; however this was regarded as a minor change and one that did not unequivocally indicate binding.

Additionally, the protein amide regions of the spectra could be compared, to determine whether the protein spectrum was changing. This amide region of a protein has a characteristic chemical shift of 5-9 ppm; here the focus was on 6.5-9 ppm to avoid the presence of nucleotide derived peaks seen at 5-6 ppm. Theoretically, if a nucleotide were to bind to a protein, this would influence the environment of the protein nuclei, in at least the region of binding and the surrounding nuclei. This change in environment would affect the resonance frequency of the nuclei, detectable by a change in the chemical shift. Whilst it is almost impossible to detect a change in specific nuclei in the amide region, analysis of this region as a whole can be sufficient to detect binding.

Samples of the YdiV protein and the protein mixed with each nucleotide were analysed, stacking the spectra for ease of analysis (Figure 4.15). The complicated series of spectra shown represent each NH group of every one of YdiV's 237 residues. Of note are the two large peaks seen in all samples at 7.3 ppm and 8.2 ppm, which are due to the nuclei in the buffer used to prepare each sample. Upon comparing the stacked spectra, the bottom five spectra (GMP, GDP, GTP, pGpG and ppGpp nucleotide and protein samples) appeared essentially identical, with the only minor change being the appearance of a small peak at ~8.25 ppm for the GMP and YdiV sample. Therefore these five spectra were taken as a baseline for comparison with the other two spectra.

Upon addition of c-di-GMP to the protein, a distinct spectrum was produced, differing in several places to the bottom five spectra. Changes at 8.1, 7.8, 7.4, 7.2 and 6.9 ppm were identified (highlighted by blue boxes), all of which are indicative of a change in environment of YdiV upon c-di-GMP addition (Figure 4.15).

Unexpectedly, the YdiV only spectrum also showed differences compared to the baseline spectra, specifically at 8.0, 7.9, 7.4, 7.2 and 6.9 ppm, indicated by red boxes (Figure 4.15).

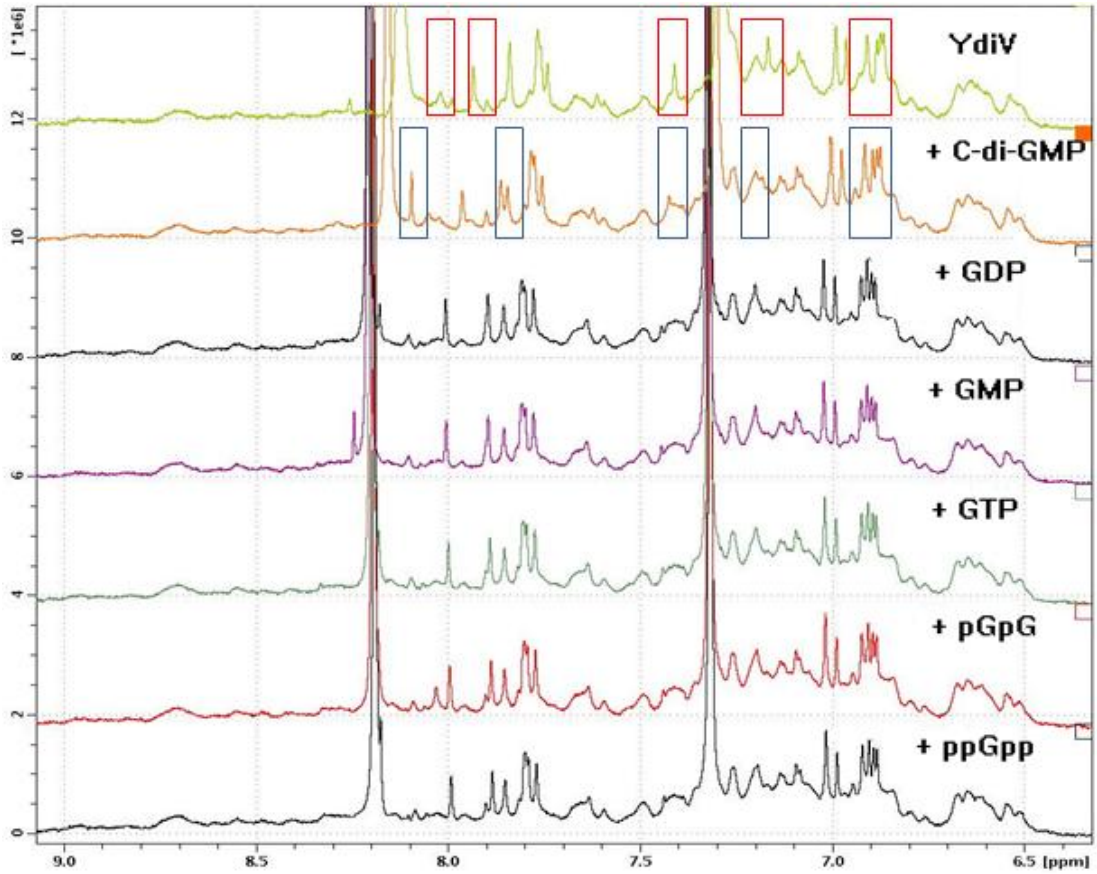


Figure 4.15: Proton NMR spectra in the 6.3-9.1 ppm region, for YdiV in the presence and absence of nucleotides

Stacked proton NMR spectra, showing the chemical shift between 6.3 and 9.1 ppm, with the protein (YdiV) only trace at the top, as labelled, and all other traces being of the protein in the presence of various ligands, again as specified.

Samples contained 30 µM YdiV, or 30 µM of both YdiV and nucleotide (as indicated) in a Tris-HCl buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM Mg^{2+} and 0.5 mM EDTA.

Boxed regions of the traces indicate particular regions of interest, which differ to the baseline (GMP/GDP/GTP/pGpG/ppGpp) traces.

Red boxes highlight areas of interest in YdiV (at 8.0, 7.9, 7.4, 7.2 and 6.9 ppm) whilst blue boxes highlight areas of interest in the YdiV + c-di-GMP sample (notably at 8.1, 7.8, 7.4, 7.2 and 6.9 ppm).

These spectral disparities indicate changes in the protein nuclei environments when the protein was present alone compared to when the protein was in the presence of nucleotides (GMP/GDP/GTP/pGpG/ppGpp). Three of the spectral changes (at locations 6.9, 7.2 and 7.4 ppm) were present in both the YdiV and YdiV + c-di-GMP sample, and therefore are c-di-GMP independent. Crucially though, the YdiV only trace lacks the peaks at 8.1 and 7.8 ppm which were present in the c-di-GMP sample, revealing that these spectral changes were c-di-GMP dependent. The peaks at 8.1 and 7.8 ppm present in the YdiV and c-di-GMP sample refer to changes in the environment of the protein amide backbone residues; however, the identity of these residues cannot be determined by ID proton NMR.

Interpreting the data was complicated by the fact that the YdiV only sample differed from all the remaining nucleotide samples, rather than being identical to some of those traces, which would then confirm the validity of any changing spectra. This therefore raises questions as to whether the observed differences are due to genuine binding interactions or weak artefacts of interactions between the sample components affecting the spectra. To clarify this Clp, a known c-di-GMP binding protein, was isolated, and the NMR spectra after c-di-GMP binding was determined for this system (Chin *et al.,* 2010).

4.5.4 Cloning and over-production of the fusion protein Clp-His_6

The transcriptional regulator Clp from *Xanthomonas campestris* was used as a positive control, as this protein is known to bind c-di-GMP (Chin *et al.,* 2010).

The gene *clp* was amplified from *X. campestris pv. vesicatoria* using primers NSW13 and NSW14 (Table 4.3), as described in Section 2.3.2. These primers were designed to contain an NdeI and an XhoI restriction site at either end of the PCR fragment (Section 2.3.2). The *clp* stop codon was not included in the primer sequence, in order to allow fusion with the C-terminal His_6 tag encoded by the pET21a vector (Figure 4.16).

Table 4.3: Primers used in amplification of *clp* gene for insertion into pET21a vector

Primer	Sequence	Function
NSW13	TTTTCATATGAGCCCAGGAAATAC GAC	Forward primer for *clp* gene
NSW14	TTTTCTCGAGGCGGGTGCCGTACA GCAC	Reverse primer for *clp* gene (no stop codon to enable Clp:His fusion)

The grey shaded sequence indicates an NdeI restriction site (which includes the start codon ATG) and the boxed sequence shows an XhoI restriction site.

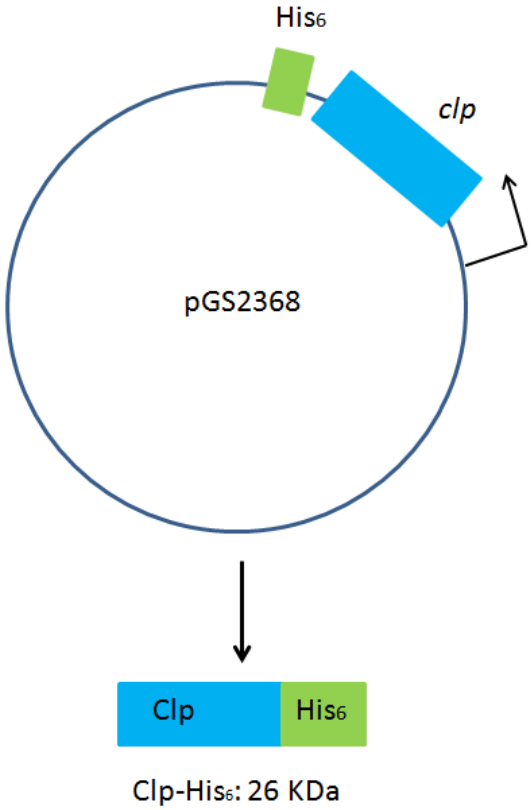


Figure 4.16: Schematic plasmid map for the insertion of *clp* into pET21a

Plasmid map for the pGS2368 constructed vector, indicating position of the His_6 tag and *clp* insertion site, and the fusion protein produced.

The direction of transcription is indicated.

The vector encodes a C-terminal His_6 fusion protein, with two amino acid residues linking *clp* and the His_6 tag, giving a total molecular weight of 26 KDa.

PCR reactions were carried out using high fidelity polymerase, *E. coli* MG1655 genomic DNA and primers NSW13 and NSW14, to amplify the *clp* gene (Section 2.3.3). The PCR product was purified and digested with the restriction enzymes NdeI and XhoI (Sections 2.3.10 and 2.3.6). The pET21a vector was purified and digested with these same restriction enzymes to create a linearised plasmid (Sections 2.3.4 and 2.3.6). The PCR product and plasmid were then ligated and used to transform electrically competent *E. coli* DH5α (Sections 2.3.9 and 2.2.7). Colonies containing the plasmid were selected using ampicillin resistance and then the size of the plasmid insert was analysed by colony PCR and T7 promoter and T7 terminator primers (Table 4.2) (Sections 2.3.11). The sequence of the insert was then confirmed by DNA sequencing using the T7 sequencing primers (Table 4.2) and plasmids containing no mutation were named pGS2368 (pET21a::*clp*). The plasmid was then used to transform electrically competent *E. coli* BL21 (λDE3) cells, ready for protein overproduction (Section 2.2.7).

Overproduction of the protein construct, Clp-His_6 was carried out in *E. coli* BL21 λ(DE3) pGS2368. The optimal overproduction conditions were not investigated, instead the Chin *et al.* protocol in which overproduction occurred for 8 h with 100 μg/ml IPTG inducer was adopted (Chin *et al.,* 2010). Overnight cultures were grown in LB, supplemented with ampicillin, which were then used to inoculate larger 500 ml cultures for overproduction. The cultures were incubated at 37^{o}C until an OD_{600} of ~0.6 was reached, at which point IPTG was added and the cultures incubated for another 8 h. Following this, the cultures were harvested and cell pellets stored at -20^{o}C until required (Sections 2.2.2, 2.2.3 and 2.6.1)

To purify the Clp-His fusion protein, the cell pellets (of strain *E. coli* BL21 λ(DE3)/pGS2368) were re-suspended in breakage buffer, prior to sonication and fractionation of the soluble and insoluble fractions (Section 2.6.2). The soluble fraction was purified by HiTrap His-chelating chromatography on an AKTA Prime protein system (Section 2.6.3). The His-chelating programme was run on the system and the elution profile recorded (Figure 4.17A). The elution profile showed two distinct peaks, of which the first usually indicates the elution of non-specifically bound protein. To confirm the presence of Clp in the second peak, eluted fractions were analysed on an SDS-PAGE gel (Section 2.6.9) and showed purified Clp-His_6 at the expected size of 26 KDa (Figure 4.17B). The purest and most concentrated fractions were then chosen for future use.

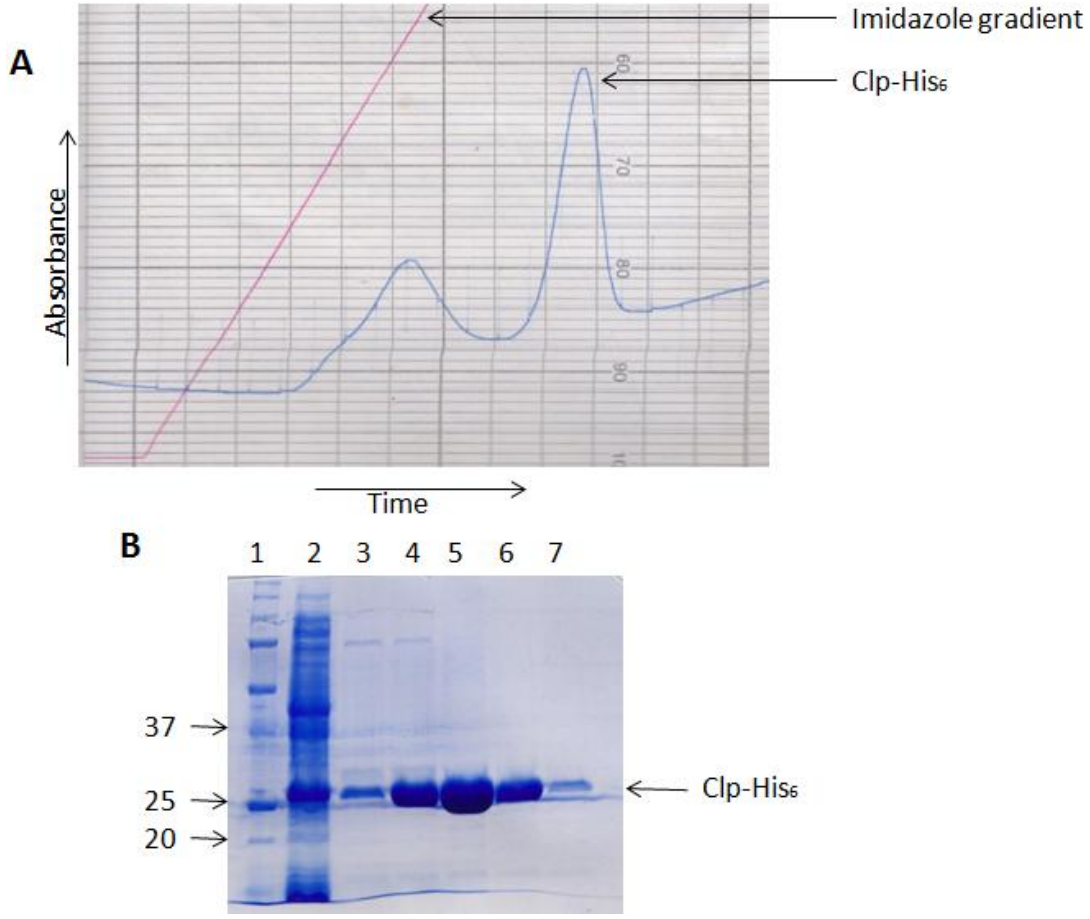


Figure 4.17: HiTrap chelating chromatography as a one-step purification of Clp-His_6 fusion protein

A) Elution profile of Clp-His_6 from HiTrap chelating chromatography.
Cell-free extract containing Clp-His_6 was applied to a 1 ml HiTrap chelating column and fractionated by an imidazole gradient (linear gradient from 0-0.5 M). The blue trace shows the detected A_{280nm} levels, indicative of the presence of protein, with the red trace representing the imidazole gradient. The expected Clp-His_6 location is indicated.

B) Coomassie- stained SDS-PAGE gel of Clp-His_6 purification.
HiTrap elution fractions spanning the second peak from the Ni-NTA trace in (A). M= Precision plus protein standard marker (molecular weights indicated in KDa), 2: soluble crude extract, 3-7: fractions 12-16 eluted from the HiTrap column.

4.5.5 Nuclear Magnetic Resonance analysis of Clp in the presence of c-di-GMP

In order to study the proton NMR spectrum of Clp, the chosen purified fractions (Figure 4.16B) were dialysed into the same Tris-HCl buffer used for the YdiV assays (20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4) (Section 2.6.7). Samples were then prepared of 30 µM Clp, 30 µM c-di-GMP and 30 µM of both Clp and c-di-GMP in the same buffer conditions as previously used to test YdiV samples. The samples were then analysed by NMR, focusing again on the nucleotide peak and the amide regions of the spectra (Figure 4.18 and 4.19 respectively) (Section 2.7.3).

Given the knowledge that Clp bound to c-di-GMP (Chin *et al.,* 2010), it was surprising that the nucleotide peak corresponding to the nuclei in the c-di-GMP molecules did not alter upon protein addition (Figure 4.18). It was expected that the nucleotide peak would broaden or shift if bound to a protein, however that did not occur, with both traces being essentially identical in the presence or absence of protein.

More differences were noted in the protein amide region of the spectra, with some of the protein peaks (identified by green boxes) altering quite significantly in the presence of c-di-GMP, specifically at regions 6.6, 6.75, 7.0, 7.8, 7.9 and 8.1 ppm (Figure 4.19). Whilst this was reassuring, it is also known that a large conformational change occurs upon Clp and c-di-GMP binding, which questions whether any interaction lacking this large conformational change, would be detectable from the NMR spectra using this method. Interestingly, the chemical shifts produced at 7.8 and 8.1 ppm upon c-di-GMP interaction with Clp were similar to those in the YdiV and c-di-GMP sample (Figure 4.15).

The relatively minor changes in the Clp amide spectra upon c-di-GMP binding (Figure 4.19), supports a hypothesis of nucleotide binding in YdiV, given that there was a similar level of chemical shifts observed (Figure 4.15).

Whilst no firm conclusion can be made, it was tentatively proposed that YdiV interacts with c-di-GMP. Both partial proteolysis and proton NMR have revealed changes in the trypsin profile and proton chemical shifts respectively upon addition of c-di-GMP. To support this hypothesis, further experimentation would be necessary, such as analysis by Isothermal Titration Calorimetry.

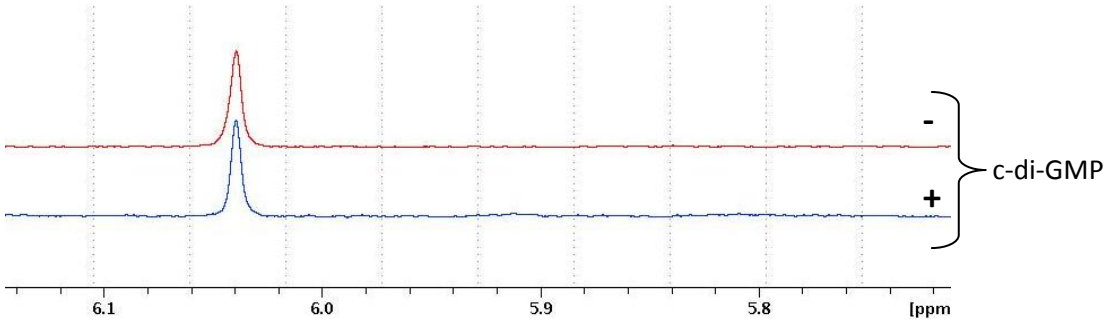


Figure 4.18: Proton NMR spectrum in the 5.71-6.15 ppm region for c-di-GMP, in the presence and absence of Clp

Stacked proton NMR trace, showing the nucleotide chemical shift region of the specta, at 5.7-6.15 (indicated here in ppm).

Samples contained 30 µM c-di-GMP, or 30 µM of both c-di-GMP and Clp in a Tris-HCl buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM Mg^{2+} and 0.5 mM EDTA.

The annotations +/- refers to presence or absence of Clp protein.

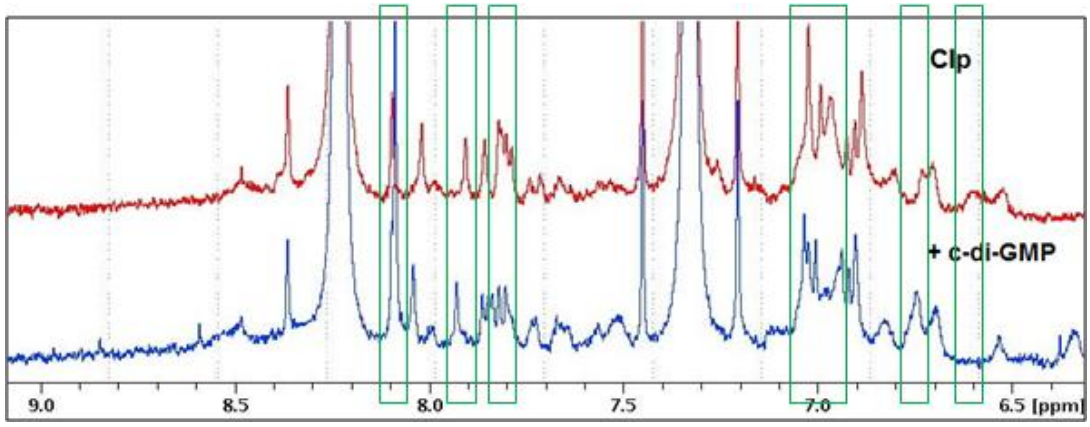


Figure 4.19: Proton NMR spectrum in the 6.3-9.1 ppm region, for Clp in the presence and absence of c-di-GMP

Stacked proton NMR spectra, showing the chemical shift between 6.3-9.1 ppm, for the red (Clp only) and blue (Clp and c-di-GMP) samples.

Samples contained 30 µM Clp, or 30 µM of both c-di-GMP and Clp in a Tris-HCl buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM arginine, pH 7.4) with the addition of 1 mM Mg^{2+} and 0.5 mM EDTA.

Green boxed regions highlight regions of spectral differences between the Clp only and Clp + c-di-GMP samples.

4.6 Characterisation of the effects of *ydiV* over-expression and mutations on bacterial motility

Following Dr Melissa Lacey's work, in which phenotyping studies showed a decrease in motility upon *ydiV* over-expression (unpublished), investigations were carried out, to determine whether YdiV was affecting the flagella levels directly, or whether some other mechanism was acting.

The strains studied were *E. coli* MG1655 (wild-type cells), JRG6058 (MG1655 pBR322), JRG6134 (MG1655 pBR322:*ydiV*), JRG6130 (MG1655Δ*ydiV*) and JRG6578 (MC1000Δ*fliC*) (Table 2.1). This represented a wild-type strain, two positive control strains (vector and *ydiV* mutant), the strain of interest (*ydiV* over-expression in strain JRG6134) and a negative control strain (*fliC* mutant).

To analyse the motility phenotypes of these strains, two approaches were taken, detection of FliC by Western blot analysis and flagella visualisation by EM.

4.6.1 Determination of FliC expression levels according to *ydiV* expression

Since FliC is the basic subunit forming *E. coli* flagellar filament, the levels of this protein were analysed by Western blot assays, probing with a FliC antibody (Kuwajima *et al.,* 1986).

The *E. coli* strains of interest were grown in motility broth as overnight cultures and then sub-cultured until OD_{600} ~0.2 reached (Sections 2.2.1, 2.2.2 and 2.2.3). The cells were harvested by centrifugation at 4000 x*g* for 10 min and the pellets resuspended carefully in dH_2O to reduce flagellar shearing, before samples were mixed with SDS-loading dye (Table 2.8). The samples were then analysed by SDS-PAGE and Western blot as described in Sections 2.6.9 and 2.6.10 respectively. Western blot analysis used a primary antibody of FliC (1/3000 dilution) with the anti-rabbit secondary antibody (diluted at 1/10,000).

The Coomassie-stained SDS-PAGE gel showed equal loading of samples of the various strains indicating comparable cell densities and components (Figure 4.20). The FliC protein, of expected size 51.3 KDa, was not identifiable from the SDS-PAGE gel. The Western blot

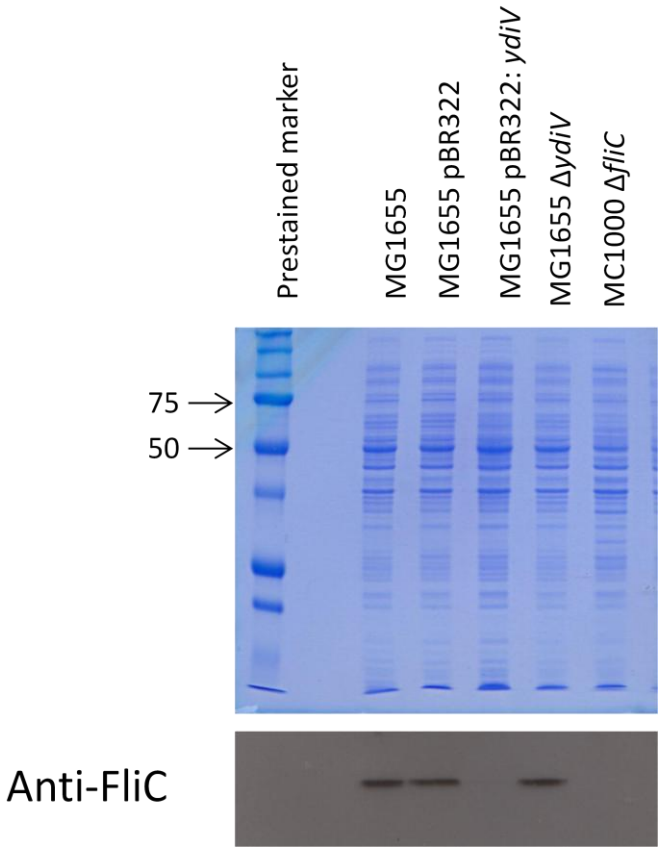


Figure 4.20: Detection of FliC production in various strains, including Δ*ydiV* and *ydiV* over-expression strains

Coomassie-stained SDS-PAGE gel (top) and Western blot probed with FliC antibody (bottom).

The SDS-PAGE gel shows equal loading of the five strains (as specified), which were then probed with the FliC antibody.

The expected size of FliC was 51.3 KDa, and is not clearly visibly from the SDS-PAGE gel, however it can clearly be identified by Western blot analysis.

however clearly showed a species indicating the FliC protein, which appeared to be present in only some of the strains. The wild-type (*E. coli* MG1655), empty vector control (MG1655 pBR322) and YdiV mutant (MG1655 Δ*ydiV*) possessed detectable FliC, at a seemingly similar level. As expected the negative control *fliC* mutant (MC1000 Δ*fliC*) lacked FliC. Of note, the strain of interest; the *ydiV* over-expression strain (MG1655 pBR322:*ydiV*) also lacked FliC in the Western blot, indicating that *ydiV* over-expression in some manner prevents FliC production.

4.6.2 Flagella number per cell changes as a result of *ydiV* over-expression or mutations

To confirm this FliC change upon *ydiV* over-expression, the *E. coli* flagella were visualised by transmission electron microscopy. Strains were once again grown in motility broth (Section 4.6.1) until comparable optical densities were achieved. The cultures were harvested at low speeds (4000xg) and resuspended slowly and very gently in order to reduce flagellar shearing. Once the cultures had been resuspended, the cell samples were fixed in a 50/50 mix of fixing solution, preserving the growth state of cells (Section 2.5.1). When ready to observe the cells, samples were placed on a formvar coated copper grid, stained with phosphotungstic acid and observed by EM as detailed in Section 2.5.1 (Mr Chris Hill, University of Sheffield). Ideally, for future work, the same batch of cell samples would be used for both of the flagellar detection techniques used (Western blots and EM) rather than separate cell cultures which were used in this work.

Representative images were recorded for each strain (Figure 4.21). Upon visualisation, it could be seen that there was a marked change in flagella number for the various strains. There were flagella protruding from almost every cell in the wild-type strain (*E. coli* MG1655), the *ydiV* mutant (MG1655Δ*ydiV*), and although less clear, the vector control strain (MG1655 pBR322). As expected there was a complete lack of flagella in the *fliC* mutant strain (MC1000 Δ*fliC*) but of interest there was a clear decrease in flagella in the *ydiV* over-expressing strain (MG1655 pBR322:*ydiV*). Quantifying the number of flagella present in a sample of 25 cells of each strain (highlighted the lack in flagella in *ydiV* overproduction strains. The samples of 25 cells showed 93 flagella (*E. coli* MG1655), 59 flagella (pBR322), 69 flagella (Δ*ydiV*) and 1 flagellum for the negative control strain (Δ*fliC*). The same sized cell sample of *ydiV* overproduction (pBR322:*ydiV*) cells showed only 9

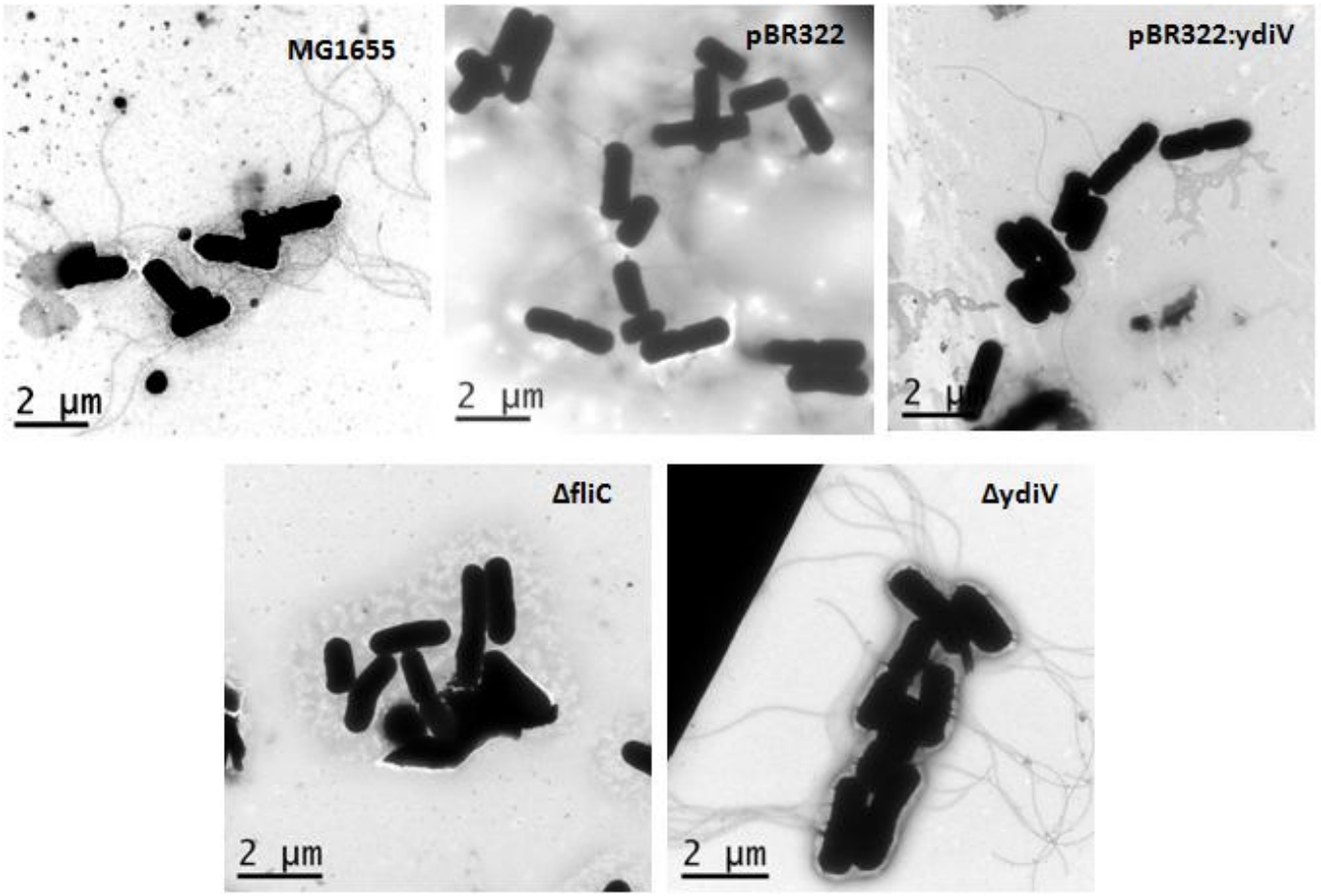


Figure 4.21: Electron Microscopy visualisation of flagella for various strains, including Δ*ydiV* and *ydiV* over-expression strains

Transmission electron microscopy images for each strain, showing a representative average image of the cells for each strain. In all cases, cell samples were placed onto formvar coated copper grids and stained with phosphotungstic acid.

Scale bars are shown for each image.

flagella, noticeably less that the empty vector control or wild-type strains. Therefore, there is an ~80% decrease in flagella numbers in the *ydiV* overproduction strain compared to the vector control or wild-type strains. This supports the theory that elevated levels of YdiV (as seen in the pBR322:*ydiV* strain) inhibit FliC production. As FliC is the basic unit which forms the visible bacterial flagella filament, this results in a lower number of flagella produced.

In hindsight, these motility assays (both Western blot and TEM visualisation) should have been carried out at 30^{o}C rather than 37^{o}C, as flagellar-based motility is inhibited above 37^{o}C (Adler & Templeton, 1967). Whilst clearly FliC was detected in the Western blot assays and flagella are detected in the TEM, abundant numbers are not apparent, which could be resolved by repetition of the assays at a lower temperature.

4.7 Identification of *ydiV* induction conditions

Following the observation that over-expression of *ydiV* caused motility changes, attempts were made to determine under which growth conditions the *ydiV* gene was highly expressed. This was carried out both by studying the protein levels and gene expression levels under various growth conditions.

4.7.1 YdiV protein production analysis

The effect of *E.coli* growth rates on YdiV protein levels was analysed by Western blot analysis. Chemostat samples were provided by Thomas Curran, in which *E. coli* MG1655 cells were cultured in Evans minimal medium (Section 2.2.1) at pH 6.95, 37^{o}C and with 400 rpm stirring. The chemostat cultures were grown at different growth rates, by controlling the medium dilution rate. Bacterial growth rates between 13.9 h (0.05 /h dilution rate) and 1.4 h (0.5 /h dilution rate) were analysed, as described in Section 2.2.2

The chemostat samples analysed came from cultures grown at four different dilution rates (0.05, 0.1, 0.2 and 0.5 /h) each with two biological replicates. These samples were analysed alongside the *E. coli* strains MG1655 (wild-type), JRG6130 (MG1655 Δ*ydiV* acting as a *ydiV* mutant) and JRG6134 (MG1655 pBR322:*ydiV* acting as a YdiV overproduction strain) all grown in LB medium at 37^{o}C (Sections 2.2.1 and 2.2.2) with these and the chemostat samples diluted to give comparable cell densities (Table 2.1). These whole-cell samples

were separated by SDS-PAGE electrophoresis (Section 2.6.9) and analysed by Western blotting (Section 2.6.10) alongside a reference sample of purified YdiV protein at 0.01 mg/ml concentration (Figure 4.22) (Section 4.2.3). Polyclonal antibodies raised against YdiV (Section 5.5), were used in the Western blot analysis, initially at 1/5000 dilution before lowering to 1/10,000 to decrease the non-specific background.

The resultant Western blots indicated that the expression levels of YdiV are below the detection limit (Figure 4.22). The 0.01 mg/ml purified sample (which was too dilute for Coomassie detection), was easily detected by Western blotting. Other than this sample, the only other sample containing detectable levels of YdiV was the YdiV overproduction strain in lane 3 (Figure 4.22). All other samples did not show the presence of YdiV. This indicates a very low level of YdiV was present at all the dilution rates tested, and gave no indication of *ydiV* induction conditions.

4.7.2 Gene expression analysis of *ydiV* in the presence of quorum sensing molecules

The expression levels of *ydiV* were further analysed in the presence of quorum sensing (QS) molecules, using β-galactosidase assays to quantify expression levels (Section 2.4.1).

It has been reported that an up-regulator of *ydiV* expression is SdiA (Suppressor of cell division inhibitor) which is part of the QS system 1. SdiA is a quorum-sensing protein, and is activated by extracellular autoinducer 1 (AI-1) which is presumably produced by other microbial species because *E. coli* does not produce AI-1 (Zhou *et al.,* 2008). The primary autoinducers produced by many gram-negative bacteria are acyl-homoserine lactone (AHL) molecules, which are lipid molecules consisting of a lactone ring and an acyl chain (Watson *et al.,* 2002; Zhou *et al.,* 2008).

Theoretically then, the levels of flagella gene expression can be altered to some extent simply by altering the autoinducer 1 (AHL) levels via a cascade of activation and inhibition (Figure 4.23).

β-Galactosidase assays (Section 2.4.1), using strain JRG6097 (a *ydiV*::*lacZ* fusion strain) were used to attempt to identify conditions in which *ydiV* was induced (Table 2.1). These assays

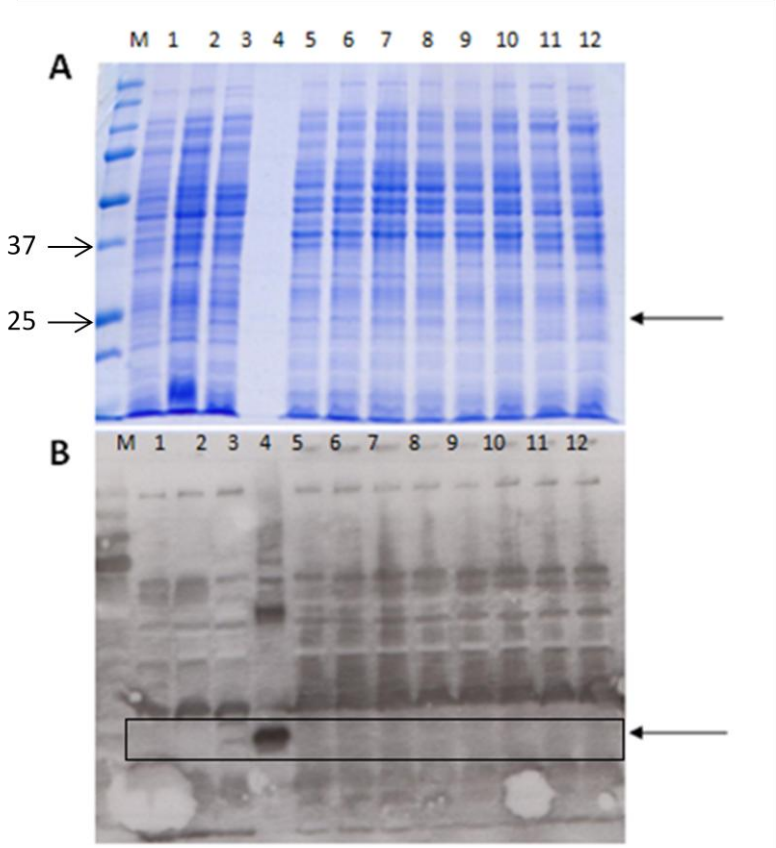


Figure 4.22: Comparison of YdiV protein expression levels at various growth rates

A) Coomassie-stained SDS-PAGE gel, B) Western blot probed with YdiV antibody.

For ease of labelling, dilution rates (/h) are denoted simply by numbers, i.e. 0.05 to 0.5, as explained previously, and the two biological replicates denoted by #1 or #2.

M:BioRad prestained marker, molecular weights indicated in KDa. Lanes 1: *E. coli* MG1655, 2: Δ*ydiV*, 3: pBR322:*ydiV*, 4: 0.01 mg/ml purified YdiV protein, 5: MG1655 0.05 #1, 6: MG1655 0.05 #2, 7: MG1655 0.1 #1, 8: MG1655 0.1 #2, 9: MG1655 0.2 #1, 10: MG1655 0.2 #2, 11: MG1655 0.5 #1, 12: MG1655 0.5 #2.

The position of YdiV is indicated by the arrow, and boxed in B.

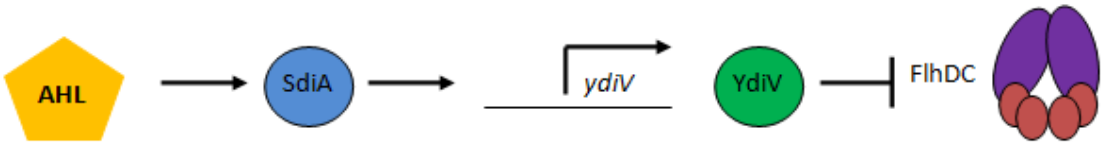


Figure 4.23: Schematic of SdiA induced YdiV expression

Schematic diagram of the series of events that result in YdiV production and the inhibition of the transcription factor FlhDC.

N-acyl-homoserine lactone molecules (AHL) are produced and activate SdiA, which upregulates the transcription of *ydiV*, causing an increased level of YdiV protein in the cell. The YdiV in turn then causes FlhDC to be dissociated from DNA, causing downregulation of flagellar genes.

were carried out in the presence of two types of AHL; *N*-(β-Ketocaproyl)-L-homoserine lactone and *N*-Hexanoyl-DL-homoserine lactone, which for ease are henceforth referred to as K and H respectively.

The *ydiV-lacZ* strain (JRG6097) of *E. coli* was cultured overnight and then sub-cultured at 28^{o}C until mid-log phase was reached (Section 2.2.2). The β-galactosidase assays were prepared, using 250 µl cells and initially 1 µM AHL at the sub-culturing stage, and assays carried out at 28^{o}C under aerobic conditions (Section 2.4.1). The concentration of AHL and stage of its addition were altered throughout the experiments, dependent on previous results.

For each condition, three biological replicates were made, and the assays performed in triplicate. From the raw data, the average MU (Miller unit) was determined, and the standard deviation calculated (shown as error bars on the graph) for each condition.

Initially a final concentration of 1 µM AHL was added at the sub-culture stage and β-galactosidase assays carried out after 3 h of incubation (Figure 4.24A). The MU values for all cultures were low and similar. There was no difference between the samples, with the error bars overlapping for all four conditions. Following this, AHL was added for the duration of cell growth, at a concentration of 1 µM AHL (Figure 4.24B). Again, very low expression of *ydiV* was detected. Despite the AHL samples having a higher MU (Figure 4.24B), there were no significant differences between conditions.

Following a lack of any *ydiV* induction, the concentration of AHL was increased to a concentration of 50 µM to determine if the concentration was simply too low previously. Assays were carried out, this time just adding either K or H to the samples (Figure 4.24C). Once again no difference was seen, providing no indication of how *ydiV* expression is activated.

At this point, it was concluded that *ydiV* was not being induced to a detectable level, even with the addition of high concentrations of AHLs. SlyA is known to be a transcriptional regulator and an inducer of cell motility and potentially an up-regulator of *ydiV* expression (Simms & Mobley, 2008). We therefore transformed the JRG6097 strain with a SlyA overproduction plasmid (pGS2468) (Section 2.2.7) (Table 2.2). The plasmid contained the

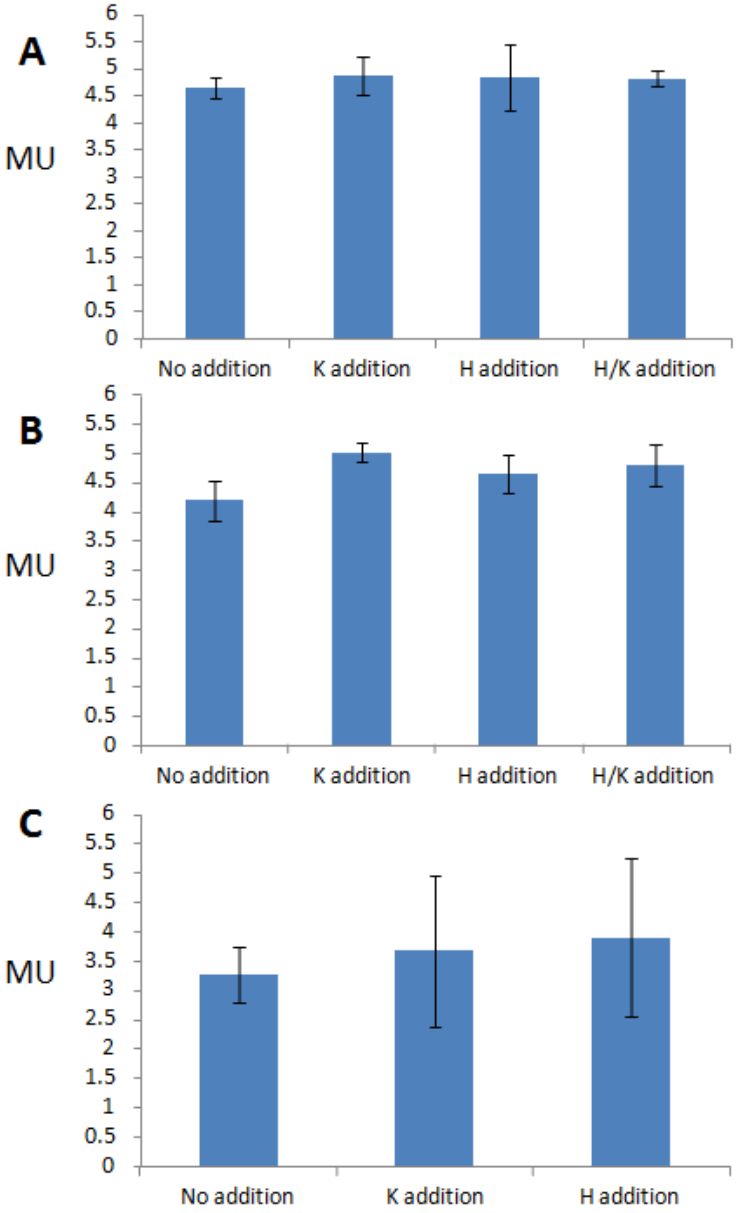


Figure 4.24: β-Galactosidase assay results using the *E. coli ydiV-lacZ* strain

Bar graphs plotted from β-galactosidase assays carried out in the presence of varying concentrations of AHL. In all cases, a negative control 'No addition' culture was used as a baseline, alongside 'K', 'H' and 'K/H' mixes, simply indicating the presence of the K (*N*-(β-Ketocaproyl)-L-homoserine lactone) or H (*N*-Hexanoyl-DL-homoserine lactone) or a mixture of both.

A: A final concentration of 1 µM AHL was added at the sub-culture stage, absent in the overnight stage. B: A final concentration of 1 µM AHL was added for duration of *E. coli ydiV-lacZ* cell growth. C: A final concentration of 50 µM AHL was added for the duration of *E. coli ydiV-lacZ* cell growth.

For A, B and C, the bars show the average MU (Miller Units) for each condition. Each condition was tested using three biological replicates, with assays done in triplicate. Average MU values are plotted, with variation shown as error bars.

slyA gene under the control of its own promoter, requiring no IPTG induction. Further β-galactosidase assays (Section 2.4.1) were carried out using this *E. coli* JRG6097/pGS2468 strain, again adding 50 µM AHL to the cells for the duration of growth (Figure 4.25).

The data (Figure 4.25), showed K addition gave a higher average MU value, however these were not significantly different from the other cultures. Therefore SlyA overproduction did not appear to affect *ydiV* expression.

4.7.3 Promoter alignment

In a final attempt to understand the regulation of *ydiV* transcription, the promoter regions upstream of the *ydiV* gene in various *E. coli* and *Salmonella* strains were analysed (Section 2.10.2). All sequenced strains of *E. coli* and *S. enterica* were analysed, retrieving the sequences from xBASE (http://www.xbase.ac.uk/). The intergenic 247 bases between *ydiV* and the upstream *nlpC* gene from *E. coli* strain K-12 MG1655 was chosen as a sufficient promoter region and the same length upstream sequence was obtained for other strains containing the *ydiV* gene. These DNA sequences were then analysed and aligned using Mega6 (http://www.megasoftware.net) and the ClustalW parameters. In addition, Softberry promoter prediction software (www.softberry.com) was used to detect regulatory promoter motifs in the alignment (Figure 4.26).

Analysing the degree of similarity first, the *E. coli* strains (except B-strains) showed a high degree of sequence identity to one another, with almost completely identical promoter sequences. The two *E. coli* B-strains (B str. REL606 and BL21(DE3)) were noticeably different to the remaining *E. coli* sequences. However, the promoter regions for the *E. coli* B-strains were identical to each other, showing conservation within this subgroup. Finally, six *S. enterica* strains were also analysed, and were almost completely alike, showing promoter conservation within the species. Of importance however, whilst the *S. enterica* strains showed clear differences to the *E. coli* sequences, there were regions of identity. Indeed there was higher sequence similarity detected between the *E. coli* K-12 derived strains and *S. enterica* sequences compared to the *E. coli* K-12 derived strains and *E. coli* B-strains, implying an unexpected evolutionary order.

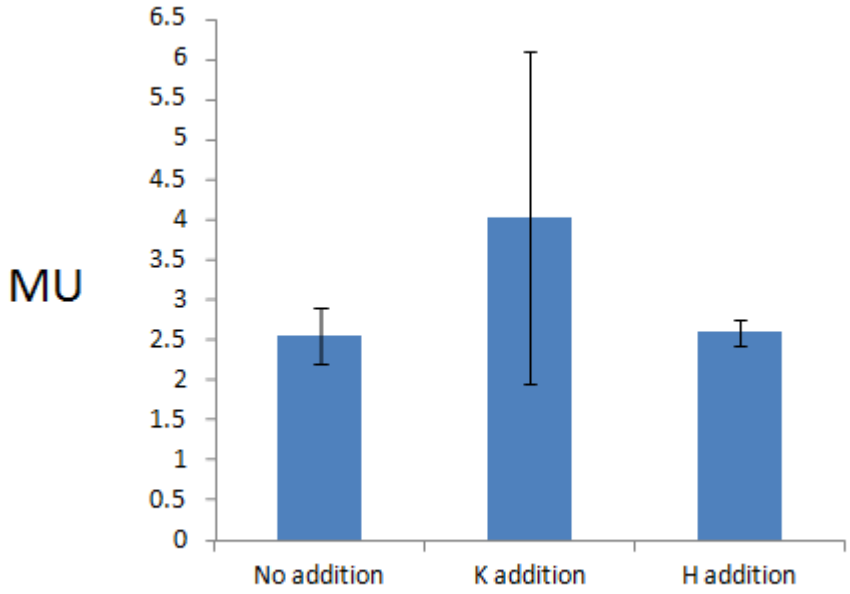


Figure 4.25: β-Galactosidase assay results using the *E. coli ydiV- lacZ*/pET28a:*slyA* strain

Bar graphs plotted from β-galactosidase assays carried out in the presence of 50 µM AHL. Control assays with no AHL are shown as 'No addition' samples, plus *N*-(β-Ketocaproyl)-L-homoserine lactone and *N*-Hexanoyl-DL-homoserine lactone were added, labelled as 'K' and 'H' addition respectively.

Here the AHL addition was carried out at the initial over-night stage, meaning presence of AHL for the duration of cell growth.

Each condition was tested using three biological replicates, and the assays done in triplicate. Average values are shown in Miller Units (MU), with error bars indicating the level of variation in the samples.

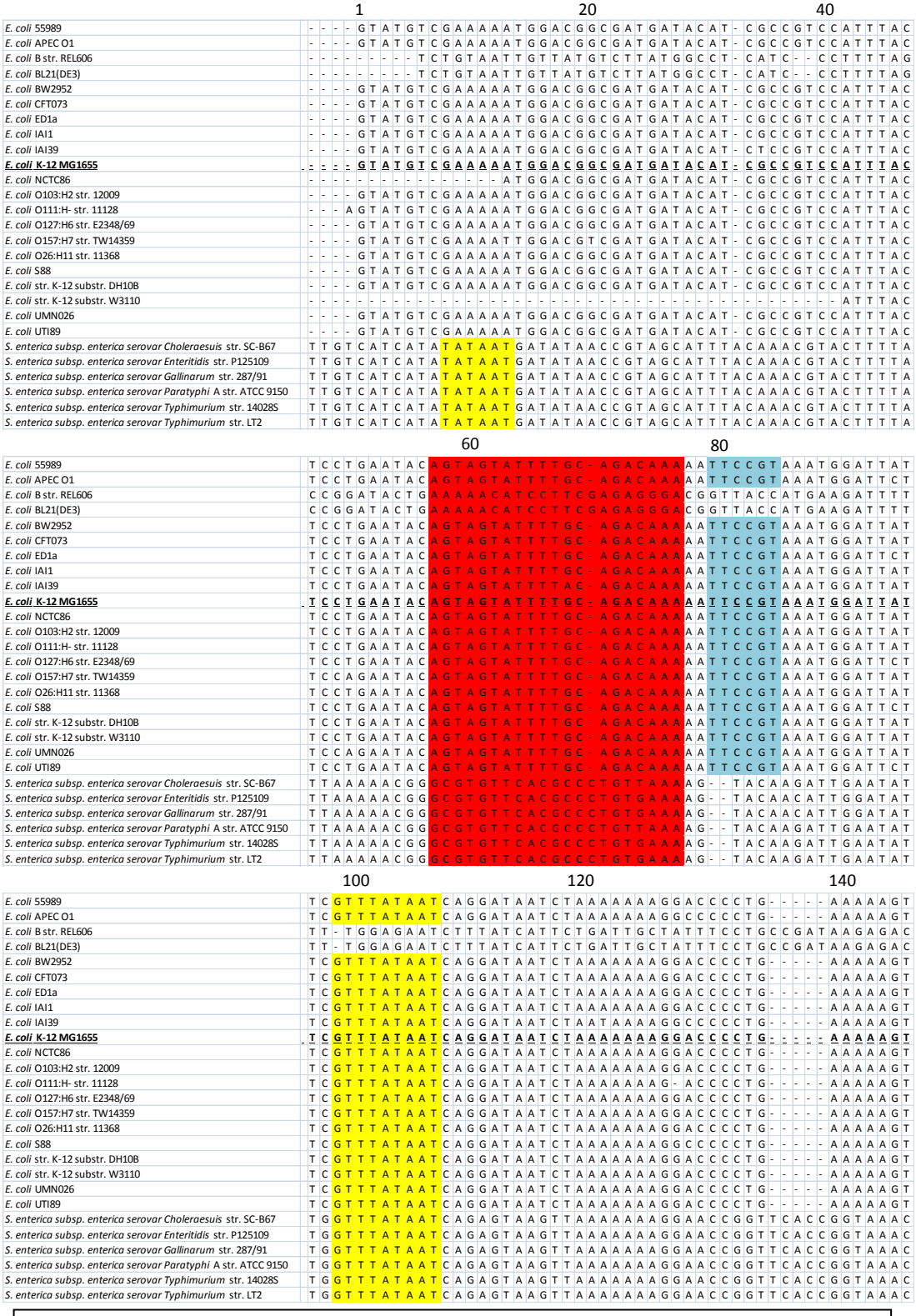


Figure 4.26: Promoter alignments for the *ydiV* upstream sequences in *E. coli* and *S. enterica* strains

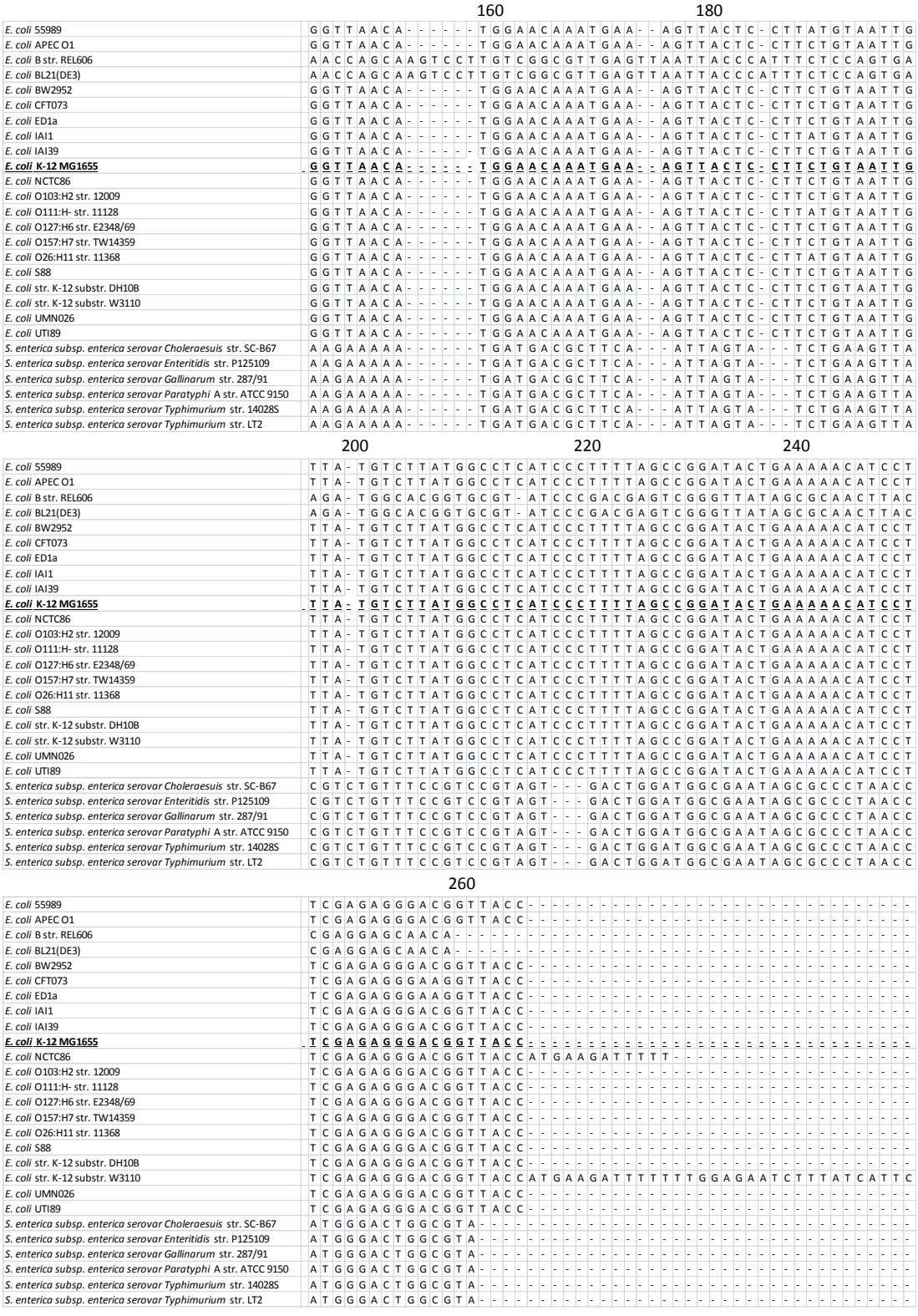


Figure 4.26: Promoter alignments for the *ydiV* upstream sequences in *E. coli* and *S. enterica* strains

Figure 4.26: Promoter alignments for the *ydiV* upstream sequences in *E. coli* and *S. enterica* strains

DNA sequence alignment of the upstream +247 bases for the *ydiV* gene, from various sequenced bacterial genomes.

The numbers indicate the base number of the alignment for reference, with 1 being the start of the intergenic gene region between *nlpC* and *ydiV* genes in *E. coli* K-12 MG1655, and base 247 being adjacent to the *ydiV* sequencing start codon. The sequence of *E. coli* K-12 MG1655 is in bold and underlined, as this is the strain studied throughout this work and thereby used for comparison with the other strains.

The highlighted regions indicate predicted regulatory promoter motifs. The yellow region at 10 bases is a -10 (TATAAT box), present only in the *Salmonella* strains. The blue region (at 80 bases) highlights the -35 box and an additional yellow region (at 100 bases) is indicative of another -10 box, both of which are involved in the recruitment of RNA polymerase (σ^{70} and σ^{24}). The red region shows the UP region (-60 to -40) which can increase σ^{24}-dependent transcription by up to 30-fold when it is particularly A/T rich.

More significant than just the DNA sequence however, was the presence or absence of promoter motifs in this DNA. These DNA motifs are required for RNA polymerase and sigma factors, as well as transcription factors, to attach to the DNA and regulate gene transcription. The *ydiV* gene is thought to be regulated by the RpoE (σ^{24}) which is a second heat stress sigma factor in *E. coli*, alongside the housekeeping RpoD (σ^{70}) transcription factor (Rouvière *et al.,* 1995; Gama-Castro *et al.,* 2008; Wada *et al.,* 2012). RpoE is known to be activated by cellular stresses, such as heat shock, the misfolding of outer membrane proteins, presence of the starvation signal ppGpp and metal ion exposure (Mecsas *et al.,* 1993; Ades *et al.,* 2003; Egler *et al.,* 2005; Costanzo & Ades, 2006). The expression of *rpoE* has been identified to be positively regulated by σ^{24} itself, and is predicted to be repressed by phosphorylated CpxR (Raina *et al.,* 1995; De *et al.,* 2002). RpoD is the major sigma factor during normal growth conditions, known to be repressed by the anti-sigma factor σ^{D} (Jishage *et al.,* 1996; Jishage & Ishihama, 1998).

The presence of strongly conserved motifs for these sigma factors (RpoE and RpoD) would therefore enable a high rate of gene transcription, whilst a less well conserved motif would reduce the level of transcription. The key motifs required for RNA polymerase (σ^{70} holoenzyme) addition are the -10 element (consensus sequence of TATAAT) and the -35 element (consensus sequence of TTGACA) (Reznikofl *et al.,*1985). Additional UP-regions have also been identified, which can increase transcription 30-fold. These UP-regions are located between -60 and -40, immediately upstream of the -35 box, and are A/T rich (Ross *et al.,* 1993). The RpoE (σ^{24}) factor recognises the same -35 and -10 DNA motifs, but the combined -35/-10 score is significant here, requiring the presence of both motifs to be fully functional (Mutalik *et al.,* 2009; Rhodius & Mutalik, 2010). In addition, the presence of a conserved 'AAC' motif in the -35 region is a good indication of gene expression, as well as A/T rich UP-regions (Mutalik *et al.,* 2009).

Using Softberry promoter prediction software, the -35 and -10 regions were identified in the alignment (Figure 4.26) and UP A/T regions were identified based on the -35 location. The *E. coli* K-12 derived strains all yielded a predicted -35 site (highlighted blue, with the identical sequence TTCCGT). Therefore, whilst relatively low identity is retained in this -35 site (compared to the consensus sequence), all the *E. coli* K-12 strains exhibit the identical motif, predicting identical σ^{70} affinity to this site for all strains. The -10 site (highlighted yellow) was then identified for all the *E. coli* K-12 derived strains (with the sequence

TATAAT) showing full sequence identity with the consensus motif, and therefore indicating strong recruitment of RNA polymerase to this site (Figure 4.26). Of significance here is the low conservation of the -35 site for all *E. coli* K-12 strains, highly important for the transcription by σ^{24}, the transcriptional activation of which is determined by the combined -35/-10 score (Rhodius & Mutalik, 2010). In addition, the -35 site lacks a conserved 'AAC' region, which is again important for σ^{24} activity. However, whilst the -10 and -35 sites are not particularly well conserved, another factor is the UP region, which can increase σ^{24}-dependent transcription by up to 30-fold. This UP-region (highlighted in red) consists of 14 A/T bases for the *E. coli* K-12 strains, indicating reasonably strong gene promotion at this site. Therefore, the *E. coli* K-12 strains have identical promoter motifs suggesting equal, but not particularly strong activation by RNA polymerases.

The *E. coli* B-strains lack the -35 and -10 site entirely and have a weaker A/T rich UP-region, questioning how RNA polymerase would be recruited in these bacterial strains. The *S. enterica* strains lack a predicted -35 site, and correspondingly lack an A/T region upstream of this. However these strains do contain two TATAAT boxes, one of which is at the very start of the analysed DNA sequence, and the other aligned to the *E. coli* -10 site. The first TATAAT box is of unknown function, with the upstream gene (*nlpC*) discrete from *ydiV*, and the promoter regions not overlapping at all. The second TATAAT box is the -10 site, perfectly conserved to the consensus; however the -35 region is poorly conserved, not identified by the promoter detection software at all. In addition to the poor -35 region, the UP-region is weak, with only 9/20 bases being A/T.

Whilst computational analysis has identified *ydiV* transcription to be σ^{24}-dependent (Gama-Castro *et al.,* 2008), the promoter alignment reveals reasonably weak conservation in the promoter DNA to the consensus motifs of σ^{24} transcription. For example, the -35/-10 score for the *E. coli* K-12 strains is not particularly high, with weak conservation of the -35 site. In addition, the 'AAC' motif characteristically found in σ^{24}-regulated genes is absent, indicating that any σ^{24}-dependant transcription of *ydiV* is likely to be weak. Alongside the poor conservation amongst the *E. coli* K-12 strains, the *S. enterica* and *E. coli* B-strains have even weaker promoter motifs. This poor promoter region may explain the difficulty in determining the expression conditions for *ydiV* transcription, possibly with poor transcription even under optimal conditions.

To investigate this further, artificially increasing the cellular σ^{24} levels could enable *ydiV* expression conditions to be identified, with significant levels of σ^{24} overriding the problematic weak *ydiV* promoter regions. To increase the σ^{24} levels, either addition of the starvation signal ppGpp, or deletion of the σ^{24} repressor proteins (RseA and RseB) would produce high levels in the cell. In this way, it could be shown whether *ydiV* transcription is really σ^{24}-dependent or whether σ^{70} is responsible for its transcription. For completion, overproduction of σ^{70}, and analysis of *ydiV* transcription would determine which σ factor was being used, and also determine whether the *ydiV* transcription rate was low irrespective of which σ factor was used.

4.7.4 Phylogenetic analysis

Using the promoter alignments for *E. coli* and *S. enterica* strains, a phylogenetic tree was produced based on maximum parsimony (Section 2.10.3), using Mega6 software (http://www.megasoftware.net), analysing just the *ydiV* promoter region from each strain (Figures 4.27, 4.28).

The phylogram visually shows the DNA differences between the various strains, with the scale bar indicating 0.2 substitutions per nucleotide site (Figure 4.27). This phylogram shows the high sequence similarity across the *E. coli* K-12 derived strains, with very short branches interlinking these species, indicative of a very small number of DNA changes across this group. The *S. enterica* group again show small DNA differences amongst the *S. enterica* strains, revealing the high level of DNA conservation amongst this group. However, the branch linking the *S. enterica* group and the *E. coli* K-12 strains is relatively long, indicative of the vast number of DNA substitutions between the two groups. The *S. enterica* strains have diverged less from the ancestor, with a probability of ~0.4 substitutions per nucleotide site, meaning approximately every other base is different in the *S. enterica* strains compared to the ancestor. Comparing this with the *E. coli* K-12 strains and the common ancestor, a much greater proportion of the DNA sequence has altered over evolution, with a probability of ~1 substitution per nucleotide, meaning almost all the DNA sequence is different in the *E. coli* K-12 compared to the common ancestor. However, the distance between the *S. enterica* and *E. coli* K-12 strains reveals fewer DNA changes between these two groups, with a probability of ~ 0.6 substitutions per nucleotide or a DNA change for 2/3 nucleotides. Therefore approximately 1/3rd of the DNA sequence

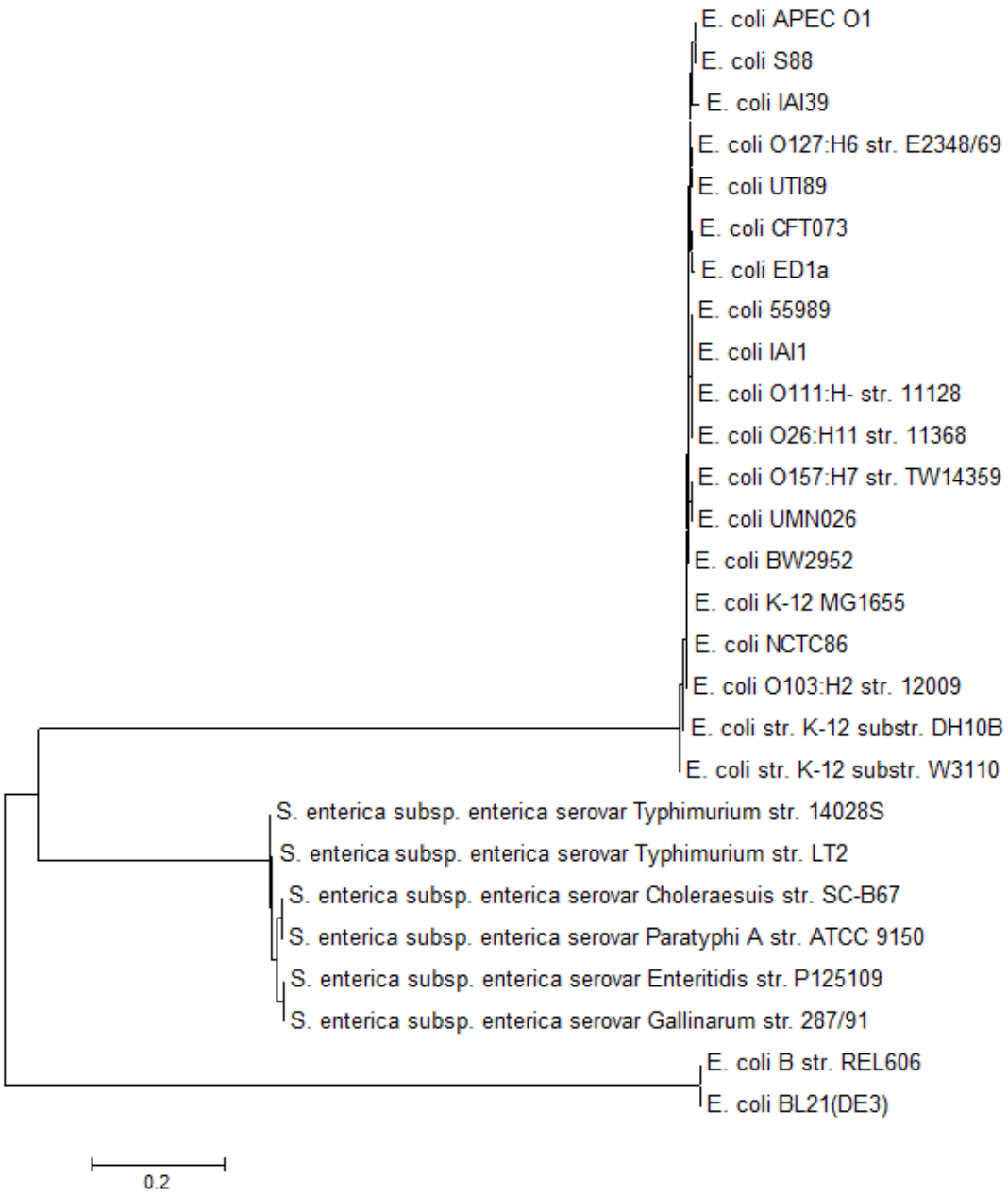


Figure 4.27: Phylogram of *E. coli* and *S. enterica* strains based on similarities in *ydiV* promoter sequences

Based on the *ydiV* promoter sequences, the strains were aligned phylogenetically based on a method of maximum parsimony. The phylogram thereby indicated the strains of greatest evolutionary similarities and differences. The length of the branches between each strain is indicative of the time between each strain, with small branches signifying highly similar strains.

The scale 0.2 refers to 0.2 substitutions per nucleotide site, and therefore enables the level of DNA similarities and differences between strains to be identified.

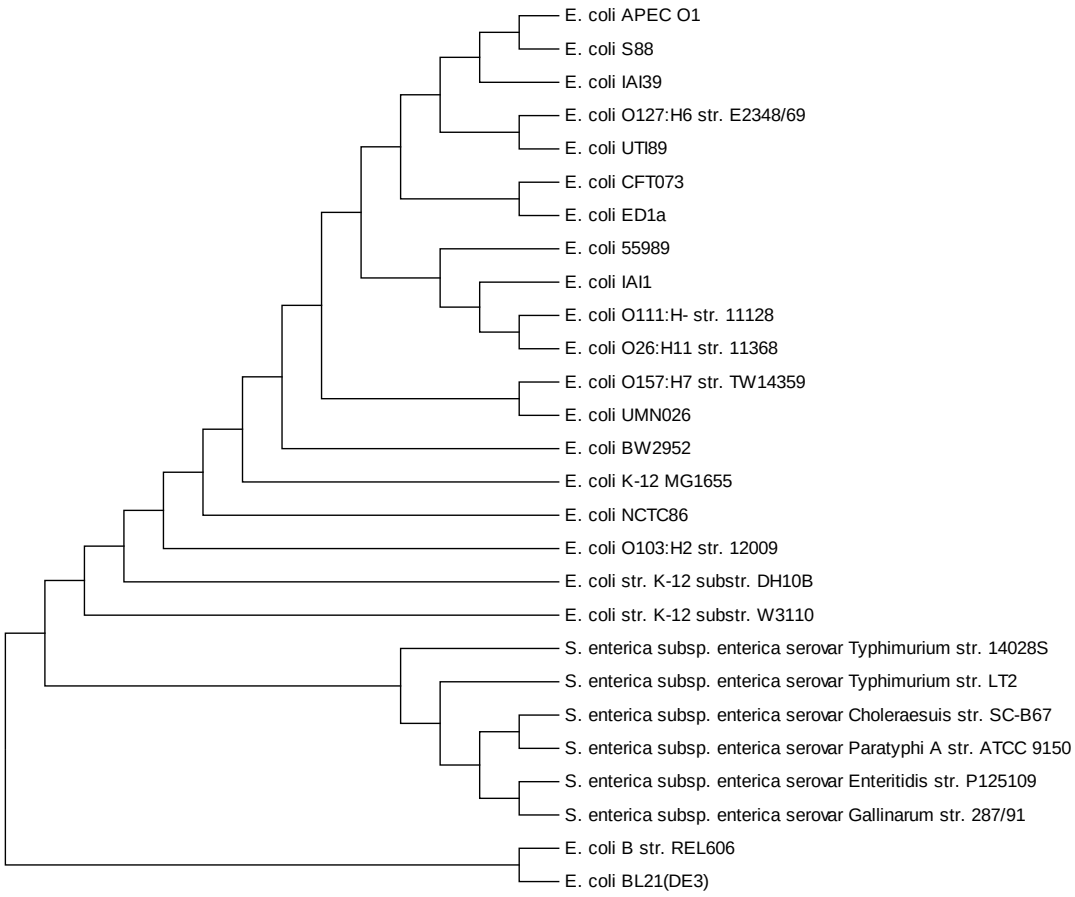


Figure 4.28: Cladogram to reveal the relatedness of *E. coli* and *S. enterica* strains based on similarities in *ydiV* promoter sequences

Based on the *ydiV* promoter sequences, the strains were aligned phylogenetically, which indicated the strains of greatest similarities and differences.

Each branch point identifies that the descending branches are produced of a common ancestor, with many strains being closely related. Whilst the relatedness of strains can be visualised here, the branches are not to scale, so no evolutionary time or divergence can be determined.

is conserved amongst the *E. coli* K-12 and *S. enterica* strains, showing the level of divergence between the groups.

Moving on to study the *E. coli* B-strains, based on the *ydiV* promoter sequences, these strains appear to be produced from an ancestor, which subsequently diverged into the *E. coli* K-12 and *S. enterica* strains (Figure 4.27). However, this is an *ydiV* promoter specific discovery, which is not supported by genome-wide studies of *E. coli* and *S. enterica* ancestory (Elena *et al.,* 2005). Based on the *ydiV* promoter phylogram, it appears that the DNA sequences in the three groups has diverged quite substantially, to account for the large branches between the three groups (Figure 4.27). Using the branch lengths, it can be seen that the *E. coli* B-strains are more similar in DNA sequence to the *S. enterica* strains than the *E. coli* K-12 strains. Using the scale, the *E. coli* B-strains have a 0.6 probability of nucleotide substitutions compared to the *S. enterica* strains, indicating only 40% sequence homology between the two groups. However, this is a greater sequence homology than that between the *E. coli* B-strains and the K-12 strains, which ranges between 0-40% sequence homology on a species-species basis.

The most apparent conclusion from this phylogram is the high level of sequence similarities within the three groups, but the relatively high levels of divergence between the groups, with the *E. coli* B-strains being more similar to the *S. enterica* strains than the *E. coli* K-12 strains.

In order to identify the strains of highest similarities, a cladogram was also produced, which revealed only the branching patterns between strains rather than the amount of DNA changes amongst strains (Figure 4.28). This revealed five pairs of highly similar *E. coli* K-12 strains, as well as two pairs of very similar *S. enterica* strains. In addition, *E. coli* K-12 MG1655, which was used throughout this work as a wild-type model, was the ancestor for 14 identified strains, including the pathogenic avian APEC strain and the O157:H7 strain.

Taking all into account, the induction conditions of *ydiV* are still unknown. Whilst experimental data has failed to reveal any conditions in which *ydiV* up-regulation occurs, promoter alignment has revealed that this may be due to weak promoter motifs, meaning that *ydiV* transcription may always be at a low level. Further analysis to improve the conservation of promoter consensus sites (by site-directed mutagenesis) could reveal

whether the low levels of *ydiV* transcription are due to simply poor RNA polymerase transcription. Equally, exploring more extreme growth conditions would presumably activate σ^{24} at a higher level, and therefore may identify *ydiV* growth conditions.

4.8 Discussion

Work carried out in this chapter has characterised the *E. coli* protein YdiV. Successful cloning and overproduction of the YdiV protein fused to a C-terminal His_6 tag was undertaken, enabling a simple one-step purification which yielded pure and concentrated protein. YdiV has been reported to be an active PDE by Hisert *et al.* (2005), but an inactive PDE by Tchigvintsev *et al.* (2010) and Wada *et al.* (2011). The data reported here supports the inactive or 'degenerate' view of YdiV. Through nucleotide reactivity assays, it was shown that YdiV did not react with c-di-GMP (the substrate of active EAL domains) nor with several other tested nucleotides, identifying YdiV as one of the four 'degenerate' GGDEF/EAL domain proteins in *E. coli* (Figure 4.9). This supported the conclusions of Tchigvintsev *et al.* (2010) and Wada *et al.* (2011).

Following the assigment of YdiV as an inactive PDE, studies were undertaken to investigate nucleotide binding to the protein, which might regulate YdiV protein function. Previous studies had shown the possibility of degenerate GGDEF/EAL proteins becoming activated by nucleotide binding (Duerig *et al.,* 2009; Lee *et al.,* 2007), and structural analysis of FimX had provided information specific to EAL domain proteins (Navarro *et al.,* 2009). Futhermore, the X-ray crystallography structure of YdiV indicates a groove similar to that seen in other EAL domain proteins to enable c-di-GMP coordination at the active site (Li *et al.,* 2012). Therefore, it seemed quite plausible for YdiV to be regulated by a nucleotide of some type.

The intrinsic fluorescence results were largely inconclusive, given that all the fluorescence changes upon nucleotide addition were quantitatively small, and of equal size to the control variations. Therefore the results of interest are those from the partial proteolysis and proton NMR. Partial proteolysis (Figure 4.11) showed a slightly varied protein fragmentation pattern upon c-di-GMP, ppGpp and cAMP additions compared to the protein alone or the protein mixed with the other nucleotides. Clear variations in the profiles were seen; however in no case did the cleavage pattern dramatically alter

suggesting no large conformational changes. There are 14 cleavage sites (K and R residues) in YdiV, resulting in 120 possible protein fragments. From the SDS-PAGE gels of the partial proteolysis assays (Figure 4.11) there appear to be two fragments between 15-20 KDa, one fragment between 10-15 KDa and one fragment smaller than 10 KDa. The identity of these fragments (in terms of residue components) is unknown, and with 120 possible fragments is difficult to predict. To identify the location of these fragments in the whole protein, N-terminal sequencing could be carried out on these SDS-PAGE gel fragments, to decipher the trypsin sites cleaved and give further insight into nucleotide-dependent changes in protein structure.

The major difficulty is determining the significance of these trypsin cleavage changes. Whilst clear differences are seen upon specific nucleotide addition, the overall fragment pattern is very similar, confirming that a major conformational change has not occurred. Since nucleotide binding is likely to occur, if at all, at a small binding pocket of the protein, this is not necessarily surprising. Therefore the differences seen in cleavage fragment number 1 and its smaller halo partner species could be very significant, indicating binding of a nucleotide and therefore concealing a trypsin cleavage site forming the more intense single band at cleavage site 1 with ppGpp present. Equally, in the case of c-di-GMP and cAMP, nucleotide binding may have revealed a trypsin cleavage site and therefore produced weaker bands at cleavage fragment 1.

Proton NMR revealed spectral changes in the amide region for all tested nucleotides in comparison with the YdiV only sample (Figure 4.15). However, 5 out of 6 tested nucleotides produced near identical spectra and therefore these were used as baseline spectra. The addition of c-di-GMP to the protein produced distinct chemical shifts, which were not present in the YdiV protein sample or for the remaining nucleotide and protein samples. This therefore highlighted c-di-GMP as a possible binding molecule, inducing chemical shifts at 7.8 and 8.1 ppm. Of note, the known c-di-GMP binder, Clp produced chemical shifts at the same spectral region upon c-di-GMP addition (Figure 4.19). This suggests the possibility that c-di-GMP is binding to YdiV.

Taken together, the partial proteolysis and proton NMR data tentatively suggest c-di-GMP binding to YdiV. The partial proteolysis assays highlighted c-di-GMP as one of a few nucleotides to subtly alter the cleavage pattern of the pattern. Proton NMR reinforced this

suggestion, with c-di-GMP producing a distinctly different spectrum to the other tested nucleotide samples, mirroring some of the changes seen for Clp:c-di-GMP binding. Whilst a firm conclusion cannot be made, the data provides the foundations for further investigations. To further investigate this binding, a wider range of experiments could be carried out, such as crosslinking methods (to capture any transient interactions) and high sensitivity techniques such as Isothermal Titration Calorimetry.

The oligomerisation state of YdiV was also investigated to determine whether YdiV retained the ability to form dimers, as is the case for other EAL domains (Tchigvintsev *et al.,* 2010). Oligomeric analysis revealed YdiV to be a monomeric protein (Figures 4.6 and 4.7). Whilst undertaking this investigation, the same conclusions were found by Li *et al.* (2012), both by X-ray crystallography and gel filtration techniques (Li *et al.,* 2012). The YdiV structure revealed a decreased 'dimerisation interface' between the YdiV monomers compared to other catalytically active EAL domains (Li *et al.,* 2012). Additionally, YdiV was identified to lack three of seven conserved residues required for EAL domain protein dimerisation (Tchigvintsev *et al.,* 2010). Therefore it is not at all surprising that YdiV cannot form dimeric species, existing solely as a monomeric protein.

Of interest then, if a nucleotide was found to bind to the protein, would be to analyse whether any conformational change was induced by binding, causing the dimerisation interface to open and enabling dimerisation to become more stable. To investigate this further, a modified 2-hybrid system could be used, for example tagging YdiV with two domains of adenylate cyclase (Karimova *et al.* 1998). If YdiV was to dimerise, the two domains of adenylate cyclase would come together, creating the active enzyme and increasing the cAMP levels in the cells. The cAMP would then activate CRP, and a reporter promoter-*lacZ* fusion would be induced. The expression of *lacZ* could be detected by β-galactosidase assays, and therefore the level of dimerisation assessed. Various possible nucleotides could then be added to the bacteria, assessing the affect the nucleotides have on YdiV dimer stability, and determining both if a nucleotide binds and whether it alters YdiV stability. Alternatively site-directed mutagenesis of the dimerisation motif to the conserved motif would be an interesting experiment, analysing whether this did restore the ability of YdiV to form dimers, and analysing the nucleotide binding capacity of this mutated protein.

Functional characterisation of YdiV revealed the importance of the protein in flagella production, supporting the motility phenotypes seen by Dr Melissa Lacey (unpublished). Dr Lacey found over-expression of *ydiV* resulted in a marked decrease in cell motility, which was confirmed here by detection of FliC levels and visualisation of flagella. Western blot analysis identified a complete lack of FliC in an *ydiV* overexpression strain (MG1655 pBR322:*ydiV*), which resulted in a marked decrease in flagella numbers visible by electron microscopy (Figures 4.20 and 4.21). Therefore, the data supports the phenotypic analysis, showing that upon *ydiV* over-expression, flagellin protein (FliC) is repressed, resulting in a decrease of flagella numbers of ~80% and hence causing the bacteria to be less motile. This is consistent with the motility inhibitory effects of YdiV overproduction seen by other groups during the course of this work (Li *et al.,* 2012; Wada, *et al.* 2012).

Whilst work in this chapter has begun to characterise the function of YdiV, the conditions of *ydiV* expression are still largely unknown. Western blot analysis of cells grown at different growth rates failed to show an optimal growth rate for high *ydiV* induction, with the YdiV levels being below the detection limit (Figure 4.22). Whilst this does not necessarily mean that *ydiV* is not induced here, it does suggest that the quantities of YdiV protein in the cell are extremely low. Equally, while this does not disregard the functional phenotypes we have determined for YdiV, it puts into question how artificial these situations are. It is obvious that experimentation using overproduced protein is not an ideal comparison with the situation *in vivo*, however the majority of the effects seen with YdiV (motility inhibition, possible nucleotide binding) are seen with overproduced levels of YdiV. Clearly this questions whether these effects would occur, at least to the same extent, if YdiV was at a much lower concentration. The lack of progress to determine *ydiV* upregulation conditions was illustrated by β-galactosidase assays (Figure 4.24) in which quorum sensing molecules AHL were added to induce *ydiV* expression. Again these experiments did not identify any growth condition under which *ydiV* was upregulated, leaving open the environmental conditions needed for up-regulation. The addition of AHL molecules did not seem to affect the level of *ydiV* expression, however despite previous evidence that SdiA upregulated *ydiV* (Zhou *et al.,* 2008), it has recently been contradicted by Spurbeck *et al.* (2013), who claim SdiA has no effect on *ydiV* expression, a conclusion also reached from the experiments carried out here (Figure 4.24).

Promoter alignments were made, to determine the degree of sequence similarities of the *ydiV* promoter regions across the *E. coli* and *S. enterica* genomes (Figure 4.26). This revealed three distinct groups; the *E. coli* K-12 strains; the *E. coli* B-strains and the *S. enterica* strains, each of which had high sequence identity within the group, but differed from the other groups. The transcription of *ydiV* is predicted to be σ^{70}- and σ^{24}-dependent. Promoter prediction software revealed a highly conserved -10 region for all the *S. enterica* and *E. coli* K-12 derived strains, but a weak or absent -35 region; a characteristic often associated with promoters that require a positively acting transcription factor. The activity of σ^{24}- is dependent on the overall strength of both the -35 and -10 boxes, plus an additional 'AAC' motif at the -35 site, which was lacking throughout. Therefore the level of RNA-polymerase recognition of the promoter motifs is likely to be reasonably weak, possibly explaining the low level of *ydiV* transcription via poor promoter recognition. Additionally, genome wide transcriptional analysis has been carried out for *E. coli*, and provided an insight into *ydiV* induction. Transcription of *ydiV* in *E. coli* K-12 is down regulated by carbon monoxide (Nobre *et al.,* 2009), hyperosmotic conditions (Gunasekera *et al.,* 2008) and acidic pH stress (King *et al.,* 2010), and up-regulated slightly by DNA damage (by Norfloxacin) and biofilm formation (Faith *et al.,* 2007; May *et al.,* 2009). However, the pathogenic strain *E. coli* 0157:H7 has shown significant *ydiV* up-regulation in the presence of cinnamaldehyde (Visvalingam *et al.,* 2013), with *N*-acyl homoserine lactones actually down-regulating the gene expression (Hughes *et al.,* 2010). Clearly therefore the *ydiV* gene expression can be altered, but possibly the growth conditions required for up or down-regulation are simply more extreme than typical laboratory growth conditions for *E. coli*, suggesting the possibility of a basal level of gene-expression which is only altered in stress conditions. This is consistent with *ydiV* expression being σ^{24}-dependent, as this factor is a heat-shock stress factor, or dependent on an unknown transcription factor. However this yields no explanation how YdiV imparts its functional roles, in motility and as an anti-FlhDC factor in non-extreme conditions.

To conclude, the *E. coli* protein YdiV has been found to be a catalytically inactive PDE, part of the 'degenerate' group of GGDEF/EAL proteins in bacteria. The protein has been successfully purified, enabling investigation of potential activation triggers for the protein, in the form of nucleotides. Nucleotide binding assays have suggested a possible nucleotide interaction with c-di-GMP, but require further investigation to increase the confidence of this result. Functional investigations have shown YdiV to regulate bacterial motility by

regulation of flagella synthesis in some manner, cementing the importance of the protein in cell function. However the induction conditions of *ydiV* have not been elucidated, but analysis of the promoter regions revealed weak conservation of consensus motifs, suggesting that transcription is generally low due to a lack of RNA polymerase recruitment at these sites.

Further work in this thesis will investigate the mechanism in which YdiV regulates flagella synthesis and impacts motility, and it's interactions with the transcriptional regulator $FlhD_4C_2$ and the effect this has on the DNA-binding capacity of $FlhD_4C_2$.

Chapter 5: Characterisation and functional investigation of the *Escherichia coli* YdiV and FlhDC interaction

5.1 Introduction

As discussed previously, the *E. coli* flagellar biosynthesis genes are arranged in a three tier system, named Class I, II and III, or early, middle and late genes (Figure 1.2). The Class I genes consist of the *flhDC* operon, Class II includes operons *flgAMN*, *flhBAE*, *fliAZY* and *fliDST*, and Class III includes *fliC* and *motAB* (Liu & Matsumura, 1994; Chilcott & Hughes, 2000).

The *flhDC* operon is responsible for flagellar gene expression. The *flhD* and *flhC* genes encode proteins which form the heterotetrameric complex $FlhD_4C_2$ (Wang *et al.,* 2006). The $FlhD_4C_2$ complex is a DNA-binding protein, that up-regulates Class II genes (Liu & Matsumura, 1994). These Class II genes encode proteins such as the σ factor FliA, basal body proteins such as FlgB, chaperones such as FliT and export machinery proteins such as FlhB (Figure 1.2) (Chilcott & Hughes, 2000; Macnab, 2003). FliA activates the expression of all Class III genes, producing proteins such as the flagella filament protein FliC and chemotaxis proteins CheR and CheB (Chilcott & Hughes, 2000; Macnab, 2003).

Due to the impact of FlhDC on this system, the complex is regarded as the 'master regulator' of flagella gene transcription. The two proteins FlhD and FlhC, which can form homodimers ($FlhD_2$ and $FlhC_2$) are functional as a complex, initially thought to be $FlhD_2C_2$, before being shown to be the hexameric form $FlhD_4C_2$ (Liu & Matsumura, 1994; Claret & Hughes, 2002; Wang *et al.,* 2006). This $FlhD_4C_2$ complex binds upstream of class II flagellar genes, at an ~50 bp region, which overlaps the -35 element of the promoters (Liu & Matsumura, 1994; Claret & Hughes, 2002). Sequence comparisons identified these binding regions as two inverted repeats of the consensus sequence $AA(C/T)G(C/G)N_{2/3}AAATA(A/G)CG$, separated by a 10-12 base pair spacer (Claret & Hughes, 2002; Stafford *et al.,* 2005). These two repeats were named as 'FlhDC binding boxes', with a single $FlhD_4C_2$ complex binding to the entire 50 bp region, interacting with both FlhDC binding boxes (Claret & Hughes, 2002). Binding of $FlhD_4C_2$ to the promoter DNA induces a

bend of 111° in the DNA, causing DNA to wrap itself around the protein complex and gene transcription to be promoted (Wang *et al.,* 2006).

Due to the extensive influence of $FlhD_4C_2$, regulation of its expression and activity is critical. The expression of *flhDC* is regulated by several systems, such as quorum sensing, heat shock and high osmolarity responses and by several regulatory proteins such as CsrA and LrhA (Sperandio *et al.,* 2002; Shin & Park, 1995; Lehnen *et al.,* 2002; Wei *et al.,* 2001; Shi *et al.,* 1992). Flagellar synthesis is further regulated by the induced transcriptional cascade produced following *flhDC* transcription (Figure 1.2). FlhDC itself regulates flagellar synthesis by up-regulating class II genes. Protein products of this class II transcription then act to activate (FliZ) and repress (FliT) the activity of FlhDC in a feedback loop, to ensure tight regulation of FlhDC activity. Furthermore, other class II transcription products (FliA and FlgM) are responsible for the transcription of class III genes, with FliA (σ^{28}) activating transcription and FlgM down-regulating the activity of FliA. These many levels of regulation ensure fine-tuning of the extent and timing of flagella production.

In *Salmonella enterica*, it has been found that $FlhD_4C_2$ binds to the STM1344 protein. This protein:protein interaction affects DNA expression, with STM1344 inhibiting $FlhD_4C_2$ binding to Class II promoter DNA, thereby repressing gene transcription (Wada *et al.,* 2011). The activity of STM1344 therefore has a huge downstream impact, affecting the production of flagella and therefore motility in *S. enterica*.

The *E. coli* YdiV and *S. enterica* STM1344 proteins have significant similarities; both being catalytically inactive PDEs with high sequence identity (52%) (Simm *et al.,* 2009). In addition STM1344 has been found to affect the c-di-GMP pool, despite neither synthesising nor degrading the molecule. Moreover, STM1344 has been found to affect the transition between biofilm formation and motility in *Salmonella*, mirroring the functions of YdiV in *E. coli* (Simm *et al.,* 2009). However, there are also noteable differences. Deletion of STM1344 prevented efficient flagellin production, which did not occur upon YdiV deletion (Wada *et al.,* 2011). Additionally, 52% amino acid sequence identity, whilst high, is considerably lower than the average protein homology between *S. enterica* and *E. coli* (of ~ 80-90%) indicating possible functional differences between the proteins (McClelland *et al.,* 2000).

Work in this chapter provides information on:

- Kinetic analysis of the YdiV: FlhDC interaction, revealing a high affinity interaction (Section 5.4)
- Identification of *in vivo* growth conditions in which this complex is likely to be formed (Section 5.6)
- Determination of YdiV to act as an anti-FlhDC factor, causing dissociation of DNA above certain concentration ratios (Section 5.7)
- Investigation of nucleotide binding to the YdiV:FlhDC complex, yielding no conclusive results (Section 5.10)
- Analysis of the FlhDC:YdiV complex in competition with other FlhDC binding proteins, discovering the formation of a FliT:FlhDC:YdiV complex (Section 5.11)

5.2 Construction and over-production of the YdiV and FlhDC proteins

Overproduction plasmids were constructed to produce various fusion proteins: YdiV-His_6, FlhDC-His_6 and $(FLAG)_3$-FlhDC, which were used in several assays in this chapter. The cloning and overproduction of YdiV-His_6 was as described in Section 4.2.

5.2.1 Cloning of FlhDC to create over-expression plasmids

The *E. coli flhDC* operon was amplified from *E. coli* MG1655 genomic DNA and inserted into the pET21a plasmid. This was achieved in two ways, simply utilising the plasmids His_6-tag to create a His_6-tagged construct, and engineering a triple FLAG-tag into the plasmid, creating a $(FLAG)_3$-tagged FlhDC (Figure 5.1).

To enable formation of a His_6-tagged protein, the *flhDC* operon was amplified by PCR from genomic *E. coli* MG1655 DNA using primers NSW35 and NSW36 (Section 2.3.3). These primers contained BamHI and XhoI restriction sites, with no stop codon included in NSW36 to enable a C-terminal His_6-tag fusion (Table 5.1).

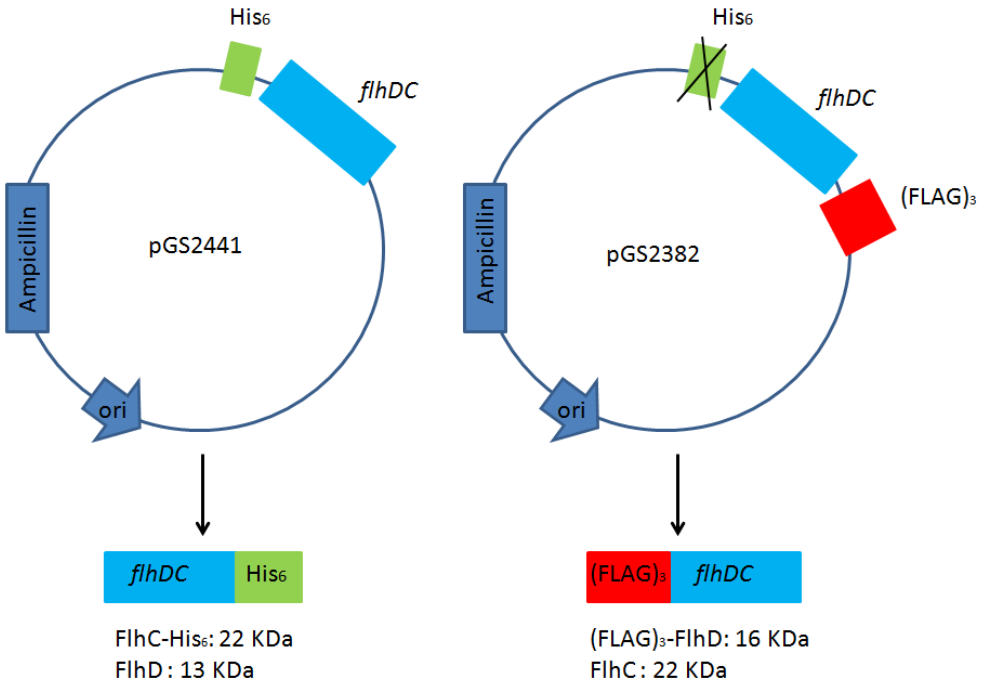


Figure 5.1: Schematic plasmid maps for the insertion of *flhDC* into pET21a constructs

Plasmid maps for the pET21a vector, indicating the position of the tags in the vector, the antibiotic resistance cassette and the fusion constructs produced. Two DNA constructs were made, encoding a C-terminal His_6 fusion, and an N-terminal $(FLAG)_3$ construct and named pGS2441 and pGS2382 respectively.

The origin of replication is indicated and labelled as ori.

The C-terminal tag produces FlhC-His_6 and FlhD of expected molecular weights of 22 and 13 KDa respectively. The N-terminal tag produces the proteins $(FLAG)_3$-FlhD and FlhC of expected molecular weights 16 and 22 KDa respectively.

Table 5.1: Primers for the production of two overproduction plasmids producing the fusion proteins FlhDC-His_6 and $(FLAG)_3$-FlhDC

Primer	Sequence	Function
NSW35	TTTTGGATCCATGCATACCTCCGAGTTG CTG	Forward primer for *flhDC* amplification to insert into pET21a
NSW36	TTTTCTCGAGAACAGCCTGTACTCTCTG TTC	Reverse primer for *flhDC* amplification to insert into pET21a
NSW15	TATGGACTACAAGGACGATGACGACAA GGACTACAAGGACGATGACGACAAGGA CTACAAGGACGATGACGACAAGG	Forward primer encoding $(FLAG)_3$ to insert into pET21a
NSW16	GATCCCTTGTCGTCATCGTCCTTGTAGTC CTTGTCGTCATCGTCCTTGTAGTCCTTGT CGTCATCGTCCTTGTAGTCCA	Reverse primer encoding $(FLAG)_3$ to insert into pET21a
NSW17	TTTTGGATCCCATACCTCCGAGTTGCTG AAA	Forward primer for *flhDC* amplification to insert into pET21a:$FLAG_3$
NSW18	TTTTAAGCTTTTAAACAGCCTGTACTCTC TGTTC	Reverse primer for *flhDC* amplification to insert into pET21a:$FLAG_3$

The sequences shaded grey indicate BamHI restriction sites, with the boxed sequence indicating a HindIII restriction site, cyan specifying an NdeI site, and green indicating an XhoI site.

The PCR product was isolated by PCR purification and digested with BamHI and XhoI restriction enzymes (Sections 2.3.10 and 2.3.6). The pET21a plasmid (Novagen) was purified (Section 2.3.4) and digested with the same restriction enzymes (Section 2.3.6) before ligation of the *flhDC* fragment and the linearised pET21a vector (Section 2.3.9). This ligated DNA was used to transform electrically competent *E. coli* DH5α cells and transformants were selected on ampicillin supplemented agar (Section 2.3.11). Following this, the colonies were screened by colony PCR, before plasmids were checked by DNA sequencing (Source Bioscience) using the T7 promoter and T7 terminator primers (Table 4.2). The resulting plasmid containing no mutations in the *flhDC* sequence was named pGS2441 (pET21a:*flhDC*).

To create the $(FLAG)_3$-tagged protein construct, a two-step process was carried out, initially creating a $FLAG_3$-tag in the pET21a vector, before amplifying the *flhDC* operon for insertion into this modified vector. Oligonucleotides NSW15 and NSW16 were designed to encode a 3x-FLAG tag and were annealed at 68°C to produce a dsDNA fragment (Table 5.1). Of note,

NSW15 contained a start codon, designed to enable transcription of the FLAG-tag in the vector. The $FLAG_3$ fragment and the pET21a vector were then digested with NdeI and BamHI restriction enzymes (Section 2.3.6) before ligation of the $(FLAG)_3$ fragment into the linearised vector (Section 2.3.9). This ligated DNA was then used to transform electrically competent *E. coli* DH5α. Transformants were selected and screened as previously described for pET21a plasmids (Section 2.3.11). The resultant plasmid was named pGS2381 (pET21a:$FLAG_3$) and was used to generate a $(FLAG)_3$-FlhDC plasmid. The *flhDC* operon was amplified from genomic *E. coli* MG1655 DNA (Section 2.3.3) using primers NSW17 and NSW18, introducing BamHI and HindIII restriction sites at either end of the PCR product (Table 5.1). A stop codon was included in primer NSW18, to terminate transcription at the 3' end of the *flhDC* operon, but no start codon was required, initiating transcription in the N-terminal $(FLAG)_3$ tag instead. Following amplification, the purified product (Section 2.3.10) and pET21a:$FLAG_3$ vector (Section 2.3.4) were digested with BamHI and HindIII restriction enzymes (Section 2.3.6) before ligation of the DNA (Section 2.3.9). Once again the ligated DNA was used to transform electrically competent *E. coli* DH5α cells (Section 2.3.11) and transformants were selected and screened before DNA sequencing (Source Bioscience) using the T7 promoter and T7 terminator primers (Table 4.2). The resultant plasmid was named pGS2382 (pET21a:$FLAG_3$:*flhDC*).

5.2.2 Over production of the FlhDC protein

Plasmids pGS2382 and pGS2441 were used to transform electrically competent *E. coli* BL21 λ(DE3) (Section 2.3.7). The transformants were used to overproduce the $(FLAG)_3$-FlhDC and FlhDC-His_6 fusion proteins respectively. In each case an aliquot (1 ml) of overnight culture was added to 100 ml of ampicillin supplemented LB, and the cells were grown at 37°C with 250 rpm agitation, until an OD_{600} of ~0.6 was reached (Section 2.6.1). IPTG was added to a final concentration of 100 µg/ml and the cultures were incubated at 25°C for the remainder of cell growth. Samples were taken at hourly intervals and the cellular protein composition analysed by SDS-PAGE (Section 2.6.9) to determine the optimal growth conditions for protein overproduction.

The $(FLAG)_3$-FlhDC protein was clearly visible in the overproduction samples, with a 3 h or more induction period giving satisfactory overproduction levels (Figure 5.2). Equally, the FlhDC-His_6 fusion protein was successfully overproduced, yielding rather modest overproduction levels after 2 or more h induction (Figure 5.3). For production of both

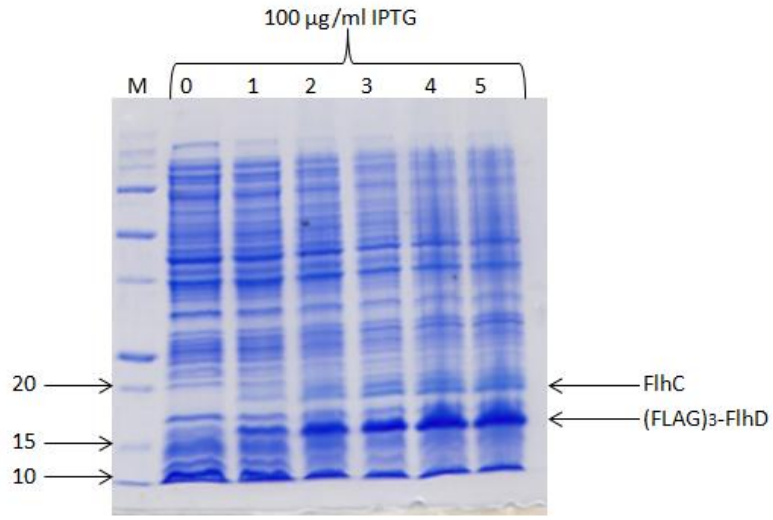


Figure 5.2: Optimisation of overproduction conditions for the $(FLAG)_3$-FlhDC protein in *E. coli* BL21 (λDE3)/ pGS2382 cells

Coomassie-stained SDS-PAGE gel of whole cell samples taken pre- and post-induction of FlhDC expression.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-5 indicate induction times in h, with FlhDC production being induced using a final concentration of 100 µg/ml IPTG, as specified. The position of FlhC and $(FLAG)_3$-FlhD are indicated, at approximately the expected sizes (FlhC: 22 KDa, $(FLAG)_3$-FlhD: 16 KDa).

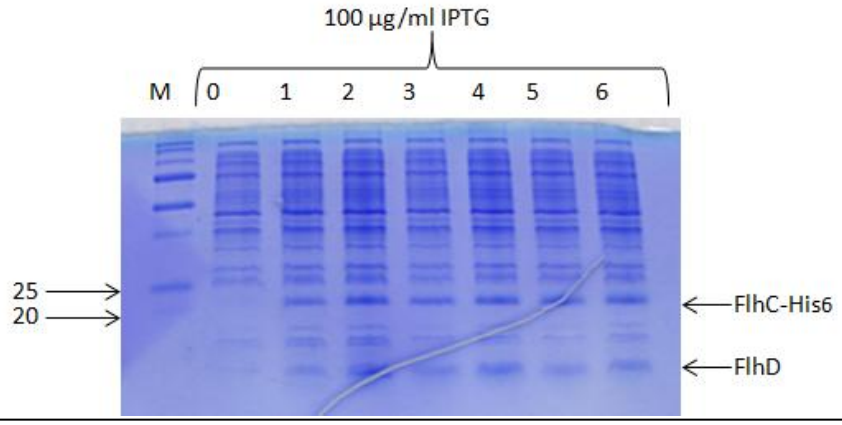


Figure 5.3: Optimisation of overproduction conditions for the FlhDC-His_6 protein in *E. coli* BL21 (λDE3)/ pGS2441 cells

Coomassie-stained SDS-PAGE gel of whole cell samples taken pre- and post-induction of FlhDC expression.

M: Precision plus protein standards (molecular weights indicated in KDa), 0-6 indicate induction times in h, with FlhDC production being induced using a final concentration of 100 µg/ml IPTG, as specified. The position of FlhC and FlhD-His_6 are indicated, at approximately the expected sizes (FlhC-His_6: 22 KDa, FlhD: 13 KDa).

proteins a 3 h induction period was chosen for future cell culture. The differing molecular weights of the FlhD and FlhC proteins, encoded by the two plasmids used for overproduction, was due to the differing nature and locations of the affinity tags in each case (Figure 5.1).

5.2.3 Purification of His-tagged proteins

Both YdiV-His_6 and FlhDC-His_6 were purified in a single-step affinity chromatography method, whilst the $(FLAG)_3$-FlhDC was used in its soluble crude state.

For all three strains (*E. coli* BL21 λ(DE3)/pGS2354, *E. coli* BL21 λ(DE3)/pGS2441 and *E. coli* BL21 λ(DE3)/pGS2382), the cell pellets (from 1-5 litres of culture) were resuspended in breakage buffer, sonicated and the phases separated by centrifugation (Section 2.6.2)(Table 2.2). The soluble fractions were then used directly in assays, or applied to chromatography columns to produce pure protein (Section 2.6.3).

YdiV purification was achieved with a one-step affinity chromatography method, which yielded >90% pure protein, as described in Section 4.2.3.

FlhDC-His_6 purification was carried out in the same manner as YdiV-His_6 purification, applying the soluble protein fraction produced (from *E. coli* BL21 λ(DE3)/pGS2441 cells) to a 1 ml HiTrap column, for affinity chromatography. The His-chelating programme was run on an AKTA, following manufacturer's instructions (Section 2.6.3). The resultant trace was multi-peaked, with non- specifically bound protein eluting initially, prior to the larger peak, representing the His_6-tagged protein (Figure 5.4A). The elution fractions corresponding to the major peak were then analysed by SDS-PAGE as described in Section 2.6.9 (Figure 5.4B) and showed the presence of the protein of interest, FlhDC, composed of two defined species (corresponding to separated FlhD and FlhC, due to the denaturing conditions of the gel). From the gel, fractions corresponding to lanes 4, 5 and 6 were chosen for future use, as these were the most pure and concentrated protein samples.

In some cases, eluted fractions containing the FlhDC-His_6 fusion protein were used directly, or they were transferred into an imidazole-free buffer using a Vivaspin 20 column for other assays (Section 2.6.7).

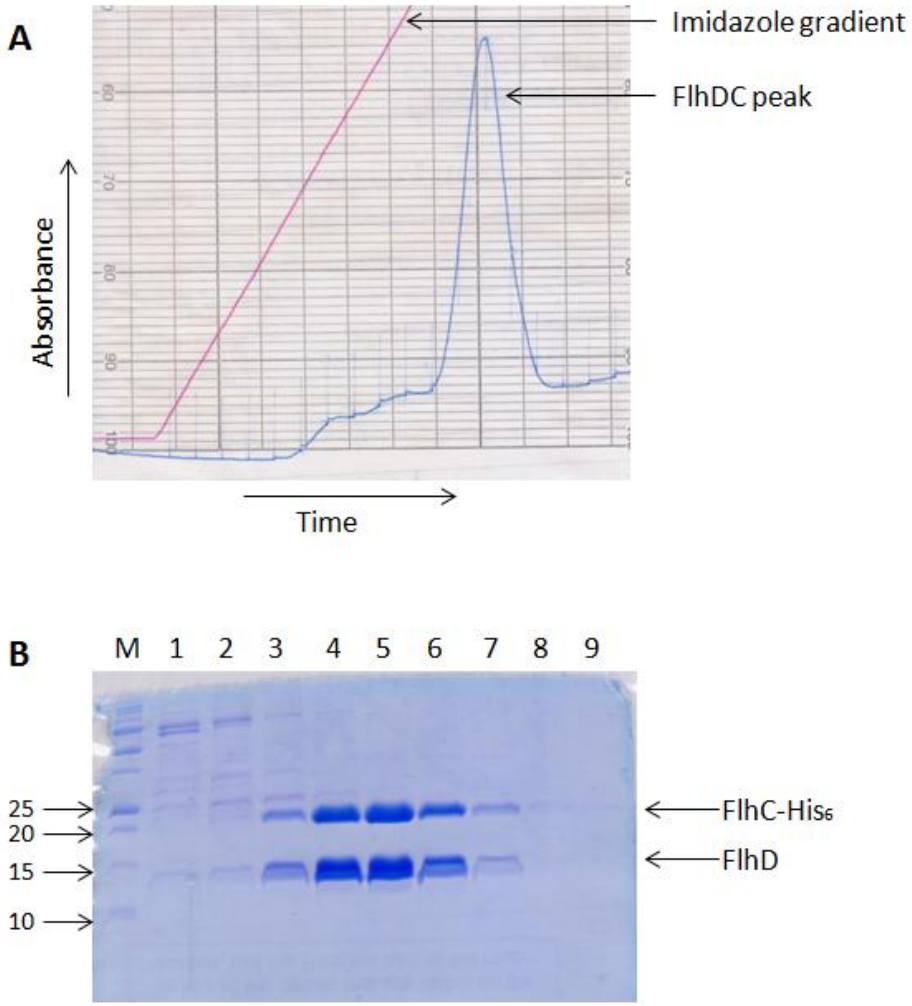


Figure 5.4: HiTrap chelating chromatography as a one-step purification of FlhDC-His$_6$ protein

A) Elution profile of FlhDC-His$_6$ from HiTrap chelating chromatography.
Cell-free extract containing FlhDC-His$_6$ was applied to a 1 ml HiTrap chelating column and fractionated by an imidazole gradient (linear gradient from 0-0.5 M). The blue trace shows the A_{280} values, indicative of protein levels, with the red trace representing the imidazole gradient. The expected FlhDC-His$_6$ location is indicated.

B) Coomassie-stained SDS-PAGE gel of FlhDC-His$_6$ purification.
HiTrap elution fractions spanning the major peak from the Ni-NTA trace. M= Precision plus protein standards (molecular weights indicated in KDa), lanes 1-9 show elution fractions collected during the imidazole gradient application, and represent fractions corresponding to the labelled peak in (A).
The positions of FlhD (expected molecular weight 13 KDa) and FlhC-His$_6$ (expected molecular weight 22 KDa) are indicated.

The denaturing conditions of SDS-PAGE gels yielded FlhD and FlhC monomers only (Figure 5.4B). However, it is known that these proteins form various oligomeric species, such as homodimers of FlhD or FlhC, as well as the reported multimers $FlhD_2C_2$ and $FlhD_4C_2$ (Liu & Matsumura, 1994, Wang et al., 2006). The hexameric $FlhD_4C_2$ structure is believed to be the functional form of the transcription factor, and therefore the molecular weight of this species was used throughout for molarity calculations. Gel filtration of purified FlhDC showed a complex consisting of 2:1 ratio of FlhD and FlhC proteins, supporting the $FlhD_4C_2$ oligomeric species (data not shown). Production of FlhDC from overproduction constructs however produced a mixture of oligomeric forms; therefore the complex is referred to as FlhDC rather than specifying a potentially misleading oligomer. When necessary the stoichiometry of FlhD and FlhC is specified.

5.3 Determination of YdiV:FlhDC binding

Knowing that *S. enterica* STM1344 (*ydiV*) and FlhDC interacted (Wada *et al.,* 2011), it was important to determine whether the same was true in *E. coli*. Pull-down assays were chosen to analyse this possible interaction, immobilising one protein to a column and adding a second protein to the column. The point of elution of the second protein was carefully analysed, with immediate elution indicative of no interaction, but retention by the column indicating a protein interaction.

Pull-down assays were carried out (Section 2.8.1) immobilising soluble overproduced YdiV-His_6 (Section 4.2.2) onto a HiTrap column, before soluble overproduced $(FLAG)_3$-tagged FlhDC was added (Section 5.2.3). Following washing, an imidazole rich elution buffer was added to the column, to elute all His-tagged proteins. Fractions collected at each stage of the pull-down were separated by SDS-PAGE (Section 2.6.9), one of which was Coomassie-stained and the other probed with a FLAG antibody in a Western blot analysis (Section 2.6.10).

The Coomassie-stained SDS-PAGE gel showed the presence of all proteins at their expected sizes, YdiV-His_6 (lane 1) and the FlhDC components (lane 4) in their respective crude extracts (Figure 5.5A). Subsequent wash steps (lane 6) following FlhDC addition lacked high levels of FlhDC, suggesting a possible interaction between FlhDC and YdiV. This interaction was confirmed by the co-elution of YdiV and FlhDC in the imidazole elution fractions (Figure 5.5A, lanes 7-11).

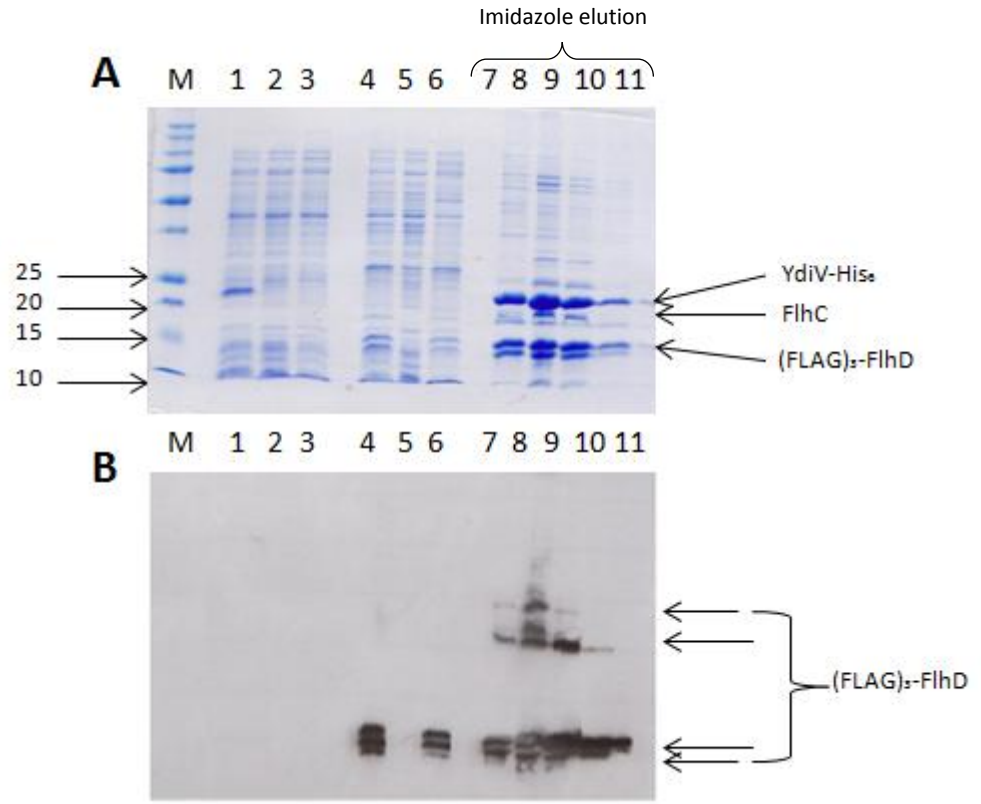


Figure 5.5: Binding interaction between the *E. coli* proteins FlhDC and YdiV

A) Coomassie-stained SDS-PAGE gel of pull-down assay fractions
B) Western blot, probed with a FLAG antibody

The method used is described in Section 2.8.1.

In both cases lanes M: BioRad All blue pre-stained marker (molecular weights shown in KDa), 1: crude soluble YdiV, 2: unbound protein from column, 3: wash step, 4: crude soluble FlhDC, 5: unbound protein, 6: wash step, lanes 7-11: elution fractions.

The YdiV-His_6 (lane 1) and $(FLAG)_3$-FlhDC proteins (lane 4) are present in the crude cell extracts and in the elution fractions (lanes 7-11), as indicated. The expected molecular weights of the proteins are: YdiV-His_6 (28 KDa), FlhC (22 KDa), $(FLAG)_3$-FlhD (16 KDa).

The Western blot with FLAG antibody, shows the presence of $(FLAG)_3$-FlhDC, appearing as multiple species in the elution fractions, as indicated.

Western blot techniques analysed the same pull-down samples, probing with a FLAG antibody to specifically detect $(FLAG)_3$ –FlhD. The Western blot confirmed the FlhDC:YdiV interaction, clearly identifying $(FLAG)_3$ –FlhD in the elution fractions (Figure 5.5B). The $(FLAG)_3$-FlhD protein was detected as multiple species, indicating the formation of different oligomers of FlhD (Figure 5.5B). This is in agreement with the known ability of FlhD to form dimers as well as multimers with FlhC of $FlhD_2C_2$ and $FlhD_4C_2$, in addition to the many oligomeric states of the FlhDC:YdiV complex (Liu & Matsumura, 1992; Wang *et al.,* 2006; Li *et al.,* 2012). As the antibody detects FlhD and given the low levels of eluted FlhC, a possibility and minor concern would be that YdiV is binding to FlhD, rather than the FlhDC complex. Given that some FlhC was detected, we were reasonably confident of the YdiV and FlhDC interaction, mirroring the situation reported for FlhDC and STM1344 in *S. enterica* (Wada *et al.,* 2011). Whilst unknown at the point of this investigation, the YdiV:FlhDC interaction was reported independently by Wada *et al.* (2012).

Crystallisation trials (Section 2.6.15) were carried out for the YdiV:FlhDC complex, in an attempt to elucidate the structure of the protein, as only the $YdiV_2$-$FlhD_2$ structure has currently been solved (Li *et al.,* 2012). The YdiV:FlhDC complex was purified by pull-down assays (Section 2.8.1), the purification confirmed by SDS-PAGE (Section 2.6.9) before dialysis of the complex into a 20 mM sodium phosphate, 500 mM NaCl, pH 7.5 buffer (Section 2.6.7). The protein complex was then concentrated to ~9 mg/ml (Section 2.6.7) and tested for crystallisation with commercial screens (PACT, PEG, MPD, Classics, pH clear and JCSG+) (QIAGEN) (Section 2.6.15). No protein crystals were identified using these screens and therefore this was not further investigation.

5.4 Characterisation of the YdiV and FlhDC binding interaction

The binding interaction between FlhDC and YdiV was characterised kinetically using a BLItz (fortebio) detector and thermodynamically by Isothermal Titration Calorimetry (TA instruments).

The BLItz system detects binding interactions by measuring biosensor thickness (nm) as a reaction proceeds, detecting protein as a physical increase in sensor thickness. Immobilisation of one protein to the biosensor is therefore required, enabling a detectable biosensor change upon protein:protein interactions. Isothermal Titration Calorimetry (ITC)

measures binding interactions by detecting the heat change upon titration of one species into a vessel containing its binding partner.

To study the FlhDC and YdiV protein interaction, both proteins were purified as His_6-tagged constructs (Sections 4.2.3 and 5.2.3) and dialysed in sodium phosphate buffers (Section 2.6.7). Absorbencies were recorded at 280 and 320 nm for both proteins and were used with calculated extinction coefficient values (ExPASy ProtParam tool) to determine molar concentrations (Section 2.6.8). The protein purity was crucial to accurate kinetic and thermodynamic calculations and therefore proteins were analysed by SDS-PAGE (Section 2.6.9) to check for protein purity. The SDS-PAGE gels showed high levels of protein purity for both the YdiV and FlhDC protein, with very little contaminating proteins (Figure 5.6). These results gave confidence in the calculated protein concentrations, thereby allowing kinetic and thermodynamic assays to be pursued.

For the BLItz assay (Section 2.8.3), chemical immobilisation of one protein to an AR2G biosensor was carried out by forming covalent amide bonds between primary amine groups of the protein and the carboxy-terminated biosensor surface. A second protein was subsequently added in solution and the protein interactions studied. Protein immobilisation is a major limitation of this technique, potentially shielding interaction interfaces. To reduce any experimental errors caused by this, both proteins were immobilised to the biosensor in turn and tested with the other partner in solution. To supplement these experiments, ITC analysed the protein interaction in solution.

5.4.1 Characterisation of YdiV binding to the immobilised FlhDC protein

The FlhDC complex was immobilised to an AR2G biosensor at a fixed concentration of 13.2 µM and YdiV protein was added at five different concentrations ranging from 16 µM to 16 nM (1000-fold range).

The biosensor thickness was recorded, identifying the points at which proteins were associating and dissociating with/from the biosensor (Figure 5.7). Five differently coloured traces are present on the figure, indicative of the five different YdiV concentrations added to the immobilised FlhDC.

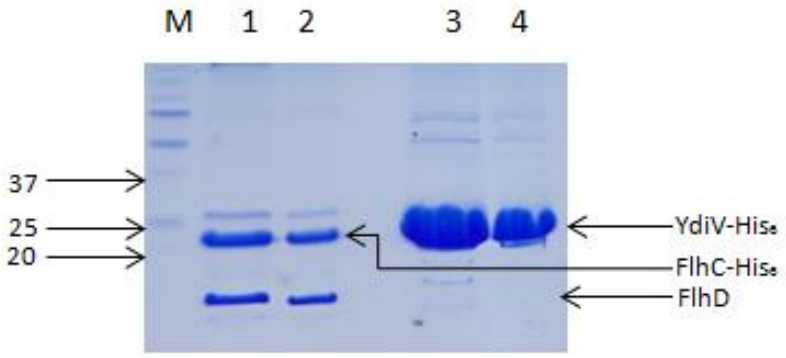


Figure 5.6: Assessment of FlhDC and YdiV protein purity for kinetic and thermodynamic interaction analysis

SDS-PAGE gel of purified protein samples. M: Precision plus protein standard (molecular weights shown in KDa), 1 + 2: purified FlhDC-His_6 (44 µM), 3 + 4: purified YdiV-His_6 (235 µM). Lanes 1 and 3 contain 10 µl of each purified protein, whilst lanes 2 and 4 contain 8 µl of protein. The concentrations stated above are those determined by using extinction coefficient values (Section 2.6.8).

The locations of proteins are indicated, at the expected molecular weights: YdiV-His_6 (28 KDa), FlhC-His_6 (22 KDa), FlhD (13 KDa).

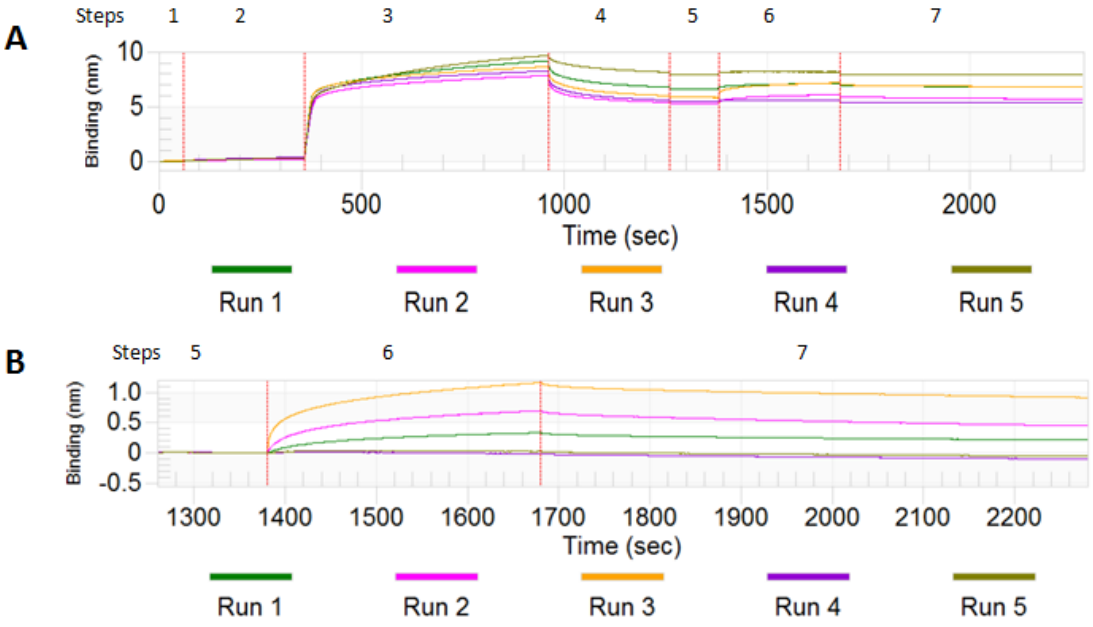


Figure 5.7: BLItz binding spectra for YdiV binding to immobilized FlhDC

Binding spectra for a series of experiments in which a fixed concentration of FlhDC was immobilised to a biosensor and soluble YdiV was added to the system. The binding reaction was composed of 7 discrete steps; 1: initial baseline, 2: chemical activation, 3: loading, 4: quenching, 5: baseline, 6: association and 7: dissociation.

A) Biosensor binding (nm) traces recorded for the entire duration of each experiment, as the reactions proceeded (sec).
B) Biosensor binding (nm) trace for the final three steps, enlarging the association and dissociation curve data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

The concentration of FlhDC was 13.2 µM, with YdiV concentrations varying for each of the five runs, identified by different coloured traces. The concentrations of YdiV tested were; run 1: 1.6 µM, run 2: 3.2 µM, run 3: 16 µM, run 4: 0.16 µM and run 5: 0.016 µM.

All five traces show biosensor thickening (step 3) indicative of FlhDC immobilisation to the biosensor (Figure 5.7A). Whilst excess protein was released in the quenching stage (step 4), the thickness of the biosensor remained above baseline, confirming that protein was bound to the sensor. Upon YdiV addition (step 6), association curves were seen (of varying gradients due to the varying YdiV concentrations) showing a distinct increase in binding (Figure 5.7B). The dissociation step (step 7) caused the binding thickness' to decrease, but gradually rather than instantaneously, indicating that the proteins are slowly dissociating. As can be seen, the higher YdiV concentrations (i.e. runs 2 and 3) elicited the largest change in amplitude of the association curve, indicative of thicker biosensors being formed, whilst lower YdiV concentrations (runs 4 and 5) produced numerically fewer YdiV and FlhDC interactions and therefore produced a smaller amplitude change in the association curve (Figure 5.7B).

Using the association and dissociation data, binding kinetics were calculated using the BLItz fortebio system. Global kinetic analysis gave an average set of kinetics, taking into account the values from all five experiments (as none were obviously outliers). The K_D value was 286 nM, with an association rate (k_a) of $1.539 \times 10^3 \pm 4.9 \times 10^1$ 1/Ms and a dissociation rate (k_d) of $4.41 \times 10^{-4} \pm 3.1 \times 10^{-5}$ 1/s. The K_D value indicates strong affinity binding between the proteins, generated by a high association and a low dissociation rate. To give a sense of scale, affinity constants for weak non-specific protein interactions are usually mM, antibody-antigen K_D values are in the region of 10-100 nM, whilst the highly specific streptavidin-biotin interactions have K_D values in the pM or even fM range (Green, 1975).

5.4.2 Characterisation of FlhDC binding to the immobilised YdiV protein

The analysis was then repeated using the two proteins in opposite roles. YdiV was immobilised to the biosensor at a concentration of 6.9 μM, adding FlhDC at concentrations ranging between 41 μM and 41 nM, again a 1000-fold range.

YdiV immobilisation was confirmed by the association curve (step 3) and raised baseline (step 5) following loading of YdiV to the AR2G surfaces (Figure 5.8A). Upon FlhDC addition (step 6), association curves were produced (again of varying gradients due to the varying FlhDC concentrations) showing a distinct increase in biosensor thickness (Figure 5.8B). The

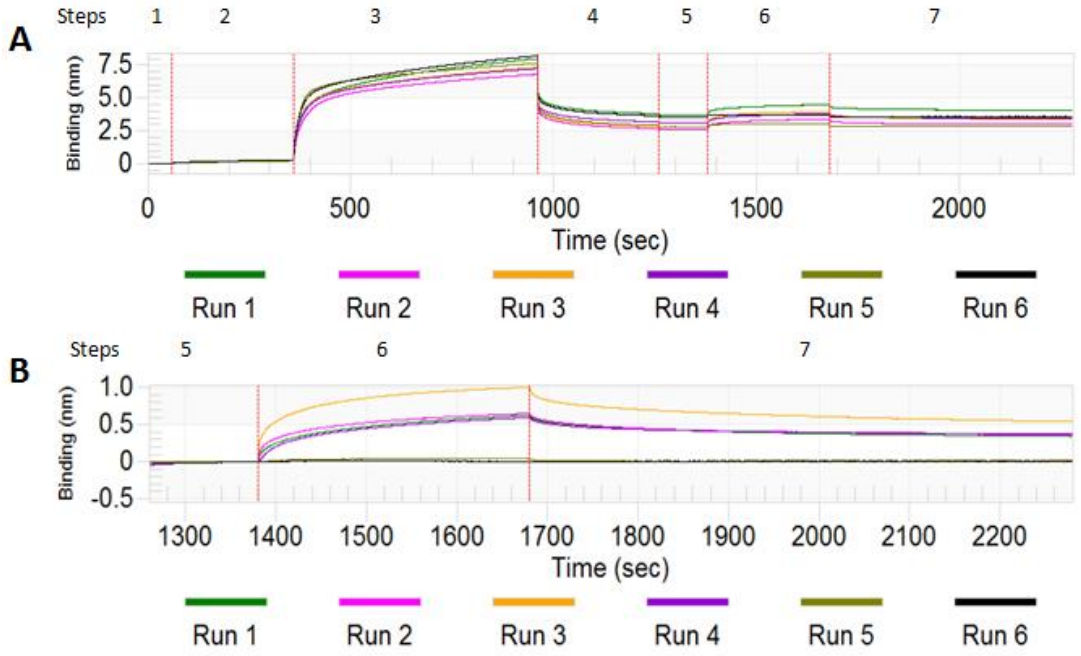


Figure 5.8: BLItz binding spectra for FlhDC binding to immobilized YdiV

Binding spectra for a series of experiments in which a fixed concentration of YdiV was immobilised to a biosensor and soluble FlhDC was added to the system. The binding reaction was composed of 7 discrete steps; 1: initial baseline, 2: chemical activation, 3: loading, 4: quenching, 5: baseline, 6: association and 7: dissociation.

A) Biosensor binding (nm) trace recorded for the entire duration of each experiment, as the reactions proceeded (sec).
B) Biosensor binding (nm) trace for the final three steps, enlarging the association and dissociation curve data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

YdiV was used at 6.9 µM, with FlhDC concentrations varying for each of the six runs, identified by differently coloured traces. The concentrations of FlhDC tested were; run 1: 4.1 µM, run 2: 8.2 µM, run 3: 41.3 µM, run 4: 412 nM, run 5: 41 nM and run 6: 206 nM.

dissociation step (step 7) again showed a gradual decrease in binding thickness, indicating protein dissociation (FlhDC) from the biosensor but a significant proportion remained bound to the YdiV coated biosensor (Figure 5.8B).

Once again, individual kinetic values were computed by the BLItz fortebio system, disregarding run 6 when calculating an average value, due to its poorly fitted analysis curve. The affinity constant K_D was calculated at 395 nM, once again a nM affinity, indicative of a strong specific binding interaction. The association constant (*k*a) was measured at $3.17x10^3 \pm 1.15x10^2$ 1/Ms and the dissociation constant (*k*d) calculated at $1.25x10^{-3} \pm 2.77x10^{-5}$ 1/s. Therefore the rate of protein association was high and the dissociation rate was low.

Taken together, the two global K_D values of 286 nM and 395 nM indicate high affinity binding between YdiV and FlhDC. The calculated *k*a and *k*d errors are relatively low, at between 2-7%, providing confidence in the global K_D values. The variation between the K_D values could be due to the artificial nature of protein immobilisation, potentially shielding binding interfaces and altering affinities.

5.4.3 Characterisation of YdiV and FlhDC interactions in solution

The binding interactions between YdiV and FlhDC complex were also studied in solution, by Isothermal Titration Calorimetry (Section 2.8.4). Both proteins were purified with a C-terminal His_6-tag (Sections 4.2.3 and 5.2.3) before dialysis of proteins into sodium phosphate dialysis buffer and determination of protein concentrations, as detailed in Section 5.4.

Following system cleaning and equilibration, FlhDC (44 µM) was loaded into the sample chamber and YdiV (235 µM) into the syringe. Titrations were then carried out at 25^{o}C with 250 rpm stirring in the sample chamber. There were 25 injections of YdiV into the chamber, each of 2 µl volume, spacing the injections 180 s apart to enable restoration of constant temperature between injections. Following completion of the experiment, and repetition in triplicate, thermodynamic analysis was undertaken. Throughout the duration of the experiment, the fluctuations in the chamber temperature were recorded, and could be corrected to a baseline. These corrected heat changes revealed an exothermic reaction

between FlhDC and YdiV, with temperature increases produced upon YdiV injection (Figure 5.9). These heat changes were then related to the protein concentrations, calculating exact concentrations of each protein as the reaction proceeded, and plotted to give a graph of final heat changes Q (μJ) against protein molar ratios. In order to calculate the thermodynamic values of the reaction, the heat changes detected were inputted into a model fitting program (Figure 5.9).

The FlhDC and YdiV titration heat changes were applied to a range of models, but the independent model gave the closest fit and was therefore deemed the most likely binding model. A single binding curve was observed, with an n value of 0.8, indicating that a 1:1 stoichiometry of $FlhD_4C_2$ and YdiV. This suggests that only one YdiV binds to the $FlhD_4C_2$ complex under these conditions, potentially requiring a trigger for additional binding of YdiV to the $FlhD_4C_2$ complex. The K_D value of 158 nM was similar to the 286-395 nM values calculated from BLItz analysis. The binding of YdiV to FlhDC was enthalpy and entropy driven (ΔH = -34.1 KJ/mol; ΔS = 24.9 J/mol). The ΔG for the reaction was -41.5 KJ/mol. These measurements do not support a recently proposed model in which four YdiV molecules bind sequentially to $FlhD_4C_2$, with addition of the 1st or 2nd YdiV units having no disruptive effect on the $FlhD_4C_2$-DNA interactions (Figure 5.10). However, further addition of the 3rd or 4th YdiV subunits were speculated to disrupt the conformation of the $FlhD_4C_2$ ring structure and trigger DNA dissociation (Li *et al.,* 2012).

5.5 Production of YdiV and FlhDC antibodies

Following characterisation of the YdiV:FlhDC complex, it was of interest to determine the conditions under which these proteins were expressed in *E. coli* and therefore when the complex would be likely to form. Antibodies against these proteins were raised, foreseeing the usefulness of these to analyse protein levels.

Polyclonal antibodies were raised against YdiV and FlhDC proteins, by Bioserve (University of Sheffield). Purified and dialysed His-tagged proteins were produced for each protein (Sections 4.2.3; 5.2.3 and 2.6.7), and used to stimulate antibody production in rabbits (Section 2.6.11).

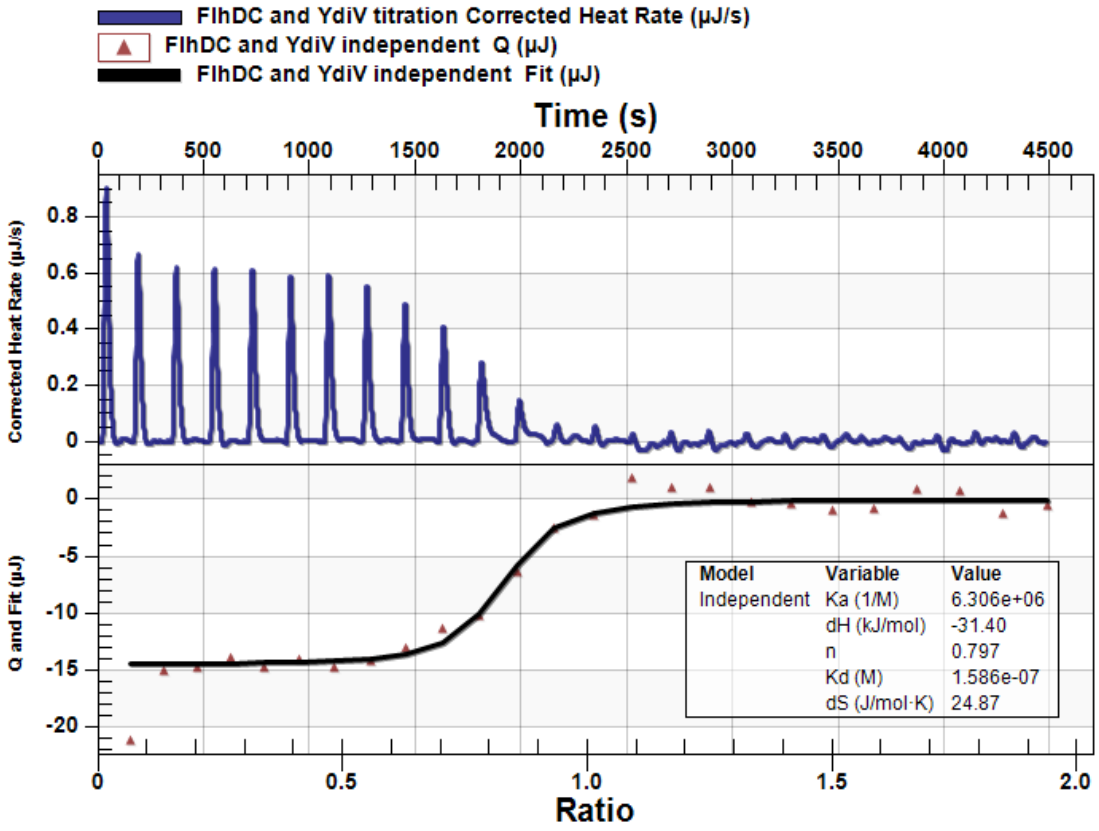


Figure 5.9: Isothermal Titration Calorimetry curve for the FlhDC YdiV titration

The top graph shows the corrected heat changes (µ/Js) for the FlhDC YdiV titration, with addition of YdiV clearly producing an exothermic reaction, denoted by an increase in chamber temperature.

Using the known concentrations of the proteins, the molar ratio of proteins interacting throughout the experiment was determined (lower graph). These points were then fitted to an independent binding site model, with the fitted curve shown. Using this fitted curve, the FlhDC and YdiV binding kinetics and molar stoichiometries were determined, shown in the figure.

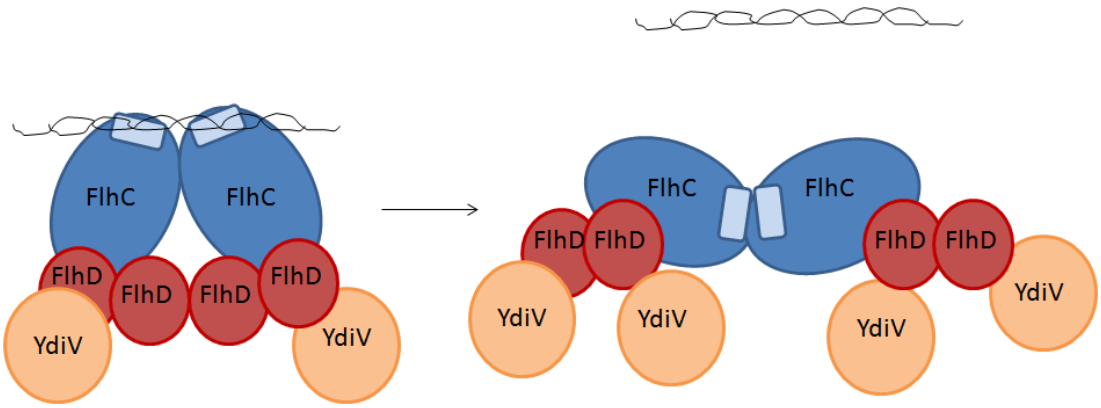


Figure 5.10: Schematic of the proposed Li *et al.* (2012) model for YdiV:$FlhD_4C_2$ interaction

Diagram to show the model of YdiV binding to the $FlhD_4C_2$ complex, as proposed by Li *et al.* (2012). This model shows a ring-structure of $FlhD_4C_2$ with DNA binding to the FlhC subunits via specific DNA-binding regions (shown in light blue). This model proposes a sequential binding interaction with YdiV, suggesting that two YdiV molecules can bind to $FlhD_4C_2$ via the outer-most FlhD units. Addition of the initial two YdiV molecules is suggested to be undisruptive to the ring structure of $FlhD_4C_2$, thereby retaining the DNA-binding ability of the complex. However, upon binding of further YdiV molecules to the inner FlhD subunits (totalling three or four), the ring-like $FlhD_4C_2$ structure is broken due to spatial restrictions. The FlhC units rotate inwards, essentially concealing the DNA-binding regions and therefore the complex is no longer able to bind DNA.

Modified from Li *et al.* (2012).

5.5.1 Bleed specificity tests

The proteins were injected into separate rabbits, producing a primary immune response against the proteins, before a second round of injections was carried out to give a heightened secondary response. Small test bleeds were taken at both stages, and Western blot analyses (Section 2.6.10) were carried out to detect the presence of FlhDC and YdiV antibodies. Samples of rabbit serum which had not been subject to antigen challenge were used as control samples.

Following the first round of antigen exposure, the rabbit antisera were used to probe for the relevant target protein (YdiV or FlhDC) using purified YdiV and crude soluble FlhDC as the target proteins. Equally loaded protein samples (Figure 5.11A) were probed by various dilutions of the rabbit sera in Western blot analyses (Figure 5.11B). Importantly, the rabbit sera contained the relevant antibodies, detecting YdiV protein (left) and FlhDC as the FlhD and FlhC proteins individually (right). This specificity was confirmed by an additional Western blot, consisting of YdiV and FlhDC proteins probed with non-stimuated rabbit sera, which failed to detect these proteins (Figure 5.11C).

Following a second round of antigen exposure, rabbit serum samples were taken once more, hoping to detect a heightened response to the protein. The anti-YdiV and anti-FlhDC sera were probed against purified protein samples (0.05 mg/ml of YdiV or FlhDC respectively) at four different dilution levels (1/100, 1/500, 1/1000 and 1/5000). For both proteins, the respective antibody clearly detected the protein of interest, with the Western blot detection becoming cleaner and better defined when using more dilute antibody (Figure 5.12). Significantly, the sera show stronger detection and therefore affinity for the proteins, compared to the first injection sera. Once more, the purified protein samples were probed with rabbit control serum and the resulting Western blots (Figure 5.12C) lacked detection of either protein, showing that the response in Figure 5.12B was a targeted and specific response.

Due to the required level of specificity being reached after the second injection, the rabbits were not further challenged, but left for blood levels to be restored, before terminal bleeds were taken, which were used for all future antibody work.

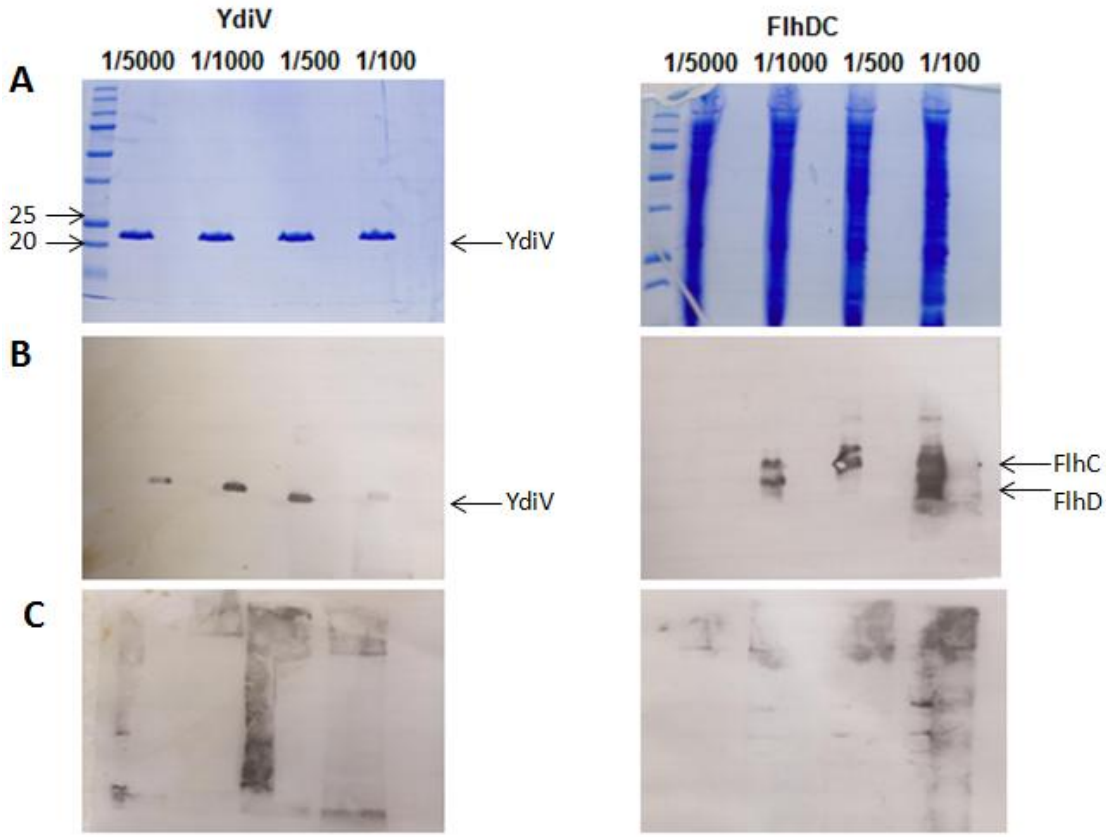


Figure 5.11: Primary immune response antibody specificity test

Following addition of the initial injection of YdiV or FlhDC antigens into rabbits, small samples of rabbit antisera were removed. These antisera were tested for specificity against YdiV and FlhDC.

A) Coomassie-stained SDS-PAGE gel of protein (either purified YdiV or crude soluble FlhDC) with All blue marker on the far left hand sides of the gel (molecular weights shown in KDa).
B) Western blot using rabbit second bleed serum at various dilutions.
C) Western blot using rabbit control serum (from non-injected rabbits) at various dilutions.

In all cases, the antibody dilution was as specified and the locations of YdiV and FlhDC are as indicated.

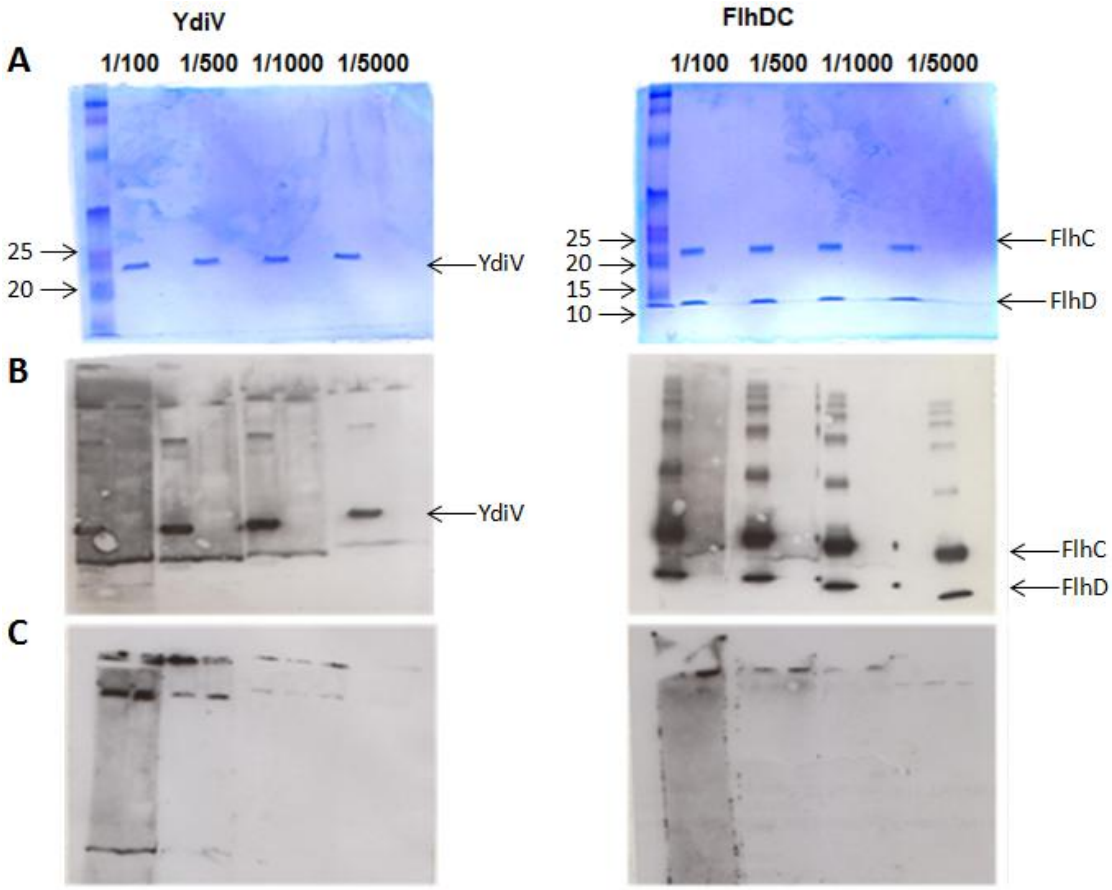


Figure 5.12: Secondary immune response antibody specificity test

Following a second antigen exposure, the rabbits were expected to produce a heightened secondary response. A sample of antiserum was removed and tested against purified YdiV and FlhDC (antigen) samples, to determine the specificity of the antibodies.

A) Coomassie-stained SDS-PAGE gel of 0.05 mg/ml purified protein (YdiV or FlhDC as specified) with the All-blue marker on the far left hand side of each gel (molecular weights indicated in KDa).
B) Western blot using rabbit second bleed serum at various dilutions.
C) Western blot using rabbit control serum (from non-injected rabbits) at various dilutions.

In all cases, the antibody dilution was as specified and the protein positions are as indicated.

5.5.2 Terminal bleed antibody sensitivity tests

To test the sensitivity of the final bleed antisera, the sera were used at dilutions of 1/5000 to probe antigen samples at 0.05 mg/ml and 0.01 mg/ml via Western blot. Strong detection was seen for both antigens, with the antisera identifying even the weakest 0.01 mg/ml protein levels (Figures 5.13 and 5.14). In the case of YdiV, the detection was very clean giving a single species (Figure 5.13). However, the FlhDC antiserum detected multiple species, which were consistent with the monomeric proteins FlhD and FlhC, as well as homo- and hetero-oligomers of these proteins (Figure 5.14), which are known to be formed (Liu & Matsumura, 1992; Wang *et al.,* 2006). Therefore, for future work, a 1/5000 dilution of both antisera was adopted as to be a suitable starting concentration.

5.6 Determination of YdiV and FlhDC concentrations at different growth rates

The effect of *E. coli* growth rates on *ydiV* and *flhDC* expression levels were analysed in an attempt to identify conditions in which FlhDC and YdiV protein levels were at their highest. Chemostat samples were provided by Thomas Curran, in which *E. coli* MG1655 cells were cultured in Evans minimal medium at pH 6.95, 37^{o}C and with 400 rpm stirring at different dilution rates (Sections 2.2.1 and 2.2.2). Four dilution rates were analysed (0.05, 0.1, 0.2 and 0.5 /h), creating approximately a 10-fold range of bacterial doubling times.

5.6.1 Analysis of YdiV expression levels

YdiV protein levels were analysed using rabbit polyclonal antibodies to probe the *E. coli* cell samples specified above. The resultant SDS-PAGE (Section 2.6.9) and Western blot (Section 2.6.10) showed no detection of YdiV in the samples, and as a consequence, indicate that the protein must be present at very low levels in the cell (Figure 4.21).

5.6.2 Analysis of FlhDC expression levels

To detect the FlhDC expression levels, the chemostat growth rate samples were again analysed, alongside the *E. coli* strains MG1655 (wild-type) and JRG6577 (an *flhDC* mutant)

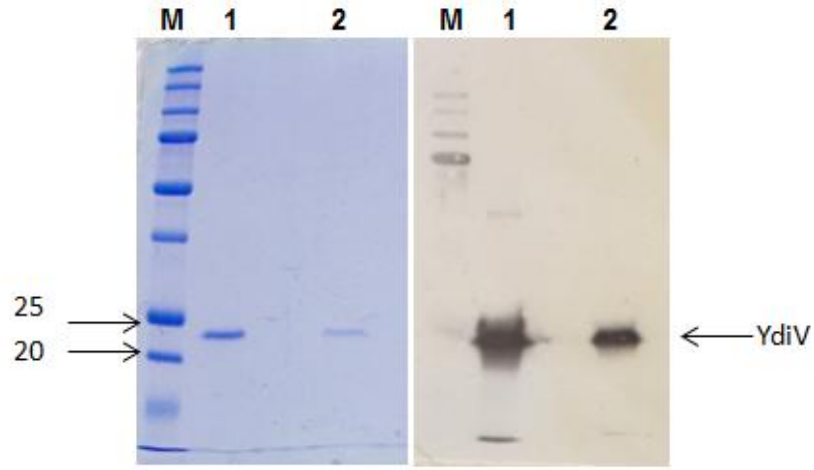


Figure 5.13: Anti-YdiV terminal bleed sensitivity test

Coomassie-stained SDS-PAGE gel (left) and Western blot (right) assessing the sensitivity of the rabbit antiserum against the YdiV antigen.

In both cases, M: BioRad all blue prestained marker (molecular weights indicated in KDa). Lane 1: Concentrated YdiV (0.05 mg/ml) 2: Diluted YdiV (0.01 mg/ml). The Western blot (R) analyses the same SDS-PAGE gel lanes, but is probed with the YdiV rabbit antiserum (at a dilution of 1/5000). The YdiV protein (of expected molecular weight 28 KDa) was clearly detected by Western blot as indicated.

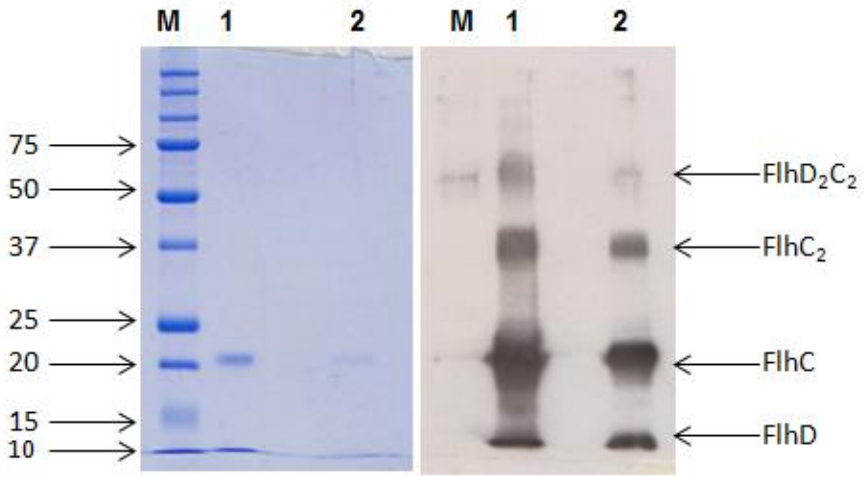


Figure 5.14: Anti-FlhDC terminal bleed sensitivity test

Coomassie-stained SDS-PAGE gels (left) and Western blot (right) assessing the sensitivity of the rabbit antiserum against the FlhDC antigen.

In both cases, M: BioRad all blue prestained marker (molecular weights indicated in KDa). Lane 1: Concentrated FlhDC (0.05 mg/ml) 2: Diluted FlhDC (0.01 mg/ml). The Western blots (R) analyses the same SDS-PAGE gel lanes but is probed with the FlhDC rabbit antiserum (at a dilution of 1/5000). The location of FlhC and FlhD are indicated at the expected molecular weights of 22 and 13 KDa respectively, as well as predicted $FlhC_2$ and $FlhD_2C_2$ species (expected molecular weight of 44 and 68 KDa respectively).

which were grown in LB medium at 37^{o}C (Section 2.2.2) (Table 2.1). Of importance, strain JRG6577 was grown in the absence of kanamycin, in order to lose the complementation plasmid containing *flhDC* (Table 2.1). The chemostat and control cell samples were then diluted to give comparable cell density, and were analysed on an SDS-PAGE gel (Section 2.6.9) and by Western blot (Section 2.6.10), alongside a reference sample of purified FlhDC protein at 0.01 mg/ml concentration (Section 5.2.3). Polyclonal antibodies raised against FlhDC (Section 5.5.2), were used in the Western blot analysis, initially at 1/5000 dilution before increasing the dilution to 1/10,000 to reduce the non-specific background.
Equal loading of the samples was confirmed by SDS-PAGE analysis (Figure 5.15A). Unlike YdiV, Western blot analysis showed expression of FlhDC (Figure 5.15B). The purified FlhDC sample which is barely detectable in the PAGE gel, was clearly detected by the Western blot and enabled identification of the protein in the test samples. The wild-type sample (lane 1) gave a weak species of the expected size, which was absent in the Δ*flhDC* sample (lane 2). In the chemostat samples, a clear species is present in all cases, the intensity of which appears to peak at a dilution rate of 0.2 /h before decreasing slightly at 0.5 /h.

The purified FlhDC sample yielded a faint species at 13 KDa (indicative of FlhD) in the Coomassie-blue stained SDS-PAGE gel (Figure 5.15A). The Western blot, however, identifies multiple species in lane 3, signifying the presence of both monomeric components of the FlhDC protein sample, as well as several of the expected oligomeric states of the complex (Figure 5.15B). For ease of analysis, only the FlhD component was quantified but is directly representative of the FlhDC complex levels.

The intensity of the species seen by Western blot, was quantified by Image-J (Section 2.10.1) to show the relative expression levels of FlhDC (Figure 5.16). Of interest, the wild-type levels of FlhDC are much lower than all the chemostat samples, presumably as a result of differing strain and growth conditions. Whilst the wild-type strain was grown in LB medium at exponential phase, the chemostat samples were grown continuously in minimal medium and samples taken at steady state.

Comparing the *E. coli* growth rate samples, the levels of FlhDC increased almost 2-fold between the 0.05 and 0.1 /h dilution rates, indicating that an increased bacterial growth

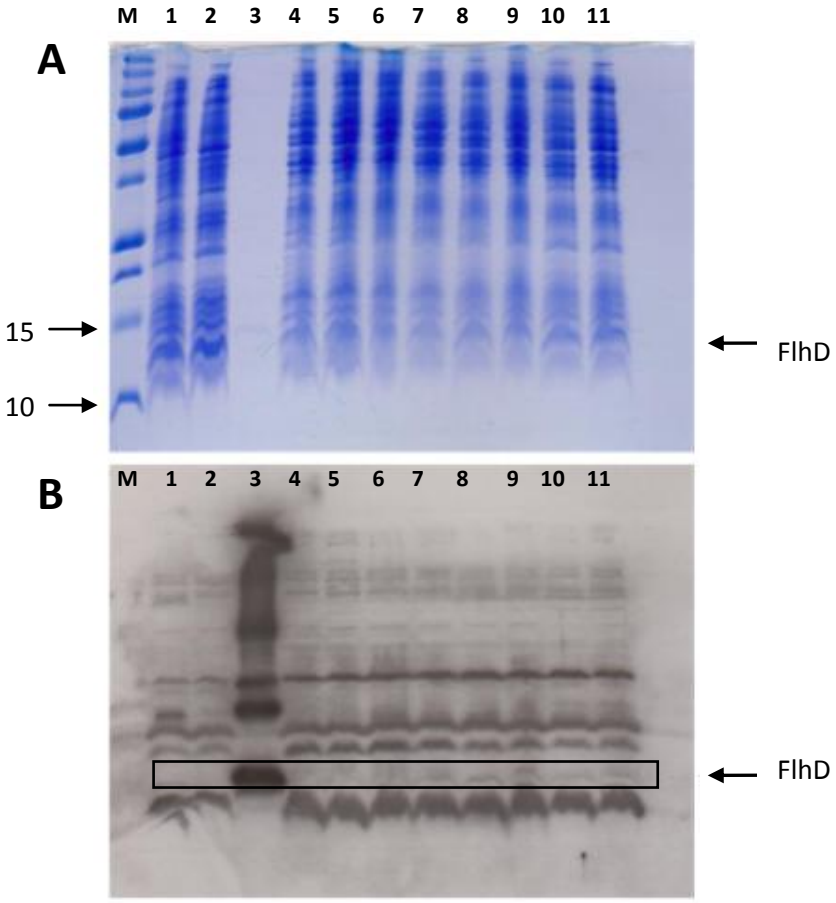


Figure 5.15: FlhDC expression changes at various growth rates

Samples of *E. coli* MG1655 chemostat cultures maintained at different dilution rates were analysed to identify whether FlhDC was up and down-regulated.

Cell samples were analysed by A) Coomassie-stained SDS-PAGE and B) Western blot analysis, probed with FlhDC antiserum.

For each sample, the dilution rates (1/h) are denoted simply by numbers, i.e. 0.05 to 0.5, as explained previously, and the two biological replicates denoted by #1 or #2.

M:BioRad Prestained marker (molecular weights shown in KDa), 1: *E. coli* MG1655, 2: *E. coli* JRG6577 (an *flhDC* mutant) 3: 0.01 mg/ml purified FlhDC protein, 4: MG1655 0.05 #1, 5: MG1655 0.05 #2, 6: MG1655 0.1 #1, 7: MG1655 0.1 #2, 8: MG1655 0.2 #1, 9: MG1655 0.2 #2, 10: MG1655 0.5 #1, 11: MG1655 0.5 #2.

The location of FlhDC (specifically FlhD) is indicated.

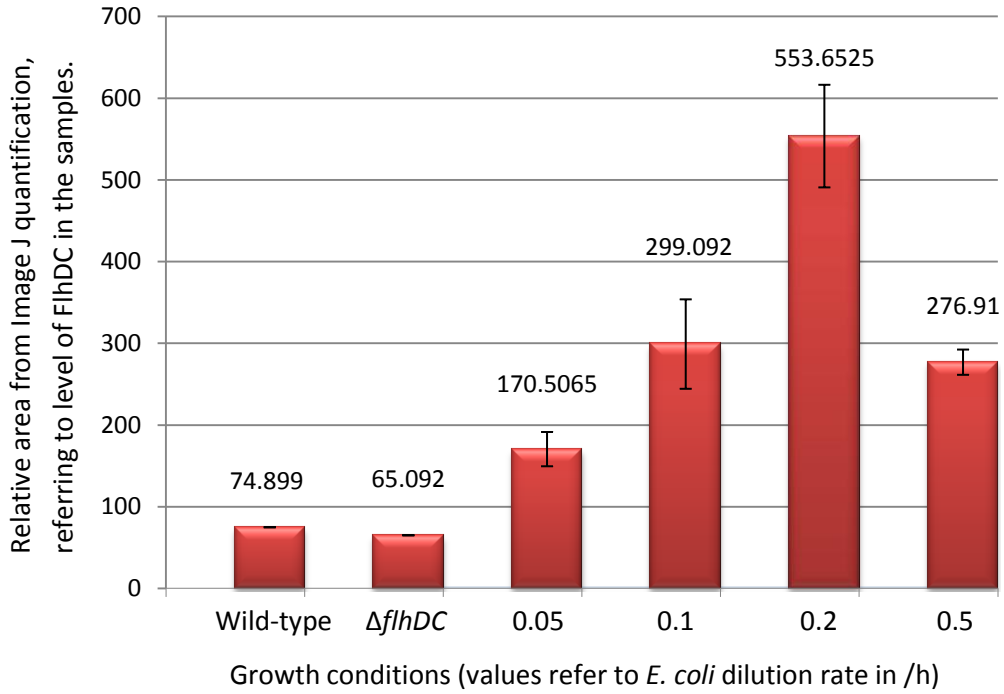


Figure 5.16: Relative intensities of FlhDC expression, as quantified by Image-J software

The relative intensities of FlhDC expression, determined from Western blot analysis are shown for each of the different *E. coli* growth conditions.

Numerical values 0.05, 0.1, 0.2 and 0.5 refer to the chemostat dilution rates (1/h), which correspond to the *E. coli* MG1655 growth rates. For each of the chemostat samples, two biological replicates were taken with an average quantified intensity plotted on the chart. The quantified level of variation between the replicates is shown by error bars.

The numbers above each bar indicate the relative intensities of FlhDC detected in each condition.

rate caused the concentration of FlhDC to be up-regulated (Figure 5.16). An approximately equal transition was present between the 0.1 and 0.2 /h dilution samples, signifying even higher levels of FlhDC at the 0.2 /h dilution rate. However, the protein levels peaked at this dilution rate, with a decrease in protein quantities in the 0.5 /h dilution rate sample. In real terms, the quantity of FlhD in the maximal growth conditions has been calculated at ~5 ng per 0.22 OD_{600} units, using the intensity of the known concentration FlhDC sample as a reference. If 1 unit of cells at OD_{600} is equivalent to 5 $x10^8$ - 1 $x10^9$ cells, this equates to ~0.03 fg of FlhDC per cell.

Therefore, these data demonstrate that the optimal growth rate for FlhDC production is equivalent to a dilution rate of 0.2 /h, equivalent to a 3.5 h doubling time. Future work could investigate the mechanism by which this maximal FlhDC production is produced; whether gene transcription or translation are maximally activated or whether protein degradation rates are maximally decreased.

5.7 DNA-binding capacity of FlhDC in the presence of YdiV

Electromobility shift assays (Section 2.9.2) were carried out to investigate the binding of the transcription factor FlhDC to flagellar type II genes in the presence and absence of the anti-FlhDC factor YdiV.

The entire promoter region of the *fliD* gene (P*fliD*) was chosen as an example of a class II gene and contained two FlhDC recognition sites or 'binding boxes' that constitute the FlhDC binding site. This promoter region was amplified from *E. coli* MG1655 genomic DNA using primers NSW27 and NSW28 (Table 5.2) by PCR (Section 2.3.3).

Table 5.2: Primers for amplification of P*fliD* for EMSA assays

Primer	Sequence	Function
NSW27	TTTTTCTAGAGATTCGTTATCCTATATT GCAAGT	Forward primer for P*fliD* amplification
NSW28	TTTTGCGATTTCCTTTTATCTTTCG	Reverse primer for P*fliD* amplification

The sequence shaded in grey indicates an XbaI site.

The amplified DNA was purified, digested using XbaI enzyme and subsequently purified and quantified (Sections 2.3.10; 2.3.6 and 2.3.8). The DNA fragment (150 ng) was radiolabelled using [α^{32}P]-dCTP and purified once more prior to use in EMSAs (Section 2.9.1). Reaction samples were produced containing radiolabelled P*fliD* with increasing quantities of FlhDC and subsequently YdiV (Section 2.9.2). Both proteins were purified with C-terminal His_6-tags and were used directly in sodium phosphate elution buffers (Sections 4.2.3 and 5.2.3). The proteins (of specified concentrations) were incubated with ~ 2 nM radioactively labelled P*fliD*, before being incubated at 37^{o}C for 30 min. Following this, the samples were mixed with EMSA loading dye (Table 2.14) and electrophoresed to separate the free P*fliD* DNA from the protein-DNA complexes.

The EMSA showed that FlhDC (at concentrations from 9 µM) bound at P*fliD*, but in the presence of YdiV (at 7.5 µM and above) was no longer able to bind P*fliD* (Figure 5.17). This indicates that the formation of the FlhDC:YdiV complex inhibits the formation of the FlhDC:P*fliD* complex, with YdiV causing the dissociation of P*fliD* DNA. This switch in FlhDC binding partners (from P*fliD* DNA to YdiV), revealed that the interactions are competitive to some extent, possibly via the DNA and YdiV interacting at the same site in the FlhDC complex, or inducing a change to FlhDC which prevents both partners binding simultaneously.

This analysis confirms that the *E. coli* YdiV:FlhDC interaction has the same DNA-binding properties as in *S. enterica* (Wada *et al.,* 2011). Whilst undertaking this study, the same conclusion was reached by competing groups, who also found YdiV to be an anti-FlhDC factor (Wada *et al.,* 2012; Li *et al.,* 2012).

Calculating the stoichiometry, the initial addition of YdiV at 4 µM yields an ~ 1:4 ratio between YdiV:FlhDC and at this molar ratio, the DNA remained shifted (Figure 5.17). Addition of 7.5 µM YdiV to samples (producing a 1:2 molar ratio of YdiV:FlhDC) however caused a complete lack of DNA shifting, confirming complete DNA dissociation. Competing groups such as Wada *et al.* (2012) and Li *et al.* (2012) also calculated the YdiV:FlhDC molar ratios necessary for DNA dissociation from FlhDC. Wada *et al.* (2012) found that DNA remained shifted in a 1:1 ratio between YdiV:FlhDC, but increasing the ratio to 2.5:1 caused DNA dissociation from the FlhDC complex. A similar stoichiometry was found by Li *et al.* (2012) who found that below a 2.5:1 molar ratio of YdiV:FlhDC, DNA binding remained but

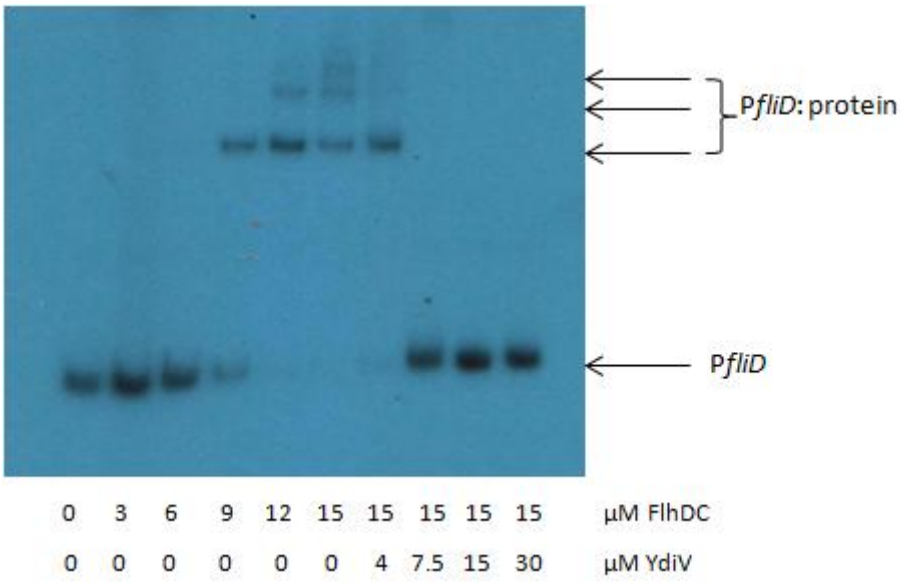


Figure 5.17: EMSA of radiolabelled P*fliD* DNA with varying quantities of FlhDC and YdiV proteins

Exposed autoradiography film, showing the labelled DNA in the presence of purified FlhDC and YdiV, with protein concentrations as stated.

Free P*fliD* DNA and bound protein:P*fliD* species are indicated, as well as the multiple protein:P*fliD* species.

a 3:1 ratio of YdiV:FlhDC protein prevented the DNA shift entirely. These stoichiometry variations could feasibly be a result of different reaction buffer conditions as well as different affinities of FlhDC for the various promoter regions. The data here are for the P*fliD* region, whilst the P*flhB* and P*fliA* regions were used in the Li *et al.* (2012) and Wada *et al.* (2012) studies respectively.

Noteably, the EMSA showed three distinct species of protein:P*fliD* complexes in the shifted region indicating that FlhDC and P*fliD* formed three different interactions (Figure 5.17). The P*fliD* DNA region contains both 'FlhDC-binding boxes' which are characteristically present in all Class II flagella gene promoters for FlhDC recognition (Claret & Hughes, 2002; Wang *et al.,* 2006). It is believed that the promoter motifs act as one single binding site, consisting of two FlhDC binding boxes and the interlinking region, all of which is necessary to bind a single $FlhD_4C_2$ (Claret & Hughes, 2002). Both FlhC and FlhD are necessary to produce a stable DNA-binding complex, with much higher concentrations of FlhC singularly required to bind to DNA (Claret & Hughes, 2002). Therefore a single FlhDC functional species is the most likely protein interacting with the P*fliD* rather than singular FlhD or FlhC. One possible reason for the multiple shifted species could be a heterogeneous mixture of FlhDC complexes in solution, producing a mixture of $FlhD_2C_2$-DNA and $FlhD_4C_2$-DNA interactions which would potentially account for two of the three species. Equally, it is unknown whether essentially 'half' of the FlhDC complex binds to one of the FlhDC-binding boxes, and the other 'half' to the other box. If this were the case, then there is a possibility of one FlhDC-binding box being occupied whilst the other is unoccupied, which would again produce different EMSA shifted species. To clarify this result, techniques such as gel filtration or analytical centrifugation could be used to determine the identity of the three distinct shifted species and determine precisely which proteins are involved in each case. Additionally, segmenting the P*fliD* DNA into multiple fragments would provide useful information, identifying whether any form of FlhDC could bind to just one of the binding boxes or whether both were essential.

An alternative and perhaps more likely explanation, is the presence of low homology binding sequences in the P*fliD* DNA fragment which also interact with the FlhDC protein. Obviously the 'FlhDC-binding box' is the prefential site for FlhDC binding, and would produce the lower shifted species which is produced initially. Once the protein concentration increases, the less mobile species are produced, as a result of binding at

alternative lower affinity sites. In order for this to be true, there would need to be two lower affinity FlhDC-binding sites in the P*fliD* region.

5.8 Characterisation of binding kinetics of DNA:FlhDC interaction

BLItz analysis was utilised once more to kinetically analyse the binding of FlhDC to P*fliD* DNA (Section 2.9.5), as shown previously by EMSA analysis (Figure 5.17).

The interactions between purified FlhDC and the "target" DNA region of P*fliD* were tested alongside the interactions between FlhDC and "control" DNA, which lacked the 'FlhDC-binding box motifs'. In addition purified YdiV was tested against both DNA fragments, acting as a protein control predicted to have no interaction with the DNA.

Both target and control DNA fragments were synthesised to produce 5' biotinylated dsDNA fragment of 160 bp long. The target DNA corresponded to a region in the *fliD* promoter (P*fliD*) including both FlhDC binding boxes, whilst the control DNA corresponded to a region further upstream of the P*fliD* sequence lacking the FlhDC binding boxes. The regions were amplified from *E. coli* MG1655 genomic DNA by PCR (Section 2.3.3) using primer pairs NSW84 and NSW28 (target) and primers NSW85 and NSW86 (control DNA) (Table 5.3).

Table 5.3: Primers for synthesis of target and control dsDNA for BLItz analysis

Primer	Sequence	Function
NSW84	Biotin-AGCCATTTTTTGTTAGTCGCCG AAATACTCTTTTCTCTGCCCCTTATTCC CGCTATTAAAAAAAACAATT	Forward primer for target P*fliD* 160 bp DNA (including FlhDC binding boxes)
NSW28	TTTTGCGATTTCCTTTTATCTTTCG	Reverse primer for target P*fliD* 160 bp DNA (including FlhDC binding boxes)
NSW85	Biotin- TGCGTCATCCTTCGCGCTGT	Forward primer for control 160 bp DNA
NSW86	TTTTGACTTGCAATATAGGATAACG	Reverse primer for control 160 bp DNA

The sequences highlighted in grey indicate FlhDC binding boxes in the P*fliD* region.

The amplified DNA was then purified using a QIAgen Gel Extraction kit (Section 2.3.7) and concentrated using a heat-vacuum system. The DNA concentration was determined,

reverting back to further concentrating of DNA, if the concentration was not deemed sufficiently high (Section 2.3.8).

Using the synthesized biotinylated DNA fragments, the "target" or "control" DNA were immobilised to streptavidin (SA) biosensors at fixed concentrations of 35 µg/ml. The FlhDC and YdiV proteins were subsequently added to the biosensor systems at multiple concentrations, to detect any binding interactions. Both proteins were produced as His_6-tagged fusion proteins and were purified and dialysed in sodium phosphate buffers (Sections 4.2.3; 5.2.3 and 2.6.7).

5.8.1 Control experiments

To assess the binding affinity of DNA and FlhDC, the DNA was attached to a biosensor (step 2) with FlhDC added at the association stage (step 4). Prior to this, control experiments were carried out, to ensure that the equipment was capable of immobilising this particular DNA and differentiating between non-interacting and interacting proteins (Figure 5.18). The first control (run 1) was an entire buffer assay, with stages 1-5 all consisting of buffer only (Figure 5.18). As expected this gave a completely flat trace, with no buffer loading onto the biosensor, and no association curve. Secondly, a DNA immobilisation control (run 2) was carried out, loading the target dsDNA to the biosensor and adding buffer at step 4. Here, a loading curve was clearly visible, confirming that DNA was immobilised to the chip. Upon addition of buffer at step 4, no association was indicated, which confirmed that the buffer components did not bind to the DNA.

The equipment was then tested for its ability to detect DNA and protein interactions, using both positive and negative controls. The target dsDNA was immobilised to the biosensor and a protein (100 µg/ml concentration) added in step 4. The positive control protein was FlhDC (run 3) and clearly showed immobilisation of DNA prior to a clear association curve upon protein addition, indicating an interaction between DNA and protein. Finally target DNA was tested against a negative control protein (run 4) mH0579 which is a protein found in *Mannheimia haemolytica* of ~130 amino acids length. The function of this protein is currently unknown, but has no predicted DNA binding ability. In this case the association step produced a flat trace, indicative of a lack of protein:DNA binding.

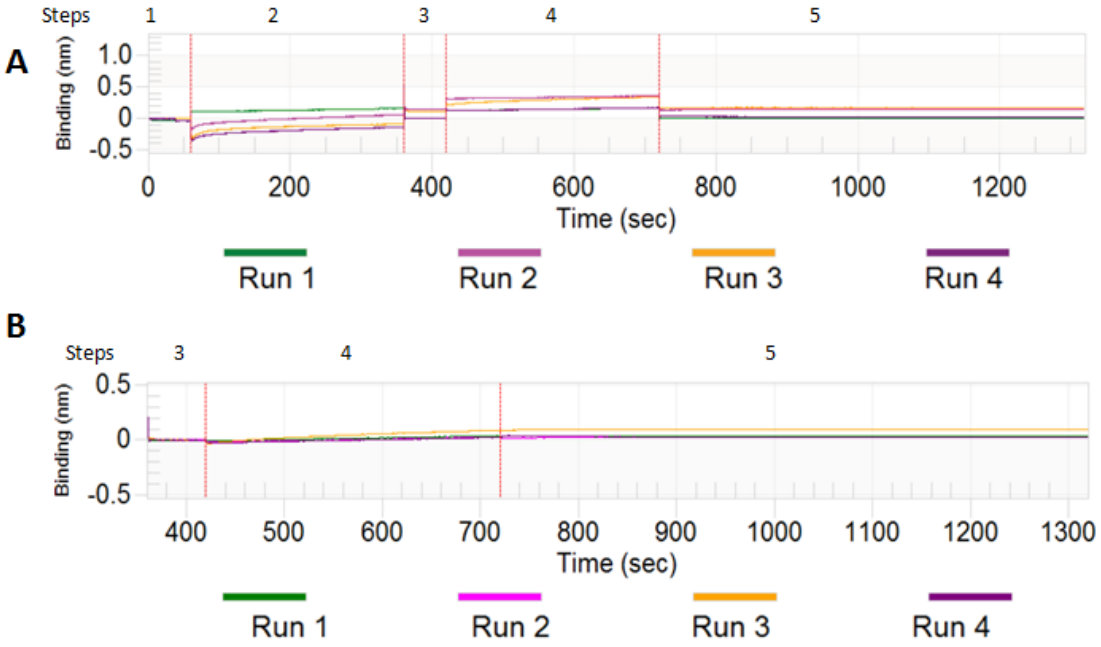


Figure 5.18: BLItz binding spectra for DNA-protein control experiments

Binding spectra for a series of control experiments in which target biotinylated DNA was immobilised to a streptavidin biosensor and positive and negative control proteins were added. The binding reaction was composed of 5 discrete steps; 1: initial baseline, 2: DNA immobilisation, 3: baseline, 4: association and 5: dissociation.

A) Biosensor binding (nm) for entire duration of the experiments, as the reactions proceeded (sec).
B) Biosensor binding (nm) trace for the final three steps, focused on the association and dissociation data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

Each of the four runs had different immobilisation and association components to act as various controls, with all DNA added at 35 µg/ml and proteins added at 100 µg/ml. The components of each run were as follows; 1: Buffer throughout, 2: Target DNA immobilisation and buffer association, 3: Target DNA immobilisation and FlhDC association, 4: Target DNA immobilisation and mH0579 association.

These four control reactions confirmed that the assay worked as expected, with DNA successfully immobilised to the streptavidin chip. Of importance, the buffer-DNA control indicated absolutely no signs of any association interaction, meaning that any protein-DNA interaction would be a direct result of the protein rather than any buffer component. In addition, the system was able to detect the interaction between FlhDC and DNA, but did not detect any interaction with mH0579, therefore confirming the ability of the equipment to differentiate between interacting and non-interacting experiments.

Following this, the kinetic analysis of the FlhDC: DNA interaction could commence.

5.8.2 Kinetics of the FlhDC:DNA interaction

The interaction between target DNA and FlhDC was measured by BLItz analysis (Section 2.9.5) to determine the binding affinities of the P*fliD* and FlhDC interaction. Alongside this, the interaction between the control DNA and FlhDC was measured. This enabled the binding kinetics of both reactions to be compared, to assess whether the FlhDC and P*fliD* interaction was genuine and sequence specific.

The target DNA and FlhDC experiment series was carried out by immobilising target DNA to a streptavidin (SA) biosensor at a concentration of 35 µg/ml. FlhDC was then added at 5 different concentrations ranging from 11 µM to 44 nM, almost a 1000-fold range (Figure 5.19). The resultant traces showed an increase in biosensor thickness (step 2) upon addition and immobilisation of target DNA to the streptavidin chip. FlhDC was subsequently added (step 4) and an upward association curve was produced, indicating the binding interaction of DNA and FlhDC, before the dissociation step (step 5) which showed a gradual loss of binding (Figure 5.19B). Of note, the association step is of varying gradients dependent on the protein concentrations used. The almost flat dissociation step indicates that the interaction is kinetically stable and plateaus at a higher biosensor thickness than the baseline, showing that even after the 10 min dissociation step some of the protein remained bound to DNA. This is in contrast to the control DNA and FlhDC experiment, which shows a steep dissociation curve (Figure 5.20), thereby indicating the genuity of the target DNA and FlhDC interaction.

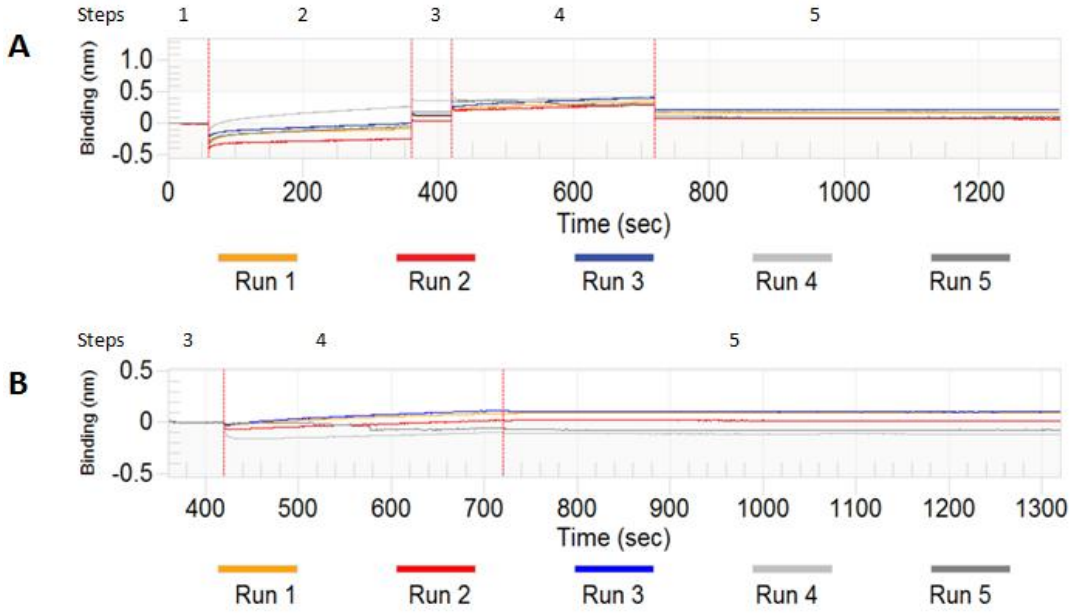


Figure 5.19: BLItz binding spectra for "target" DNA immobilisation and FlhDC addition

See p215 for details.

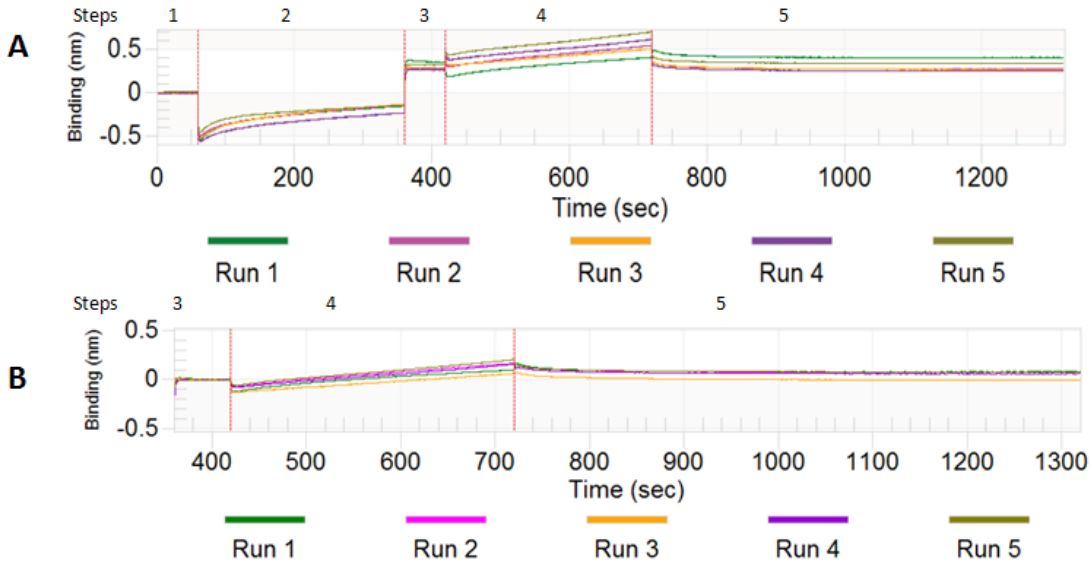


Figure 5.20: BLItz binding spectra for "control" DNA immobilisation and FlhDC addition

See p215 for details.

Figure 5.19: BLItz binding spectra for "target" DNA immobilisation and FlhDC addition

Binding spectra for a series of control experiments in which target biotinylated DNA was immobilised to a streptavidin biosensor and FlhDC protein added. The binding reaction was composed of 5 discrete steps; 1: initial baseline, 2: DNA immobilisation, 3: baseline, 4: association and 5: dissociation.

A) Biosensor binding (nm) trace for entire duration of the experiments, as the reactions proceeded over time (sec).
B) Biosensor binding (nm) trace for the final three steps, enlarging the association and dissociation curve data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

The concentration of dsDNA was fixed at 35 µg/ml, with FlhDC concentrations varying for each of the five runs. The concentration of FlhDC tested was as follows; run 1: 4.4 µM, run 2: 2.2 µM, run 3: 11 µM, run 4: 440 nM and run 5: 44 nM.

Figure 5.20: BLItz binding spectra for "control" DNA immobilisation and FlhDC addition

Binding spectra for a series of control experiments in which control biotinylated DNA was immobilised to a streptavidin biosensor and FlhDC protein added. The binding reaction was composed of 5 discrete steps; 1: initial baseline, 2: DNA immobilisation, 3: baseline, 4: association and 5: dissociation.

A) Biosensor binding (nm) trace for entire duration of the experiments, as the reactions proceeded over time (sec).
B) Biosensor binding (nm) trace for the final three steps, enlarging the association and dissociation curve data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

The concentration of dsDNA was fixed at 35 µg/ml, with FlhDC concentrations varying for each of the five runs. The concentration of FlhDC tested was as follows; run 1: 22 µM, run 2: 4.4 µM, run 3: 11 µM, run 4: 2.2 µM and run 5: 440 nM.

The global kinetics suggested a K_D of 158 nM, which is indicative of tight binding and a strong affinity between the DNA and FlhDC. The association rate constant k_a was calculated at $1.88 \times 10^2 \pm 3.8 \times 10^1$ 1/Ms and the dissociation rate constant k_d was $2.98 \times 10^{-5} \pm 2.5 \times 10^{-5}$ 1/s. These values indicate a reasonable rate of association but an extremely low off-rate, which may suggest that YdiV is necessary for the efficient removal of DNA from FlhDC.

The K_D was calculated for the interaction between FlhDC with four other flagella class II promoters, giving binding values between 12-43 nM (Stafford *et al.,* 2005). Comparing these binding affinities with that of FlhDC to P*fliD* reveals a weaker binding constant to P*fliD*. The major hypothesis to account for this difference is due to the large calculated errors of this analysis. Taking these factors into account, the calculated affinity constant of FlhDC for the *fliD* promoter is consistent with other class II promoter regions (Stafford *et al.,* 2005).

The control DNA series utilised control dsDNA (35 µg/ml) immobilised to a biosensor to test whether the FlhDC:target DNA interactions were specific or not (Figure 5.20). Once more, a range of FlhDC concentrations were tested, between 22 µM and 440 nM. The resultant traces again confirmed DNA immobilisation (step 2), this time the control DNA to the biosensor. Upon FlhDC addition, an association curve was seen, representing some interaction between the protein and DNA (Figure 5.20B). However, the traces of the association curve run parallel to one another, indicating that the interaction is not protein concentration dependent and is purely a result of the protein and DNA being in the same area for a given time. Once the biosensor is removed from the protein sample (step 5), there is a clear rapid drop in the biosensor thickness signifying that any interaction between the protein and control DNA is quickly dissociated and therefore non-specific. The calculated kinetic values yield a much higher global K_D value, indicating much weaker binding, of 4 mM. For completeness, the k_a value was measured as $3.22 \times 10^{-1} \pm 2.1 \times 10^2$ 1/Ms and k_d calculated to be $1.3 \times 10^{-3} \pm 9.12 \times 10^{-5}$ 1/s.

This mM K_D constant is indicative of non-specific binding interactions, with this interaction likely to only occur in these artificial conditions rather than *in vivo*. Of note, this control K_D is dramatically different from the K_D of 158 nM measured for the target DNA binding, confirming that the initial FlhDC: P*fliD* interaction was sequence specific.

5.8.3 Kinetics of the YdiV:DNA interaction

For completeness, the YdiV and P*fliD* binding interaction was analysed, predicting no affinity at all, as shown by EMSA analysis (Figure 5.21). Both the target P*fliD* and control dsDNA were tested, to determine whether YdiV interacted more favourably with the P*fliD* DNA than a random sequence control DNA.

Both types of DNA were immobilised to a streptavidin chip at a fixed 35 µg/ml concentration. Two concentrations of YdiV were then added (5355 nM and 1500 nM), to the reactions and were tested against both types of DNA.

Both the target and control DNA were clearly immobilised to the streptavidin chip, with clear upwards traces visible (Figure 5.21A). Upon YdiV addition to the test dsDNA (run 1 and 2), the association curves were flat, with the dissociation trace level with the baseline. This indicated that any association between YdiV and P*fliD* is temporary, with no or little interaction remaining following the dissociation step (Figure 5.21B). Upon addition of YdiV to the immobilised control dsDNA (run 3 and 4) a marked increase in the association gradient was seen, compared to the much flatter response to the target DNA. However the dissociation step shows a quick drop in the biosensor thickness before forming a plateau at a baseline level. This indicates that any binding that occurred in these reactions was non-specific, and therefore were quickly dissociated.

The kinetics were then calculated for these experiments, and consistently showed a high K_D in the range of 0.8 mM and 14.4 mM for YdiV with both types of DNA. The k_a values for all assays were between 1.36 and 7.73 $x10^{-1}$ 1/Ms (errors of ~4 $x10^{2}$ 1/Ms) and k_d values calculated ranging between 1.03 $x10^{-2}$ and 6.5 $x10^{-4}$ (errors of ~1.5$x10^{-4}$).

This demonstrates that YdiV has no preferential binding for the P*fliD* region than the further upstream control region, and indicates that this affinity range shows non-specific background interaction. In addition, the interactions between FlhDC and the control DNA were of the same K_D range, again convincing evidence that these are simply background interactions rather than anything significant.

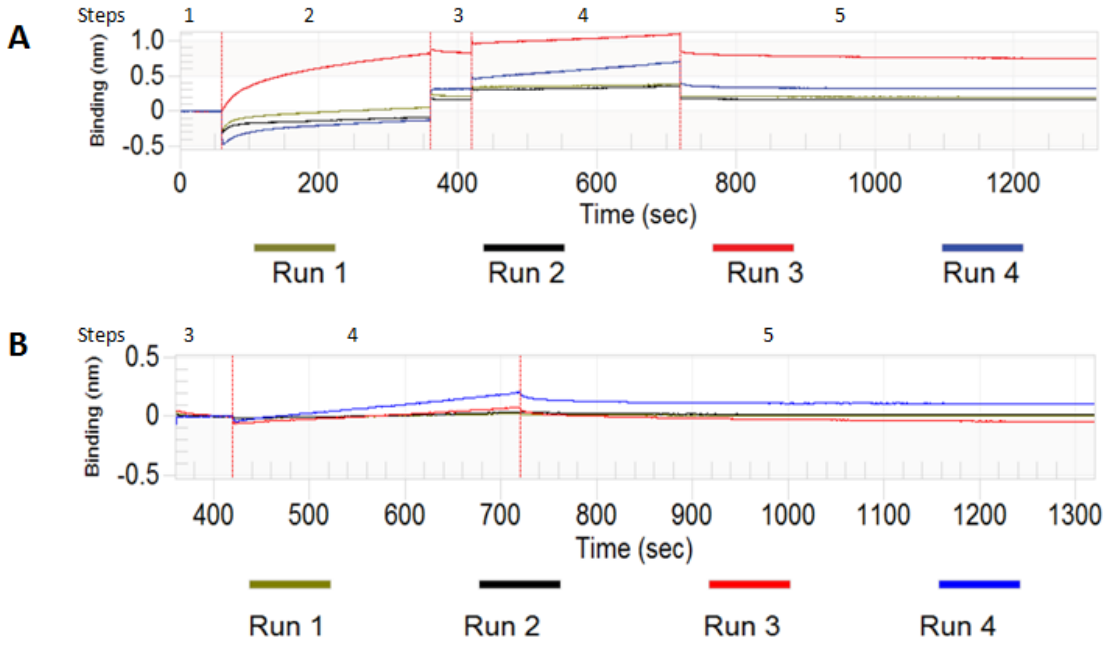


Figure 5.21: BLItz binding spectra for "control" and "target" DNA immobilisation and YdiV addition

Binding spectra for a series of control experiments in which both control and target biotinylated DNA were immobilised to streptavidin biosensors and YdiV protein added. The binding reaction was composed of 5 discrete steps; 1: initial baseline, 2: DNA immobilisation, 3: baseline, 4: association and 5: dissociation.

A) Biosensor binding (nm) trace for entire duration of the experiments, as the reactions proceeded over time (sec).
B) Biosensor binding (nm) trace for the final three steps, enlarging the association and dissociation curve data.

For both traces, the steps are specified, with each vertical line indicating the start of a new step.

Target dsDNA was immobilised in run 1 and 2 and control DNA immobilised in run number 3 and 4 (both at 35 µg/ml). YdiV was added at the association step in all four experiments, at 5355 nM in run 1 and 3, and 1500 nM in run 2 and 4.

5.9 Effect of YdiV concentration on *in vitro* transcription of *fliD*

Following the observation that YdiV inhibited FlhDC:P*fliD* interactions, transcription of *fliD* was studied using *in vitro* transcription assays (Section 2.9.3). Radio-labelled RNA corresponding to the *fliD* DNA was synthesised during the experiment, with regulatory proteins such as FlhDC and YdiV included in the reactions, enabling the impact of these proteins on transcription to be determined. The extent of *fliD* transcription could then be analysed by electrophoresis of the samples on an IVT gel and detection by autoradiography (Section 2.9.4).

DNA was synthesised to cover the entire promoter region and 200 bp of the *fliD* gene by amplifying the region from *E. coli* MG1655 genomic DNA using primers NSW39 and NSW40 (Table 5.4) (Section 2.3.3).

Table 5.4: Primers for IVT P*fliD*- *fliD* DNA synthesis

Primer	Sequence	Function
NSW39	TTTTGGATCCGATTCGTTATCCTATATTG CAAGT	Forward primer for P*fliD* (leading to *fliD* gene)
NSW40	TTTTGAATTCGCGCGCTTTTCAGCGTACC	Reverse primer for *fliD* gene and P*fliD*

The amplified DNA was then purified and quantified, prior to storage at -20^{o}C until required (Section 2.3.8 and 2.3.10). Reaction mixtures were made, consisting of 1 pmole RNA polymerase (specifically as a holoenzyme with σ^{70}), 0.1 pmole *fliD* DNA and varying concentrations of the purified proteins FlhDC and YdiV in an IVT Tris-HCl buffer, pH 8 (Section 2.9.3). Both proteins were purified as His_6-tagged fusion proteins and were used in these assays directly in sodium phosphate elution buffer (Sections 4.2.3 and 5.2.3).

A control reaction containing neither FlhDC nor YdiV was analysed, alongside various samples containing 12 μM (seen previously to be the minimum concentration of FlhDC to cause DNA-binding in the EMSA assays; Figure 5.17), with increasing quantities of the anti-FlhDC factor YdiV. These samples were incubated at 37^{o}C for 15 min prior to transcription initiation by addition of a nucleotide mix, containing [α-^{32}P]-labelled UTP. Once again samples were incubated for 15 min at 37^{o}C to enable transcription to proceed. The

reactions were stopped by addition of stop/loading dye (Table 2.15). The samples were then electrophoresed on a 6% IVT gel alongside a 0.1-1 Kb RNA marker (Section 2.9.3).

In the absence of FlhDC, no *fliD* transcript was detected, indicating that P*fliD* transcription was completely dependent on FlhDC (Figure 5.22, lane 1). In the presence of FlhDC, two transcripts were detected, suggesting two FlhDC-dependent transcript starts (Figure 5.22, lane 2). Interestingly, concentrations of YdiV caused an initial increase in the abundance of both transcripts (Figure 5.22, lanes 3-5) followed by a decrease (Figure 5.22, lanes 6-10). Of note, equimolar concentrations of FlhDC and YdiV caused repression of P*fliD* transcription (Figure 5.22, lane 8).

This data suggests that maximal transcription occurred when a small quantity of YdiV was present, perhaps by inhibiting the formation of the minor alternate P*fliD*:FlhDC complexes observed in EMSAs (Figure 5.17). Alternatively, YdiV could be part of the FlhDC:P*fliD* binding complex, directly aiding transcription in some manner, possibly by altering the DNA conformation. The possibility that YdiV might act to aid FlhDC:P*fliD* transcription under some conditions is in contrast to the previous hypothesis in which YdiV was reported solely as an anti-FlhDC factor, with no evidence of YdiV having any positive effect on the FlhDC:DNA complex (Wada *et al.,* 2011; Wada *et al.,* 2012). However, a theoretical model of the YdiV:FlhDC interaction has been hypothesised, in which a transitional $YdiV_2$:$FlhD_4C_2$:DNA complex was proposed (Li *et al.,* 2012). Therefore attempting to understand the manner in which YdiV increased *fliD* transcription was of interest.

5.9.1 Deciphering the role of YdiV in the FlhDC:DNA complex

Following the IVT assays, the role of YdiV in the FlhDC, P*fliD* and YdiV reaction mix was analysed, carrying out EMSAs in the presence of YdiV and FlhDC antibodies. It was hypothesised that a supershift would occur in the EMSA if an antibody bound to a DNA-binding complex protein, and therefore would identify which proteins were present in this DNA complex.

Electromobility shift assays (Section 2.9.2) were carried out using radiolabelled P*fliD* DNA (Section 5.7). Reaction mixes of 12 µM FlhDC and 1.5 µM YdiV were incubated with 2 nM

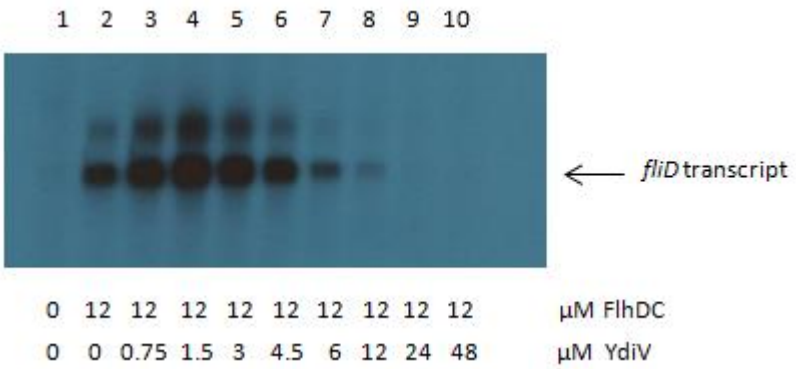


Figure 5.22: *In vitro* transcription of *fliD* in the presence of the transcription factor FlhDC and anti-FlhDC factor YdiV

Autoradiograph showing the extent of radiolabelled *fliD* transcript formed under various conditions.

IVT reactions containing RNA polymerase, *fliD* DNA, a radiolabelled NTP mix and varying concentrations of the purified proteins FlhDC and YdiV as specified. Following the initial control reaction (containing no FlhDC or YdiV), the remaining samples contained 12 µM FlhDC as this was the minimum concentration required for a P*fliD*:protein shift in an EMSA (Figure 5.17). The YdiV protein was then titrated in to the samples, with molar ratios of FlhDC: YdiV progressing from 16:1 right to 1:4.

The location of the *fliD* transcript is indicated.

radiolabelled P*fliD* (Section 2.9.1) for 30 min at 37^{o}C (as these protein concentrations caused maximum transcription in the IVT assays). Following this, YdiV and FlhDC antibodies (Section 5.5) were added to the relevant samples (at an antiserum dilution of 1/10), just prior to addition of EMSA loading dye (Table 2.14) and electrophoresis of the samples. The EMSA showed the free P*fliD* DNA (lane 1) being shifted to indicate a protein:P*fliD* complex upon FlhDC and YdiV addition (lanes 2 and 3) (Figure 5.23). Upon addition of both the YdiV antibody (lane 4) and the FlhDC antibody (lane 5), the protein:P*fliD* species was itself shifted, giving two supershifted species. This indicated that both the FlhDC and YdiV proteins were present in the DNA binding complex. To confirm this result, control reactions of P*fliD* and the YdiV and FlhDC antibodies (lanes 6 and 7 respectively) were carried out and caused no DNA shifting, thereby authenticating the supershifts seen in lanes 4 and 5.

An additional control was carried out to ensure that the YdiV antibody had no affinity for the FlhDC protein (Figure 5.24). Here, as expected, FlhDC addition produced a shifted species upon binding to the P*fliD* DNA. However, upon addition of the anti-YdiV serum to the protein:P*fliD* species, no supershift was produced indicating no cross-affinity of the anti-YdiV for the FlhDC complex. This confirms that the supershift seen in Figure 5.23 was directly due to the presence of YdiV in the DNA:protein complex.

These assays clarify the IVT result, suggesting that the YdiV protein has a direct effect on the FlhDC:DNA interaction, and is part of the DNA-binding complex, forming a YdiV:FlhDC:P*fliD* complex. This conclusion is in agreement with that of Takaya *et al.* (2012) who identified the existence of the same DNA:YdiV:FlhDC complex. Whilst the formation of this complex is conclusive, identifying the mechanism of the interaction and the possible upregulation ability of YdiV is a crucial area for investigation.

5.10 Investigation of ligand interaction to the YdiV:FlhDC complex

At present, there is no known trigger for the FlhDC:YdiV interaction. Analysis of YdiV tentatively suggested a possible interaction with c-di-GMP (Section 4.5) which provided a reason to investigate nucleotide binding. Therefore, a number of assays were carried out, each with its own limitations, and hence the data were interpreted as a collective.

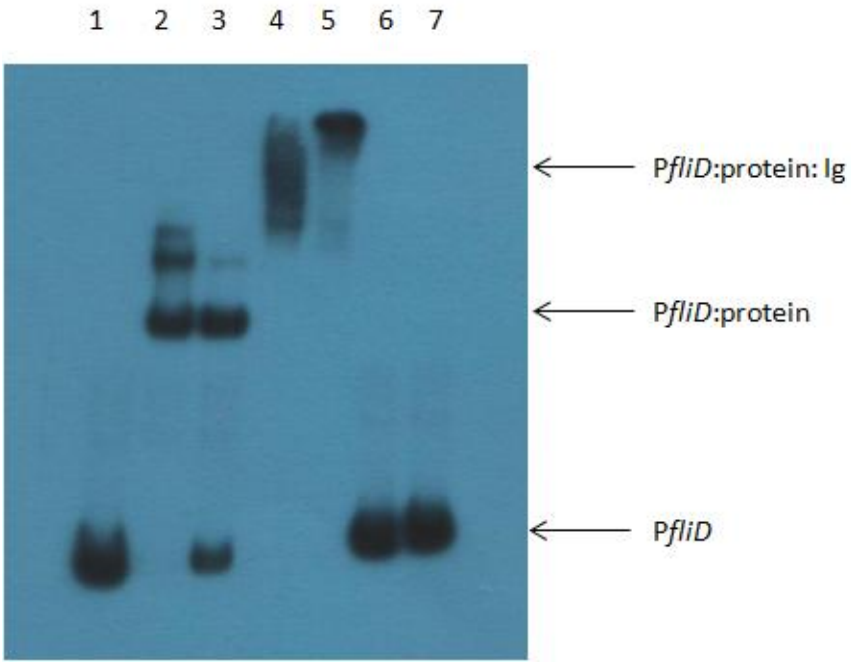


Figure 5.23: EMSA to determine the role of YdiV in the P*fliD* , FlhDC and YdiV reaction mix

Autoradiograph, showing labelled DNA in the presence of purified FlhDC and YdiV, as well as YdiV or FlhDC antibodies.

Lanes 1-7 all contain 2 nM radiolabelled P*fliD* DNA, with the additional constituents detailed below:

Lane 1: DNA only
Lane 2: DNA + 12 µM FlhDC
Lane 3: DNA, 12 µM FlhDC +1.5 µM YdiV
Lane 4: DNA, 12 µM FlhDC, 1.5 µM YdiV + anti-YdiV
Lane 5: DNA, 12 µM FlhDC, 1.5 µM YdiV + anti-FlhDC
Lane 6: DNA + anti-YdiV
Lane 7: DNA + anti-FlhDC

The free P*fliD*, P*fliD*:protein and P*fliD*:protein:Ig species are indicated.

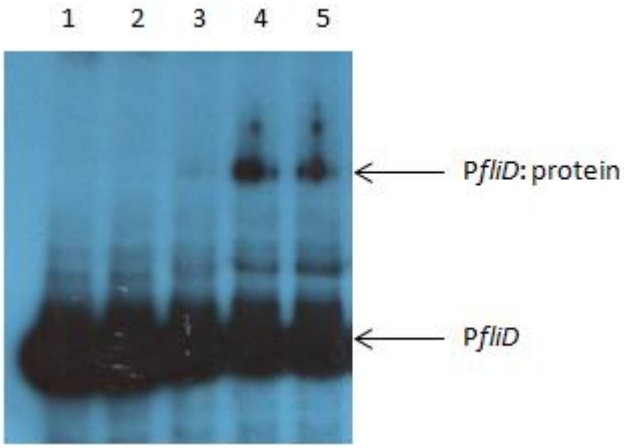


Figure 5.24: Control EMSA for FlhDC and anti-YdiV

Autoradiograph, showing labelled DNA in the presence of purified FlhDC and YdiV antibodies.

Lanes 1-5 all contain 2 nM radiolabelled P*fliD* DNA, with the additional constituents detailed below:

Lane 1: DNA only
Lane 2: DNA + 12.6 µM FlhDC
Lane 3: DNA + 38 µM FlhDC
Lane 4: DNA + 50 µM FlhDC
Lane 5: DNA, 50 µM FlhDC + anti-YdiV

The free P*fliD* and P*fliD*:protein species are indicated.

5.10.1 Protein and nucleotide pull-down assays

Pull-down assays were carried out to determine whether the addition of a nucleotide altered the YdiV:FlhDC interaction (Section 2.7.5). The YdiV:FlhDC complex was formed prior to the addition of nucleotides, therefore enabling analysis of whether specific nucleotides disrupted this complex. However, feasibly a ligand could also enhance the YdiV:FlhDC complex formation, in which case the quantity of the complex or speed of its formation may increase. Whilst any changes in the quantity of complex could be detected, any changes in the speed of formation are undetectable, thereby highlighting the limitations of this technique.

Pull-down assays were carried out using small aliquots (100 µl) of pre-washed Ni-NTA resin (Section 2.7.5). Soluble crude YdiV-His_6 (Section 4.2.2) was immobilised to the resin, before addition of crude $(FLAG)_3$-FlhDC (Section 5.2.2) to form the YdiV:FlhDC complex. Various individual nucleotides (30 µl of 50 mM concentration) were then added. Samples were taken from the unbound fraction and the resin was thoroughly washed. Following this, addition of an imidazole-rich elution buffer eluted the His-tagged YdiV protein from the resin. Hypothetically if a nucleotide were to interact and disrupt the FlhDC:YdiV complex, it was predicted that the FlhDC component would be lost from the pull-down assay in the wash step following ligand addition, whilst the His-tagged YdiV would remain bound until imidazole elution.

Samples were collected at each stage and analysed by SDS-PAGE (Section 2.6.9) to identify which proteins were eluted at the various stages of the protocol (Figure 5.25). In all cases, there was no obvious indication of YdiV:FlhDC disruption upon nucleotide addition (lane 5), with the two proteins clearly eluted together in lanes 6 and 7. In some cases, such as upon GMP or ATP addition, protein species were present in the excess nucleotide sample (lane 5). However both YdiV and FlhDC were detected indicating that this could simply be a technical error i.e. due to uptake of a small volume of resin along with the supernatant.

Whilst this assay did not indicate FlhDC dissociation upon nucleotide addition, there was a noticeable variation in the proportion of YdiV and FlhDC proteins eluted (lane 6) and resin-bound (lane 7) (Figure 5.25). This difference appeared to be nucleotide dependent and the relative amounts of both proteins were quantified by Image-J analysis (Section 2.10.1) of

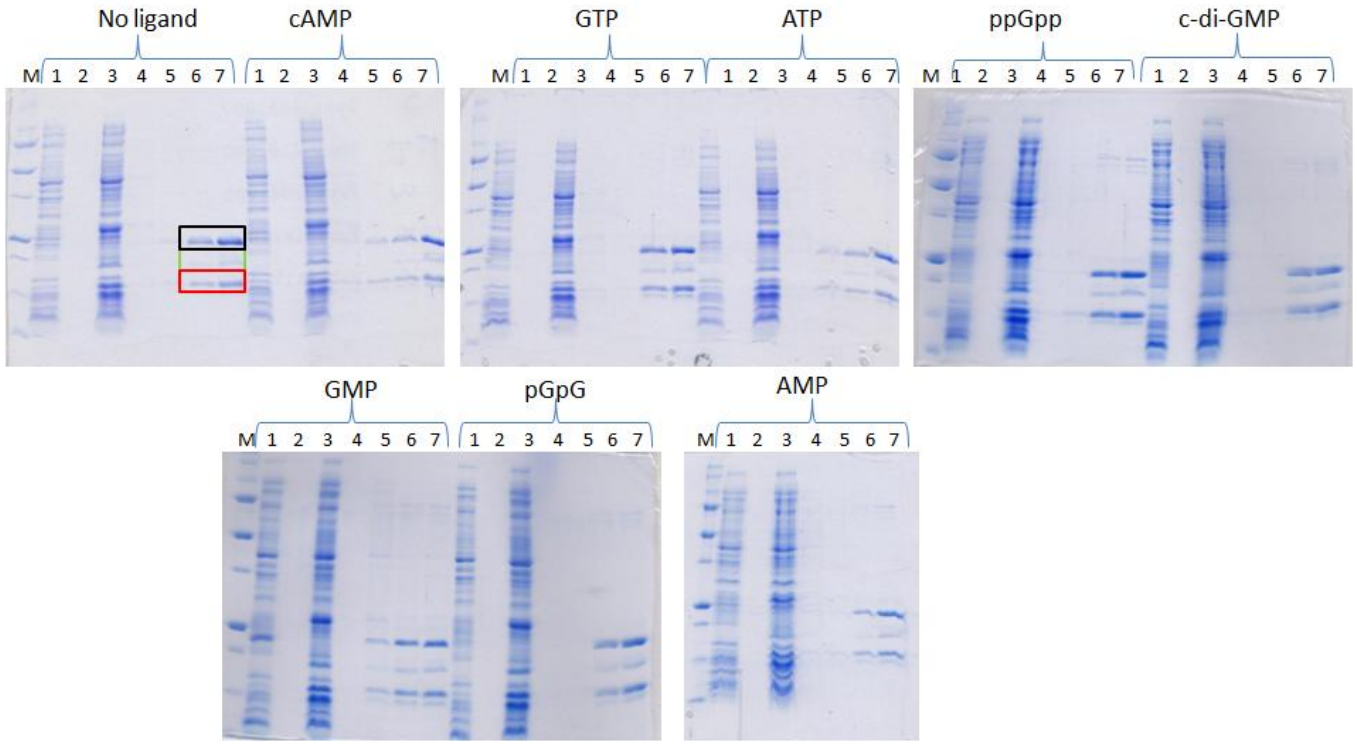


Figure 5.25: YdiV:FlhDC pull-down assays in the presence of nucleotides

Samples were taken at all stages during the pull-down assays, and analysed via SDS-PAGE. In all cases, M: Precision plus protein marker, Lane 1: Soluble YdiV, 2: final wash, 3: soluble FlhDC, 4: final wash, 5: post-nucleotide wash, 6: elution, 7: resin.

Nucleotides (30 µl of 50 mM) were added to the pull-down assays (Section 2.7.5), as indicated. YdiV, FlhD and FlhC are present at their expected sizes of 28, 16 and 22 KDa respectively. The locations of these proteins are indicated by coloured boxes in the no ligand samples, with YdiV highlighted with a black box, FlhD shown in a red box and FlhC in a green box.

the SDS-PAGE gels (Figure 5.26). Quantification of these proteins indicated variation amongst samples, with the proportion of YdiV fluctuating from 0.5 to 0.65. Reactions containing pGpG, ppGpp and c-di-GMP to the FlhDC:YdiV complex were quantified and indicated a 50/50 elution mix of FlhDC and YdiV, possibly indicating a 1:1 stoichiometry of the proteins. Contrastingly, the no ligand sample or the addition of cAMP yielded a 65/35% elution mix of YdiV:FlhDC, possible indicating a 2:1 stoichiometry of the eluted proteins.

These stoichiometry changes could be highly significant, with models hypothesising that four distinct YdiV:FlhDC complexes (specifically $YdiV_1$:$FlhD_4C_2$, $YdiV_2$:$FlhD_4C_2$, $YdiV_3$:$FlhD_4C_2$ and $YdiV_4$:$FlhD_4C_2$) could form (Li *et al.,* 2012). Therefore the 1:1 stoichiometry of YdiV:FlhDC seen upon pGpG, ppGpp and c-di-GMP addition could feasibly indicate the formation of a $YdiV_1$-$FlhD_4C_2$ complex whilst the 2:1 stoichiometry could therefore indicate the formation of an $YdiV_2$-$FlhD_4C_2$ complex. If the oligomeric state was genuinely altered upon nucleotide addition, this would indicate that nucleotide-protein interactions were occurring and presumably reduced the ability of the proteins to interact, either by inducing a conformational change or by shielding an interaction interface.

Whilst this is one theory, the significance of these protein fluctuations is unknown, requiring assay repetitions in order to increase confidence in these data. Equally, one of the limitations of this assay was the fixed order in which the pull-down assay was carried out, with the FlhDC-YdiV complex forming before the nucleotide addition. This complex formation could therefore prevent nucleotide binding which might have previously occurred to the independent proteins. Equally, as discussed previously, nucleotide-dependent enhancement of YdiV:FlhDC complex formation may not be detected by this method. A further experiment would therefore be to investigate the YdiV-FlhDC complex formation with nucleotide addition throughout the entire assay.

5.10.2 Fluorescence spectroscopy

Intrinsic tryptophan fluorescence (Section 2.7.4) was measured for the FlhDC:YdiV complex before and after nucleotide addition, as a means to determine nucleotide binding. In principal the technique can detect nucleotide binding, providing that this causes a conformational change within the FlhDC:YdiV complex that altered the fluorescent properties of the tryptophan residues. Thus the major limitation is that any nucleotide

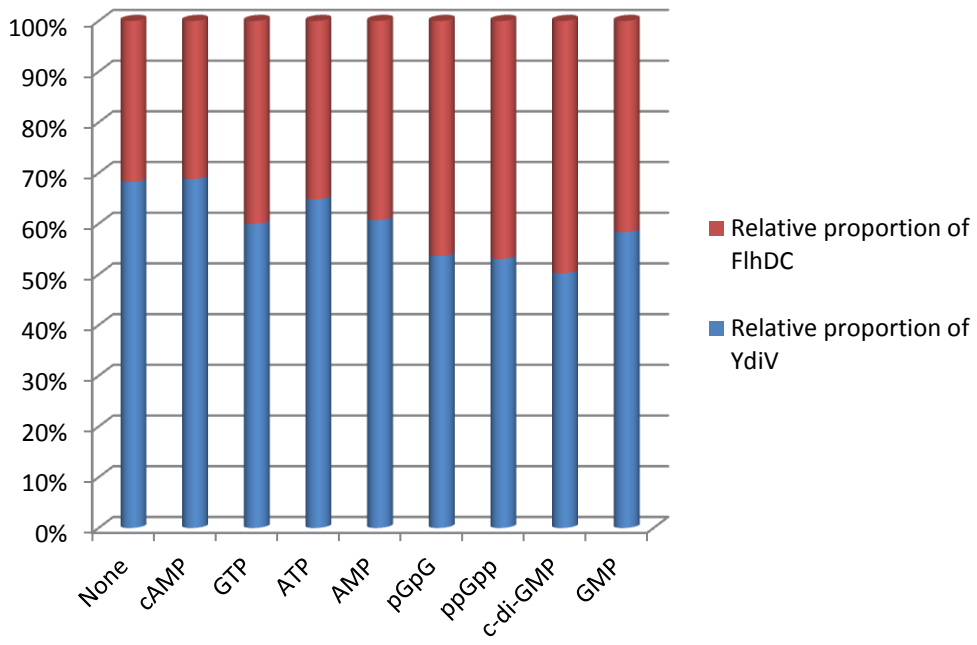


Figure 5.26: Quantification of the relative amounts of eluted FlhDC and YdiV proteins resulting from nucleotide pull-down assays

Image-J quantification of the FlhDC and YdiV proteins eluted from pull-down assays and analysed by SDS-PAGE (Figure 5.25).

Following quantification of the relative amounts of each protein, a bar chart was produced visually indicating each proteins proportion of the total elution mix.

Nucleotides were added to the pull-down assays, as specified.

binding which did not affect the tryptophan residues of the FlhDC:YdiV complex would not be detected.

Samples containing the protein complex alone or equimolar mixtures of the protein complex and nucleotide were tested, to analyse fluorescence changes upon nucleotide addition. The samples consisted of 50 µM FlhDC:YdiV complex (sodium phosphate buffer), 30-fold molar excess of Mg^{2+} in a 20 mM sodium phosphate pH 7.5 buffer, with nucleotides also added at 50 µM (Section 2.7.4).

The FlhDC:YdiV complex was formed, by carrying out a pull-down assay, to produce a native stoichiometry of complex, rather than adding a fixed amount of each (Section 2.8.1). Complex formation and purification was confirmed by SDS-PAGE (Section 2.6.9) prior to protein quantification by BioRad protein assay (Section 2.6.8).

Samples were made, added to a 3 ml cuvette and fluorescence readings taken at 30^{o}C, with an excitation wavelength of 280 nm, and emission readings recorded between 300-400 nm. The fluorescence spectrum for each sample was recorded and enabled comparisons between the complex in the presence and absence of various nucleotides. Generally the resultant traces have overlying protein only (blue) and protein and nucleotide (red) traces indicating no fluorescence differences upon nucleotide addition (Figure 5.27). For example, the ATP, AMP, pGpG and GMP traces showed almost no differences in the presence of these nucleotides. Other spectra showed greater differences, such as the cAMP, GTP and GDP traces, in which the blue protein only trace showed slightly more fluorescent emission than the red protein and nucleotide trace. The most dramatic difference was for the c-di-GMP sample, in which the YdiV:FlhDC complex fluorescence seemed to be enhanced upon nucleotide addition, which could indicate a tryptophan residue becoming more exposed following a nucleotide induced protein conformational change. However upon triplicate repetition of the c-di-GMP samples, the experimental variation was clear which meant these results should be treated with caution (Figure 5.28).

From these data, no firm indication of nucleotide binding was detected, however due to the limitations of this technique this did not rule out nucleotide binding.

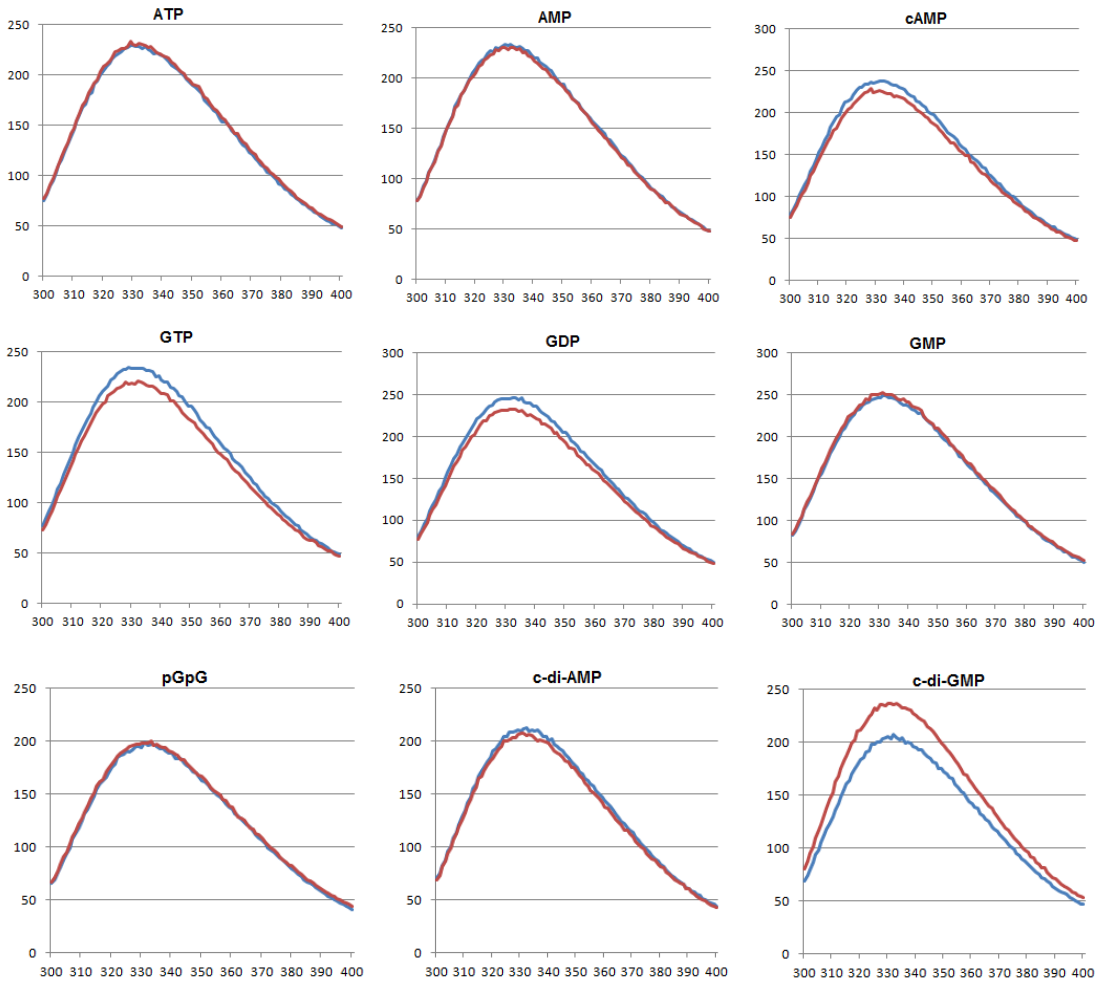


Figure 5.27: Fluorescence spectroscopic analysis of the YdiV:FlhDC complex in the presence of various nucleotides

All graphs show the fluorescence spectra at 30 ^{o}C.

In all cases, the blue traces show the protein complex (50 µM) only sample, whilst the red traces show the same complex (50 µM) with the addition of a specified nucleotide (50 µM). For all spectra, the fluorescence intensity in mAU (y axis) is shown against wavelength (x axis) in nm.

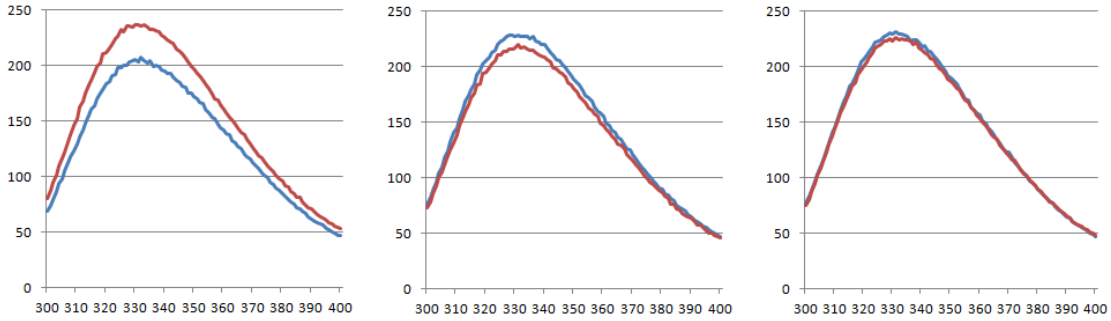


Figure 5.28: Fluorescence spectroscopic analysis of the YdiV:FlhDC complex and c-di-GMP addition for three biological replicates

The three traces show three independent c-di-GMP assays, once again with the protein complex only trace being the blue trace and the red trace showing the complex plus nucleotide sample. In all cases, nucleotides were added at a fixed concentration, from a single stock solution and a single protein preparation was used, with all readings taken at 30^{o}C.

For all spectra, the fluorescent intensity in mAU (y axis) is shown against the wavelength in nm (x axis).

5.10.3 Determination of the DNA-binding capacity of FlhDC in the presence of nucleotides

Electromobility shift assays (Section 2.9.2) were used to analyse what effect, if any, the nucleotides c-di-GMP, c-di-GMP and pGpG had on the DNA-binding capacity of FlhDC.

Previous EMSA assays showed that FlhDC bound to the P*fliD* DNA at a concentration of 12 µM but binding was repressed fully at equimolar concentrations of YdiV (Figure 5.17). Conditions were therefore chosen of 12 µM FlhDC (intended to bind DNA) and 12 µM FlhDC plus 12 µM YdiV (expected to inhibit DNA binding). Both proteins were purified in their His_6 fusion forms (Sections 4.2.3 and 5.2.3) and used along with a control sample containing no protein, incubating all samples with 2 nM radioactively labelled P*fliD* (Section 5.7) and excess nucleotide (100 µM) at 37^{o}C for 30 min (Section 2.9.2). Following this, the samples were mixed with EMSA loading dye (Table 2.14) and electrophoresed.

The exposed film (Section 2.9.4) showed the expected DNA shift when FlhDC was present, whilst only free unbound DNA was present in samples containing both FlhDC and YdiV (Figure 5.29). Once again, three distinct shifted species were present, indicating differing protein-DNA species. The abundance of each shifted species appeared to be nucleotide dependent, with the no ligand control producing the two largest shifted species, c-di-AMP producing approximately equal quantities of each species and both pGpG and c-di-GMP producing only the smallest shifted species.

The identity of these species and their significance is currently unknown, but they are likely to represent binding of the FlhDC complex to the high affinity 'FlhDC-binding box' as well as additional lower affinity regions in the P*fliD* DNA, producing multiple shifted species. It appears that c-di-GMP and pGpG resolve the multiple complexes into the single high-affinity complex, but do not inhibit YdiV-mediated dissociation of the P*fliD*:FlhDC complex.

The highest priority is therefore to understand the significance of these three shifted species and to determine the exact components of each. Following this, the role of nucleotides in altering the proportions of the species can be studied. If a nucleotide were to bind to the FlhDC complex, this could potentially alter or limit which oligomeric complex could be formed and in this manner could influence DNA binding. Equally, another

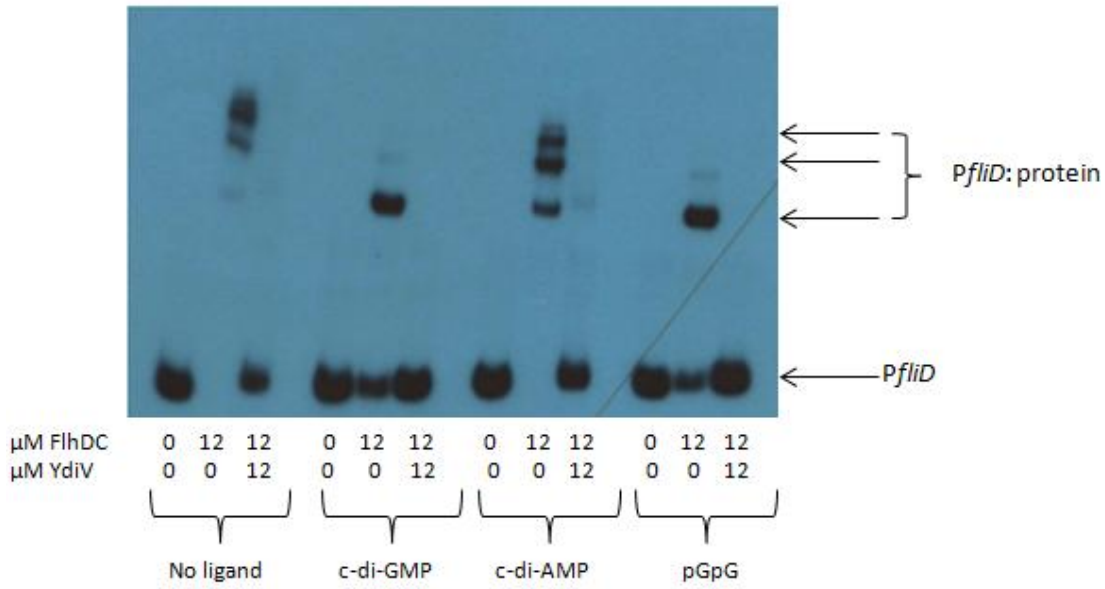


Figure 5.29: EMSA to show the P*fliD*:FlhDC binding capacity in the presence of various nucleotides

EMSA gel consisting of radiolabelled P*fliD* DNA, in the presence or absence of FlhDC, YdiV and nucleotides. Concentrations of the two proteins and nucleotide additions (100 µM) are stated.

The location of the free P*fliD* and P*fliD*:protein species are indicated.

possibility would be that nucleotide interactions with FlhDC caused a DNA:FlhDC interaction site to be shielded, therefore altering how many binding regions in the P*fliD* region could be occupied by FlhDC.

Clearly at this stage the mechanism in which nucleotides alter FlhDC:DNA binding is only speculation, however the presence of nucleotides does appear to affect the FlhDC:DNA interaction.

5.10.4 Analysis of the effect, if any, of nucleotides upon FlhDC mediated *in vitro* transcription

The previous EMSA analysis revealed that the presence of nucleotides (specifically c-di-GMP, c-di-AMP and pGpG) appeared to alter FlhDC:DNA binding, specifically by affecting the proportion of the three P*fliD*:FlhDC species being produced (Figure 5.29). These three nucleotides were further investigated, to identify whether the various FlhDC:DNA species (and the proportion of each) had any effect on FlhDC-mediated transcription from P*fliD*.

In vitro transcription assays (Section 2.9.3) were carried out, assessing the level of *fliD* transcription promoted by FlhDC in the presence and absence of nucleotides. DNA corresponding to the entire promoter region and 200 bp of the *fliD* gene was synthesised as previously described (Section 5.9). This DNA (0.1 pmole) was then incubated with 1 pmole RNA polymerase and 12 µM purified FlhDC-His_6 as described in Section 5.2.3 (directly in sodium phosphate elution buffer) in an IVT Tris-HCl buffer, pH 8. A no ligand reaction was carried out, as well as samples containing FlhDC and c-di-AMP, c-di-GMP and pGpG, with ligands at 1 mM. These four reactions were incubated at 37^{o}C for 15 min and analysed on an IVT gel.

The exposed film (Section 2.9.4) showed the level of *fliD* RNA produced under the various conditions (Figure 5.30). It appeared that the no ligand, c-di-AMP and pGpG samples (lanes 1, 2 and 4) produced very similar RNA intensity species, indicating similar *fliD* transcription. In contrast to the IVT experiments shown in Figure 5.22, only one transcript was detected. Interestingly, c-di-GMP addition (lane 3) appeared to result in a decreased level of transcription. The intensities of the transcripts in the IVT were quantified using Image-J

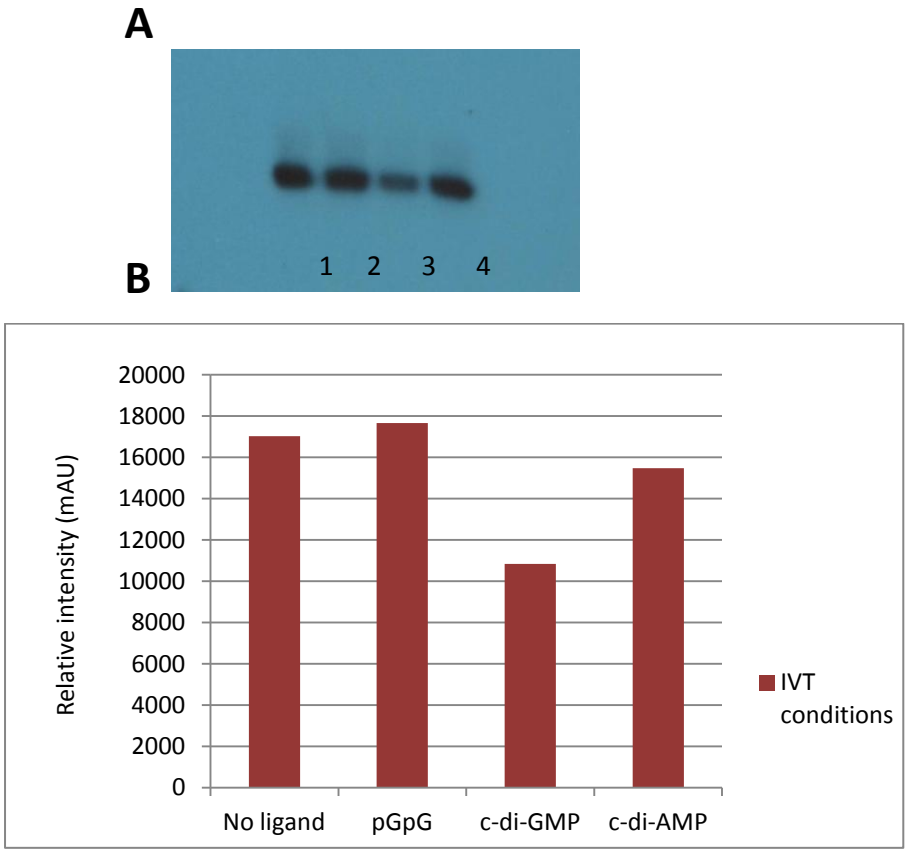


Figure 5.30: *In vitro* transcription levels of *fliD* in the presence of FlhDC and excess nucleotides

A) Exposed *in vitro* transcription film, showing the extent of *fliD* transcription *in vitro*.
Lanes 1-4 contain 12 µM FlhDC, with lane 1: no nucleotide addition, lane 2: excess pGpG, lane 3: excess c-di-GMP, lane 4: excess c-di-AMP.

B) Quantification of bands on IVT film, as quantified by Image J software.

(Section 2.10.1; Figure 5.30A). Image-J quantified the intensity of lane 3 to be ~11,000, compared to ~15-17,000 for the remaining three samples (Figure 5.30B). Therefore quantitative analysis showed that the transcriptional differences are significant, indicating a possible role of c-di-GMP in *fliD* transcriptional repression.

5.10.5 Concluding remarks regarding nucleotide binding to the FlhDC or FlhDC:YdiV complex

Taking into account the four assay methods utilised here, the results are ambiguous. Intrinsic fluorescence spectroscopy suggested that c-di-GMP interacted with the FlhDC:YdiV complex, but this was not always observed (Figures 5.27; 5.28). NMR experiments suggested YdiV binds c-di-GMP (Figure 4.14), EMSAs showed that c-di-GMP altered the interaction of FlhDC with P*fliD* (Figure 5.29) and c-di-GMP inhibited *fliD* transcription *in vitro* (Figure 5.30). Although there are several caveats associated with these data, there is a body of evidence to suggest that c-di-GMP interacts with YdiV and FlhDC, perhaps as the FlhDC:YdiV complex, resulting in conformational changes in the complex that alter DNA-binding/transcriptional activation. Further experimentation would therefore be essential to fully understand the role, if any, of nucleotides in the FlhDC:YdiV complex.

5.11 Competition assays with FliT/FliZ

FliT and FliZ were chosen as potential competitors of the YdiV:FlhDC interaction, due to the confirmed direct interaction between FliT and FlhDC and the fact that the potential FlhDC interactor FliZ has been shown to increase the stability of FlhDC (Imada *et al.,* 2010; Saini *et al.,* 2008).

5.11.1 Cloning of FliT and FliZ overproduction plasmids

The *E. coli* genes *fliT* and *fliZ* were cloned into the overproduction plasmid pGEX-KG to give N-terminal GST-tag fusion constructs (Figure 5.31).

Primers NSW59-62 were designed to amplify the *fliT* and *fliZ* genes from genomic DNA (Table 5.5). These primers were designed to contain BamHI and HindIII restriction sites at

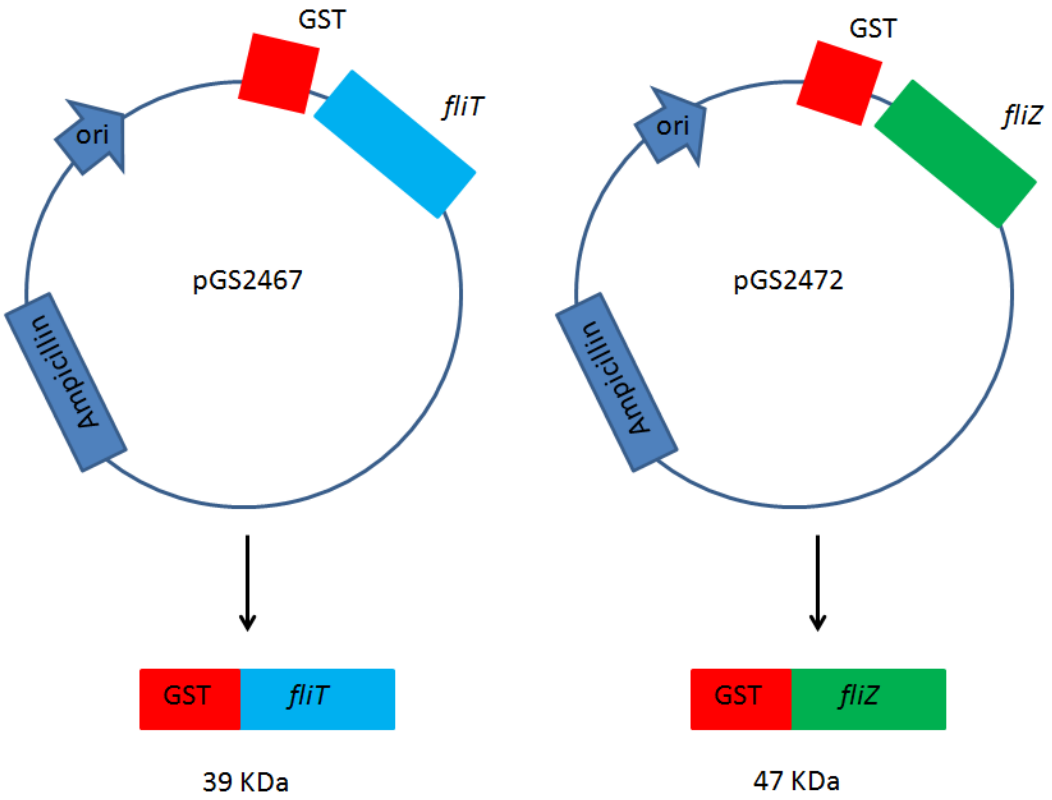


Figure 5.31: Schematic plasmid maps for the insertion of *fliT and fliZ* into pGEX-KG vectors

Maps for the constructed pGS2467 and pGS2472 plasmids, indicating the position of the GST tag, the antibiotic resistant cassette and the fusion constructs produced.

The origin of replication is indicated and labelled as ori.

Each fusion construct encodes the relevant fusion protein of expected molecular weights 39 KDa (GST-FliT) and 47 KDa (GST-FliZ).

either end of the PCR fragment respectively (Table 5.5). A start codon was absent in the forward primers to enable fusion with the N-terminal plasmid-derived GST tag.

Table 5.5: Primers used for amplification of the *fliT and fliZ* genes for insertion into the pGEX-KG plasmid

Primer	Sequence	Function
NSW59	TTTTGGATCCAACCATGCACCGCATTTATATTTC	Forward primer for *fliT* gene
NSW60	TTTTAAGCTTTCAAAAGAGGTTATCCTGCGGAG	Reverse primer for *fliT* gene
NSW61	TTTTGGATCCATGGTGCAGCACCTGAAAAGAC	Forward primer for *fliZ* gene
NSW62	TTTTAAGCTTTTAATATATATCAGAAGAAGGCAGGC	Reverse primer for *fliZ* gene

The grey shaded sequences indicate BamHI restriction sites and the boxed sequences show HindIII restriction sites.

PCR reactions were carried out using high fidelity polymerase, *E. coli* MG1655 genomic DNA and primers in pairs NSW59 and 60 (*fliT*) and NSW61 and 62 (*fliZ*) to amplify the respective genes (Sections 2.3.1 and 2.3.3). The resultant PCR products were cleaned by PCR purification and digested using the BamHI and HindIII restriction enzymes (Sections 2.3.10 and 2.3.6). The pGEX-KG plasmid (Lab stock) was purified and digested with the same restriction enzymes (BamHI and HindIII) (Sections 2.3.4 and 2.3.6). The *fliT* and *fliZ* fragments and the linearised pGEX-KG vector were ligated (Section 2.3.9) via their sticky ends and used to transform electrically competent *E. coli* DH5α cells (Section 2.2.7). Transformants were selected on ampicillin supplemented agar and screened by colony PCR using primers NSW65 and NSW66 (Table 5.6) to assess which plasmids contained the correctly sized inserts (Section 2.3.11). Plasmids containing appropriate length inserts were checked by DNA sequencing (Source Bioscience) using the same pGEX-KG sequencing primers (Table 5.6).

Table 5.6: Sequencing primers for pGEX-KG

Primer	Sequence	Function
NSW65	GACCATCCTCCAAAATCGGA	Forward sequencing primer for pGEX-KG
NSW66	GAGGTTTTCACCGTCATCAC	Reverse sequencing primer for pGEX-KG

Plasmids containing no mutations in the *fliT* and *fliZ* sequences were named pGS2467 (pGEX-KG::*fliT*) and pGS2472 (pGEX-KG::*fliZ*) and the plasmids used to transform electrically competent *E. coli* BL21 λ(DE3) cells (Section 2.2.7).

5.11.2 Overproduction of GST-FliT and GST-FliZ

Overproduction of the proteins, GST-*fliT* and GST-*fliZ* was carried out in *E. coli* BL21 (λDE3) (Section 2.6.1). The optimal overproduction conditions were not investigated, instead simply inducing cultures for 2.5 h with 100 µg/ml IPTG, which upon SDS-PAGE gel analysis (Section 2.6.9) confirmed successful overproduction of the respective proteins at the expected molecular weights (data not shown).

When required, the cell pellets of the overproduction strains (*E. coli* BL21 λ(DE3)/pGS2467 and BL21 λ(DE3)/pGS2472) were resuspended in sodium phosphate breakage buffer, cells broken by sonication and the phases separated (Section 2.6.2), using the soluble fraction for the pull-down assays (Section 2.8.2).

5.11.3 Control assays

Control pull-down assays (Section 2.8.2) were carried out to determine whether the FliT and FliZ proteins interacted with solely FlhDC, prior to interaction analysis with the FlhDC:YdiV complex.

Crude soluble extracts of overproduced FlhDC-His_6 protein (Section 5.2.2) were immobilized on a 1 ml HiTrap column before crude extracts of GST-FliT or GST-FliZ were added (Section 5.11.2). The columns were thoroughly washed before an imidazole-rich elution buffer was added. Fractions from each stage of the assay were analysed by SDS-PAGE (Section 2.6.9), to identify the point of elution for each of the proteins.

Both SDS-PAGE gels clearly showed the overproduced FlhDC initially being immobilised onto the column (lanes 1-3) with species at approximately the expected sizes for the FlhD-His_6 and FlhC components (Figure 5.32). Lanes 4-6 then showed the addition of the overproduced FliT (Figure 5.32A) and FliZ (Figure 5.32B) to the column, again with GST-FliT and GST-FliZ at the expected sizes. For both Figure 5.32A and 5.32B, lanes 7-11 show the

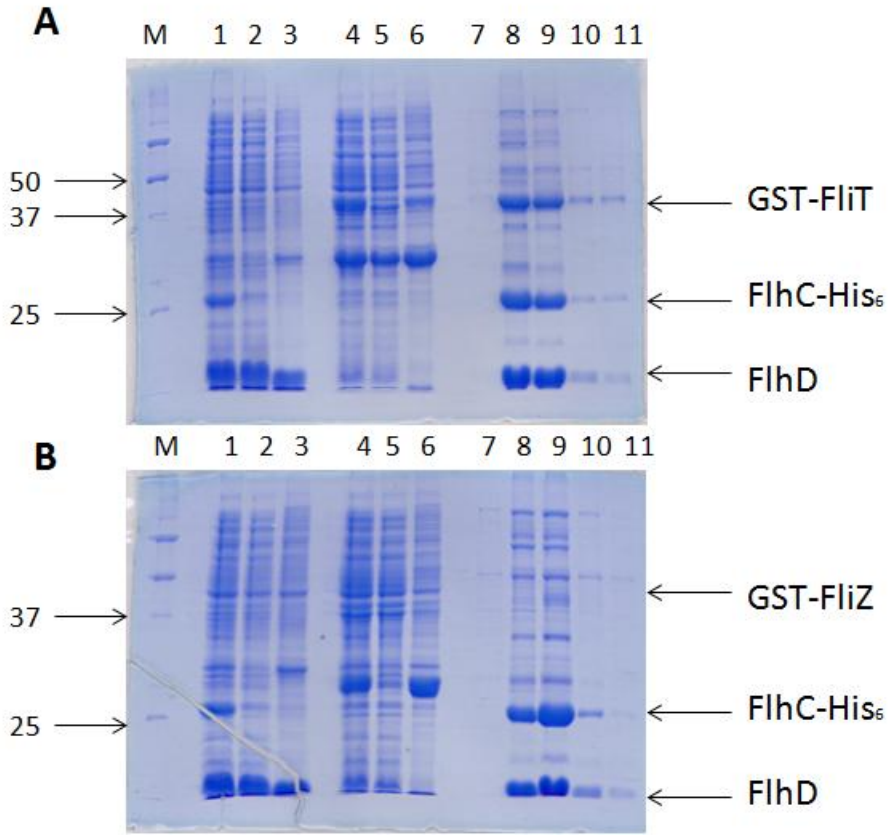


Figure 5.32: Control pull-down assay, testing potential binding partners of the transcription factor FlhDC

Coomassie-stained SDS-PAGE gel analysing fractions of the control pull-down assays.

In both cases, FlhDC-His_6 was immobilised to a 1 ml HiTrap column, prior to addition of GST-tagged FliT (A) and FliZ (B). The columns were then thoroughly washed before an imidazole-rich elution buffer was added (Section 2.8.2).

In both cases, M: Precision plus protein standard (molecular weights indicated in KDa), 1: Soluble FlhDC-His, 2: FlhDC flow-through, 3: Wash step, 4: Soluble FliT/Z, 5: FliT/Z flow-through, 6: Wash step 2, 7-11: Elution fractions.

The positions of FlhC-His_6 (22), FlhD (13), GST- FliT (39) and GST-FliZ (47) are indicated, with their respective expected sizes shown here in KDa.

column elution fractions, with clear FlhDC elution in these fractions. Of note, the GST-FliT protein is present in these elution fractions indicating a direct interaction between FliT and FlhDC (Figure 5.32A). However the GST-FliZ protein was not present in the elution fractions, instead being eluted in a wash step (lane 6) suggesting it does not have a direct interaction with FlhDC (Figure 5.32B).

These results were as expected from literature reports confirming a direct interaction between FliT and FlhDC (Saini *et al.,* 2008; Imada *et al.,* 2010). Following this outcome, it was deemed highly unlikely that the FliZ protein would interact with the FlhDC:YdiV complex, given that it did not directly interact with FlhDC and there were no reports of any interactions between FliZ and YdiV. However, for completeness the competition assays were carried out for both FliT and FliZ.

5.11.4 Competition assays

To determine whether FliT or FliZ interacted or interfered with the known YdiV:FlhDC complex, pull-down assays were carried out once more (Section 2.8.2). Soluble YdiV-His_6 (Section 4.2.2) was firstly immobilized onto a 1 ml HiTrap column before $(FLAG)_3$- FlhDC was added (Section 5.2.2), to produce the immobilised complex YdiV:FlhDC. Following a thorough wash of the column, the potential competitor proteins FliT and FliZ were added (Section 5.11.2). Another wash step was carried out, before an imidazole-rich elution buffer was added to strip His-tagged proteins from the column.

Fractions were retained from each step in the assay and were analysed by SDS-PAGE electrophoresis (Section 2.6.9). The resulting gels showed the overproduction of each protein at the expected sizes in the relevant fractions, with lane 1 showing YdiV-His_6, lane 4 showing the FlhD and FlhC components and lane 7 containing the FliT or FliZ protein (Figure 5.33).

In the case of FliT, the elution fractions in lanes 11-13, show the presence of all four individual proteins, indicating the combined elution of FliT, YdiV and FlhDC (Figure 5.33A). This signifies that an interaction has occurred between these three components, forming a multispecies complex. Of note, the FliT:FlhDC interaction did not break the YdiV:FlhDC interaction, therefore the binding of FliT and YdiV to the FlhDC probably occurs at

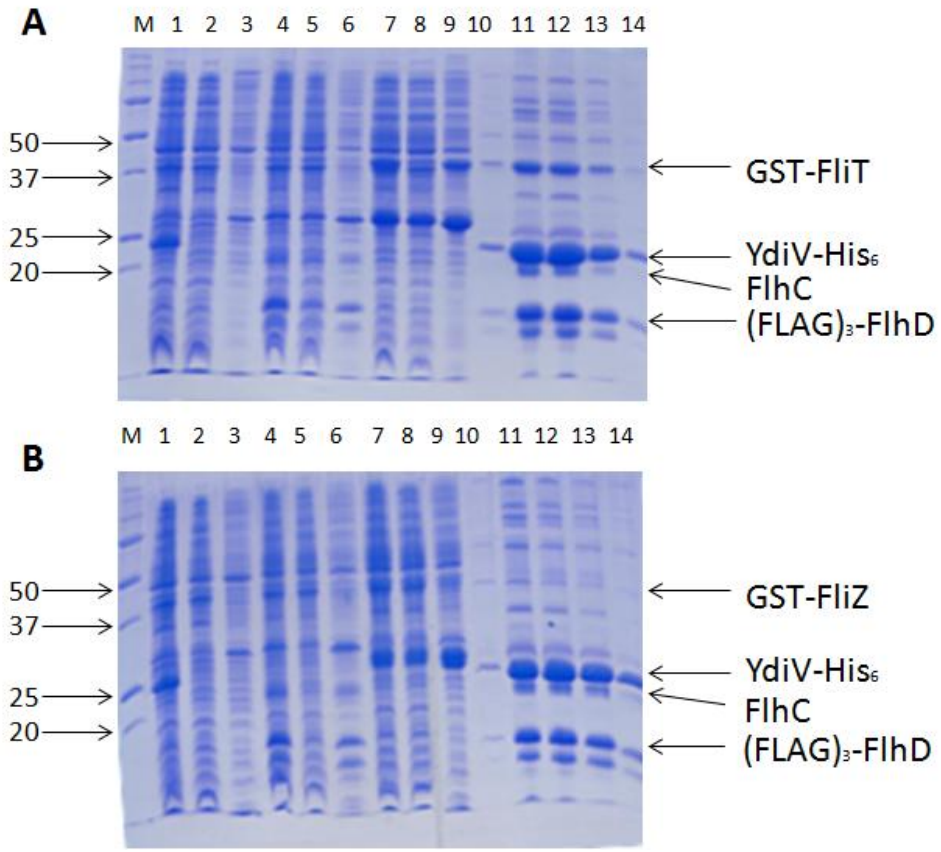


Figure 5.33: Competition pull-down assay, testing potential competitors of the FlhDC:YdiV interaction

Coomassie stained SDS-PAGE gel of the competition pull-down assays (Section 2.8.2), immobilising YdiV-His_6 to a 1 ml HiTrap column. Soluble $(FLAG)_3$-FlhDC was subsequently added to the columns, producing the FlhDC:YdiV complex. Finally potential competitors were tested, adding GST- FliT (A) and GST-FliZ (B) to the column, to detect whether these out-competeted the FlhDC:YdiV interaction.

In both cases, M: Precision plus protein standard (molecular weights indicated in KDa), 1: Soluble YdiV, 2: YdiV flow-through, 3: Wash step, 4: Soluble FlhDC, 5: FlhDC flow-through, 6: Wash step 2, 7: Soluble FliT/Z, 8: FliT/Z flow-through, 9: Wash step 3, 10-14: elution fractions.

The positions of YdiV-His_6 (28), $(FLAG)_3$ –FlhD (16), FlhC (22), GST-FliT (39) and GST-FliZ (47) are indicated, with the expected sizes shown here in KDa.

independent sites, which do not overlap or shield the other binding site. Further investigations *in vivo* would reveal whether this complex is formed in the cell, and if so functional assays could be carried out, assessing whether the complex altered the DNA binding properties of FlhDC.

In contrast, FliZ was present in the wash step in lane 9, indicating the loss of protein from the column (Figure 5.33B). Additionally the protein was absent in the elution fractions in lanes 11-13. This signifies that FliZ neither binds to FlhDC singularly, or to the complex YdiV:FlhDC, and therefore does not regulate FlhDC directly, or indirectly via YdiV.

5.12 Discussion

The interactions between the inactive PDE protein YdiV and the transcription factor FlhDC in *E. coli* have been explored. Successful cloning and overproduction of FlhDC and YdiV proteins was achieved which subsequently enabled the interactions and functions of these proteins to be studied.

At the outset, complex formation between FlhDC and STM1344 (YdiV) was only known in *S. enterica* (Wada *et al.,* 2011) and therefore it was a priority to analyse the interaction in *E. coli*. Pull-down assays using overproduced FlhDC and YdiV protein (Figure 5.5) revealed complex formation in *E. coli*, a result which now supports the same conclusion reached by Wada *et al.* (2012) and Li *et al.* (2012). Analysis of this FlhDC:YdiV interaction by BLItz methods measured the affinity of binding (K_D of 286-395 nM) (Figures 5.7 and 5.8). Isothermal calorimetry revealed the FlhDC:YdiV interaction to be exothermic, with a K_D of 158 nM, similar to the value determined by BLItz analysis. The data fitted an independent binding model, with a stoichiometry of 1:1, indicating that only a single YdiV was binding to the $FlhD_4C_2$ complex under these conditions (Figure 5.9).

At this point, attempts were carried out to determine the growth conditions in which this FlhDC:YdiV complex would form *in vivo*. Polyclonal antibodies were successfully raised against both FlhDC and YdiV proteins and enabled Western blot analysis of the FlhDC and YdiV expression levels in *E. coli*. However, under all conditions tested the levels of YdiV were below the detection limit (Figure 4.21). FlhDC protein expression could be detected, with the highest levels produced when *E. coli* doubled ~ every 3.5 h (Figure 5.15).

Following the identification that an FlhDC:YdiV complex was produced in *E. coli*, and an indication of the growth conditions in which it would form, the role of this complex was investigated. It was previously known that FlhDC bound to class 2 promoter regions of flagellar genes, controlling the expression and synthesis of flagella in *Bacteria* (Liu & Matsumara, 1994). Analysis in *S. enterica* had revealed STM1344 to act as an anti-FlhDC factor, inhibiting FlhDC binding to DNA (Wada *et al.,* 2011). The ability of the protein complex FlhDC:YdiV to bind to DNA was therefore investigated, revealing by EMSA analysis that YdiV was an anti-FlhDC factor in *E. coli* (Figure 5.17). This was in agreement with the conclusions determined by Wada *et al.* (2012) and Li *et al.* (2012), with approximately similar molar ratios of YdiV:FlhDC necessary for DNA dissociation in all three analyses.

Noteably, three distinct species of FlhDC:DNA complexes were observed in the EMSA assays, potentially resulting from various oligomeric forms of FlhDC binding to the promoter region, or FlhDC binding to multiple regions in the DNA (at the 'FlhDC-binding boxes as well as other low affinity sites). Further experimentation is therefore essential to understand the significance of these three species. Analytical centrifugation would enable the molecular weights of the three species to be determined, whilst promoter fragments containing only part of the motif region would reveal exactly which DNA bases were required for FlhDC to bind. Additionally, if cofactors were involved in the DNA:FlhDC interactions, these could be identified by mass spectroscopy techniques.

Kinetic analysis of the DNA and FlhDC interaction was undertaken using BLItz methods, comparing the affinity of the FlhDC complex to a target piece of DNA (a region of P*fliD* containing two FlhDC-binding boxes) and a control piece of DNA (with no FlhDC-binding sites). A clear sequence-specific interaction between FlhDC and the target DNA was detected (K_D of 158 nM) (Figures 5.19 and 5.20). The measured K_D value (with errors) for FlhDC to P*fliD* was reasonably consistent with the K_D values reported for the binding interactions between FlhDC and four other Class II DNA regions (Stafford *et al.,* 2005).

In vitro transcription analysis was subsequently carried out, investigating the effects of YdiV on the FlhDC-mediated *fliD* transcription rate. Rather surprisingly, the highest level of gene transcription occurred in sample conditions of 12 µM FlhDC and 1.5 µM YdiV rather than in conditions lacking any YdiV (Figure 5.22). The role of YdiV in this reaction was clarified by EMSA analysis, revealing that YdiV was part of the FlhDC:DNA binding complex and was

having a direct effect on DNA transcription (Figure 5.23). This FlhDC:DNA:YdiV complex was subsequently identified by Takaya *et al.* (2012) too, supporting these findings. Additionally, this detection of YdiV binding to the $FlhD_4C_2$: DNA complex is in agreement with the transitional YdiV:FlhDC:DNA complex proposed in a model by Li *et al.* (2012). Whilst our data does not fully support the Li *et al.,* (2012) model, it does certainly agree with the presence of the proposed transitional species.

The mechanism in which YdiV enhances transcription is unknown and therefore open to speculation. One possible explanation is that YdiV binding to the FlhDC:DNA complex causes the DNA to increase the degree in which it bends around the FlhDC complex, thereby increasing the accessibility of RNA polymerase to the DNA. From X-ray crystallography of the FlhDC:DNA complex, It is known that upon binding to FlhDC, DNA bends 111^{o} around the complex (Wang *et al.,* 2006). Therefore hypothetically, if this angle of DNA curvature was to increase, this could increase the exposure of DNA to RNA polymerase, thereby increasing transcription. Alternatively, as discussed previously, three shifted species are produced upon FlhDC binding to the P*fliD* DNA, which may be produced by FlhDC binding to lower homology regions of DNA. YdiV binding to the FlhDC:DNA complex, may therefore, prevent the binding of FlhDC to these lower affinity binding sites, therefore increasing the proportion of FlhDC complexes binding at the 'FlhDC-binding box' and therefore initiating gene transcription from this site. As previously stated, investigations into FlhDC binding to short P*fliD* fragments would reveal whether multiple DNA sites were binding to FlhDC, which would subsequently impact the putative role of YdiV in restricting the additional species formation. Analysis of DNA curvature around the FlhDC complex in the presence and absence of YdiV could additionally be studied to assess the veracity of this hypothesis.

In this work, kinetic and thermodynamic parameters have been determined for various interactions. Bringing all this data together, a model has been proposed for the mechanism by which FlhDC and YdiV control DNA transcription in the cell. The BLItz data for the interaction between $FlhD_4C_2$ and P*fliD* DNA yielded a K_D value (~158 nM k_a = 1.9 x 10^2 M^{-1} s^{-1}, k_d = 2.9 x 10^{-5} s^{-1}) (Figure 5.19). Isothermal calorimetry (ITC) indicated that the stoichiometry of the $FlhD_4C_2$:YdiV interaction was 1:1 and the K_D value for this interaction was ~150 nM (Figure 5.9). The BLItz analyses showed that the K_D values for $FlhD_4C_2$ – YdiV interactions were similar when either partner protein was immobilized (K_D ~280 nM when

$FlhD_4C_2$ was immobilized, k_a = 1.6 x 10^3 M^{-1} s^{-1}, k_d = 4.4 x 10^{-4} s^{-1}; K_D ~395 nM, k_a = 3.2 x 10^3 M^{-1} s^{-1}, k_d = 13 x 10^{-4} s^{-1} when YdiV was immobilized). Therefore the K_D value when $FlhD_4C_2$ was immobilised was closer to that obtained by ITC and was the mid-range value and hence the rate constants obtained from BLItz when $FlhD_4C_2$ was immobilised were used in the following analysis. The EMSA experiments (with 2 nM P*fliD*, 15 μM $FlhD_4C_2$ and 0-30 μM YdiV incubated for 30 min before electrophoresis) were used to determine the amount of $FlhD_4C_2$:P*fliD* complex formed in the presence of increasing concentrations of YdiV (Figure 5.17). A simple model (Model 1) was constructed in Copasi 4.8 Build 35 (http://www.copasi.org/tiki-index.php?page=Build35Changelog) to simulate competitive binding of P*fliD* and YdiV to $FlhD_4C_2$ (Figure 5.34; Table 5.7). Steady-state simulations were unable to match the experimental data (i.e. the amount of free DNA in reactions with different ratios of $FlhD_4C_2$ and YdiV; Figure 5.35) and hence simple YdiV sequestration of $FlhD_4C_2$ leading to liberation of free DNA was not sufficient to capture the behaviour of the system.

Supershift EMSAs showed that an $FlhD_4C_2$:P*fliD*:YdiV complex was formed (Figure 5.23), consistent with published work (Takaya *et al.,* 2012). Therefore, Model 1 was extended to include the formation and breakdown of an $FlhD_4C_2$:P*fliD*:YdiV complex (Figure 5.34; Model 2). Analysis of the EMSA data of Takaya *et al.* (2012) suggested a K_D for the $FlhD_4C_2$:P*fliD*:YdiV complex of ~200 nM, similar to the K_D for the $FlhD_4C_2$:YdiV complex measured here. On this basis the k_a and k_d values for this reaction were assigned as k_a = 2 x 10^3 M^{-1} s^{-1}, k_d = 4 x 10^{-4} s^{-1}. There are no experimental data for the breakdown of the $FlhD_4C_2$:P*fliD*:YdiV complex but the EMSAs show that that this complex degrades more readily than the $FlhD_4C_2$:P*fliD* binary complex and so a k_d value (1 x 10^{-2} s^{-1}) ~300-fold greater than that for $FlhD_4C_2$:P*fliD* was assigned. Steady-state simulations of the extended model (Model 2; Figures 5.34; 5.35) were run in Copasi 4.8 Build 35 (http://www.copasi.org/tiki-index.php?page=Build35Changelog) for comparison with the EMSA data shown in Figure 5.17. The reactions and parameters of Model 2 are summarized in Figure 5.34 and Table 5.7. Plotting the steady-state values of free DNA predicted by Model 2 as a function of YdiV concentration matched the experimental data much more closely than those of Model 1 (Figure 5.35). Thus, it was concluded that the formation of the $FlhD_4C_2$:P*fliD*:YdiV ternary complex and its subsequent breakdown are crucial components for YdiV-mediated inhibition of $FlhD_4C_2$-mediated gene expression.

Table 5.7: Model reactions and parameter values

Reaction	Parameter value	Measurement technique	Designation in model
1. F + D = FD	k_a = 1.9 x 10^2 M^{-1} s^{-1} k_d = 2.9 x 10^{-5} s^{-1}	BLItz	k_1 k_2
2. F + Y = FY	k_a = 1.6 x 10^3 M^{-1} s^{-1} k_d = 4.4 x 10^{-4} s^{-1}	ITC (stoichiometry) and BLItz (immobilized $FlhD_4C_2$)	k_3 k_4
3. FD + Y = FYD	k_a = 2 x 10^3 M^{-1} s^{-1} k_d = 4 x 10^{-4} s^{-1}	EMSAs (Takaya *et al.,* 2012) combined with BLItz data for the F + Y = FY reaction	k_5 k_6
4. FYD → FY + D	k_d = 1 x 10^{-2} s^{-1}	Assigned	k_7

Abbreviations: D, P*fliD* DNA (2 nM); F, $FlhD_4C_2$; FD, $FlhD_4C_2$:DNA complex; FY, $FlhD_4C_2$:DNA:YdiV complex; Y, YdiV.

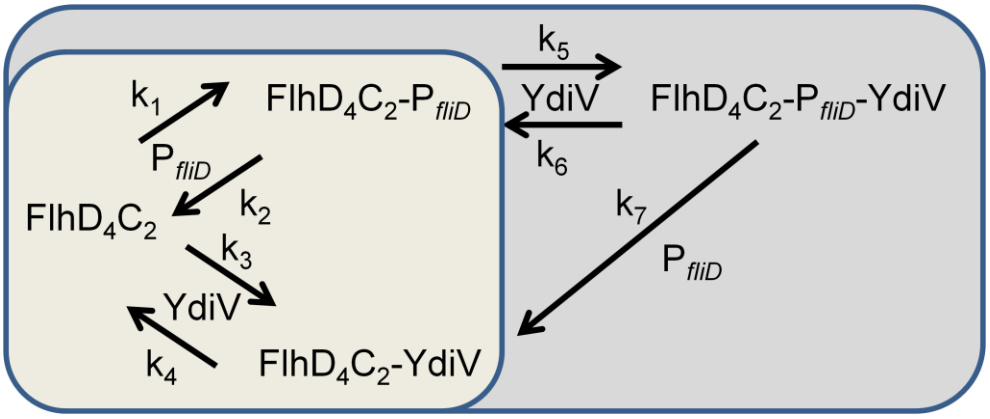


Figure 5.34: The structure of models developed to simulate inhibition of $FlhD_4C_2$-mediated gene expression by YdiV

Components of Model 1 are shown against a pale background; Model 2 incorporates all the components shown. The corresponding reactions and parameters are listed in Table 5.7.

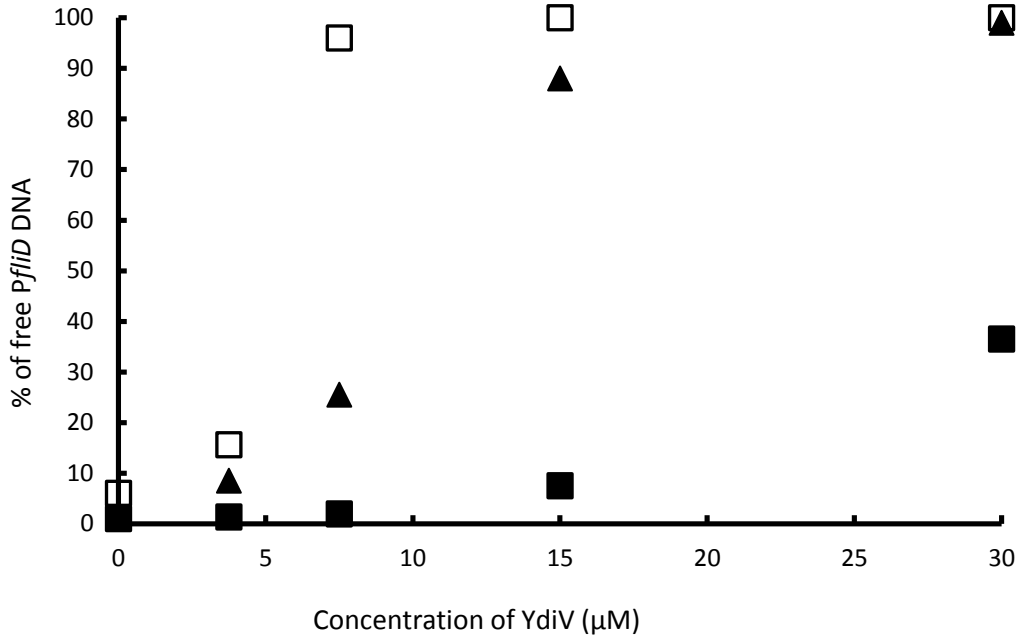


Figure 5.35: Graphical representation of the effect of YdiV on DNA:FlhDC dissociation

The amount of free P*fliD* in reactions containing a fixed concentration of $FlhD_4C_2$ and variable concentrations of YdiV measured in EMSAs compared the output from Model 1 and Model 2. The relative amounts of free DNA (% of total, 2 nM) in the EMSA shown in Figure 5.17 were determined by analysis with ImageJ (Section 2.10.1) and plotted against the concentration of YdiV in each reaction (open squares). The predicted steady-state concentrations of free DNA from simulations of the EMSA reactions using Model 1 (filled squares) and Model 2 (filled triangles) were converted to % values and plotted against YdiV concentration.

Previous studies had revealed a possible YdiV binding interaction with c-di-GMP, thereby giving a reason to investigate the role of nucleotides in the YdiV:FlhDC complex. Ambiguious results were produced, finding no indication of nucleotide binding at all from fluorescence analysis. Pull-down assays revealed varying stoichiometries of FlhDC and YdiV in elution fractions, which may be nucleotide-induced (Figure 5.26). However, there are several experimental limitations which would need to be overcome before any conclusions could be made about this stoichiometry difference. EMSA analysis indicated that the presence of nucleotides (specifically pGpG, c-di-GMP and c-di-AMP) altered the proportion of the three FlhDC:DNA species produced (Figure 5.29). Hypothetically nucleotides binding to the FlhDC complex could feasibly shield DNA binding sites, and therefore influence FlhDC binding to the P*fliD* DNA. The primary concern is therefore determining the identity of each species, which will subsequently put these proportional changes into context. Finally IVT analysis indicated a decrease in *fliD* transcription in the presence of c-di-GMP rather than in the presence of alternative tested nucleotides or in the absence of nucleotides (Figure 5.30). Whilst there are many limitations to these assays, there are small indications that suggest c-di-GMP may interact with FlhDC and YdiV. There was a body of evidence to previously suggest that YdiV may be interacting with c-di-GMP (Section 4.5). This data further supports this notion, indicating a possible interaction of c-di-GMP with YdiV and FlhDC, either individually or as the FlhDC:YdiV complex. Therefore given this data, there is significant reason to further investigate the role of c-di-GMP in the FlhDC and YdiV system.

Competition assays were carried out, using the known FlhDC binding protein FliT, to determine which protein (FliT or YdiV) had the stronger affinity for FlhDC. FliZ was additionally investigated as this is known to indirectly increase the stability of FlhDC, and therefore could be imparting its function via YdiV. Whilst no binding was detected for FliZ with the FlhDC:YdiV complex, an interaction between FliT and FlhDC:YdiV was produced. Rather than FliT outcompeting YdiV to bind to FlhDC, a larger complex was formed consisting of YdiV:FlhDC:FliT (Figure 5.33). This indicates that FliT and YdiV bind to FlhDC at independent sites (FlhC and FlhD respectively), and clearly binding at one site does not impair binding at the other site. Of interest would be to determine whether this complex was formed *in vivo* and if so, to determine the role of this complex. It has been suggested that FliT weakens the FlhD-FlhC interactions within the FlhDC complex and therefore analysis as to whether YdiV reduces this destabilisation would be a logical progression (Imada *et al.,* 2010).

Chapter 6: Discussion

Work in this thesis has investigated the roles of the putative 'degenerate' DGC and PDE proteins in *E. coli* K-12, specifically YeaI and YdiV. Both proteins were chosen because of their non-conserved active site (GGDEF and EAL) motifs, which suggested that they were probably catalytically inactive and possible regulatory or effector proteins. The aims were therefore to characterise both proteins, aiming to determine whether they had regulatory roles in the cell or even if they were c-di-GMP effector proteins (as some degenerate GGDEF/EAL proteins are).

YeaI was shown to be a catalytically inactive DGC protein, in agreement with Sommerfeldt *et al.* (2009) and contradicting Sanchez-Torres *et al.* (2011). It was therefore hypothesised that the protein could have a regulatory role in the cell, however a role for YeaI was not identified. One of the major limitations was the amount of soluble, purified YeaI that could be produced using His-tagged protein constructs of $YeaI_{GGDEF}$ (both with C- and N-terminal His tags). The solubility of $YeaI_{GGDEF}$ was significantly increased upon fusion with a trigger factor chaperone using a pCOLD TF vector, however purification of the protein was still problematic. Thrombin was used to cleave the large TF tag from the $YeaI_{GGDEF}$ domain at a thrombin cleavage site. Whilst cleavage was successful, separation of the TF tag and $YeaI_{GGDEF}$ domain was not achieved, suggesting that the two proteins were interacting. A possible mechanism to disrupt this TF:$YeaI_{GGDEF}$ interaction could be the use of detergents or urea which may enable the isolation of $YeaI_{GGDEF}$. If detergent addition was not successful, it may be necessary to study YeaI intact, including the N-terminal membrane bound domain. Whilst isolation of membrane proteins is difficult, this domain may provide structural stability required for the native folding of the GGDEF domain. The longstanding purification problems identified have clearly limited progress in characterising the protein, however major improvements have been made in YeaI solubility which should advance further analysis of the protein. Comparisons of wild-type *E. coli* and an *E. coli* strain in which the *yeaI* gene is upregulated, could reveal genome-wide phenotypic changes upon *yeaI* upregulation.

Investigations into the degenerate PDE protein, YdiV, were more productive, identifying a cellular role for this protein. The scientific interest in this protein has noticeably grown since the initiation of this project, with many competing research groups becoming active and the protein becoming increasingly well characterised.

The YdiV protein was shown to be catalytically inactive as a c-di-GMP PDE, in agreement with the data from Tchigvintsev *et al.* (2010) and Wada *et al.* (2011), but in contrast to Hisert *et al*. (2005). Nucleotide binding assays were not entirely consistent but the balance of the evidence suggested that c-di-GMP was a putative interacting nucleotide. Both the NMR spectrum of the protein and the partial proteolysis cleavage pattern of YdiV were altered upon c-di-GMP addition. It is therefore tempting to speculate that c-di-GMP binds to YdiV and regulates the activity of this protein in some manner. This possible c-di-GMP binding has not been reported previously, however the X-ray crystallographic structure of the protein did identify a surface groove, which in other EAL domains is the active-site for c-di-GMP hydrolysis (Li *et al.,* 2012). Taken together, the c-di-GMP binding of YdiV and the FlhDC:YdiV complex is certainly a focus point for further investigation. High sensitivity methods such as Isothermal Titration Calorimetry or BLItz techniques could be used to further study this possible interaction.

One concern, which may put the validity of nucleotide-binding assays into doubt, is that the YdiV protein may have been purified in complex with a ligand. All overproduced proteins (including YdiV, Clp and $YeaI_{GGDEF}$) were produced using the *E. coli* BL21 λ(DE3) strain, the genome of which encodes all DGCs and PDEs encoded by *E. coli* MG1655 K-12 except the DGC YedQ (Figure 1.6) (http://www.genome.jp/kegg/kegg2.html). Therefore, there is a high probability that the *E. coli* BL21 λ(DE3) cells will contain c-di-GMP, given the presence of many DGCs and PDEs. This is a major concern when studying the nucleotide-binding character of a protein, as there is a strong likelihood that c-di-GMP binding proteins would be purified already in complex with the ligand, thereby potentially shielding all detection of nucleotide-binding in the assays. Several of the nucleotide-binding assays yielded results which indicated possible c-di-GMP binding but were slightly ambiguous (NMR and partial proteolysis in particular), which could be due to the heterogeneity of the protein (partly bound to c-di-GMP). Additionally, some proteins bind 2 c-di-GMP molecules such as PilZ domains, thereby questioning whether any c-di-GMP binding detected was a single c-di-GMP binding to the protein or whether it was the 2nd c-di-GMP binding to an already c-di-

GMP-YdiV complex. Furthermore, the NMR control protein Clp exists as a dimer, so whilst the NMR spectra showed differences between the Clp sample and the Clp + c-di-GMP sample, this could have been due to a 2nd c-di-GMP binding and may therefore produce a different signal to a single c-di-GMP binding to YdiV (Section 4.5). Whilst this is clearly a concern, all overproduced proteins were produced in *E. coli* BL21 λ(DE3) so therefore the protein assays were at least consistent with one another. There is no doubt that partial proteolysis and NMR assays produced differing results upon YdiV and YdiV + c-di-GMP, however a greater confidence would be achieved by ensuring no c-di-GMP was pre-bound to the protein before these assays were carried out. Futhermore, if all ligand was removed from the protein before the assays were carried out, the stoichiometry of any nucleotide-protein binding could be determined accurately.

One way to overcome this problem in future, would be to incubate the overproduced protein with a known active PDE, such as YfgF, to sequester any c-di-GMP present in the sample and degrade it to pGpG. The proteins could subsequently be separated by gel filtration and the pre-treated overproduced protein used for all nucleotide-binding assays.

Oligomerisation assays of YdiV revealed a monomeric species by gel filtration and glutaraldehyde assays. This was consistent with the conclusions made by Li *et al*. (2012), who independently identified a monomeric species by gel filtration. Based on both sequence and structural analysis this native monomeric species is not at all surprising; YdiV has only 4/7 conserved residues necessary for protein dimerisation, and a much reduced 'dimerisation interface' was observed by X-ray crystallography (Li *et al.,* 2012; Tchigvintsev *et al.,* 2010). Further analysis of the oligomeric state of YdiV could therefore include site-directed mutagenesis of the three non-conserved amino acid residues in YdiV which are required for dimerisation, to analyse whether native dimerisation occurs.

At the beginning of this project, one of the few known characteristics of YdiV was that over-expression of the *ydiV* gene resulted in a marked decrease in cell motility (Dr. M. Lacey, unpublished). Upon *ydiV* over-expression, there was a complete lack of FliC detection and an 80% decrease in flagella numbers. The slight disparity between the results from both assays is likely to be simply due to timescales and therefore not significant, with *fliC* expression presumably repressed more quickly than pre-formed flagella are degraded. Hence, these data are in agreement with the motility effects seen by Dr. M. Lacey.

Additionally, the data (showing that *ydiV* over-expression reduces motility and flagella production) supports the same conclusions made by other groups during the course of this work (Li *et al.,* 2012; Wada, *et al.*, 2012).

Bacterial culture conditions have been a continuing issue throughout this work, with several assays carried out at non-optimal conditions, thereby requiring repetition at optimal conditions. Bacteria induce flagellar expression when they sense environments where motility might give them a fitness advantage, which in *E. coli* is in nutrient starvation conditions and at ~30^{o}C (Wada *et al.,* 2012; Adler & Templeton, 1967).

In this work, motility assays (specifically Western blot and TEM assays) were carried out at 37^{o}C, which in hindsight is a too high temperature for flagellar studies (Section 4.6). It has previously been identified that flagellar synthesis is downregulated above 37^{o}C, with 30^{o}C reported as the optimal temperature for bacterial flagellation (Adler & Templeton, 1967). Additionally, the cultures were grown in 2 L conical flasks, whereas in hindsight round-bottom flasks would be preferential as they would provide a more gentle culture vessel to limit flagellar shearing. Chemostat samples were later used in a bid to determine YdiV and FlhDC expression conditions (Sections 4.7; 5.6), in which bacterial samples were produced at 37^{o}C, in aerobic conditions in a stirred chemostat (Section 2.2.2). Western blot assays were carried out with these samples, however whilst the chemostat samples were readily available, they were not a suitable source for this work and have not generated usable data. When *E. coli* resides in the gut, it respires anaerobically and thereby the bacterial growth conditions should have been anaerobic and without stirring in order to mimic these conditions. Additionally, as discussed previously, flagellar gene expression is maximal at 30^{o}C and therefore samples should have been cultured at 30^{o}C rather than 37^{o}C. Given these significant differences between ideal growth conditions and the chemostat conditions used, the Western blot data should not be used for future work and YdiV and FlhDC protein analysis should be repeated using bacterial samples grown in more optimal conditions.

The *S. enterica* YdiV homolog STM1344 had previously been shown to bind to the transcription factor FlhDC and cause repression of flagellar genes (Wada *et al.,* 2011). Pull-down analysis proved that this interaction did occur in *E. coli*, again in agreement with competing research groups (Wada *et al.,* 2012; Li *et al.,* 2012). Going a step further, kinetic

and thermodynamic analyses showed the proteins to interact specifically and strongly in a 1:1 stoichiometry (K_D value of 286-395 nM from BLItz analysis, 158 nM by ITC).

The stoichiometry of the YdiV:FlhDC complex is a topic of much dispute. Li *et al.* (2012) have reported YdiV to bind to $FlhD_4C_2$, creating four complexes (YdiV:$FlhD_4C_2$, $YdiV_2$:$FlhD_4C_2$, $YdiV_3$:$FlhD_4C_2$ and $YdiV_4$:$FlhD_4C_2$). These species have been identified by size-exclusion chromatography however the physiological relevance of these complexes may be debatable given the low levels of YdiV in the cell. Takaya *et al.* (2012) identified a stable $YdiV_2$:$FlhD_4C_2$ complex upon incubation of the proteins and gel filtration. The ITC analysis performed in this work identified a 1:1 molar ratio of YdiV and $FlhD_4C_2$, indicating that only one YdiV molecule was binding to the complex. Notably, whilst structures have been elucidated for YdiV, $FlhD_4C_2$ and $YdiV_2$-$FlhD_2$, no crystal structure has been produced for YdiV:$FlhD_4C_2$, presumably indicating the long-term instability or transient nature of the various $YdiV_{1-4}$:$FlhD_4C_2$ complexes (Li *et al.,* 2012; Wang *et al.,* 2006). Clearly there are discrepancies in the reported YdiV:FlhDC complex stoichiometries, indicating an area for future investigation and one which would have significant impact on the proposed models for this interaction.

After determining that YdiV and FlhDC interacted, the DNA-binding capacity of the complex was analysed. *S. enterica* studies had shown STM1344 to induce DNA dissociation from FlhDC, producing free DNA upon formation of the STM1334:FlhDC complex (Wada *et al.,* 2011). EMSA assays revealed that the *E. coli* YdiV acted in a similar manner to STM1344, causing DNA (specifically P*fliD*) dissociation, which was also shown by competing research groups during this work (Wada *et al.,* 2012; Li *et al.,* 2012; Takaya *et al.,* 2012). Kinetic analysis of this FlhDC:DNA interaction revealed a strong and sequence specific interaction, K_D of ~158 nM.

EMSA analysis revealed three FlhDC:DNA species. In this case, the chosen DNA was the *fliD* promoter region, a class II promoter of ~250 bases long. Given that this trio of shifted species was novel, this may be a DNA-specific effect rather than a general FlhDC-effect. The most likely hypotheses for the multiple shifted species is the presence of low-specificity weaker binding sites in the DNA (Section 5.7). To clarify this hypothesis, further EMSA assays could be carried out using fragmented DNA, with regions spanning the FlhDC-

binding box region only, and regions other sides of this to visualise whether the pattern of shifted species was altered.

In vitro transcription assays revealed a maximal transcription rate in the presence of FlhDC and substoichiometric YdiV. Control assays confirmed both proteins were part of a DNA-binding complex, indicating the presence of a novel FlhDC:YdiV:DNA complex. Whilst this was previously not identified, the same complex was later identified by Takaya *et al.* (2012). The manner in which this complex causes increased gene transcription is currently unknown. Again, further assays using various different fragments of DNA would enable clarification here.

Given the putative YdiV:c-di-GMP interaction, the FlhDC:YdiV complex was studied in the presence of nucleotides to analyse whether c-di-GMP (or other nucleotides) were involved in the complex formation/dissociation or regulated activity. EMSA and IVT assays provided the most useful data. In both cases, the presence of c-di-GMP had an effect on FlhDC activity, with c-di-GMP altering the proportion of the three FlhDC:DNA species produced compared to a no ligand control sample (Figure 5.29). The *in vitro* transcription levels of *fliD* were decreased in the presence of c-di-GMP (Figure 5.30). Taken together, in addition to the experimental data suggesting YdiV may interact with c-di-GMP and the X-ray structure of YdiV, this provides multiple pieces of evidence to suggest that the proteins YdiV and FlhDC, possibly as a YdiV:FlhDC complex, are interacting with c-di-GMP (Li *et al.,* 2012). Thereby, there is significant reason to further investigate the role of c-di-GMP, and nucleotide analogues, with the complex to determine the role of this putative nucleotide interaction. As discussed previously, pre-treatment of the overproduced protein with an active PDE protein would ensure that the protein was not already bound to c-di-GMP, and therefore yield more conclusive data of the nucleotide-binding character of YdiV and FlhDC.

In order to understand the significance of the data characterising YdiV and the YdiV:FlhDC interaction, both YdiV and FlhDC expression levels were investigated to identify growth conditions in which YdiV:FlhDC would likely be produced. The optimal conditions for YdiV production could not be determined, with YdiV protein levels below the detection limit. Whilst this is a concern, as discussed previously, the chemostat bacterial culture conditions were far from optimal and therefore these assays should be repeated in different conditions. Promoter alignment revealed the presence of both the -35 and -10 consensus sites in the P*ydiV* region but weak -35 sequence conservation, possibly explaining the low

level of *ydiV* transcription due to poor promoter recognition. The low protein concentrations of YdiV were also noted by Wada *et al.* (2012), who identified efficient gene transcription but poor translation in cells. FlhDC expression levels were also analysed but again using sub-optimal bacterial growth conditions and therefore any results produced using this work are far from conclusive.

In order to understand the cellular mechanism in which YdiV affects FlhDC transcriptional activity, various models have been produced. The Li *et al.* (2012) group, identified four YdiV and FlhDC complexes ($YdiV_{1-4}$:$FlhD_4C_2$) which led to the hypothesis of a sequential binding model (Figure 5.10). Based on the structural data known, the group hypothesised that 2 YdiV molecules would bind to the ring-like $FlhD_4C_2$ complex without disrupting the DNA-binding regions of the complex. Therefore a YdiV:FlhDC:DNA complex would be formed. Upon addition of the third and fourth YdiV molecules, the ring-structure of FlhDC would be altered and the DNA-binding regions of the protein would become less exposed, thereby releasing the DNA. This model thereby accounts for the YdiV:FlhDC:DNA complex which has been identified and the DNA dissociation seen by EMSAs. However, the basis of this model is on a sequential binding model with no known trigger for YdiV binding or understanding how YdiV has bias for certain FlhD subunits. Furthermore, this model assumes that YdiV would be in enough of an excess to FlhDC that four molecules could bind to the complex, requiring a 3:1 ratio of YdiV:FlhDC before YdiV-mediated gene repression was produced.

A contrasting model has been produced by Takaya *et al.* (2012) who identified that YdiV not only caused DNA dissociation of FlhDC, but targeted FlhDC for ClpXP-mediated proteolysis. This paper reported that YdiV bound to free- and DNA-bound FlhDC and delivered FlhDC to ClpXP for proteolysis. The half-life of FlhDC was vastly increased in the absence of both YdiV and ClpXP. This model indicates that at low levels of YdiV, the ClpXP machinery causes slow proteolysis of FlhDC, thereby activating flagellar gene expression. Upon an increase in YdiV concentration, YdiV binds to FlhDC (indicated to form an $YdiV_4$-$FlhD_4C_2$ complex), which both prevents FlhDC binding to DNA (and dissociates the DNA already bound) and targets FlhDC for ClpXP proteolysis (Takaya *et al.,* 2012). Whilst this model accounts for the DNA-dissociation of FlhDC:YdiV complexes, it does not indicate in any way the manner in which YdiV binds to the complex.

In this work, a model has been proposed to account for the experimental data obtained (Figure 5.34; 5.35). Using the calculated rate constants of various reactions (FlhDC and YdiV; FlhDC and DNA), a steady-state model was determined and matched to the experimental data. An FlhDC:P*fliD*:YdiV complex was identified by supershift EMSA assays (Figure 5.23), however it was unknown whether this complex was a key component of the molecular system. The model proposes that the FlhDC:P*fliD*:YdiV complex is a crucial step in the YdiV-mediated inhibition of FlhDC-mediated DNA transcription, binding to FlhDC:P*fliD* and promoting the dissociation of DNA from FlhDC.

All three models clearly require further experimentation for verification. The Li *et al.* (2012) model depends on sequential YdiV binding inducing an FlhDC conformational change. Crystallographic studies of an FlhDC complex bound to YdiV at low and high concentrations would therefore be a useful next step, potentially supporting this hypothesis if an FlhDC conformational change was induced. Equally binding studies would be helpful, if the order in which FlhD subunits being occupied could be determined. Further analysis into the ClpXP proteolysis of FlhDC would be useful. Whilst Takaya *et al.* (2012) report on the increased degradation of FlhDC upon YdiV binding to the complex, the exact stoichiometry of this complex is not known, thereby questioning whether there is any difference in the rate of adaptor-mediated proteolysis when $YdiV_2$-$FlhD_4C_2$ is produced rather than $YdiV_4$-$FlhD_4C_2$. Equally, the biochemical manner in which YdiV delivers $FlhD_4C_2$ to the ClpXP protein is another unknown, requiring further analysis for a complete understanding. The model proposed in this work was based on experimentally-derived reaction constants, which provided the basis for the modelling of a steady-state system. Whilst the majority of reaction constants were known, one value was assigned (for the FlhDC:P*fliD*:YdiV dissociation). Kinetic analysis of the rate of FlhDC:P*fliD*:YdiV dissociation is therefore a crucial step to advancing this model. Additionally, this model does not take into account the role of ClpXP, or the rate at which this ClpXP-mediated FlhDC degradation occurs.

Clearly there are several key questions which need to be answered in order to advance our understanding of YdiV; the mechanism of regulation of YdiV binding to FlhDC, the manner in which YdiV binds to FlhDC (whether it is sequential or random and the range of possible stoichiometries), the *in vivo* levels of YdiV and conditions in which high levels of YdiV are produced and a clear understanding of exactly which complexes are formed. Additionally,

investigations into the possible nucleotide binding of YdiV, and whether the FliT:FlhDC:YdiV complex (identified by pull-down analysis) is produced *in vivo*, would add to our understanding of the YdiV and FlhDC cellular roles. Kinetic analysis of the FlhDC interactions with both DNA and YdiV revealed strong high affinity interactions. Of interest however, would be to analyse the binding affinity of YdiV to the FlhDC:DNA complex, and of the FlhDC:YdiV complex to DNA, to determine whether the affinity's were altered. Additionally, given the strong affinity interactions of FlhDC to DNA, it can be assumed that an energy input is required to break apart this interaction. The question then is whether the FlhDC:YdiV interaction alone provides this energy or whether a nucleotide or trigger of some form is involved as well.

To summarise, the work in this thesis has increased our knowledge of the putative 'degenerate' EAL domain protein YdiV, and has revealed this protein to have an important role in flagellar gene expression and motility. In this manner, YdiV is acting as a regulatory protein, as predicted upon its identification as a non-consensus EAL protein (Sommerfeldt *et al.,* 2009). The YdiV protein has been shown to act as an anti-FlhDC factor, causing DNA dissociation from FlhDC at certain concentrations. Upon addition of YdiV at lower concentrations to the FlhDC:DNA system, a large FlhDC:YdiV:DNA complex was produced, which is proposed to have a crucial role in FlhDC:DNA dissociation. Methods in this thesis have produced a reliable system in which to study the YdiV and FlhDC proteins. One area in which to develop may be the buffer components in which to purify and dialyse FlhDC, as currently the stability of this protein is relatively poor, only able to be stored for a few days. Improvements in this would reduce the workload necessary to purify the protein, thereby increasing the ease in which experiments could be carried out. Several suggestions for future experiments have been detailed, however key foci may be the putative YdiV:c-di-GMP interaction, clarification of the EMSA shifted species and investigations into the *ydiV* expression conditions and *in vivo* protein levels.

References

Ades, S.E., Grigorova, I.L. & Gross, C.A. (2003) *Journal of Bacteriology*; 185(8):2512-9. Regulation of the alternative sigma factor sigma (E) during initiation, adaptation, and shutoff of the extracytoplasmic heat shock response in *Escherichia coli*

Adler, J. & Templeton, B. (1967) *Journal of General Microbiology*; 46(2):175-84. The effect of environmental conditions on the motility of *Escherichia coli*

Aldridge, P., Paul, R., Goymer, P., Rainey, P & Jenal, U. (2003) *Molecular Microbiology;* 47(6):1695-708. Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus*

Amikam, D. & Galperin, M.Y. (2006) *Bioinformatics*; 22(1):3-6. PilZ domain is part of the bacterial c-di-GMP binding protein

An, S.-Q., Chin, K.-H., Febrer, M., McCarthy, Y., Yang, J.-G., Liu, C.-L., Swarbreck, D., Rogers, J., Dow, J.M., Chou, S.-H. & Ryan, R.P. (2013) *EMBO;* 32(18):2430-8. A cyclic GMP-dependent signalling pathway regulates bacterial phytopathogenesis

Aravind, L. & Ponting, C. P. (1997) *Trends in biochemical sciences;* 22(12):458-9. The GAF domain: an evolutionary link between diverse phototransducing proteins

Bachmann, B.J. (1972). *Bacteriological reviews*; 36(4):525-557. Pedigrees of some mutant strains of *Escherichia coli* K-12

Baraquet, C., Murakami, K., Parsek, M.R. & Harwood, C.S. (2012) *Nucleic Acids Research;* 40(15):7207-18. The FleQ protein from *Pseudomonas aeruginosa* functions as both a repressor and an activator to control gene expression from the pel operon promoter in response to c-di-GMP

Barembruch, C. & Hengge, R. (2007) *Molecular Microbiology;* 65(1):76-89. Cellular levels and activity of the flagellar sigma factor FliA of *Escherichia coli* are controlled by FlgM-modulated proteolysis

Barrick, J.E. & Breaker, R.R. (2007) *Genome biology;* 8(11):R239. The distributions, mechanism, and structures of metabolite-binding riboswitches

Bassler, B.L. & Losick, R. (2006) *Cell;* 125(2):237-46. Bacterially speaking

Bjarnsholt, T. (2013) *APMIS Suppplementary;* (136):1-51. The role of bacterial biofilms in chronic infections

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shou, Y. (1997) *Science*; 277(5331):1453-1462. The complete genome sequence of *Escherichia coli* K-12

Boehm, A., Kaiser, M., Li, H., Spangler, C, Kasper, C.A., Ackermann, M., Kaever, V., Sourjik, V., Roth, V. & Jenal, U. (2010) *Cell*; 141:107-116. Second messenger mediated adjustment of bacterial swimming velocity

Bomchil, N., Watnick, P. & Kolter, R. (2003) *Journal of Bacteriology;* 185(4):1384-90. Identification and characterisation of a *Vibrio cholerae* gene, *mbaA*, involved in maintenance of biofilm architecture

Borrelli, E., Montmayeur, J.P., Foulkes, N.S. & Sassone-Corsi, P. (1992) *Critical reviews in Oncogenesis;* 3(4):321-38. Signal transduction and gene control: the cAMP pathway

Botsford, J.L. & Harman, J.G. (1992) *Microbiology Review;* 56(1):100-22. Cyclic AMP in prokaryotes

Bradford, M.M. (1976) *Analytical Biochemistry;* 72:248-54. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding

Brombacher, E., Baratto, A., Dorel, C. & Landini, P. (2006) *Journal of Bacteriology;* 188(6):2027-37. Gene expression regulation by the curli activator CsgD protein: Modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion

Chan, C., Paul, R., Samoray, D., Amiot, N.C., Giese, B., Jenal, U. & Schirmer, T. (2004) *Proceedings of the National Academy of Science;* 101(49):17084-9. Structural basis of activity and allosteric control of diguanylate cyclase

Chang, A.L., Tuckerman, J.R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. & Giles-Gonzalez, M.-A. (2001) *Biochemistry;* 40(12):3420-6. Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinus*, is a heme-based sensor

Chang, W.S., van de Mortel, M., Nielsen, L., Nino de Guzman, G., Li, X. & Halverson, L.J. (2007) *Journal of Bacteriology;* 189(22):8290-9. Alginate production by *Pseudomonas putida* creates a hydrated microenvironment and contributes to biofilm architecture and stress tolerance under water-limiting conditions

Chatterji, D & Ojha, A.K. (2001) *Current opinion in Microbiology;* 4(2):160-5. Revisiting the stringent response, ppGpp and starvation signalling

Chilcott, G.S. & Hughes, K.T. (2000) *Microbiology and Molecular Biology Reviews;* 64(4):694-708. Coupling of flagellar gene expression to flagellar gene expression in *Salmonella enterica* Serovar Typhimurium and *Escherichia coli*

Chin, K.H., Lee, Y.C., Tu, Z.L., Chen, C.H., Tseng, Y.H., Yang, J.M., Ryan, R.P., McCarthy, Y., Dow, J.W., Wang, A.H. & Chou, S.H. (2010) *Journal of Molecular Biology;* 396(3):646-62. The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signalling to virulence gene expression in *Xanthomonas campestris*

Christen, B., Christen, M., Paul, R., Schmid, F., Folcher, M., Jenoe, P., Meuwly, M. & Jenal, U. (2006) *Journal of Biological Chemistry;* 281(42):32015-24. Allosteric control of cyclic di-GMP signalling

Christen, M., Christen, B., Allan, M.G., Folcher, M., Jeno, P., Grzesiek, S., Jenal, U. (2007) *Proceedings of the National Academy of Sciences;* 104(10):4112-7. DgrA is a member of a new family of cyclic diguanosine monophosphate receptors and controls flagellar motor function in *Caulobacter crescentus*

Christen, M., Christen, B., Folcher, M., Schauerte, A. & Jenal, U. (2005) *Journal of Biological Chemistry;* 280(35):30829-37. Identification and characterisation of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP

Claret, L. & Hughes, C. (2002) *Journal of Molecular Biology;* 321(2):185-99. Interaction of the atypical prokaryotic transcription activator $FlhD_2C_2$ with early promoters of the flagellar gene hierarchy

Corrigan, R.M. & Gründling, A. (2013) *Nature Reviews in Microbiology;* 11(8):513-24. Cyclic di-AMP: another second messenger enters the fray

Costanzo, A. & Ades, S.E. (2006) *Journal of Bacteriology*; 188(13):4627-34. Growth phase-dependent regulation of the extracytoplasmic stress factor, sigmaE, by guanosine 3', 5'-bispyrophosphate (ppGpp)

Costerton, J.W., Stewart, P.S. & Greenberg, E.P. (1999) *Science;* 284(5418):1318-22. Bacterial biofilms: a common cause of persistent infections

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D. & von Heijne, G. (2005) *Science;* 308(5726):1321-3. Global topology analysis of the *Escherichia coli* inner membrane proteome

De Wulf, P., McGuire, A.M., Liu, X. & Lin, E.C. (2002) *Journal of Biological Chemistry*; 277(29):26652-61. Genome-wide profiling of promoter recognition by the two-component response regulator CpxR-P in *Escherichia coli*

De, N., Navarro, M.V., Raghavan, R.V. & Sondermann, H. (2009) *Journal of molecular biology;* 393(3):619-33. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR

Duerig, A., Abel, S., Folcher, M., Nicollier, M., Schwede, T., Amiot, N. Giese, B. & Jenal, U. (2009) *Genes and Development;* 23(1):93-104. Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression

Egler, M., Grosse, C., Grass, G. & Nies, D.H. (2005) *Journal of Bacteriology*; 187(7):2297-307. Role of the extracytoplasmic function protein family sigma factor RpoE in metal resistance of *Escherichia coli*

Elasri, M.O. & Miller, R.V. (1999) *Applied environmental microbiology;* 65(5):2025-31. Study of the response of a biofilm bacterial community to UV radiation

Elena, S.F., Whittam, T.S., Winkworth, C.L., Riley, M.A. & Lenski, R.E. (2005) *International Microbiology*; 8:271-278. Genomic divergence of *Escherichia coli* strains: evidence for horizontal transfer and variation in mutation rates

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, T., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J. & Gardner, T.S. (2007) *PLoS Biology;* 5(1):e8. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles

Fazli, M., O'Connell, A., Nilsson, M., Niehaus, K., Dow, J.M., Givskov, M., Ryan, R.P., Tolker-Nielsen, T. (2011) *Molecular Microbiology;* 82(2):327-41. The CRP/FNR family protein Bcam1349 is a c-di-GMP effector that regulates biofilm formation in the respiratory pathogen *Burkholderia cenocepacia*

Forouhar, F., Lew, S., Seetharaman, J., Sahdev, S., Xiao, R., Ciccosanti, C., Foote, E.L., Wang, H., Everett, J.K., Nair, R., Acton, T.B., Rost, B., Montelione, G.T., Hunt, J.F. & Tong, L. Unpublished, structure via PDB(www.rcsb.org/pdb), assigned code 3HVW. Crystal structure of the GGDEF domain of the PA2567 protein from *Pseudomonas aeruginosa*

Fux, C.A., Costerton, J.W., Stewart, P.S. & Stoodley, P. (2005) *Trends in Microbiology;* 13(1):34-40. Survival strategies of infectious biofilms

Galperin, M.Y. (2004) *Environmental microbiology;* 6(6):552-67. Bacterial signal transduction network in a genomic perspective

Galperin, M.Y. (2006*) *Journal of Bacteriology;* 188(12):4169-82. Structural classification of bacterial response regulators: diversity of output domains and domain combinations

Galperin, M.Y., Natale, D.A., Aravind, L. & Koonin, E.V. (1999) *Journal of Molecular Microbiology and Biotechnology;* 1(2):303-5. A specialized version of the HD hydrolase domain implicated in signal transduction

Galperin, M.Y., Nikolskaya, A.N. & Koonin, E.V. (2001) *FEMS Microbiology letters;* 203(1):11-21. Novel domains of the prokaryotic two-component signal transduction systems

Gama-Castro, S., Jiménez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Peñaloza-Spinola, M.I., Contreras-Moreira, B., Segura-Salazar, J., Muñiz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-MartÍnez, C., Abreu-Goodger, C., RodrÍgues-Penagos, C., Miranda-Ríos, J., Morett, E., Merino, E., Huerta, A.M., Treviño-Quintanilla, L & Collado-Vides, J. (2008) *Nucleic Acids Research;* 36(Database issue):D120-4. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation

García, B., Latasa, C., Solano, C., García-del Portillo, F., Gamazo, C. & Lasa, I. (2004) Molecular *Microbiology;* 54(1):264-77. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation

Green, N.M. (1975) *Advanced Protein Chemistry;* 29: 85-133. Avidin

Gunasekera, T.S., Csonka,L.N. & Paliy,O. (2008) *Journal of Bacteriology;* 190(10):3712-20. Genome-wide transcriptional responses of *Escherichia coli* K-12 to continuous osmotic and heat stresses

Hammer, B.K. & Bassler, B.L. (2003) *Molecular Microbiology;* 50(1):101-4. Quorum sensing controls biofilm formation in *Vibrio cholera*

Hantke, K., Winkler, K. & Schultz, J.E. (2011) *Journal of Bacteriology*; 193(5): 1086-9. *Escherichia coli* exports cyclic AMP via TolC

Heikaus, C.C., Pandit, J. & Klevit, R.E. (2009) *Structure;* 17(12):1551-7. Cyclic nucleotide binding GAF domains from phosphodiesterases- structural and mechanistic insights

Hengge, R. (2009) *Nature reviews;* 7(4):263-73. Principles of c-di-GMP signalling in bacteria

Henry, J.T. & Crosson, S. (2012) *Annual review of Microbiology;* 65:261-86. Ligand binding PAS domains in a genomic, cellular and structural context

Hickman, J.W, Tifrea,D.F. & Harwood,C.S. (2005) *Proceedings of the National Academy of Science;* 102(40):14422-7. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels

Hickman, J.W. & Harwood, C.S. (2008) *Molecular Microbiology;* 69(2):376-89. Identification of FleQ from *Pseudomonas aeruginosa* as a c-di-GMP-responsive transcription factor

Hisert, K.B., MacCoss, M., Shiloh, M.U., Darwin, K.H., Singh, S., Jones, R.S., Ehrt, S., Zhang, Z., Gaffney, B.L., Gandotra, S., Holden, D.W., Murray, D. & Nathan, C. (2005) *Molecular Microbiology;* 56(5):1234-45. A glutamate-alanine-leucine (EAL) domain protein of *Salmonella* controls bacterial survival in mice, antioxidant defence and killing of macrophages: role of cyclic diGMP

Hoffmann, A., Bukau, B. & Kramer, G. (2010) *Biochemica et Biophysica Acta;* 1803(6):650-61. Structure and function of the molecular chaperone trigger factor

Hou, S., Freitas, T., Larsen, R.W., Piatibratov, M., Sivozhelezov, V., Yamamoto, A., Meleshkevitch, E.A., Zimmer, M., Ordal, G.W. & Alam, M. (2001) *Proceedings of the National Academy of Science;* 98(16):9353-8. Globin-coupled sensors: A class of heme-containing sensors in Archaea and Bacteria

Hughes, D.T., Terekhova, D.A., Liou, L., Hoyde, C.J., Sahl, J.W., Patankar, A.V., Gonzalez, J.E. Edrington, T.S., Rasko, D.A. & Sperandio, V.(2010) *Proceedings of the National Academy of Science;* 107(21):9831-6. Chemical sensing in mammalian host-bacterial commensal associations

Iguchi, A., Thomson, N.R., Ogura, Y., Saunders, D., Ooka, T., Henderson, I.R., Harris, D., Asadulghani, M., Kurokawa, K., Dean, P., Kenny, B., Quail, M.A., Thurston, S., Dougan, G., Hayashi, T., Parkhill, J. & Frankel, G. (2009) *Journal of Bacteriology*; 191(1):347-354. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* 0127:H6 strain E2348/69

Imada, K., Minamino, T., Kinoshita, M., Furukawa, Y. & Namba, K. (2010) *Proceedings of the National Academy of Science;* 107(19):8812-7. Structural insight into the regulatory mechanisms of interactions of the flagellar type III chaperone FliT with its binding partners

Jenal, U. & Malone,J. (2006) *Annual review of genetics;* 40:385-407. Mechanisms of cyclic-di-GMP signalling in Bacteria

Jishage, M. & Ishihama, A. (1998) *Proceedings of the National Academy of Science*; 95(9):4953-8. A stationary phase protein in *Escherichia coli* with binding activity to the major sigma subunit of RNA polymerase

Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. (1996) *Journal of Bacteriology*; 178(18):5447-51. Regulation of RNA polymerase sigma subunit synthesis is *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions

Johnson, J.G., Murphy, C.N., Sippy, J., Johnson, T.J. & Clegg, S. (2011) *Journal of Bacteriology;* 193(14):3453-60. Type 3 fimbriae and biofilm formation are regulated by the transcriptional regulators MrkHI in *Klebsiella pneumoniae*

Kalia, D., Merey, G., Nakayama, S., Zheng, Y., Zhou, J., Luo, Y., Guo, M., Roembke, B.T. & Sintim, H.Q. (2013) *Chemical Society Reviews;* 42(1):305-41. Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP, (p)ppGpp signalling in bacteria and implications in pathogenesis

Kaplan, J.B., Izano, E.A., Gopal, P., Karwacki, M.T., Kim, S., Bose, J.L., Bayles, K.W. & Horswill, A.R. (2012) *mBio;* 3(4):e00198-12. Low levels of β-lactam antibiotics induce extracellular DNA release and biofilm formation in *Staphylococcus aureus*

Karimova, G., Pidoux,J. Ullmann,A. & Ladant,D. (1998) *Proceedings of the National Academy of Science;* 95(10):5752-6. A bacterial two-hybrid system based on a reconstituted signal transduction pathway

Kazmierczak, B.I., Lebron, M.B. & Murray, T.S. (2006) *Molecular Microbiology;* 60(4):1026-43. Analysis of FimX, a phosphodiesterase that governs twitching motility in *Pseudomonas aeruginosa*

King, T., Lucchini, S., Hinton, J.C. & Gobius, K. (2010) *Applied Environmental Microbiology;* 76(19):6514-28. Transcriptomic analysis of *Escherichia coli* 0157:H7 and K-12 cultures exposed to inorganic and organic acids in stationary phase reveals acidulant-and stress-specific acid tolerance responses

Kirillina, O., Fetherston, J.D., Bobrov, A.G., Abney, J. & Perry, R.D. (2004) *Molecular Microbiology;* 54(1):75-88. HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*

Ko, M. & Park, C. (2000) *Journal of Molecular Biology;* 303(3):371-82. Two novel flagellar components and H-NS are involved in the motor function of *Escherichia coli*

Kolter, R., Siegele, D.A. & Tormo, A. (1993) *Annual review of microbiology;* 47:855-74. The stationary phase of the bacterial life cycle

Krasteva, P.V., Fong, J.C.N., Shikuma, N.J., Beyhan, S., Navarro, M.V.A.S, Yildiz, F.H. & Sondermann, H. (2010) *Science;* 327(5967):866-8. *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP

Kumar, M. & Chatterji, D. (2008) *Microbiology;* 154(Pt 10):2942-55. Cyclic di-GMP: a second messenger required for long-term survival, but not for biofilm formation, in *Mycobacterium smegmatis*

Kutsukake, K. & Iino, T. (1994) *Journal of Bacteriology;* 176(12):3598-605. Role of the FliA-FlgM regulatory system on the transcriptional control of the flagellar regulaon and flagellar formation in *Salmonella typhimurium*

Kutsukake, K., Ikebe, T. & Yamamoto, S. (1999) *Genes & Genetic systems*; 74(6):287-92. Two novel regulatory genes, fliT and fliZ, in the flagellar regulon of *Salmonella*

Kuwajima, G., Asaka, J., Fujiwara, T., Fujiwara, T., Node,K. & Kondo,E. (1986) *Journal of Bacteriology*; 168(3):1479-83. Nucleotide sequence of the hag gene encoding flagellin of *Escherichia coli*

Lacey, M., Agasing, A., Lowry, R. & Green, J. (2013) *Open Biology*; 3(6):130046. Identification of the YfgF MASE1 domain as a modulator of bacterial responses to aspartate

Lacey, M.M., Partridge, J.D. & Green, J. (2010) *Microbiology*; 156(Pt 9):2873-86. *Escherichia coli* K-12 YfgF is an anaerobic cyclic di-GMP phosphodiesterase with roles in cell surface remodelling and the oxidative stress response

Lee, V.T., Matewish, J.M., Kessler, J.L., Hyodo, M., Hayakawa, Y. & Lory, S. (2007) *Molecular Microbiology*; 65(6):1474-84. A cyclic di-GMP receptor required for bacterial exopolysaccharide production

Lehnen, D., Blumer, C., Polen, T., Wackwitz, B., Wendisch, V.F. & Unden, G. (2002) *Molecular Microbiology;* 45(2):521-32. LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in *Escherichia coli*

Li, W., Li, N., Wang, F., Guo, L., Huang, Y., Liu, X., Wei, T., Zhu, D., Liu, C., Pan, H., Xu, S., Wang, H-W & Gu, L. (2012) *Nucleic Acids Research*; 40(21):11073-85. Structural insight of a concentration-dependent mechanism by which YdiV inhibits *Escherichia coli* flagellum biosynthesis and motility

Liu, X. & Matsumura, P. (1994) *Journal of Bacteriology*; 176(23):7345-51. The FlhD/FlhC complex, a transcriptional activator of the *Escherichia coli* flagellar class II operons

Macnab, R.M. (1977) *Proceedings of the National Academy of Science*; 74(1):221-5. Bacterial flagella rotating in bundles: A study in helical geometry

Macnab, R.M. (1996) '*Escherichia coli and Salmonella: cellular and molecular biology'* in Neidhardt *et al.,* 2nd ed. Washington, D.C.: American Society for Microbiology; pp. 123–145. Flagella and Motility

Macnab, R.M. (2003) *Annual review of microbiology*; 57:77-100. How bacteria assemble flagella

Maguire, B.A. & Wild, D.G.(1997) *Molecular Microbiology*; 23(2):237-45. The roles of proteins L28 and L33 in the assembly and function of *Escherichia coli* ribosomes *in vivo*

Martinez-Hackert, E. & Hendrickson, W.A. (2009) *Cell*; 138(5):923-934. Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone

Matz, C., McDougald, D., Moreno, A.M., Yung, P.Y., Yildíz, F.H. & Kjelleberg, S. (2005) *Proceedings of the National Academy of Science;* 102(46):16819-24. Biofilm formation and phenotypic variation enhance predation-driven persistence of *Vibrio cholera*

May, T., Ito, A. & Okabe, S. (2009) *Antimicrobial agents and chemotherapy*; 53(11):4628-39. Induction of multidrug resistance mechanism in *Escherichia coli* biofilms by interplay between tetracycline and ampicillin resistance genes

McClelland, M., Florea, L., Sanderson, K., Clifton, S.W., Parkhill, J., Churcher, C., Dougan, G., Wilson, R.K. & Miller, W. (2000) *Nucleic Acids Research*; 28(24):4974-86. Comparison of the *Escherichia coli* K-12 genome with sampled genomies of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typthimurium, Typhi and Paratyphi

Mecsas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J. & Gross, C.A. (1993) *Genes and development*; 7(12B):2618-28. The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins

Méndez-Ortiz, M.M., Hyodo, M., Hayakawa, Y. & Membrillo-Hernández, J. (2006) *Journal of Biological Chemistry*; 281(12):8090-9. Genome-wide transcriptional profile of *Escherichia coli* in response to high levels of second messenger 3',5'-cyclic diguanylic acid

Merighi, M., Lee, V.T., Hyodo, M., Hayakawa, Y. & Lory, S. (2007) *Molecular Microbiology*; 65(4):876-95. The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-comtaining receptor Alg44 are required for alginate biosynthesis in *Pseudomonas aeruginosa*

Minasov, G., Padavattan, S., Shuvalova, L., Brunzelle, J.S., Miller, D.J., Baslé, A., Massa, C., Collart, F.R., Schirmer, T. & Anderson, W.F. (2009) *Journal of Biological Chemistry*; 284(19):13174-84. Crystal structures of YkuI and its complex with second messenger cyclic di-GMP suggest catalytic mechanism of phosphodiester bond cleavage by EAL domains

Mutalik, V.K., Nonaka, G., Ades, S.E., Rhodius, V.A. & Gross, C.A. (2009) *Journal of Bacteriology*; 191(23):7279-87. Promoter strength properties of the complete sigma E regulon of *Escherichia coli* and *Salmonella enterica*

Nambu, J.F., Lewis, J.O., Wharton, Jr. K.A. & Crews, S.T. (1991) *Cell*; 67(6):1157-67. The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development

Navarro, M.V., Newell, P.D., Krasteva, P.V., Chatterjee, D., Madden, D.R., O'Toole, G.A., Sondermann, H. (2011) *PLoS biology;* 9(2):e1000588. Structural basis for c-di-GMP-mediated inside-out signalling controlling periplasmic proteolysis

Navarro, M.V.A.S., De, N., Bae, N., Wang, Q. & Sondermann, H. (2009) *Structure;* 17(8):1104-16. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimX

Newell, P.D., Monds, R.D. & O'Toole, G.A. (2009) *Proceedings of the National Academy of Science;* 106(9):3461-6. LapD is a bis (3', 5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonad fluorescens* Pf0-1

Nikolskaya, A.N., Mulkidjanian, A.Y., Beech, I.B. & Galperin, M.Y. (2003) *Journal of Molecular Microbiology and biotechnology*; 5(1):11-6. MASE1 and MASE2: two novel integral membrane sensory domains

Nobre, L.S., Al-Shahrour, F., Dopazo, J. & Saraiva, L.M. (2009) *Microbiology*; 155(Pt 3):813-24. Exploring the antimicrobial action of a carbon monoxide-releasing compound through whole-genome transcription profiling of *Escherichia coli*

O'Toole, G., Kaplan, H.B. & Kolter, R. (2000) *Annual Review of Microbiology*; 54:49-79. Biofilm formation as microbial development

Olsén, A., Jonsson, A. & Normark, S. (1989) *Nature*; 338(6217):652-5. Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coli*

Parsek, M.R. & Greenberg, E.P. (2005) *Trends in Microbiology*; 13(1):27-33. Sociomicrobiology: the connections between quorum sensing and biofilms

Paul, K, Nieto, V., Carlquist, W.C., Blair, D.F. & Harshey, R.M. (2010) *Molecular Cell*;38:128-139. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism

Paul, R., Abel, S., Wassmann, P., Beck, A., Heerklotz, H. & Jenal, U. (2007) *Journal of Biological Chemistry*; 282(40):29170-7. Activation of the diguanylate cyclase PleD by phosphorylation-mediated dimerization

Paul, R., Weiser, S., Amiot, N.C., Chan, C., Schirmer, T., Giese, B. & Jenal, U. (2004) *Genes and Development*; 18(6):715-27. Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain

Pesavento, C., Becker, G., Sommerfeldt, N., Possling, A., Tschowri, N., Mehlis, A. & Hengge, R. (2008) *Genes and Development*; 22(17):2434-46. Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*

Petters, T., Zhang, X., Nesper, J., Treuner-Lange, A., Gomez-Santos, N., Hoppert, M., Jenal, U. & Søgaard-Andersen, L. (2012) *Molecular Microbiology*; 84(1):147-65. The orphan histidine protein kinase SgmT is a c-di-GMP receptor and regulates composition of the extracellular matrix together with the orphan DNA binding response regulator DigR in *Myxococcus xanthus*

Ponting, C.P. & Aravind, L. (1997) *Current Biology*; 7(11):R674-7. PAS: a multifunctional domain family comes to light

Povolotsky, T.L. & Hengge, R. (2012) *Journal of Biotechnology*; 160:10-16. 'Life-style' control networks in *Escherichia coli*: Signaling by the second messenger c-di-GMP

Pratt, J.T., Tamayo, R., Tischler, A.D. & Camilli, A. (2007) *Journal of Biological Chemistry*; 282(17):12860-70. PilZ domain proteins bind cyclic diguanylate and regulate diverse processes in *Vibrio cholerae*

Pratt, L.A. & Kolter, R. (1998) *Molecular Microbiology*; 30(2):285-93. Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type 1 pili

Prigent-Combaret, C., Brombacher, E., Vidal, O., Ambert, A., Lejeune, P., Landini, P. & Dorel, C. (2001) *Journal of Bacteriology*; 183(24):7213-23. Complex regulatory network controls initial adhesion and biofilm formation of Escherichia coli via regulation of the *csgD* gene

Qi, Y., Chuah, M. L. C., Dong, X., Xie, K., Luo, Z., Tang, K. & Liang, Z.-X. (2011) *Journal of Biological Chemistry;* 286(4):2910-7. Binding of cyclic diguanylate in the non-catalytic EAL domain of FimX induces a long-range conformational change

Raina, S., Missiakas, D. & Georgopoulos, C. (1995) *EMBO*; 14(5):1043-55. The rpoE gene encoding the sigma E (sigma 24) heat shock sigma factor of *Escherichia coli*

Rao, F., Yang, Y., Qi, Y. & Liang, Z.-X. (2008) *Journal of Bacteriology*; 190(10):3622-31. Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*

Redl, B., Walleczek, J., Stöffler-Meilicke, M & Stöffler, G. (1989) *European journal of biochemistry;* 181(2):351-6. Immunoblotting analysis of protein-protein crosslinks within the 50 S ribosomal subunit of *Escherichia coli*. A study using dimethylsuberimidate as crosslinking reagent

Reuter, M., Mallett, A., Pearson, B.M. & van Vliet, A.H.M. (2010) *Applied environmental microbiology*; 76(7):2122-8. Biofilm formation by *Campylobacter jejuni* is increased under aerobic conditions

Reznikofl, W.S., Siegele, D.A., Cowinl, D.W. & Gros, C.A. (1985) *Annual review of Genetics*; 19:355-87. The regulation of transcription initiation in bacteria

Rhodius, V.A. & Mutalik, V.K. (2010) *Proceedings of the National Academy of Science*; 107(7):2854-9. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE

Richter, A.M., Povolotsky, T.L., Wieler, L.H. & Hengge, R. (2014) *EMBO Molecular Medicine*; 6(12):1622-1637.Cyclic-di-GMP signalling and biofilm-related properties of the *Shiga* toxin-producing 2011 German outbreak *Escherichia coli* O104:H4

Römling, U., Bian, Z., Hammar, M., Sierralta, W.D. & Normark, S. (1998) *Journal of Bacteriology*; 180(3):722-31. Curli fibres are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation

Römling, U., Galperin, M.Y. & Gomelsky, M. (2013) *Microbiology and Molecular Biology Reviews*; 77(1):1-52. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger

Römling, U., Gomelsky, M. & Galperin, M.Y. (2005) *Molecular Microbiology*; 57(3):629-39. C-di-GMP: the dawning of a novel bacterial signalling system

Ross, P., Mayer, R., Weinhouse, H., Amikam, D., Hyggirat, Y & Benziman, M. (1990) *The Journal of Biological Chemistry;* 265(31):18933-43. The cyclic diguanylic acid regulatory system of cellulose synthesis in *Acetobacter xylinum*

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana,P., Mayer, R., Braun, S., De Vroom, E., Van der Marel, G.A., Van Boom, J.H. & Benziman, M. (1987) *Nature*; 325(6101):279-81. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. & Gourse, R.L. (1993) *Science;* 262(5138):1407-13. A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase

Rouvière, P.E., De Las Penãs, A., Mecsas, J., Lu, C.Z., Rudd, K.E. & Gross, C.A. (1995) *EMBO*; 14(5):1032-42. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*

Ryan, R.P., Fouhy, Y., Lucey, J.F., Crossman, L.C., Spiro, S., He, Y.W., Zhang, L.H., Heeb, S., Camara, M., Williams, P. & Dow, J.M. (2006) *Proceedings of the National Academy of Science*; 103(17):6712-7. Cell-cell signalling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover

Ryjenkov, D.A., Simm, R., Römling, U. & Gomelsky, M. (2006) *Journal of Biological Chemistry*; 281(41):30310-4. The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria

Ryjenkov, D.A., Tarutina, M., Moskvin, O.M. & Gomelsky, M. (2005) *Journal of Bacteriology*; 187(5):1792-8. Cyclic diguanylate is a ubiquitous signalling molecule in Bacteria: insights into biochemistry of the GGDEF protein domain

Saini, S., Brown, J.D., Aldridge, P.D. & Rao, C.V. (2008) *Journal of Bacteriology*; 190(14):4979-88. FliZ is a posttranslational activator of $FlhD_4C_2$-dependent flagellar gene expression

Sanchez-Torres, V., Hu, H. & Wood, T. (2011) *Applied microbiology and biotechnology;* 90(2):651-8. GGDEF proteins YeaI, YedQ, and YfiN reduce early biofilm formation and swimming motility in *Escherichia coli*

Schleif, R. (2000) *Trends in Genetics*; 16(12):559-65. Regulation of the L-arabinose operon of *Escherichia coli*

Schmidt, A.J., Ryjenkov, D.A. & Gomelsky, M. (2005) *Journal of Bacteriology*; 187(14):4774-81. The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains

Schröder, H.C., Wang, X., Manfrin, A., Yu, S.-H., Grebenjuk, V.A., Korzhev, M., Wiens, M., Schlossmacher, U. & Müller, W.E.G. (2012) *Journal of Biological Chemistry*; 287(26):22196-22205. Acquisition of structure-guiding and structure-forming properties during maturation from the pro-silicatein to the silicatein form

Serres, M.H., Kerr, A.R.W., McCormack, T.J. & Riley, M. (2009) *Biology Direct*; 4(46). Evolution by leaps: gene duplication in bacteria

Shi, W., Zhou, Y., Wild, J., Adler, J. & Gross, C.A. (1992) *Journal of Bacteriology*; 174(19):6256-63. DnaK, DnaJ and GrpE are required for flagellum synthesis in *Escherichia coli*

Shin, S. & Park, C. (1995) *Journal of Bacteriology*; 177(16):4696-702. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR

Simm, R., Morr, M., Kader, A., Nimtz, M & Römling, U. (2004) *Molecular Microbiology*; 53(4):1123-34. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility

Simm, R., Remminghorst, U., Ahmad, I., Zakikhany, K. & Römling, U. (2009) *Journal of Bacteriology*; 191(12):3928-37. A role for the EAL-like protein STM1344 in regulation of CsgD expression and motility in *Salmonella enterica* Serovar *Typhimurium*

Simms, A.N. & Mobley, H.L.T. (2008) *Journal of Bacteriology*; 190(10):3747-56. Multiple genes repress motility in uropathogenic *Escherichia coli* constitutively expressing type 1 fimbriae

Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N. & Hengge, R. (2009) *Microbiology*; 155(Pt 4):1318-31. Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in *Escherichia coli*

Sperandio, V., Torres, A.G. & Kaper, J.B. (2002) *Molecular Microbiology*; 43(3):809-21. Quorum sensing *Escherichia coli* regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in *E. coli*

Spurbeck, R.R., Alteri, C.J., Himpsl, S.D. & Mobley, H.L.T. (2013) *Journal of Bacteriology*; 195(14):3156-64. The multifunctional protein, YdiV represses P Fimbriae-Mediated adherence in uropathogenic *Escherichia coli*.

Stafford, G.P., Ogi, T. & Hughes, C. (2005) *Microbiology*; 151(Pt 6):1779-88. Binding and transcriptional activation of non-flagellar genes by the *Escherichia coli* flagellar master regulator $FlhD_2C_2$

Stoodley, P., Sauer, K., Davies, D.G. & Costerton, J.W. (2002) *Annual review of Microbiology*; 56:187-209. Biofilms as complex differentiated communities

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H. & Breaker, R.R. (2008) *Science;* 321(5887):411-3. Riboswitches in eubacteria sense the second messenger cyclic di-GMP

Suzuki,K., Babitzke,P., Kushner,S.R. & Romeo,T. (2006) *Genes and development*; 20(18):2605-17. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E

Takaya, A., Erhardt, M., Karata, K., Winterberg, K., Yamamoto, T. & Hughes, K.T. (2012) *Molecular Microbiology*; 83(6):1268-84. YdiV: a dual function protein that targets FlhDC for ClpXP-dependent degradation by promoting release of DNA-bound FlhDC complex

Tal, R., Wong, H. C., Calhoon, R., Gelfand, D., Fear, A. L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P. & Benziman, M. (1998) *Journal of Bacteriology*; 180(17):4416-25. Three cdg operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organisation and occurrence of conserved domains in isoenzymes

Tamayo, R., Tischler, A.D. & Camilli, A. (2005) *Journal of Biological Chemistry*; 280(39):33324-30. The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase

Taylor, B.L.& Zhulin, I.B. (1999) *Microbiology & Molecular Biology Reviews*; 63(2):479-506. PAS domains: internal sensors of oxygen, redox potential, and light

Tchigvintsev, A., Xu, X., Singer, A., Chang, C., Brown, G., Proudfoot, M., Cui, H., Flick, R., Anderson, W.F. & Joachimiak, A. (2010) *Journal of Molecular Biology*; 402(3):524-38. Structural insight into the mechanism of c-di-GMP hydrolysis by EAL domain phosphodiesterases

Tischler, A.D. & Camilli, A. (2004) *Molecular Microbiology*; 53(3):857-69. Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation

Tischler, A.D. & Camilli, A. (2005) *Infection and immunity*; 73(9):5873-82. Cyclic diguanylate regulates *Vibrio cholerae* virulence gene expression

Tschowri, N., Busse, S. & Hengge, R. (2009) *Genes and development*; 23(4):522-34. The BLUF-EAL protein YcgF acts as a direct anti-repressor in a blue-light response of *Escherichia coli*

Tucker, N.P., D'Autrèaux, B., Yousafzai, F.K., Fairhurst, S.A., Spiro, S. & Dixon, R. (2008) *Journal of Biological Chemistry;* 283(2):908-18. Analysis of the nitric oxide-sensing non-heme iron centre in the NorR regulatory protein

Tuckerman, J.R., Gonzalez, G., Sousa, E.H., Wan, X., Saito, J.A., Alam, M. & Gilles-Gonzalez, M.A. (2009) *Biochemistry*; 48(41):9764-74. An oxygen-sensing diguanylate cyclase and phosphodiesterase couple for c-di-GMP control

Visvalingam, J., Hernandez-Doria, J.D. & Holley, R.A. (2013) *Applied and environmental microbiology*; 79(3):942-50. Examination of the genome-wide transcriptional response of *Escherichia coli* 0157:H7 to cinnamaldehyde exposure.

Wada, T., Hatamoto, Y & Kutsukake, K. (2012) *Microbiology*; 158(Pt 6):1533-42. Functional and expressional analyses of the anti-$FlhD_4C_2$ factor gene *ydiV* in *Escherichia coli*.

Wada, T., Morizane, T., Abo, T., Tominaga, A., Inoue-Tanaka, K. & Kutsukake, K. (2011) *Journal of Bacteriology*; 193(7):1600-11. EAL domain protein YdiV acts as an anti-$FlhD_4C_2$ factor responsible for nutritional control of the flagellar regulon in *Salmonella enterica* serovar typhimurium

Walleczek, J., Redl, B., Stöffler-Meilcke, M & Stöffler, G. (1989) *Journal of Biological Chemistry*; 264(7):4231-7. Protein-protein cross-linking of the 50 S ribosomal subunit of *Escherichia coli* using 2-iminothiolane. Identication of cross-links by immunoblotting techniques

Wang, S., Fleming, R.T., Westbrook, E.M., Matsumura, P. & McKay, D.B. (2006) *Journal of Molecular Biology*; 355(4):798-808. Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription

Watson, W.T., Minogue, T.D., Val, D.L., von Bodman, S.B. & Churchill, M.E. (2002) *Molecular Cell*; 9(3):685-94. Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing

Weber, H., Pesavento, C., Possling, A., Tischendorf, G. & Hengge, R. (2006) *Molecular Microbiology*; 62(4):1014-34. Cyclic-di-GMP-mediated signalling within the σ^S network of *Escherichia coli*

Wei, B.L., Brun-Zinkernagel, A.-M., Simecka, J.W., Prüβ, B.M., Babitzke, P & Romeo, T. (2001) *Molecular Microbiology*; 40(1):245-56. Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*

Wüthrich, K. (1986) *Wiley Interscience*. NMR of proteins and nucleic acid.

Yamamoto, K., Hirao, K., Oshima, T., Aiba, H., Utsumi, R. & Ishihama, A. (2005) *Journal of Biological Chemistry*; 280(2):1448-56. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*

Zhou, B., Peng, D., Tan, X., Yuan, D., Liu, X. & Zhang, L. (2014) *Protein expression and purification*; pii: S1046-5928(14)00109-0. doi: 10.1016/j.pep.2014.05.008. Expression of *Sapium sebiferum* (L.) Roxb stearoyl-acyl carrier protein desaturase in *Escherichia coli*. [Epub ahead of print]

Zhou, X., Meng,X., & Sun,B. (2008) *Cell Research*; 18(9):937-48. An EAL domain protein and cyclic AMP contribute to the interaction between the two quorum sensing systems in *Escherichia coli*

Zhulin, I.B., Taylor, B.L. & Dixon, R. (1997) *Trends in Biochemical Science*; 22(9):331-3. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox

Zogaj, X., Nimtz, M., Rohde, M., Bokranz, W. & Römling, U. (2001) *Molecular Microbiology*; 39(6):1452-63. The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix