



Strathclyde Institute for Pharmaceutical and Biomedical Sciences

Towards Structural Dynamics of Bacterial GntR Proteins

Kirsty A. Robb

Thesis presented in fulfilment of the requirement for the degree of
Doctor of Philosophy

2015

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Publications

Several papers have been published as a result of work carried out during this project. These are as follows and can be found printed at the back of this thesis.

Adamczyk, K., Candelaresi, M., Kania, R., Robb, K., Bellota-Antón, C., Greetham, G. M., Pollard, M. R., Towrie, M., Parker, A. W., Hoskisson, P. A., Tucker, N. P. & Hunt, N. T. (2012). *Phys Chem Chem Phys* **14**, 7411-7419.

Adamczyk, K., Candelaresi, M., Robb, K. A., Walsh, M., Parker, A., PA, Tucker, N. & Hunt, N. (2012). *Meas. Sci. Technol.* **23**.

Candelaresi, M., Gumiero, A., Adamczyk, K., Robb, K., Bellota-Antón, C., Sangal, V., Munnoch, J., Greetham, G. M., Towrie, M., Hoskisson, P. A., Parker, A. W., Tucker, N. P., Walsh, M. A. & Hunt, N. T. (2013). *Org Biomol Chem* **11**, 7778-7788.

Shaw, D. J., Adamczyk, K., Frederix, P. W., Simpson, N., Robb, K., Greetham, G. M., Towrie, M., Parker, A. W., Hoskisson, P. A. & Hunt, N. T. (2015). *J Chem Phys* **142**, 212401.

Acknowledgements

I would like to thank all the people whose support and have made this thesis possible. First and foremost, I would like to thank my supervisors, Dr Paul Hoskisson and Dr Martin Walsh for their support, patience and guidance throughout this project.

Particular thanks to Dr Daniel Shaw for help with 2D-IR analysis, Dr Simone Culurgioni for help with crystallography and Dr Petra Lukacik for helpful (or not so helpful sometimes!) discussions and advice about crystallography, Dr Louise Bird and Anil Verma for assistance with HTP cloning and crystallisation. Thank you also to everyone in the lab, HW601, for making the last four years so enjoyable.

I must also thank the BBSRC and Diamond Light Source for providing funding, without which this project wouldn't have happened.

I also have to mention Dr Moira Wilson, without whom I would never have found an interest in protein science.

And finally, none of this would have been possible without the love and support of my family, friends and partner, John. Thank you for believing in me when I didn't believe in myself.

To John, thank you for your and love and support. I couldn't have done it without you.

To my best friends, Olivia and Rosslyn, thank you for putting up with some really crazy rants from me at times and making me many cups of tea. I promise never to rant about it again!

Abstract

The GntR super-family of transcriptional regulators are ubiquitously distributed throughout the prokaryotic world. There are currently 231,015 sequences in the Pfam database pertaining to GntR proteins (PF000392) from over 17,000 bacterial species. The proteobacteria, firmicutes and actinobacetrria account for 95% of all GntR sequences in the database. Yet despite their wide distribution, very few have been studied in depth.

The general structure of GntR proteins is a highly homologous helix-turn-helix (HTH) domain linked to a C-terminal effector binding/oligomerisation (Eb/O domain) by a flexible linker. Binding of effector molecules to the Eb/O domain causes conformational changes within the protein allowing binding or release of specific DNA operator sequences which controls gene transcription or repression.

The work described here aims to address some of the unknowns relating to GntR-like regulators in terms of structural and dynamical information about their general mechanism of function of this highly diverse family of proteins. Thirty target GntR proteins were selected from sequence analysis of PF000392. These thirty targets were extensively analysed *in silico* revealing four proteins with novel C-terminal topologies (Gp26, FucR, Reut_B4779 and Colbol_00895) for which there is no structural information. Four proteins (HutC, DevA, DevE and Gp26) were also studied experimentally by using protein biochemistry and x-ray crystallography. The structure of HutC has been solved. However, it has not been fully refined due to the publication of a homologous structure during the course of this work. DevE crystals were obtained and examined by x-ray, further refinement of crystallography conditions and selenomethionine protein preparation resulted in poor phasing and low resolution diffracting crystals. Work in this thesis also aimed to refine methodology of 2D-IR spectroscopy to examine protein molecular dynamics at femtosecond resolution with a view to applying this to GntR-like proteins. The model system for this work was InhA and isoniazid.

Contents

Chapter 1 Introduction	1
1.1 DNA binding proteins	1
1.2 History of GntR Proteins	1
1.3 Structure and function of GntRs.....	5
1.4 Subfamilies	9
1.4.1 FadR	9
1.4.2 HutC	12
1.4.3 MocR.....	13
1.4.4 YtrA	15
1.4.5 AraR and PlmA	15
1.4.6 DevA	15
1.5 Evolution of GntR-like regulators	17
1.6 DNA binding and regulation	18
1.7 Studying protein structure/function relationships	20
1.7.1 Bioinformatics.....	21
1.7.2 Size Exclusion Chromatography Multiple Angle Laser Light Scattering	
21	
1.7.2 X-ray crystallography.....	22
1.7.3 Infrared spectroscopy techniques	24

1.8	Aims and objectives	30
	Aims of the project	30
Chapter 2: Materials and Methods		32
2.1	Strains and their cultivation.....	32
2.1.1	Bacterial strains, culture and storage	32
2.1.2	Plasmids	32
2.1.3	Media.....	32
2.2	Molecular biology	36
2.2.1	Genomic DNA Isolation	36
2.2.2	Primers and PCR.....	36
2.2.3	Agarose Gel Electrophoresis	42
2.2.4	Cloning.....	43
2.2.5	Chemically competent <i>E. coli</i>	45
2.2.6	Transformation of <i>E. coli</i>	45
2.2.7	Complementation of PA14 HutC::Tn7 with pKR034.....	46
2.2.8	Alkaline lysis Plasmid isolation.....	47
2.3	Protein techniques	48
2.3.1	Protein Over-expression.....	48
2.3.2	Production of Selenomethionine DevE.....	48
2.3.3	Protein Purification	49
2.3.4	SDS-PAGE.....	50

2.3.5 Bradford Assay (Bradford, 1976)	51
2.3.6 Electrophoretic mobility shift assay (EMSA)	51
2.3.7 Size exclusion chromatography multiple angle laser light scattering (SEC-MALLS).....	52
2.4 Crystallisation of GntR-like proteins	53
2.4.1 Crystallisation trials	53
2.4.2 X-ray diffraction and data collection	55
2.5 Two Dimensional Infra-Red Spectroscopy (2D-IR)	56
2.6 Bioinformatics	56
Chapter 3: Selection, phylogenetics and <i>in silico</i> characterisation of GntR-like protein targets.....	57
3.1.1 Bioinformatics and target selection	57
3.1.2 Secondary structure analysis	66
3.1.3 Phylogenetic analysis of selected GntR regulators	70
3.2 Target Cloning strategies and optimisation of sequences for crystallisation	76
3.2.1 Cloning into pET vectors	76
3.2.2 Surface entropy reduction (SERp) mutants	76
3.2.3 High-throughput cloning.....	79
3.3 Construction of GntR-like protein overexpression plasmids	90
3.3.1 Cloning of full length GntR target genes into pET100 vectors	90

3.3.2	High throughput cloning of sequence optimised GntR targets	96
3.3.3	Cloning of potential upstream promoter regions	106
3.4	Protein overexpression and optimisation of purification	113
3.4.1	Overexpression testing of GntR proteins from pET100 vectors	113
3.4.3	Overexpression testing of GntR proteins from HTP constructs.....	121
3.4.3	Optimisation of protein purification	131
Chapter 4 Purification, biophysical analysis and crystallisation.....		134
4.1	DevA	134
4.1.1	Purification.....	134
4.1.2	Size Exclusion Chromatography Multi-Angle Laser Light Scattering Analysis	134
4.1.3	DevA Crystallisation.....	138
4.2	DevE.....	144
4.2.1	Purification of <i>apo</i> -DevE.....	144
4.2.2	SEC-MALLS	144
4.2.3	Crystallisation of DevE.....	147
4.2.4	Co-purification of DevE-18mer complex	153
4.3	HutC	163
4.3.1	Purification.....	163
4.3.2	SEC-MALLS	163
4.3.3	Crystallisation	166

4.3.4 Data Collection, structure solution and refinement	171
4.4 Gp26	183
4.4.1 Purification.....	183
4.4.2 SEC-MALLS	183
4.4.3 Crystallisation	186
Chapter 5 Functional analysis of the GntR-like proteins HutC, DevA and DevE	188
5.1 Examination of protein-DNA binding by EMSA.....	188
5.1.1 HutC-DNA binding is diminished by intermediates of histidine utilisation	191
5.1.2 DevE binds to the promoter region of DevA but not vice versa.....	195
5.1.2.2 DevE binding studies	198
5.2 Phenotypic analysis of a PA14 HutC transposon insertion mutant.....	203
Chapter 6 Two- Dimensional Infrared Spectroscopy (2D-IR).....	212
6.1 2D-IR of Isoniazid (INH)	212
6.2 2D-IR of InhA as a model for future protein dynamic analysis	217
6. 4 Preliminary 2D-IR spectroscopic analysis of HutC	232
Chapter 7 General discussion	234
Appendix 1 Secondary structure predictions	243

Appendix 2 Expression test results for HTP plasmids	275
Appendix 3 Vector maps	277
Appendix 4 Thermofluor screening conditions	280
Appendix 5 Upstream DNA fragments used in EMSAS	281
Appendix 6 Tables of normalised peak intensities obtained for 2D-IR spectroscopy data	282
References	285

Chapter 1 Introduction

1.1 DNA binding proteins

DNA binding proteins are essential to gene control within living organisms. They include polymerases, nucleases and transcription factors. These are responsible for a wide variety of functions including DNA packaging, replication, and repair as well as transcription and repression of genes. Transcription factors, specifically, facilitate the transcription of DNA to RNA. These include a wide variety of proteins that initiate and regulate gene transcription; a process that is essential for information transfer in cells, allowing organisms to respond rapidly to their surrounding environment. Their importance is highlighted by the vast number of transcription factors in bacterial genomes and the diversity of functions (Aravind *et al.*, 2005). The defining feature of transcription factors is that they contain a sequence specific DNA binding domain (DBD; (Ohlendorf *et al.*, 1983). The DBD is not limited to one type of motif with the leucine zipper, zinc finger and helix-turn-helix being common motifs in transcription factors; while zinc fingers and leucine zippers are most common in eukaryotes, the helix-turn-helix domain is the most widely distributed domain in prokaryotes (Koonin *et al.*, 1995, Luscombe *et al.*, 2000).

1.2 History of GntR Proteins

The understanding of DNA binding proteins was greatly enhanced by pioneering work in the 1980s that led to the identification of a tri-helical domain in the bacteriophage Lambda proteins (cI and cro) and *lac* operon repressor, LacI. These proteins were identified as having essential roles in DNA binding (Ohlendorf *et al.*, 1982, Ohlendorf & Matthews, 1983, Sauer *et al.*, 1982). The helix-turn-helix DNA

binding motif is comprised of three α helices forming a tri-helical bundle. Helix-turn-helix (HTH) DNA binding proteins are amongst the most common proteins within bacteria allowing gene expression to be modified rapidly in response to their surrounding environment. Gene expression is generally modulated by a metabolic intermediate molecule (effector) that binds to the HTH transcription factor allowing transcription or repression of the gene which they regulate. The GntR superfamily is one of the most ubiquitously distributed groups of HTH transcription factors in bacteria (>200,000 sequences; Pfam database; Aug 2015).

The family of GntR proteins takes its name from the repressor of the gluconate operon of *Bacillus subtilis*. This protein family was first identified by David Haydon and John Guest in 1991, when they discovered significant amino acid sequence similarities amongst the N-terminal regions of several bacterial proteins (Haydon & Guest, 1991). In particular, P30 and GenA from *Escherichia coli* were noted to closely resemble the GntR protein (Buck & Guest, 1989). Furthermore, the N-terminal regions conferred a helix-turn-helix (HTH) motif indicating that these were DNA binding proteins and probably sequence specific binding proteins due to their similarities with λ Cro protein. BLAST searches against GntR, P30 and GenA resulted in a number of proteins from various organisms showing similarities between the sequences and the sequence of λ Cro protein; a DNA binding protein first characterised in the 1980s (Ohlendorf *et al.*, 1982).

GntR regulatory proteins are transcription factors which function to regulate gene expression in response to environmental stimuli. GntR proteins hold great interest as they are ubiquitously spread across the prokaryotic world although there are a few members in other phyla. There are 231,015 sequences over 17,666 species (Pfam,

Aug 2015). The majority of GntRs belong to the Proteobacteria, Firmicutes and Actinobacteria (**Figure 1.1**). In fact, GntR sequences in these phyla alone account for almost 95% of all known bacterial GntR sequences. This is most likely an evolutionary response to the complex and diverse environments that these bacteria are found in; where the internal environment is required to respond rapidly to changing conditions in the surrounding environment (Hoskisson & Rigali, 2009). Archaea and eukaryotes are also known to have GntR regulators but they are not as prevalent as in the bacterial phyla. There also two known GntRs in bacteriophages; *Streptomyces* phage ϕ C31 has a GntR known as Gp26 and there is another known in enterophage ϕ p27. Gp26 has been shown to interact with the bacteriophage proteins required for integration/excision of the genome in to the bacterial host (Khaleel *et al.*, 2011). The function of the latter within the phage is unknown, as is the case for many GntR proteins, although these have been most likely acquired from host genomes.

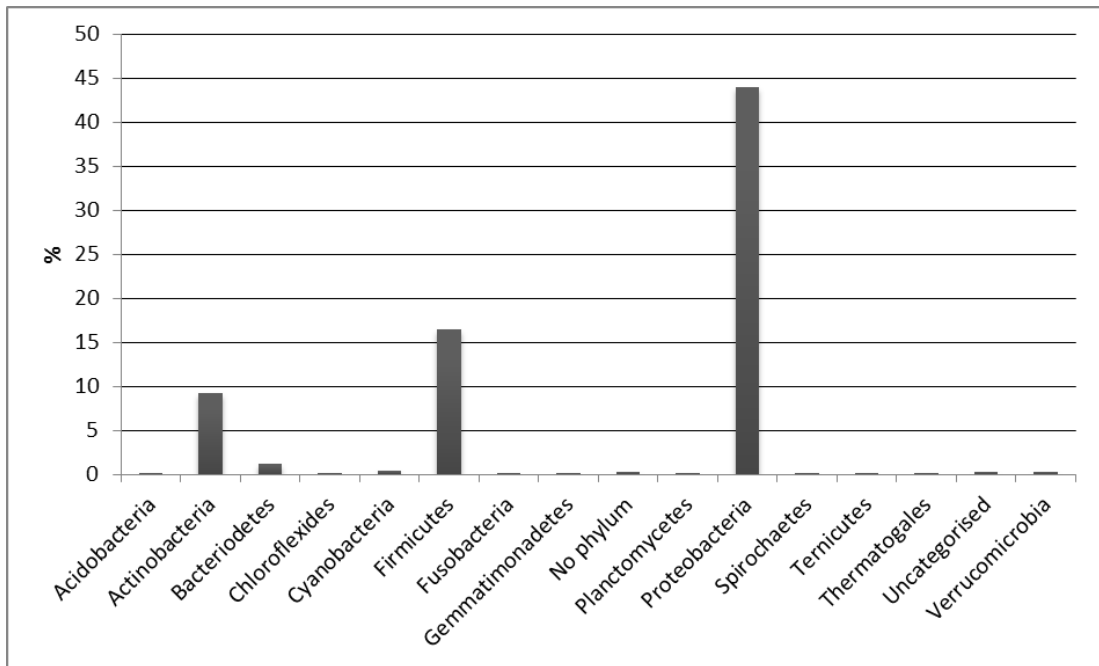


Figure 1.1 Distribution of GntR-like regulators present in bacteria phyla. Data were taken from sequences available in the Pfam database (August 2015) and are expressed as a percentage present in available sequenced bacterial genomes (Ensembl Bacteria; Kersey *et al.*, 2016).

1.3 Structure and function of GntRs

The general structure of GntRs is an N-terminal DNA binding domain (HTH) and a C-terminal effector/oligomerisation (Eb/O) binding domain (**Figure 1.2 A**). Within the HTH are three α -helices (**Figure 1.2 B**), at least one of which binds to the major groove of DNA and is known as the recognition helix (Aravind *et al.*, 2005). This helix confers sequence specificity to GntR proteins and as well as enabling protein-DNA interaction to take place. The effector molecule binds to the C-terminal domain causing a conformational change in the protein which in turn causes the protein to either bind or release DNA thereby allowing transcription or repression of transcription of genes.

The N-terminal domain of the GntR-like regulators are extremely well conserved and the HTH is well known as the classification “marker” for the family (Haydon & Guest, 1991). The HTH domain shows very little sequence homology overall; the secondary structure prediction is highly conserved, however, showing the characteristic tri-helical arrangement (**Figure 1.3**). In stark contrast, however, the C-terminal domains are highly heterologous (Rigali *et al.*, 2002). When considering the vast number of metabolic processes that these proteins help regulate this should not be unusual. Most of the GntR superfamily remains largely uncharacterised with the exception of a few (FadR, AraR, PhnF, GabR, YvoA, YydK). BLAST searches of the full-length protein demonstrate the high conservation of the GntR HTH domain as thousands of hits are reported.

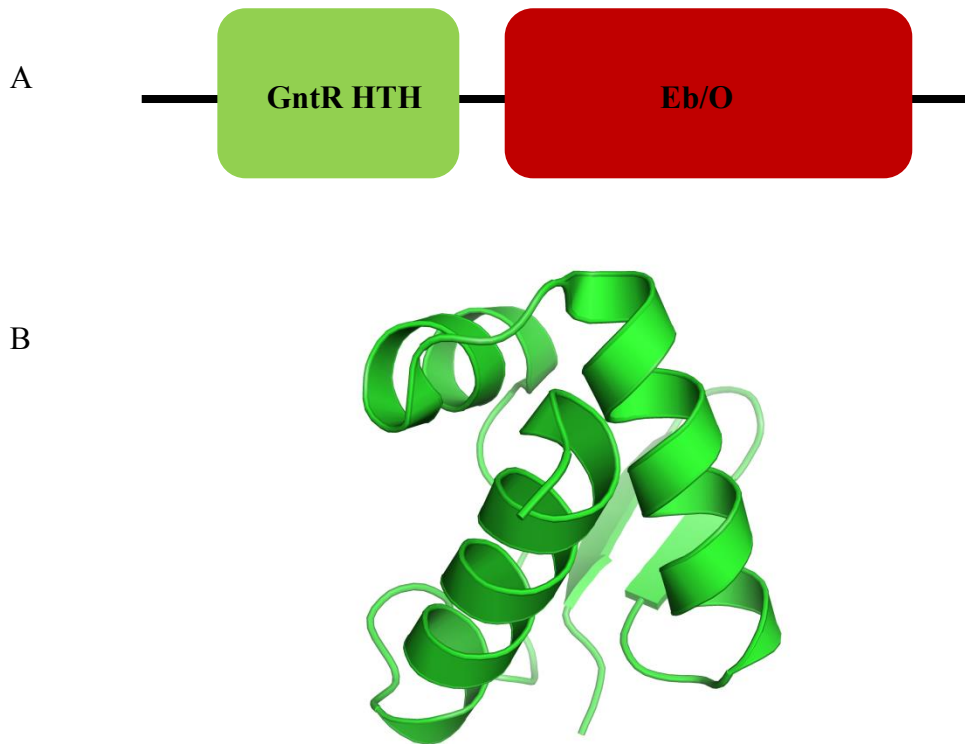


Figure 1.2(A) Structural representation of a GntR protein showing the N-terminal DNA-binding domain and the longer effector binding/oligomerisation domain. **(B)** Cartoon representation of the tri-helical HTH domain created in Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC) from FadR (1E2X; (van Aalten *et al.*, 2000)).

B8H254_CAUCN/19-81
 A0A059MTD8_9NOCA/28-90
 YIN1_STRAM/9-71
 Q98DC7_RHILO/19-81
 Q98MD3_RHILO/11-73
 Q9RHW8_COMTE/15-77
 Q982I7_RHILO/42-104
 Q98LF9_RHILO/21-83
 Y3073_RHOS4/13-76
 A0A017KZ04_ECOLX/10-73
 C9MHP0_HAEIF/25-88
 A0A017I5J4_ECOLX/9-72
 GNTR_BACLI/19-81
 A0A017QV01_SALET/6-69
 A0A017HZ81_ECOLX/7-70
 A0A009DS42_ECOLX/8-71
 KORA_STRLI/8-71
 A0A017HZN0_ECOLX/2-65
 A0A017J6T7_ECOLX/13-75
 A0A017ISS1_ECOLX/6-69
 A0A016QXY5_KLEPN/13-76
 J7JJC4_BACIU/18-81
 A0A017I1M0_ECOLX/12-75
 NANR_ECOLC/32-95
 A0A009ELE0_ECOLX/11-74
 A0A017I0L9_ECOLX/8-71

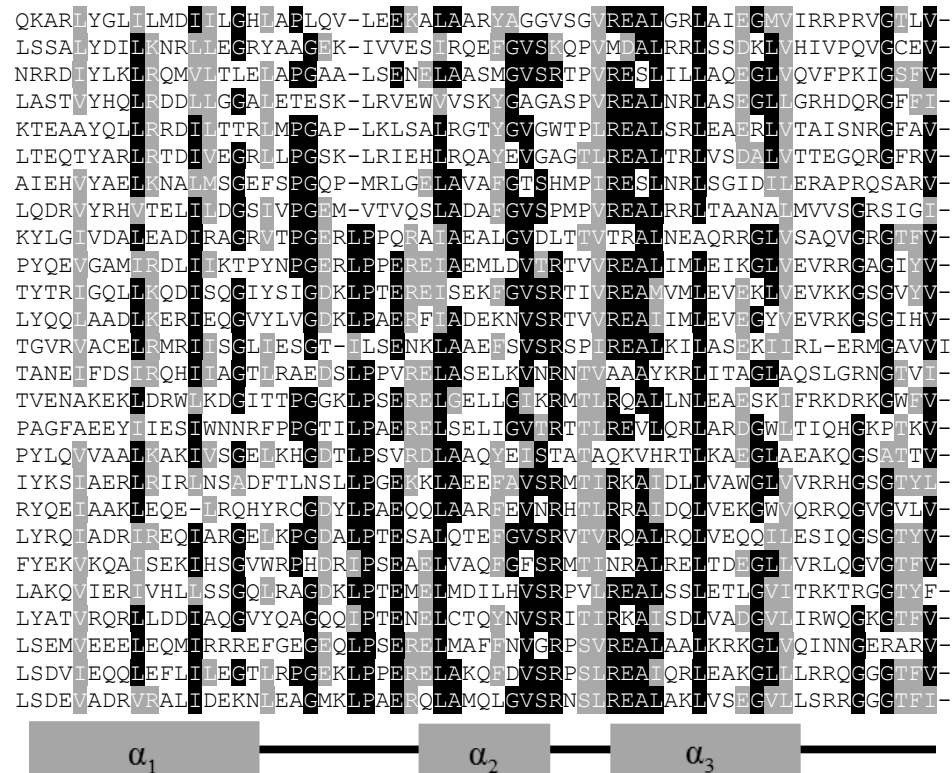


Figure 1.3 Alignment of the HTH domain of GntR-like regulators demonstrating the conservation of sequence structure within the domain. Alignment was performed with ClustalW (Larkin *et al.*, 2007) and residues coloured using Boxshade (http://www.ch.embnet.org/software/BOX_form.html). Secondary structure prediction was performed using PRALINE (Simossis & Heringa, 2005).

The importance of GntR proteins is evidenced by the great many diverse biological processes which they are known to regulate, for example, plasmid maintenance (Reuther *et al.*, 2006), virulence (Casali *et al.*, 2006), antibiotic resistance (Truong-Bolduc & Hooper, 2007), development (Hoskisson *et al.*, 2006), motility (Jaques & McCarter, 2006) and antibiotic production (Hillerich & Westpheling, 2006). However, the function of the majority of GntR-like regulators remains unknown. Identification of the molecules which bind the effector binding domains is, without a doubt, a critical bottleneck in the understanding of the function of these proteins. Structural analysis of C-terminal domains reveals some insight into effector molecules by revealing conserved protein folds such as the chorismate lyase fold (Gallagher *et al.*, 2001, Aravind & Anantharaman, 2003) but GntR regulators containing these conserved folds don't necessarily bind the same metabolic products.

Transcription of the fatty acid degradation genes (*fadL*, *fadD*, *fadE*, *fadH*, *fadA*, *fadB*, *yfcX*, *yfcY*) is negatively regulated by FadR. In contrast, FadR also positively regulates transcription of *fabA* and *fabB*, the fatty acid biosynthesis genes (DiRusso *et al.*, 1993, Raman *et al.*, 1997, DiRusso *et al.*, 1998). Additionally, FadR also regulates *iclR* expression, which functions to negatively regulate the *aceB-aceA-aceK* operon encoding the glyoxylate shunt enzymes (Maloy & Nunn, 1981, Gui *et al.*, 1996), as well as, *yhcX* and *yhcY*, of which the gene products are homologues of FadA and FadB, respectively that are involved with anaerobic growth on fatty acids (Campbell *et al.*, 2003).

The clear importance of FadR regulation in fatty acid metabolism led to structural studies to elucidate the mechanisms of gene control. FadR from *E. coli* was the first GntR-like protein to be crystallised and its 3-D structure revealed that the C-terminal domain comprises seven α -helices which create a binding pocket for the effector molecule, acyl-CoA (**Figure 1.4 A & B**). The structure of FadR revealed that the functional form was a homodimer (van Aalten *et al.*, 2000, van Aalten *et al.*, 2001). Binding of acyl-CoA at the C-terminal causes a conformational change within the protein resulting in a 7.2 Å movement in the DNA binding domain which prevents DNA binding and therefore transcriptional repression is altered and the relevant genes are transcribed (van Aalten *et al.*, 2001). The structure of FadR has revealed a wealth of information about GntR regulators however; there are still gaps in the knowledge of how these conformational changes come about, raising interest in the dynamic environment of the protein.

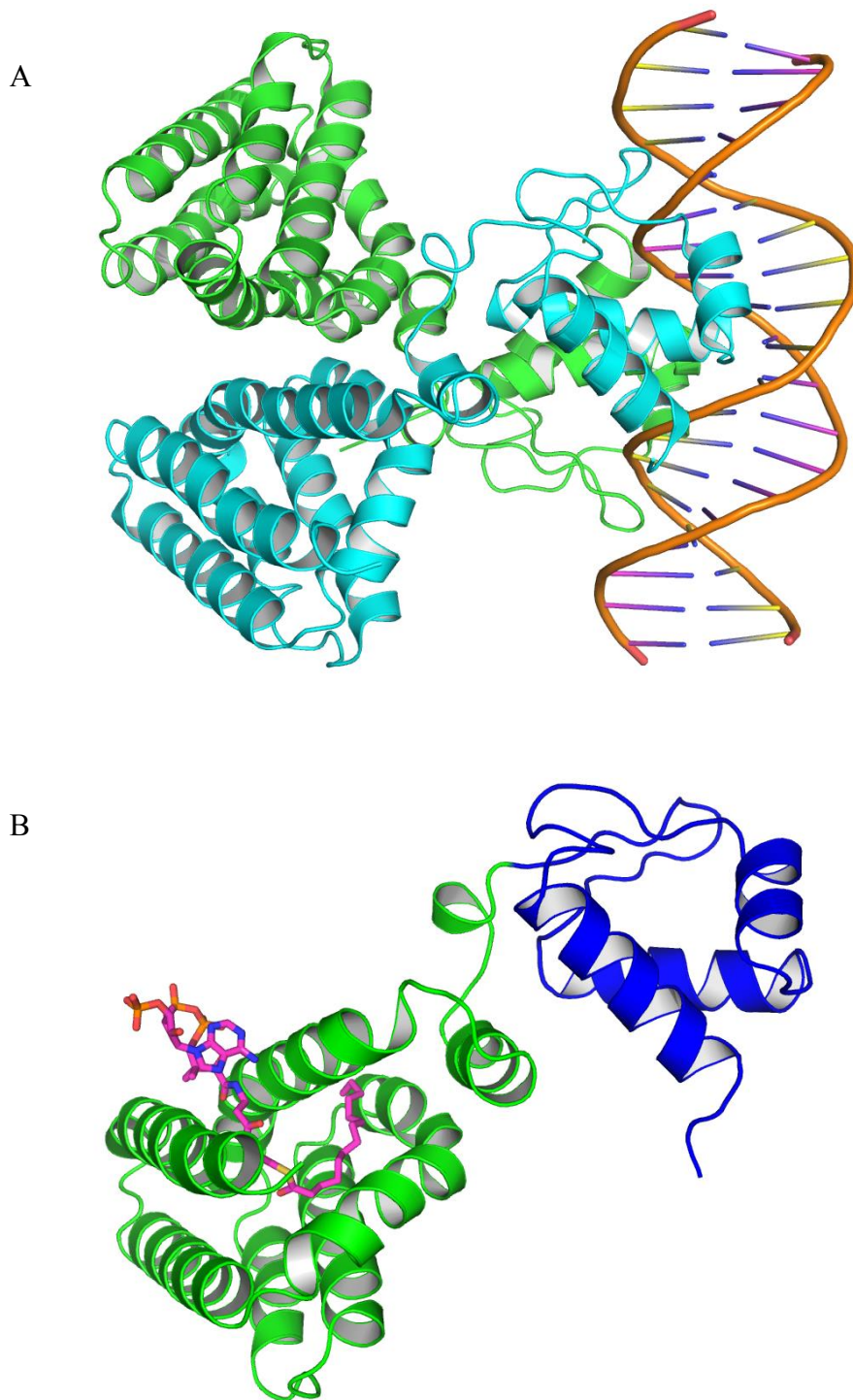


Figure 1.4 (A) FadR dimer bound to *fadB-A* operator sequence (PDB code 1H9T)

(B) FadR monomer in complex with miristoyl-CoA (PDB code 1H9G).

1.4.2 HutC

The HutC subfamily was first identified in *Klebsiella aerogenes* and *Pseudomonas putida*. They are involved in the regulation of histidine utilisation whereby bacteria can use histidine as their only carbon and nitrogen source when other sources are limited, however, HutC control is not only limited to histidine utilisation but also plasmid maintenance and transfer (Kendall & Cohen, 1988). The C-terminal effector molecule in HutC of *P. putida* is known to be urocanate which upon binding releases the protein from the DNA allowing transcription of the histidine utilisation genes (Allison & Phillips, 1990, Hu & Phillips, 1988). The C-terminal domain comprises 6 α -helices and 7 β -strands which has an average length of around 170 amino acids. Analysis of the crystal structure of PhnF, a HutC subfamily member from *E.coli* revealed the arrangement of the C-terminal domain to form a binding pocket showing high homology to the chorismate lyase fold (Aravind & Anantharaman, 2003, Gorelik *et al.*, 2006). There is a putative binding pocket on the surface of the C-terminal domain which also shows high conservation amongst other HutC members indicating that this is the probable effector molecule docking site. HutC orthologs exist in almost all bacterial species demonstrating their importance in regulation of metabolism. Amongst *Pseudomonas* species there are 22 orthologs alone (<http://www.pseudomonas.com>; (Winsor *et al.*, 2011). Most recently, the crystal structure of NagR has been published (Fillenberg *et al.*, 2015). It is the first reported HutC family member of which the structure has been solved in complex with its DNA operator sequence (PDB code 4WWC; **Figure 1.5 A**). NagR is the homologue of HutC from *B. subtilis*, which controls the uptake and metabolism of N-acetylglucosamine (GlcNAc). Furthermore, the structure of NagR in complex with

its effector molecule, N-acetylglucoasamine-6-phosphate (GlcNAc-6-P), reveals distinct conformational changes within the secondary structure of the Eb/O domain and extraordinary repositioning of the HTH domain in relation to the Eb/O domain (PDB code 4U0W; **Figure 1.5 B**).

1.4.3 MocR

The MocR subfamily is interesting as it has an unusually long C-terminal domain of around 350 amino acids long, which is considerably longer than that of the FadR or HutC subgroups. This subfamily accounts for around 16% of sequences deposited in Pfam. The C-terminal is homologous to the class I aminotransferase enzymes. These enzymes catalyse the transamination of amino acids to α -keto acids using pyridoxal 5'-phosphate (PLP) as a co-factor. Some MocR family members are known to have a requirement for PLP (TauR; (Wiethaus *et al.*, 2008), GabR; (Belitsky, 2004); furthermore PdxR in *Streptomyces* species is involved in regulation of PLP synthesis (Magarvey *et al.*, 2001). This leads to the question of whether or not the MocR subfamily has catalytic properties. There is no evidence for this in the current literature but it would seem that the Eb/O domain is a factor in the dimerisation arrangement of these proteins (Rigali *et al.*, 2002).

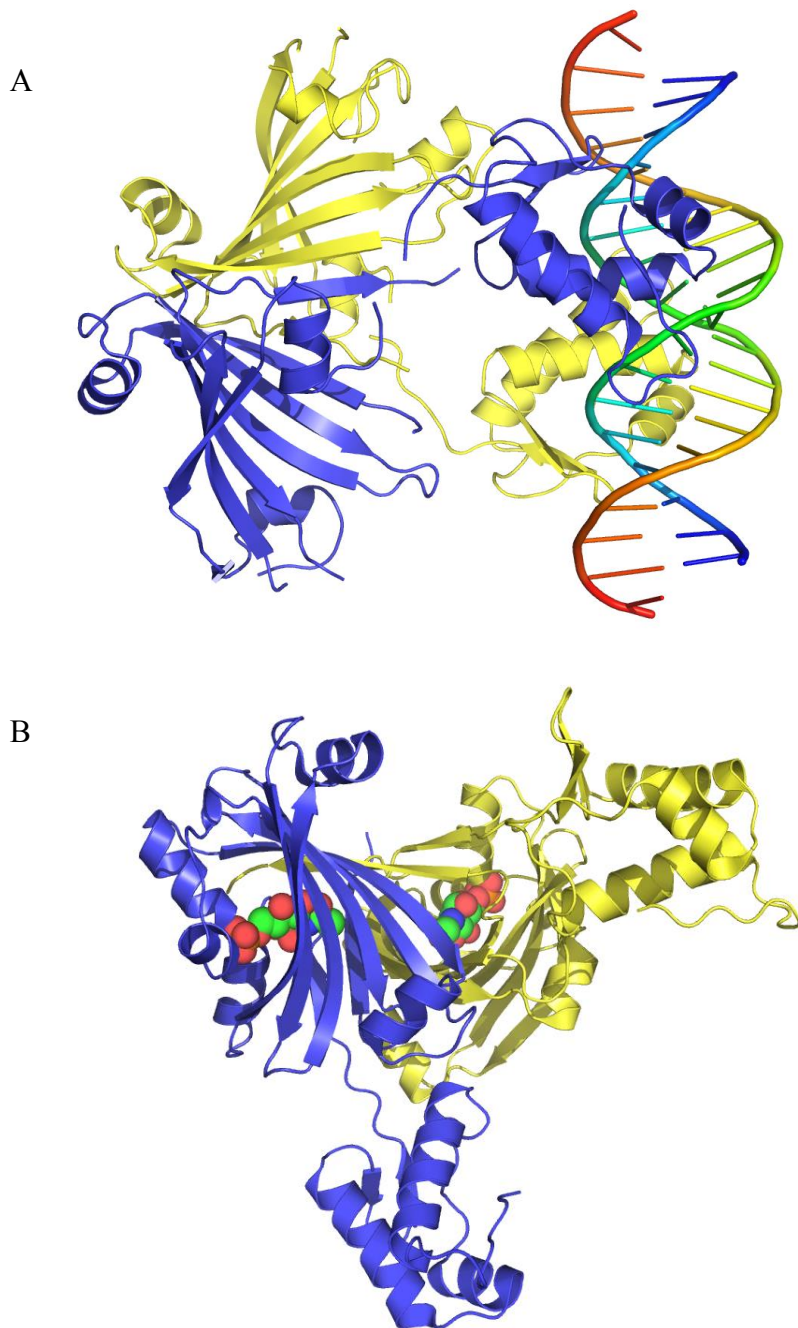


Figure 1.5 (A) Dimeric NagR bound to its 19 bp operator sequence (PDB code 4WWC)

(B) NagR dimer in complex with GlcNAc-6-P (PDB code 4UOW)

1.4.4 YtrA

The YtrA subfamily has an average C-terminal length of 50 amino acids, which forms two α -helices. It is most likely too small to accommodate efficient effector molecule binding but is large enough to still be able to dimerise (Rigali *et al.*, 2002). Members of the YtrA subfamily have been implicated in negative regulation of the *ytrABCDEF* ABC transporter operon in *B. subtilis* (Yoshida *et al.*, 2000). More recently YtrA has been shown to be required for repression and responsiveness to antibiotic stress in *B. subtilis* (Salzberg *et al.*, 2011). Another study shows IndYR, a YtrA family member in *Streptomyces globisporus*, is implicated in sporulation and antibiotic production (Ostash *et al.*, 2011).

1.4.5 AraR and PlmA

The AraR and PlmA subfamilies account for a very small proportion of GntR subfamilies. AraR subfamily members have significant homology in their C-terminal to LacI/GalR family of proteins. In *B. subtilis*, AraR is known to regulate carbon catabolism genes (Mota *et al.*, 2001).

The PlmA subfamily was identified in *Anabaena sp.* and is responsible for regulation of plasmid maintenance (Lee *et al.*, 2003). This subfamily appears to be limited thus far to cyanobacteria. N-terminal alignments reveal similarities with MocR and YtrA suggesting the C-terminal has been replaced at some point during evolution (Hoskisson & Rigali, 2009).

1.4.6 DevA

The most recent subfamily to be identified is DevA, which has a novel C-terminal structure to those already known. This subfamily is limited to Streptomyces thus

far. There are only 4 members of this subfamily – SCO4190 (DevA) and SCO4188 (DevE) from *S. coelicolor* and their orthologues, SAV4023 and SAV4021, from *Streptomyces avermitilis*. No effector molecules have been conclusively identified for any DevA subfamily members.

DevA and DevE have been identified as regulators for the correct development of *S. coelicolor* (Hoskisson *et al.*, 2006, Clark & Hoskisson, 2011). Arising from an ancestral gene in *Salinispora*, DevA is most likely an evolutionary response as the lifecycle of *S. coelicolor* is more complex than that of *Salinispora*. DevA and DevE share 57.6% identity at the amino acid level with both genes conserved across *S. avermitilis* and *S. scabies*. Phylogeny analysis of DevA and its homologues in actinomycetes revealed a gene duplication event gave rise to *devE* (Clark & Hoskisson, 2011). Following the duplication event, DevA and DevE have developed distinct developmental roles despite their homology. Expression of *devA* and *devE* occur at different temporal stages during the lifecycle of *S. coelicolor* further confirming their differing functions. Disruption of *devA* results in reduced ariel hyphae with irregular septum formation and misshapen spores compared to the wild type strain. DevA has been identified as negative auto-regulator. Furthermore, DevA has also been identified as a regulator of *devB* (SCO4191), a putative phosphatase/hydrolase downstream of *devA* (Hoskisson *et al.*, 2006). The two genes, *devA* and *devB* are co-transcribed. This leads to the possibility that a phosphorylated metabolic product, produced by DevB, is the effector molecule for DevA.

DevE is also implicated in hyphal growth in *S. coelicolor*; however these developmental characteristics are morphologically different to those shown by *devA* mutants. Disruption of *devE* results in normal length ariel hyphae; however, these

are mis-septate hyphae. It is likely that DevE is involved in regulation of initiation of septa formation

1.5 Evolution of GntR-like regulators

The high conservation of the HTH domain in several transcription factor families indicates its importance in DNA binding. The GntR HTH domain is thought to be extremely ancient as the distribution of GntR regulators suggests pan-bacterial distribution which can be traced back to the last universal common ancestor (Aravind *et al.*, 2005). The diversity observed in the C-terminal Eb/O domain suggest that fusion of different domains to the HTH domain has occurred frequently throughout evolution resulting in great diversity of function allowing response to novel molecules in response to changing environmental stimuli. This explains the vast diversity of function found within the GntR family. These novel C-terminals give rise to the GntR sub-families discussed in the previous section. Horizontal gene transfer and gene duplication are the most likely scenarios for the increase in regulators belonging to each sub-family; furthermore it can also account for sequence similarities within the sub-families.

The MocR sub-family has high homology with the aminotransferase enzymes while the HutC sub-family share homology with the chorismate lyase enzyme (UbiC) of *E. coli*. Gene fission, horizontal gene transfer and gene decay all give rise to gene fusion (Suhre & Claverie, 2004). It's clear, therefore, that adjacent genes may become fused with HTH domains resulting in the chimeric gene products of the GntR family. Within these sub-families, enzymatic activity doesn't exist despite homology to enzymatic folds. *In silico* structural analysis revealed that the chorismate lyase fold is widespread, suggesting that they evolved for ligand binding

and likely evolved enzymatic activity whilst another version of the domain was fused to the HTH domain resulting in diversity to interact with a wide variety of ligands through which gene expression can be controlled (Aravind & Anantharaman, 2003).

Gene duplication provides an opportunity for organisms to develop new functions. One of the duplicated copies diverges to acquire new function resulting in evolution of a new functional gene product. In general, gene duplication is generally thought to allow duplicates to become specialised in different developmental stages or tissues. Functional variation and differential regulation obtained as a result of gene duplication can confer a fitness advantage in complex environments. This appears to be the case in the DevA subfamily, which is so far limited to Streptomyces which have complex lifecycles resulting from their environmental niche. Furthermore, *devA*, and its paralogue, *devE* are expressed at different temporal stages of the *S. coelicolor* lifecycle (Clark & Hoskisson, 2011), fulfilling the paradigm of gene duplication allowing duplicates to become specialised.

1.6 DNA binding and regulation

The HTH DNA binding domain is the best characterised of all transcription factors. Prokaryotic transcriptional regulators are generally dimeric in arrangement or become dimerised upon DNA binding (Raman *et al.*, 1997). The crystal structures of FadR (PDB code 1H9T), AraR (PDB codes 5DDG; 4H0E) and NagR (PDB code 4WWC) are the only GntR structures bound to DNA operator sequences available in the Protein Data Bank (PDB; (Berman *et al.*, 2000). All of these structures show a dimeric arrangement when bound to DNA. Furthermore, crystal structures of PhnF (PDB code 2FA1), a HutC subfamily member and GabR (PDB code 4N0B), a MocR type regulator, also show dimeric arrangements indicating they are most likely stable

dimers in solution before binding to DNA. No crystal structures are available for any PlmA or DevA type regulators to date. There is an indication that DevA is most likely a dimer by mass spectrometry analysis (Hoskisson & Rigali, 2009).

GntR transcription factors are generally auto-regulatory in nature and as such the promoter region for the gene is usually found within 300 bp upstream of the gene where operator sites are classically found as inverted repeats or directed repeats (Rigali *et al.*, 2002). Directed repeats have an impact on the dimeric arrangement of the regulator due to the lack of symmetry in the operator sequence. Several GntR-like proteins are known to have directed repeat arrangements in their operator sites including TraR (Rossbach *et al.*, 1994), FucR (Hooper *et al.*, 1999), BphS (Watanabe *et al.*, 2000), AphS (Arai *et al.*, 1999), NanR (Condemine *et al.*, 2005) and NagQ (Yang *et al.*, 2006).

Steric constraints are placed upon the HTH-DNA interaction by the Eb/O domain upon effector molecule binding which causes conformational changes within the protein structure (Rigali *et al.*, 2004). By considering this factor, it has been shown that the general operator pattern within the GntR superfamily is palindromic 5' (N)_yGT(N)_xAC(N)_y-3' where (x) and (y) are a variable number of nucleotides and (N) is variable nucleotides. Common ancestry is observed for the FadR, HutC and YtrA subfamilies when operator sites are aligned. The FadR consensus sequence is 5'-*t.GTa.tAC.a-3'* and the HutC consensus sequence is 5'-*GT.ta.AC-3'*. The YtrA subfamily only shows 5'-*GT.AC-3'* over a significantly longer palindromic sequence. It is possible that this arrangement in YtrA operator sequences is in response to the shortened length of the Eb/O domain and thus an unusual dimerisation arrangement.

There are examples of FadR and HutC sub family members which recognise motifs that have little or no symmetry as mentioned previously.

The AraR subfamily has a proposed consensus sequence of 5'-(N)_yTNG(N)_xCNA/T(N)_y-3' (Jain & Nair, 2013) which is based on contacts formed within the DNA bound crystal structure (PDB code 4H0E).

The MocR, PlmA and DevA subfamilies have no defined consensus sequences and don't share sequence homology with the previous subfamilies. The MocR subfamily doesn't have any conserved palindromic sequences common to either the GntR superfamily or within the MocR subfamily (Rigali *et al.*, 2002). Directed repeats do appear to be a common motif pattern within the MocR subfamily, however, with examples including ATACCA (GabR; (Belitsky, 2004), CTGGACYTAA (TauR; (Wiethaus *et al.*, 2008) and AAAGTGGWCTA (PdxR; (Jochmann *et al.*, 2011).

The DevA subfamily currently has no defined consensus sequence. Furthermore, no candidate sequences have been reported to date.

1.7 Studying protein structure/function relationships

Protein structure/function relationships are essential to understand the processes underpinning life. The understanding of metabolic and signalling pathways or gene regulation is reliant on knowledge of protein-metabolite, protein-protein and protein-DNA interactions (Russo Krauss *et al.*, 2013). The primary amino acid sequence of proteins gives very limited information about protein function. The key to elucidating protein function lies in the three dimensional structure of the molecule; seemingly unrelated amino acid sequences may have conserved structural homology (Aravind & Anantharaman, 2003). Structural data can help to assign function to

proteins with previously unknown function. Furthermore, elucidating structure/function relationships of proteins has great potential to allow rational drug design, which is of great clinical significance (Cachau & Podjarny, 2005, Deschamps, 2005, Hoffman, 2012). Several techniques are available that can be used to glean information on structure/function relationships within proteins; some of these are detailed in the following paragraphs.

1.7.1 Bioinformatics

There are now many bioinformatics servers that have the ability to take primary amino acid sequences and predict both secondary and tertiary structure of proteins. Secondary structure of proteins gives the tertiary structure stability with hydrophobic regions found in the centre of the molecule while hydrophilic residues tend to be found on the surface. The secondary structure predictions have given much information on the GntR sub-families by revealing distinct C-terminal secondary structure architecture (Rigali *et al.*, 2002, Hoskisson & Rigali, 2009).

1.7.2 Size Exclusion Chromatography Multiple Angle Laser Light Scattering

Size exclusion chromatography coupled to multiple angle laser light scattering (SEC-MALLS) is a technique that can provide information about the molecular weight of molecules in solution. This is particularly useful in protein biochemistry where the oligomeric state of the protein contributes to its function (Griffin & Gerrard, 2012). The monodispersity of the protein solution can also be assessed by SEC-MALLS which is one of many factors involved in protein crystallisation.

1.7.2 X-ray crystallography

Since the determination of the first 3-D protein structure (sperm whale myoglobin in the 1950s), a vast number of protein structures have been solved by x-ray crystallography (93,594; PDB, Aug 2015). This is contrast to other techniques of structure determination which have significantly fewer structures attributed to them (NMR - 9736 structures; electron microscopy – 613 structures) making x-ray crystallography, by far, the best technique to obtain high resolution protein structural data. High resolution structures have revolutionised our understanding of structure function relationships in proteins (McPherson, 2004) and is becoming much more widely used as tool for rational drug design (Deschamps, 2005, Hoffman, 2012, Franklin *et al.*, 2015).

In x-ray crystallography, protein crystals are illuminated by a beam of monochromatic x-rays. Atoms within the crystals diffract the beam generating a distinct diffraction pattern. Typically a complete diffraction data set is collected by rotating a crystal(s) in the x-ray beam. As the crystal is made up of a periodic array of the protein molecules then the resulting electron density can be represented as a Fourier series. X-rays, as with all electromagnetic radiation, have wave properties, i.e. they have amplitude and a phase. Only amplitudes can be recorded experimentally, phase information is lost which is required for the FT calculation. In crystallography, this is known as “the phase problem”. In other words, in order to solve a structure, the phase information must be determined in order to calculate an electron density map which allows us to build a model of the structure which best fits this experimental data. The phase problem can be resolved by a number of methods depending on the problem in hand. If the structure of interest has a similar fold to

know structures then the method of Molecular Replacement can be used to obtain initial phase information (Scapin, 2013). If a homologous structure doesn't exist obtaining phases by incorporation of heavy atoms and location of these to provide initial phases can be applied through the use of isomorphous replacement or anomalous scattering methods or a combination of both (Taylor, 2010). The advantages of x-ray crystallography are clear in terms of detailed structural information obtained. Crystallisation of proteins allows the structure to be studied at the atomic level in great detail allowing structural information to be gathered, allowing new structure/function relationships in proteins to be uncovered. There is, however, the issue of obtaining crystals of diffraction quality which still involves a significant amount of trial and error and the successful crystallisation of the protein of interest is not guaranteed. That being said careful assessment of the protein target of interest can aid in the success rate. For example it has been shown that many proteins contain regions of disorder which could affect crystallisability. Other factors such as the size of the protein, predicted solubility, isoelectric point can also provide data to assess the crystallisability of the protein of interest. For the latter the program XtalPred (Slabinski *et al.*, 2007), has grouped together a range of parameters and devised a scoring system of crystallisability. For the former the web server RONN (Regional Order Neural Network; (Yang *et al.*, 2005) can identify disordered regions and this information can inform construct design. Other approaches that can aid crystallisability can take a more directed approach e.g. the reductive methylation of lysine residues can increase crystal hit rates (Walter *et al.*, 2006) or site directed mutation of surface amino acids with a high likelihood of disorder known as SERp (Surface Entropy Reduction prediction; (Goldschmidt *et al.*, 2007). However, good

as this technique is at solving structures, it does not give much information in terms of the dynamics of the molecule; therefore other techniques can complement and inform the information obtained from crystallography.

1.7.3 Infrared spectroscopy techniques

1.7.3.1 An overview of infrared spectroscopy

Fourier Transform Infrared Spectroscopy (FT-IR)

FT-IR is a spectroscopy technique that detects chemical bond vibrations (**Figure 1.6**). Absorption of infrared light causes bonds to vibrate within a molecule. Characteristic bonds or functional groups within a molecule only absorb infrared photons of very specific frequencies, which cause them to oscillate. Differences in atom size, bond length and bond strength all contribute to the frequency at which infrared light is absorbed.

Limitations of FT-IR

FT-IR is useful for determination of functional groups and bond vibrations within molecules however it only gives information on average of all vibrational modes within the whole molecule in one 'snapshot' in time. Therefore the information gathered in FT-IR is limited in terms of evolution of time (dynamics) and distance between functional groups, which gives information on structural arrangements and vibrational coupling.

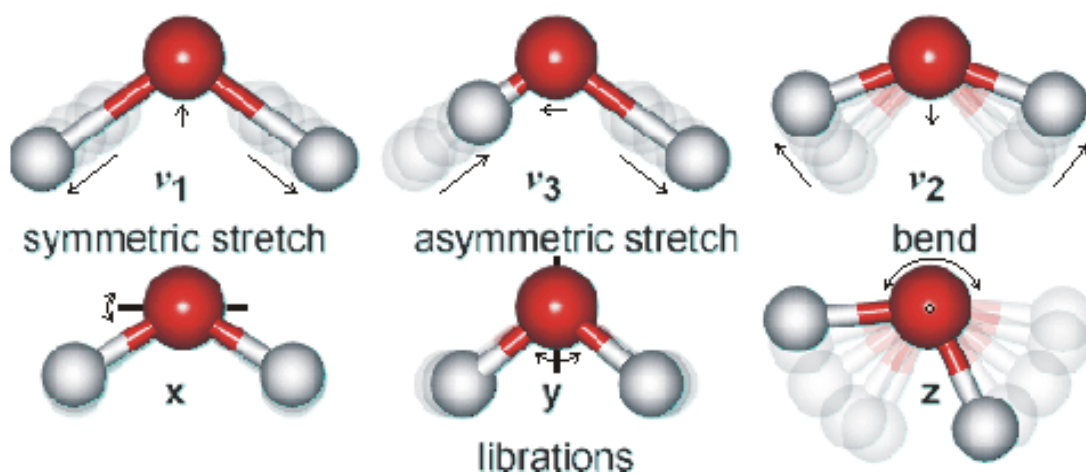


Figure 1.6 Different types of vibrations exist in chemical bonds, all of which contribute to some degree in the infrared spectra, creating distinct patterns for different molecules. Reproduced from www.rsc.org

1.7.3.2 IR Pump probe

Time resolved Infrared Spectroscopy (TRIR)

Time resolved IR spectroscopy is time-resolved derivative of FT-IR where a sample is excited by an intense pump pulse have a frequency tuned to the absorption band of interest followed shortly after by another pulse (probe pulse) which measures the response of the vibration. This technique can give information on the vibrational lifetime of bonds within a molecule.

Limitations of TRIR

Like FTIR, there are limitations on the information that can be obtained on a molecule. TRIR addresses the problem of obtaining dynamical information however; both FTIR and TRIR are one dimensional which limit these techniques to studying an average picture of the molecular system as a whole.

1.7.4 Introduction to 2D-IR

Although TRIR is useful to determine a more dynamic picture of a molecular system, two-dimensional infrared spectroscopy (2D-IR) is a relatively new measurement technique which can characterise protein dynamics (Hamm *et al.*, 1998, Zanni & Hochstrasser, 2001, Zheng *et al.*, 2007, Hunt, 2009). 2D-IR takes the principles of TRIR and moves them into two dimensions which allows vibrational coupling to be observed between different parts of the molecule.

In terms of protein 2D-IR, the amide I band is the area that holds most interest. The amide I band lies in the frequency range $1600 - 1700 \text{ cm}^{-1}$. Vibrations arising primarily from the C=O stretch and somewhat from the C-N stretch of the amide

bond in the protein backbone give rise to this region. The C-C and C-H stretching also contributes to the IR spectra as well as hydrogen bonding both within the protein but also with the surrounding solvent and other small molecules e.g. drugs, DNA, co-factors etc.

Protein secondary structure elements can be detected in the amide I region with β -sheet structures having the lowest absorption frequencies of 1615-1630 cm^{-1} , followed by an overlap of α -helices and random coils over a 1630-1650 cm^{-1} range. Inter-strand β -sheets absorb around 1660-1670 cm^{-1} and finally a second, weaker β -sheet absorption can usually be (though not always) seen at 1680 cm^{-1} depending if the β -sheet is in a parallel or anti-parallel conformation

Photon-echo 2D-IR spectroscopy has been used extensively to probe molecular interactions within the University of Strathclyde in collaboration with the Central Laser Facility (CLF) based at the Rutherford Appleton Laboratory, Didcot. The technology works by two pump pulses exciting a mode of interest (e.g. C=O) followed by a probe pulse which measures the response. A schematic diagram of the system can be found in **Figure 1.7**. Examples of molecular systems which I have been involved with at the University of Strathclyde include the interaction of NO with haem proteins; myoglobin, catalase and cytochrome c (Bellota-Antón *et al.*, 2011, Adamczyk, Candelaresi, Robb, *et al.*, 2012, Adamczyk, Candelaresi, Kania, *et al.*, 2012, Candelaresi *et al.*, 2013). More recently, the enoyl acyl carrier protein reductase, InhA, from *Mycobacterium tuberculosis* and its interaction with and the anti-tuberculosis pro-drug isoniazid (INH) have been probed with 2D-IR (Shaw *et al.*, 2015). The InhA co-factor, NADH, has been also been analysed extensively by

2D-IR and the study of the InhA/INH interaction builds on this foundation (Simpson *et al.*, 2013).

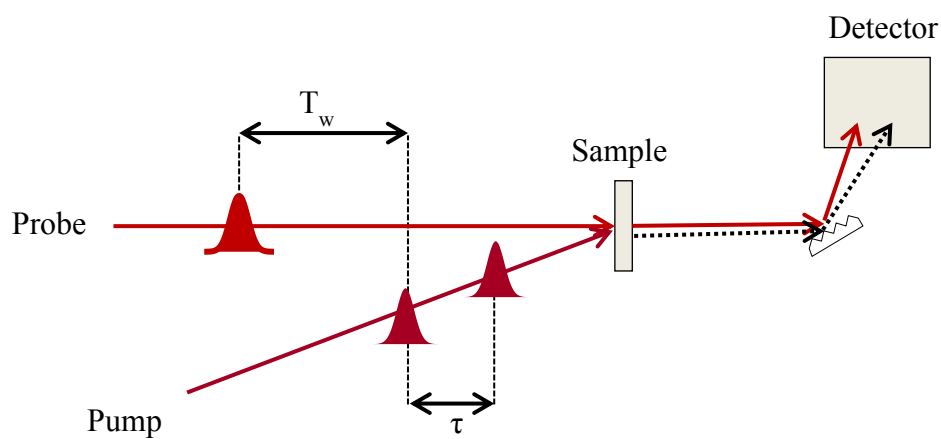


Figure 1.7 Schematic diagram of pump probe 2D-IR spectroscopy. τ denotes the time between pump pulses used to excite the mode of interest. The time delay between these two pulses is used to obtain dynamic information about the system. T_w denotes the waiting time between pump and probe pulses; usually in the sub-picosecond range.

1.8 Aims and objectives

This project aims to address some of the unknowns in terms of structural and dynamical characterisation of GntR proteins. The function of many GntR proteins remains unknown and as previously mentioned, there are very few effector molecules known. Elucidating these molecules may give clues to protein function as known effector molecules, so far, appear to be molecules which are present in the metabolic pathways that these proteins regulate or sense key metabolic checkpoints to coordinate gene regulation (Hoskisson & Rigali, 2009).

Aims of the project

The project was divided in to four main aims.

1. Identification of new novel GntR-like targets

Bioinformatic analysis was used to identify proteins that are novel in terms of secondary structure analysis and phylogeny, representing poorly studied groups of GntR-like proteins. There are a few subfamilies with relatively few members when compared to the vast number of GntR-like sequences in the database. This leads to the possibility of new structural folds that are currently unknown in relation GntR proteins. These are currently limited to a few folds such as the UbiC (HutC), FadR and amino transferase (MocR).

2. Crystallisation and structural characterisation of identified GntR-like targets to assign new structure/function relationships

Since many of the GntR proteins in the Pfam database are of unknown function, solving structure can help assign functions through understanding of ligand binding. Structural information could also give clues to what the effector molecules may be. Despite lack of sequence homology, structural homology can exist and seemingly vast differences in sequence may be irrelevant as structurally similar folds can give clues as to the function of the protein. For example, the globin fold is present in mammalian myoglobin and also HmpA, a nitric oxide reductase, from *E. coli*, which share very little sequence homology (Ilari & Boffi, 2008).

3. Dynamic characterisation of identified GntR-like targets

Two-dimensional infrared spectroscopy (2D-IR) is a relatively new technique that has been around for about 15 years but is not widely used in protein studies. The technique relies on exciting a molecule to cause a vibration. The evolution of these vibrations can be monitored and a picture of molecular dynamics can be built. Proteins are known to be dynamic environments so 2D-IR spectroscopy is well suited for probing their interactions with each other and with ligands.

4. Identification of potential promoter regions of targets

Very few of the DNA consensus sequences that GntRs bind to have been identified. Examination of the upstream regions of genes can be useful in identifying inverted repeats or directed repeats that confer operator binding sites (Rigali *et al*, 2002). Upstream regions of candidate sequences will be analysed to determine new candidate consensus sequences.

Chapter 2: Materials and Methods

2.1 Strains and their cultivation

2.1.1 Bacterial strains, culture and storage

The bacterial strains used during this study are shown in **Table 2.1**. All microbiology work was carried out aseptically. Strains were streaked on solid Luria-Bertani (LB) media and single colonies were picked for growth in liquid LB culture overnight. Antibiotics were added to media where appropriate. Glycerol stocks of strains were prepared for storage by taking a 0.5 ml aliquot of overnight culture and adding 0.5 ml of 50% (v/v) glycerol (final concentration 25% glycerol). Stocks were stored at -80°C.

2.1.2 Plasmids

The plasmids used during this study are shown in **Table 2.2**. All plasmids used during this thesis were confirmed by sequencing (Eurofins Genomics, Ebersberg, Germany).

2.1.3 Media

Media was prepared by dissolving required chemicals/reagents in distilled water and autoclaving for 121°C, 15 psi for 15 minutes. Media used are listed in **Table 2.3**. For auto-induction and M9 media, stock solutions of the relevant components were prepared by dissolving in distilled water and sterilised by 0.22 µm filter or by autoclaving and mixed after sterilisation.

Table 2.1 Bacterial strains used during this study

Strain	Description	Genotype	Source or reference
<i>Escherichia coli</i>			
DH5 α	General plasmid propagation	F ⁻ Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK ⁻ , mK ⁺) phoA supE44 λ ⁻ thi-1 gyrA96 relA1	(Hanahan, 1983)
BL21 (DE3)	General over expression strain	F ⁻ ompT gal dcm lon hsdS _B (r _B ⁻ m _B ⁻) λ (DE3) pLysS(cm ^R)	(Studier <i>et al.</i> , 1990)
Rosetta 2 (DE3)	Over expression strain for proteins not native to <i>E. coli</i>	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3) pRARE2 (Cam ^R)	Novagen
OmniMAX TM	General plasmid propagation, T1 phage resistant	F' {proAB+ lacIq lacZ Δ M15 Tn10(TetR) Δ (ccdAB)} mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80(lacZ) Δ M15 Δ (lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD	Invitrogen
B834 (DE3)	Methionine auxotroph	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm met (DE3)	Invitrogen
<i>Pseudomonas aeruginosa</i>			
UCBPP PA14	Genomic DNA		(Lee <i>et al.</i> , 2006)
PA14 HutC:: <i>Tn7</i>	HutC transposon insertion mutant		(Liberati <i>et al.</i> , 2006)
PA01	Genomic DNA		(Stover <i>et al.</i> , 2000)
<i>Pseudomonas fluorescens</i>			
SBW25	Genomic DNA		(Silby <i>et al.</i> , 2009)
<i>Streptomyces coelicolor</i>			
M145	Genomic DNA M145; Prototrophic	SCP1 ⁻ SCP2 ⁻	(Bentley <i>et al.</i> , 2002)

Table 2.2 Plasmids used during this study

Plasmid	Description	Antibiotic resistance	Source or reference
pET100	T7 expression vector, N terminal hexa-his tag	Ampicillin	Invitrogen
pUC19	M13 cloning vector	Ampicillin	(Yanisch-Perron <i>et al.</i> , 1985)
pOPINF	T7 expression vector, N terminal hexa-his tag	Ampicillin	(Berrow <i>et al.</i> , 2007)
pOPINE	T7 expression vector, C terminal hexa-his tag	Ampicillin	(Berrow <i>et al.</i> , 2007)
pOPINS3C		Ampicillin	(Berrow <i>et al.</i> , 2007)
pKR003	pET100 containing DevA gene insert	Ampicillin	This work
pKR006	pET100 containing HutC gene	Ampicillin	This work
pKR007	pET100 containing Gp26 gene	Ampicillin	This work
pKR008	pET100 containing DevE gene	Ampicillin	This work
pKR010	pEX vector containing synthetic DevA gene	Ampicillin	This work
pKR011	pEX vector containing synthetic HutC gene	Ampicillin	This work
pKR012	pEX vector containing synthetic DevE gene	Ampicillin	This work
pKR013	pOPINF containing synthetic DevA gene	Ampicillin	This work
pKR014	pOPINF containing synthetic DevE gene	Ampicillin	This work
pKR015	pOPINF containing synthetic HutC gene	Ampicillin	This work
pKR016	pOPINE containing optimised DevA gene	Ampicillin	This work
pKR024	pOPINF containing optimised DevE gene	Ampicillin	This work
pKR034	pOPINE containing optimised HutC gene	Ampicillin	This work
pKR049	pOPINF containing optimised Gp26 gene	Ampicillin	This work
pKR063	pUC19 containing DevA upstream region	Ampicillin	This work
pKR064	pUC19 containing DevE upstream region	Ampicillin	This work
pKR065	pUC19 containing HutC upstream region	Ampicillin	This work
pKR066	pUC19 containing Gp26 upstream region	Ampicillin	This work

Table 2.3 Media used for cultivation of microorganisms

Luria Bertani* (Sambrook <i>et al.</i> , 1989)	1% (w/v) tryptone 0.5% (w/v) yeast extract 1% (w/v) NaCl
ZYP-5052 medium for auto induction (AIM; (Studier, 2005)	1% tryptone 0.5% yeast extract 50 mM Na ₂ HPO ₄ ·7H ₂ O 50 mM KH ₂ PO ₄ 25 mM (NH ₄) ₂ SO ₄ 0.5% glycerol 0.05% glucose 0.2% α-lactose 2 mM MgSO ₄
5X M9 Salts	64g Na ₂ HPO ₄ ·7H ₂ O 15g KH ₂ PO ₄ 2.5g NaCl Make up to 1 litre with distilled water. Sterilise by autoclaving
M9*(Sambrook <i>et al.</i> , 1989)	1X M9 salts 2 mM MgSO ₄ 0.1 mM CaCl ₂ 0.4 % glucose
Nutrient Broth*	13g Nutrient broth mix (Oxoid) Make up to 1 litre with distilled water and autoclave
2X YT* (Sambrook <i>et al.</i> , 1989)	1.6% (w/v) tryptone 1% (w/v) yeast extract 0.5 (w/v) NaCl
SOC (Sambrook <i>et al.</i> , 1989)	20 g Tryptone 5 g yeast extract 0.5 g NaCl 10 ml KCl (250 mM) Make up to 1 litre with distilled water pH to 7.0 with NaOH After autoclaving add 20 ml sterile glucose solution (1 M) Immediately before use add 5 ml sterile MgCl ₂ (2 M)
MD media (Molecular Dimensions)	Dissolve 21.6 g of MD medium base in 1 litre of distilled water and autoclave, Dissolve 5.1g MD Nutrient Mix in 50 ml of distilled water and sterile filter and add to base medium.
* 2% agar was added when solid media was required	

2.2 Molecular biology

2.2.1 Genomic DNA Isolation

Genomic DNA was isolated from *Pseudomonas* species by following a standard protocol designed for *E. coli* (Sambrook *et al.*, 1989). Briefly, overnight cultures (5 ml) were inoculated using a single colony from a streak plate. Cells were pelleted by centrifugation at 14,000 x *g* for 5 minutes. The supernatant was removed and the pellet re-suspended in 500 µl TE buffer (10 mM Tris-HCl, 0.1 mM Na₂EDTA, pH 8.0). Cells were lysed by the addition of 50 µl 10% SDS and 14.3 µl proteinase K (35 mg ml⁻¹), mixed by inversion and incubated at 55°C for 30 minutes.

An equal volume of phenol/chloroform (1:1) was added to the mixture and mixed by inversion until a homogeneous solution was observed, followed by centrifugation at 14,000 x *g* for 10 minutes. The aqueous phase was transferred to a new tube and phenol/chloroform was repeated. The upper (aqueous) phase was transferred to a clean tube and a 10% volume of 3M sodium acetate (pH 5.2) added and mixed. DNA precipitation was performed by addition of 330 µl 100% isopropanol. DNA was collected by using a closed, glass Pasteur pipette and washed by dipping in 70% ethanol. DNA was air-dried for 20 minutes before re-suspending in 250 µl TE buffer containing 50 µg ml⁻¹ RNase and dissolving at 37°C for 15 minutes.

2.2.2 Primers and PCR

The primer sets used for TOPO cloning and upstream regions of genes for binding studies are shown in **Table 2.4**. Primers were designed for the upstream regions of genes using Genefisher software (Giegerich *et al.*, 1996). The primers used for

optimised cloning carried out at the Oxford Protein Production Facility (OPPF UK) are shown in **Table 2.5**.

Table 2.4 Primers used for TOPO cloning and blunt end cloning of genetic upstream regions

Primer	Forward sequence	Reverse sequence	Product size (bp)
Full length gene			
DevA	<u>CACCTTGGT</u> CGTGACTCAGGAGAA	CTAGGAGAGTGTCATGTCCG	876
Gp26	<u>CACCATGCCGGCTCCGGCCCAAAT</u>	TCAGTTGGGGGAGAGCGTGA	525
HutC PA14	<u>CACCGTGACGTCCTCTTCTTCCGA</u>	TCATGAGCTGAAGCGTCCTT	753
DevE	<u>CACCTTGGT</u> CGTGGAGCCGGAACA	TCAGCCCGCGAGCACCAGCT	912
Upstream regions			
DevA_US	CTGCTCGAAGGCGATGACGA	TCTGCTGCCGTTACGGACA	233
DevE_US	CCCGTACTTCCACTGCACGA	CGGCTCCACGACCAAACCCTA	240
PA14_HutC_US	CGGGACGAATCTCGGCGAGA	CCAGCGGGGAACGATCGGAA	237
Gp26_US	TTGTAACCCACAGCTTTGCAGA	TGACCGACCTCTATTTCCCTGA	223
M13Cy5	GTAAAACGACGGCCAGTGAATTC	CAGGAAACAGCTATGACCATGATTA	Variable

Table 2.5 High Throughput primers for optimised plasmids

Well	Gene name	Forward Primer	Tm	Reverse Primer
A1	DEVA_native	aggagatataccatgGTCGTGACTCAGGAGAACGTGTCCG	40	gtgatggtgatgtttGGAGAGTGCATGTCCGTGGTGATGG
B1	DEVA_native	aagttctgtttcaggccccgGTCGTGACTCAGGAGAACGTGTCCG	45	atggtctagaaagctttaGGAGAGTGCATGTCCGTGGTGATGG
C1	DEVA_native	aggagatataccatgTCGTCCCAGGAGATCGCCGACG	37	gtgatggtgatgtttGGAGAGTGCATGTCCGTGGTGATGG
D1	DEVA_native	aagttctgtttcaggccccgTCGTCCCAGGAGATCGCCGACG	42	atggtctagaaagctttaGGAGAGTGCATGTCCGTGGTGATGG
E1	DEVA_native	aagttctgtttcaggccccgTCGTCCCAGGAGATCGCCGACG	42	atggtctagaaagctttaGGAGAGTGCATGTCCGTGGTGATGG
F1	DEVE_native	aggagatataccatgGTCGTGGAGCCGGAACACGC	35	gtgatggtgatgtttGCCCGCGAGCACCAGCTCCG
G1	DEVE_native	aagttctgtttcaggccccgGTCGTGGAGCCGGAACACGC	40	atggtctagaaagctttaGCCCGCGAGCACCAGCTCCG
H1	DEVE_native	aggagatataccatgGCAACACATCGAGAGGTGGCCG	37	gtgatggtgatgtttGCCCGCGAGCACCAGCTC
A2	DEVE_native	aagttctgtttcaggccccgGCAACACATCGAGAGGTGGCCG	42	atggtctagaaagctttaGCCCGCGAGCACCAGCTC
B2	DEVE_native	aggagatataccatgCCCACGCAGGCCAGTTGGC	35	gtgatggtgatgtttGCCCGCGAGCACCAGCTC
C2	DEVE_native	aagttctgtttcaggccccgACGCAGGCCAGTTGGCCGAG	41	atggtctagaaagctttaGCCCGCGAGCACCAGCTCCG
D2	STRCO Putative	aggagatataccatgACCGCGCCCGTCGTCCAC	33	gtgatggtgatgtttGGGGCGGGGTGCGCAGCCC
E2	STRCO Putative	aagttctgtttcaggccccgACCGCGCCCGTCGTCCAC	38	atggtctagaaagctttaGGGGCGGGGTGCGCAGCCC
F2	STRCO Putative	aggagatataccatgTCGCTGCGCGAACAGATCCG	35	gtgatggtgatgtttGGGGCGGGGTGCGCAGCCC
G2	STRCO Putative	aagttctgtttcaggccccgTCGCTGCGCGAACAGATCCG	40	atggtctagaaagctttaGGGGCGGGGTGCGCAGCCC
H2	HUT_native	aggagatataccatgACGTCCTCTTCTTCCGATCGTTCCC	40	gtgatggtgatgtttTGAGCTGAAGCGTCTTCCAGACGG
A3	HUT_native	aggagatataccatgGCGCCGCTCTACGCGCGGGTC	36	gtgatggtgatgtttTGAGCTGAAGCGTCTTCCAGACG

B3	HUT_native	aagttctgttcagggcccGCGCCGCTCTACGCGGGGTC	41	atggtctagaagctttaTGAGCTGAAGCGTCCTTCCAGACG
C3	HUT_native	aggagatataccatgGCGCCGCTCTACGCGGGTCAAG	39	gtgatggtgatgtttGCTGAAGCGTCCTTCCAGACGGTG
D3	HUT_native	aagttctgttcagggcccGCGCCGCTCTACGCGGGTCAAG	44	atggtctagaagctttaGCTGAAGCGTCCTTCCAGACGGTG
E3	PA14_34660 GntR gene	aggagatataccatgAGCATCACCAAGAACGACAAGAACACG	42	gtgatggtgatgtttGGTGCTCTCCCGCGCCATCAGC
F3	PA14_34660 GntR gene	aagttctgttcagggcccAGCATCACCAAGAACGACAAGAACAC G	47	atggtctagaagctttaGGTGCTCTCCCGCGCCATCAGC
G3	PA14_34660 GntR gene	aggagatataccatgCCCACCCTCAACGAAGTCGCG	36	gtgatggtgatgtttGGTGCTCTCCCGCGCCATCAG
H3	PA14_34660 GntR gene	aagttctgttcagggcccACCCTCAACGAAGTCGCGCGC	39	atggtctagaagctttaGGTGCTCTCCCGCGCCATCAG
A4	PA14_34660 GntR gene	aggagatataccatgGCCGGGGTCAGCCCGATCAC	36	gtgatggtgatgtttGGTGCTCTCCCGCGCCATCAG
B4	PA14_34660 GntR gene	aagttctgttcagggcccGCCGGGGTCAGCCCGATCAC	35	atggtctagaagctttaGGTGCTCTCCCGCGCCATCAG
C4	P. fluorescens	aggagatataccatgACTATAAAAGCAATTGGCCGACGCGATC	43	gtgatggtgatgtttGAACCCATAGAGCTTTGCAGGGTTATC AAC
D4	P. fluorescens	aagttctgttcagggcccACTATAAAAGCAATTGGCCGACGCGA TC	48	atggtctagaagctttaGAACCCATAGAGCTTTGCAGGGTTA TCAAC
E4	P. fluorescens	aggagatataccatgGATCACTTCTCGGTTGAAATTTTTTCGT CAC	46	gtgatggtgatgtttGAACCCATAGAGCTTTGCAGGGTTATC AAC
F4	P. fluorescens	aagttctgttcagggcccGATCACTTCTCGGTTGAAATTTTTTCGT CACC	51	atggtctagaagctttaGAACCCATAGAGCTTTGCAGGGTTA TCAAC
G4	Gp26	aggagatataccatgCCGGCTCCGGCCCAAATTTTTACTC	40	gtgatggtgatgtttGTTGGGGGAGAGCGTGACGAAAATTC
H4	Gp26	aagttctgttcagggcccGCTCCGGCCCAAATTTTTACTCAGCG	46	atggtctagaagctttaGTTGGGGGAGAGCGTGACGAAAATT CC
A5	Gp26	aggagatataccatgCCGAAGACGCAAGCGGCGTAC	36	gtgatggtgatgtttGTTGGGGGAGAGCGTGACGAAAATTC
B5	Gp26	aagttctgttcagggcccAAGACGCAAGCGGCGTACGTG	41	atggtctagaagctttaGTTGGGGGAGAGCGTGACGAAAATT C

C5	DEVA_SERp	aggagatataccatgTCGTCCCAGGAGATCGCCGACG	37	gtgatggtgatgtttGGAGAGTGTCATGTCCGTGGTGATGG
D5	DEVA_SERp	aagttctgtttcaggccccgTCGTCCCAGGAGATCGCCGACG	42	atggtctagaaagctttaGGAGAGTGTCATGTCCGTGGTGATGG
E5	DEVA_SERp	aagttctgtttcaggccccgTCGTCCCAGGAGATCGCCGACG	42	atggtctagaaagctttaGGAGAGTGTCATGTCCGTGGTGATGG
F5	DEVE_SERp	aggagatataccatgGCAACACATCGAGAGGTGGCCG	37	gtgatggtgatgtttGCCCGCGAGCACCAGCTCC
G5	DEVE_SERp	aagttctgtttcaggccccgGCAACACATCGAGAGGTGGCCG	42	atggtctagaaagctttaGCCCGCGAGCACCAGCTCC
H5	DEVE_SERp	aggagatataccatgACGCAGGCCAGTTGGCCG	34	gtgatggtgatgtttGCCCGCGAGCACCAGCTC
A6	DEVE_SERp	aagttctgtttcaggccccgACGCAGGCCAGTTGGCCG	39	atggtctagaaagctttaGCCCGCGAGCACCAGCTC
B6	HUT_SERp	aggagatataccatgGCGCCGCTCTACGCGCGGGTC	36	gtgatggtgatgtttTGAGCTGAAGCGTCCTTCCAGACG
C6	HUT_SERp	aagttctgtttcaggccccgGCGCCGCTCTACGCGCGGGTC	41	atggtctagaaagctttaTGAGCTGAAGCGTCCTTCCAGACG
D6	HUT_SERp	aggagatataccatgCTCTACGCGCGGGTCAAGCAG	36	gtgatggtgatgtttTGAGCTGAAGCGTCCTTCCAGACG
E6	HUT_SERp	aagttctgtttcaggccccgCTCTACGCGCGGGTCAAGCAG	41	atggtctagaaagctttaTGAGCTGAAGCGTCCTTCCAGACG
F6	HUT_SERp	aggagatataccatgGCGCCGCTCTACGCGCGGGTC	36	gtgatggtgatgtttGCTGAAGCGTCCTTCCAGACGGTG
G6	HUT_SERp	aagttctgtttcaggccccgGCGCCGCTCTACGCGCGGGTC	41	atggtctagaaagctttaGCTGAAGCGTCCTTCCAGACGGTG
	pOPIN sequencing	gaccgaaattaatacgaactactataggg	60	-

2.2.3 Agarose Gel Electrophoresis

Confirmation of PCR products and plasmids was carried out by gel electrophoresis. Gels contained 1% (w/v) agarose (unless stated otherwise) dissolved by heating in a microwave in 1X TAE buffer (diluted from a 50 X Stock; see **Table 2.6**). Ethidium bromide was added to a final concentration of 1 $\mu\text{g ml}^{-1}$ when the gel was cool enough to hold comfortably. Gels were run at 80 volts for between 30 minutes and 1.5 hours depending on the size of DNA and visualised by UV transillumination. Reagents used are shown in **Table 2.6** below.

Table 2.6 Reagents used for agarose gel electrophoresis

6X DNA loading dye	0.25% Bromophenol blue 0.25% Xylene cyanol FF 30% glycerol
50X TAE	2 M Tris-acetate 50 mM EDTA

2.2.4 Cloning

Initial cloning for protein over expression was carried out using a Champion™ pET Directional TOPO® Expression Kit from Invitrogen. This was useful for testing over-expression conditions. PCR was carried out using the primers listed in **Table 2.4**.

The TOPO ligation was carried out according to the manufacturer's instructions. Where larger quantities of protein were required for crystallisation and 2D-IR studies, we made use of the high throughput (HTP) facility at the Oxford Protein Production Facility (OPPF), Research Complex at Harwell, UK. A range of truncations, mutations and tags were tried for each protein in order to find the most suitable construct for crystallisation (See **Chapter 3** sections for details). A list of truncations can be found in **Table 3.5**. HTP cloning was carried out using a standard operating procedure from the OP PF. The SOP uses the In-Fusion® system from Clontech using methods described previously (Berrow *et al.*, 2007). Briefly, gene specific primers are designed with 15 bp extensions homologous to linearised vector ends. The gene of interest is amplified by PCR. Following PCR clean up, PCR product and linearised vector are mixed with the In-Fusion enzyme which digests double stranded homologous ends to single strands allowing homologous recombination to occur and a new construct is created (**Figure 2.1**)

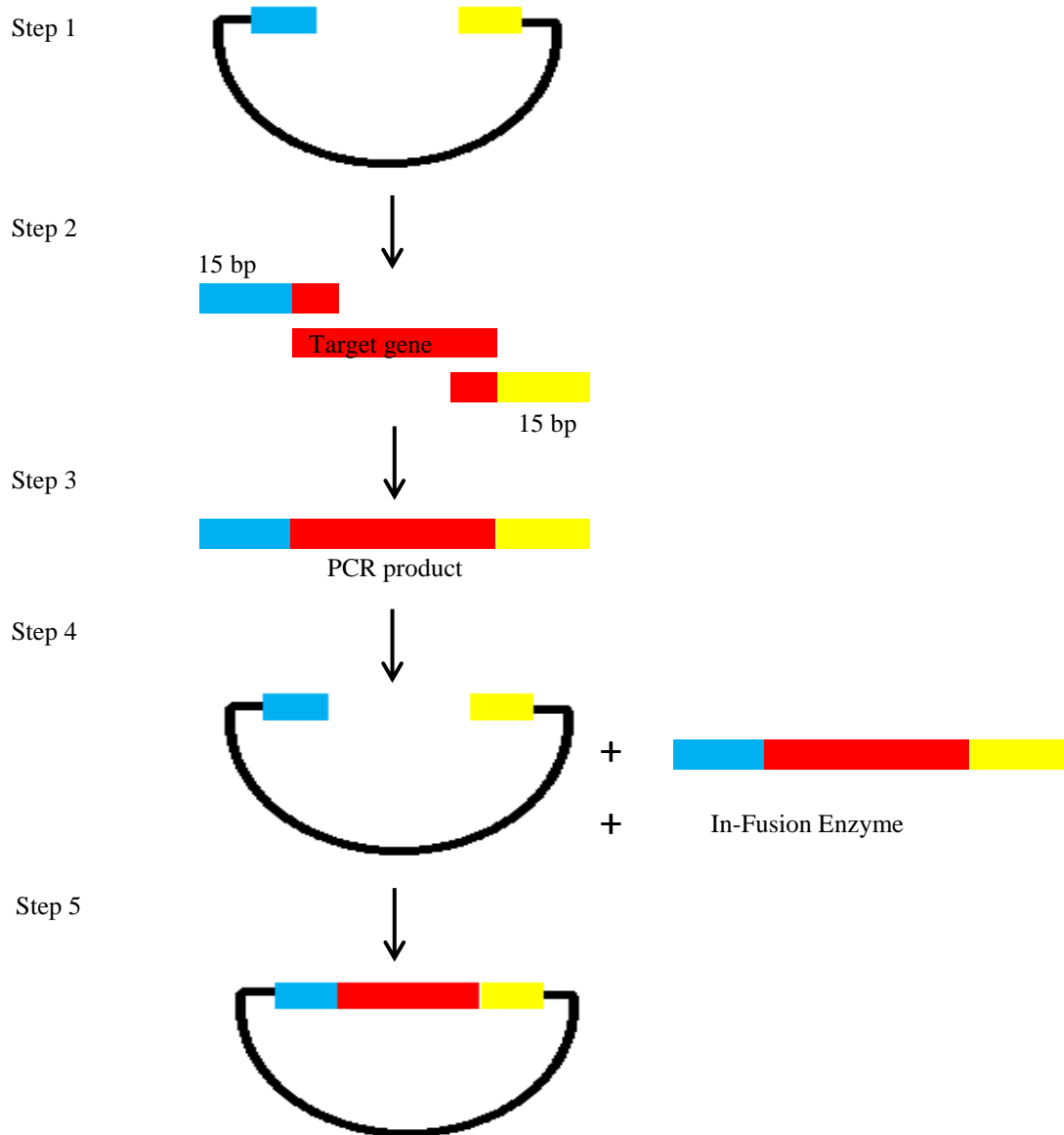


Figure 2.1 In-Fusion cloning technology allows directional cloning of target DNA quickly and efficiently. The double stranded primer extensions are digested by the In-Fusion enzyme to single stranded DNA allowing homologous recombination to take place. **Step 1:** Vector is linearised **Step 2:** Primers are designed for target gene with 15 bp extensions homologous to vector ends **Step 3:** PCR amplification of target gene to produce PCR product **Step 4:** Linearised vector, PCR product and In-Fusion enzyme are incubated at 50°C for 15 minutes **Step 5:** New construct

2.2.5 Chemically competent *E. coli*

Cultures (100 ml) were inoculated using 250 µl of fresh overnight culture containing antibiotics where appropriate. Cells were grown at 37°C with shaking until the optical density at 600 nm (OD₆₀₀) reached 0.4 approximately. Cells were pelleted by centrifuging at 4000 x g for 10 minutes at 4°C. The supernatant was discarded. Cells were re-suspended slowly in ice cold 0.1 M MgCl₂ (25 ml). The cells were again pelleted by centrifuging at 4000 x g for 10 minutes. The pellet was then re-suspended in ice cold 0.1 M CaCl₂ (25 ml) and stored on ice for 20 minutes. Cells were then pelleted at 4000 x g for 10 minutes, and re-suspended in 100 µl 0.1 M CaCl₂ containing 20% (v/v) glycerol. Competent cells were stored in 100 µl aliquots at -80°C until required.

2.2.6 Transformation of *E. coli*

Approximately 1 ng of plasmid DNA was pipetted in to an Eppendorf tube containing 50 µl chemically competent *E. coli*. This was incubated on ice for 30 minutes followed by a heat shock at 42°C for 60 seconds. The tube was placed back on ice for 2 minutes followed by the addition of 450 µl sterile LB media. Cells were then incubated at 37°C, with shaking, for 1 hour to allow the cells to recover. Aliquots of 100 µl and 200 µl were plated on solid LB media containing the appropriate antibiotic and incubated overnight at 37°C

2.2.7 Complementation of PA14 HutC::Tn7 with pKR034 (Cadoret *et al.*, 2014)

2.2.7.1 Preparation of electrocompetent PA14 HutC::Tn7

Aliquots (1 mL) of fresh overnight culture of PA14 HutC::Tn7 were used to inoculate 50 mL LB containing 15 $\mu\text{g ml}^{-1}$ gentamicin. Cultures were grown at 37°C until OD600 reached 0.4. Cells were harvested by centrifugation (2300 $\times g$, 10 minutes) Cells were resuspended in 10 mL ice cold sucrose solution (300 mM) followed by centrifugation. Supernatant was discarded and cells resuspended in 5 mL ice cold sucrose solution. Cells were again centrifuged, supernatant discarded and finally resuspended in 100 μL ice cold sucrose solution. Competent cells were used immediately after preparation.

2.2.7.2 Electroporation of competent PA14 HutC::Tn7 with PKR034

An aliquot (80 μL) of electrocompetent cells was mixed with $\sim 1 \mu\text{g}$ pKR034 and stored on ice for 30 minutes. The mixture was transferred to an electrocuvette (0.1 cm) and pulse applied by a Genepulser II electroporator (BioRad) using setting Ec1 (1.8 kV, 5 ms). Cells were immediately recovered in 2 mL SOC medium and incubated at 37°C for 4 hours.

Recovered cells (200 μL) were plated on LB containing 15 $\mu\text{g ml}^{-1}$ gentamicin and 100 $\mu\text{g ml}^{-1}$ carbenicillin. Plates were incubated overnight at 37°C.

2.2.7.3 Phenotypic analysis of complemented PA14 HutC::Tn7 with PKR034

A single colony was picked from the transformation plate following electroporation and streaked aseptically on to minimal M9 media which was supplemented with 15 mM L-histidine, 15 mM urocanic acid or 18.7 mM NH_4Cl and 22.2 mM glucose.

M9 without supplements were used as the negative control. 2xYT was used as the positive control. Plates were incubated overnight at 37°C.

2.2.7.4 Growth curves

Bacterial growth curves were carried out using liquid M9 media supplemented as above. Media was aliquoted (250 µL) into a 96-well culture plate and inoculated with 1% of a fresh overnight culture grown at 37°C. Each growth curve was carried out in triplicate. Growth was measured automatically every 15 minutes in a Bio-Tek multi detection microplate reader (Synergy HT) over a period of 24 hours.

2.2.8 Alkaline lysis Plasmid isolation (Sambrook *et al.*, 1989)

Aliquots of LB (5 ml) containing the appropriate antibiotic were inoculated with *E. coli* containing the relevant plasmid and grown overnight at 37°C with shaking (250 rpm). Cells were harvested in an Eppendorf tube by centrifugation at 12500 x g for 1 minute and then re-suspended in 100 µl ice cold Solution 1 (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0). Solution 2 (200 µl; 0.2N NaOH, 1% SDS) was added and mixed by inversion followed by 150 µl Solution 3 (3 M Potassium acetate, 2 M glacial acetic acid). The contents of the tube were thoroughly mixed by vortexing and then stored on ice for 10 minutes. Tubes were then centrifuged for 10 minutes at 12500 x g. The supernatant was transferred in to a fresh Eppendorf tube and two volumes of ethanol added. The contents were mixed by inversion then incubated at room temperature for 2 minutes. The tube was again centrifuged at 12500 x g for 10 minutes and the resulting pellet was washed in 1 ml 70% v/v ethanol. The pellet was allowed to air dry at room temperature for 15 minutes. DNA pellets were re-suspended in 50 µl TE buffer and stored at -20°C. Small scale

plasmid purifications utilised commercially available kits from Promega and Bioline were also used according to the manufacturer's instructions.

2.3 Protein techniques

2.3.1 Protein Over-expression

Constructs were initially tested in two different expression strains, *E. coli* BL21 or *E. coli* Rosetta, using two different media (Auto induction media (Studier, 2005) and LB) and at two different temperatures to ascertain the best conditions for over-expression. Media (50 ml) were inoculated with 1% (v/v) of fresh overnight culture and cells were grown to an OD₆₀₀ of between 0.4 and 0.6 then the temperature was reduced to 20°C or 25°C. At this point, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to the LB culture to induce protein expression at a final concentration of 1 mM. Cultures were incubated at reduced temperatures for a minimum of 12 hours. Cells were harvested by centrifugation (6000 x g, 4°C, 15 minutes) and resulting pellets were stored at -80°C until purification. Scaled up over-expression of proteins was carried out in 1 L cultures after establishing optimum conditions.

2.3.2 Production of Selenomethionine DevE

Transformation of *E. coli* B834 cells was carried out as per **section 2.2.6**. A single colony was picked from the transformation plate and used to inoculate 5 mL LB medium. This culture was grown overnight at 37°C. The overnight culture was then used to inoculate (1% v/v) 50 mL of fresh LB medium. Cultures were grown at 37°C until OD₆₀₀ = 1. Cells were harvested by centrifugation at 6000 x g for 10 minutes then washed four times in MD medium before cell were finally used to inoculate 1 L

MD medium containing 40 mg ml⁻¹ selenomethionine (Sigma Aldrich). Cells were cultured to OD₆₀₀ = 0.6 and induced by addition of IPTG (final concentration 1 mM) and culturing overnight at 25°C. Cells were harvested as above.

2.3.3 Protein Purification

Stored cell pellets were resuspended in Buffer A (**Table 2.7**) with addition of Benzonase (100 U/mL; Sigma Aldrich) and a Complete protease inhibitor tablet (one tablet/10 ml buffer; Roche) according to manufacturer's instructions. Cells were lysed using pressure (10 kpsi) and the lysate was clarified by centrifugation (7000 x g, 4°C, 40 mins).

Proteins were purified by nickel affinity chromatography and size exclusion chromatography (SEC). Clarified lysate was applied to pre-equilibrated columns on an AKTA Purifier system using Unicorn 5.11 software (GE Healthcare). Following application to the columns over-expressed proteins were washed with 5 column volumes of Buffer A, followed by elution using an isocratic gradient of Buffer B. Fractions were collected automatically in volumes of 1 ml. Buffers used for purifications are detailed in **Table 2.7**. HisTrap FF crude column (5 ml; GE Healthcare) was used for initial nickel affinity purification and a HiLoad 16/60 Superdex 75 prep grade (GE Healthcare) for further size exclusion chromatography.

Table 2.7 Buffers used during protein purification

HisTrap Buffer A	50 mM Tris pH 7.5 0.5 M NaCl 30 mM imidazole
HisTrap Buffer B	50 mM Tris pH 7.5 0.5 M NaCl 0.5 M imidazole
GF Buffer 1	20 mM MES pH 6.0 0.5M NaCl 1 % glycerol

2.3.4 SDS-PAGE

Protein samples were mixed with 2X sample dye at a 1:1 ratio then heated at 75°C for 5 minutes before being loaded on to either a precast 4 – 12 % Amersham ECL polyacrylamide gel (GE Healthcare) or a precast 12% Novex gel (Invitrogen).

Gels were run at 160V (ECL gel) or 200V (Novex gel) for 1 hour or 40 minutes respectively. Gels were stained using Coomassie blue G250. The buffers used for SDS-PAGE are detailed in **Table 2.8**.

Table 2.8 Reagents used for SDS-PAGE

2X SDS sample buffer	63 mM Tris pH 6.8 2% w/v SDS 10% v/v glycerol 5% v/v β -mercaptoethanol 0.001% w/v bromophenol blue
10X SDS running buffer	150g Tris 720g Glycine 50g SDS Make up to 5L with distilled water
MES running buffer	50 mM MES 50 mM Tris 1 mM EDTA 0.1% (w/v) SDS
Coomassie Blue stain	208 ml distilled water 208 ml methanol 84 ml glacial acetic acid 0.1% w/v Coomassie blue G250
Destain	50% distilled water 40% methanol 10% acetic acid

2.3.5 Bradford Assay (Bradford, 1976)

Standards were prepared by diluting DevE in GF buffer 1 to 0.1 – 1.0 mg ml⁻¹ (linear range for the assay). Bradford reagent (1.5 mL; Sigma) was added to protein samples (50 μ L), vortexed briefly and incubated at room temperature. Sample absorbance was measured at 595 nm and compared to standard to determine concentration.

2.3.6 Electrophoretic mobility shift assay (EMSA) [Adapted from (Hutchings & Drabble, 2000) and (Craig *et al.*, 2012)]

Reagents used for EMSAs are listed in **Table 2.9**. Upstream promoter regions were cloned into pUC19 then amplified using an M13 primer labelled with Cy5 (**Table 2.4**). A fixed amount of Cy5 labelled promoter DNA (1.5 ng) was used for the assay. Protein concentrations varied from 0 – 1000 nM initially before further refinement.

Reactions were carried out in 10 μ l volumes containing 7 μ l binding buffer, 1 μ l BSA (10 mg/ml), 1 μ l DNA (15 ng/ μ l), 1 μ l protein of appropriate concentration. This mixture was incubated at 30°C for 15 minutes then 1 μ l bromophenol blue dye was added and loaded on to 6% PAGE gels. Gels were run at 120 V for 40 minutes in 1X TAE then visualised on a Typhoon 9200 scanner (excitation 633 nm, emission 670 nm; Amersham).

Table 2.9 Reagents used for EMSA

EMSA binding buffer	10 mM Tris HCl pH 7.5 5 mM MgCl ₂ 0.1 mM EDTA 60 mM KCl
6% PAGE gel	1 ml 10X TAE 2 ml 30% Bis-acrylamide 7 ml distilled water 100 μ l 25% APS 20 μ l TEMED
Bromophenol blue loading dye	0.25% Bromophenol blue 30% glycerol

2.3.7 Size exclusion chromatography multiple angle laser light scattering (SEC-MALLS)

Protein samples (100 μ L) were diluted in the appropriate buffer (**Table 2.10**). SEC-MALLS data was collected on Dawn Helios II multi angle light scattering detector and Optilab T-rEX differential refractive index detector (both Wyatt Technologies) connected to a Superdex 10/30 column. Flow rate was set at 0.7 ml min⁻¹. Data was analysed in Astra v6.0 software (Wyatt Technologies); refractive index increment was set to 0.185.

Table 2.10 SEC-MALLS protein samples

Protein	Buffer	Final protein concentration
DevA	GF buffer 1	2 mg ml ⁻¹
DevE	GF buffer 1	1 mg ml ⁻¹
HutC	GF buffer 1	2 mg ml ⁻¹
Gp26	GF buffer 2	2 mg ml ⁻¹

2.4 Crystallisation of GntR-like proteins

2.4.1 Crystallisation trials

Crystallisation conditions for each of the target proteins (HutC, DevA, DevE, and Gp26) were screened using a range of commercially available crystallisation screens detailed in **Table 2.11**. This was facilitated by the high-throughput crystallisation facility at the OPPF. Crystallisation trials were set up in 200 nl drops in a 96-well format using the sitting drop vapour diffusion method (McPherson, 2004); **Figure 2.2**) All proteins were tested over a range of protein concentrations, with the exception of DevE, detailed in **Table 2.12**. Trials were set-up in 96-well Greiner plate using a Cartesian MicroSys (Digilab, Marlborough, MA) capable of setting crystallisation drops of 200 nl (100 nl protein/100 nl reagents; (Walter *et al.*, 2005). Crystal trays were stored in a Rock Imager (Formulatrix, Bedford, MA) imaging system at 20 °C which imaged drops at regular intervals from a very early stage in the crystallisation process. Crystal images were accessed via the remote access camera at <http://www.oppf.rc-harwell.ac.uk/xtalpims/>.

Conditions with potential ‘crystal hits’ were followed up and optimised manually. Typically optimisation was carried out in 24 well vapour diffusion sitting drop trays. Optimisation strategies for each target will be discussed in the relevant section in **Chapter 4**.

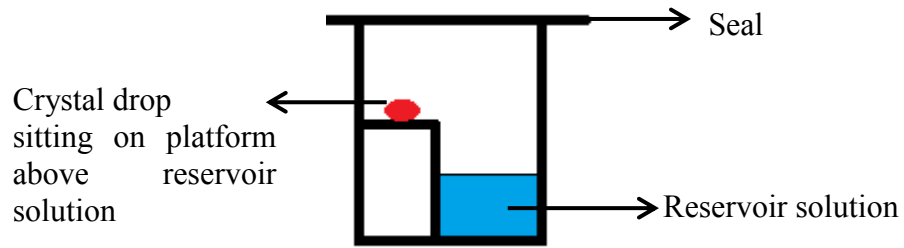


Figure 2.2 Schematic of the sitting drop vapour diffusion method of crystallisation. Typically the protein solution is mixed in equal volumes with a reservoir solution containing precipitants, buffer and other chemicals and dispensed either as a sitting or hanging drop and sealed in a container with the reservoir solution. The crystal drop has a lower precipitant concentration and gradually equilibrates over time with the reservoir solution. If the conditions are favourable, crystals form.

Table 2.11 Commercially available crystallisation screens used in the OPPF

Screen	Manufacturer	Features
Morpheus™	Molecular Dimensions	3D protein crystallisation screen. All conditions are cryo-protected.
Index HT	Hampton Research	Sparse matrix and grid screen and incomplete factorial.
JCSG- <i>plus</i> ™	Molecular Dimensions	Sparse matrix screen with reduced redundancy.
PACT <i>premier</i> ™	Molecular Dimensions	Tests the effect of pH, anions and cations, using PEG as the precipitant.
Wizard III & IV	Emerald BioSystems	Sparse matrix
PEG/ Ion HT™	Hampton Research	Polymer, salt and pH matrix screen.

Table 2.12 Concentration of proteins used in crystallisation trials

Protein	Concentration trialled
HutC	22mg ml ⁻¹ , 40 mg ml ⁻¹
HutC + urocanic acid	2 mg ml ⁻¹ (74 µM) + 20 mg ml ⁻¹ (145 µM)
DevA	16 mg ml ⁻¹ , 18 mg ml ⁻¹
DevE	22.6 mg ml ⁻¹
DevE + 18mer	16 mg ml ⁻¹ (0.52 mM) + 1.2 mM DNA
Gp26	24 mg ml ⁻¹ , 28 mg ml ⁻¹

2.4.2 X-ray diffraction and data collection

Diffraction data were collected on beamlines I03, I04 and I04-1 at Diamond Light Source (DLS). Data were integrated and scaled using tools available in the CCP4 software suite (Winn *et al.*, 2011) and the *PHENIX* software suite (Adams *et al.*, 2010). In particular *iMOSFLM* (Battye *et al.*, 2011) and *xia2* (Winter *et al.*, 2013) was used for data reduction and integration; *PHASER* (McCoy *et al.*, 2007, McCoy, 2007) for molecular replacement; *Buccaneer* (Cowtan, 2006) and *Autobuild* (Terwilliger *et al.*, 2008) for automated model building.

2.5 Two Dimensional Infra-Red Spectroscopy (2D-IR)

HutC was measured using 2D-IR spectroscopy. Protein was concentrated to 27 mg/ml (~1 mM) and buffer exchanged into deuterated phosphate buffer (pH7.5). The sample was then aliquoted (20 μ L) in to a Harrick cell with CaF₂ windows and a 50 μ m PTFE spacer. The sample was measured using FTIR prior to performing 2D-IR to determine the frequency range to target during 2D-IR.

2D-IR measurements were carried out on the ULTRA LASER system at the Rutherford Appleton Laboratory, Oxfordshire.

2.6 Bioinformatics

For bioinformatic methodologies and analyses, please refer to **Chapter 3**.

Chapter 3: Selection, phylogenetics and *in silico* characterisation of GntR-like protein targets

3.1.1 Bioinformatics and target selection

There are currently 231,015 sequences (August 2015) in the GntR protein family according to the Pfam database (Bateman *et al.*, 2002, Finn *et al.*, 2014), conversely only 76 crystal structures of GntR proteins are available in the RCSB Protein Data Bank (www.rcsb.org; (Berman *et al.*, 2000). Clearly with more than 200,000 GntR sequences, there is great scope for the discovery of new protein folds leading to identification of new GntR subfamilies in addition to the 7 subfamilies established previously (Rigali *et al.*, 2002, Rigali *et al.*, 2004, Hoskisson & Rigali, 2009).

Given the diversity of the sequences it would not be practical to attempt to characterise them all during the course of this project therefore *in silico* analysis became an important analytical tool during this project.

It has been shown previously that there is a high level of similarity in the N-terminus of GntR family proteins (Rigali *et al.*, 2004, Hoskisson & Rigali, 2009) with the initial characterisation of the family being based on this homology (Haydon & Guest, 1991). Given the separation of GntR proteins into subfamilies has been validated based on the C-terminal sequence variation our target selection was based on this characteristic. Targets were selected based on their novelty in terms of secondary structural predictions and C-terminal sequence.

The N-terminal HTH region is very highly conserved and when subjecting this region to BLAST, no novel GntR targets were identified. Cloning targets were

selected via BLAST searching against the PDB. GntR regulators were not selected for further study if they fell in to the two main subfamilies, FadR and HutC particularly or if crystal structures were already present in the PDB. The full bioinformatics pipeline is detailed in **Figure 3.1**.

HutC from *Pseudomonas aeruginosa* PA14 was used as a control protein as the effector molecule is known to be urocanic acid and the mechanism of gene control is well established (Allison & Phillips, 1990, Zhang & Rainey, 2007, Bender, 2012). At the time of initial target selection there was no crystal structure for this particular protein, however a structure has since been published of NagR, the homologue from *Bacillus subtilis* (Fillenberg *et al.*, 2015).

A total of 30 GntR targets were selected and ranked in order of interest (sequence, secondary structural and biological novelty) for cloning. These are detailed **Table 3.1**. The last 11 targets in **Table 3.1** have crystal structures, available in PDB, which were solved during the course of this study by other groups. The PDB accession number is shown in the table with the reference although many of these remain unpublished despite being deposited in the PDB.

GntR-like sequences that were chosen from *S. coelicolor* were generally implicated in carbon metabolism. The majority of these were putative proteins which have arisen from data that is now available due sequenced genome annotations. These hold interest when considering the role of *Streptomyces* in production of antibiotics and the rise in antibiotic resistance. Understanding the mechanisms of some of these proteins may provide greater knowledge in terms of rational drug design as a means

to find new novel antimicrobials through activation of cryptic biosynthetic pathways for example (Liu *et al.*, 2013).

Targets were also so selected from *Pseudomonas aeruginosa* due to its clinical significance in the cystic fibrosis (CF) patient (Banerjee & Stableforth, 2000). Biofilm formation in the lung of the CF patient makes *P. aeruginosa* very difficult to treat. Antibiotics have limited success and eventually infection becomes permanent (Lipuma, 2010). More effective antibiotics could help this situation hence the selection of these targets as potential anti-microbial targets. Again, these are putative proteins that have been identified from genome sequencing.

The reference sequences which define the 7 main subfamilies are listed in **Table 3.2**. These sequences were included in both secondary structure and phylogenetic analysis in order that new members may be assigned to them. FadR was used to root Neighbour-Joining and Maximum Parsimony trees and to check the accuracy of the secondary structure prediction given the high-resolution crystal structure available and the functional experimental analysis it has been subjected to (van Aalten *et al.*, 2000, van Aalten *et al.*, 2001).

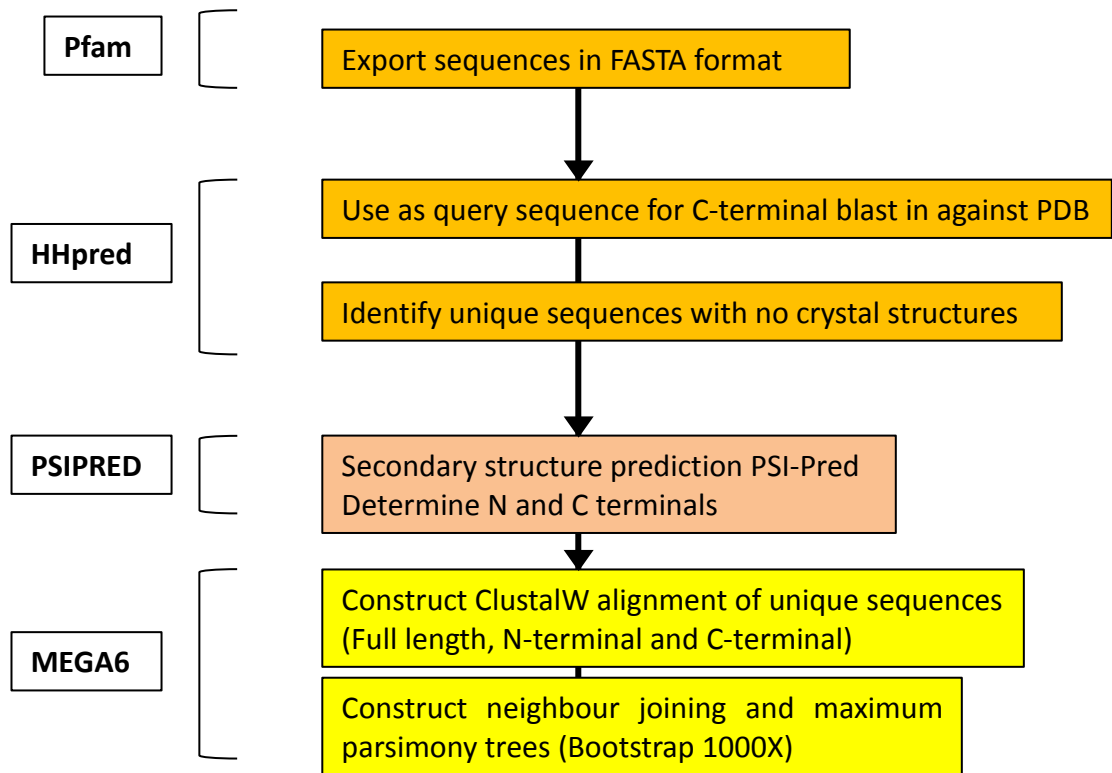


Figure 3.1 Bioinformatics pipeline for selection of cloning targets and phylogenetic tree assembly

Table 3.1 Cloning targets selected from analysis of Pfam and PDB

Rank	Organism	Gene	Protein	UniProt/PDB code	N-terminal	C-terminal	Comments
1	<i>P. aeruginosa</i> PA14	<i>PA14_67420</i>	HutC	Q02ER1	1-95	90-250	Control, mutant also available
2	<i>S. coelicolor</i> A3(2)	<i>SCO4190</i>	DevA	Q9FCH9	1-87	82-291	Implicated in spore development, effector molecule and C-terminal unknown
3	<i>S. coelicolor</i> A3(2)	<i>SCO4188</i>	DevE	Q9FCI0	1-95	90-303	C-terminal and effector molecule unknown
4	<i>Streptomyces</i> Phage <i>phi-C31</i>	<i>gp26</i>	Gp26	Q9T219	1-89	87-174	Only GntR regulator known in a virus
5	<i>P. aeruginosa</i> PA14	<i>PA14_71680</i>	Putative transcriptional regulator, GntR	Q02DT0	1-110	100-491	
6	<i>S. coelicolor</i> A3(2)	<i>SCO2182</i>	Putative GntR	Q9S2Q4	1-74	73-207	Upstream of SCO2183 (pyruvate dehydrogenase)
7	<i>S. coelicolor</i> A3(2)	<i>SCO7702</i>	Putative transcriptional	Q93JH0	1-73	72-236	Homology to lactate dehydrogenase

			regulator, GntR				regulator in <i>E. coli</i>
8	<i>S. coelicolor</i> A3(2)	<i>SCO7168</i>	Putative GntR transcriptional regulator	Q9FBS4	1-88	87-224	
9	<i>S. coelicolor</i> A3(2)	<i>SCO2442</i>	Putative GntR	Q9L0A4	1-82	80-235	Homology to glucuronate utilisation regulator (UxuR) in <i>E. coli</i>
10	<i>S. coelicolor</i> A3(2)	<i>SCO7056</i>	Putative GntR	Q9FC28	1-95	90-253	
11	<i>Bacteroides thetaiotaomicron</i>	<i>BT_1272/fucR</i>	FucR	Q9RQ14	1-85	80-326	Fucose utilisation repressor
12	<i>C. diphtheriae</i>	<i>DIP2241</i>	Putative GntR	Q6NEN1	1-120	115-278	
13	<i>P. aeruginosa</i> PA14	<i>PA14_70710</i>	Transcriptional regulator GlcC	Q02E06	1-77	75-251	Glycolate utilisation regulator
14	<i>P. aeruginosa</i> PA14	<i>PA14_34880</i>	Putative transcriptional regulator, GntR	Q02M96	1-85	80-249	Predicted to regulate an operon containing an oxoreductase and ferredoxin (Mao <i>et al.</i> , 2009)
15	<i>Corynebacterium</i>	<i>DIP2081</i>	Putative transcriptional	Q6NF19	1-77	75-232	

	<i>diphtheria</i>		regulator				
16	<i>Mycobacterium tuberculosis</i> H37Rv	<i>Rv0494</i>	HTH-type transcriptional regulator Rv0494/MT0514	P67739	1-87	85-242	
17	<i>Olsenella uli</i> (<i>Lactobacillus uli</i>)	<i>Olsu_0403</i>	Transcriptional regulator, GntR	E1QYR3	1-87	85-243	
18	<i>Clostridium bolteae</i>	<i>CLOBOL_00895</i>	Putative uncharacterised protein	A8RJE9	1-80	75-490	
19	<i>S. coelicolor</i>	<i>SCO1177</i>	Putative GntR	Q9RJZ3	1-85	80-246	Homology to galactone operon repressor (GdoR) in <i>E. coli</i>
20	<i>Coxiella burnettii</i>	<i>CBU_0775</i>	Transcriptional regulator, GntR	Q83DG1 (PDB 3tqn)	1-77	74-113	(Franklin <i>et al.</i> , 2015)
21	<i>Ralstonia eutropha</i> JMP134	<i>Reut_B4779</i>	Phenylacetic acid degradation related protein	Q46RV7 (PDB 2pimA)	1-57	50-140	(Genomics, To be published)
22	<i>Mycobacterium smegmatis</i>	<i>phnF</i>	HTH-type transcriptional repressor, GntR	A0QQ72 (PDB 3f8M)	1-80	75-244	(Gebhard <i>et al.</i> , 2014)

23	<i>Corynebacterium glutamicum</i>	<i>cg3261</i>	Regulatory protein, GntR	Q8NLJ5 (PDB 2ek5)	1-73	69-121	(Gao <i>et al.</i> , 2007)
24	<i>Enterococcus faecalis</i>	<i>EF_1328</i>	Transcriptional regulator, GntR	Q835P8 (PDB 3ddv)	1-80	75-235	(Zhang, Zhou, <i>et al.</i> , To be published)
25	<i>E. coli O6</i>	<i>c4276</i>	Putative regulator	Q8FCM7 (PDB 3hfi)	1-90	85-251	(Zhang, Xu, <i>et al.</i> , To be published)
26	<i>Streptomyces phaeochromogeness</i>	<i>traR</i>	TraR	Q54677 (PDB 1v4r)	1-75	72-245	(Tanaka <i>et al.</i> , To be published)
27	<i>Ralstonia eutropha</i> JMP134	<i>Reut_B4629</i>	Transcriptional regulator, GntR	Q46SA5 (PDB 3ihuA)	1-83	80-221	(Genomics., To be published)
28	<i>P. syringae</i>		Regulatory protein, GntR	Q19AK4 (PDB 3c7j)	1-90	88-245	(Nocek <i>et al.</i> , To be published)
29	<i>Rhodococcus jostii</i> RHA1	<i>RHA1_ro03477</i>	Probable transcriptional regulator, GntR	Q0SB06 (PDB 2hs5)	1-95	90-235	(Tan <i>et al.</i> , To be published)
30	<i>Oenococcus oeni</i>	<i>OEOE_1803</i>	Transcriptional regulator, GntR	Q04D30 (PDB 3by6A)	1-75	74-123	(Zhang, Volkart, <i>et al.</i> , To be published)

Table 3.2 Reference sequences used during bioinformatic analyses

Organism	Gene	Protein	UniProt/PDB code	N-terminal	C-terminal	Comments
<i>B. subtilis</i>	<i>araR</i>	AraR	P96711	1-100	95-362	(Franco <i>et al.</i> , 2006)
<i>Nostoc sp.</i>	<i>all1076</i>	PlmA	Q8YXY0	1-90	85-328	(Lee <i>et al.</i> , 2003)
<i>E. coli</i> K12	<i>fadR</i>	FadR	P0A8V6	1-73	57-239	(Rigali <i>et al.</i> , 2002)
<i>P. putida</i>	<i>hutC</i>	HutC	P22773	1-86	67-248	(Rigali <i>et al.</i> , 2002)
<i>Rhizobium meliloti</i>	<i>mocR</i>	MocR	P49309	1-79	47-493	(Rigali <i>et al.</i> , 2002)
<i>B. subtilis</i>	<i>ytrA</i>	YtrA	O34712	1-79	71-130	(Rigali <i>et al.</i> , 2002)
<i>S.coelicolor</i> A3(2)	<i>SCO4190</i>	DevA	Q9FCH9	1-87	82-291	(Hoskisson <i>et al.</i> , 2006)

3.1.2 Secondary structure analysis

Secondary structure analysis was carried out using the PSI-PRED web server (Jones, 1999, Buchan *et al.*, 2013). The sub-family reference sequences were included in secondary structure analysis to confirm the integrity of the prediction. All of the full length secondary structure predictions can be found in **Appendix 1**. The secondary structure architecture of the C-terminal domains for each target is listed in **Table 3.3**. Despite avoiding the main sub-families of FadR and HutC during BLAST selection of targets, it is clear that several targets fall in to these groups based on secondary structure predictions. Only four targets show truly unique secondary structure architecture (Gp26, FucR, CLOBOL_00895, Reut_B4779). Eleven targets can be attributed to the FadR sub-family and another nine to the HutC family. Three belong to the YtrA subfamily, two to DevA and one to MocR.

Table 3.3 C-terminal secondary structure predictions of selected GntR regulators

Rank	Organism	Gene	Protein	UniProt/PDB code	C-terminal secondary structure prediction	Sub-family designation
1	<i>P. aeruginosa</i> PA14	<i>PA14_67420</i>	HutC	Q02ER1	- α - β - α - β - β - α - β - β - β -	HutC
2	<i>S. coelicolor</i> A3(2)	<i>SCO4190</i>	DevA	Q9FCH9	- α - α - α - α - α - β - β - β - α -	DevA
3	<i>S. coelicolor</i> A3(2)	<i>SCO4188</i>	DevE	Q9FCI0	- α - α - α - α - α - β - β - β - α -	DevA
4	<i>Streptomyces</i> Phage <i>phi-C31</i>	<i>gp26</i>	Gp26	Q9T219	- α - α - α - β - β - β -	Novel topology
5	<i>P. aeruginosa</i> PA14	<i>PA14_71680</i>	Putative transcriptional regulator, GntR	Q02DT0	- α - α - β - α - β - α - β - β - α - β - β - α - β - α - β - α - β - α -	MocR-like
6	<i>S. coelicolor</i> A3(2)	<i>SCO2182</i>	Putative GntR	Q9S2Q4	- α - α - α - α - α - α - α -	FadR
7	<i>S. coelicolor</i> A3(2)	<i>SCO7702</i>	Putative transcriptional regulator, GntR	Q93JH0	- α - α - α - α - α - α - α -	FadR
8	<i>S. coelicolor</i> A3(2)	<i>SCO7168</i>	Putative GntR transcriptional regulator	Q9FBS4	- α - α - α - α - α - α - α -	FadR
9	<i>S. coelicolor</i> A3(2)	<i>SCO2442</i>	Putative GntR	Q9L0A4	- α - α - α - α - α - α - α -	FadR
10	<i>S. coelicolor</i> A3(2)	<i>SCO7056</i>	Putative GntR	Q9FC28	- α - β - α - β - β - α - β - α - β - β - β - β -	HutC

23	<i>Corynebacterium glutamicum</i>	<i>cg3261</i>	Regulatory protein, GntR	Q8NLJ5 (PDB 2ek5)	- α - α - α -	YtrA
24	<i>Enterococcus faecalis</i>	<i>EF_1328</i>	Transcriptional regulator, GntR	Q835P8 (PDB 3ddv)	- α - β - α - β - β - α - β - α - β - β - β -	HutC
25	<i>E. coli</i> O6	<i>c4276</i>	Putative regulator	Q8FCM7 (PDB 3hfi)	- α - β - α - β - β - α - β - α - β - β - β -	HutC
26	<i>Streptomyces phaeochromogeness</i>	<i>traR</i>	TraR	Q54677 (PDB 1v4r)	- α - β - α - β - β - α - β - α - β - β - β -	HutC
27	<i>Ralstonia eutropha</i> JMP134	<i>Reut_B4629</i>	Transcriptional regulator, GntR	Q46SA5 (PDB 3ihuA)	- α - α - α - α - α - α -	FadR
28	<i>P. syringae</i>		Regulatory protein, GntR	Q19AK4 (PDB 3c7j)	- α - α - α - α - α - α -	FadR
29	<i>Rhodococcus jostii</i> RHA1	<i>RHA1_ro03477</i>	Probable transcriptional regulator, GntR	Q0SB06 (PDB 2hs5)	- α - α - α - α - α - α -	FadR
30	<i>Oenococcus oeni</i>	<i>OEOE_1803</i>	Transcriptional regulator, GntR	Q04D30 (PDB 3by6A)	- α - α -	YtrA

3.1.3 Phylogenetic analysis of selected GntR regulators

It is common throughout nature that protein sequences have become very diverse whereas structures can still be highly conserved and can perform differing functions to the original sequence. This is the case with some members of the GntR superfamily. One such example of this is the chorismate lyase fold. Chorismate lyase is an enzyme encoded by the *ubiC* gene, which catalyses the removal of the pyruvate from chorismate forming in 4-hydroxybenzoate (4HB) in the ubiquinone biosynthesis pathway of *E. coli* (Gallagher *et al.*, 2001, Aravind & Anantharaman, 2003).

The chorismate lyase fold is so highly conserved, it gives rise to the HutC subfamily via the UTR; UbiC transcription regulator domain whereby it functions as a small ligand binding domain causing a conformational change when ligand is bound.

There is also evidence in the literature which suggests that events have taken place which give rise to different Eb/O domains (representing the sub-families) being fused with a common HTH domain. Following the fusion event it would seem that proteins within the sub-family have arisen from duplication (Rigali *et al.*, 2002). Phylogenetic analysis of the selected GntR proteins was carried out as a means to determine new subfamilies based on C-terminal domains as the HTH domain remains highly conserved throughout the GntR superfamily and, indeed, defines it (Haydon & Guest, 1991).

Phylogenetic analyses were carried out using MEGA v.6 software (Tamura *et al.*, 2013). Alignments were carried out using the ClustalW algorithm (Larkin *et al.*, 2007) on full length, N-terminal and C-terminal amino acid sequences. Neighbour

joining and maximum parsimony trees were built for all alignments except from the C-terminal which showed too much diversity after alignment and gap deletion for a NJ tree to be built. All phylogenetic trees were rooted with FadR as the outgroup.

Analysis of the full length amino acid sequences reveals high diversity between the 30 targets (**Figure 3.2A**). Some subfamilies such as HutC, DevA, MocR and YtrA are obvious from both trees. Both the neighbour joining tree and maximum parsimony tree show similar patterns in the clustering of proteins showing the robustness of the trees. However, due to the highly conserved nature of the N-terminal no conclusions can be drawn about further sub-families from this.

Phylogenetic analysis of the N-terminal alone has similar clustering patterns to the full length trees due to the high homology in the N-terminal domain (**Figure 3.2B**). Of particular note here is that Gp26 appears as a distinct branch, indicating that the HTH domain probably arose from a distinct genetic event. The MocR, YtrA and PlmA subfamilies cluster in the same clade reflecting amino acid similarities. These subfamilies have most likely arisen from replacement of the C-terminal domain. Several targets appear to cluster with these sub-families, further evidencing the hypothesis of C-terminal domain replacement. This could also be the case with SCO7056 which is a FadR subfamily member but clusters closely with DevA during phylogenetic analysis of the HTH domain. HutC also forms a distinct branch, however, other sub-family members appear elsewhere in tree indicating recruitment of the C-terminal domain to perform different a variety of biological functions.

Due to the diversity in the C-terminal domain, only a maximum parsimony could be calculated. This tree generally reflects the observations for C-terminal secondary

structure predictions. The targets which showed novel C-terminal secondary structure architecture are marked with arrows in **Figure 3.2C**.

In this case, Gp26 clusters together with DevA and DevE. Through secondary structure analysis, Gp26 has unique C-terminal topology. This indicates that the sequence of the C-terminal may have been acquired from the genome of *S. coelicolor* at some point in evolutionary history (**Figure 3.2C**). In general, all of the targets which were predicted to contain FadR-like C-terminal secondary structures clustered closely together, indicating the conserved nature of the FadR structural domain. Interestingly, SCO2182 appears to cluster with the PlmA sub-family although secondary structure analysis reveals it to be a FadR subfamily member. The same can be said for those targets which were predicted to be HutC family members.

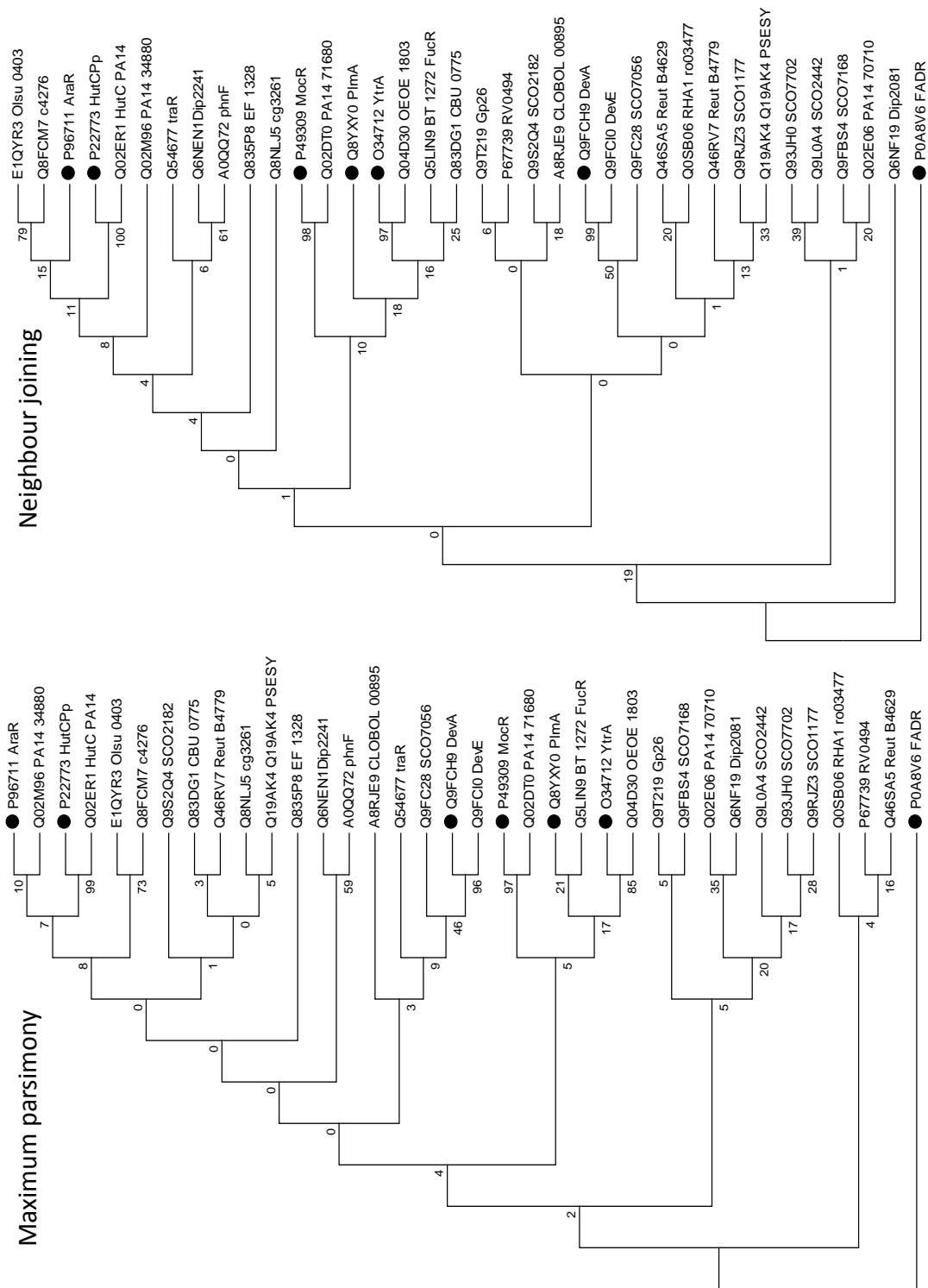


Figure 3.2 (A) Neighbour-joining (R) and maximum parsimony (L) tree inferring evolutionary history of full length GntR genes. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Reference sequences relating to the seven sub-families are represented by dots.

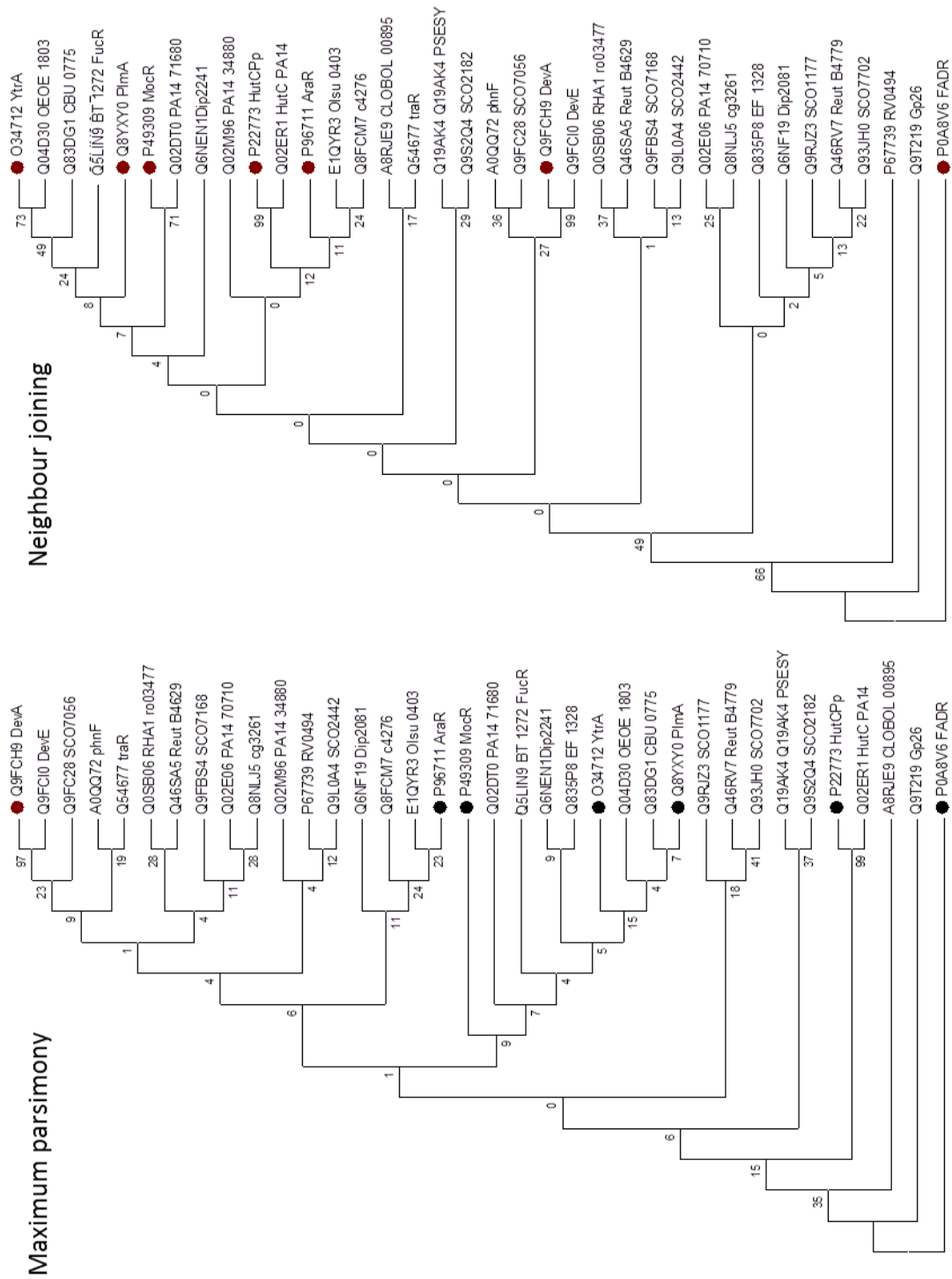


Figure 3.2 (B) Neighbour joining (R) and maximum parsimony (L) tree inferring evolutionary history of the N-terminal of GntR genes (N and C termini are shown in **Table 3.1**). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Reference sequences relating to the seven sub-families are represented by dots.

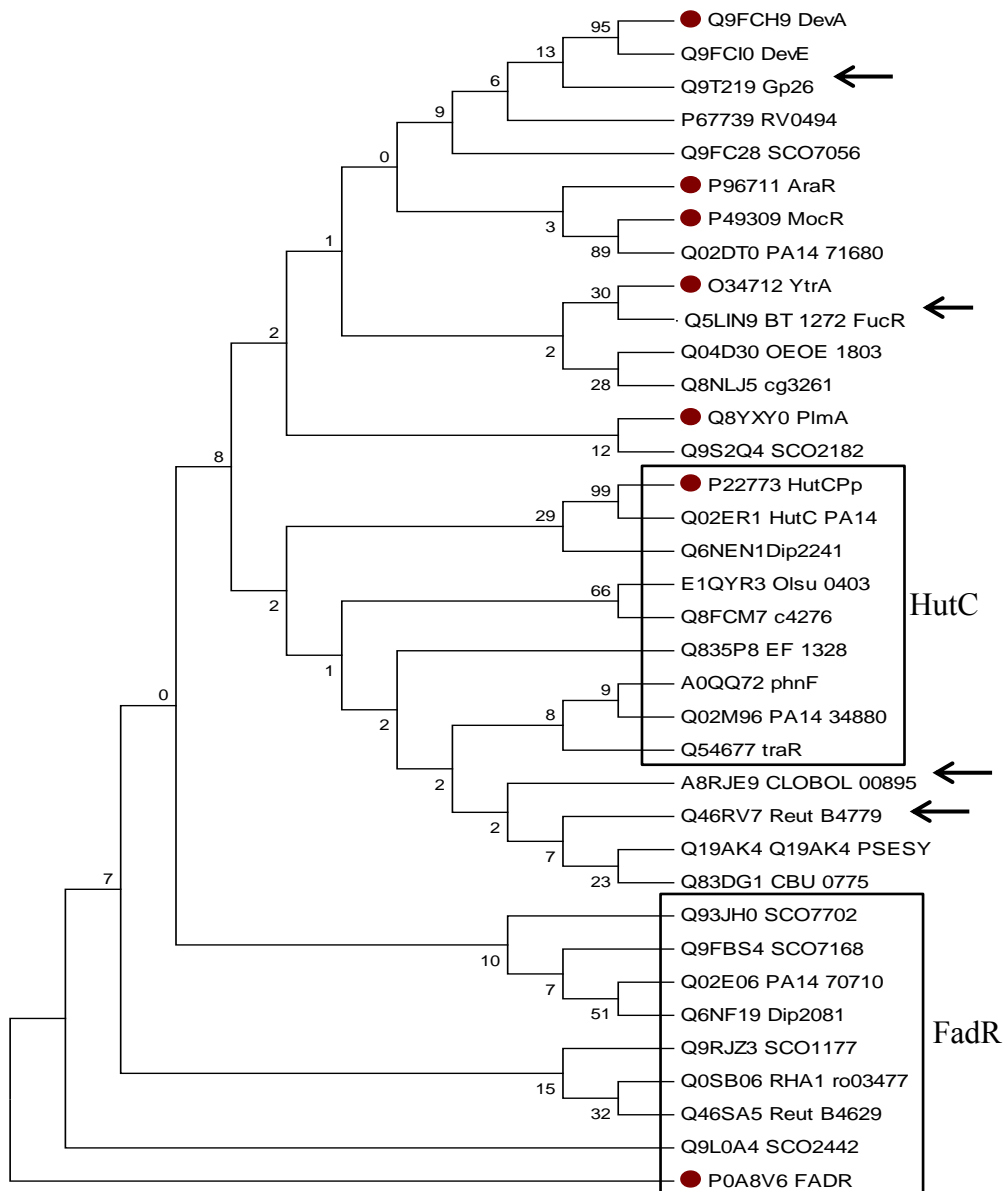


Figure 3.2 (C) Maximum parsimony tree showing evolutionary history of the C-terminal of GntR genes (N and C termini are shown in **Table 3.1**). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Reference sequences relating to the seven subfamilies are represented by dots. Arrows represent GntR-like regulators which showed unique C-terminal architecture during secondary structure prediction. Neighbour joining tree could not be calculated

3.2 Target Cloning strategies and optimisation of sequences for crystallisation

Construction of plasmids to yield the best chance of successful overexpression, purification and crystallisation requires a range of differently designed constructs to be assessed in order to find the best construct. A number of strategies were employed to achieve this including various truncations to eliminate disordered structural regions in the proteins and different amino and carboxy-terminal affinity tags. The success of plasmids was based on the yield and stability of the protein produced.

3.2.1 Cloning into pET vectors

The initial cloning strategy involved cloning full length target genes into the pET100 vector from Invitrogen due ease and speed of cloning. Full length genes were successfully cloned into pET100 vectors yielding the plasmids listed in **Table 2.2**. All plasmids were confirmed by sequencing. However, the protein yield from these plasmids was low and, although usable for functional studies in Electrophoretic Mobility Shift Assays (EMSA), did not yield the amount of protein required for crystallisation or 2D-IR spectroscopy studies. The proteins expressed from the pET100 vectors were found to have poor solubility and precipitated out of solution when concentrated above 2 mg ml⁻¹.

3.2.2 Surface entropy reduction (SERp) mutants

To obtain a higher yield of soluble protein and to produce protein which may be more likely to crystallise, surface entropy mutations were introduced in to DevA, DevE, and HutC. The basic principle of reducing surface entropy relies on mutating high entropy amino acid such as lysine, glutamine and arginine to alanine in order to

force lower entropy thus making it easier and therefore more likely to form protein crystals. Using this approach a non-crystallisable GntR regulator, TM0439, from *Thermotoga maritima* was successfully crystallised (Zheng *et al.*, 2009)

The SERp server (Goldschmidt *et al.*, 2007) was used to determine the residues with highest entropy values. SERp scores which are greater than 3 are good candidate residues for mutation. Where a cluster of residues were found, all residues within the cluster were mutated. The candidate residues for entropy mutation for HutC, DevA, DevE and Gp26 are shown in **Table 3.4**.

Gp26 was not a candidate for SERp mutation because all the amino acid residues identified had scores of less than 3 therefore their entropic value was not judged to be a factor to affect the likelihood of crystallisation. All seven of the SERp suggested mutations were introduced to HutC, DevA and DevE.

Synthetic genes were designed for the SERp mutants and were codon optimised for expression in *E. coli*. The synthetic genes were designed to contain 5' *KpnI* and a 3' *HindIII* site for ease of cloning. Excision and cloning of the synthetic gene in to pOPINF (vector maps can be found in **appendix 3**) was carried out using the *KpnI* and *HindIII*. Genes were synthesised by Integrated DNA Technologies and supplied in pEX vectors (**appendix 3**). The SERp approach did not affect protein expression yields for the three proteins for which the SERp method was applied optimisation, although overall solubility was improved and the proteins did not precipitate as easily during concentration, therefore higher concentrations of protein were achieved. Differences in expression of WT proteins compared to SERp optimised proteins are detailed in **Section 3.4 (Table 3.10 and Figures 3.20-3.24)**.

Table 3.4 Candidate mutations to lower surface entropy in selected GntR proteins.

Protein		Cluster #1	Cluster #2	Cluster #3	Cluster #4	Cluster #5	Cluster #6	Cluster #7	Cluster #8
PA14_HutC	Residues	200 - 206	120 – 121	86 - 89					
	Mutations	E201A E204A K206A	E120A E121A	E87A K89A					
	SERp Score	4.28	4.03	3.19					
DevA	Residues	235 -237	166 -167	271 - 272					
	Mutations	E235A E236A E237A	E166A E167A	E271A E 272A					
	SERp Score	5.63	5.11	3.66					
DevE	Residues	246 – 248	282 – 283	270 - 271					
	Mutations	E246A Q247A E248A	K282A Q283A	E270A Q271A					
	SERp Score	5.09	3.04	2.95					
Gp26	Residues	126	102 – 103	22	164	96	144	155	71 – 72
	Mutations	E126A	E102A	K22A	E164A	E96A	K44A	E155A	E72A
	SERp Score	2.63	2.54	2.35	2.32	2.11	2.09	2.09	2.02

3.2.3 High-throughput cloning

A major issue was encountered with protein expression levels and the amount of soluble protein obtained was still relatively low and many litres of overexpression culture were required to obtain enough protein for crystallisation trials and 2D-IR spectroscopy experiments. In order to explore and assess many more construct designs to overcome the low protein issue in a quick and efficient way, the high-throughput facility at the Oxford Protein Production Facility (OPPF) was utilised.

Various constructs were designed for seven different proteins in order to find the best construct to express maximum protein. Construct design included N and C terminal truncations based on disorder prediction using the RONN server (Yang *et al.*, 2005) and different affinity tags to try to increase solubility as protein expressed in pET vectors previously precipitated during concentration. A full list of protein vectors, truncations and tags designed can be found in **Table 3.5**. N- and C-terminal truncations listed are in relation to the native amino acid sequence. Further information on the vectors and tags used can be found in the **section 3.3**. To ensure that truncations did not result in disruptions to overall protein structure, secondary structure predictions were carried out to ensure the HTH domain remained intact. Secondary structure analysis was carried out using PSIPRED v3.3 (Jones, 1999) within the PSIPRED Protein Sequence Analysis Workbench server (Buchan *et al.*, 2013).

The truncations to HutC from residue 2- 250 amino acids (ID 16305) and 20 -250 (ID 16303) were predicted not to alter overall secondary structure (**Fig. 3.3**).

The secondary structure prediction for DevA (**Fig 3.4**) shows that the N-terminal 17-291 (ID 16281) truncation may have an effect on the secondary structure of the second β strand in the HTH domain. The confidence prediction of this section is low (score = 4), however, which may be an effect of the secondary structure prediction algorithm.

DevE secondary structure was affected by the N-terminal 47 -303 amino acid truncation (ID 16296) with the HTH domain appearing as H-H- β -H- β rather than the traditional motif of 3 helices followed by 2 β strands (**Fig. 3.5**)

Gp26 remains unaffected by N-terminal truncations at the HTH domain (**Fig. 3.6**). Minor changes in the secondary structure appear in the C-terminal although the confidence prediction for the 21-174 (ID 16314; score = 3) is higher than for the 2-174 truncation (ID 16312; score = 1). Therefore, as with DevA, this may be an effect of the prediction algorithm.

To summarise, sequence optimisation was an important step to achieve the highest yield of protein and highest level of solubility allowing the best version of the protein to be determined by taking a logical approach to protein production. Codon optimisation had potential to increase yield of protein produced, while surface entropy reduction had potential to increase likelihood of protein crystallisation. Moreover, using the HTP pipeline at OPPF was useful due to the high number of different constructs that could be made simultaneously. Optimisation at this level eliminated disordered regions and enabled various different affinity tags (N and C terminal hexa-histidine and SUMO) and vectors to be assessed systematically.

Table 3.5 Table of HTP construct vectors, tags and truncations

Well	Gene name	ID No.	pOPIN vector	Tag	N terminal amino acid truncation	C terminal amino acid truncation
A1	DEVA_native	16279	E	C - terminal 6x His	2	291
B1	DEVA_native	16280	S3C	SUMO	2	291
C1	DEVA_native	16281	E	C - terminal 6x His	17	291
D1	DEVA_native	16282	F	N -terminal 6x His	17	291
E1	DEVA_native	16283	S3C	SUMO	17	291
F1	DEVE_native	16291	E	C - terminal 6x His	2	303
G1	DEVE_native	16292	S3C	SUMO	2	303
H1	DEVE_native	16293	E	C - terminal 6x His	21	303
A2	DEVE_native	16294	F	N -terminal 6x His	21	303
B2	DEVE_native	16295	E	C - terminal 6x His	46	303
C2	DEVE_native	16296	F	N -terminal 6x His	47	303

D2	STRCO Putative	16322	E	C - terminal 6x His	2	207
E2	STRCO Putative	16323	F	N -terminal 6x His	2	207
F2	STRCO Putative	16324	E	C - terminal 6x His	8	207
G2	STRCO Putative	16325	F	N -terminal 6x His	8	207
H2	HUT_native	16305	E	C - terminal 6x His	2	250
A3	HUT_native	16303	E	C - terminal 6x His	20	250
B3	HUT_native	16304	F	N -terminal 6x His	20	250
C3	HUT_native	16306	E	C - terminal 6x His	20	249
D3	HUT_native	16307	F	N -terminal 6x His	20	249
E3	PA14_34660 GntR gene	16316	E	C - terminal 6x His	2	343
F3	PA14_34660 GntR gene	16317	F	N -terminal 6x His	2	343

G3	PA14_34660 GntR gene	16318	E	C - terminal 6x His	16	343
H3	PA14_34660 GntR gene	16319	F	N -terminal 6x His	17	343
A4	PA14_34660 GntR gene	16320	E	C - terminal 6x His	25	343
B4	PA14_34660 GntR gene	16321	F	N -terminal 6x His	25	343
C4	<i>P. fluorescens</i>	16308	E	C - terminal 6x His	2	536
D4	<i>P. fluorescens</i>	16309	F	N -terminal 6x His	2	536
E4	<i>P. fluorescens</i>	16310	E	C - terminal 6x His	10	536
F4	<i>P. fluorescens</i>	16311	F	N -terminal 6x His	10	536
G4	Gp26	16312	E	C - terminal 6x His	2	174
H4	Gp26	16313	F	N -terminal 6x His	3	174
A5	Gp26	16314	E	C - terminal 6x His	21	174

B5	Gp26	16315	F	N -terminal 6x His	22	174
C5	DEVA_SERp	16284	E	C - terminal 6x His	2	276
D5	DEVA_SERp	16285	F	N -terminal 6x His	2	276
E5	DEVA_SERp	16286	S3C	SUMO	2	276
F5	DEVE_SERp	16287	E	C - terminal 6x His	2	284
G5	DEVE_SERp	16288	F	N -terminal 6x His	2	284
H5	DEVE_SERp	16289	E	C - terminal 6x His	28	284
A6	DEVE_SERp	16290	F	N -terminal 6x His	28	284
B6	HUT_SERp	16297	E	C - terminal 6x His	2	232
C6	HUT_SERp	16298	F	N -terminal 6x His	2	232
D6	HUT_SERp	16299	E	C - terminal 6x His	4	232
E6	HUT_SERp	16300	F	N -terminal 6x	4	232

				His		
F6	HUT_SERp	16301	E	C - terminal 6x His	2	231
G6	HUT_SERp	16302	F	N -terminal 6x His	2	231
H6	GFP positive control	n/a	F	N-terminal 6x His	n/a	n/a

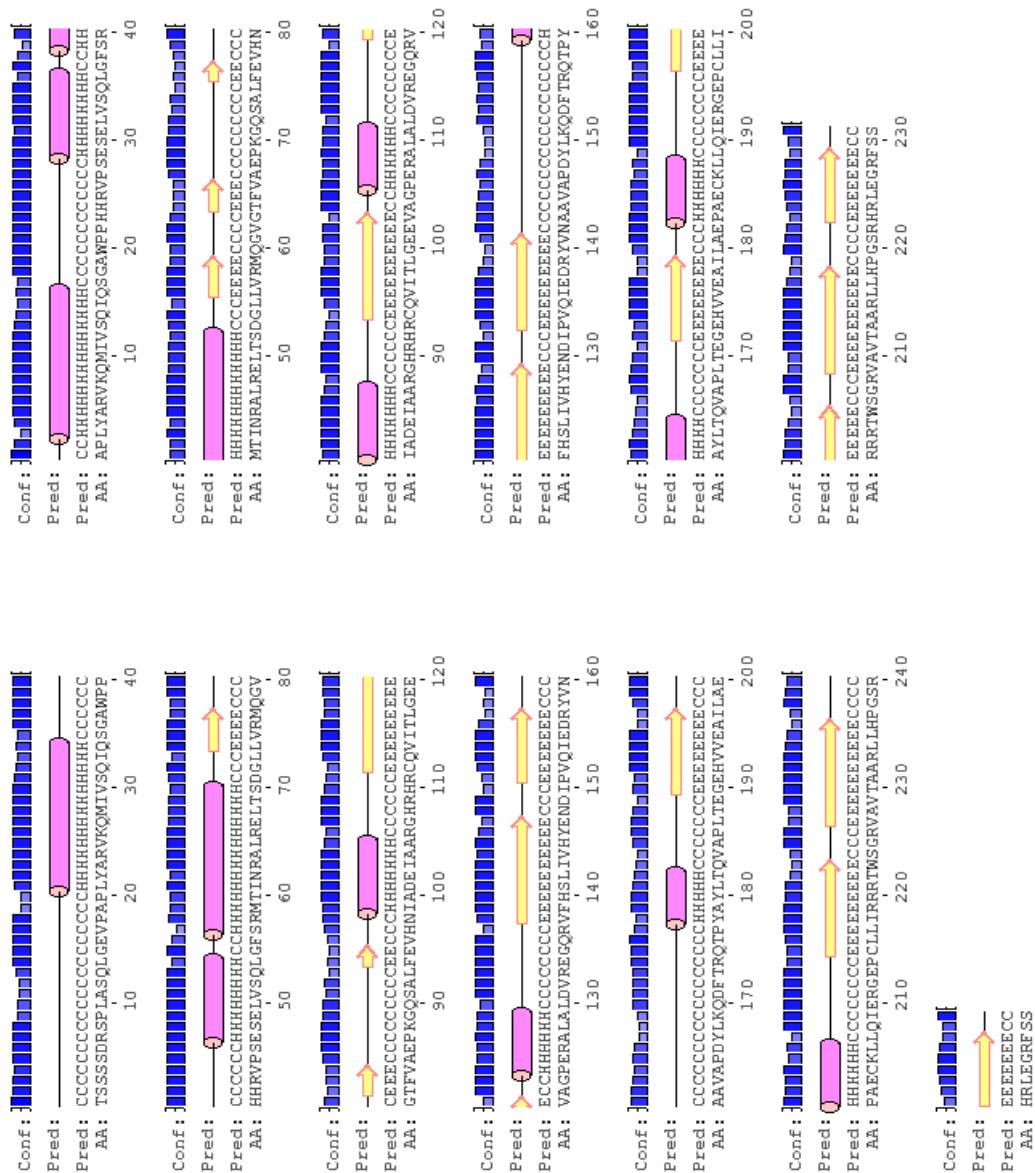


Figure 3.3 Secondary structure predictions of truncated HutC amino acid sequences. The 2-250 (ID 16305) amino acid truncation is shown on the left and the 20-250 (ID 16303) truncation on the right. The truncations do not appear affect the secondary structure of HutC. Structure analysis was carried out using PSIPRED v3.3. Pink cylinders = helix; Yellow arrow = beta strand; Black line = coil.

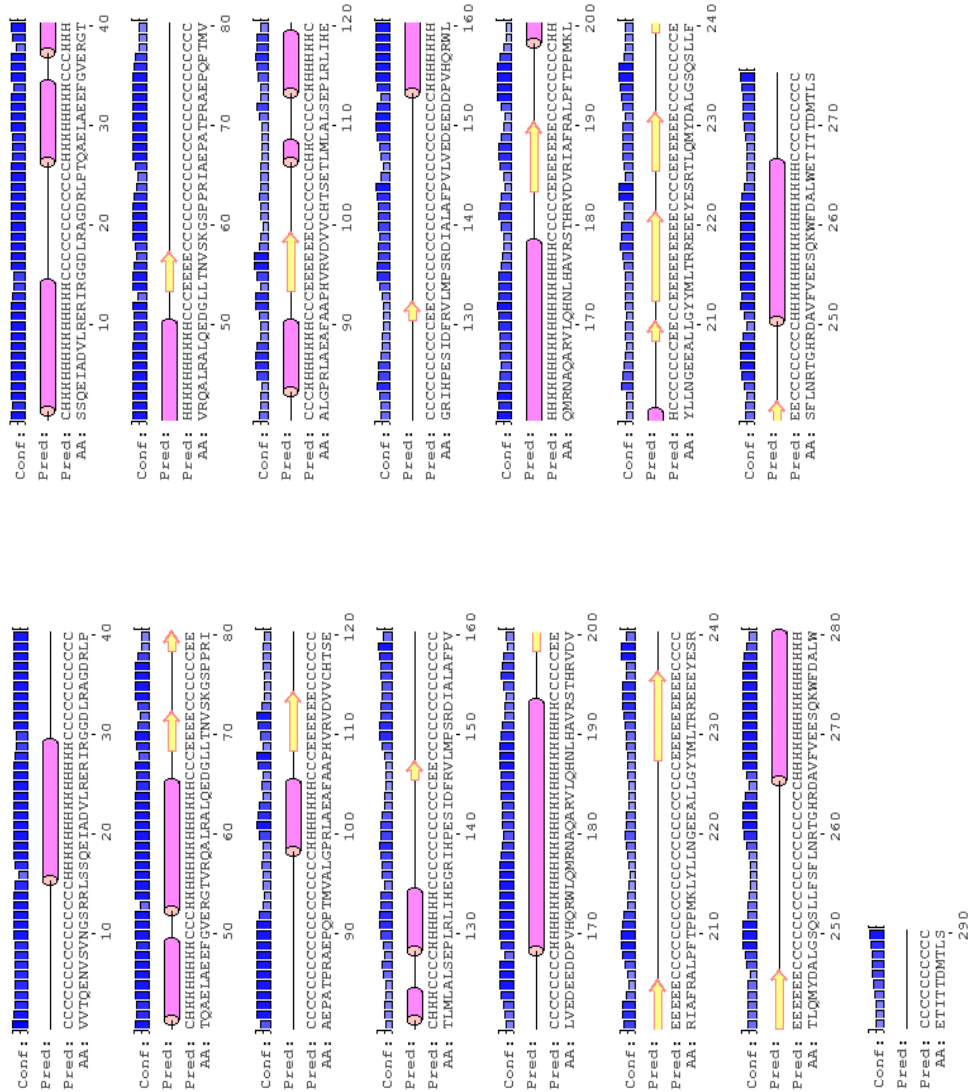


Figure 3.4 Secondary structure predictions of truncated DevA amino acid sequences. The 2-291 (ID 16279) amino acid truncation is shown on the left and the 17-291 (ID16281) truncation on the right. The 17-291 truncation appears to slightly affect the secondary structure of DevA. Structure analysis was carried out using PSIPRED v3.3. Pink cylinders = helix; Yellow arrow = beta strand; Black line = coil

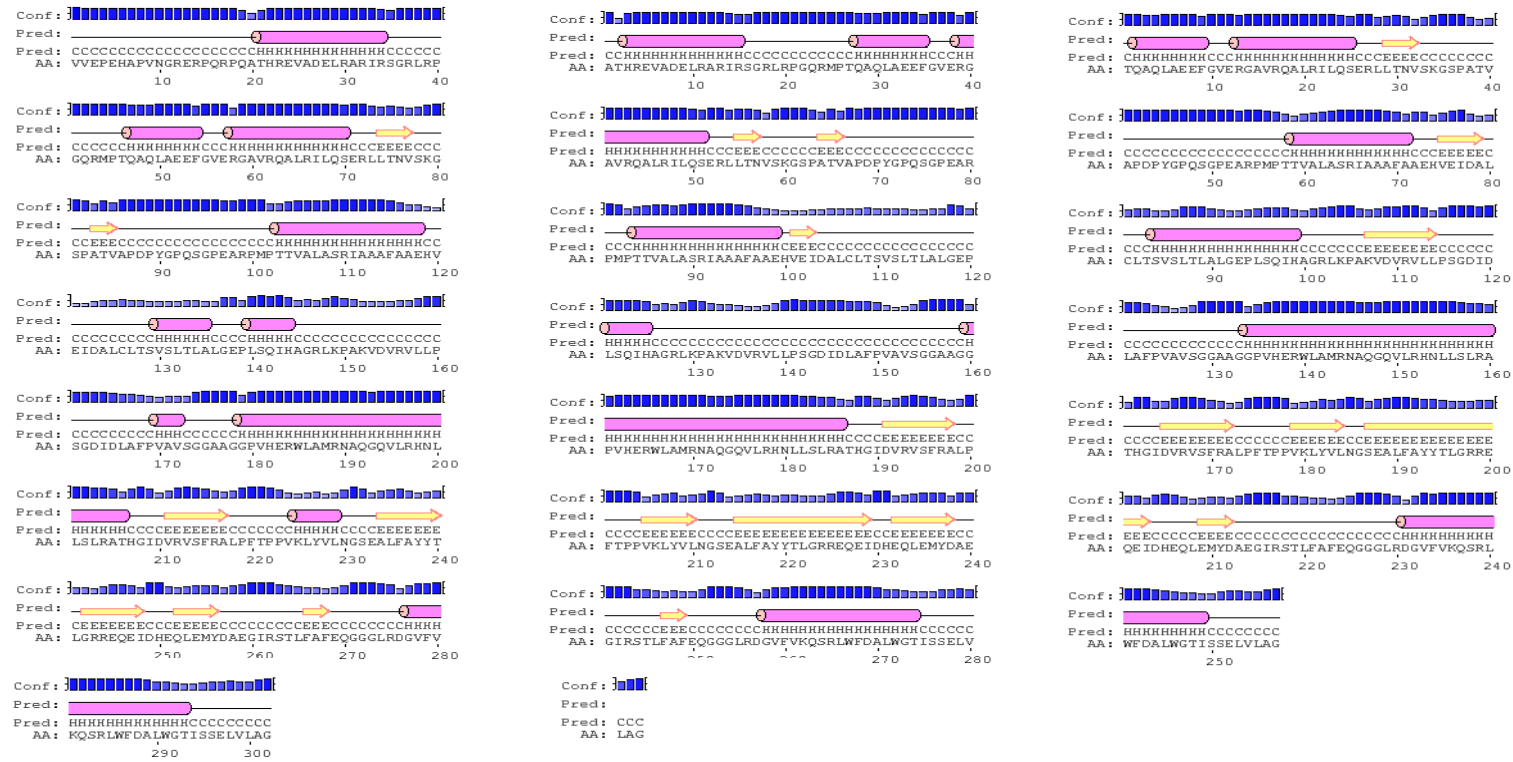


Figure 3.5 Secondary structure predictions of truncated DevE amino acid sequences. The 2-303 (ID 16291) amino acid truncation is shown on the left, 21-303 (ID 16294) in the middle and the 47-303 (ID 16296) truncation on the right. The 47-303 truncation appears to slightly affect the secondary structure of the HTH domain. Structure analysis was carried out using PSIPRED v3.3. Pink cylinders = helix; Yellow arrow = beta strand; Black line = coil

3.3 Construction of GntR-like protein overexpression plasmids

This section describes the construction and verification of GntR overexpression plasmids. Initially, a commercial kit from Invitrogen was used to clone the four genes that make the main focus of the project. Following this, a high throughput approach was used to clone a range of variants of each gene such as truncations into a range of vectors with a view to obtaining the best construct yielding the most soluble and highest yield of protein for further functional and structural studies. This section also details the cloning of potential upstream (promoter) regions for HutC, DevA, DevE and Gp26 for further functional analysis in **Chapter 5**.

3.3.1 Cloning of full length GntR target genes into pET100 vectors

The pET100 vector from Invitrogen was chosen as a start point for cloning due to the speed and efficiency of cloning. It offers advantages such as directional cloning and an N-terminal hexa-histidine purification tag which can be cleaved with enterokinase (EK) if required. It also has an inducible promoter (T7) which allows control of protein overexpression through IPTG (Studier *et al.*, 1990).

Full-length DNA sequences encoding the GntR-like proteins (*hutC*, *SCO4190*, *SCO4188* and *gp26*) were subjected to PCR using the primers detailed in **Table 2.4**. Gene sequences were obtained from NCBI and gene specific primers were designed which contained a 5' CACC overhang on the forward primer for directional cloning. PCR products of the expected size were purified from the gel and were used to clone into pET100. Amplicons are highlighted in **Figure 3.7**. The negative control for Gp26 (lane 3) appears to have PCR product at the same size as *gp26* gene. The PCR

product was extracted and sequencing confirmed it was Gp26. This was assumed to be contaminated with PhiC31 genomic DNA.

The resulting plasmids from the cloning reaction were designated pKR003 (DevA gene), pKR006 (HutC gene), pKR007 (Gp26 gene) and pKR008 (DevE gene). pET100 backbone plasmids were confirmed by restriction digest with *NdeI* and *Sall* (pKR003, pKR008) or *NdeI* and *BglII* (pKR006, pKR007; **Figure 3.8**). **Table 3.6** details the expected band sizes from these restriction digests. The directionality and sequence were confirmed by sequencing (MWG Eurofins, London). Plasmids maps were assembled using SnapGene software (GSL Biotech LLC) and are shown in **Figure 3.9A** and **3.9B**.

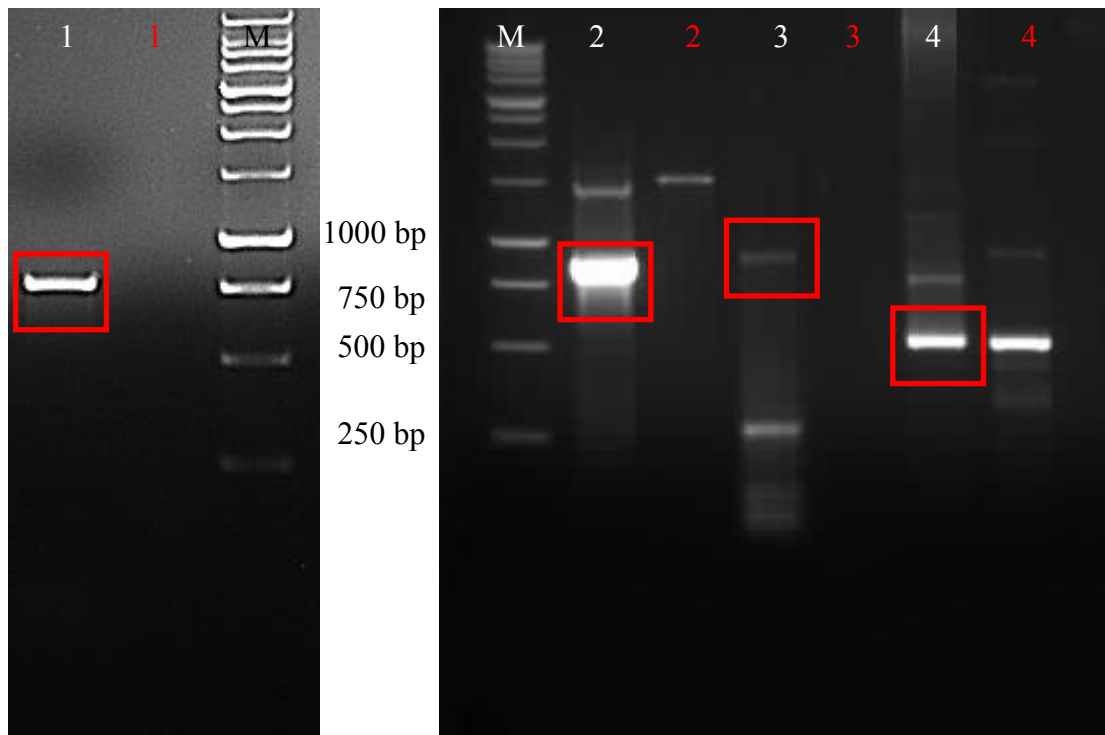


Figure 3.7 PCR products obtained from genomic DNA **1)** HutC 753bp; **2)** DevA 876bp; **3)** DevE 912bp; **4)** Gp26 525bp. Lane numbers in white represent samples which contained template DNA, numbers in red represent samples which contained water instead of template DNA as a negative control. **Marker:** 1 kB Promega; Gel: 0.8% agarose, 80V.

Table 3.6 Band sizes of pET100 backbone plasmids after restriction digest

pKR003	pKR006	pKR007	pKR008
<i>NdeI, Sall</i>	<i>NdeI, BglII</i>	<i>NdeI, BglII</i>	<i>NdeI, Sall</i>
5936	2316	2316	5781
446	1885	1885	566
258	1372	1408	329
	326	446	
	270	234	
	234		
	114		

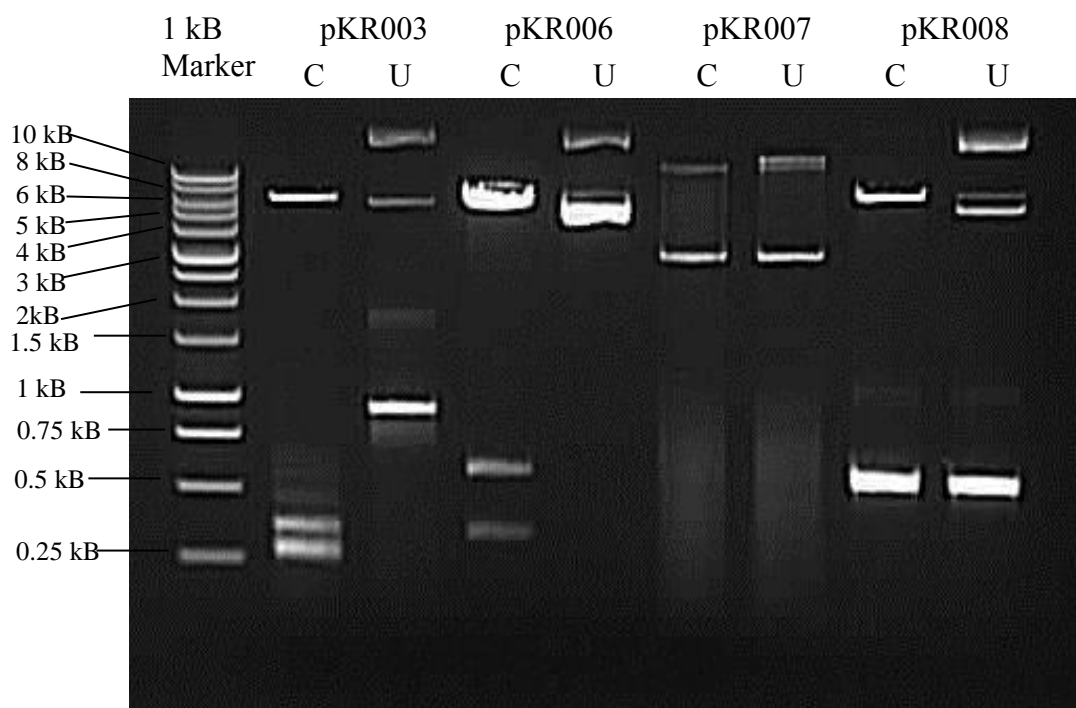


Figure 3.8 Restriction digest of pKR003 and pKR006 with *NdeI* and *Sall*. pKR007 and pKR008 were digested with *NdeI* and *BglIII*. (C) denotes cut plasmid DNA, (U) denotes uncut plasmid DNA. See **Table 3.6** for exact band sizes. **Marker:** 1 kB Promega; Gel: 0.8% agarose, 80V.

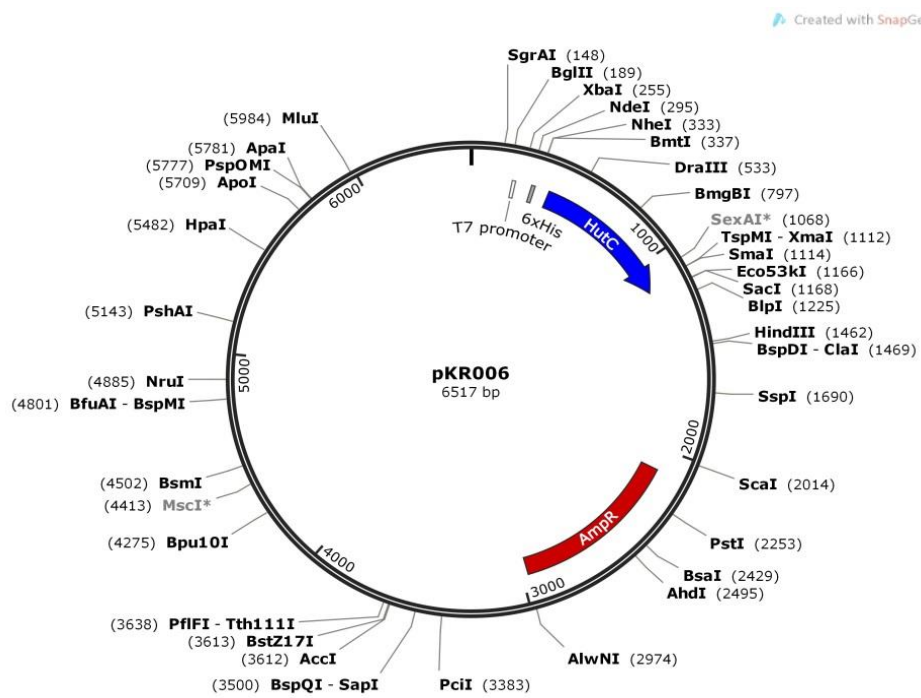
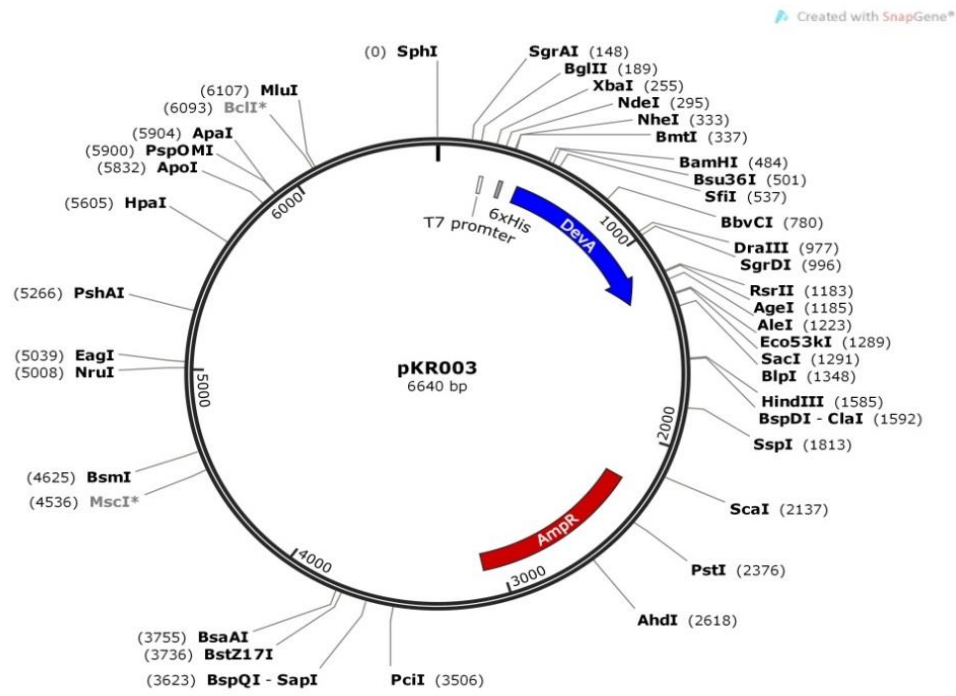


Figure 3.9A Plasmid maps of pKR003 and pKR006, resulting from cloning DevA and HutC genes, respectively, into the pET100 vector via Topo directional cloning. The maps were constructed *in silico* with Snapgene software.

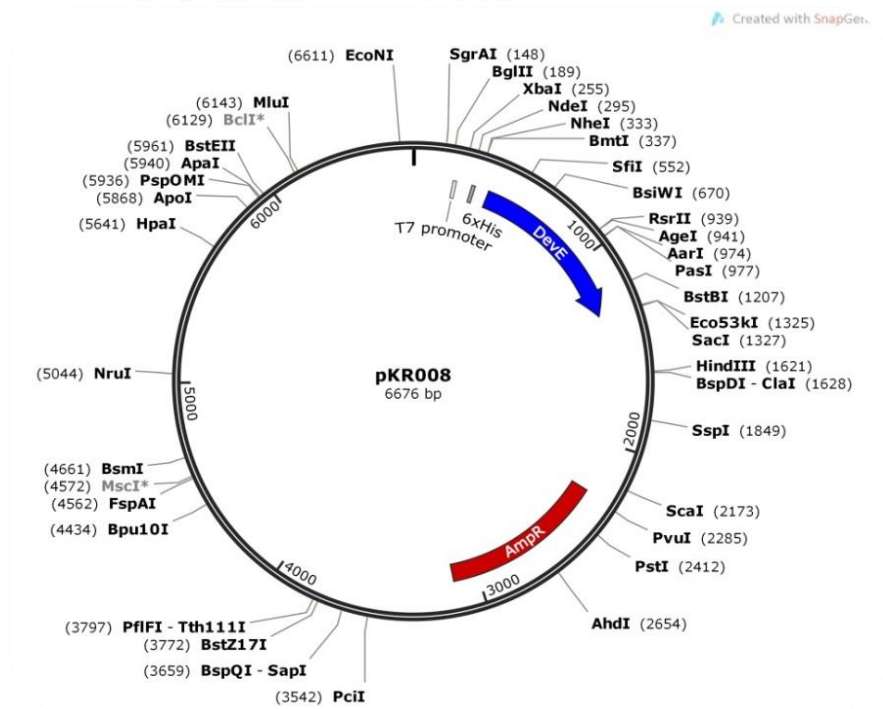
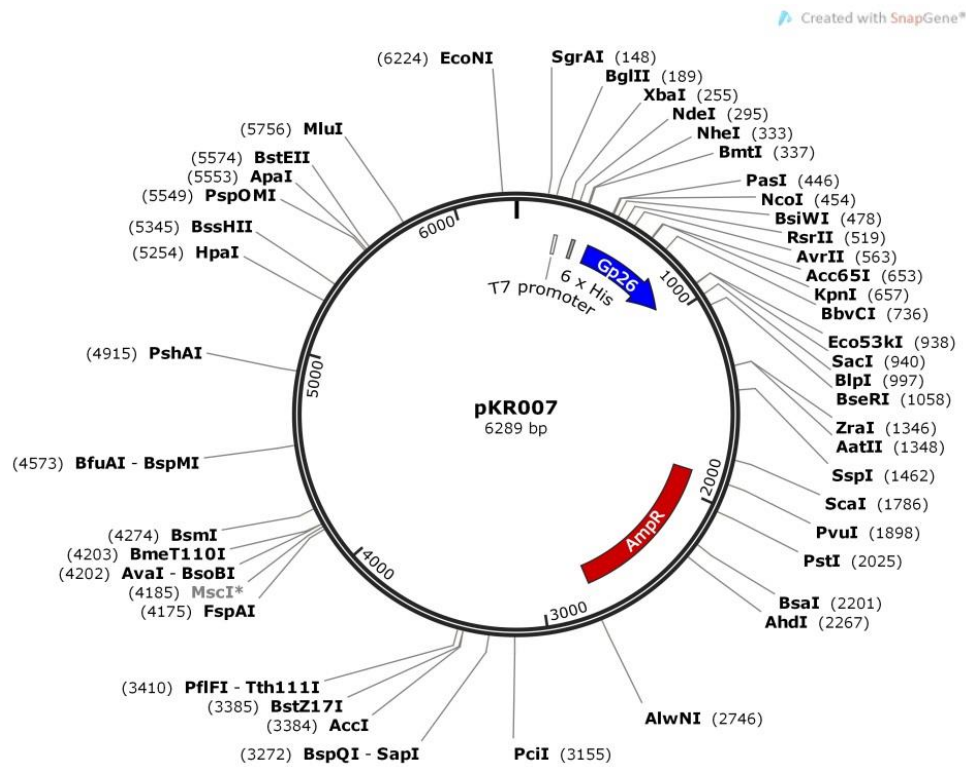


Figure 3.9B Plasmid maps of pKR007 and pKR008, resulting from cloning Gp26 and DevE genes, respectively, into the pET100 vector via Topo directional cloning. The maps were constructed *in silico* with Snapgene software.

3.3.2 High throughput cloning of sequence optimised GntR targets

In order to obtain higher amounts of protein optimised for crystallography, construction of further overexpression plasmids was carried out using the high throughput facility at OPPF, Harwell. The workflow for cloning of constructs is detailed in **Figure 3.10**. The HTP system utilises the In-Fusion cloning reaction from Clontech with pOPIN vectors which were developed in-house (Berrow *et al.*, 2007). The pOPIN vectors have the advantage of the *lacZ* gene allowing positive clones to be selected by blue/white screening on media containing X-gal.

The pOPIN vectors have different tags which can be used to aid solubility of the resulting GntR protein; The pOPINF vector has an N-terminal hexa-histidine tag which can be cleaved quite readily by C3 protease while pOPINE has a C-terminal tag which can be cleaved albeit with slightly less efficiently than with C3 protease. The pOPINS3C vector was also used as it has been shown that the SUMO tag has the ability to enhance protein overexpression and stabilise proteins (Wagner *et al.*, 2008). **Table 3.5** details which vectors were used for each construct.

A wide range of PCR cycling conditions were used to obtain a good yield of the desired amplicon using the primers detailed in **Table 2.5**. Final successful PCR conditions which were used to amplify the various constructs are detailed in **Table 3.7**. The PCR product size for each construct is detailed in **Table 3.8**. The final PCR products used for the In-Fusion reaction are shown in **Figure 3.11**. Six of the final PCR products had a very low yield of amplicons, which was true throughout all the PCR reactions carried out. These 6 genes are shown in **Figure 3.11**, marked with an arrow.

Following the In-Fusion cloning reaction, constructs were transformed into Omnimax cells. Two white colonies were picked from each transformation plate and were designated clone 1 and clone 2. These were cultured overnight and plasmids were then isolated. The resulting plasmids from clone 1 and clone 2 were subject to PCR to verify gene insertion into the vector. The In-Fusion cloning reaction ensures correct directionality of the gene insert during cloning; therefore PCR was able to be used to verify successful clones using the pOPIN sequencing primer (**Table 2.5**). Successfully verified plasmids from clone 1 and 2 are marked in **Figure 3.12**. Twelve out of 47 constructs (E1, D2, F2, H3, A4, B4, G4, B5, E5, B6, D6 and E6) were not verified from clone 1 and 2. Where plasmids were not successfully verified, a further 8 white colonies were picked, cultured, plasmids isolated and subject to PCR verification. Successful clones from these are marked in **Figure 3.13**. Plasmid maps were assembled using SnapGene software (GSL Biotech LLC) for the plasmids that were finally used for large scale protein production of DevA, DevE, HutC and Gp26 (**Figure 3.14** and **3.15**).

Using this system, 41 out of 47 constructs were successfully verified by PCR equating to an 87.2% success rate. Of the 6 constructs which did not return a positive result from PCR verification, all had poor initial PCR product yield which, in turn led to poor cloning results with many blue colonies. This indicates that the *lacZ* gene had not been disrupted, thus the gene was unlikely to have been inserted, making it difficult to pick individual white colonies from the plates.

Table 3.7 PCR conditions, In-Fusion cloning

Construct	Polymerase	Annealing temperature *	Cycles	PCR conditions
All	Phusion	60°C	30	98°C – 10 s 98°C – 1 s Variable annealing temp * – 5 s 72°C – 40s 72°C -1 min
C4, D4, E4, F4	Phusion	55°C	30	
B1, D2- G2, C5-H5	Phusion	65°C	30	
F2, G2	Phusion, PhGC buffer	60°C	40	
C5-H5	Phusion	60°C	40	
All	KOD Xtreme	60°C	30	
B1, G1	KOD Xtreme, high GC buffer	60°C	40	94°C – 2 min 98°C – 10 s 60°C – 30s 68°C – 2 min 68°C – 3 min
A6-G6	KOD Xtreme	60°C	40	

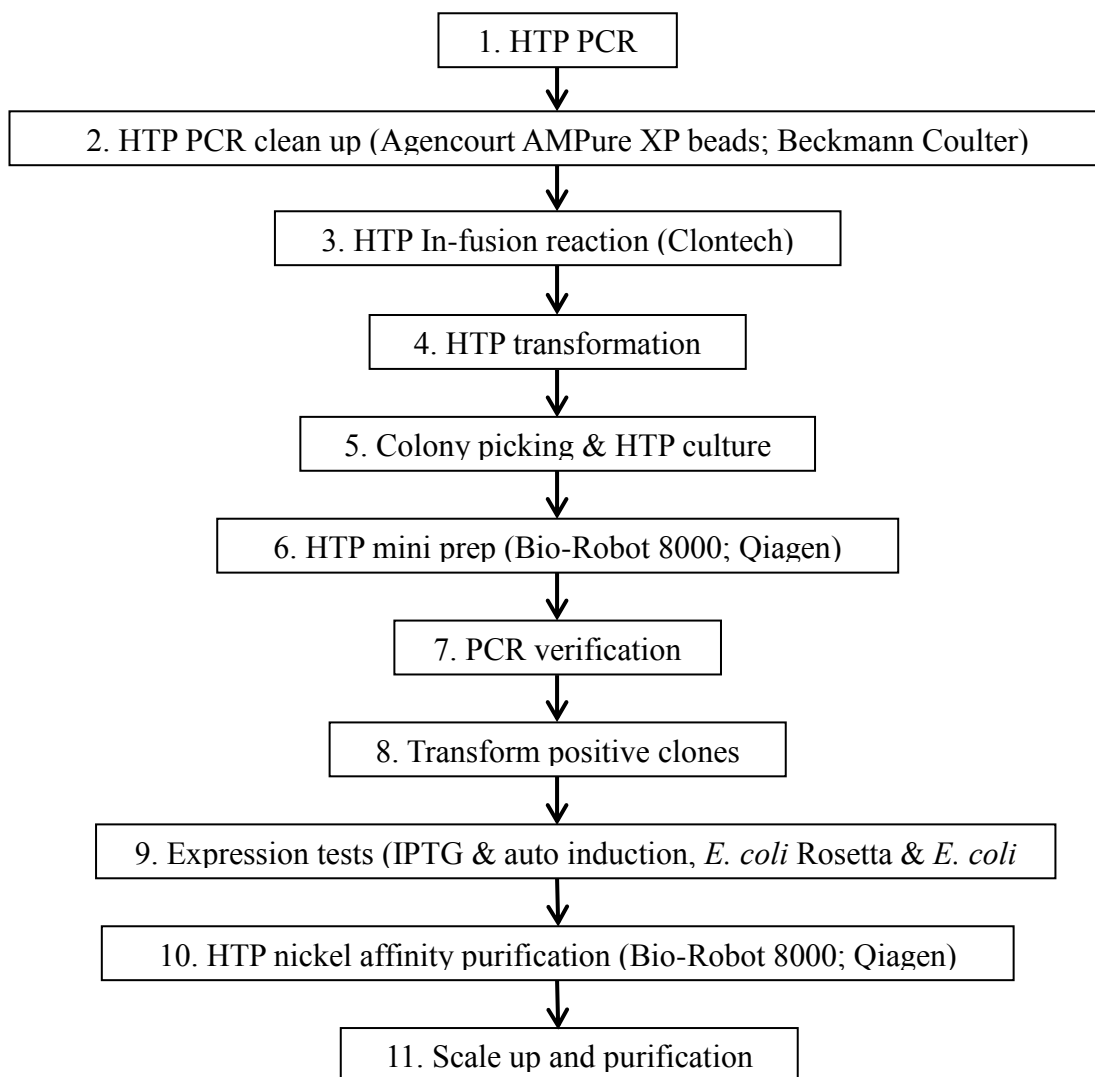


Figure 3.10 Workflow for HTP construction of GntR overexpression plasmids at the Oxford Protein Production Facility (OPPF), Harwell, Oxfordshire. Steps 5 – 7 were repeated if PCR verification was not successful for some clones.

Table 3.8 PCR product sizes for HTP gene cloning using OPPF

Well	Gene name	AA_N	AA_C	pOPIN Vector	PCR/bp
A1	DEVA_native	2	291	E	900
B1	DEVA_native	17	291	E	855
C1	DEVA_native	17	291	F	855
D1	DEVA_native	2	291	S3C	900
E1	DEVA_native	17	291	S3C	855
F1	DEVE_native	2	303	E	936
G1	DEVE_native	21	303	E	879
H1	DEVE_native	46	303	E	804
A2	DEVE_native	21	303	F	879
B2	DEVE_native	47	303	F	801
C2	DEVE_native	2	303	S3C	936
D2	STRCO Putative	2	207	E	648
E2	STRCO Putative	8	207	E	630
F2	STRCO Putative	2	207	F	648
G2	STRCO Putative	8	207	F	630
H2	HUT_native	20	250	F	723
A3	HUT_native	20	249	F	720
B3	HUT_native	2	250	E	777
C3	HUT_native	20	250	E	723
D3	HUT_native	20	249	E	720
E3	PA14_34660 GntR gene	2	343	E	1056
F3	PA14_34660 GntR gene	16	343	E	1014
G3	PA14_34660 GntR gene	25	343	E	987
H3	PA14_34660 GntR gene	2	343	F	1056
A4	PA14_34660 GntR gene	17	343	F	1011
B4	PA14_34660 GntR gene	25	343	F	987
C4	<i>P. fluorescens</i>	2	536	F	1635
D4	<i>P. fluorescens</i>	10	536	F	1611
E4	<i>P. fluorescens</i>	2	536	E	1635
F4	<i>P. fluorescens</i>	10	536	E	1611
G4	Gp26	2	174	E	549
H4	Gp26	21	174	E	492
A5	Gp26	3	174	F	546
B5	Gp26	22	174	F	489
C5	DEVA_SERp	2	276	F	855
D5	DEVA_SERp	2	276	S3C	855
E5	DEVA_SERp	2	276	E	855
F5	DEVE_SERp	2	284	E	879
G5	DEVE_SERp	28	284	E	801
H5	DEVE_SERp	2	284	F	879
A6	DEVE_SERp	28	284	F	801
B6	HUT_SERp	2	232	F	723
C6	HUT_SERp	4	232	F	717
D6	HUT_SERp	2	231	F	720
E6	HUT_SERp	2	232	E	723
F6	HUT_SERp	4	232	E	717
G6	HUT_SERp	2	231	E	720



Figure 3.11 HTP PCR products used for In-Fusion cloning. DevA native (A1-E1), DevE native (F1-H1, A2-C2), STRCO putative (D2-F2), HutC native (G2, H2, A3-C3), PA14_34660 (D3-H3, A4, B4), *P. fluorescens* (C4-F4), Gp26 (G4, H4, A5, B5), DevA SERp (C5-E5), DevE SERp (F5-H5, A6), HutC SERp (B6-G6). The various truncations and PCR products are detailed previously in **Table 3.9**.

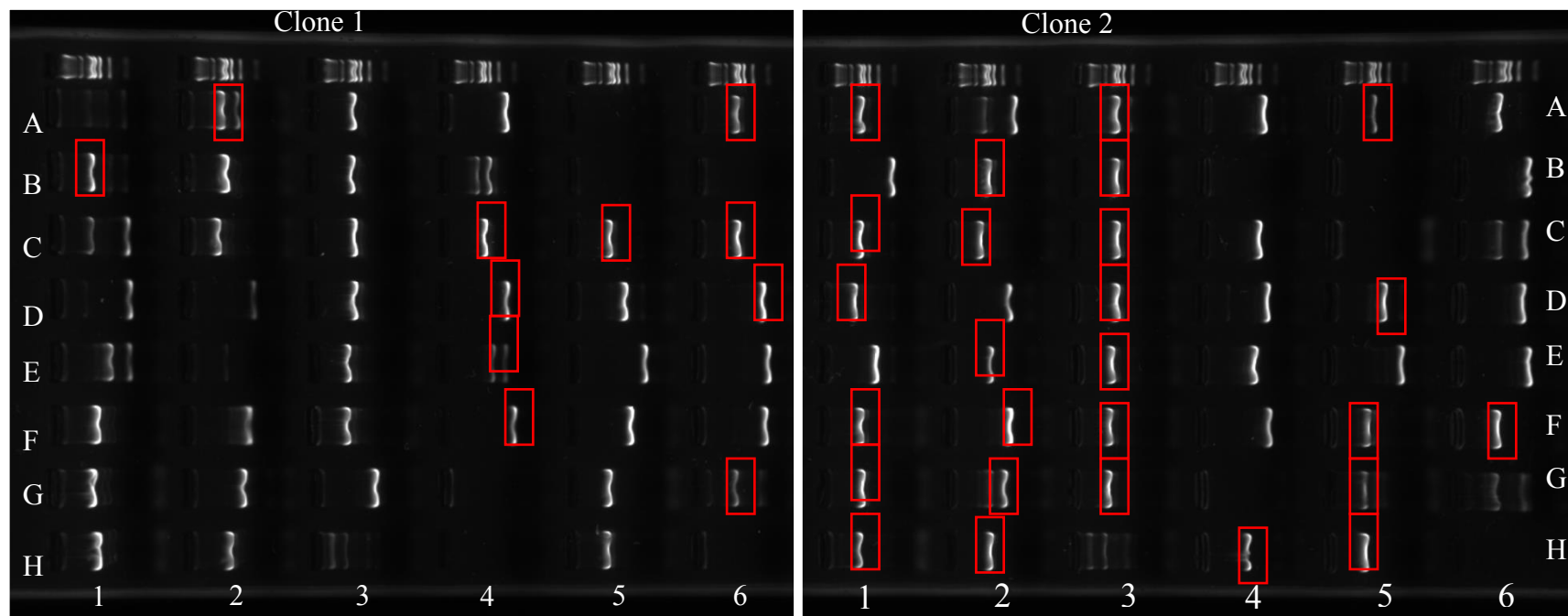


Figure 3.12 PCR verification of 47 HTP over-expression plasmids from 2 clones selected from white colonies from transformation. Successful clones are highlighted by red boxes. DevA native (A1-E1), DevE native (F1-H1, A2-C2), STRCO putative (D2-F2), HutC native (G2, H2, A3-C3), PA14_34660 (D3-H3, A4, B4), *P. fluorescens* (C4-F4), Gp26 (G4, H4, A5, B5), DevA SERp (C5-E5), DevE SERp (F5-H5, A6), HutC SERp (B6-G6). The various truncations and PCR product sizes are detailed previously in **Table 3.9**.

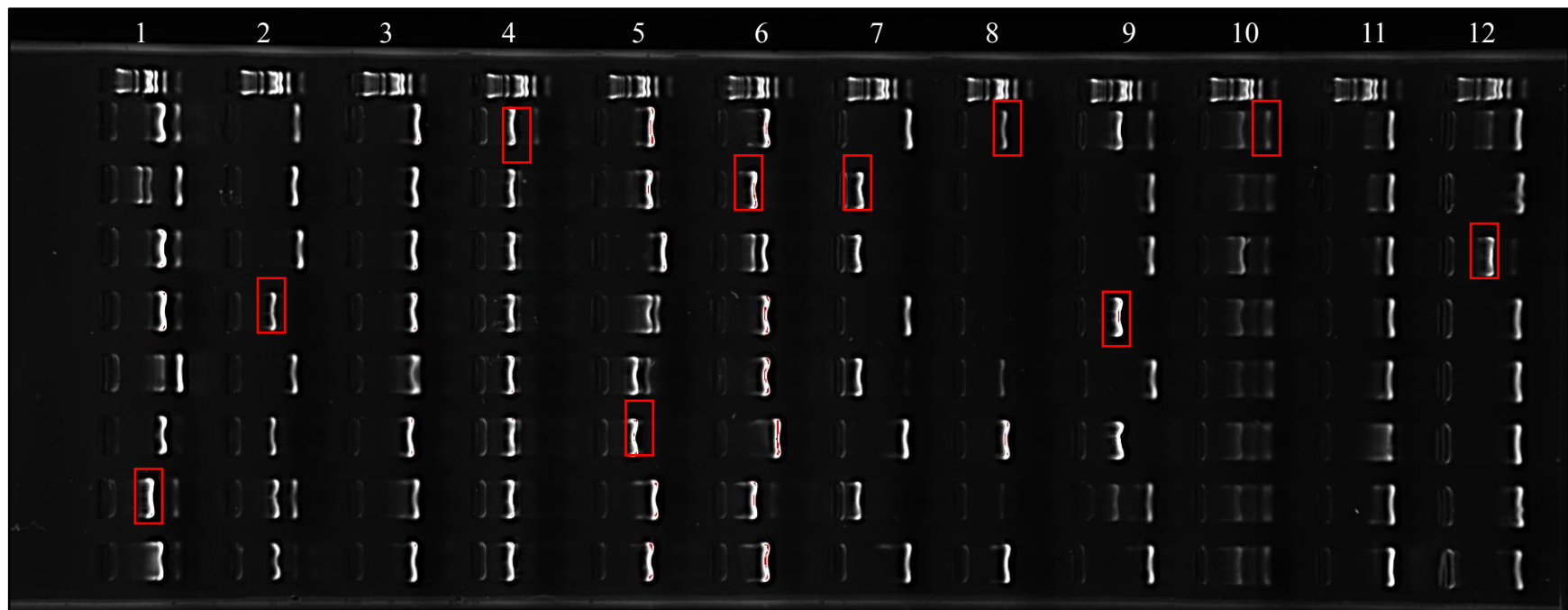


Figure 3.13 PCR verification of 12 GntR plasmids (8 clones). Plasmids were isolated using Promega Wizard SV96 plasmid kit on a Bio-Robot 8000 (Qiagen). Red boxes designate successful verification. 1) DevA Native E1; 2) STRCO Putative D2; 3) STRCO Putative F2; 4) PA14_34660 H3; 5) PA14_34660 A4; 6) PA14_34660 B4; 7) Gp26 G4; 8) Gp26 B5; 9) DevA SERp E5; 10) Hut SERp B6; 11) Hut SERp D6; 12) Hut SERp E6

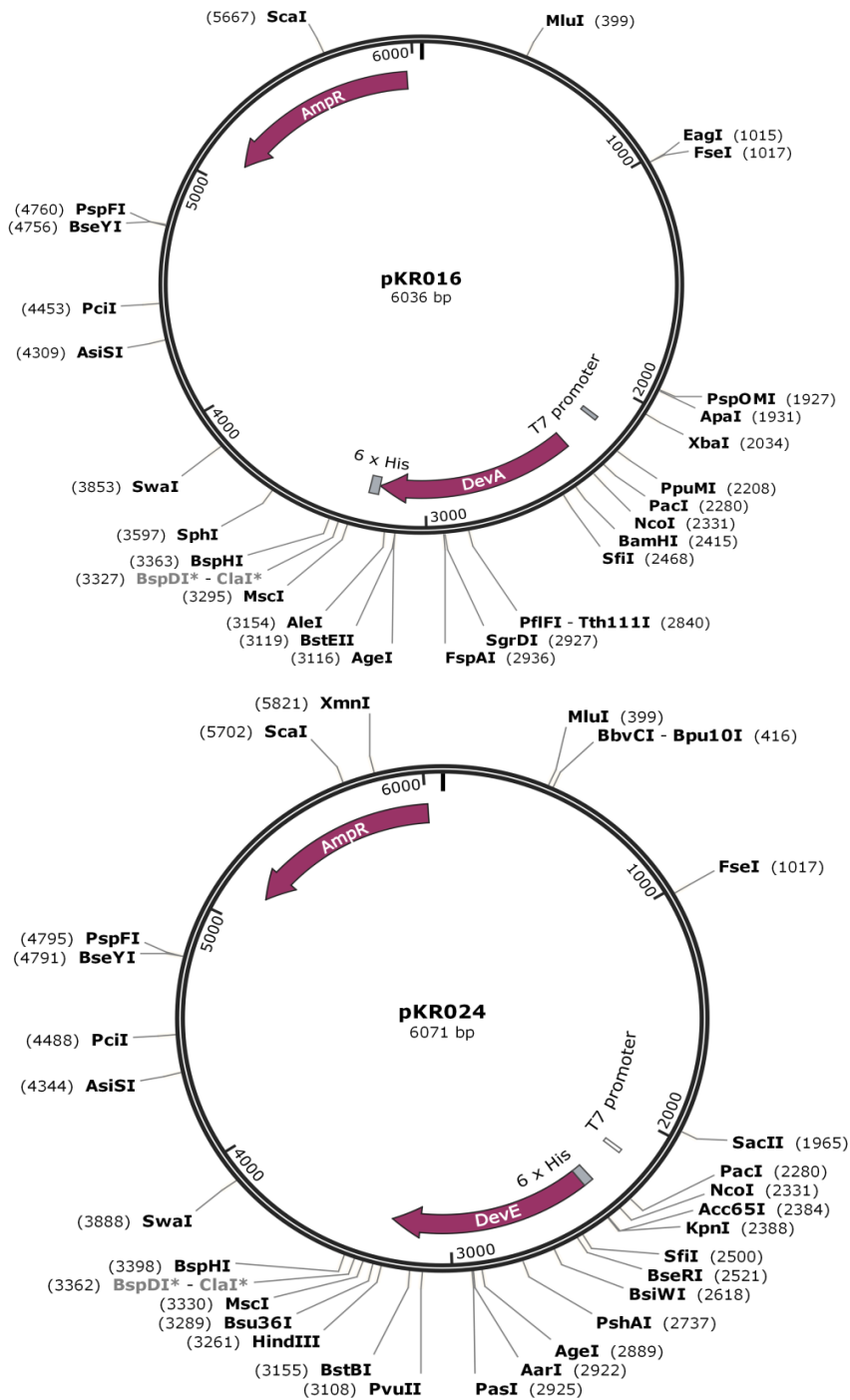


Figure 3.14 Plasmids that were used for large scale protein production were produced using In-Fusion cloning technology. pKR016 resulted from pOPINE containing DevA 2-291 AA gene insert with a C-terminal hexa-histidine tag. pKR024 resulted from pOPINF containing DevE 21-303 AA gene insert with an N-terminal hex-histidine tag (OPPF 16294). The maps were constructed *in silico* with Snapgene software.

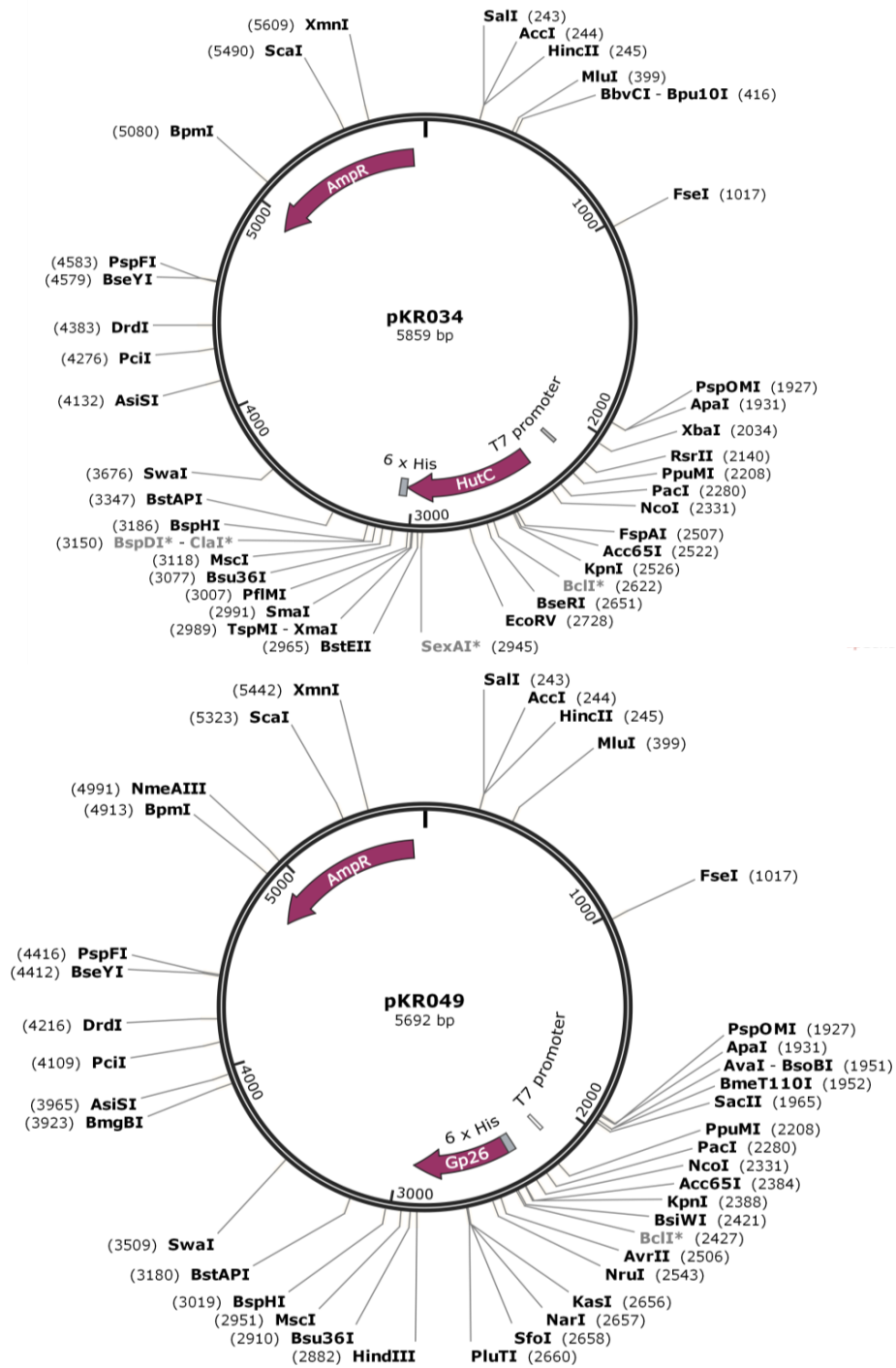


Figure 3.15 Plasmids that were used for large scale protein production were produced using In-Fusion cloning technology. pKR034 resulted from pOPINE containing HutC 20-250 AA gene insert. pKR049 resulted from pOPINE containing Gp26 21-174 AA gene insert. The maps were constructed *in silico* with Snapgene software.

3.3.3 Cloning of potential upstream promoter regions

GntR proteins are known to be generally auto-regulatory (Hoskisson & Rigali, 2009, Hoskisson *et al.*, 2006, Rigali *et al.*, 2004, Rigali *et al.*, 2002, Klaffl *et al.*, 2013). The majority appear to be negative auto-regulators although there are an increasing number that are known to be activators. For example, FadR is known to negatively regulate catabolic fatty acid genes as well as positively regulating anabolic fatty acid genes depending on whether the effector molecule, acyl-coA is bound to FadR or not (DiRusso *et al.*, 1993). Other examples of negative/positive regulation within the GntR family include NorG, a member of the FadR sub-family, from *Staphylococcus aureus* which represses cell wall autolysis and activates drug efflux proteins (Truong-Bolduc & Hooper, 2007). Great interest lies in this area of transcriptional regulation. The great number of GntR regulators in Pfam also indicates there may be potential for many more positive regulators to be discovered within the GntR superfamily.

As a starting point to look for potential promoter regions the 300 base pair upstream region of target GntR genes were extracted from NCBI. Primers were designed using Genefisher software (Giegerich *et al.*, 1996) and are detailed in **Table 2.4**. PCR was carried out and products at the correct size (**Figure 3.16A & B**) were extracted from the gel and used for blunt end cloning into linearised pUC19 (*SmaI* cut). The pUC19 vector was used for cloning upstream promoter fragments as it allowed a PCR to be carried out using M13 primers which were labelled with Cy5 dye in order to eventually carry out fluorescent EMSAs (see **Chapter 5**).

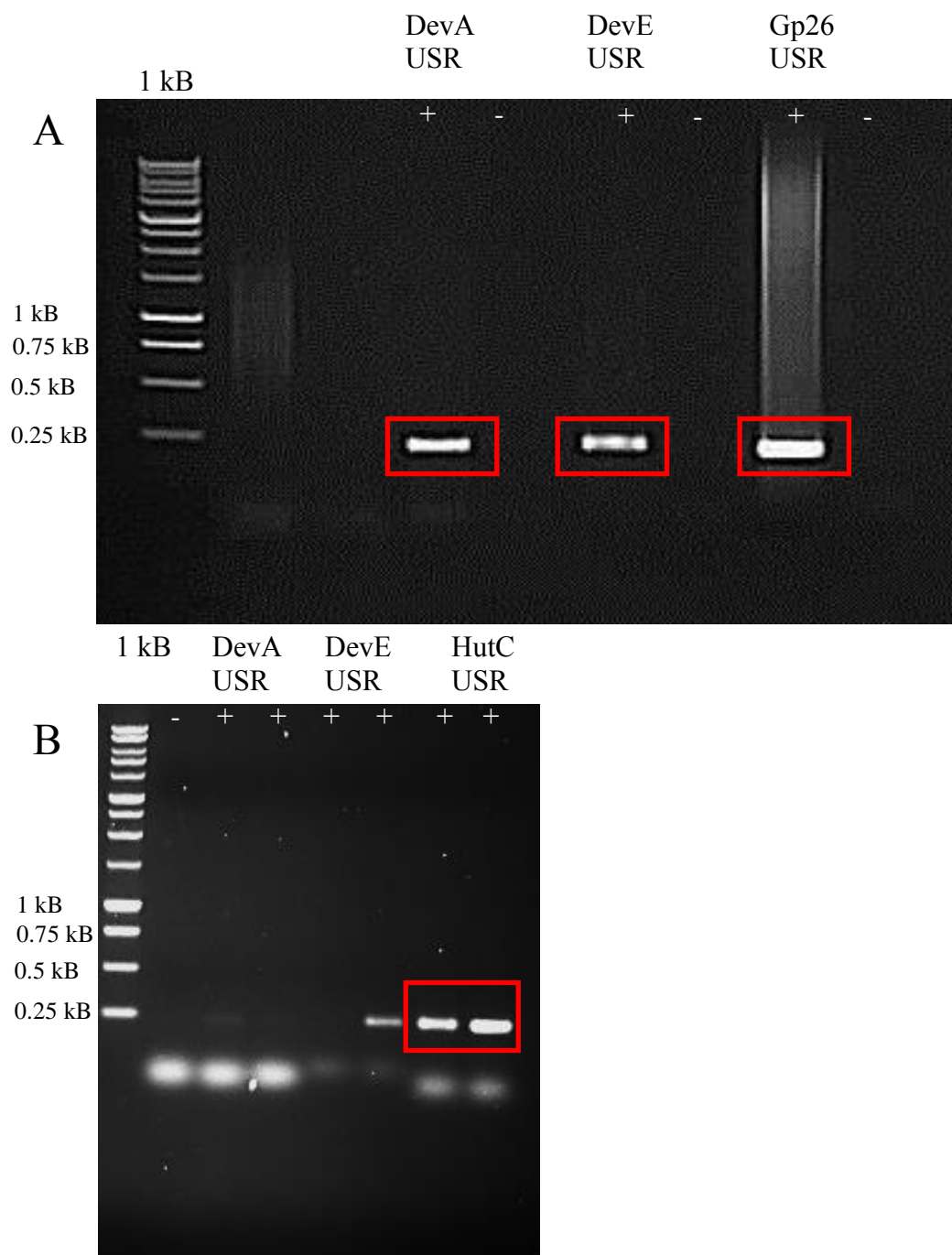


Figure 3.16(A) PCR of potential upstream promoter regions for DevA (233 bp), DevE (240bp) and Gp26 (223 bp). **(B)** PCR of upstream promoter region of HutC (237 bp) (+) denotes lanes containing DNA, (-) denotes negative control. **Marker:** Promega 1Kb; Gel: 0.8% agarose, 80 V.

Following the cloning reaction, the resulting plasmids were designated pKR063 (DevA USR [UpStream Region]), pKR064 (DevE USR), pKR065 (HutC USR) and pKR066 (Gp26 USR). Plasmids were subjected to restriction digests with *NdeI* and *PvuI* to confirm the DNA fragment had inserted in to the vector (**Figure 3.17**). Band sizes are shown in **Table 3.9**. Directionality and sequence of the insert were confirmed by sequencing (MWG Eurofins, London). Resulting plasmid maps were assembled *in silico* and are available in **Figure 3.18** and **3.19**.

Summary

Bioinformatics analyses were conducted with a view to selecting GntR cloning targets and for subsequent downstream characterisation of these targets. Secondary structure prediction from the sequences revealed that four of the 30 selected targets showed novel C-terminal secondary structure topology and one of these (Gp26) may represent a new sub-family.

Several GntR cloning targets were subject to sequence optimisation in order to obtain the maximum amount of protein that was most likely to crystallise. Sequence optimisation strategies included surface entropy reduction mutations, codon optimisation and truncations of disordered regions within the proteins whilst maintaining the overall secondary structure.

As a result, a number of overexpression constructs were created during the course of this project, including 47 from high throughput methods. By creating a number of constructs it was hoped that some of these would yield sufficient amounts of protein to move forward into structural studies. After successful cloning of constructs,

expression testing was carried out on all constructs followed by optimisation of protein purification procedures.

The upstream regions of HutC, DevA, DevE and Gp26 genes were also successfully cloned into pUC19 for further functional analysis of protein-DNA interactions by fluorescent EMSAs.

Table 3.9 Band sizes of restriction digest products for pUC19 backbone plasmids

pKR063	pKR064	pKR065	pKR066
<i>Bam</i> HI, <i>Pvu</i> I			
1652	1652	1652	1652
896	896	896	896
372	378	359	361
		16	

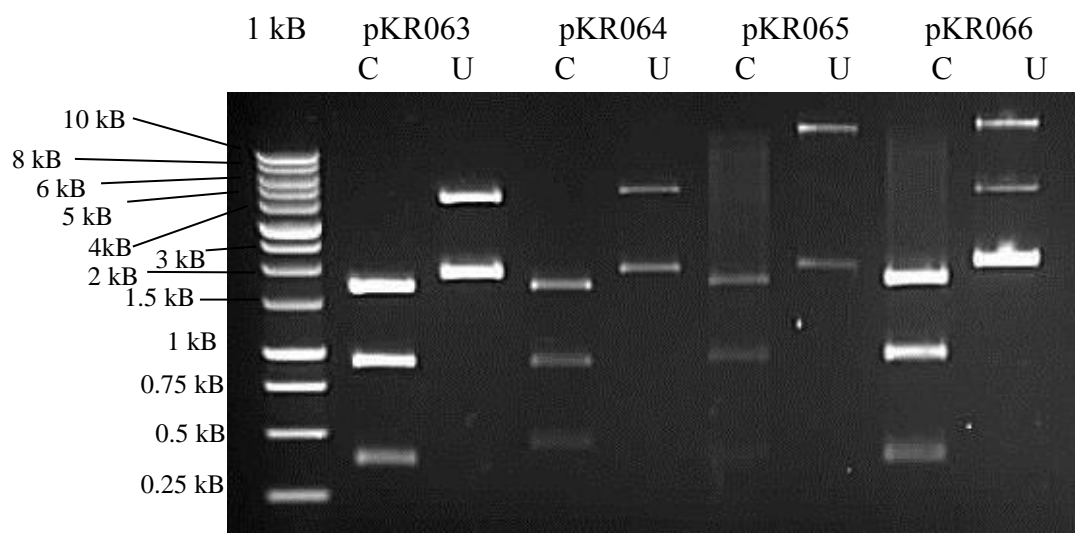


Figure 3.17 Restriction digest (*Bam*HI, *Pvu*I) of pKR063, pKR064, pKR065, pKR066 to confirm cloning into pUC19 vector. See **Table 3.9** for band sizes. (C) denotes cut plasmid DNA, (U) denotes uncut plasmid DNA. **Marker:** Promega 1 kb; Gel: 0.8% agarose, 80V.

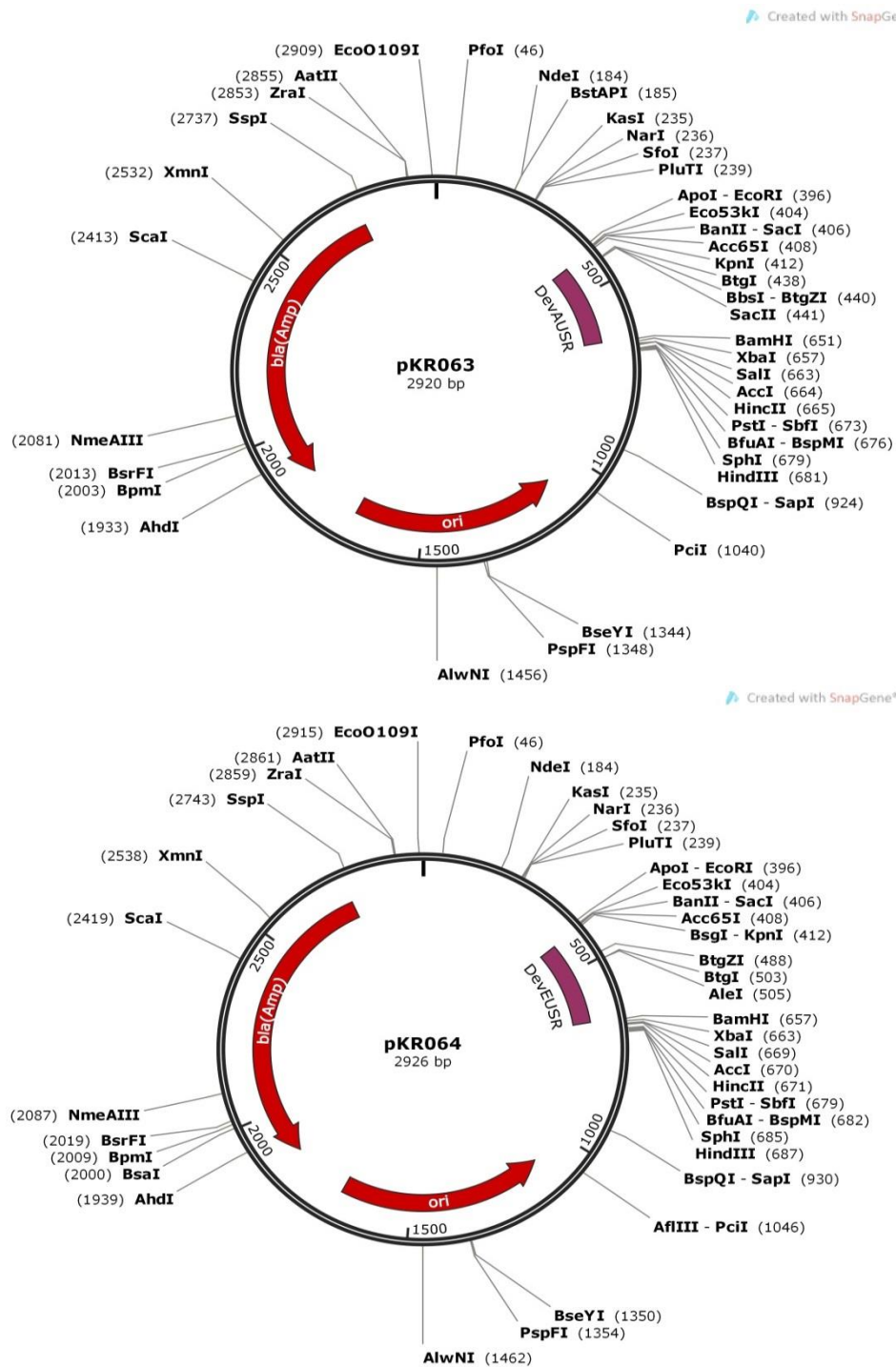


Figure 3.18 Plasmid maps of pKR063 and pKR064, resulting from cloning DevAUSR (233 bp) and DevEUSR (240 bp) fragments, respectively, into the pUC19 vector via blunt end cloning. The maps were constructed *in silico* with Snapgene software.

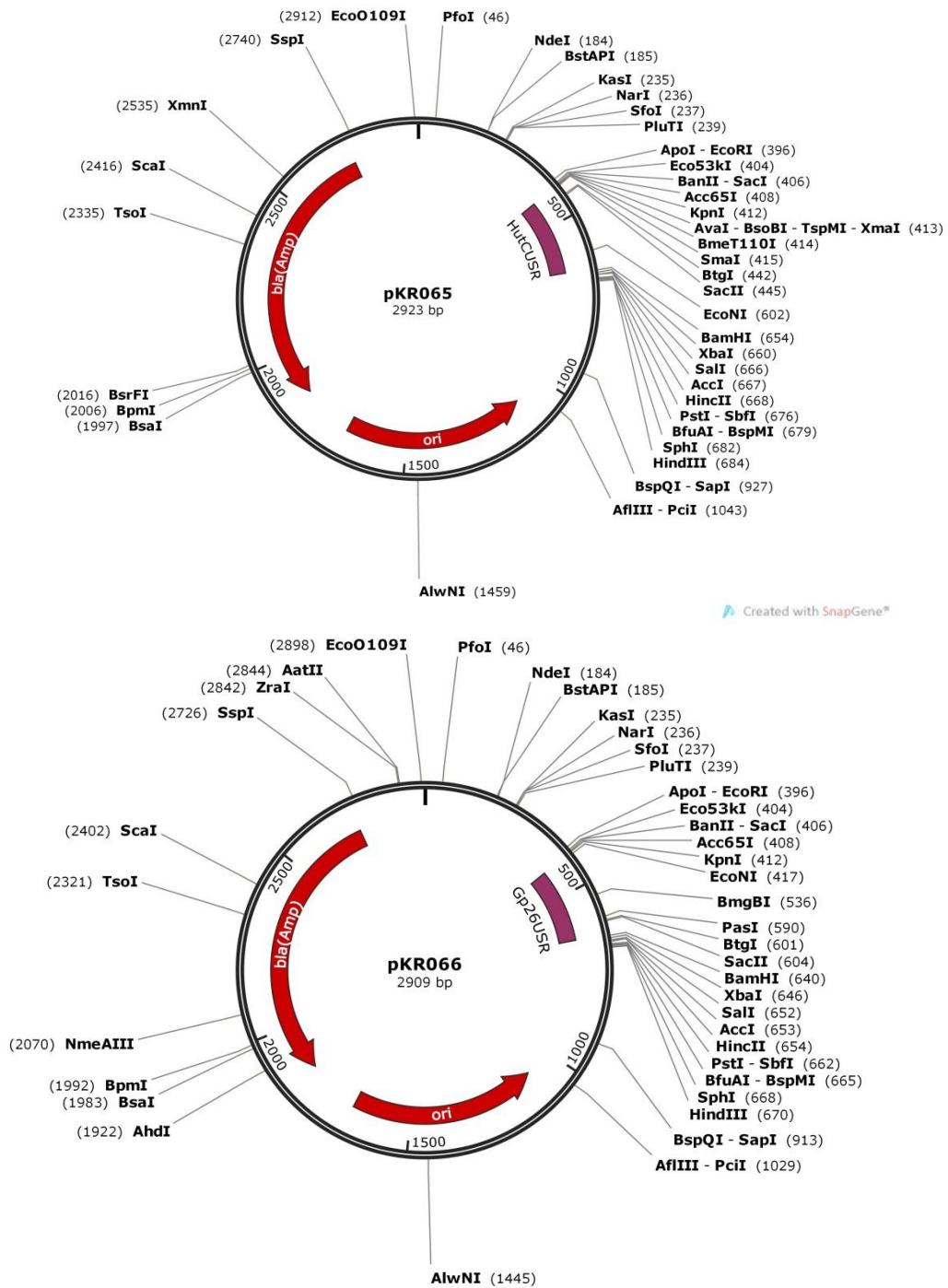


Figure 3.19 Plasmid maps of pKR065 and pKR066, resulting from cloning HutCUSR (237 bp) and Gp26USR (233 bp) fragments, respectively, into the pUC19 vector via blunt end cloning. The maps were constructed *in silico* with Snapgene software.

3.4 Protein overexpression and optimisation of purification

Optimisation of protein production and purification procedures was an important development step for this project, given the ultimate aim was to crystallise target GntR regulators. To achieve this, a large amount of high quality pure protein is required. It is well known in terms of crystallisation that protein production is the bottle neck (Derewenda, 2004a, Derewenda, 2004b, Goldschmidt *et al.*, 2007).

Initial protein overexpression was tested in small scale cultures using both IPTG induction and auto induction media. Cultures were tested at different induction temperatures to determine the best conditions for optimum protein yield.

Upon the establishment of optimal conditions cultures were scaled-up to 1 litre. When larger quantities of protein were required up to 6 litre cultures were used although still in individual 1L volumes.

3.4.1 Overexpression testing of GntR proteins from pET100 vectors

The four GntR proteins which are the main focus of this project (DevA, DevE, HutC and Gp26) were tested in two expression strains, *E. coli* BL21 (DE3) and *E. coli* Rosetta (DE3) and both rely on T7 promoter induction of expression. The BL21 strain is a good general purpose overexpression strain to start with. It is deficient in Lon and OmpT proteases which prevent over expressed protein from being degraded by the cells. *E. coli* Rosetta (DE3), while also being deficient in Lon and OmpT proteases, carries the pRARE plasmid which encodes 6 tRNAs for rare codons (AGG, AGA, AUA, CUA, CCC, CGA) and is therefore useful for expressing non-native *E. coli* proteins. Two different induction methods were also tested, LB with IPTG induction at 20°C and 37°C and auto induction at 20°C and 25°C.

IPTG induction offers the advantage of tuneable protein overexpression but relies on the OD₆₀₀ of cells being checked often to determine the optimum time for induction i.e. during the log phase of growth, which may be different for different strains and those expressing different proteins. In contrast, auto-induction can be carried out without too much monitoring of the cell density and relies on lactose contained within the media to act as the inducer.

When the T7 promoter has been induced and protein expression is induced, the temperature of the culture was reduced to either 20°C (IPTG induction and auto-induction) or 25°C (auto-induction). The lower temperature allows protein folding to be more precise and prevents or reduces the formation of inclusion bodies; therefore a better quality protein should be produced. For both induction systems, temperature reduction and addition of IPTG was at cell densities between OD₆₀₀ between 0.3 and 0.5.

Despite the variety of conditions tested, all proteins had very poor soluble expression. Overexpression was evidenced by SDS-PAGE for all proteins, in either the total extract or soluble fraction, except Gp26 which didn't have detectable expression at all in any condition. HutC was found to have no detectable expression in *E. coli* Rosetta with auto-induction. The proteins were identified as being mainly present as inclusion bodies within the total cell extract. Overexpression bands are indicated by red underlines in **Figure 3.20 -3.24**. Protein expression was assessed in terms of soluble protein and is detailed in **Table 3.10**.

Protein production was scaled up to 1 litre cultures which allowed enough soluble protein to be obtained for EMSAs from pET100 backbone plasmids, including Gp26;

however crystallisation required much greater amounts of protein which led to the design of multiple constructs at the OPPF detailed in **section 3.3.2**.

Table 3.10 Expression test results for pET100 / pOPINF backbone plasmids

	Protein	IPTG 20°C	IPTG 37°C	Auto-induction 20°C	Auto-induction 25°C
<i>E. coli</i> Rosetta	DevA	+	-	-	+
	DevE	-	-	-	-
	HutC	+	-	-	-
	Gp26	-	-	-	-
	DevAS	-	-	+	+
	DevES	+	-	-	+
	HutCS	-	-	-	-
<i>E. coli</i> BL21	DevAS	-	+	-	+
	DevES	+	-	++	+
	HutCS	-	-	-	+
No detectable expression (-), Low expression (+), Good Expression (++), Very good expression (+++)					

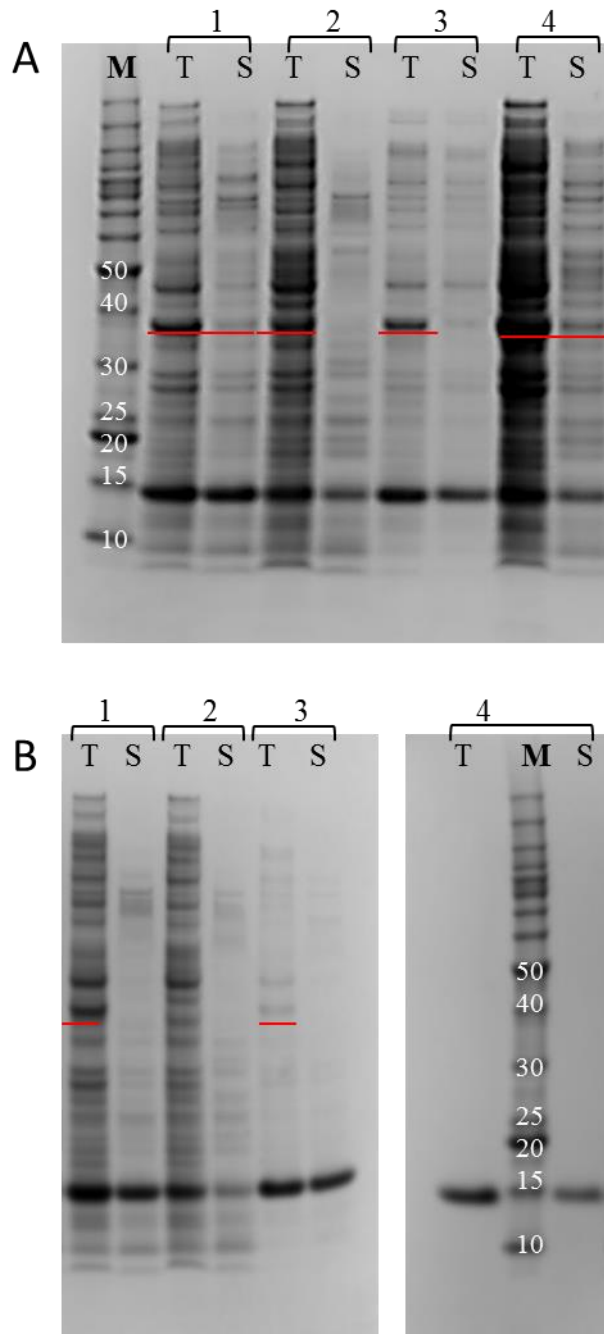


Figure 3.20 (A) Overexpression testing of pKR003 (DevA) and (B) pKR008 (DevE). Constructs were tested in *E. coli* Rosetta using different induction methods (1) IPTG 20°C (2) AIM 20°C (3) IPTG 37°C (4) AIM 25°C. (T) designates total cell extract; (S) designates the soluble fraction. Overexpression bands are underlined in red. Gel - 4-12% NuPage (R) Bis-Tris, MES buffer, 200V. **Marker** - Benchmark™ Protein ladder; molecular weights shown are in kDa.

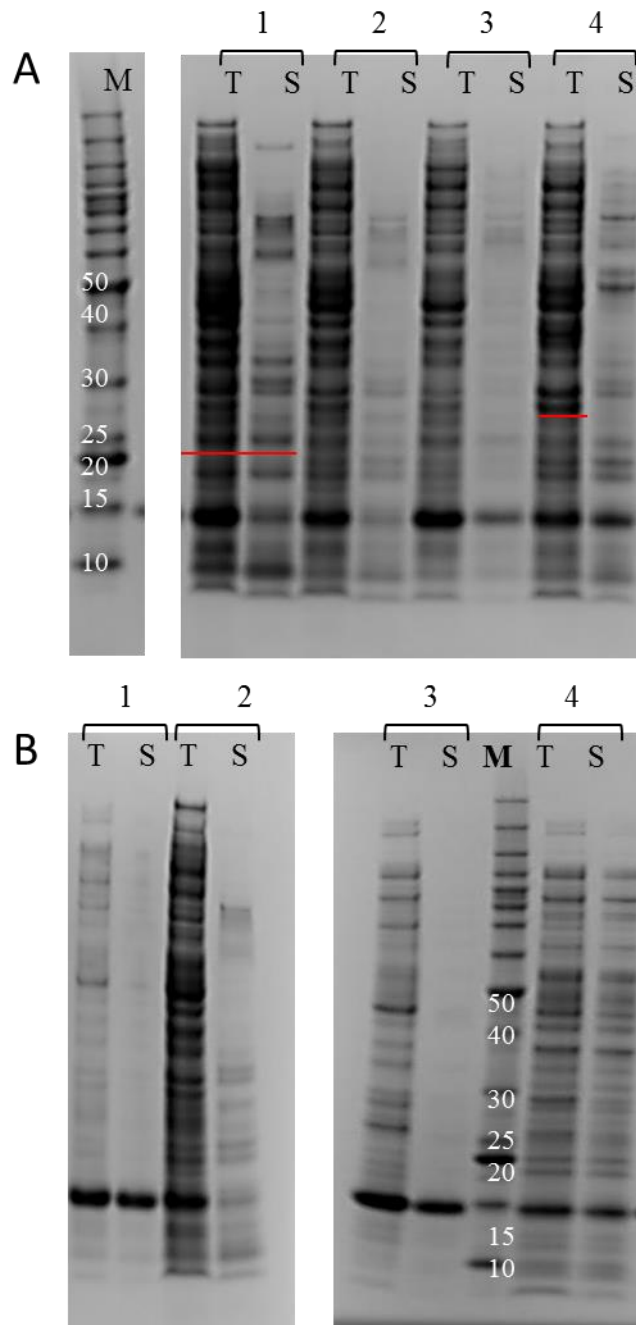


Figure 3.21 (A) Overexpression testing of pKR006 (HutC) and (B) pKR007 (Gp26). Constructs were tested in *E. coli* Rosetta using different induction methods (1) IPTG 20°C (2) AIM 20°C (3) IPTG 37°C (4) AIM 25°C. (T) designates total cell extract; (S) designates the soluble fraction. Overexpression bands are underlined in red. Gel - 4-12% NuPage (R) Bis-Tris, MES buffer, 200V. **Marker** - BenchmarkTM Protein ladder; molecular weights shown are in kDa.

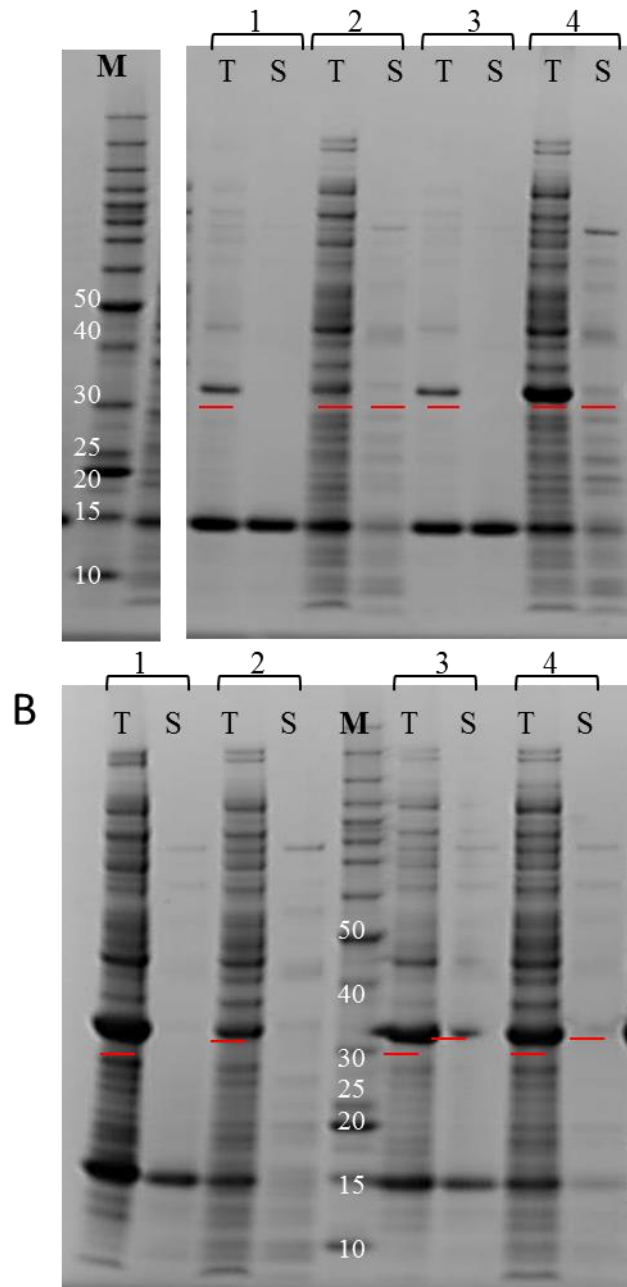


Figure 3.22 (A) Overexpression testing of pKR013 (DevAS) in *E. coli* Rosetta and (B) *E. coli* BL21. Constructs were tested using different induction methods (1) IPTG 20°C (2) AIM 20°C (3) IPTG 37°C (4) AIM 25°C. (T) designates total cell extract; (S) designates the soluble fraction. Overexpression bands are underlined in red. Gel - 4-12% NuPage (R) Bis-Tris, MES buffer, 200V. **Marker** - BenchmarkTM Protein ladder; molecular weights shown are in kDa.

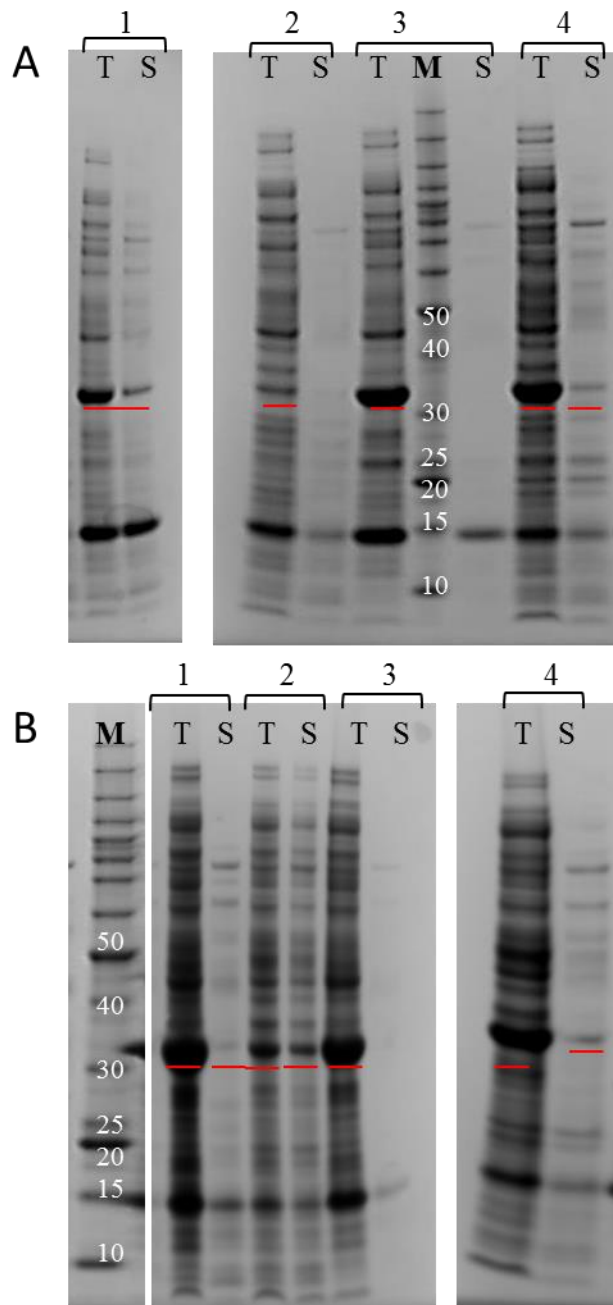


Figure 3.23 (A) Overexpression testing of pKR014 (DevES) in *E. coli* Rosetta and (B) *E. coli* BL21. Constructs were tested using different induction methods (1) IPTG 20°C (2) AIM 20°C (3) IPTG 37°C (4) AIM 25°C. (T) designates total cell extract; (S) designates the soluble fraction. Overexpression bands are underlined in red. Gel - 4-12% NuPage (R) Bis-Tris, MES buffer, 200V. **Marker** - BenchmarkTM Protein ladder; molecular weights shown are in kDa.

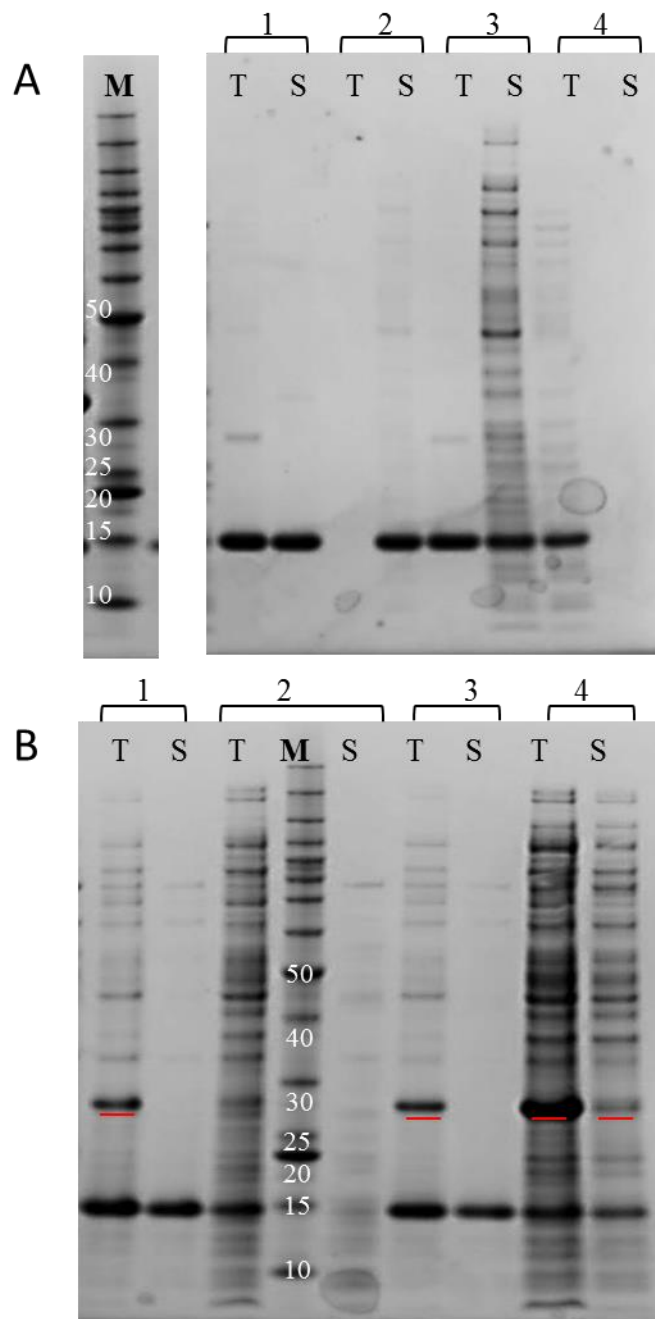


Figure 3.24 (A) Overexpression testing of pKR015 (HutCS) in *E. coli* Rosetta and (B) *E. coli* BL21. Constructs were tested using different induction methods (1) IPTG 20°C (2) AIM 20°C (3) IPTG 37°C (4) AIM 25°C. (T) designates total cell extract; (S) designates the soluble fraction. Overexpression bands are underlined in red. Gel - 4-12% NuPage (R) Bis-Tris, MES buffer, 200V. **Marker** - BenchmarkTM Protein ladder; molecular weights shown are in kDa.

3.4.3 Overexpression testing of GntR proteins from HTP constructs

Overexpression conditions were tested with two strains, *E. coli* Lemo21 (DE3) and *E. coli* Rosetta2 (DE3), using two different induction methods, IPTG induction and auto-induction. Different strains were used as they both offer different advantages. *E. coli* Rosetta2 carries the pRARE2 plasmid which encodes seven rare tRNAs (CGG, in addition to those encoded on the original pRARE plasmid) as mentioned previously. *E. coli* Lemo21 is particularly useful during overexpression of proteins with solubility issues as it was originally designed for overexpression of membrane proteins which are inherently insoluble (Wagner *et al.*, 2008). This strain carries the pLemo plasmid that can be induced by rhamnose to express T7 lysozyme, the natural inhibitor of T7 RNA polymerase, therefore there can be more control over the control of the T7 promoter and optimal protein can be expressed. Although, the addition of rhamnose was not tested in this case, this strain appeared to be particularly useful in expressing proteins from *S. coelicolor* perhaps due to the high GC content of the *S. coelicolor* genome (Bentley *et al.*, 2002).

Successful HTP plasmids were scored on their expression of soluble protein (detailed in **Table 3.11**). A full summary of all 47 constructs tested is detailed in **Appendix 2**. **Figures 3.25 - 3.28** show soluble protein expression bands from cultures which were induced with IPTG; **Figures 3.29 - 3.32** show soluble protein expression bands induced by auto-induction media. Successful overexpression bands are underlined in red. Due to the large number of constructs produced from HTP cloning, proteins in this section are referred to with their unique identifier number in brackets.

DevA (16279), DevE (16292) and HutC (16303) were noted to have particularly good expression profiles across all conditions and strains tested. DevA (16282), DevE (16295) and Gp26 (16312) were better expressed in Lemo21 than Rosetta2.

The four constructs which were taken forward to be scaled up were DevA (16279), DevE (16294), HutC (16303) and Gp26 (16314) are highlighted in red in **Table 3.11**. These were chosen as they all expressed well in similar conditions (Rosetta2, auto-induction media). From here on, all reference to DevA, DevE, HutC and Gp26 proteins refer to protein obtained from these constructs. None of these proteins were affected by truncations at the secondary structure level as detailed in **section 3.2**.

Table 3.11 Expression test results for successful HTP plasmids

Well	Gene name	ID No.	MW	Lemo21		Rosetta 2	
				IPTG	Auto	IPTG	Auto
A01	DEVA_native	16279	32890	++	++	++	++
D01	DEVA_native	16282	44890	++	++	+	+
F01	DEVE_native	16291	34210	-	+	+	++
G01	DEVE_native	16292	32120	++	++	++	++
H01	DEVE_native	16293	29370	-	-	++	+
A02	DEVE_native	16294	32120	+	+	+++	+++
C02	DEVE_native	16296	46210	++	++	-	+
E02	STRCO Putative	16323	22990	-	-	+	++
A03	HUT_native	16303	26290	++	++	++	++
B03	HUT_native	16304	28380	+	+	++	++
C03	HUT_native	16306	26400	-	+	+	+++
D03	HUT_native	16307	26290	-	+	++	+++
A04	PA14_34660 GntR gene	16320	36960	++	-	+	-
G04	Gp26	16312	20020	-	++	-	-
H04	Gp26	16313	17930	+	++	++	++
A05	Gp26	16314	19910	++	+	++	+
H06	GFP positive control	N/A	27000	+++	+++	+++	++

No detectable expression (-), Low expression (+), Good Expression (++), Very good expression (+++)

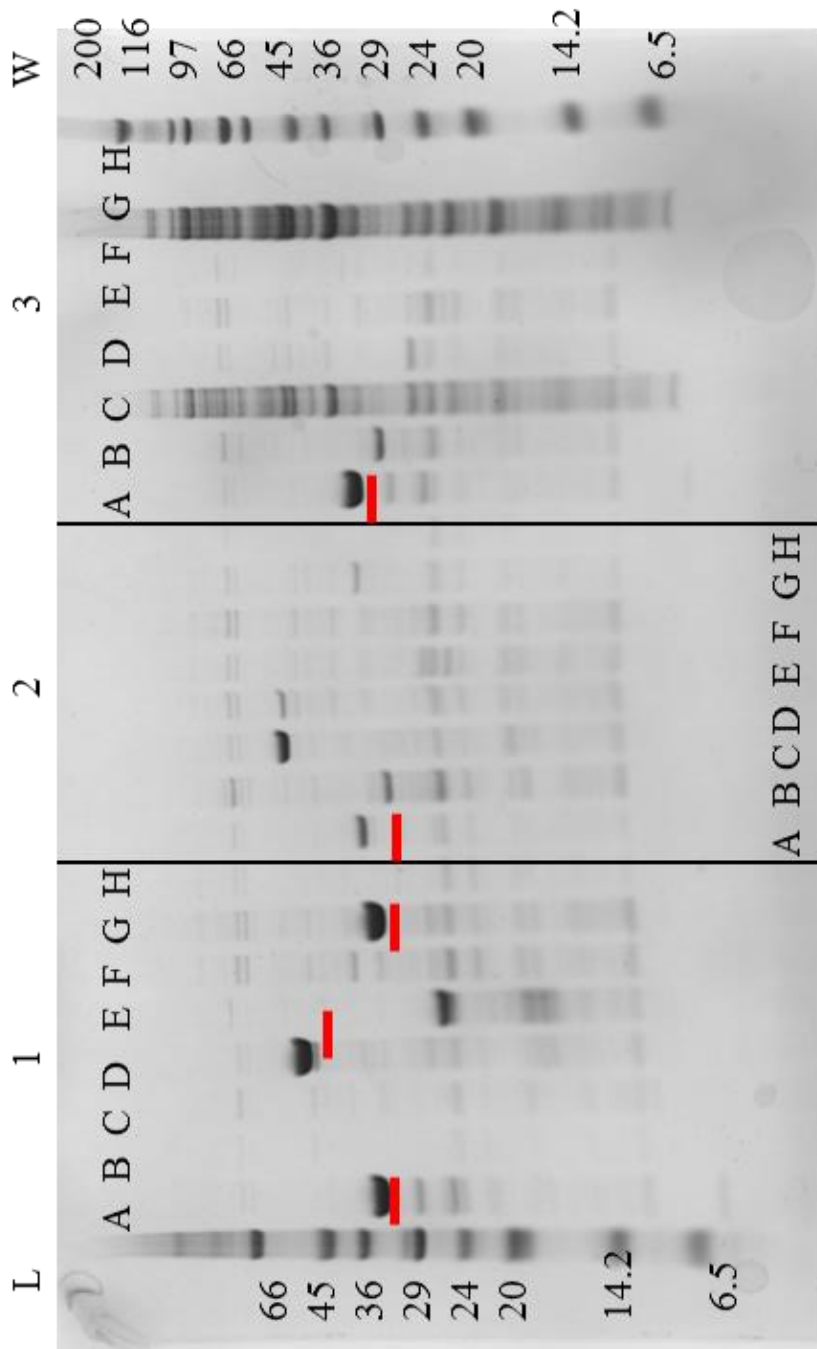


Figure 3.25 Overexpression test of HTP constructs in *E. coli* Lemo21 induced with 1 mM IPTG. Samples were purified by nickel affinity purification on BioRobot 8000. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa.

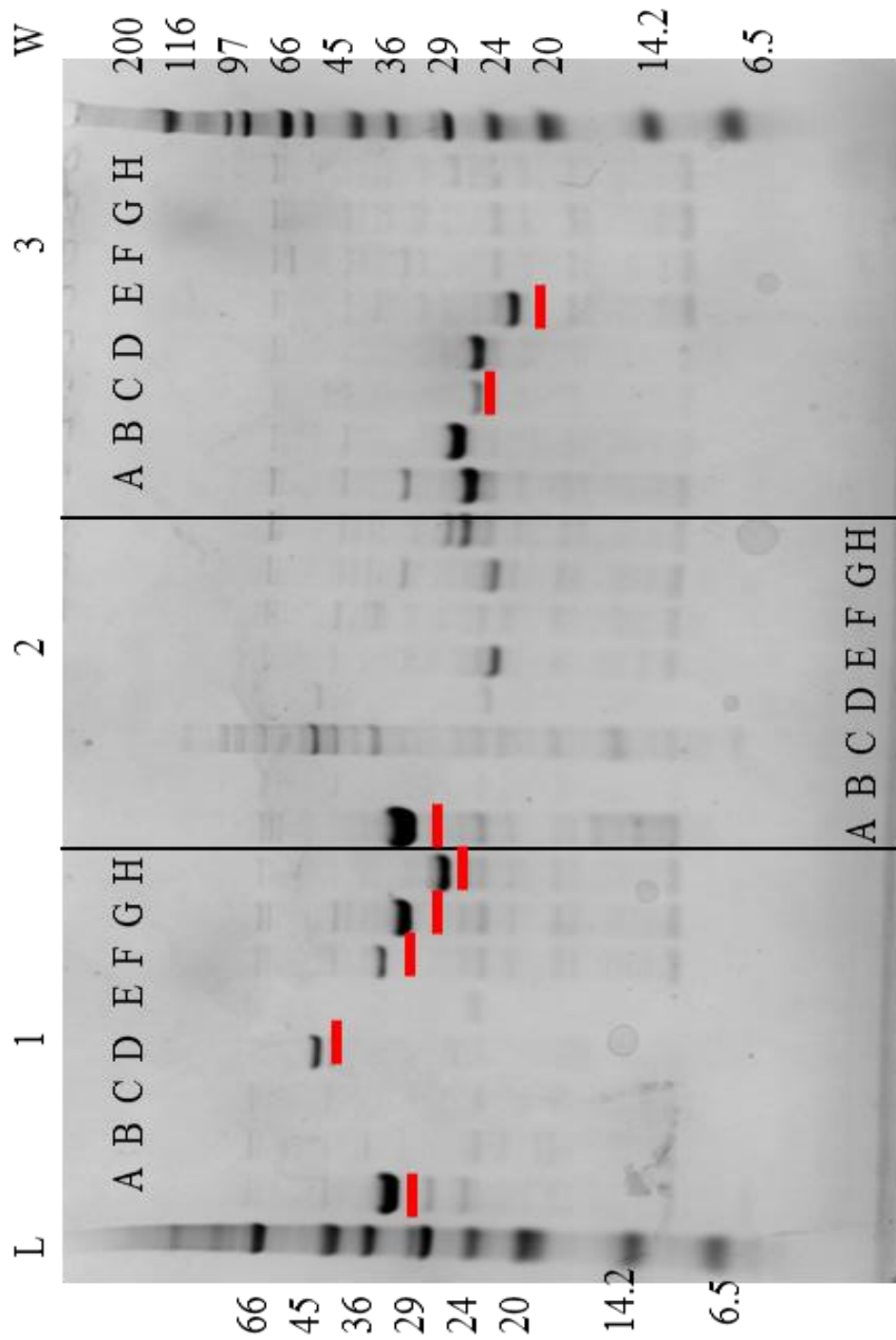


Figure 3.26 Overexpression test of HTP constructs in *E. coli* Rosetta2 induced with 1 mM IPTG. Samples were purified by nickel affinity purification on BioRobot 8000. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa.

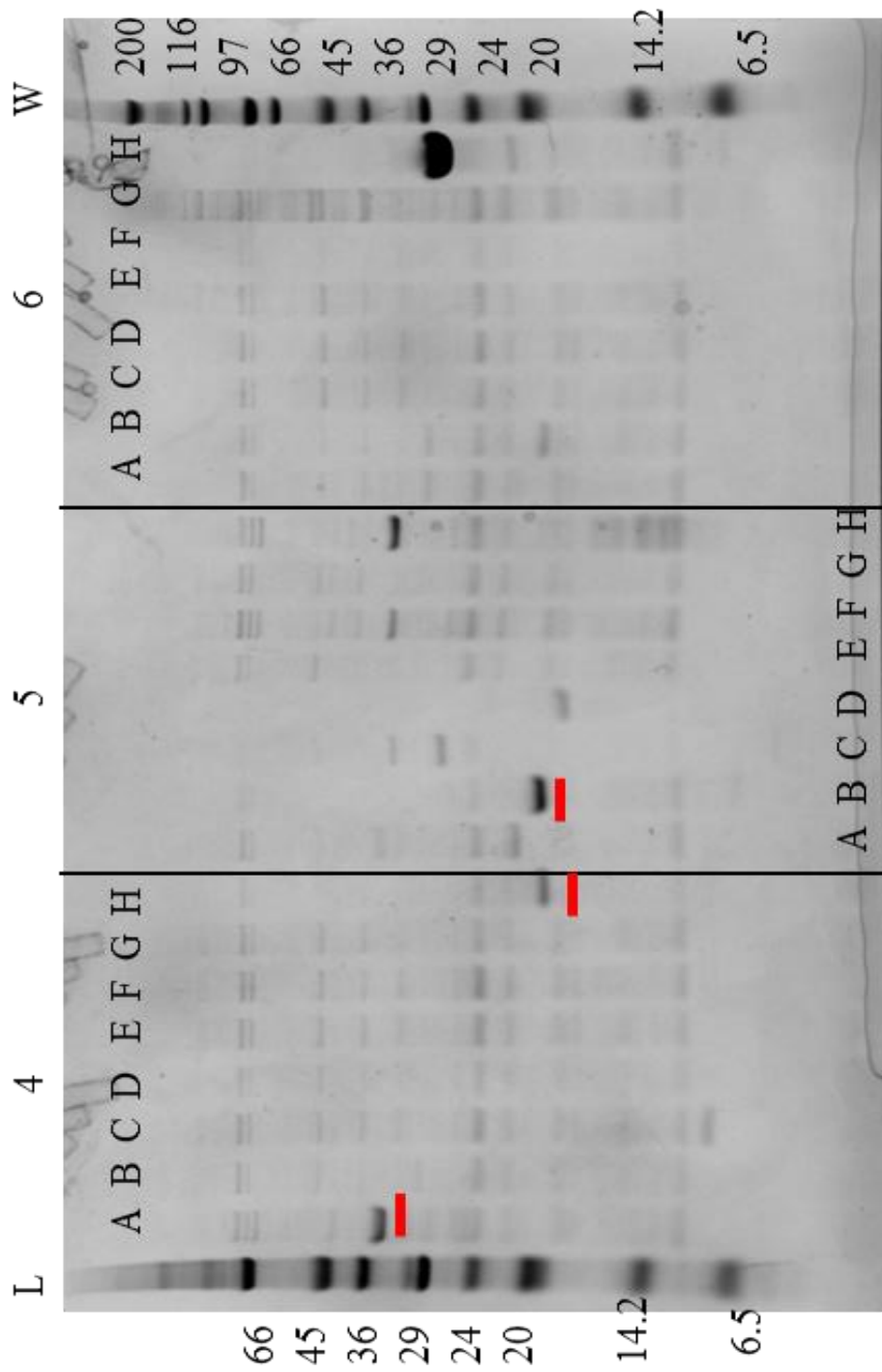


Figure 3.27 Overexpression test of HTP constructs in *E. coli* Lemo21 strains induced with 1 mM IPTG. Samples were purified by nickel affinity purification on BioRobot 8000. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa.

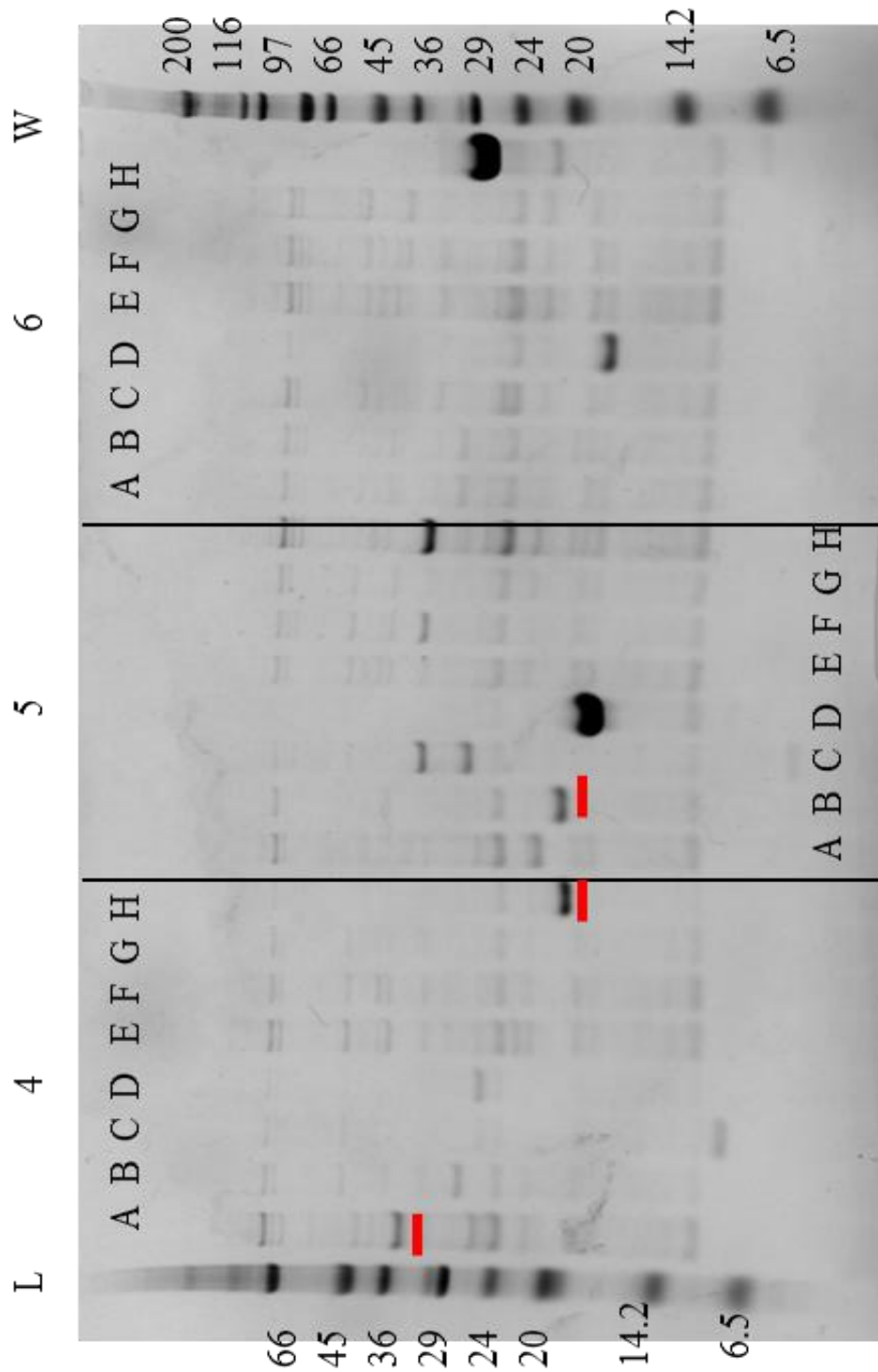


Figure 3.28 Overexpression test of HTP constructs in *E. coli* Rosetta2 induced with 1 mM IPTG. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa. Samples were purified by nickel affinity purification on BioRobot 8000.

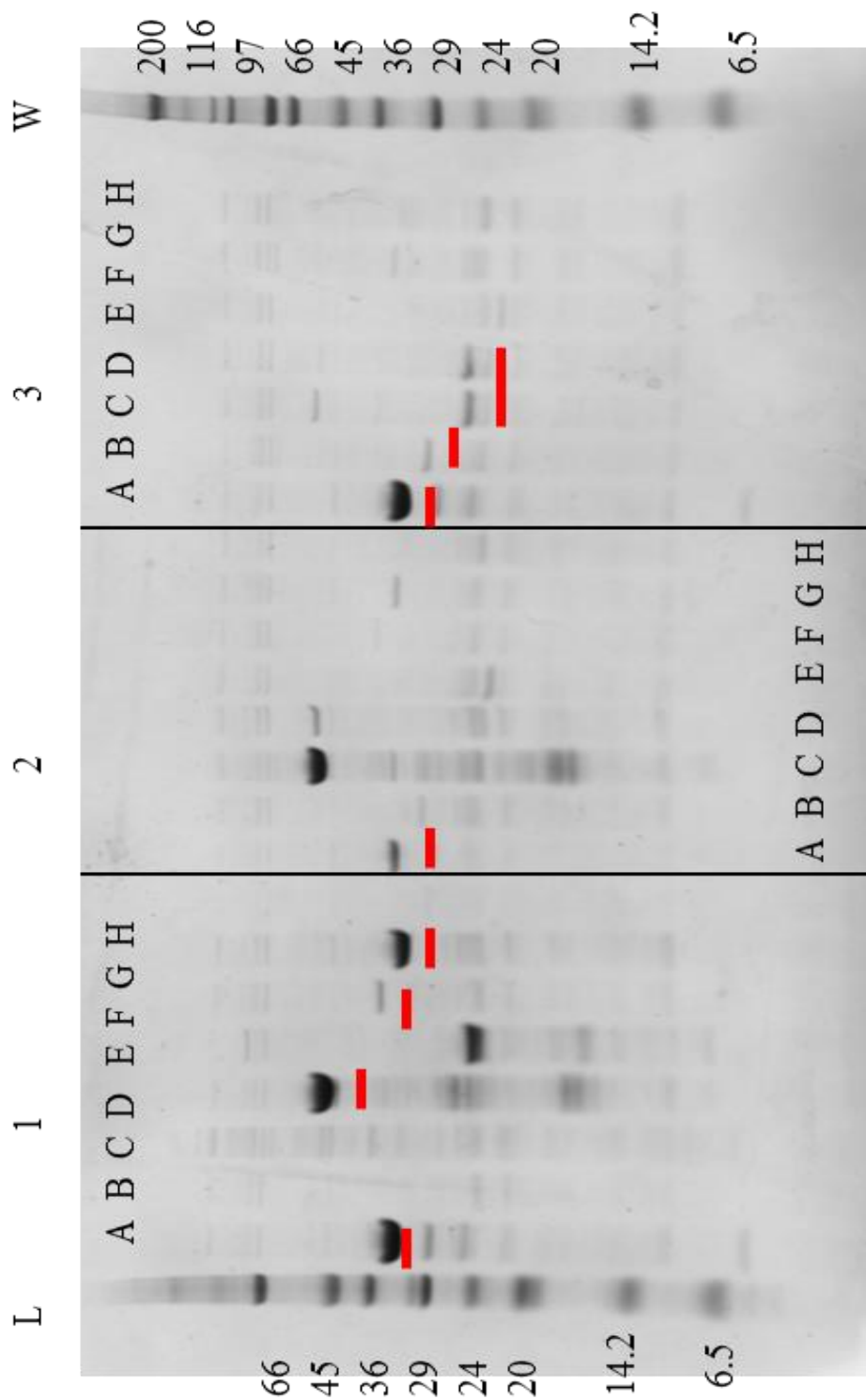


Figure 3.29 Overexpression test of HTP constructs in two *E. coli* Lemo21 induced by auto induction media. Samples were purified by nickel affinity purification on BioRobot 8000. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa.

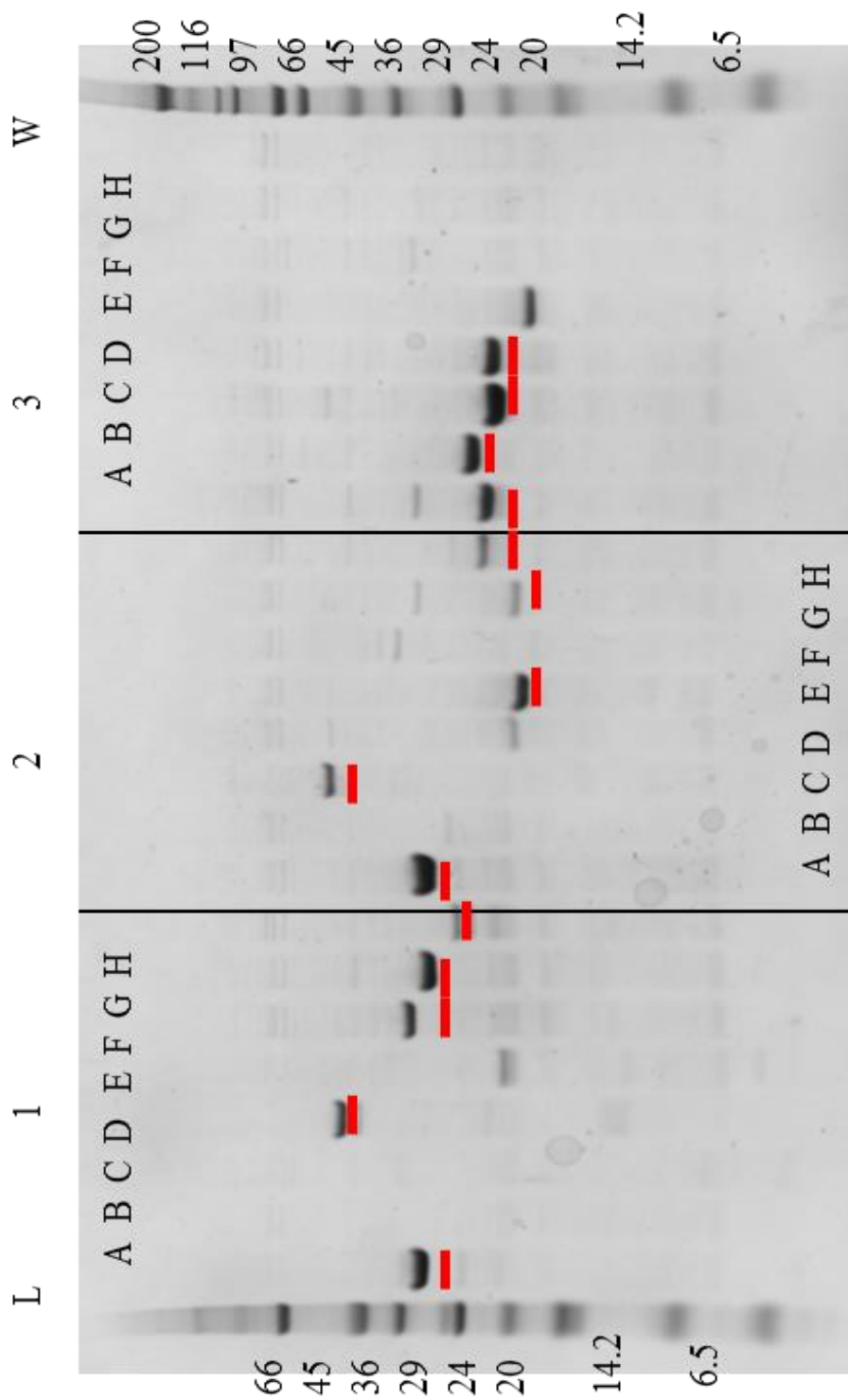


Figure 3.30 Overexpression test of HTP constructs in *E. coli* Rosetta2 induced with auto induction media. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa. Samples were purified by nickel affinity purification on BioRobot 8000.

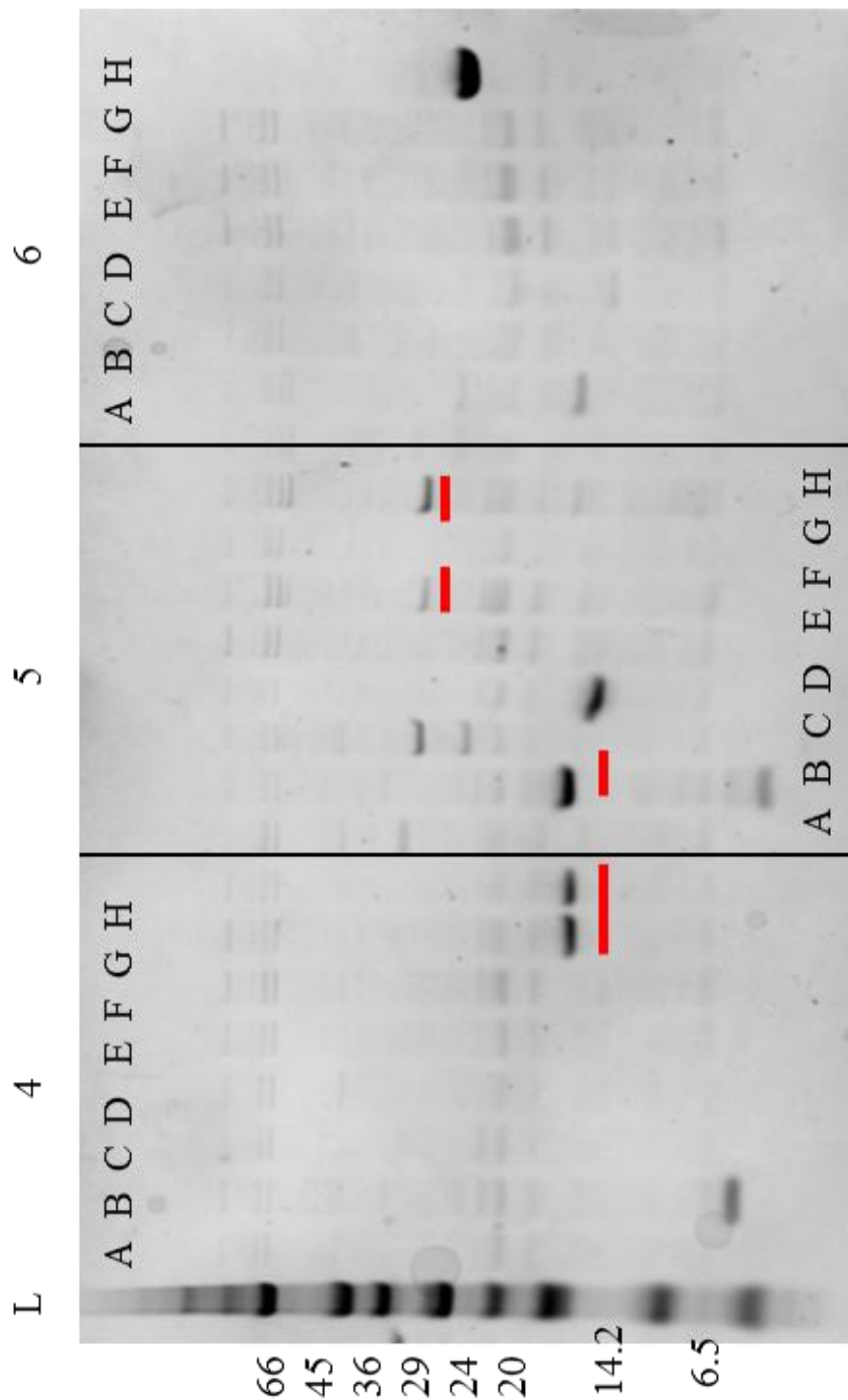


Figure 3.31 Overexpression test of HTP constructs in two *E.coli* Lemo21 strain induced by auto induction media. Samples were purified by nickel affinity purification on BioRobot 8000. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma), molecular weights in kDa.

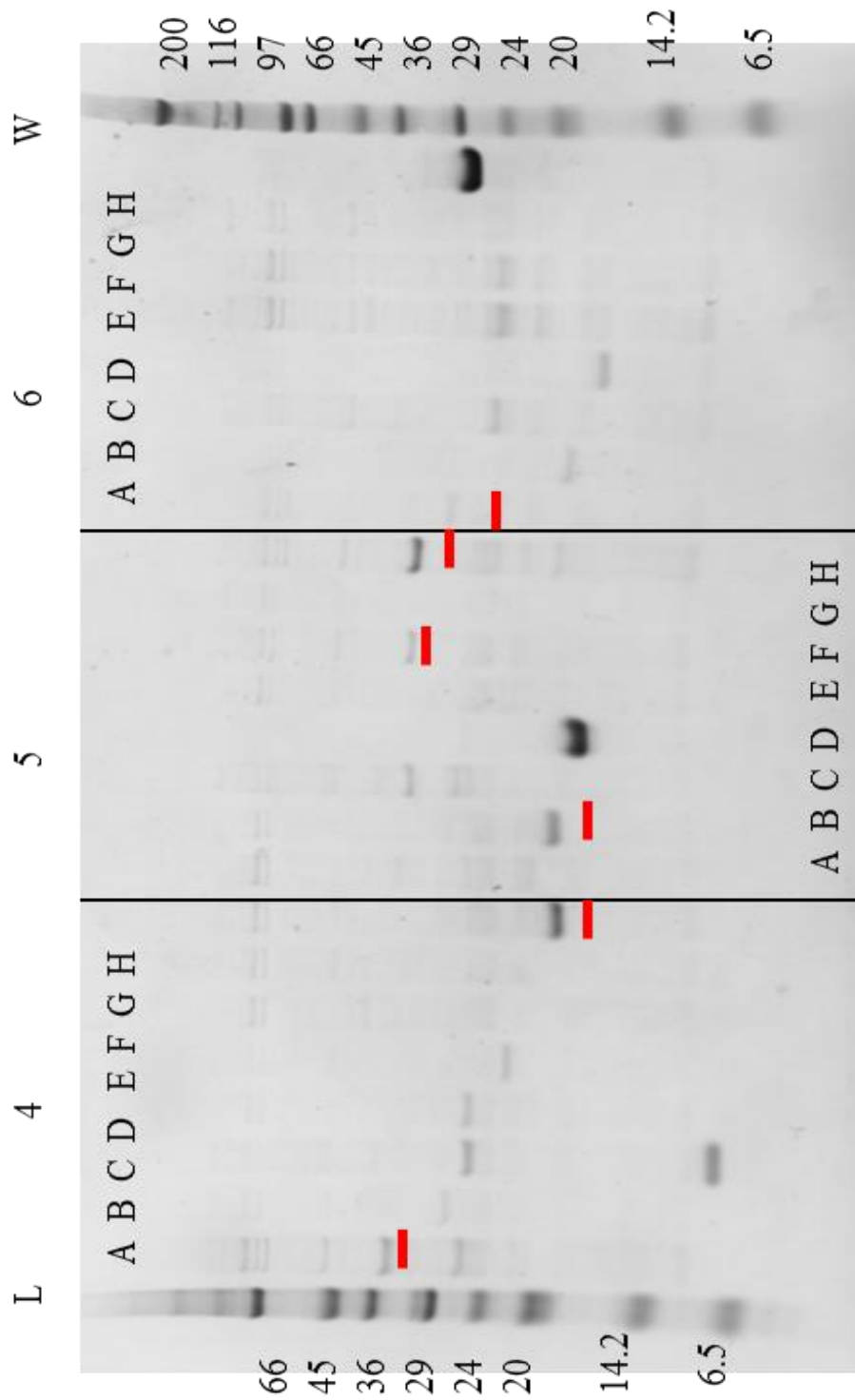


Figure 3.32 Overexpression test of HTP constructs in *E. coli* Rosetta2 induced with auto induction media. Gel: Invitrogen Bis/Tris 4-16%; **Marker** (L) Low Range (Sigma); (W) Wide Range (Sigma), molecular weights in kDa. Samples were purified by nickel affinity purification on BioRobot 8000.

3.4.3 Optimisation of protein purification

The importance of having stable protein cannot be underestimated when it eventually comes to crystallisation and buffer conditions play a very important part in this. In a study of 25 *E. coli* proteins that were assessed for stability by thermal shift assays, with and without 40 additives (inhibitors, co-factors, metal ions etc), the data suggested a 2-fold increase in crystallisation hits when proteins were in a favourable buffer with stabilising additives (Ericsson *et al.*, 2006).

Purification buffers were optimised to ensure proteins were stable prior to crystallisation. Proteins were purified by nickel affinity purification (5 mL HisTrap FF) and size exclusion chromatography (Superdex 75 16/60). All proteins were initially purified using the general purification buffers detailed in **Table 2.7**. The HisTrap buffers remained consistent across DevA, DevE, HutC and Gp26 as proteins were stable in these buffers. Proteins were assessed to be stable based on solubility when concentrating. The four proteins were also initially purified by SEC using generalised gel filtration buffer. HutC, DevA and DevE were moved forward into crystallisation trials in the generalised gel filtration buffer.

3.4.3.1 Thermofluor analysis reveals protein stability in different buffer conditions

Thermofluor is a biophysical assay which allows the relative stability of proteins to be assessed based on temperature melt curves (Nettlehip *et al.*, 2008, Geerlof *et al.*, 2006, Pantoliano *et al.*, 2001). Proteins are heated to 95°C in the presence of a SYPRO® Orange fluorophore (Molecular Probes™, Life Technologies). As protein unfolds, the fluorophore can bind to the open conformation of the protein causing an increase in fluorescence as the temperature rises. Increase in fluorescence is plotted

against melting temperature (T_m). Therefore the higher the T_m , the more stable the protein. This technique can be used to assess if different conditions (e.g. buffers, ligands, drug interactions) cause differences in the T_m , thus a difference in protein stability. Thermofluor analysis was employed to determine the best buffers for protein stability during gel filtration and, indirectly, crystallisation.

DevA, DevE, HutC and Gp26 were subject to thermofluor analysis to find the most stable buffer conditions. Only data for Gp26 is presented here as it was the only protein of the four which showed a preference for a different buffer (HEPES pH 7.5) from GF buffer 1 (**Table 2.7**). Gp26 was observed to be most stable in a buffer with pH 7.5 although salt did not appear to be a particularly important factor (**Figure 3.33**). Gp26 was purified in GF buffer 2 (20 mM HEPES, 0.5M NaCl, 2% glycerol) and was assessed to be more stable as concentrations of up to 40 mg ml^{-1} could be achieved compared to 24 mg ml^{-1} in GF buffer 1.

Summary

The results presented in this chapter demonstrate the difficulties that are often present on the road to obtaining “crystallisable” protein. Optimisation of protein production and purification was arguably the most important part of this project. Going forward into structural studies would most certainly not have been possible with these steps due to a lack of soluble protein. The following chapter deals with optimised protein purification and crystallisation of DevA, DevE, HutC and Gp26.

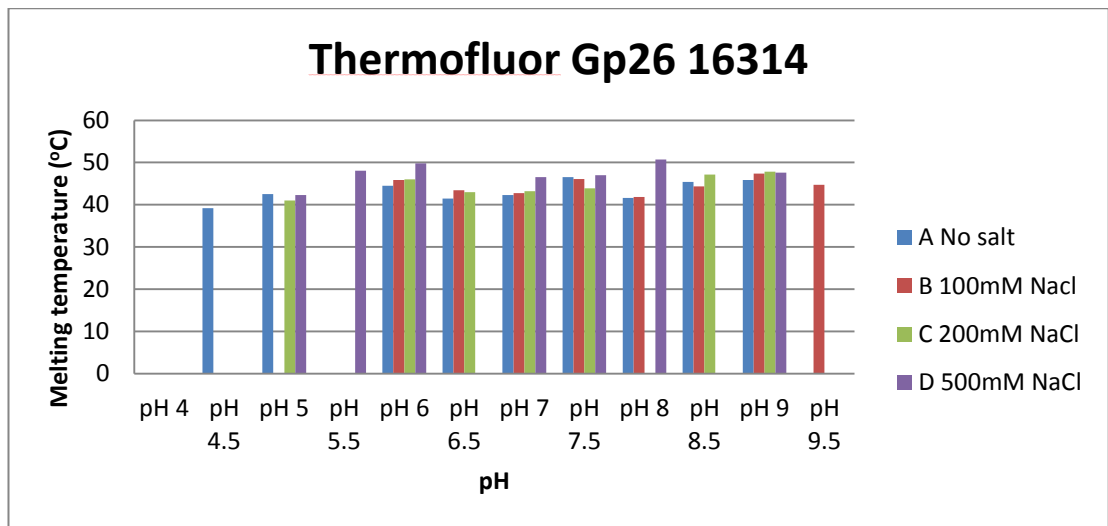


Figure 3.33 Thermofluor analysis of Gp26 purified by size exclusion chromatography (Superdex 200 16/60). Where no bars are visible, no fluorescence was measured. Gp26 is observed to be more stable at the more “biological” pH 7.5.

Chapter 4 Purification, biophysical analysis and crystallisation

4.1 DevA

4.1.1 Purification

Recombinant DevA was recovered from the supernatant of lysed *E. coli* Rosetta2 cells and purified by nickel affinity chromatography (5 mL HisTrap FF Crude) followed by gel filtration (Superdex 75 16/60). DevA eluted from the column at a retention volume of 50 mL (**Figure 4.1A**). When compared to a standard curve of proteins of known mass (ferritin, aldolase, ovalbumin and RNase), the molecular weight of the elution product was calculated to be ~136 kDa which would correspond to a tetrameric form of DevA. A shoulder is also observed on the main elution peak, which may reflect a degradation product or aggregated proteins.

Eluted fractions were analysed by SDS-PAGE. The shoulder on the main peak corresponds to fractions A6-A8 that contains some higher molecular weight proteins, perhaps aggregated protein. Some degradation bands were also evident in the main peak despite the addition of protease inhibitors (**Figure 4.1B**). Fractions were pooled and concentrated up to 20 mg ml⁻¹, flash frozen and stored at -80 °C for further use. Following concentration, SDS-PAGE was used to assess the purity of the DevA produced. A single band at ~33 kDa was observed indicating protein was pure (**Figure 4.1C**).

4.1.2 Size Exclusion Chromatography Multi-Angle Laser Light Scattering Analysis

Size Exclusion Chromatography Multi-Angle Laser Light Scattering Analysis (SEC-MALLS) was used to analyse absolute molecular weights and mono-dispersity of

DevA. The main peak, marked with an arrow in **Figure 4.2**, corresponds to a molecular weight of ~68 kDa. This indicates DevA is in a dimeric form in contrast to gel filtration data, which indicated a tetramer. The analysis also showed that the protein was monodisperse with $M_w/M_n = 1.003$ thus providing pure protein of sufficient quality for crystallisation.

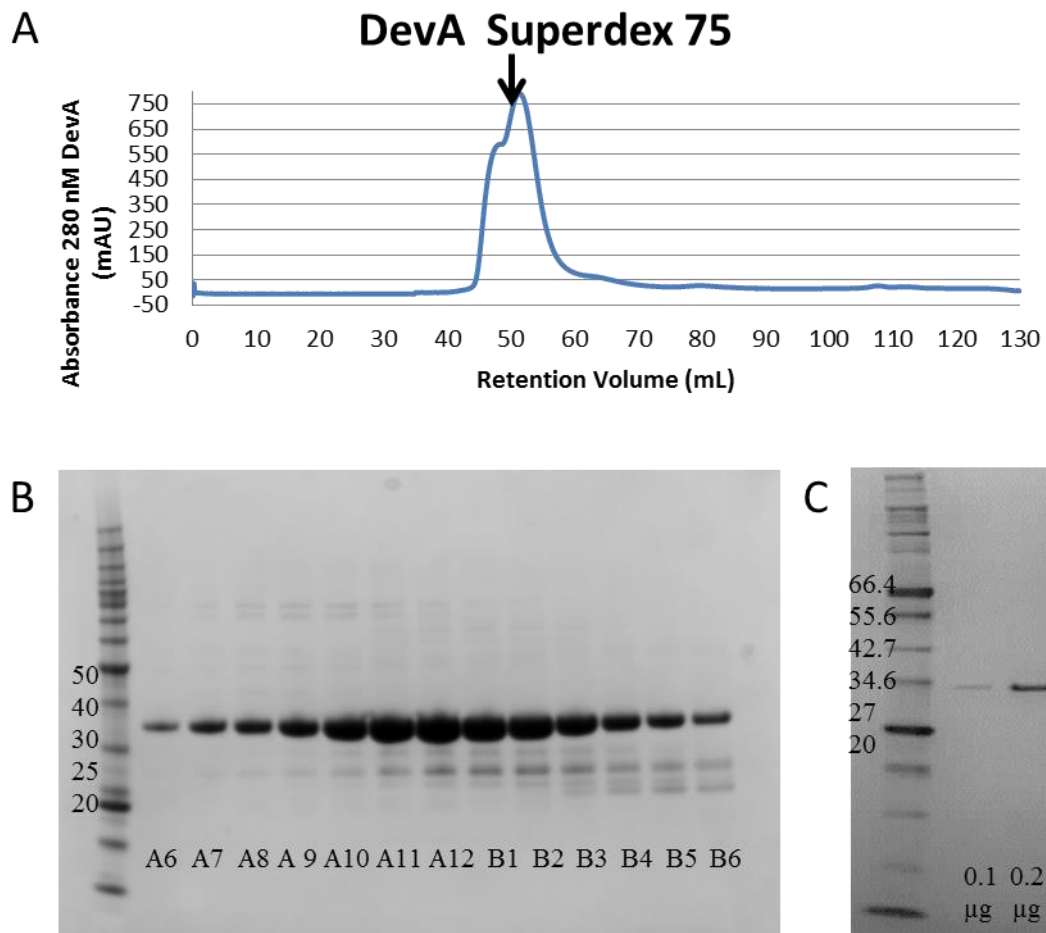


Figure 4.1(A) Elution profile of DevA overexpression in *E. coli* Rosetta2 induced by auto-induction. Column – HiLoad 16/60 Superdex 75, flow rate – 1 ml min^{-1} , buffer - GF buffer 1. **(B)** SDS-PAGE analysis of elution peak fractions. Fractions A6-A8 correspond to the shoulder on the main peak; **Marker** – Benchmark (Invitrogen). **(C)** 0.1 μg and 0.2 μg DevA showing purity of protein after concentration; **Marker** - Broad Range Protein Marker (NEB). . **Gel** - NuPage 4-12% Bis-tris, 1X SDS-MES running buffer, 200V. Molecular weights shown are in kDa.

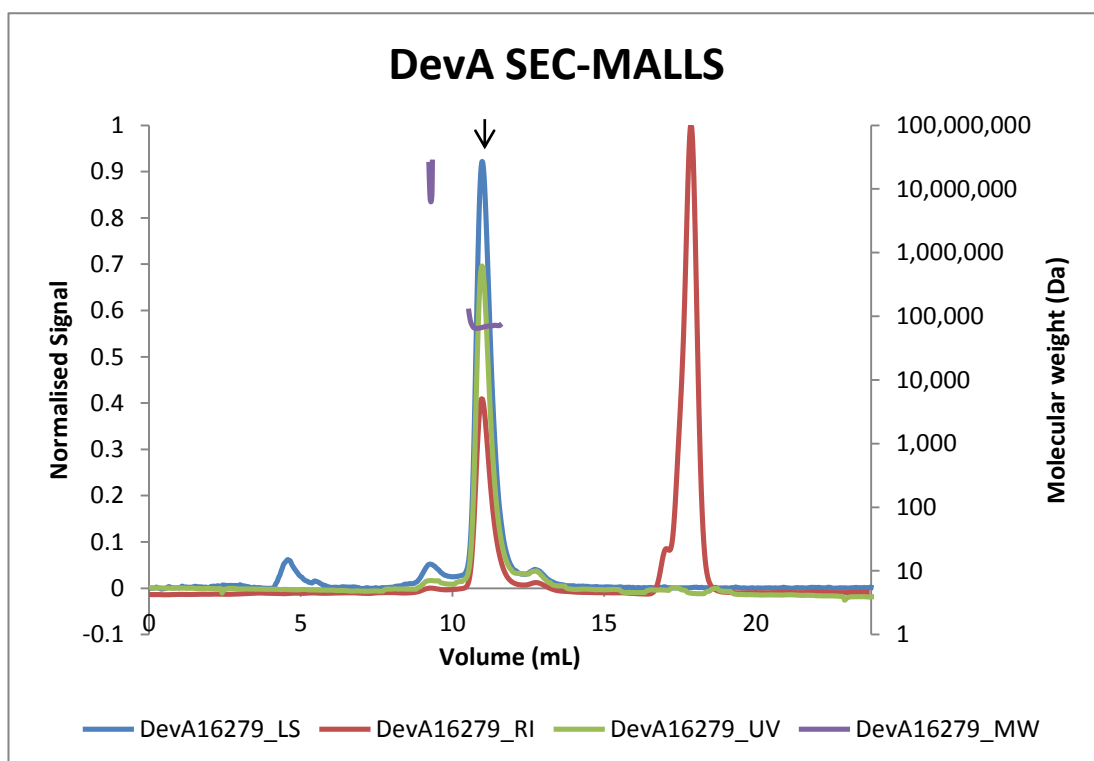


Figure 4.2 Molar mass vs volume plot of SEC-MALLS of DevA. The main peak (arrow) corresponds to a molecular weight ~ 68 kDa. Samples ($100 \mu\text{L}$; 2 mg ml^{-1}) were run at $25 \text{ }^\circ\text{C}$ with a flow rate of 0.7 ml min^{-1} . RI = refractive index of the sample; LS = light scattering of sample; UV = UV 280 nm of sample; MW = molecular weight

4.1.3 DevA Crystallisation

Extensive crystallisation trials were performed with DevA following from an initial screening run consisting of a total of 596 different conditions (6 commercially available screens) and under two different concentrations (16 and 18 mg ml⁻¹). Initial crystal hits identified from this HTP screening trial are detailed in **Table 4.1**. These hits typically showed potential crystals as long filaments or needles with the exception of one rod shaped crystal (**Figure 4.3**). These needles/crystals appeared after approximately 6 days with trials set up at 18 mg ml⁻¹ generating a greater number of reproducible hits.

Table 4.1 DevA initial crystal hits obtained from HTP crystal screening

Screen	Well	Concentration mg ml ⁻¹	Morphology	Time to appear
Wizard III & IV	D7 (0.2 M ammonium sulphate, 30% PEG 8000)	18	Needles	7 days
Morpheus*	E5 (0.12 M ethylene glycols, Buffer 2 pH 7.5 30% P550MME_P20K)	18	Filaments	1 day
JCSG+	E4 (0.2 M lithium sulphate, 0.1 M Tris pH 8.5, 1.26 M ammonium sulphate)	18	Filaments	15 days
	F7 (0.8 M succinic acid pH 7.0)	18	Filaments	1 day
	G2 (0.02 M magnesium chloride, 0.1 M HEPES pH 7.5, 22% polyacrylic acid 5100 sodium salt)	18	Rock	2 days
Index	C2 (1.1 M ammonium tartrate dibasic pH 7.0)	16	Needles	5 days
	A9 (0.1 M Bis-Tris pH 5.5, 3 M sodium chloride)	18	Filaments	6 hours
	C2 (1.1 M ammonium tartrate dibasic pH 7.0)	18	Filaments/needles	2 days
	C8 (1 M ammonium sulphate, 0.1 M Bis-tris pH 5.5, 1% PEG 3350)	18	Filaments/needles	2 days

* **Ethylene glycols** (0.3M Diethylene glycol; 0.3M Triethylene glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol), **Buffer 2 pH 7.5**; Sodium HEPES; MOPS (acid), **P550MME_P20K** (40% v/v PEG 550 MME; 20 % w/v PEG 20000)

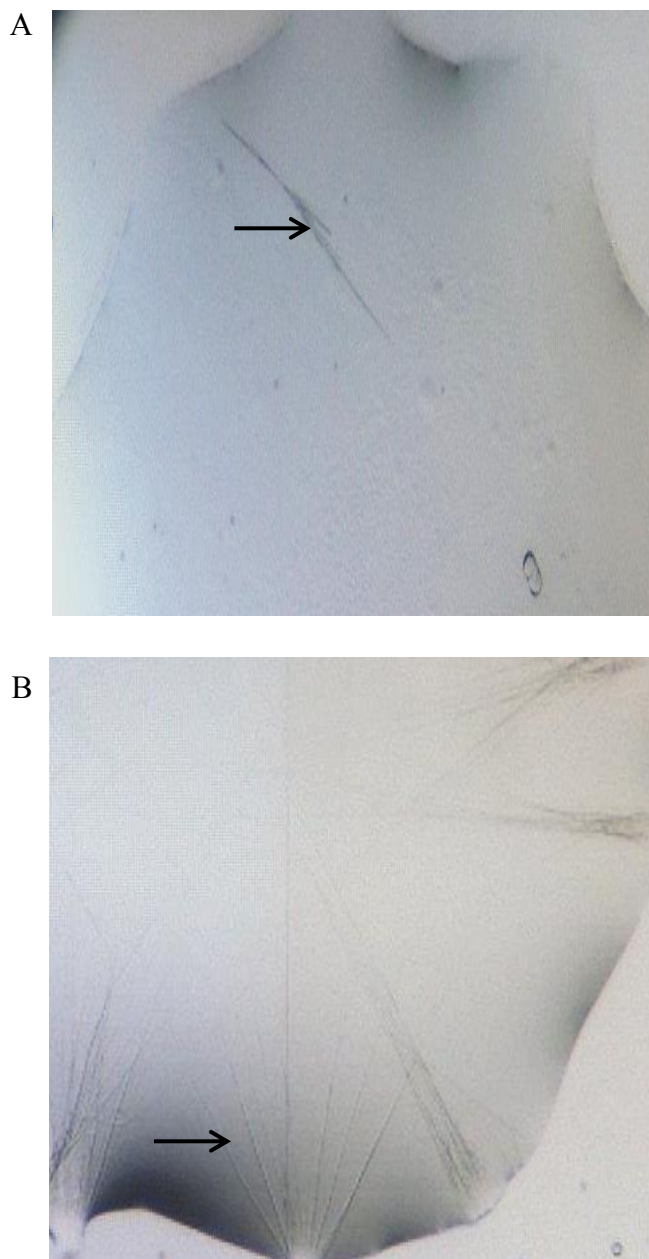


Figure 4.3 A range of crystal hits obtained for crystallisation trials of DevA grown in a variety of commercially available crystal screens at 20°C. **(A)** Wizard III & IV **D7** and **(B)** Index **C8**. Images taken automatically via Rock Imager from XtalPIMS (<https://www.oppf.rc-harwell.ac.uk/xtalpims/>).

Optimisation of DevA crystals

Based on these crystal hits several sitting drop optimisation trays were designed, encompassing the upper and lower pH ranges (pH 7.4 – pH 8.4 and pH 5.1 – pH 6.1, respectively) and 1 M – 1.75 M $(\text{NH}_4)_2\text{SO}_4$ at 20°C. The reservoir volume dispensed was 400 μL and the crystallisation drop volume was 1 μL (1:1 ratio protein:reservoir) in vapour diffusion sitting drop trays. Tray design is detailed in **Fig. 4.4 (A & B)**. After approximately 2 days, many micro crystals were observed in these optimisation trays. These micro crystals, however, did not grow any bigger over the course of two weeks therefore these initial crystals were used in seeding experiments in an attempt to obtain larger crystals. Briefly, the microcrystals from the optimisation trays (above) were crushed using an acupuncture needle whilst in the sitting drop and 1 μL of reservoir solution was added. This sample was designated UD (undiluted). From the UD stock 1/10, 1/100 and 1/1000 dilutions were made. The fresh reservoir solution comprised of 1 M ammonium sulphate, 10 % PEG 3350, 20 mM MgCl_2 and 20 mM CaCl_2 . Again, the lower pH range was used. **Figure 4.4C** details the layout of a micro-seeding optimisation tray. Crystallisation drops were set in 24 well vapour diffusion sitting drop trays with a reservoir volume of 400 μL and crystallisation drop volume of 1 μL . Reservoir solution was pipetted on to the sitting drop platform and were allowed to equilibrate in the sitting drop trays for 3 hours before seed stocks were streaked over. The control row contained no protein, only reservoir solution. Trays were incubated at 20°C. Although significant effort was put into improving the initial hits for DevA, no crystals have yet been obtained of sufficient quality for X-ray diffraction analysis. Hence work remains to be done to obtain suitable crystals for DevA. Future work directed at attempting to crystallise

the C-terminal domain of DevA alone to eliminate the expected high degree of flexibility between the HTH and C-terminal domains could be envisioned. Other strategies to pursue for the structural characterisation of the full length protein could include lysine methylation (Walter *et al.*, 2006) or surface entropy reduction (Cooper *et al.*, 2007, Goldschmidt *et al.*, 2014) in an attempt to improve crystallisability of the protein.

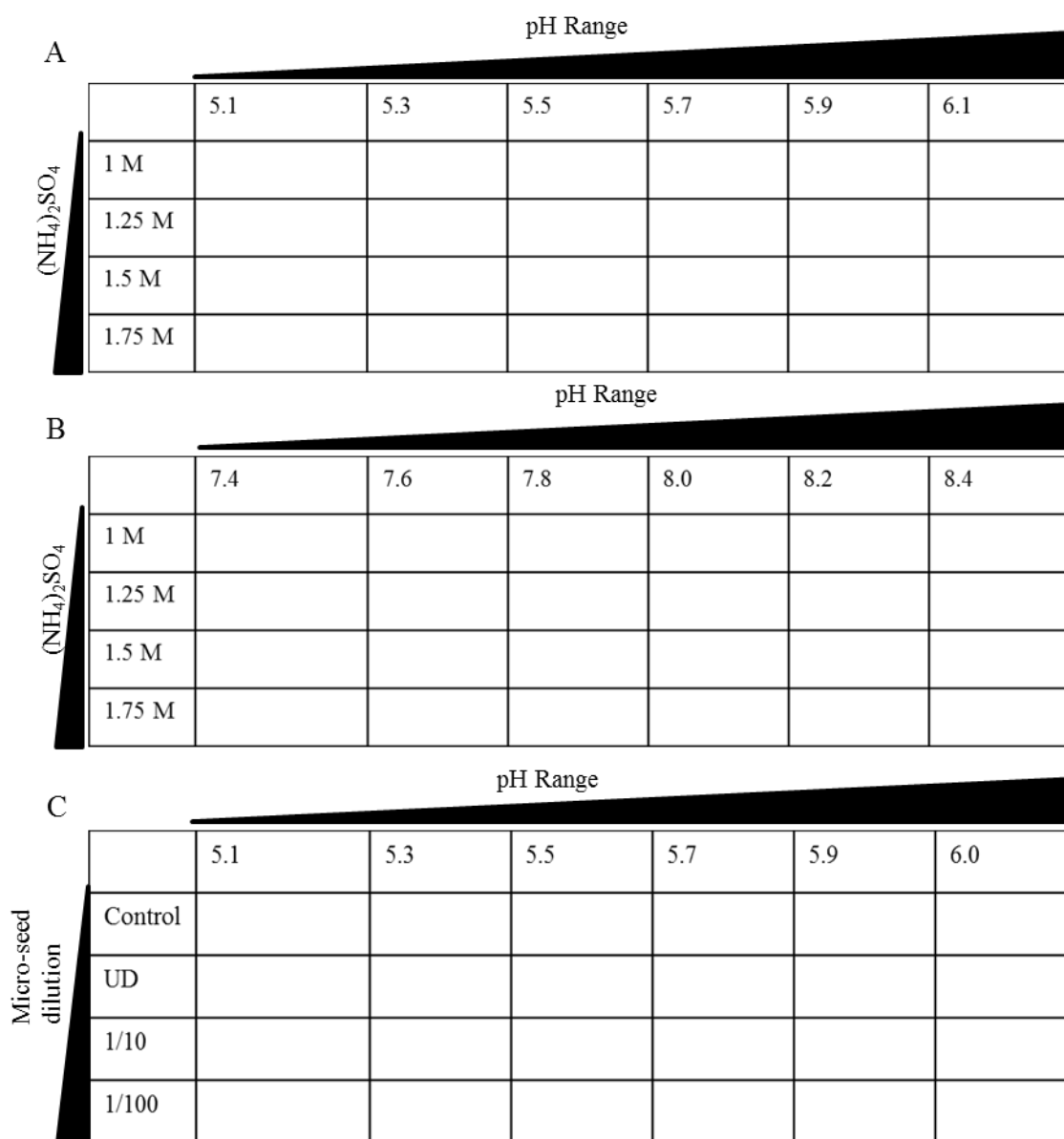


Figure 4.4 Layout of optimisation trays of DevA crystals by varying the ammonium sulphate concentration and lower (**A**) and upper (**B**) pH range of observed crystal hits from HTP screening. (**C**) Micro-seeding tray layout

4.2 DevE

4.2.1. Purification of apo-DevE

DevE was isolated from *E.coli* Rosetta2 and purified by nickel affinity chromatography (5 mL HisTrap FF) followed by gel filtration (Superdex 75 16/60). There were four elution peaks evident (**Figure 4.5A**). Peaks 1, 3 and 4 were not found to contain DevE when analysed by SDS-PAGE (**Figure 4.5B**). Peaks 3 and 4 were found to contain protein bands <20 kDa, which could probably be attributed to degraded protein. Peak 2 however, elutes at ~51 ml which corresponds to a molecular weight of ~62 kDa when calculated from the standard curve. This would indicate that DevE is a dimer in solution. Fractions were pooled and concentrated up to 31 mg ml⁻¹, flash frozen and stored at -80°C. Concentrated DevE was analysed for purity by SDS-PAGE (**Figure 4.5 C**).

4.2.2 SEC-MALLS

SEC-MALLS was carried out to determine absolute molecular weights and monodispersity of DevE. Two peaks were observed corresponding to molecular weights of ~47.5 kDa and ~61 kDa (**Figure 4.6**; peaks 1 and 2, respectively). Peak 2 (61 kDa) is consistent with a dimeric form of DevE, which is in agreement with the observations during gel filtration. The 47.5 kDa peak could be indicative of the equilibrium state of the protein in solution as this represents neither a dimeric nor a monomeric form of the protein and possibly represents an intermediate state of the equilibrium. DevE was observed to be monodisperse with a Mw/Mn value of 1.001 providing high quality protein for crystallisation trials.

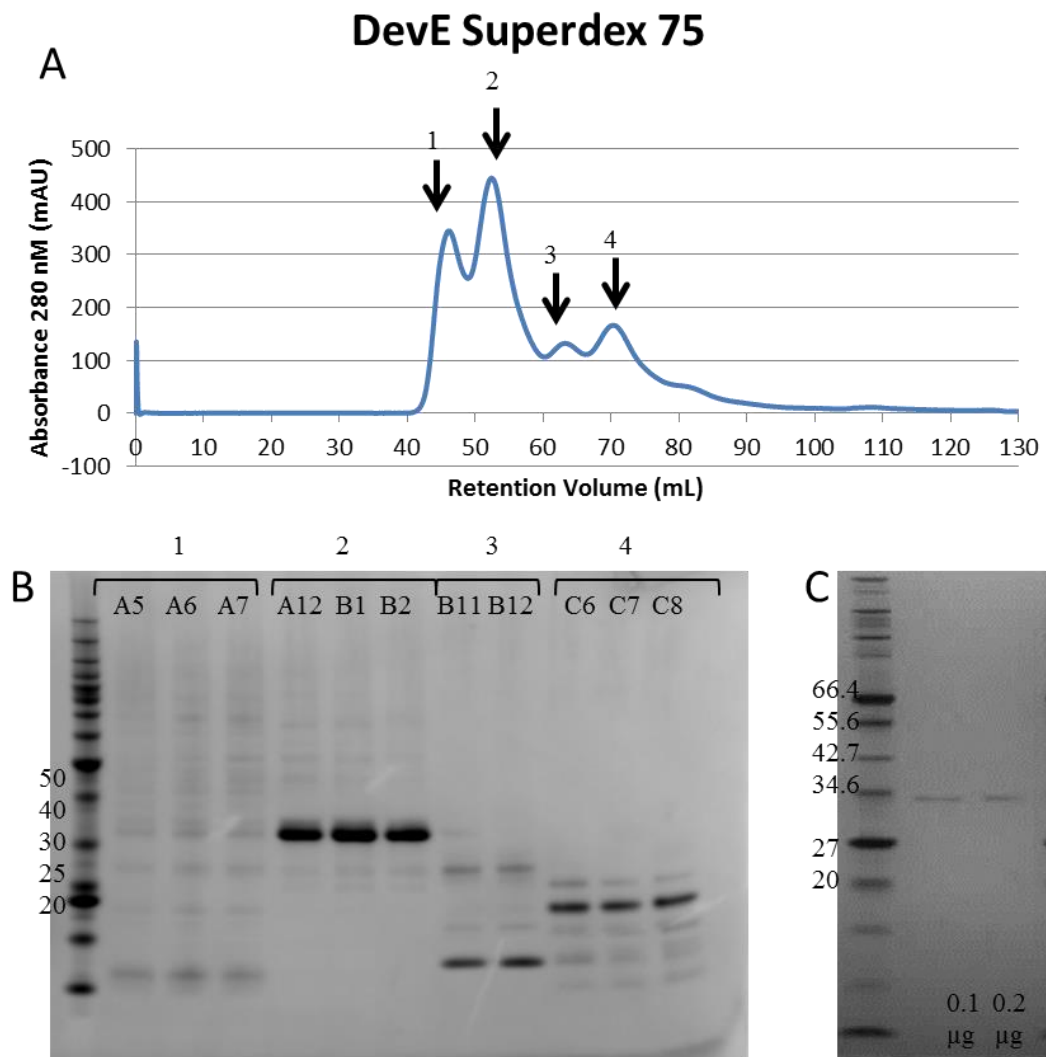


Figure 4.5 (A) Elution profile of DevE overexpression in *E. coli* Rosetta2 induced by auto-induction. Column – HiLoad 16/60 Superdex 75, flow rate - 1 ml min⁻¹, buffer - GF buffer 1. (B) SDS-PAGE analysis of fractions from multiple elution peaks (1-4); **Marker** – Benchmark (Invitrogen). (C) 0.1 µg and 0.2 µg DevE showing purity of protein after concentration; **Marker** – Broad Range Protein Marker (NEB). Gel NuPage 4-12% Bis-tris, 1X SDS-MES running buffer, 200V. Molecular weights shown are in kDa.

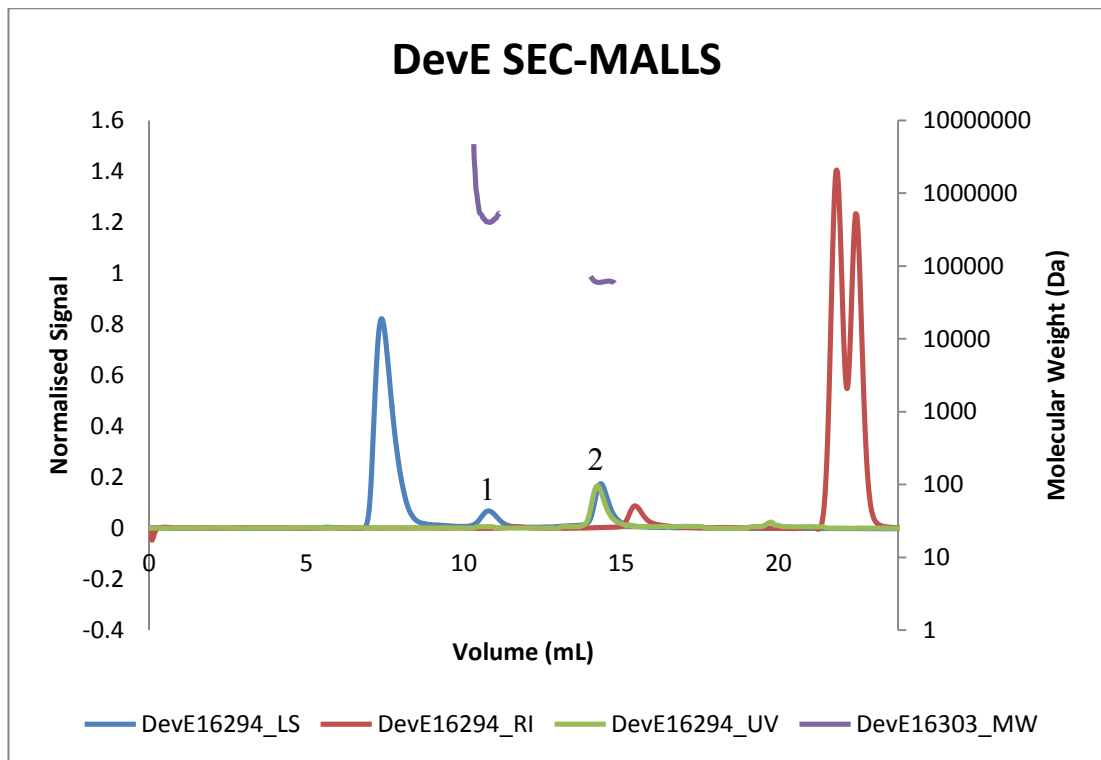


Figure 4.6 Molecular weight vs volume plot of SEC-MALLS of DevE. Two peaks were identified as containing protein. DevE is observed as a dimer with a peak corresponding to a molecular weight of 61kDa. Samples were run at 25 °C with a flow rate of 0.7 ml min⁻¹. RI = refractive index of the sample; LS = light scattering of sample; UV = UV 280 nm of sample; MW = molecular weight

4.2.3 Crystallisation of DevE

4.2.3.1 Crystallisation and characterisation of *Apo*-DevE crystals for structural analysis

DevE did not readily crystallise. Crystal hits obtained for DevE are detailed in **table 4.2**. The best diffracting crystal typically appeared after 40-60 days with crystals demonstrating different morphologies (**Figure 4.7**). The best diffracting crystal was grown in JCSG+ condition **H11** (0.2 M MgCl₂, 0.1 M Bis-Tris pH 5.5, 25% PEG w/v 3350). Efforts to reproduce these crystals, however, have been hampered by the crystal growth times and reproducibility issues but are continuing. Nevertheless we were fortunate enough to characterise a single crystal grown from this condition which is detailed in the next section.

Table 4.2 DevE crystal hits obtained from HTP crystal screening

Screen	Well	Concentration mg ml ⁻¹	Morphology	Time to appear	Diffraction
Apo-DevE					
Index	G4 (0.2 M lithium sulphate, 0.1 M HEPES pH 7.5, 25% PEG 3350)	30.7	Micro crystals	38 days	
JCSG+	H11 (0.2M magnesium chloride, 0.1 M Bis-Tris pH 5.5), 25% PEG 3350)	22.6	Rod	38 days	2.7 Å
PACT	C8 (0.2 m ammonium sulphate, 0.1 M HEPES pH 7.5, 20% PEG 6000)	22.6	Needles	38 days	
	D7 (0.2 M sodium chloride, 0.1 M Tris pH 8.0, 20% PEG 6000)	22.6	Needles	38 days	
	G2 (0.2 M sodium bromide, 0.1 M Bis-Tris propane pH 7.5, 20% PEG 3350)	22.6	Needle cluster	15 days	9 Å
Selenomethionine DevE					
Wizard III & IV	H5 (0.2 M ammonium sulphate, 0.1 M HEPES pH 7.5, 10% 2-propanol, 20% PEG 8000)	15	Rods	9 days	
Morpheus*	C9 (0.09 M NPS pH 8.5, 30% P550MME_P20K)	15	Micro needles	8 days	
DevE + 18mer					
PACT	E11 (0.2 M sodium citrate tribasic dihydrate, 20% PEG 3350)	16	Rhomboid	38 days	3 Å
	F12 (0.2 M sodium malonate dibasic monhydrate, 0.1 M Bis-Tris propane pH 6.5, 20% PEG 3350)	16	Needles	8 days	
	H2 (0.2 M sodium bromide, 0.1 M Bis-Tris propane, pH 8.5, 20% PEG 3350)	16	Needles	38 days	
Morpheus	C9 (0.09 M NPS pH 8.5, 30% P550MME_P20K)	16	Rods	15 days	
Index	E11 (0.02 M magnesium chloride heptahydrate, 0.1 M HEPES pH 7.5, 22% polyacrylic sodium salt 5100)	16	Rock	15 days	
Peg/Ion	F8 (0.2 M succinic acid pH 7.0, 20 % PEG 3350)	16	Needles	8 days	

* **NPS** (0.3M Sodium nitrate, 0.3 Sodium phosphate dibasic, 0.3M Ammonium sulfate), **Buffer 3 pH 8.5**; Tris (base); **BICINE**, **P550MME_P20K** (40% v/v PEG 550 MME; 20 % w/v PEG 20000)

4.2.3.2 Data Collection and analysis

Apo-DevE diffraction data were collected on beamline I04-1 at Diamond Light Source using a Pilatus 2M pixel array detector (Dectris). A total of 250 degrees of data were collected from 2,500 diffraction images using a 0.1° oscillation angle. A representative diffraction image can be found in **Figure 4.8**. Data were automatically integrated and scaled using the *xia2* pipeline to a resolution of 2.7 Å. The crystal was characterised to belong to the monoclinic space group P2₁ with unit cell dimensions $a = 53.07$ Å, $b = 43.49$ Å, $c = 73.22$ Å and $\alpha = 90.00^\circ$, $\beta = 106.09^\circ$, $\gamma = 90.00^\circ$. Based on these cell parameters the contents of crystallography asymmetric unit is a DevE monomer. This gives a Matthews coefficient (Matthews, 1968) of $2.38 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of 48.4%. Data collection statistics are shown in **Table 4.3**.

Table 4.3 Data collection statistics for DevE

	DevE
Data collection	
Space group	P12 ₁ 1
Cell dimensions	
a, b, c (Å)	53.07, 43.49, 73.22
α, β, γ (°)	90° 106.09° 90°
Resolution (Å)	70.35-2.66 (2.73-2.66)*
^a R _{merge}	0.058 (0.650)
$I / \sigma I$	12.5 (2.0)
Completeness (%)	97.6 (97.4)
Multiplicity	4.7 (4.8)
Wavelength (Å)	0.9200

*Values in parentheses are for highest-resolution shell. $^a R_{\text{merge}} = \sum |I_{\text{JHKL}} - \langle I \rangle_{\text{HKL}}| / \sum I_{\text{JHKL}}$

BLAST searches of the PDB identified two regions of homology; the N-terminal domain (HTH) and the C-terminal domain. YvoA (2WV0) was identified as having 37% identity with DevE at the HTH domain (residues 9 – 67 and residues 17 – 75, respectively). The FadR (1E2X) HTH was also used as a homology model due to the highly conserved nature of this domain and the previous characterisation of FadR. The C-terminal domain of DevE has 30% structural homology (residues 174-266) with a human sulfiredoxin (residues 9-106; 1YZS). These were identified as suitable candidates for homology models to be used in molecular replacement. Efforts have been made to solve this structure using molecular replacement by *PHASER* (McCoy *et al.*, 2007) using these models, however no solutions could be found. Further attempts to solve the apo-DevE structure have been hampered by the length of time for crystals to initially appear as well as the low rate of reproducibility of the crystals. Efforts continue to obtain a larger number of crystals to facilitate structure solution.

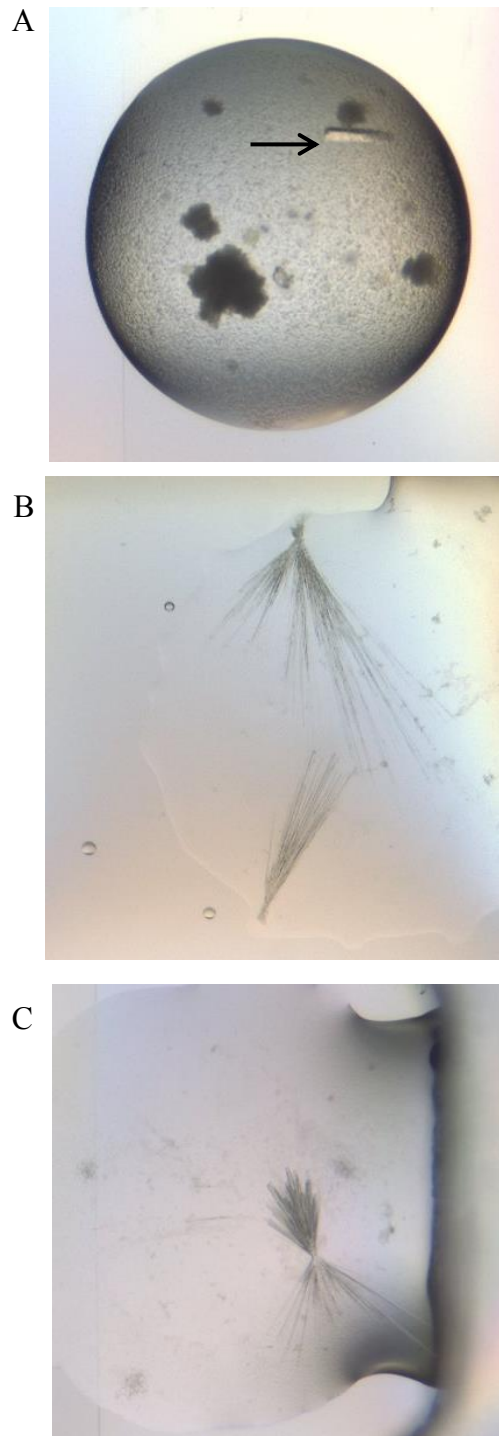


Figure 4.7 Crystals of *apo-DevE* demonstrating different morphologies (A) JSCG+ H11; (B) Pact D9; (C) Pact G2

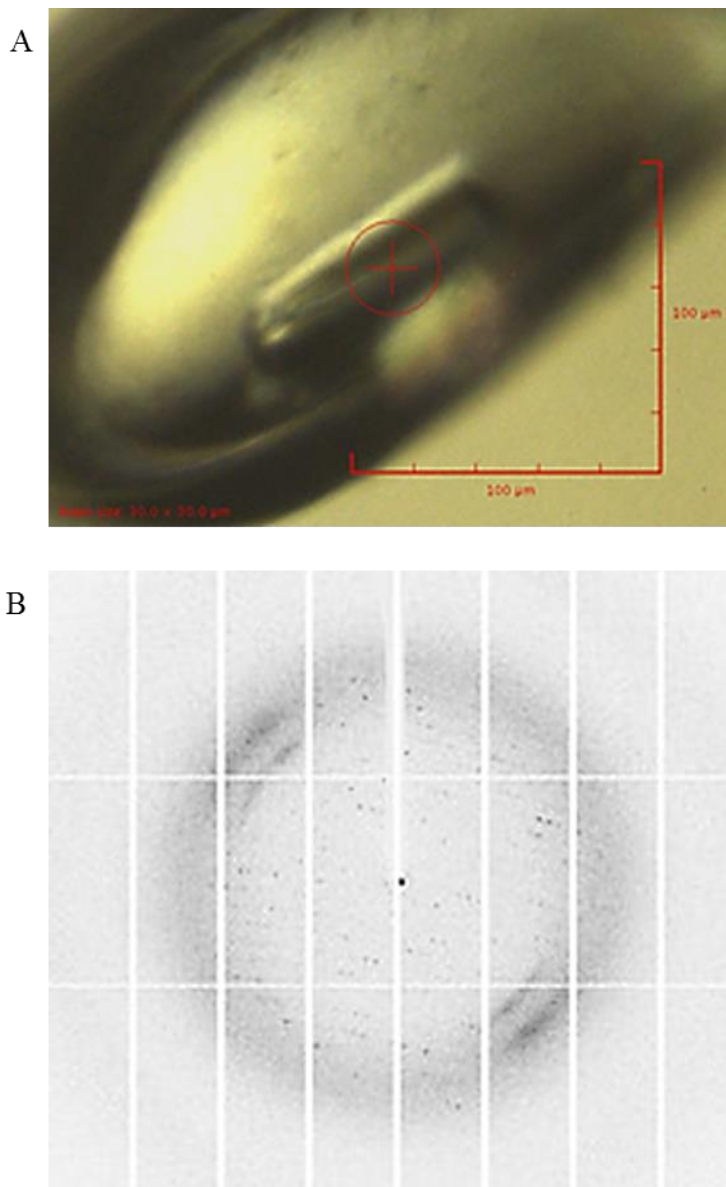


Figure 4.8 (A) A single crystal mounted on beamline I04-1 in a nylon loop holder grown in JCSG+ condition **H11**. The red circle denotes the beam incident on the crystal (100 μm). Diffraction data were collected on beamline I04-1 at Diamond Light Source and diffracted to a resolution of 2.66 Å. (B) Representative diffraction data collected from the crystal in (A)

4.2.4 Co-purification of DevE-18mer complex

Crystallisation of protein-DNA complexes can provide a great deal of information with regards to the mechanism of protein function. We identified potential promoter regions within the 300 bp upstream region of *devE* (see **Chapter 5** for details). In order to characterise the binding interaction between DevE and its promoter region, efforts were made to crystallise the complex.

Gel filtration buffer conditions were optimised following a thermofluor analysis which revealed that DevE-DNA complex was more stable at pH 7.0 and a reduced salt content compared with GF buffer 1 which *apo*-DevE was purified in and appeared to be stable. GF buffer 3 contained 20 mM MOPS pH 7.0, 0.2 M NaCl and 2% glycerol. Small amounts of MgCl₂ (5 mM) and CaCl₂ (5 mM) were also included as a means to counteract the anionic nature of the DNA that may affect the protein environment in solution. The absorbance at 280 nm and 254 nm was monitored during purification to assess if the complex was stable (**Figure 4.9 A**). The protein-DNA interaction did not appear to be stable during gel filtration and PAGE analysis revealed that protein and DNA were not bound (**Figure 4.9 B & C**). The presence of aromatic amino acid residues (primarily tryptophan) is responsible for fluorescence under UV light.

4.2.3.1 Crystallisation of DevE + 18mer dsDNA

Purified DevE protein was diluted with 18mer dsDNA equating to final concentrations of 65 μM and 160 μM respectively. The protein: DNA complex was incubated together for 15 minutes at room temperature. The complex was then concentrated using 3K MWCO centrifugal filter (Millipore). The final concentration of DevE was determined by a Bradford assay using a standard curve with known

concentrations of DevE. The DNA concentration was assumed to be proportional to the concentration of DevE. Crystallisation trials were set up with the DevE-18mer as per **section 2.4.1**.

Crystals of DevE + 18mer DNA crystallised in a variety of conditions (**Table 4.2**) and were observed to demonstrate a number of morphologies (**Figure 4.10**). Crystals appeared between 8 and 38 days, with the best diffracting crystal appearing typically after 40-60 days. These crystals have, like *apo*-DevE, been difficult to reproduce and the turn-around time of 1-2 months has contributed to the difficulty of obtaining a cohort of crystals for diffraction analysis. Nevertheless, a small number of crystals were tested for diffraction quality. The most promising crystal was grown from the PACT Premier crystal screen condition **E11** (**Table 4.2**) which is made up of 0.2 M sodium citrate and 20% w/v PEG 3350.

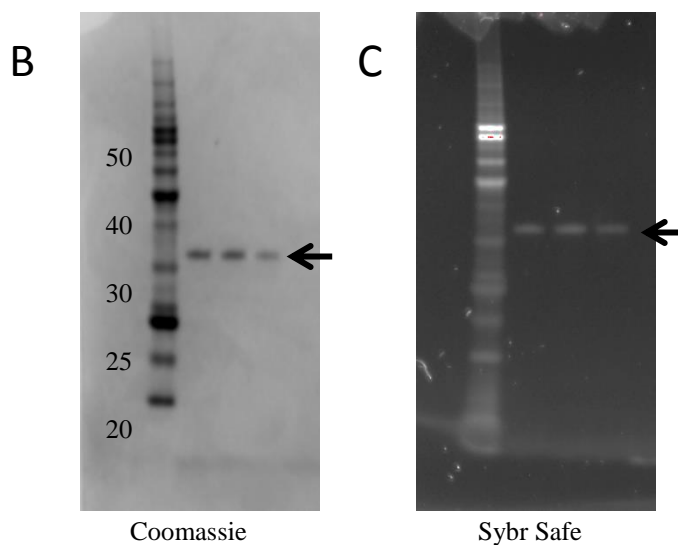
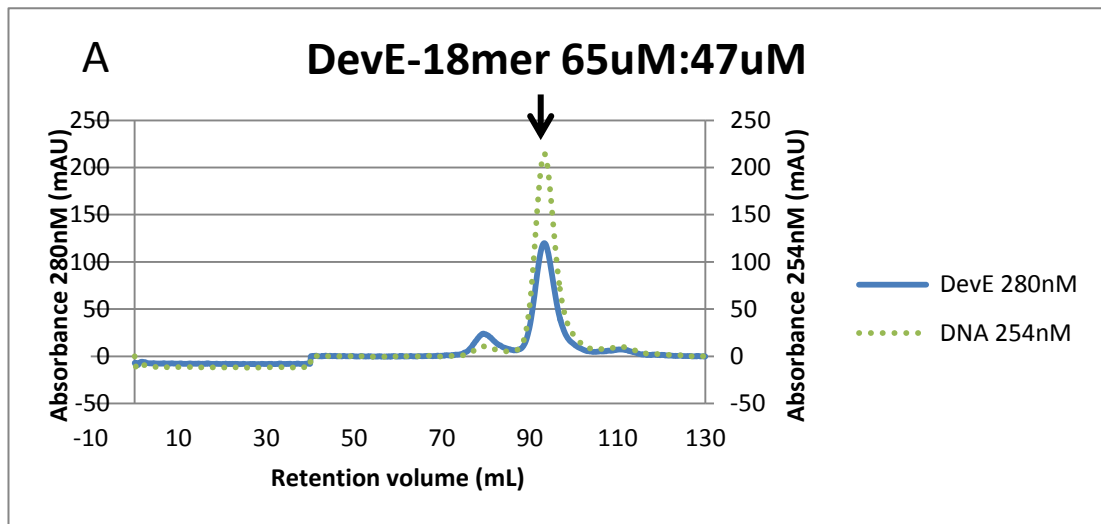


Figure 4.9 (A) Elution profile of co-purified DevE and 18mer promoter region. Column – HiLoad 16/60 Superdex 200, flow rate – 1 ml min^{-1} , buffer - GF buffer 3. (B) SDS-PAGE analysis of fraction from main elution peak stained with Instant Blue. (C) SDS-PAGE gel stained with SYBR Safe to assess DNA binding to protein. Fluorescence is due to aromatic residues within protein rather than DNA indicating binding is not stable during co-purification. Gel NuPage 4-12% Bis-tris, 1X SDS-MES running buffer, 200V. Marker – Benchmark (Invitrogen). Molecular weights shown are in kDa.

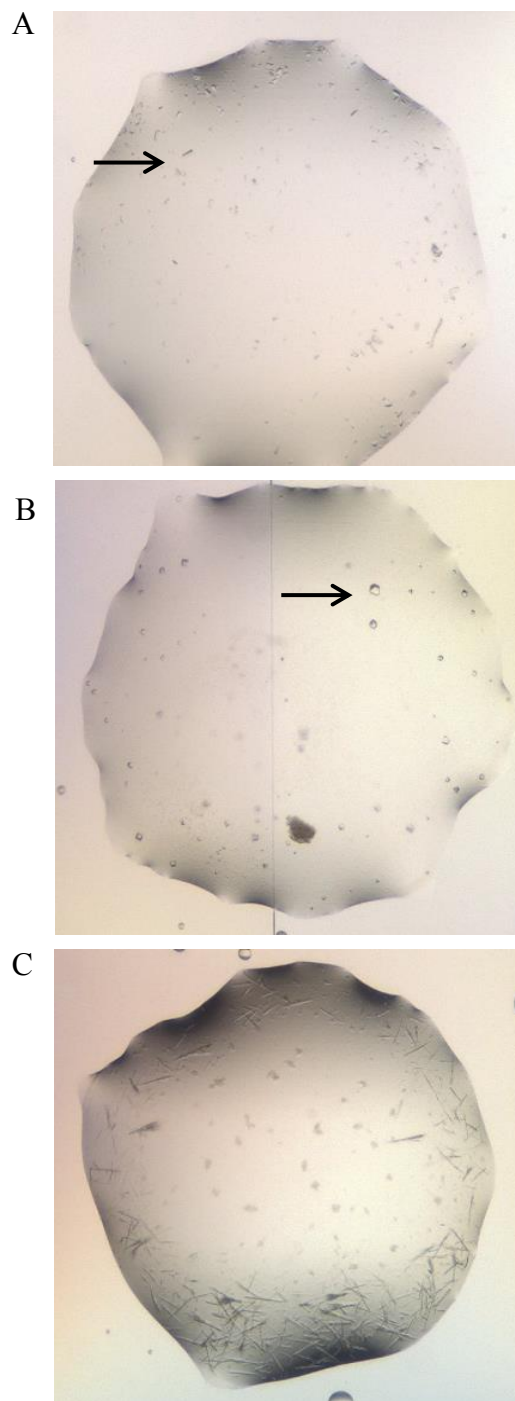


Figure 4.10 Crystals of DevE+18mer DNA demonstrating different morphologies
(A) Morpheous C9; (B) Pact E11; (C) Pact F12

Data Collection and analysis

Diffraction data from a single crystal of the putative DevE-18mer complex were collected on beamline I03 using a Pilatus3 6M detector (Dectris). Diffraction data was collected over a total 180 degrees from 1,200 images using a 0.15° oscillation angle. Data were integrated and scaled by the *xia2* pipeline to a resolution of 3 Å. The crystal was characterised as belonging to the tetragonal space group P4₁2₁2 (or its enantiomorph P4₃2₁2) with unit cell dimensions a = 52.17 Å, b = 52.17 Å, c = 295.88 Å and $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$. This provides enough space for only a single DevE monomer which would give a Matthews coefficient of 3.02 Å³ Da⁻¹ and a solvent content of 59.3%. Data collection statistics are detailed in **table 4.4**.

Table 4.4 Data collection statistics for DevE-18mer

	DevE-DNA
Data collection	
Space group	P4 ₁ 2 ₁ 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.17, 52.17, 295.88
α , β , γ (°)	90° 90° 90°
Resolution (Å)	52.17-2.69
^a R _{merge}	0.135 (2.971)*
<i>I</i> / σI	13.7 (1)
Completeness (%)	99.8 (99.4)
Multiplicity	12.2 (13.1)
Wavelength (Å)	0.97625

*Values in parentheses are for highest-resolution shell (2.79 – 2.69 Å).

$$^a R_{\text{merge}} = \sum | I_{\text{JHKL}} - \langle I \rangle_{\text{HKL}} | / \sum I_{\text{JHKL}}$$

Structure determination was attempted using the molecular replacement method for DevE. For this the homology models that were identified during efforts to solve apo-DevE from the monoclinic P2₁ crystal form were also used.

Unfortunately, no molecular replacement solutions were found with these crystal forms, hence as crystals were scarce then it was decided to characterise selenomethionine labelled DevE protein with the hope that this would crystallize and allow phases to be obtained from the anomalous scattering generated by the incorporated selenium atoms (Hendrickson *et al.*, 1990).

Crystallisation trials of Selenomethionine labelled DevE

Selenomethionine (SeMet) labelled DevE protein was produced as described in **section 2.3.2**. Crystallisation conditions for the SeMet labelled DevE protein were rescreened in addition to setting up conditions around the original crystal hits for the unlabelled protein. Crystals were observed to appear more rapidly for SeMet-DevE than was the case for *apo*-DevE; within 9 days rod shaped crystals had appeared in Wizard III & IV **H5** (**Table 4.2/Figure 4.11**). Unfortunately, these crystals did not diffract when tested so conditions were re-screened around Wizard III & IV condition **H5** (0.1 M HEPES/NaOH pH 7.5, 0.2 M (NH₄)₂S₀₄, 10% v/v 2-propanol, 20% w/v PEG 8000). The concentration of 2-propanol and PEG 8000 using 10 mg ml⁻¹ SeMet DevE were initial variables in optimisation. Protein to reservoir ratios in the drops was also changed to 2:1 (1:1 in original condition). This was followed by optimisation of the ammonium sulphate concentration and variation of nucleation inhibitors (ethylene glycol, glycerol and tacsimate). Variation of protein: reservoir ratios (1:1 and 2:1) were included in this optimisation step that allows varying

concentrations of protein and crystallisation reagents to be tested simultaneously. Optimisation drops were set up in 24 well vapour diffusion sitting drop trays with 200 μL reservoir volume and 2 μL crystallisation drops. Optimisation tray layouts can be found in **Figure 4.12**.

Unlike *apo*-DevE crystals SeMet crystals were easily reproduced but despite efforts to optimise these, no diffraction was observed from these crystals. To assess if the harvesting and process (mounting and cryocooling of crystals) was a factor in this lack of diffraction, crystals were tested for diffraction directly from the crystallisation trays. In this *in situ* method the crystallization plate is placed directly in the X-ray beam at the beamline then by use of a X-Y translation stage mounted on the beamline rotation axis the crystallization drops contained in the tray can be systematically translated into the X-ray beam. Crystals therein can then be tested for diffraction without being disturbed from the crystallization mother liquor. Again, no diffraction was observed for crystals tested directly in the crystallisation plates from where they were grown, indicating a lack of crystallinity.

Although the structure of DevE has not been solved during the course of this thesis the initial data collections from single crystals is encouraging. Solitary crystals of DevE have enabled the collection of data sets from two different crystal forms of DevE to approximately 2.7 \AA resolution. Unfortunately attempts to solve the structure of DevE by molecular replacement using these data have not been successful to date. Structure solution has been further hampered by the failure of SeMet DevE crystals to diffract. However, efforts are continuing with the availability of a couple of DevE crystals and putative DevE-18mer complex crystals to obtain phases using the anomalous scattering from the sulphur atoms inherently

present within the protein as well as potentially from phosphorus atoms within the DNA backbone of the DevE-18mer complex at the time of this thesis being submitted.

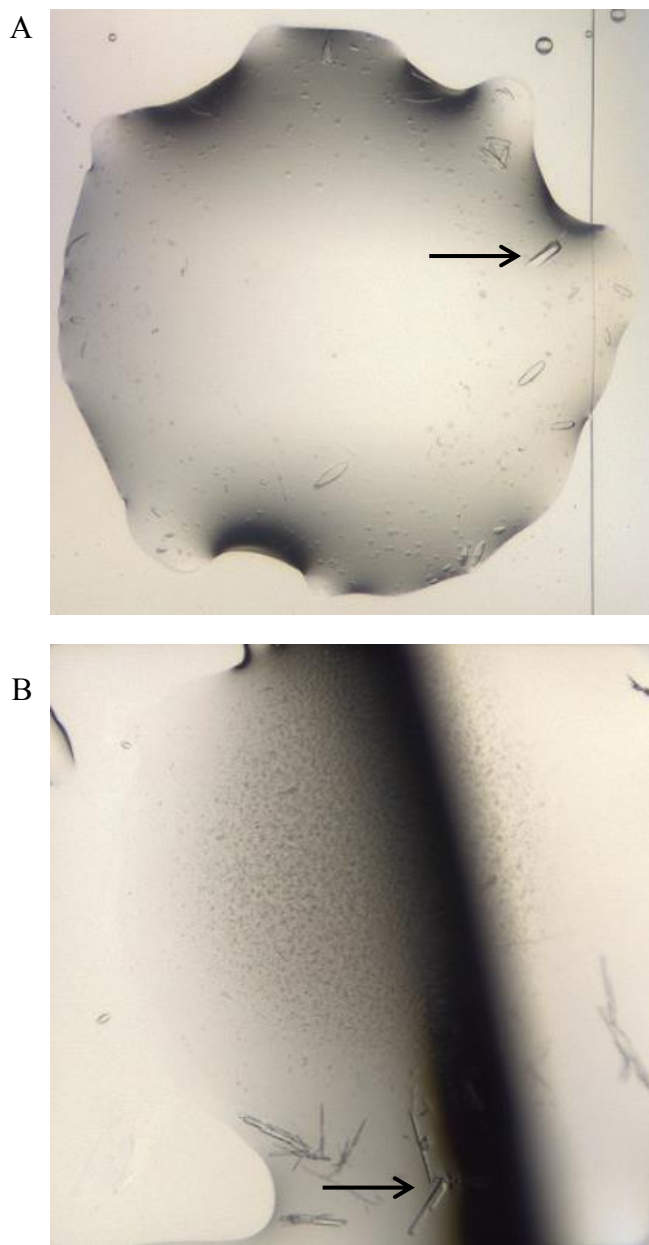


Figure 4.11 Crystal morphology of SeMet-DevE. (A) Wizard III & IV **H5**; (B) JCSG+ **H4**

		5 mg ml ⁻¹ SeMet DevE Concentration 2-propanol			10 mg ml ⁻¹ SeMet DevE Concentration 2-propanol		
Concentration PEG 8000		2.5%	5%	7.5%	2.5%	5%	7.5%
	10%						
	15%						
	20%						

		Concentration (NH) ₂ SO ₄					
Protein: reservoir ratio		0.1 M	0.15 M	0.2 M	0.25 M	0.3 M	0.35 M
	1:1						
	1:2						

		Concentration nucleation inhibitor					
Nucleation inhibitor		0.5%	1%	2%	3%	4%	5%
	Ethylene glycol						
	Glycerol						
	Tascimate						

Figure 4.12 SeMet DevE crystal optimisation strategy (A) testing different concentrations of PEG 8000 and 2-propanol at 5 and 10mg ml⁻¹ SeMet DevE (B) variation of (NH₄)₂SO₄ and protein: reservoir ratio (C) testing different additives at various concentrations to slow down nucleation.

4.3 HutC

4.3.1 Purification

HutC was isolated from *E. coli* Rosetta2 cells induced in auto-induction medium. Purification was by nickel affinity chromatography (5 mL HisTrap FF) followed by gel filtration (Superdex 75 16/60). There are two main elution peaks (**Figure 4.13 A**). These were calculated to correspond to molecular weights of ~108 kDa and ~54 kDa, indicating tetrameric and dimeric forms of HutC respectively. Analysis of fractions eluted in these peaks revealed that Peak 1 (A6-A12) contained higher molecular weight contaminants as well as HutC. The second peak (B1-B7) corresponded to dimeric HutC and contained the majority of HutC overexpressed (**Figure 4.7B**). Fractions B3-B7 were pooled and concentrated to >40 mg ml⁻¹. Samples were flash frozen and stored at -80 °C for further use.

4.3.2 SEC-MALLS

SEC-MALLS data for HutC indicates that the protein to be monodisperse ($M_w/M_n = 1.001$). Peak 1 corresponds to ~98kDa indicating a tetrameric form of HutC. The main peak (peak 2) represents a molecular weight of ~49 kDa corresponding to a dimeric form of HutC (**Figure 4.14**). Again, this is in general agreement with the observations made during gel filtration of HutC.

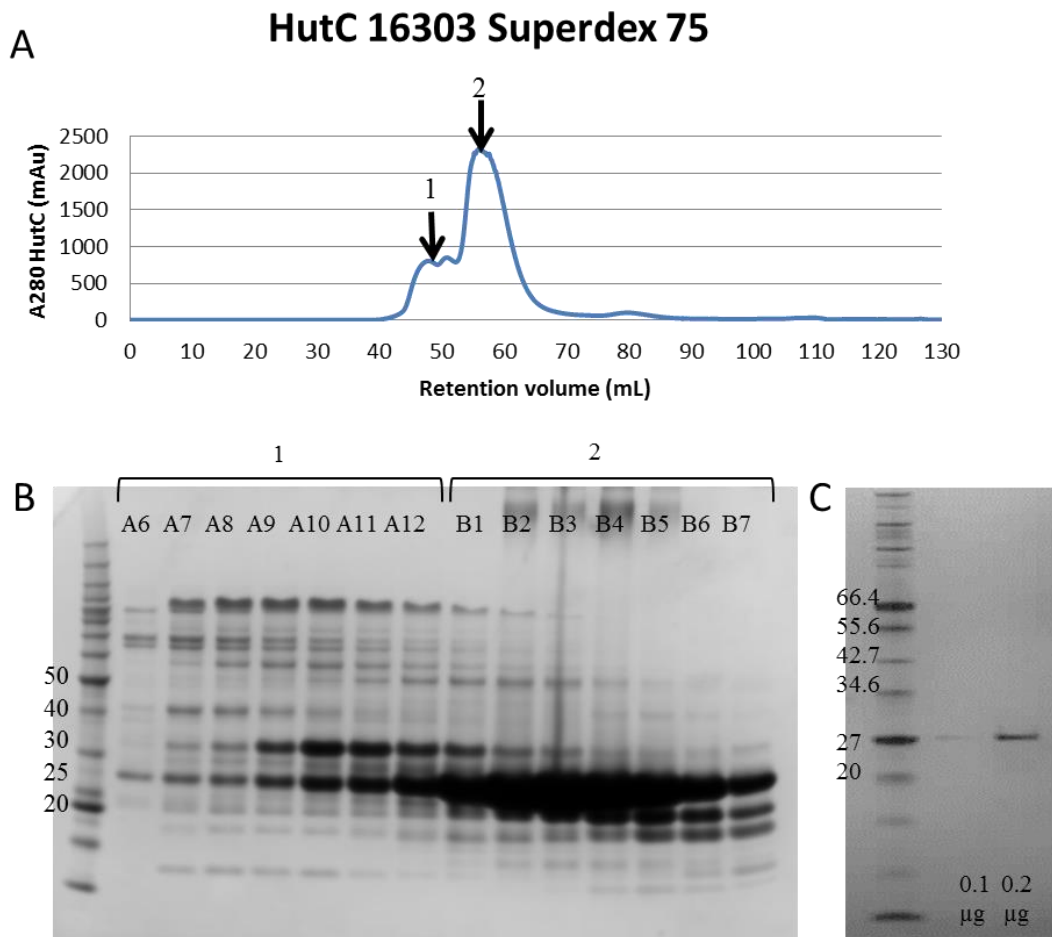


Figure 4.13 (A) Elution profile of HutC overexpression in *E. coli* Rosetta2 induced by auto-induction. Column – HiLoad 16/60 Superdex 75, flow rate - 1 mL/min, buffer - GF buffer 1. **(B)** SDS-PAGE analysis of fraction from multiple elution peaks from gel filtration column (1, 2). **Marker** – Benchmark (Invitrogen). **(C)** 0.1 µg and 0.2 µg HutC showing purity of protein after concentration; **Marker** – Broad Range Protein Marker (NEB).

Gel NuPage 4-12% Bis-tris, 1X SDS-MES running buffer, 200V. Molecular weights shown are in kDa.

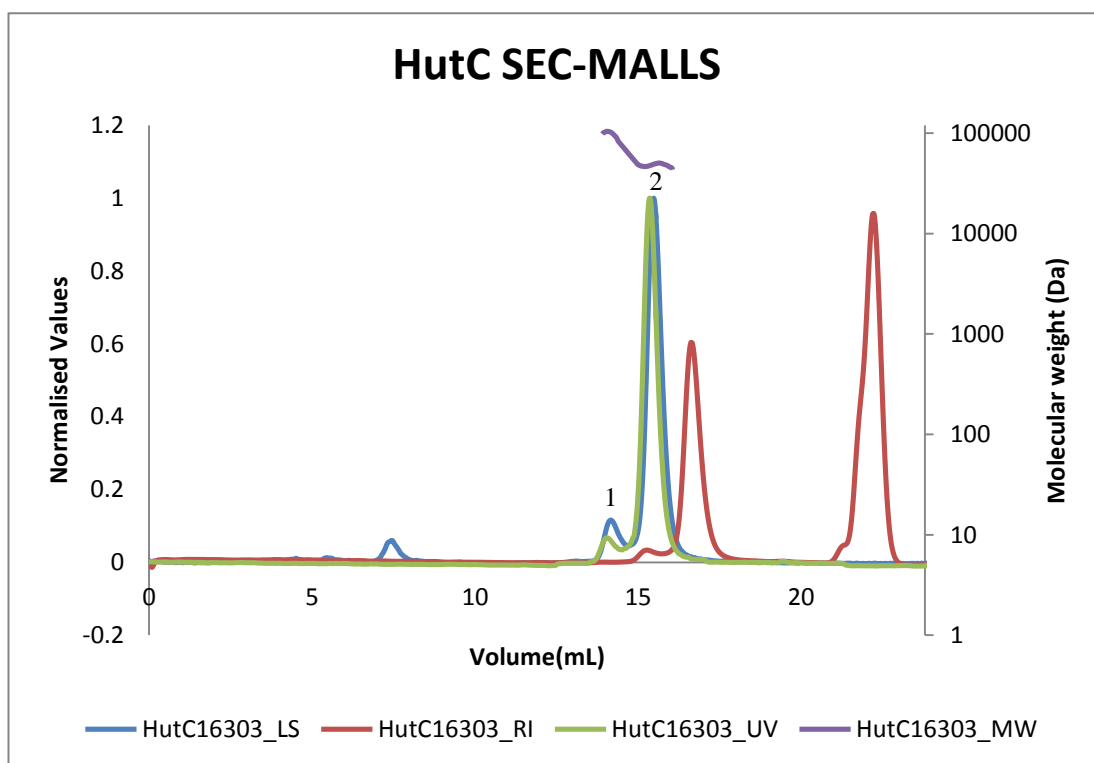


Figure 4.14 Molecular weight vs volume plot of SEC-MALLS of HutC. Two peaks were identified as containing protein. Peak 1 corresponds to ~98 kDa while peak 2 is ~49 kDa. Samples were run at 25 °C with a flow rate of 0.7 ml min⁻¹. RI = refractive index of the sample; LS = light scattering of sample; UV = UV 280 nm of sample; MW = molecular weight

4.3.3 Crystallisation

Native HutC crystallised quite readily. Crystal trials were set up with the 6 standard commercially available screening blocks using protein concentrations of 22 and 40 mg ml⁻¹ using the vapour diffusion sitting drop method. Many promising crystallisation hits were obtained from the HTP crystal screening and are detailed in **Table 4.5**. Various crystal morphologies were also observed (**Figure 4.15**). Of these hits, two conditions showed protein diffraction. These were from crystals grown in Wizard III & IV **H5** condition (0.1 M HEPES/NaOH pH 7.5, 0.2 M ammonium sulphate, 20% (w/v) PEG 8000, 10% (v/v) 2-propanol) that diffracted to 9 Å. Another was grown in Index **C12** condition (15% (v/v) Tascimate pH 7.0, 0.1 M HEPES pH 7.0, 2% (w/v) PEG 3350) that diffracted to 7 Å. Thus both these initial crystal hits were followed up by optimisation trials.

Table 4.5 HutC crystal hits obtained from HTP crystal screening

Screen	Well	Concentration mg ml ⁻¹	Morphology	Time to appear	Diffraction
HutC					
Morpheus*	D3 (0.12 M alcohols, pH 6.5, 30% GOL_P4K)	22	Needles	2 hours	
	E10 (0.12 M ethylene glycols, pH 8.5, 30% EDO_P8K)	22	Needles	2 hours	
	G11 (0.1 M carboxylic acids, pH 8.5, 30% GOL_P4K)	22	Needles	2 hours	
	A9 (0.06 M divalents, pH 8.5, 30% P550MME_P20K)	40	Needle clusters	2 hours	
Wizard III & IV	B4 (0.2 M potassium citrate dibasic)	22	Rhomboid	7 days	
	H5 (0.2 M ammonium sulphate, 0.1 M HEPES pH 7.5, 20% PEG 8000, 10% 2-propanol)	22	Rounded rods	2 hours	9Å
	B10 (0.1 M Tris pH 8.5, 20% reagent alcohol)	40	Micro crystals	2 hours	
	E2 (0.1 M sodium chloride, 0.1 M Tris pH 8.0, 5% MPD, 15% reagent alcohol)	40	Micro crystals	2 hours	
	G9 0.8 M potassium phosphate dibasic, 0.1 M HEPES pH 7.5, 0.8 M sodium phosphate monobasic)	40	Rhomboid	2 days	
	H5 (0.2 M ammonium sulphate, 0.1 M HEPES pH 7.5, 20% PEG 8000, 10% 2-propanol)	40	Teardrops	1 hour	
Index	C12 (15% tacsimate, 0.1 M HEPES pH 7.0, 2 % PEG 3350)	22	Rods	3 days	7Å
PACT	A3 (0.1 M SPG pH 5, 25% PEG 1500)	40	Needles	47 days	

* **Divalent**s (0.3M Magnesium chloride hexahydrate; 0.3M Calcium chloride dehydrate), **Alcohols** (0.2M 1,6-Hexanediol; 0.2M 1-Butanol 0.2M 1,2-Propanediol; 0.2M 2-Propanol; 0.2M 1,4-Butanediol; 0.2M 1,3-Propanediol), **Carboxylic acids** (0.2M Sodium formate; 0.2M Ammonium acetate; 0.2M Sodium citrate tribasic hydrate; 0.2M Sodium potassium tartrate tetrahydrate; 0.2M Sodium oxamate), **Ethylene glycols** (0.3M Diethylene glycol; 0.3M Triethylene glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol), **Buffer 1 pH 6.5**; Imidazole; MES monohydrate (acid), **Buffer 3 pH 8.5**; Tris (base); BICINE, **GOL_P4K** (40% v/v Glycerol; 20% w/v PEG 4000), **EDO_P8K** (40% v/v Ethylene glycol; 20% w/v PEG 8000), **P550MME_P20K** (40% v/v PEG 550 MME; 20% w/v PEG 20000)

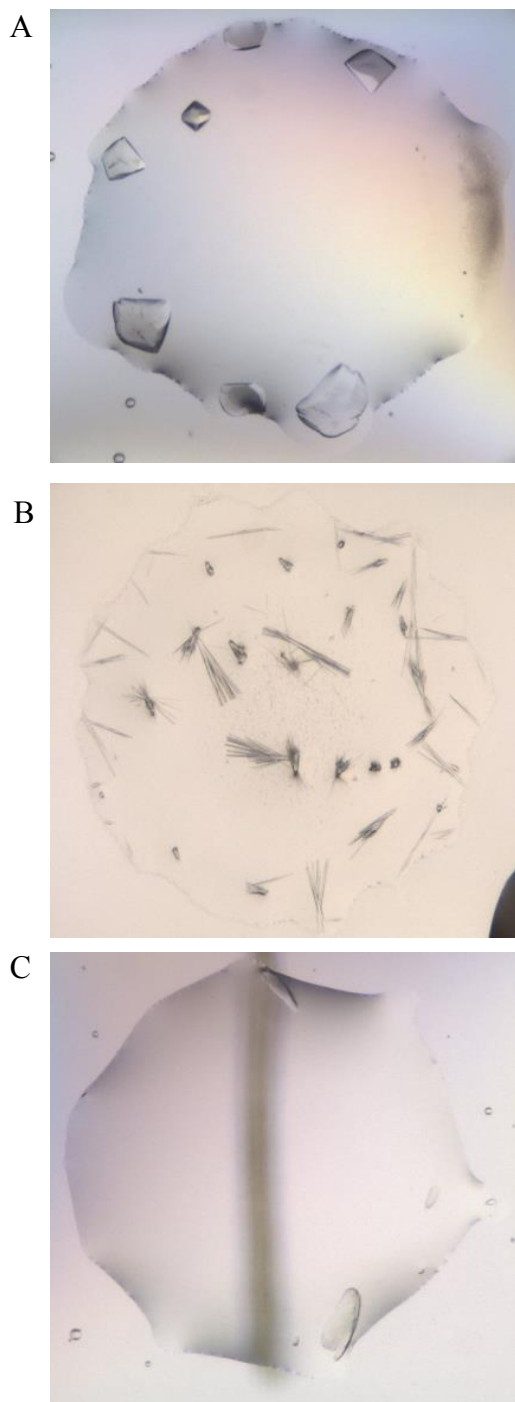


Figure 4.15 Different crystal morphologies of HutC (L-R Wizard III & IV **G9**; Morpheous **A9**; Wizard III & IV **H5**)

Optimisation of HutC crystals

Based on the crystal hits observed for HutC (**Table 4.5**) crystal optimisation was carried out around these conditions. The optimisation strategy was based around the protein: reservoir ratio and alcohol concentrations. The protein concentration was also tested as a high level of nucleation was observed during HTP crystal screening. Reservoir components remained the same as the original screening reservoir (see above). Reservoir volumes were 400 μL with a 2 μL crystallisation drop in tray 2. The crystallisation drop volume was variable in tray 1 that tested different protein: reservoir ratios. The layout of the optimisation trays are shown in **Figure 4.16**.

Optimised conditions were found to yield bigger crystals that also diffracted better with the best initial crystals diffracting up to a resolution of 2.3 \AA . The condition providing these crystals were grown from 25 mg ml^{-1} HutC in a vapour diffusion sitting drop containing 100 mM Tris pH 8, 100 mM NaCl, 13% ethanol, 5% MPD.

		Concentration HutC (mg ml ⁻¹)					
		40	37	33	30	25	22
Protein: reservoir ratio	1:1.5						
	1:1						
	2:1						
	3:1						

		Concentration HutC (mg ml ⁻¹)					
		40	37	33	30	25	22
Volume H2O (μL)	0						
	20						
	50						
	100						

Figure 4.16 Layout of HutC crystal optimisation trays testing various protein concentrations, protein: reservoir ratios and various alcohol concentrations.

4.3.4 Data Collection, structure solution and refinement

Native HutC diffraction data were collected on beamline I03 at DLS using a Pilatus3 6M detector (Dectris). In total 180 degrees of data were collected from 1,200 images using a 0.15° oscillation angle. A snapshot of the crystal mounted at the I03 beamline and a representative diffraction image is shown in **Figure 4.17 (A & B)**. Data were automatically integrated and scaled by the *xia2* pipeline. The native crystal diffracted to a resolution beyond 2.0 \AA and was characterised as belonging to the orthorhombic space group $C222_1$ with unit cell dimensions $a = 73.50 \text{ \AA}$, $b = 92.59 \text{ \AA}$, $c = 154.31 \text{ \AA}$ and $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$. Data collection statistics are shown in **Table 4.6**.

As a result of BLAST searches against the PDB with the HutC amino acid sequence, YvoA (2WV0) was revealed to show the highest homology. Efforts were made to solve the structure using YvoA as model with the program *PHASER* (McCoy, 2007). Unfortunately although different strategies were used, this failed to generate a solution. As crystals of HutC grew readily we immediately tried heavy atom soaking experiments with the view of solving the structure by the heavy atom isomorphous replacement technique (Taylor, 2010).

As a first pass to obtain a HutC heavy atom derivative, we decided to use the JBS Tantalum Cluster Derivization phasing kit (Jena Biosciences) to attempt incorporation of Ta_6Br_{12} in to the HutC crystals (Banumathi *et al.*, 2003). Solid Ta_6Br_{12} was added directly to the crystallisation drop containing the crystals by taking a few crumbs of the solid tantalum bromide cluster with a micro-spatula and carefully mixing into the drop. The solution was then left for 3-4 hours after which

the crystals were observed to turn blue-green in colour. These Ta₆Br₁₂ soaked crystals were then cryo-protected in 25% glycerol.

HutC crystals soaked with Ta₆Br₁₂ diffracted to a resolution of 2.8 Å and subsequent analysis of the reduced data clearly showed an anomalous signal from the tantalum bromide cluster. The crystal and a representative diffraction image are shown in **Figure 4.17 (C & D)**. Data from this crystal was automatically integrated and scaled by the *xia2* pipeline. Ta₆Br₁₂ soaked crystals were characterised as belonging to orthorhombic space group C222₁ with unit cell dimensions $a = 73.69 \text{ \AA}$, $b = 92.31 \text{ \AA}$, $c = 155.2 \text{ \AA}$ and $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 90^\circ$. Data collection and refinement statistics are detailed in **Table 4.6**.

This anomalous signal was used to solve the structure of HutC within the *PHENIX* suite of programs. Initial phases obtained from the Ta₆Br₁₂ soaked crystal were extended against the 1.98 Å resolution native data set. This allowed a significant proportion of the HutC model to be built automatically by the *Autobuild* module of *PHENIX*. However, in the absence of DNA the structure was observed to be highly flexible with only about 80% of the polypeptide chain successfully traced by manual building and a combination of automated model building programs; particularly the *Buccaneer* program within the CCP4 suite (Cowtan, 2006, Winn *et al.*, 2011) along with *Autobuild* within the PHENIX suite (Terwilliger *et al.*, 2008, Adams *et al.*, 2010) were utilised. The problem parts of the structure were centred in particular around the N-terminal DNA binding domains and the linker between the N-terminal and C-terminal domains which due to their high flexibility presented poorly defined electron density, especially for one of the monomers in the crystallographic asymmetric unit. Iterative cycles of manual model building in *COOT* followed by

refinement resulted in a relatively complete model. Refinement statistics are detailed in **Table 4.6**.

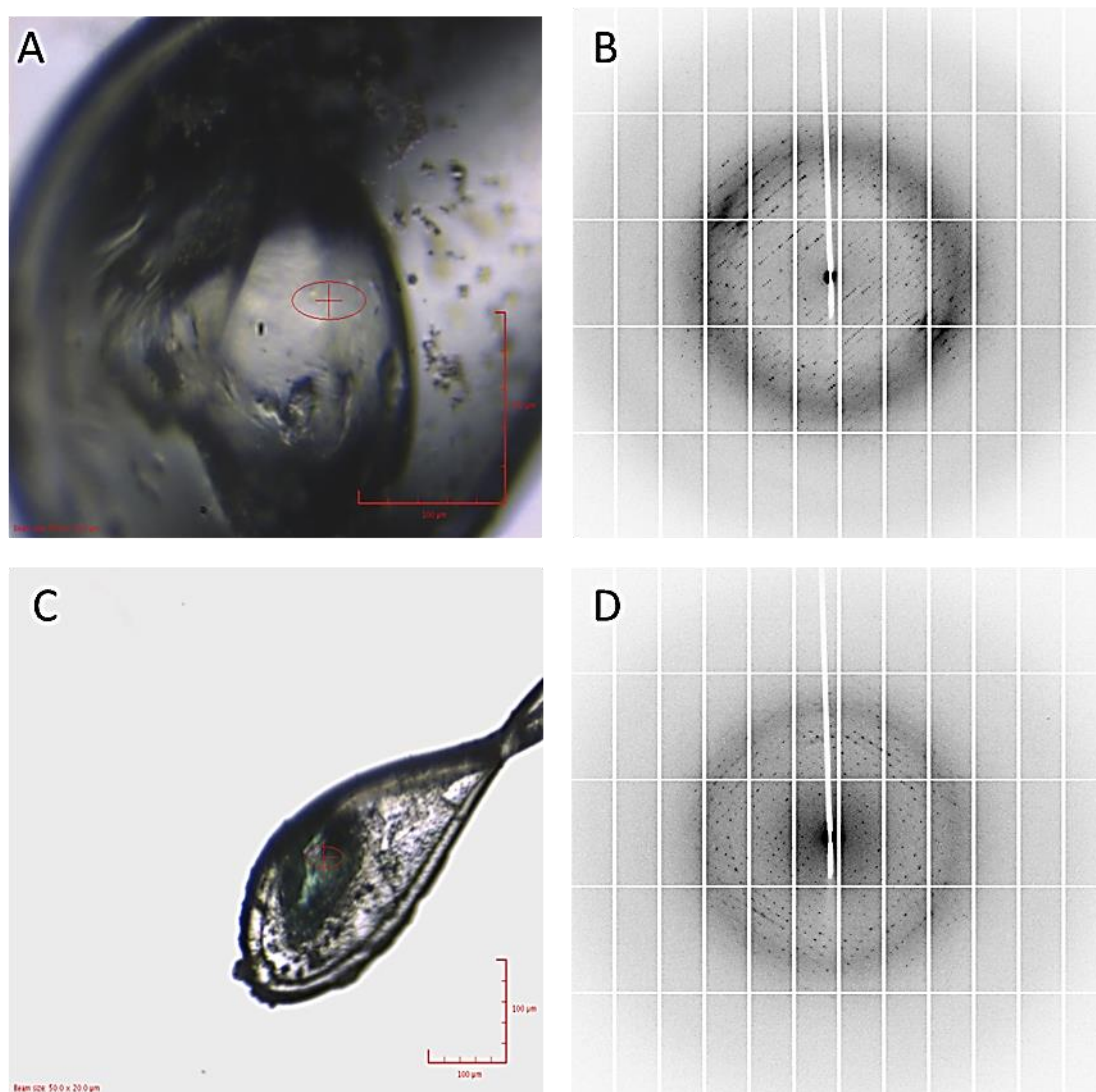


Figure 4.17 (A) A single native HutC crystal mounted on beamline I03 in a nylon loop holder (B) representative diffraction data collected from the crystal in (A). (C) A single HutC crystal soaked in $\text{Ta}_6\text{Br}_{12}$ resulting in a blue-green colour. The crystal shown is mounted on beamline I03 in a nylon loop holder. (D) Representative diffraction data collected from crystal in (C). Data were collected on beamline I03 at Diamond Light Source. Crystals diffracted to a maximum resolution of 1.98 Å and 2.8Å, respectively.

Table 4.6 Data collection and refinement statistics for HutC

HutC	Native	Ta ₆ Br ₁₂ soaked
Data collection		
Space group	C222 ₁	C222 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.50, 92.59, 154.31	73.69, 92.32, 155.20
α , β , γ (°)	90 90 90	90 90 90
Resolution (Å)	53.94-1.98 (2.03-1.98)*	51.73-2.81 (2.88-2.81)*
^a R _{merge}	0.034 (0.683)	0.071 (0.844)
<i>I</i> / σ <i>I</i>	25.3 (2.9)	20.0 (2.5)
Completeness (%)	99.8 (100)	99.9 (99.9)
Multiplicity	5.4 (5.6)	6.5 (6.8)
Wavelength (Å)	0.9795	1.2549
Refinement		
Resolution (Å)	53.94 - 1.98	
No. reflections	36,988	
^b R _{cryst} / R _{free}	0.212 / 0.261	
No. atoms		
Protein	3581	
Water	218	
<i>B</i> -factors (Å ²)		
Overall	41.5	
Water	60.34	
R.m.s deviations		
Bond lengths (Å)	0.019	
Bond angles (°)	1.998	

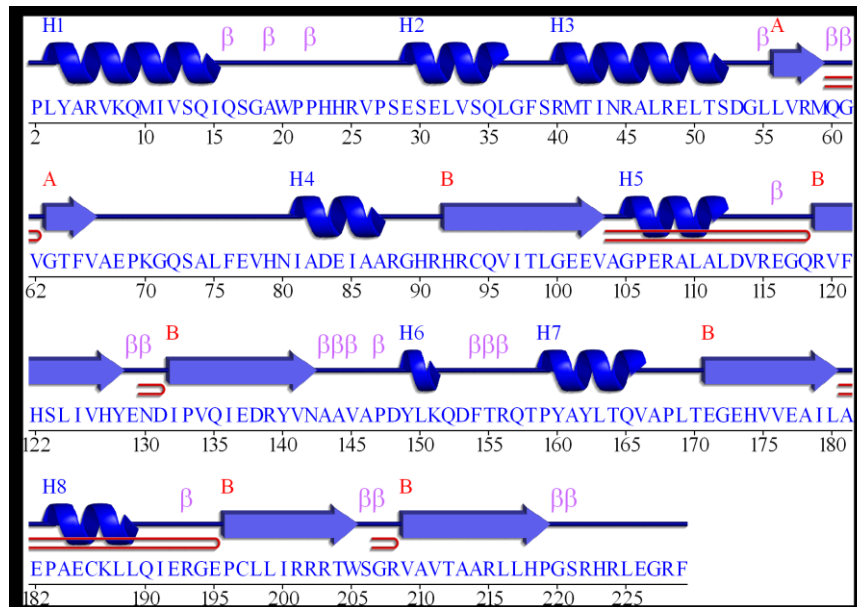
*Values in parentheses are for highest-resolution shell which is indicated in the Resolution row of the table.

^aR_{merge} = $\sum |I_{j\text{HKL}} - \langle I \rangle_{\text{HKL}}| / \sum I_{j\text{HKL}}$ where $I_{j\text{HKL}}$ is the *j*'th observation of reflection HKL and $\langle I \rangle_{\text{HKL}}$ is the average of symmetry related reflections of a unique HKL reflection.

^bR_{cryst} = $\sum | |F_{\text{obs}}| - |F_{\text{calc}}| | / \sum |F_{\text{obs}}|$, F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes. R_{free} as for R_{cryst}, using a random 5% subset of the data, excluded from the refinement

P. aeruginosa HutC crystallized with the HutC homodimer in the crystallographic asymmetric unit. The HutC monomer consists of eight α -helices and 2 β -sheets, one of two strands, in the N-terminal domain, and then another of six strands in the C-terminal domain (**Figure 4.18 A**). A topological diagram on the HutC monomer, which shows the arrangement of the secondary structural elements, is shown in **Figure 4.18 (B)**. The canonical winged helix-turn-helix N-terminal DNA binding domain of HutC is connected to the C-terminal effector-binding domain by a thirteen amino acid long linker. As expected the C-terminal domain displays a chorismate lyase fold as seen in the structure of the *B. subtilis* NagR transcription regulator (see below). A cartoon representation of the HutC dimer is shown in **Figure 4.19**.

(A)



(B)

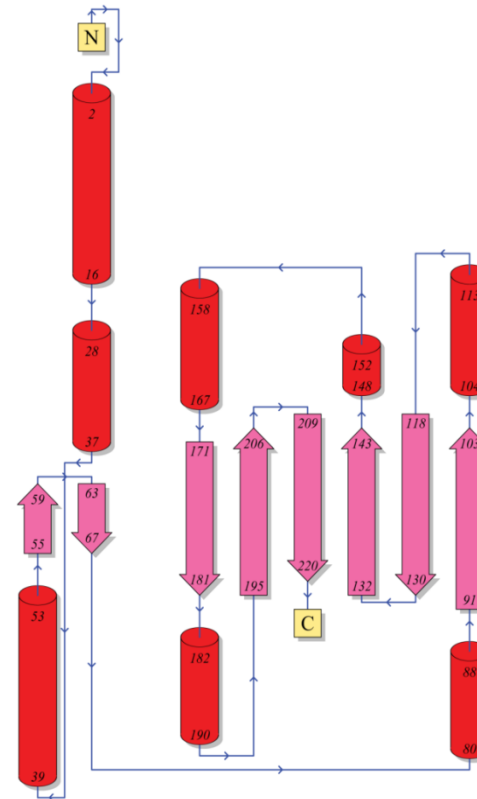



Figure 4.18 (A) Secondary structure wiring diagram of HutC. Secondary structural elements (Helices and β -strands are indicated above the HutC sequence with helices labelled H1,H2...and strands by their sheets, A and B in this case. Beta turns are indicated by β and β hairpins by  (B) Topology diagram of HutC. Analysis and figures were produced with the program PDBsum (de Beer *et al.*, 2014).

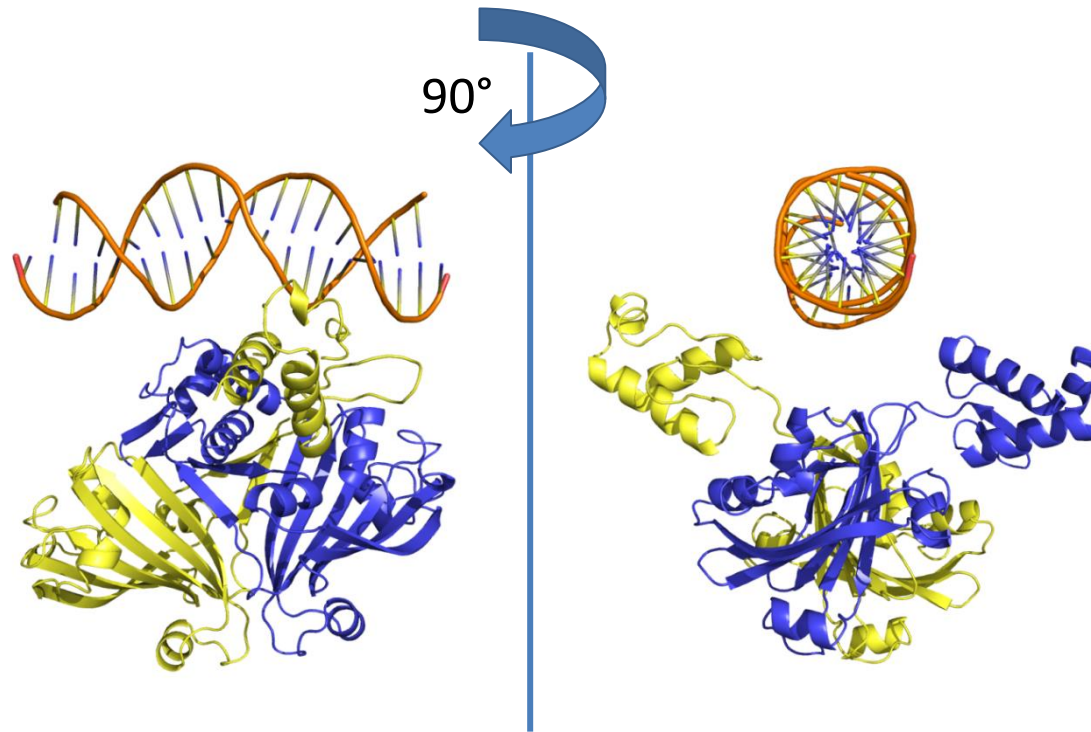


Figure 4.19 Cartoon representation of the *P. aeruginosa* HutC (HutC) dimer with the monomers coloured in blue and yellow. Two views 90° apart are shown as indicated in the figure. The 19mer palindromic dsDNA bound to the *Bacillus subtilis* NagR (NagR) is also shown in cartoon representation based on superposition of the C-terminal domains of the NagR and HutC structures.

The C-terminal β -sheet is involved in dimer formation and the dimerization mode is the same as that observed in structures determined of individual effector domains e.g. the C-terminal domain of PhnF from *E. coli* which was the first structural data available for a HutC subfamily member (Gorelik *et al.*, 2006) as well as the full length GntR/HutC transcription regulator NagR (Fillenberg *et al.*, 2015). The HutC monomer subunits were superposed as well as the C-terminal domains of PhnF and NagR and the resulting root-mean-square deviations are summarized in **Table 4.7**. Visualisation of the overall similarity in the C-terminal domains of HutC and NagR through colouring of global structural differences using the structural alignment program *ProSMART* (Nicholls *et al.*, 2012) shows the overall fold and core β -sheet structure to be very well conserved with the main differences in structure for helices H5-7 and the β -strand connecting H7 and H8 (**Figure 4.20**).

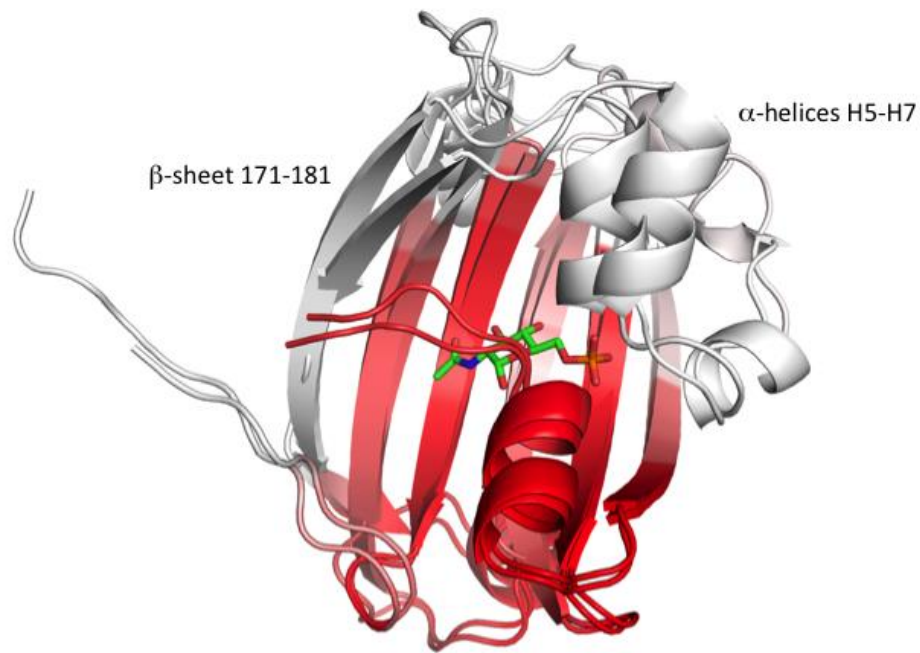


Figure 4.20 Structural comparison of the C-terminal of HutC with the C-terminal of NagR (4UOV). Global conformation changes between the two structures were generated by *ProSMART* and visualised in *PyMOL*. The residues are coloured according to the similarity of their local coordinate frames. Residues that relate closely to the rigid substructure are coloured red, fading to white for regions that adopt a different global conformation.

Table 4.7 Root-mean-square deviations (r.m.s.d) of various superpositions of HutC and other GntR/HutC structures based on C- α atoms

Model comparison	Reference atoms	R.m.s.d. (Å)
HutC monomers	C-terminal C- α atoms only (residues 72 - 229)	0.84
HutC: NagR	C-terminal C- α atoms only (residues 76 – 227)	1.8
HutC: PhnF	C-terminal C- α atoms only (residues 98 – 227)	1.8

Unfortunately, during this project the structures of NagR, the HutC homologue from *B. subtilis* were published in complex with the operator DNA and in complex with putative effector molecules glucosamine-6-phosphate and N-acetylglucosamine-6-phosphate (Fillenberg *et al.*, 2015). As our preliminary structural data showed the structures to be globally the same, despite the large conformational changes induced in the N-terminal domains by effector molecules and DNA binding, the focus of the structural studies was moved from HutC to DevE. However, the structure of NagR in complex with DNA allowed the modelling of bound DNA to the *P.aeruginosa* HutC structure determined here. The structure clearly shows that the HutC structure has been determined in a conformation primed to bind DNA and consolidates the observations made by Fillenberg *et al.* (Fillenberg *et al.*, 2015). **Figure 4.21** compares the structures of HutC with the effector molecule induced structure of NagR as well as the NagR: 19mer DNA complex.

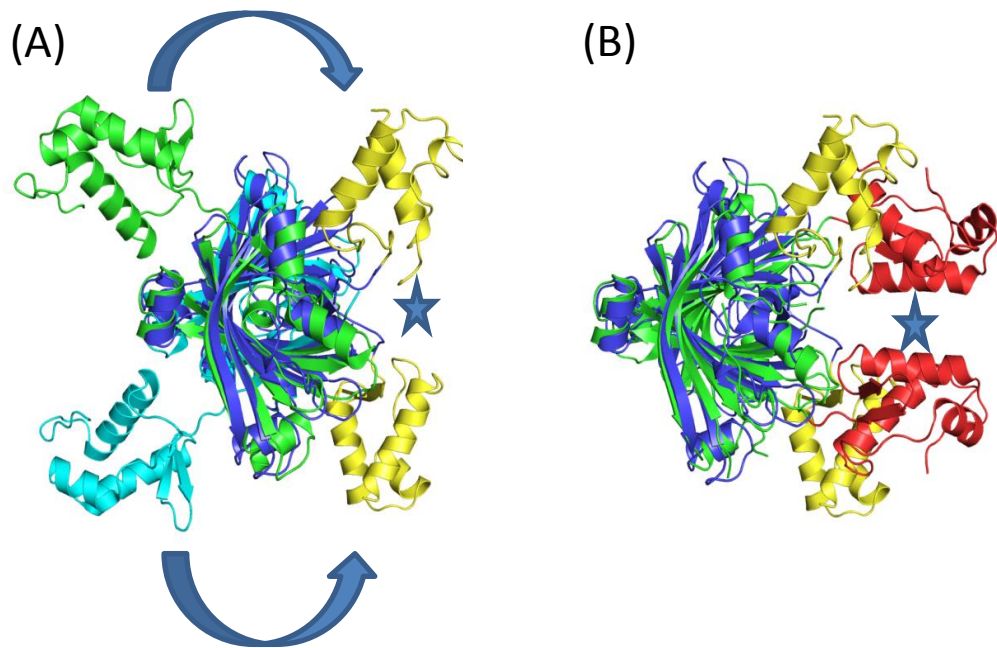


Figure 4.21 (A) Superposition of the NagR: N-acetylglucosamine-6-phosphate complex (PDB ID 4U0W) onto HutC showing the large conformation change induced by the effector molecule on the arrangement of the N-terminal DNA binding domains. HutC N- and C- terminal domains are coloured yellow and blue, respectively. The monomer subunits of the NagR homodimer are coloured in green and cyan. The DNA binding site is indicated by a blue star. (B) Superposition of the NagR: 19mer DNA complex (PDB ID 4WWC) onto HutC. The N-terminal DNA binding domains make a dramatic shift as indicated by the arrows in (A) and shows that the HutC structure determined here has been captured in its non-induced state with the N-terminal domains of HutC in a more open conformation than that of the NagR:19mer DNA complex

4.4 Gp26

4.4.1 Purification

Gp26 was purified by nickel affinity chromatography (5 mL HisTrap FF) followed by gel filtration (Superdex 7516/60). The elution profile contained 3 peaks which corresponded to molecular weights of ~400 kDa, ~36 kDa and ~20kDa (**Figure 4.22A**). This would indicate that Gp26 was present as a dimeric and a monomeric form (35.6 kDa. and 17.8 kDa, respectively). The 400 kDa peak was attributed to aggregated protein. SDS-PAGE analysis of peak 1 (A7-A9), peak 2 (B5-B9) and peak 3 (C5-C11) revealed Gp26 was present in all elution peaks further supporting the theory that peak 1 was aggregated protein (**Figure 4.22B**). Fractions C5-C11 (peak 3) were pooled and concentrated as they contained the major form of Gp26. Samples were flash frozen and stored at -80°C for further use.

4.4.2 SEC-MALLS

Analysis of Gp26 by SEC-MALLS again revealed two peaks (**Figure 4.23**). Peak 1 corresponds to a molecular weight of ~49 kDa. Peak 2 is a molecular weight of around 30 kDa. From this data, the oligomeric state of Gp26 is less clear, however peak 1 is probably dimeric and peak 2 most likely is monomeric forms of Gp26 which would be in agreement with gel filtration data. Protein was monodisperse in the monomeric peak but less so in the dimeric peak (Mw/Mn 1.001 and 1.408 respectively).

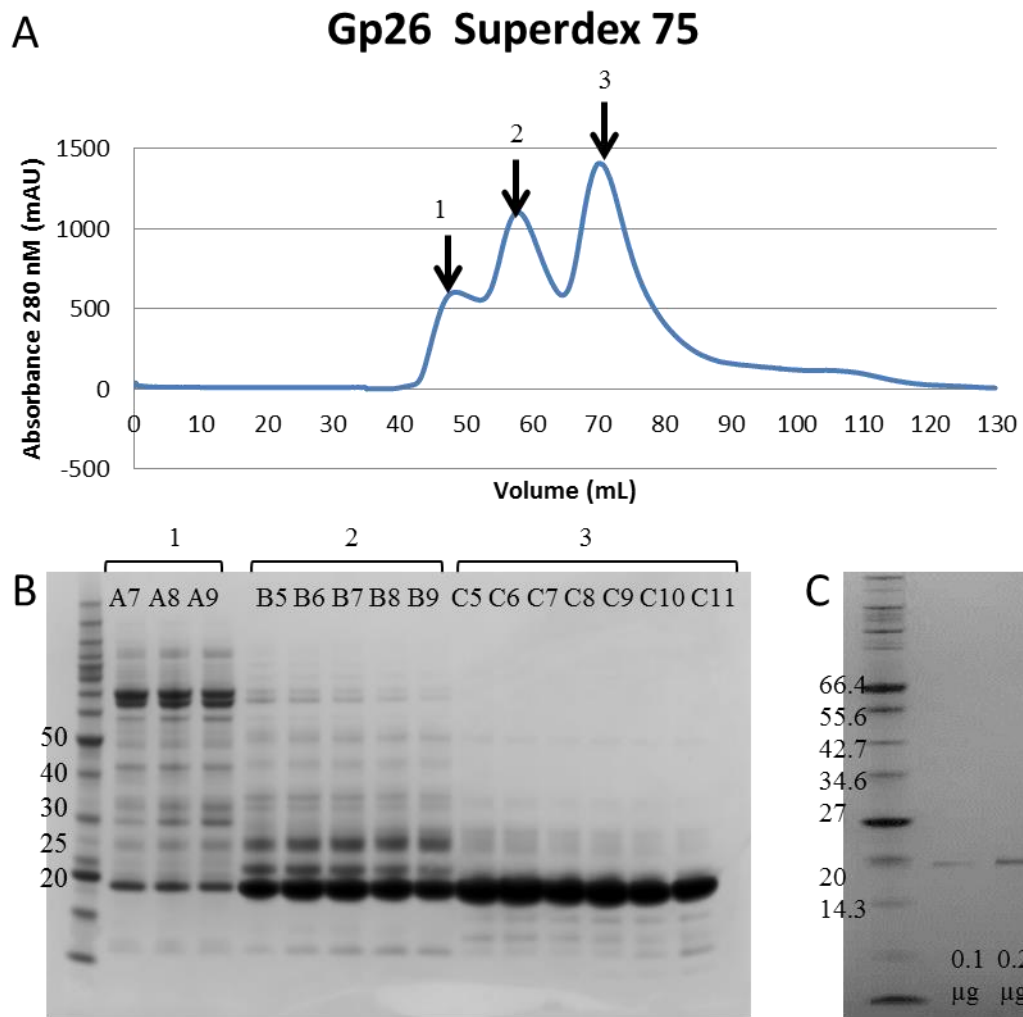


Figure 4.22 (A) Elution profile of Gp26 overexpression in *E. coli* Rosetta2 induced by auto-induction. Column – HiLoad 16/60 Superdex 75, flow rate - 1 ml min⁻¹, buffer - GF buffer 2. (B) SDS-PAGE analysis of fraction from multiple elution peaks on gel filtration column (1-3). (C) 0.1 µg and 0.2 µg Gp26 showing purity of protein after concentration. Gel NuPage 4-12% Bis-tris, 1X SDS-MES running buffer, 200V. Marker – (B) Benchmark (Invitrogen), (C) Broad Range Protein Marker (NEB). Molecular weights shown are in kDa.

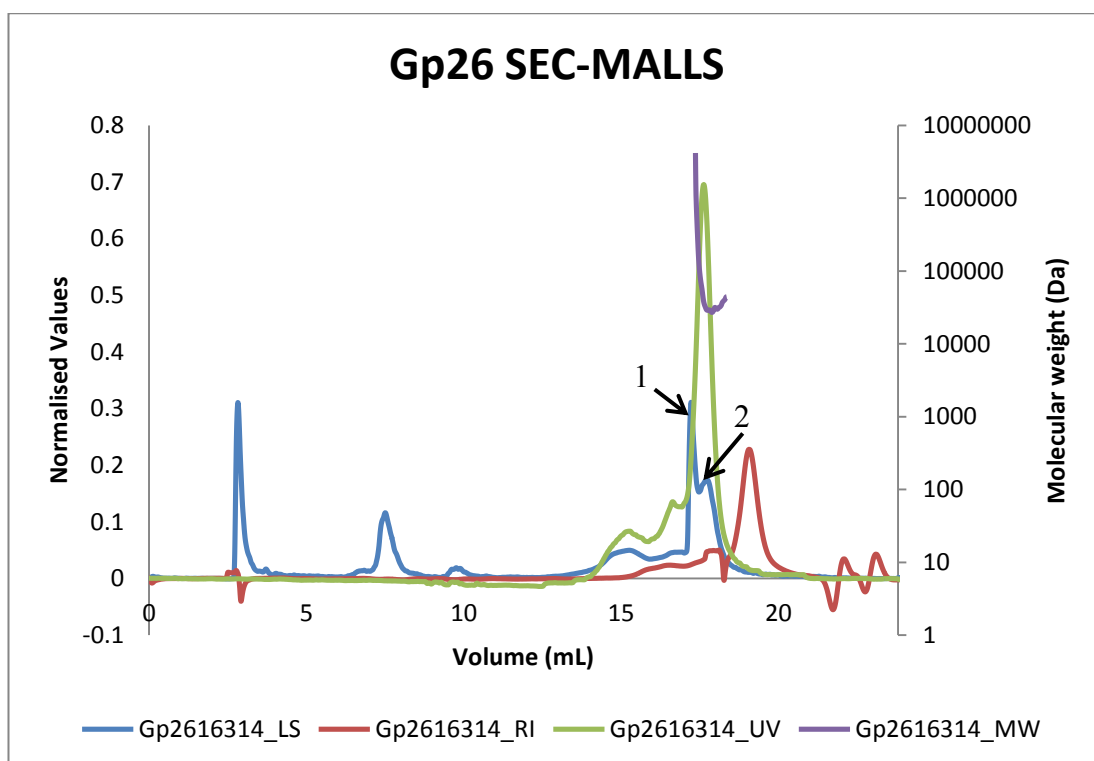


Figure 4.23 Molecular weight vs volume plot of SEC-MALLS of Gp26. The main peak corresponds to a molecular weight of ~70 kDa. Samples were run at 25°C with a flow rate of 0.7 ml min⁻¹. RI = refractive index of the sample; LS = light scattering of sample; UV = UV 280 nM of sample; MW = molecular weight

4.4.3 Crystallisation

Crystallisation trials for Gp26 were performed as for the other proteins, using as a start point, 6 commercially available screens using two different concentrations (26 and 28 mg ml⁻¹). Despite testing 1192 conditions only one crystal hit was observed in JCSG+ condition **D3** (0.1 M Na phosphate pH 6.2, 0.2 M NaCl, 50% v/v PEG 200; **Figure 4.24 A**). This crystal was grown from 28 mg ml⁻¹ in a leaf-like pattern that began to appear after 5 hours. A portion of this crystal was tested for diffraction and diffracted as protein to 4Å on beamline I03 at Diamond Light Source. The diffraction data is of very poor quality but confirms these initial crystals are protein (**Figure 4.24 B**). While this initial hit is very encouraging, further optimisation of crystals is required to obtain better quality diffraction data for future analysis.

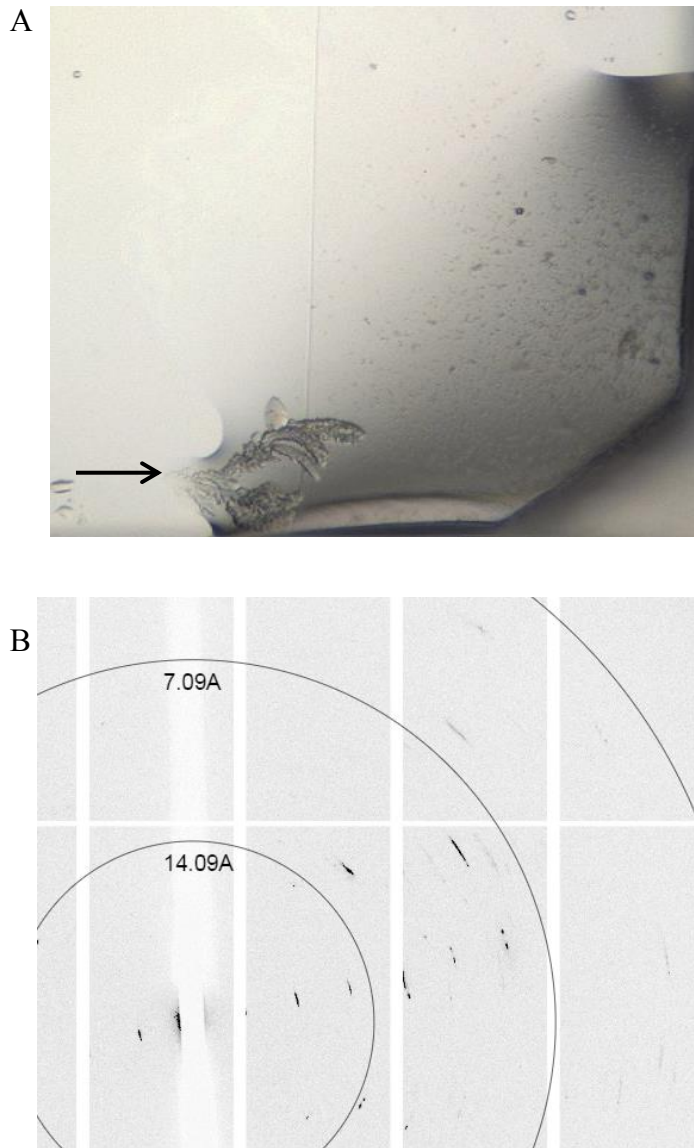


Figure 4.24 (A) Crystal morphology of Gp26 crystal grown in JCSG+ condition **D3** (0.1 M Na phosphate pH 6.2, 0.2 M NaCl, 50% v/v PEG 200)

(B) Diffraction image obtained from the crystal in **(A)**

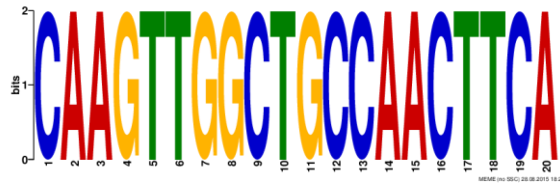
Chapter 5 Functional analysis of the GntR-like proteins HutC, DevA and DevE

Experimental protein function analysis was an integral part this project. Functional analysis of proteins, including EMSAs and complementation of mutant PA14 strains has led to a better understanding of the DNA binding characteristics of HutC, DevA, DevE and Gp26. Although the mechanisms of DNA and effector molecule binding in HutC are largely known, we have used this protein as validation for methodologies to examine DevA, DevE and Gp26.

5.1 Examination of protein-DNA binding by EMSA

Protein-DNA interactions were analysed using electrophoretic mobility shift assays (EMSA). It is known that GntR-like proteins generally bind to relatively short inverted repeat or directed repeat sequences and are often auto regulatory (Rigali *et al.*, 2002). The 250 bp (approximately) region upstream of *hutC*, *SCO4190* and *SCO4188* were analysed to find potential inverted repeat promoter regions. The DNA consensus sequence for the HutC subfamily is known to be 5'-GT(X)TA(X)AC-3', where X generally represents A or T residues (Rigali *et al.*, 2002). The DevA subfamily, in contrast, does not have a defined DNA consensus sequence. DevA and DevE upstream sequences were analysed using the EMBOSS bioinformatics software suite (Rice *et al.*, 2000). In particular, the 'einverted', 'palindrome' and 'equicktandem' programs were utilised. The einverted and palindrome programs find inverted repeats in nucleotide sequences while equicktandem finds tandem repeats with nucleotide sequences. The equicktandem program was the only program which returned results and only for the DevE upstream region (maximum repeat size, 30; threshold score, 10). These were short

sequences that have potential to be promoter regions. The DevA upstream repeats (USR) were obtained by aligning with the potential DevE promoter sequences and very similar sequences were found. These are detailed in **Section 5.1.2**. The DevE USR was subject to BLAST and several *Streptomyces* species were found to have a similar motif. *S. ambofaciens* and *Streptomyces* species PBH53 are of particular note because these sequences are found upstream of a DevE homolog and putative GntR regulator respectively. The alignment and predicted motif from MEME (Bailey & Elkan, 1994) are shown in **Figure 5.1**.



DevE22bp	-TCAAGTTGGCTGCCAACTTCAC
DevA19bp	---AAGTTGGCAACCAACTCTC-
Streptomyces ambofaciens ATCC	GTGAAGTTGGCAGCCAACCTTGA-
Streptomyces lividans TK24	GTGAAGTTGGCAGCCAACCTTGA-
Streptomyces coelicolor A3(2)	-TCAAGTTGGCTGCCAACTTCAC
Streptomyces sp. PBH53	--CAAGTTGGCTGCCAACTTCAC

Figure 5.1 Alignment of a 22 bp potential GntR-like promoter region in *Streptomyces* species and their predicted consensus sequence. Alignment by MEGA 6.0. Motif prediction MEME Suite 4.10.1.

5.1.1 HutC-DNA binding is diminished by intermediates of histidine utilisation

Purified HutC protein was tested in an EMSA assay using a fragment of DNA identical to its own upstream region. HutC was observed to bind to the 237 bp region upstream of the *hutC* gene thereby demonstrating the auto-regulatory nature of the protein. The DNA fragment used in the assay contained the consensus sequence 5'-GTATATAC-3'. Protein-DNA interactions appear to happen at concentrations greater than 100 nM HutC, however, a super-shift is observed at very high protein concentrations (**Figure 5.2A**). The K_D is estimated to be between 200 nM and 400 nM from the EMSA (**Figure 5.2B**). The K_D estimated from the EMSA is apparently relatively high when compared to other HutC family members (**Table 5.1**). Using longer DNA fragments sometimes may result in non-specific binding of the protein to DNA; there is a large difference between binding constants of 15 bp and 226 bp fragments in **Table 5.1**, although the methods used to obtain these values are different. Moreover, increased specificity of binding may be observed with smaller DNA fragments, however the stability of the complex may be compromised (Tucker *et al.*, 2010). High K_D values may be obtained because HutC may be binding non-specifically as the DNA fragment is long (237 bp).

Table 5.1 K_D of other HutC family members bound to dsDNA

Protein	Uniprot ID	Sequence identity	DNA length (bp)	K_D	Method	Reference
HutC (PA14)	Q02ER1	-	237	200 – 400 nM	EMSA	This work
HutC (<i>Brucella abortus</i>)	Q2YIL3	33.5%	226	0.75 nM	EMSA	(Sieira <i>et al.</i> , 2010)
NagR (<i>Bacillus subtilis</i>)	S6FUZ8	23.5%	15	14.5 nM (DBD only)	SPR	(Fillenberg <i>et al.</i> , 2015)
YvoA (<i>Bacillus subtilis</i>)	O34817	25.5%	18	131.2 nM	ITC	(Resch <i>et al.</i> , 2010)

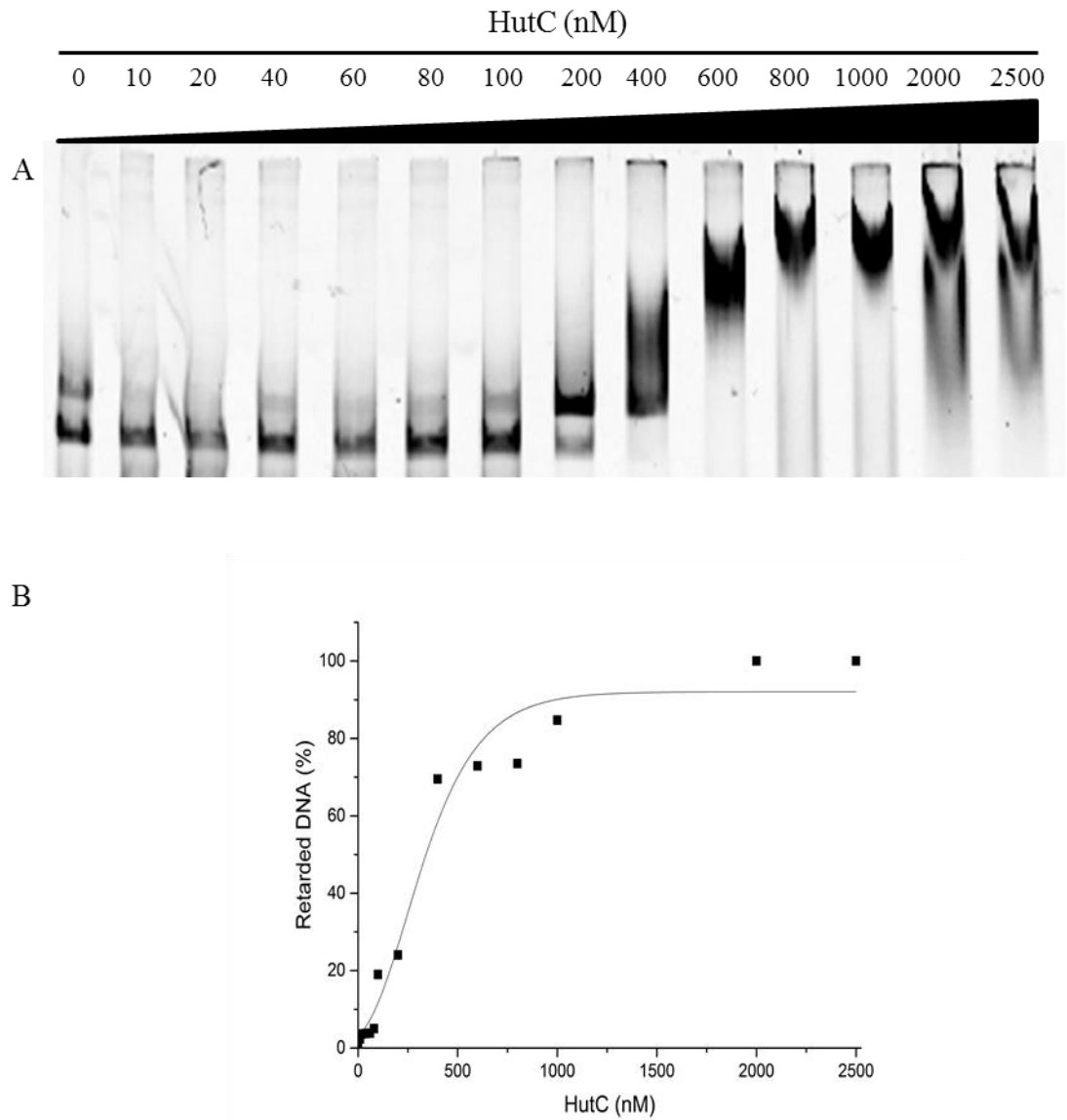


Figure 5.2(A) EMSA showing HutC binding 1.5 ng of Cy5 labelled DNA (233bp). Gel 6% acrylamide, 120V. Imaged on Typhoon 9200. **(B)** EMSA was quantified using GelQuant.NET software (BiochemLab Solutions). The total amount of retarded DNA is plotted as a percentage of fluorescence present in each lane.

HutC is a repressor of gene transcription and thus is bound to DNA in the absence of an effector molecule. The effector molecule for HutC is known to be cis-urocanic acid (Magasanik, 1976), the first intermediate in the histidine utilisation pathway, produced by the action of histidase (*hutH*). It follows, therefore, that addition of urocanic acid to the EMSA binding reaction should be able to bind to HutC and no protein-DNA complex should be formed. EMSA reactions were set up using 200 nM and 400 nM HutC, the apparent upper and lower levels of estimated K_D . DNA that did not have the Cy5 fluorophore linked to it was used as a control (**Figure 5.3A**, lane 3). The unlabelled competition DNA inhibited binding of labelled DNA although not fully. Urocanic acid diminished the HutC-DNA interaction, again not fully, however at higher concentrations of urocanic acid (1000 and 2000 nM) the DNA shift was observed to be greater. This is in contrast to previous literature shows that with 0.55 nM HutC (*B. abortus*), which showed an almost complete inhibition of protein-DNA binding with 50 μ M cis-urocanic acid (Sieira *et al.*, 2010). Interestingly, using 400 nM HutC, a reduction in the protein-DNA interaction was observed with 400 nM imidazole and 400 nM L-histidine (**Figure 5.3A**). Conversely, no reduction in the protein-DNA interaction was observed using 400 nM, 2000 nM or 4000 nM urocanic acid (**Figure 5.3B**). The binding pocket of HutC may be flexible (Fillenberg *et al.*, 2015, Resch *et al.*, 2010) and thus recognises the imidazole ring structure of urocanic acid, imidazole and histidine although this does not explain why no shift is observed at higher concentrations of urocanic acid.

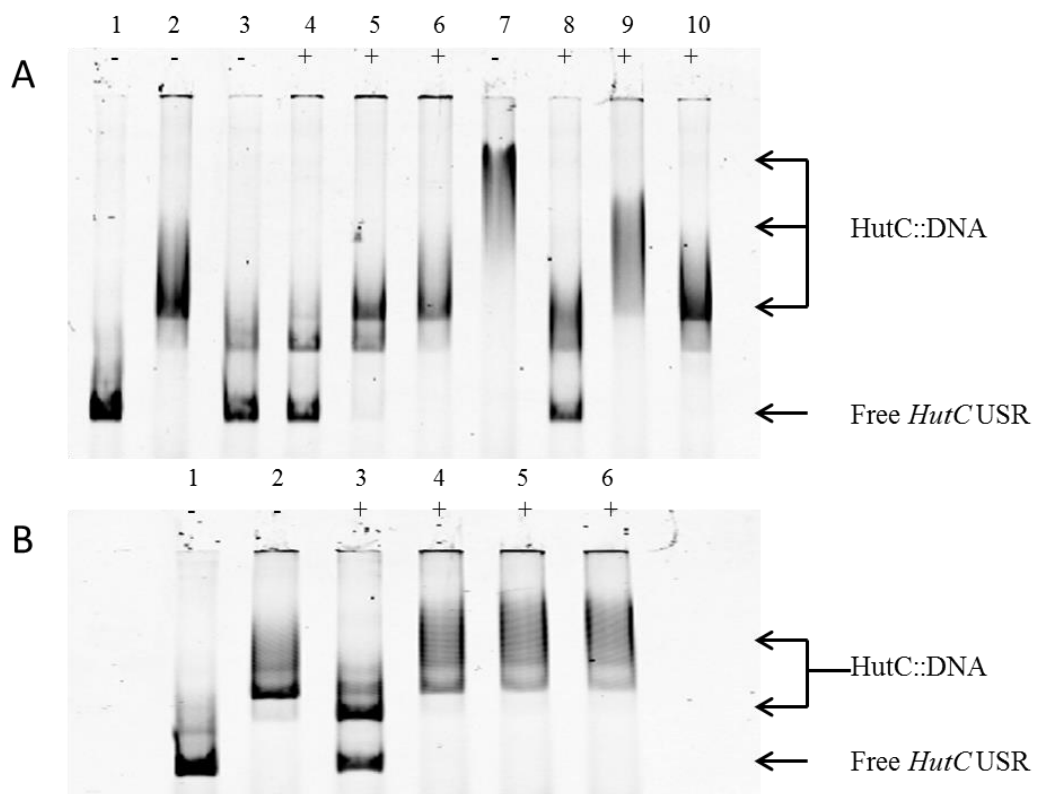


Figure 5.3 (A) EMSA showing HutC-DNA interaction with intermediates of the histidine utilisation pathway. (+) denotes the lanes which contain intermediates of the histidine biosynthesis pathway. Lanes are as follows **1)** 0 nM HutC; **2)** 200 nM HutC; **3)** 200 nM HutC+ 15 ng unlabelled DNA; **4)** 200 nM HutC + 200 nM urocanic acid ; **5)** 200 nM HutC + 1000 nM urocanic acid; **6)** 200 nM HutC + 2000 nM urocanic acid; **7)** 400 nM HutC; **8)** 400 nM HutC + 15 ng unlabelled DNA; **9)** 400 nM HutC + 400 nM imidazole; **10)** 400 nM HutC + 400 nM L-histidine.

(B) EMSA of HutC-DNA interaction with 400 nM HutC **1)** 0 nM HutC; **2)** 400 nM HutC; **3)** 400 nM HutC + 15 ng unlabelled DNA; **4)** 400 nM urocanic acid; **5)** 2000 nM urocanic acid; **6)** 4000 nM HutC. Gel 6% acrylamide, 120V. Imaged on Typhoon 9200.

5.1.2 DevE binds to the promoter region of DevA but not vice versa

5.1.2.1 DevA binding studies

EMSAs were carried out on DevA using the protein concentrations that had been established from the HutC EMSA. No binding was observed for nanomolar concentrations (≤ 1000 nM) of DevA to the promoter region (not shown) however; micromolar concentrations were observed to cause shift in the 233 bp upstream region of DevA (**Figure 5.4A**).

DevA and *DevE* arose from a gene duplication event and both play a role in the correct development of *S. coelicolor*. Although expressed at different temporal stages of the *S. coelicolor* lifecycle, both affect the correct sporulation of the bacteria (Clark & Hoskisson, 2011). It seems reasonable to suggest that as both arose from the same ancestral gene, appear to perform a similar function and share 56.7% identity at the protein level, they may have similar promoter sequences and thus may bind to each other's promoter regions.

Analysis of the upstream regions of DevA (233 bp) and DevE (240 bp) were each revealed to have two potential promoter regions. The potential core sequences are underlined in **Table 5.2**. They are imperfect, inverted tandem repeats that are slightly different to the general pattern in the GntR family. The DevA sub family is the most recently discovered so perhaps other patterns are possible.

Table 5.2 Potential promoter regions of DevA and DevE

Name	Sequence	Length (bp)
DevA_1	5' AAGTTGGCAACCAACTCTC 3'	19
DevA_2	5' CACTTGGCTACTACCTATA 3'	19
DevE_1	5' AAGTTGGCTGCCAACTTC 3'	18
DevE_2	5' CGCTTGGAAGCGTCTCATA 3'	19

EMSAs were used to assess if DevA could bind to the 240 bp DevE upstream promoter region. No binding was observed at any concentration of DevA to the DevE upstream fragment (**Figure 5.4B**). This would indicate that there is specificity in the HTH of the DevA that recognises a different DNA sequence.

Due to the micromolar concentrations of DevA used during this experiment, no K_D value has been estimated for this binding reaction due to potential non-specific binding of protein to the large DNA fragment.

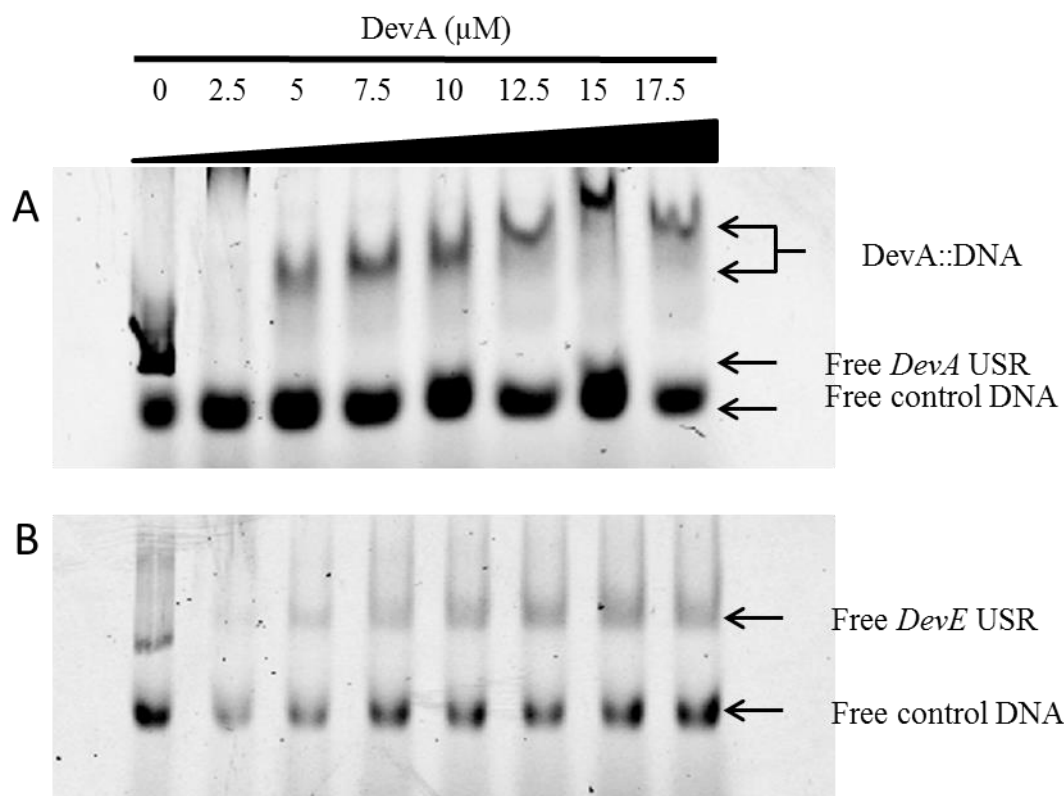


Figure 5.4 (A) EMSA was carried out with varying concentrations of DevA and upstream DNA fragment (233 bp). (B) EMSA of varying concentrations of DevA with the upstream DNA fragment of DevE (240 bp). Gel 6% acrylamide, 120V. Imaged with Typhoon 9200.

5.1.2.2 DevE binding studies

DevE binding interactions to the 240 bp upstream region of the gene were also assessed by EMSA. Like DevA, no binding was observed at sub-micromolar concentrations of protein. The similar nature of these two proteins allowed for the DevA EMSA to be used as the basis for DevE so the same concentrations of protein were used.

DevE is observed to bind to its own promoter region (**Figure 5.5A**). Even at 2.5 μM , a large shift is evident, with super shifts at higher concentrations. Therefore the binding constant is probably somewhere between 1 μM and 2.5 μM as no shift was seen up to 1 μM when tested previously (not shown). Again, SPR or ITC would most likely to be the best way forward to obtain a K_D in this case.

Furthermore, when binding is tested against the DevA promoter, DevE is also observed to bind up to 2.5 μM with super shifts appearing at higher concentrations. This indicates that as *devE* arose from duplication of *devA* perhaps DevE could be postulated to have retained some specificity for the promoter region of DevA.

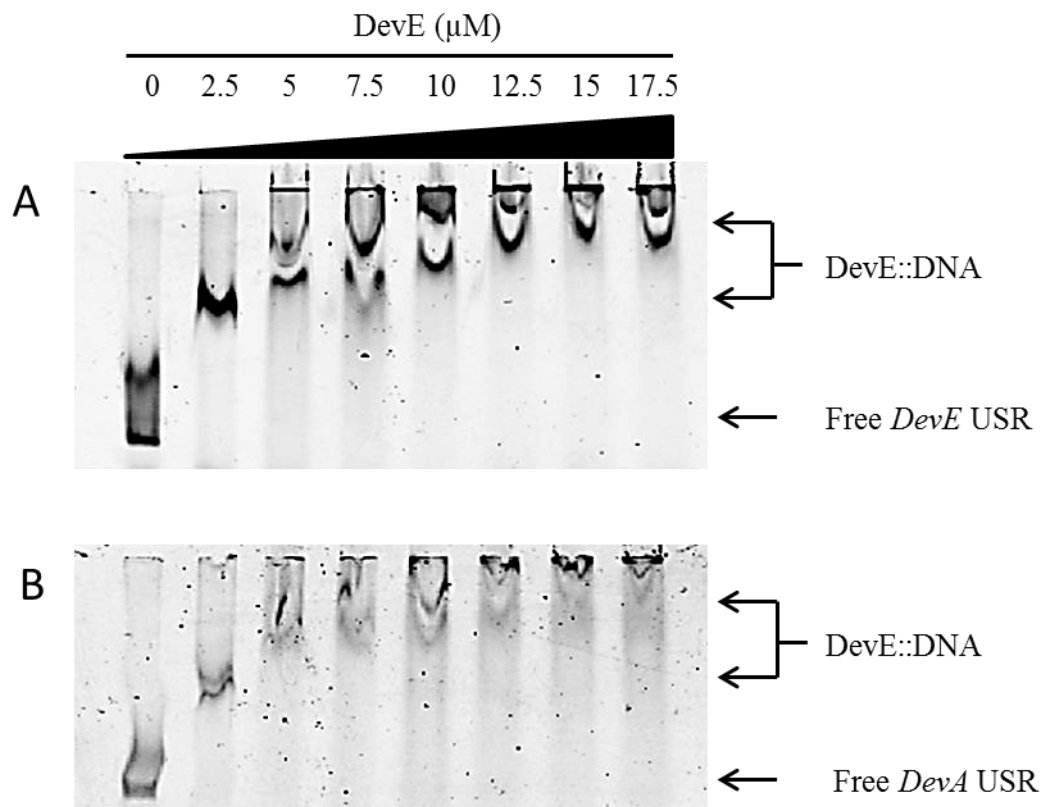


Figure 5.5 (A) EMSA was carried out with varying concentrations of DevE and upstream DNA fragment (240 bp). (B) EMSA of varying concentrations of DevE with the upstream DNA fragment of DevA (233 bp). Gel 6% acrylamide, 120V. Imaged with Typhoon 9200.

5.1.2.3 Further refinement of DevE promoter sequences

For further crystallisation studies it was important to find the shortest possible sequence which DevE would bind to as this is well documented to be an important factor during crystallisation of protein-DNA complexes (Jordan *et al.*, 1985). Before crystallisation it is essential to know if binding takes place between protein and DNA so EMSAs were employed to test this further.

The sequences used for this EMSA are detailed in **Table 5.3**. These sequences were synthesised by Integrated DNA Technologies and were labelled with a 5' Cy5 fluorophore.

Table 5.3 Sequences used for pre-crystallisation EMSA

	Sequence 5'-3'	Size (bp)
DevE_1	<u>AAGTTGGCTGCCAACTTC</u>	18
DevE_3	T <u>CAAGTTGGCTGCCAACTTC</u> AC	22
DevE_7	CGCTTGGGAAGCGT <u>CTCATA</u>	19
DevE_9	CCCGCTTGGGAAGCGT <u>CTCATA</u> AC	23

Micromolar concentrations of DevE were again used as previous studies indicated no interaction at lower concentrations of protein. **Figure 5.6** indicates that the protein-DNA complexes are fully formed at concentrations of DevE above 1.5 μ M with DevE_1 and DevE_2 which contain the same core sequences. Partial binding is observed for DevE_7 and DevE_8 with free probe evident in all lanes. This suggests that the probable promoter sequence for DevE is core sequence of DevE_1 that is underlined in **Table 5.3**.

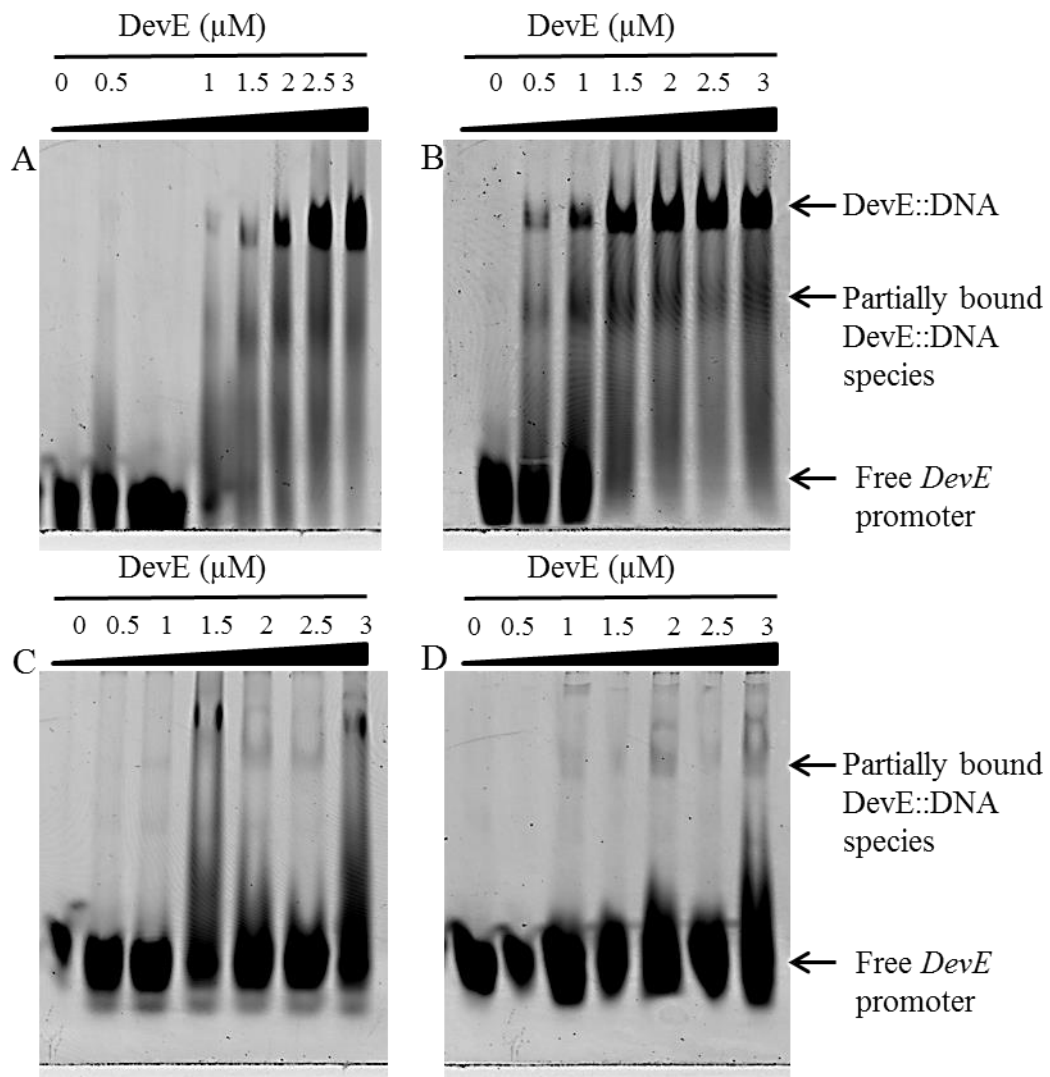


Figure 5.6 Electrophoretic mobility shift assay showing differences between binding of different size promoter regions of DevE (A) 18 bp (B) 22 bp (C) 19 bp (D) 23 bp. Binding assays were carried out using 1.5 ng Cy5 labelled DNA and varying concentrations of DevE. Gel 6% acrylamide, 120V. Imaged with Typhoon 9200.

Summary

Protein-DNA interactions have been analysed and new data has been obtained for DevE-DNA and DevA-DNA interaction. This includes putative promoter regions which were previously unknown. Analysis of protein-DNA interactions by EMSA is a valuable tool in molecular biology however results may be skewed by addition of high concentrations of protein which are not present *in vivo*. As such, the next logical step in this project was to test if functionality could be restored to a PA14 HutC insertional mutant strain by complementation with pKR034.

5.2 Phenotypic analysis of a PA14 HutC transposon insertion mutant

Phenotypic analysis was carried out of a PA14 insertional mutant strain in which the HutC gene is disrupted by the MAR2xT7 transposon. The strain was created as part of the *Pseudomonas* Transposon Insertion Library (Liberati *et al.*, 2006) which was created to facilitate the further study of *P. aeruginosa*.

This strain was complemented with pKR034 and tested on various minimal (M9) media to test for phenotypic differences. *P. aeruginosa* will biosynthesise histidine if a nitrogen source is available therefore the M9 media did not contain any NH₄Cl as the effect of the knocked out HutC gene may be compromised.

Growth was observed for PA14 WT, PA14 HutC::Tn7 and PA14 HutC::Tn7+pKR034 on M9 supplemented with glucose and ammonia, L-histidine and 2xYT (**Figure 5.7 B, C, E**). No phenotypic differences were observed between the strains on these media however it was noted that PA14 HutC::Tn7 and PA14 HutC::Tn7+pKR034 appeared to grow slower than the WT.

No growth was observed on the negative control M9 plate without additives which was expected as there was no carbon or nitrogen source available. There was also no observable growth on the plates supplemented with urocanic acid even for the wild type strain (**Figure 5.7 A & D**).

The difference in growth in solid media was followed up by growth curves in liquid media. Growth curves were carried out in a 96 well plate using eight biological replicates for each condition.

Very limited growth was observed for all strains in M9 medium but this is within the standard error (**Figure 5.8**). The wild type strain grew better than either the insertion

mutant or the complemented strain in M9 medium supplemented with histidine (**Figure 5.9**). Conversely, during a previous study, a *P. fluorescens* $\Delta hutC$ mutant was observed to grow better than the wild type strain in M9 supplemented by histidine (Zhang & Rainey, 2007). No specific growth rates have been calculated for this condition. In M9 supplemented with urocanic acid, only the wild type strain was observed to have growth albeit very limited growth (**Figure 5.10**). PA14 HutC::Tn7 was observed to grow slightly better than either the wild type or complemented strain in M9 supplemented with ammonia and glucose (**Figure 5.11**). Interestingly, the complemented strain was observed to grow better than the wild type in rich media (2xYT; **Figure 5.12**). Specific growth rates (μ) are shown in **Table 5.4**. No specific growth rates could be calculated for PA14 HutC::Tn7 and PA14 HutC::Tn7+pKR034 grown in M9 medium supplemented with urocanic acid.

Table 5.4 Specific growth rates of PA14 strain grown in various media

Strain	M9 + L-histidine	M9 + urocanic acid	M9 + NH ₄ Cl & glucose	2xYT
PA14 WT	0.226	0.234	0.136	0.168
PA14 HutC::Tn7	0.185	n/a	0.178	0.160
PA14 HutC::Tn7 + pKR034	0.194	n/a	0.136	0.214

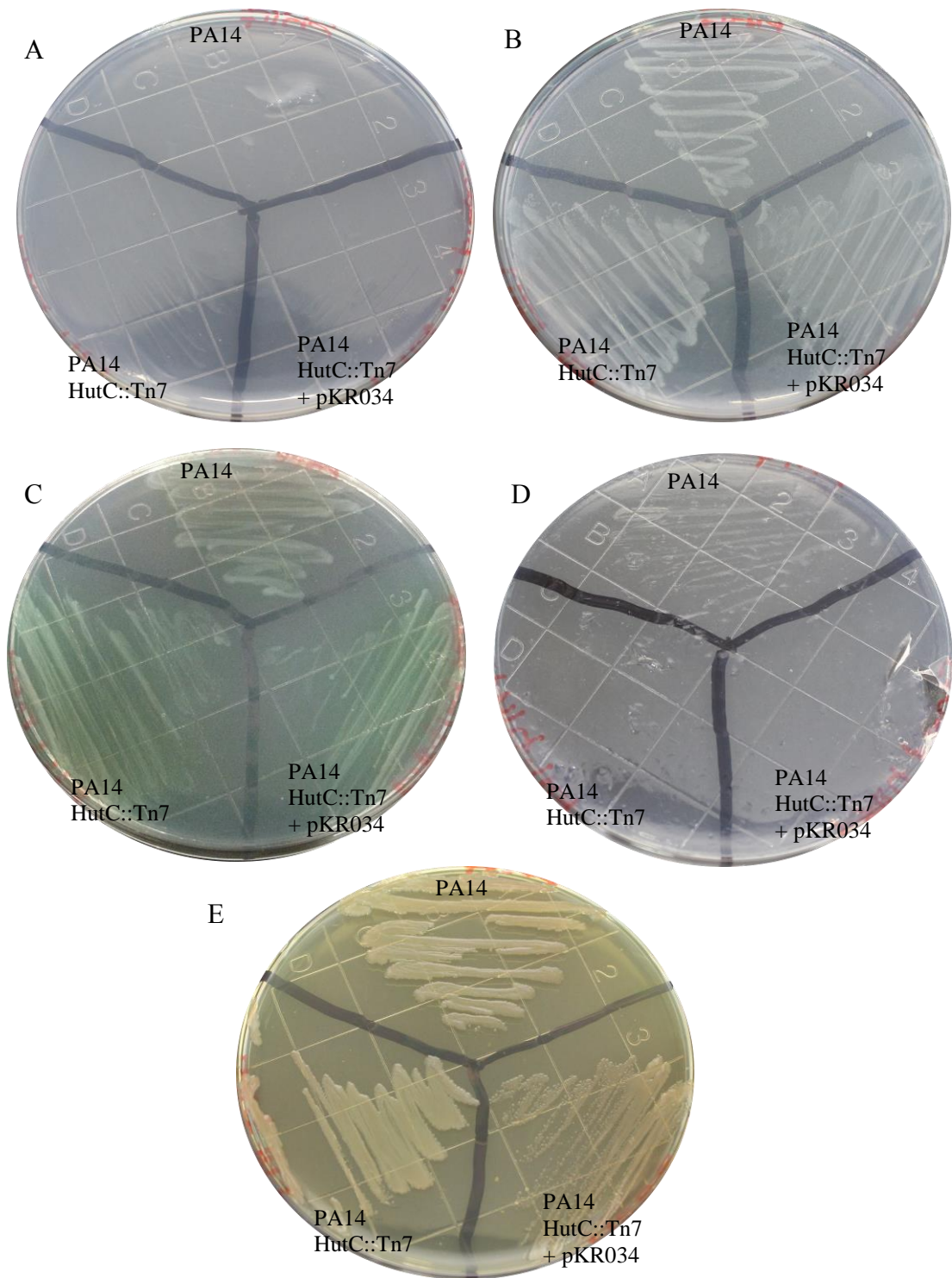


Figure 5.7 Phenotypic comparison of PA14 WT, PA14 HutC::Tn7 and PA14 HutC::Tn7 + pKR006 on various media (A) M9 (B) M9 + 18.7 mM NH₄Cl & 22.2 mM glucose (C) M9 + 15 mM L-histidine (D) M9 + 15 mM urocanic acid (E) 2xYT

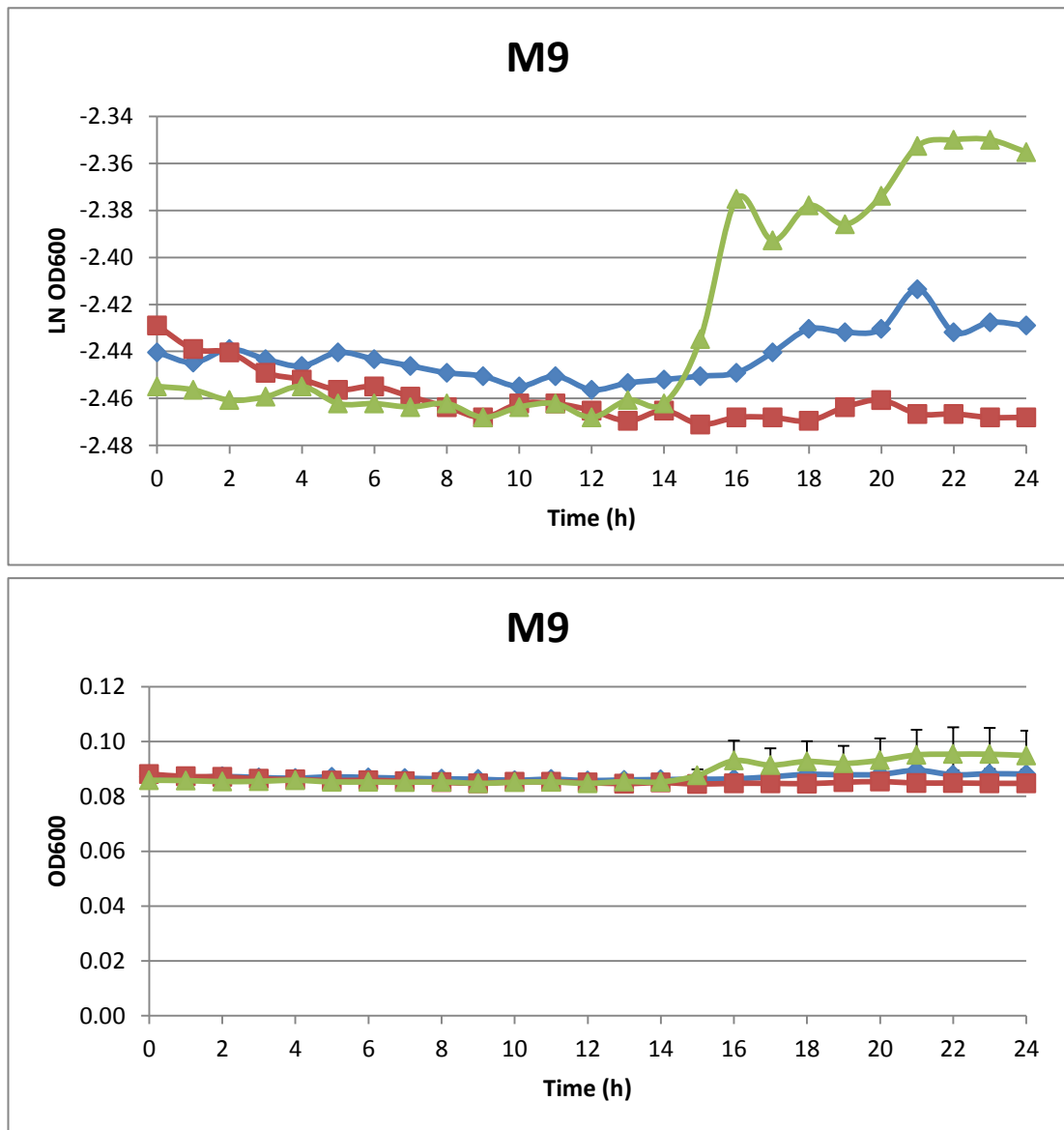


Figure 5.8 Growth curves for PA14 WT (blue diamonds), PA14 HutC::Tn7 (red squares) and PA14 HutC::Tn7 + pKR034 (green triangles) on M9 media. Natural log plots are shown in the top panel while raw data plots are shown in the bottom panel with standard error. Where error bars are not visible, the standard error is within the symbols and therefore not visible. Data were collected every 15 minutes however hourly time points are shown for clarity.

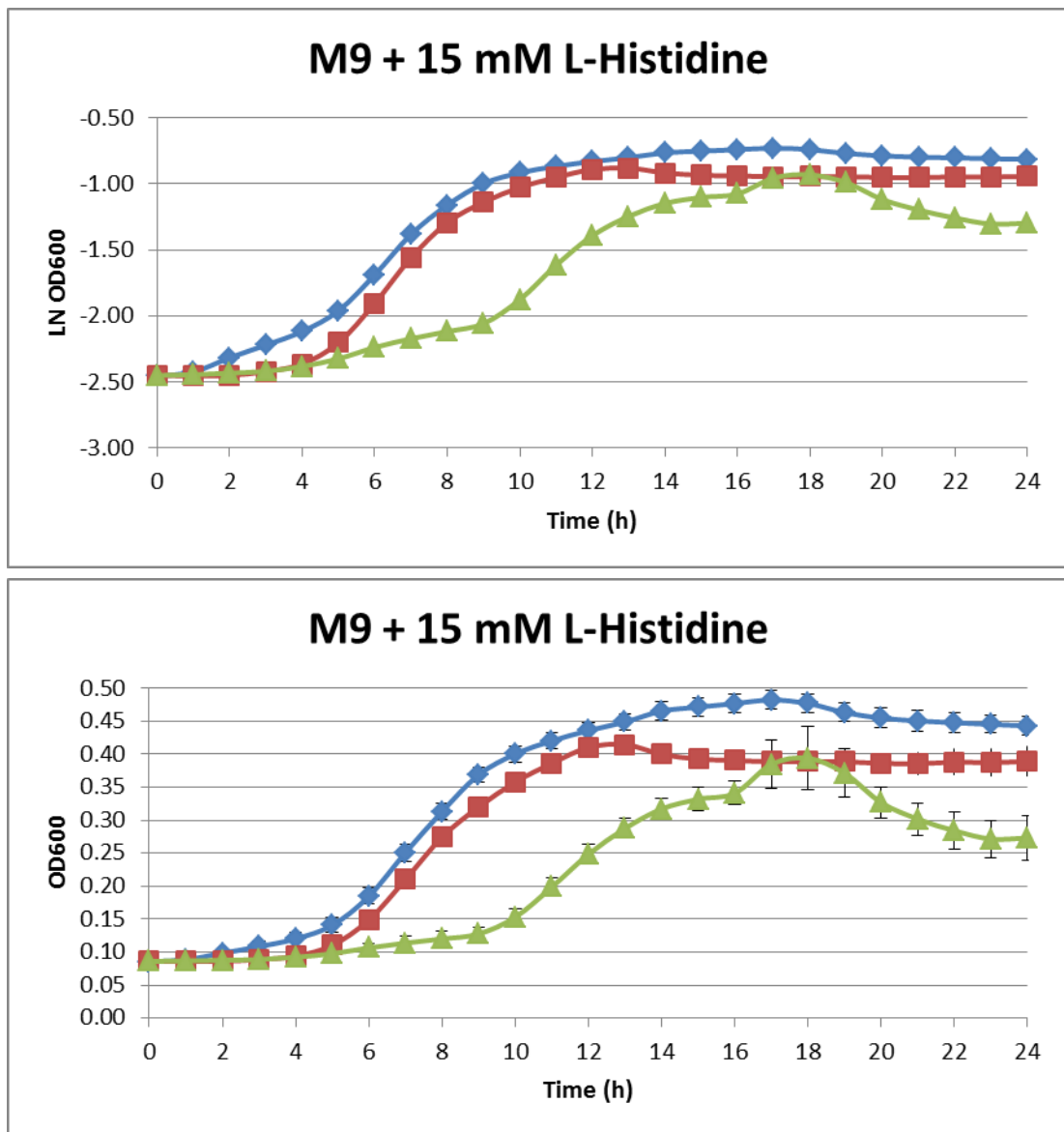


Figure 5.9 Growth curves for PA14 WT (blue diamonds), PA14 HutC::Tn7 (red squares) and PA14 HutC::Tn7 + pKR034 (green triangles) on M9 media + 15 mM L-histidine. Natural log plots are shown in the top panel while raw data plots are shown in the bottom panel with standard error. Where error bars are not visible, the standard error is within the symbols and therefore not visible. Data were collected every 15 minutes however hourly time points are shown for clarity.

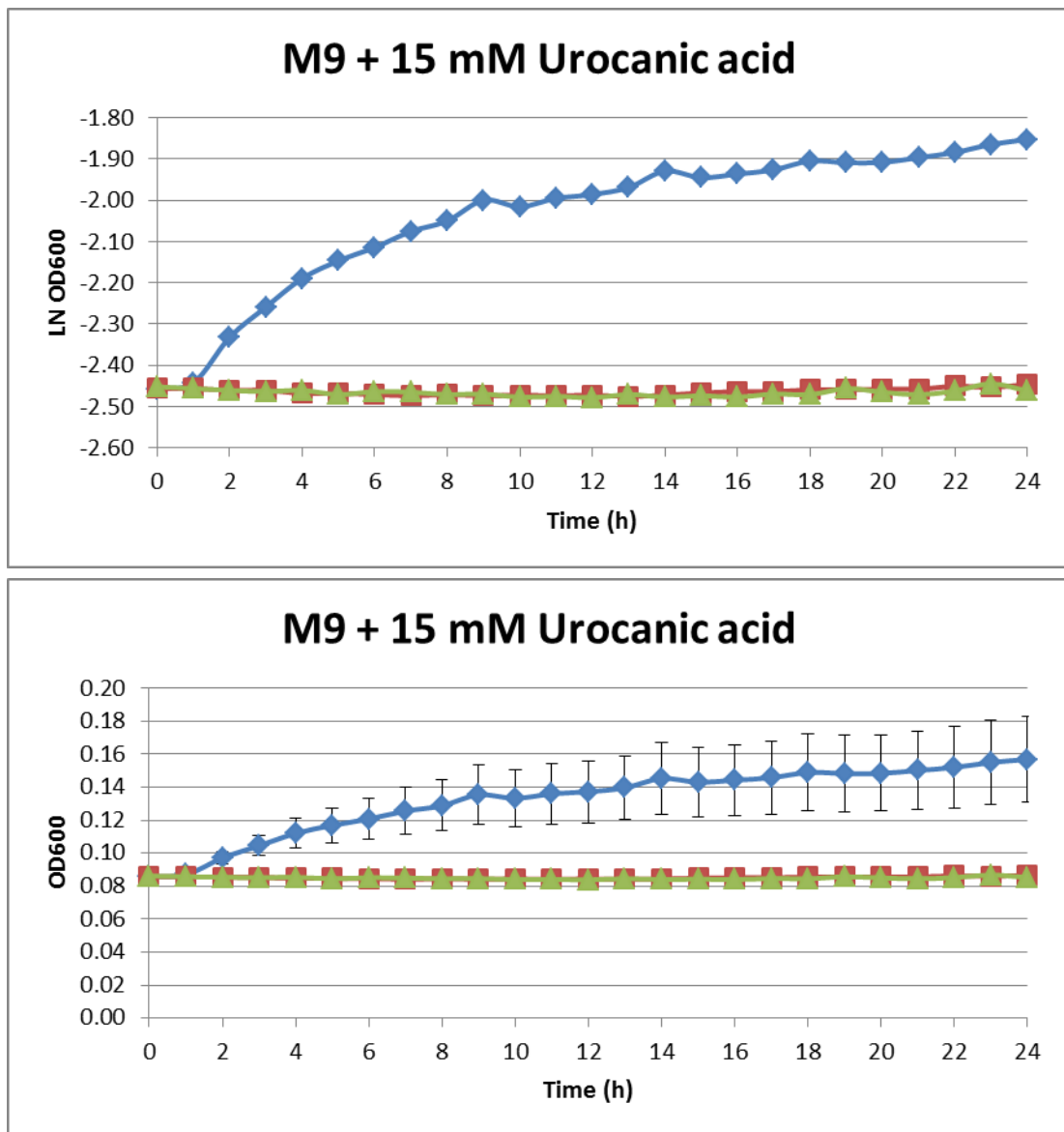


Figure 5.10 Growth curves for PA14 WT (blue diamonds), PA14 HutC::Tn7 (red squares) and PA14 HutC::Tn7 + pKR034 (green triangles) on M9 media + 15 mM urocanic acid. Natural log plots are shown in the top panel while raw data plots are shown in the bottom panel with standard error. Where error bars are not visible, the standard error is within the symbols and therefore not visible. Data were collected every 15 minutes however hourly time points are shown for clarity.

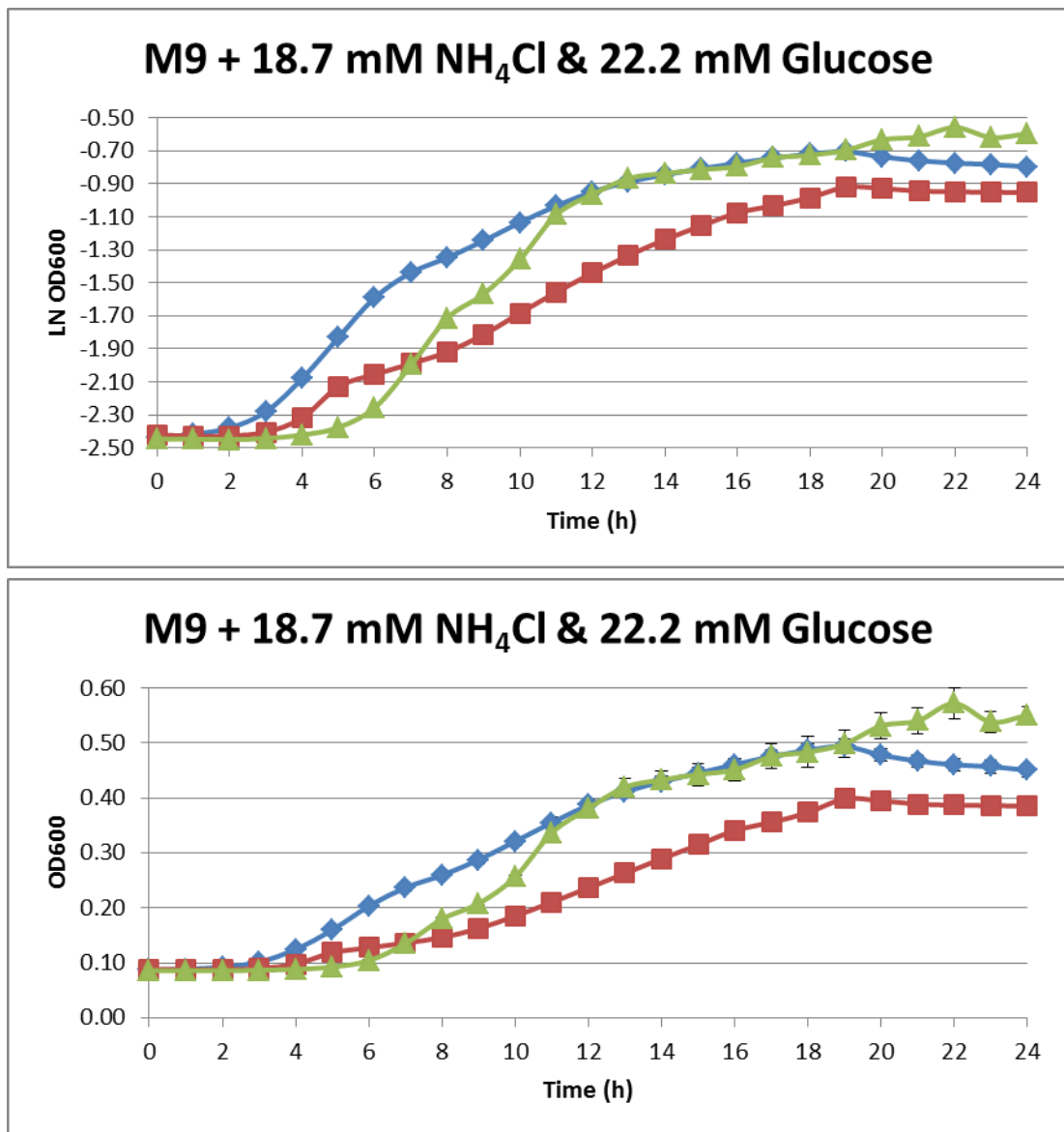


Figure 5.11 Growth curves for PA14 WT (blue diamonds), PA14 HutC::Tn7 (red squares) and PA14 HutC::Tn7 + pKR034 (green triangles) on M9 media + 22.2 mM glucose and 18.7 mM NH₄Cl. Natural log plots are shown in the top panel while raw data plots are shown in the bottom panel with standard error. Where error bars are not visible, the standard error is within the symbols and therefore not visible. Data were collected every 15 minutes however hourly time points are shown for clarity.

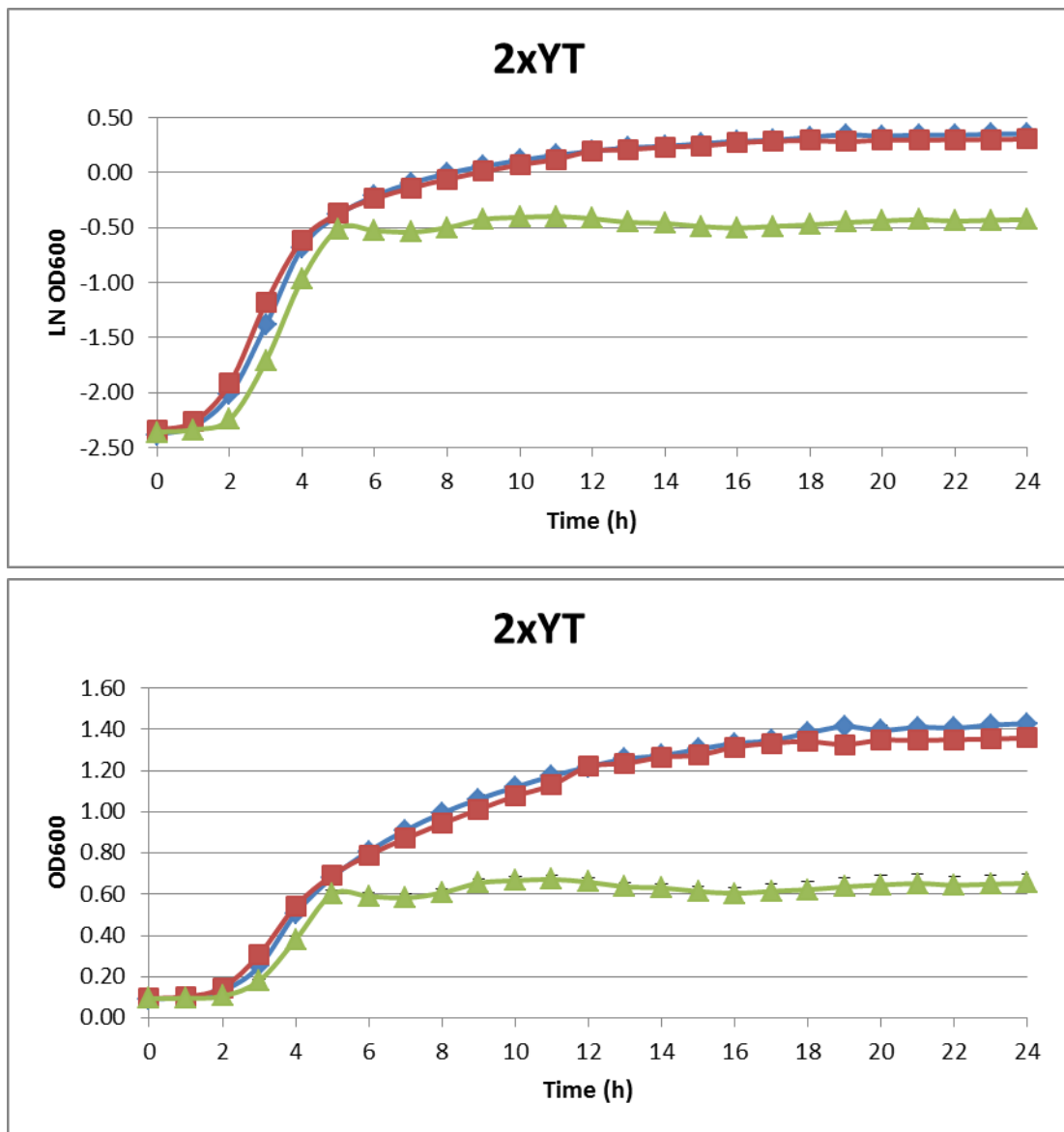


Figure 5.12 Growth curves for PA14 WT (blue diamonds), PA14 HutC::Tn7 (red squares) and PA14 HutC::Tn7 + pKR034 (green triangles) on 2xYT media. Natural log plots are shown in the top panel while raw data plots are shown in the bottom panel with standard error. Where error bars are not visible, the standard error is within the symbols and therefore not visible. Data were collected every 15 minutes however hourly time points are shown for clarity.

Summary

Phenotypic and growth analysis has revealed subtle differences between wild-type, PA14 HutC::Tn7 and PA14 HutC::Tn7 complemented with pKR034. Functionality was not conclusively restored in the complemented strain during this project. The growth experiments for PA14 mutants in this study were based on previous work on *P. fluorescens* SBW25 HutC mutants (Zhang & Rainey, 2007). These two strains occupy different ecological niches. PA14 is a virulent pathogen which most commonly infects the lungs of cystic fibrosis patients (Doring, 1993, Lee *et al.*, 2006). *P. fluorescens*, on the other hand, is found in the roots of plants where it exists symbiotically, providing the plant with an advantage against pathogens and providing growth promoting compounds. (Cook *et al.*, 1995, Artursson *et al.*, 2006).

The differences in the ecological niches between these strains goes some way, perhaps, to account for the differences that were observed in growth between PA14 HutC mutants in this work and *P. fluorescens* mutants in previous work.

Chapter 6 Two- Dimensional Infrared Spectroscopy (2D-IR)

6.1 2D-IR of Isoniazid (INH)

Isoniazid (isonicotinic acid hydride; INH) is a frontline treatment for tuberculosis and the primary inhibitor of the enoyl acyl carrier protein InhA. Isoniazid is a pro-drug which requires activation to become functional. The action of the oxidative peroxidase enzyme KatG is responsible for oxidation of the pro-drug by removal the -NH-NH₂ moiety. There are then two proposed pathways which have been proposed (**Figure 6.1**) which form the biologically active INH-NAD adduct (Zhang *et al.*, 1992, Rozwarski *et al.*, 1998, Kruh *et al.*, 2007, Molle *et al.*, 2010). This is of great significance when it is considered that tuberculosis is still such a global health problem; 8.6 million cases and 1.3 million deaths worldwide (Eurosurveillance editorial team 2013). The available treatments include INH, rifampicin, ethionamide (ETH), and isoxyl (ISO) which have been in use for more than 50 years (Favila *et al.*, 2007, Klopman *et al.*, 1996) and as such *M. tuberculosis* has developed significant resistance to these treatments. It is vital that drug-protein binding dynamics are better understood in order to design new effective therapeutic strategies. The following work gives an insight into the InhA/INH interaction by 2D-IR spectroscopy as a model to develop the study of whole proteins by 2DIR, with a view to applying this method to the complex dynamics of GntR-like regulators and specifically to HutC interactions with urocanic acid and DNA.

The molecular structure of INH lends itself well to 2DIR spectroscopy as it has functional groups which give relatively large distinguishable signals in IR

spectroscopy e.g. C=O, NH₂, NH (**Figure 6.2**). The absorption frequencies are also well separated, allowing the kinetics of each moiety to be monitored independently.

Table 6.1 details the vibrational frequencies for molecular motions of INH in D₂O.

The molecular dynamics of INH were analysed on ULTRA based at the CLF.

2D-IR spectra are shown with the corresponding FTIR spectra for reference for a range of waiting times (0.25 – 1.5 ps) which demonstrates vibrational coupling to different moieties within the INH molecule (**Figure 6.3 a - f**). These are known as cross peaks; peaks which appear off the diagonal, indicated by dashed lines in **Fig. 6.3**. Additionally, the cross peaks intensify progressively as the waiting times also increase due to vibrational energy transfer.

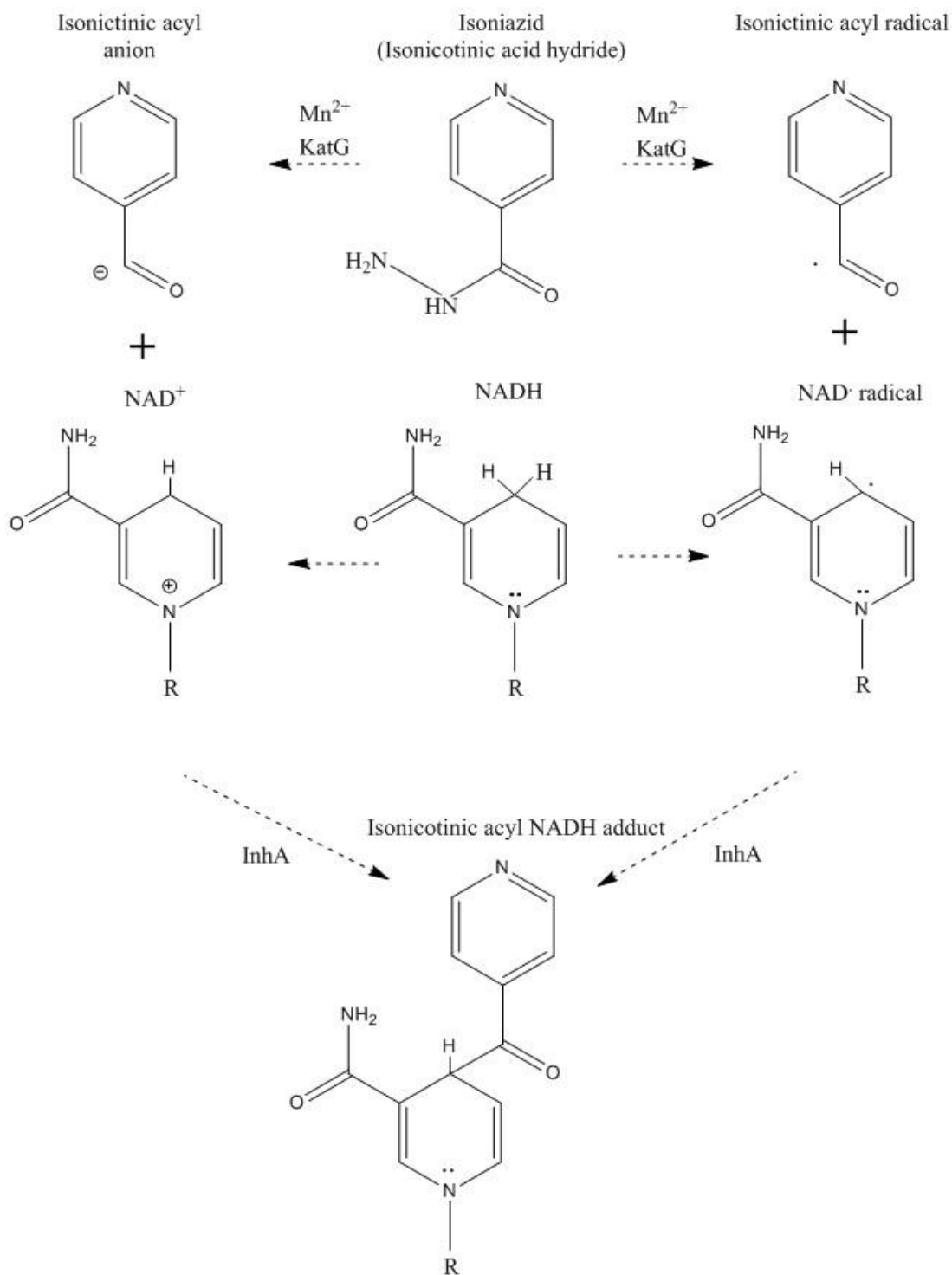


Figure 6.1 Two proposed pathways for activation of INH. INH is oxidised by KatG/Mn²⁺ forming either an anion or radical which covalently binds to the C4 position of a form of NADH (NAD⁺ or radical) in the active site of InhA. Produced by ChemDraw v15 (Perkin Elmer Informatics)

Table 6.1 IR spectroscopy data for INH in D₂O

Frequency (cm ⁻¹)	Molecular motion
1644	$\nu(\text{CO})$
1608	$\nu(\text{CC})_{\text{ring, sym}}$
1555	$\nu(\text{CC}, \text{CN})_{\text{ring, asym}}$
1502	$\nu(\text{CC})_{\text{ring, asym}}$
1447	$\nu(\text{CN})-\beta(\text{ND})$
1415	$\nu_{\text{ring}}, \beta(\text{CH})$

ν = stretching vibration, β = bending vibration

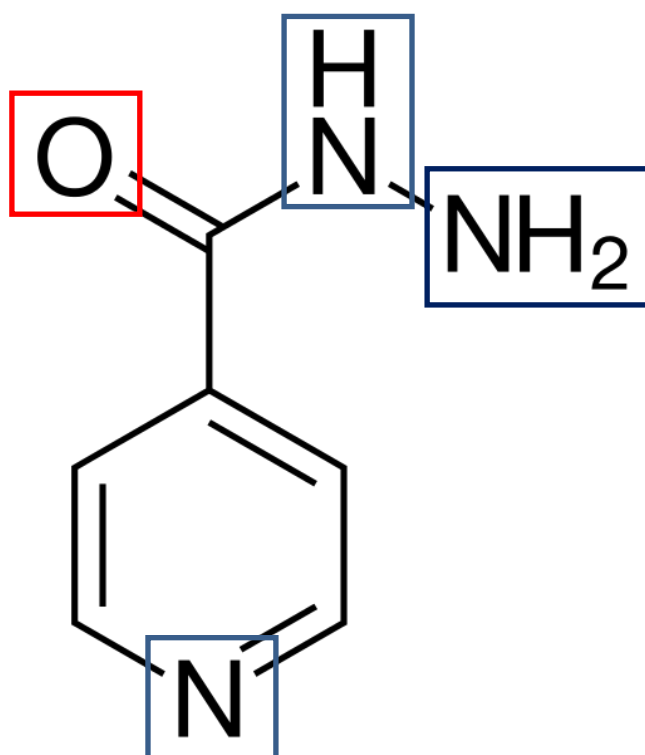


Figure 6.2 Molecular structure of Isoniazid. Functional groups which contribute to IR spectra are highlighted by coloured boxes. These are also highlighted **Table 6.1**.

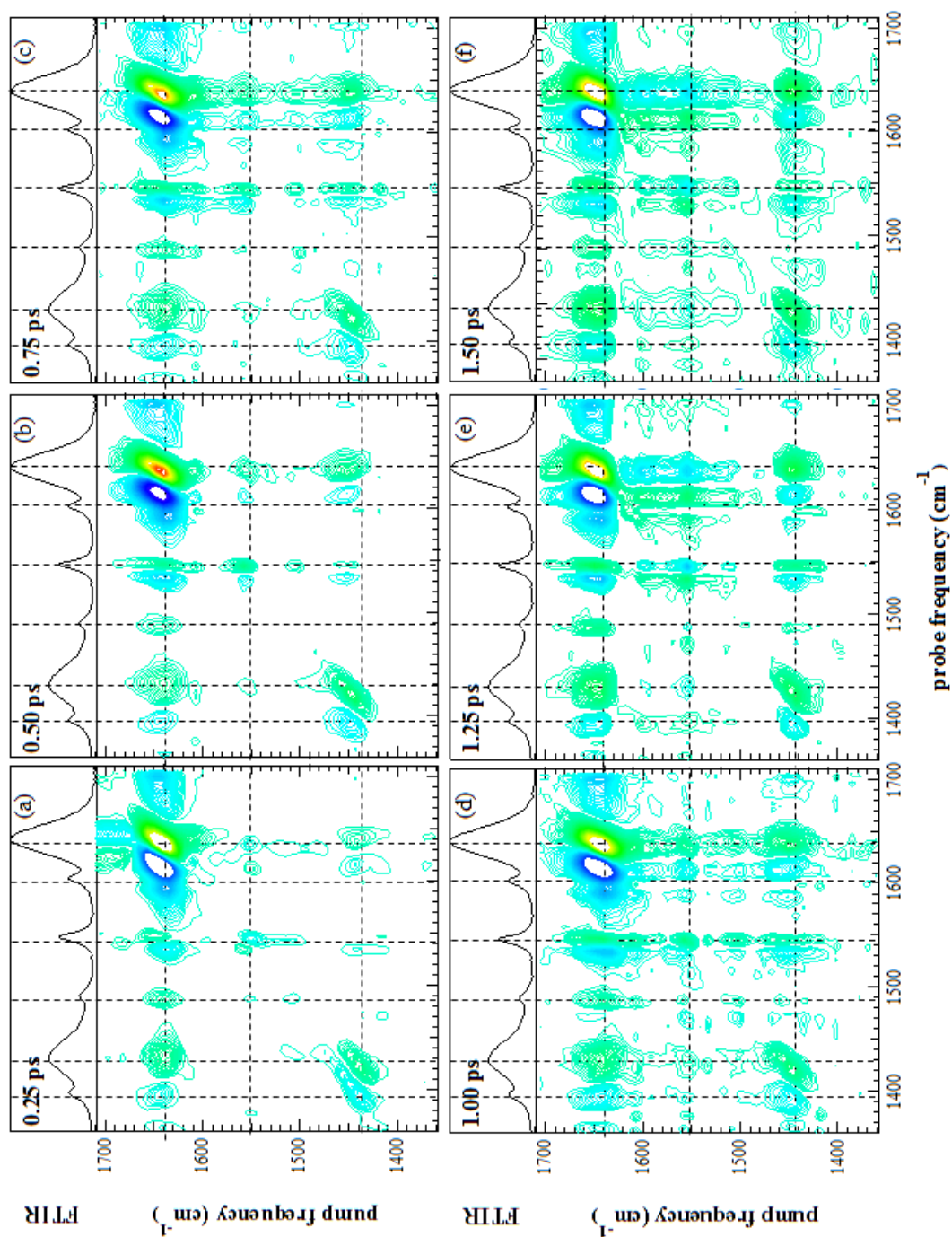


Figure 6.3 2D-IR spectra of INH in D₂O solution at a range of experimental waiting times (Tw). Tw are shown in each figure. FTIR spectra are given in each case for reference. Dashed lines indicate vibrational coupling between different moieties within INH. Figure reproduced from (Shaw *et al.*, 2015)

6.2 2D-IR of InhA as a model for future protein dynamic analysis

InhA is a 269 amino acid enoyl acyl carrier protein reductase which is primarily responsible for biosynthesis of mycolic acids; a vital component of *M. tuberculosis* cell walls. It is the primary target of INH which inhibits the binding of NADH in the single active site and thus inhibits biosynthesis of mycolic acids. As mentioned previously *M. tuberculosis* has developed significant resistance to current treatments. One such resistance mutation is an S94A mutation within InhA however given that multidrug resistance is now prominent it seems reasonable to probe all molecular contacts within the 20 amino acid active site of InhA in order to find new novel inhibitors.

Molecular dynamics of InhA and its interaction with INH were interrogated using the ULTRA laser in the photo-echo 2D-IR configuration based within the CLF. A number of point mutations (M155A, F149A, P193A, W222A and S94A) were introduced by site directed mutagenesis using plasmids created during previous studies of InhA (Molle *et al.*, 2010). These mutations are within the active site of InhA to investigate their roles in activity. InhA requires NADH, which is oxidised within the active site to NAD⁺, for mycolic acid synthesis. INH blocks this reaction by binding to the nicotinamide ring of NADH. The INH-NAD adduct cannot enter the active site as it is no longer in the correct conformation. Of particular interest is the S94A mutation which is a known natural mutation and confers significant resistance to INH; it is, however, not part of the active site but rather lies near to the entrance of the active site presumably conferring a conformational change which prevents INH binding to NADH in the active site. The focus of this chapter is on InhA WT, as a control, S94A, due to its natural significance and P193A which had

almost no activity in turnover of DD-CoA or NADH during activity assays (**Figure 6.4 A & B**).

InhA samples (WT, S94A and P193A) were measured at concentrations of 26 mg ml⁻¹. The frequency region from 1550 – 1700 cm⁻¹ which spans the amide I region was probed and revealed the secondary structure features present in InhA which were compared with the available crystal structures (4DTI, 2AQ8, 2NV6 and 1ZID). Secondary structure features of *apo*-InhA variants were compared to those of InhA variants with NAD⁺ or INH in the binding pocket.

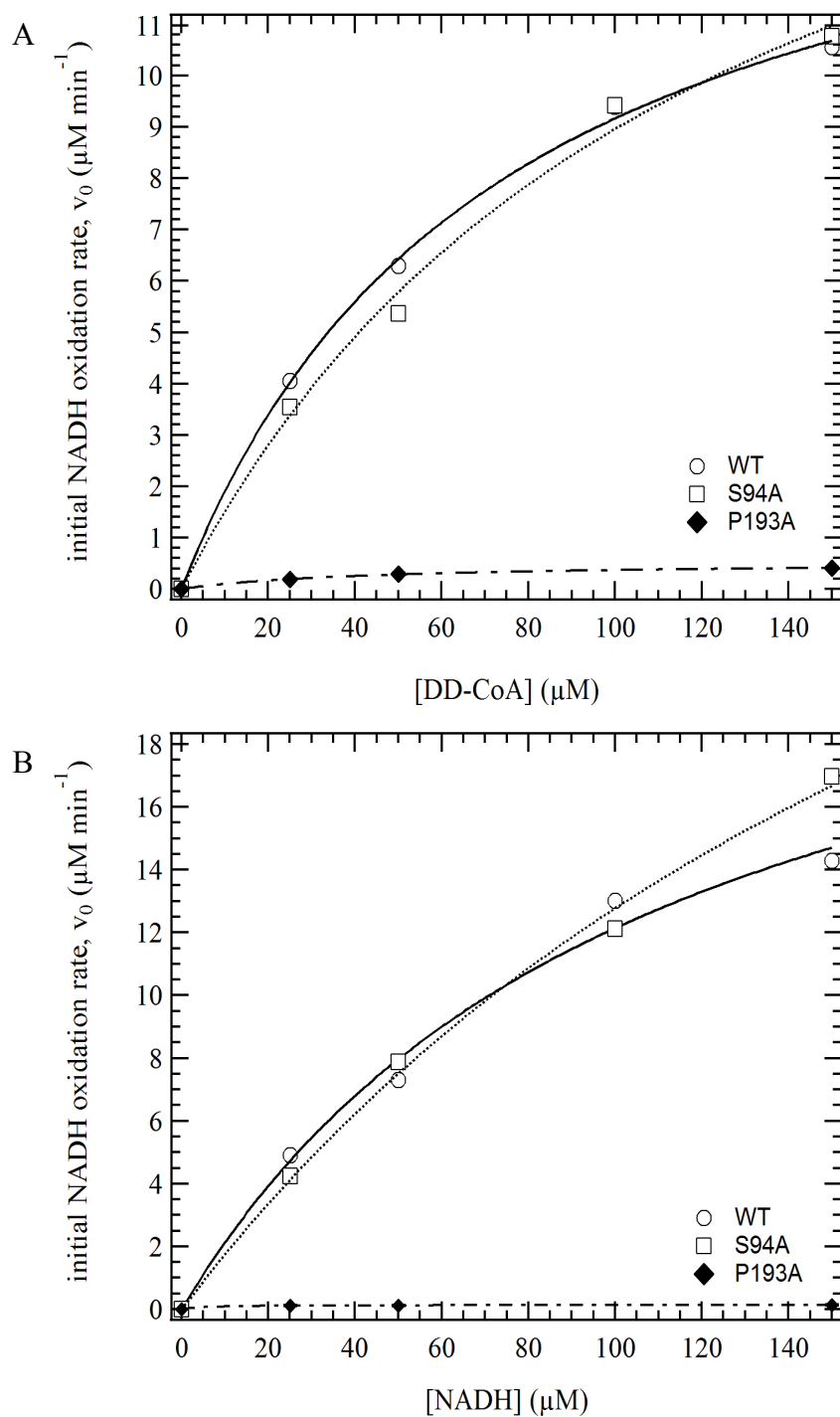


Figure 6.4 (A) The enzymatic activity of the InhA variants in the presence of fixed 100 μM NADH with increasing DD-CoA concentration (25-150 μM). The dashed line is a fit to the data described by $v_0 = V_{\text{max}} [\text{DD-CoA}] / (k_m + [\text{DD-CoA}])$ (B) The enzymatic activity of the InhA variants in the presence of fixed 50 μM DD-CoA with increasing NADH concentration (25-150 μM). The dashed line is a fit to the data described by $v_0 = V_{\text{max}} [\text{NADH}] / (k_m + [\text{NADH}])$

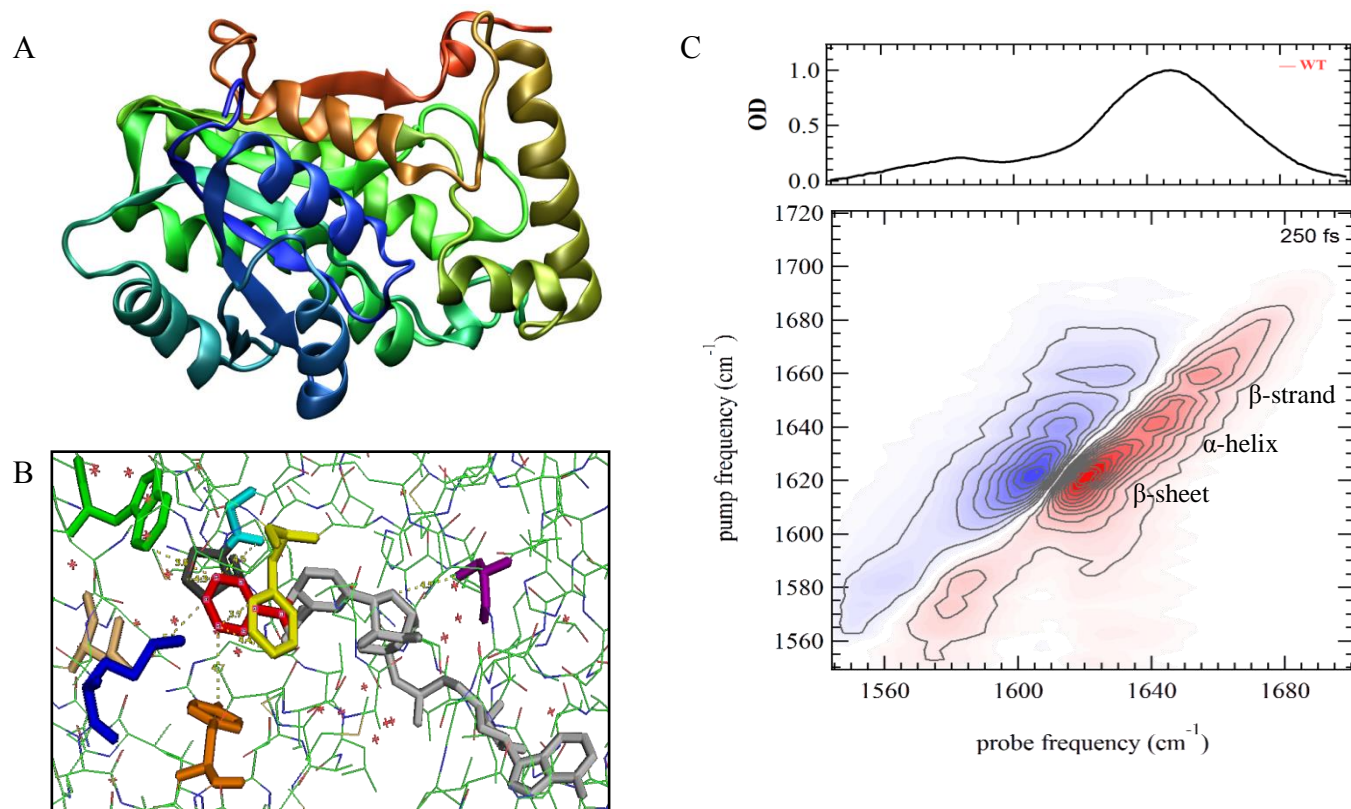


Figure 6.5 (A) Crystal structure of InhA (PDB code 1ZID); (B) The active site of InhA containing an NADH molecule (grey) covalently attached to the INH moiety (red). SDM was used to introduce the following point mutations methionine M155A (dark blue), phenylalanine F149A (yellow), proline P193A (black), tryptophan W222A (green) and serine S94A (purple); (C) representative FT-IR spectra and its corresponding 2D-IR spectra of InhA WT which show the secondary structure features of (A)

InhA P193A

Activity assays with DD-CoA and NADH reveals that InhA P193A has no turnover of either DD-CoA or NADH indicating that this mutation causes the active site to lose function (**Figure 6.4**). This is reflected in the 2D-IR data as there are no spectral differences observed for the P193A mutation either in the *apo* form, with NAD⁺ bound or INH bound (**Figure 6.5**). Cross sections of the 2D spectra also reveal no differences in peak intensities at 1620 cm⁻¹ (β -sheet) and 1650 cm⁻¹ (α -helix) between NAD⁺ and drug binding (**Figure 6.6 and 6.7**). As a result of the activity assay InhA P193A is used as a reference sample. Data have been normalised to the β -sheet intensities for a clear comparison of the interaction however as the β -sheet is also changing data normalised to the α -helix have also been provided for comparison. The β -strands mentioned previously have not been included in analysis of peak intensities as they don't participate directly in the binding interactions with NAD⁺ or INH but just interconnect the β -sheets and α -helices which do. Full tables with all peak intensities can be referred to in **Appendix 6**.

InhA WT

Clear spectral differences are observed (**Figure 6.8**) for InhA WT when the *apo* form is compared to the NAD⁺ and drug bound complexes. Secondary structure features are obvious in all three spectra (1620cm⁻¹ = β sheet, 1647 cm⁻¹ = α helix, 1661 cm⁻¹ = β strand) for the wild type however the peak intensities are lower in InhA WT bound to NAD⁺ and InhA WT bound to INH (**Figure 6.9**). Comparison of peak intensities reveals that changes occur within the β -sheet- α -helix interactions of InhA upon binding of NAD⁺ or INH (**Figure 6.10**)

InhA S94A

InhA S94A again has obvious spectral differences between the various forms of the protein (**Figure 6.11**). The intensity of the bleach is clearly more intense in the *apo*-InhA S94A than InhA S94A + NAD⁺. This is indicative of the limited binding of INH to InhA S94A. The cross section peak intensities are shown in **Figure 6.12**. The differences are clear to see when the peak intensities are compared. There are changes in both the β -sheet and α -helix in the drug bound form compared with the apo and NAD⁺ bound form (**Figure 6.13**). Furthermore, when this data is compared with the data for the WT protein, it is clear that the β -sheet- α -helix interaction is important in drug binding therefore this data can go some way to explain the mechanism of resistance to INH observed with the S94A mutation.

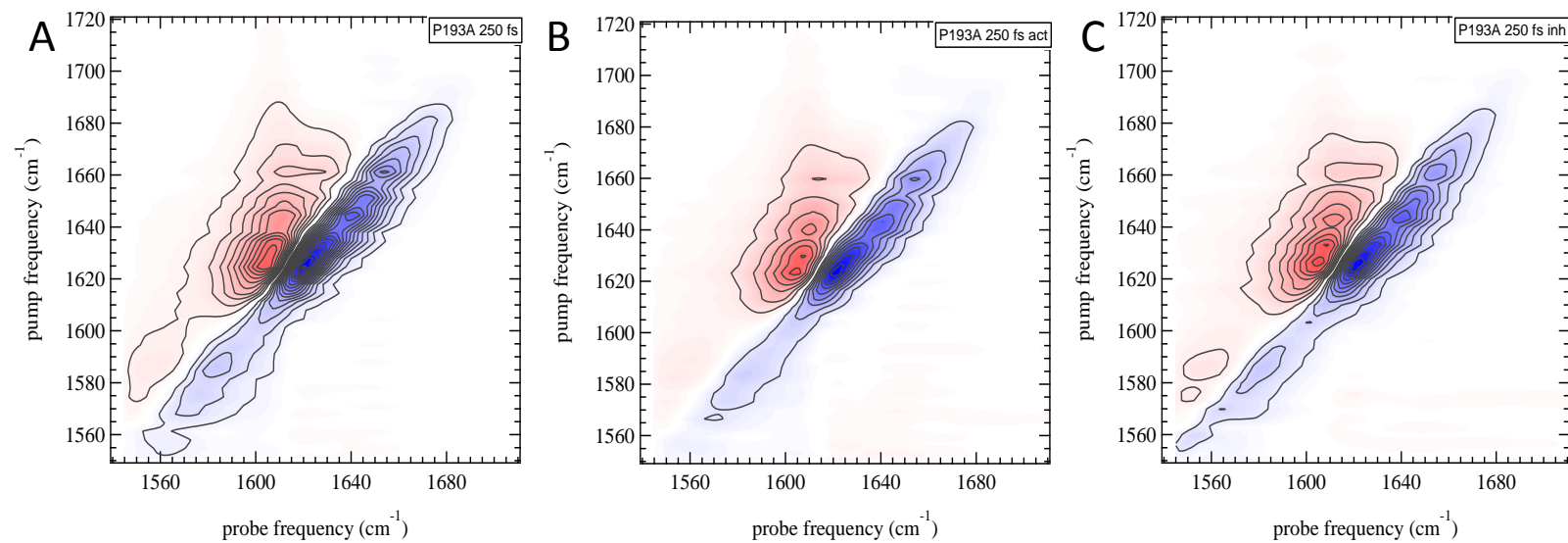


Figure 6.5 2D-IR spectra showing differences in secondary structure elements of **(A)** InhA P193A **(B)** InhA P193A with NAD⁺ **(C)** InhA P193A with INH. Data were collected on ULTRA with a T_w of 250 fs.

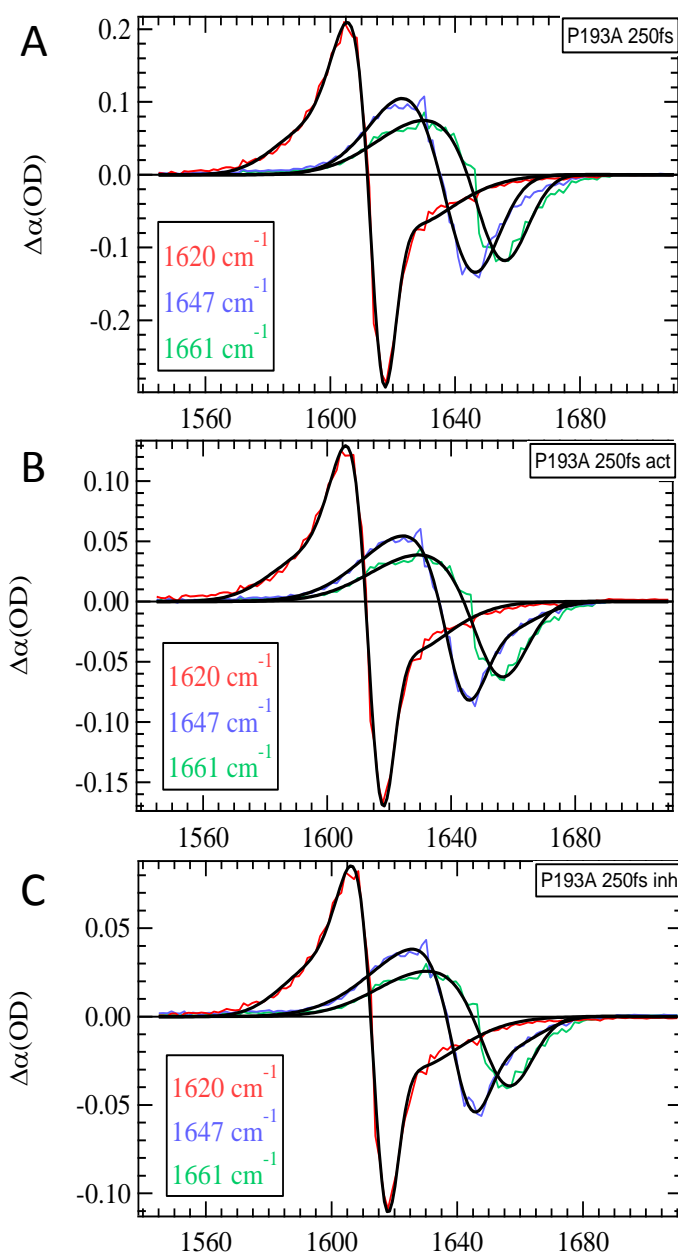


Figure 6.6 Cross section slices of 2D-IR spectra reveal differences in (A) InhA P193A (B) InhA P193A + NAD⁺ (C) InhA P193A + INH. Data were collected on ULTRA with $T_w = 250$ fs.

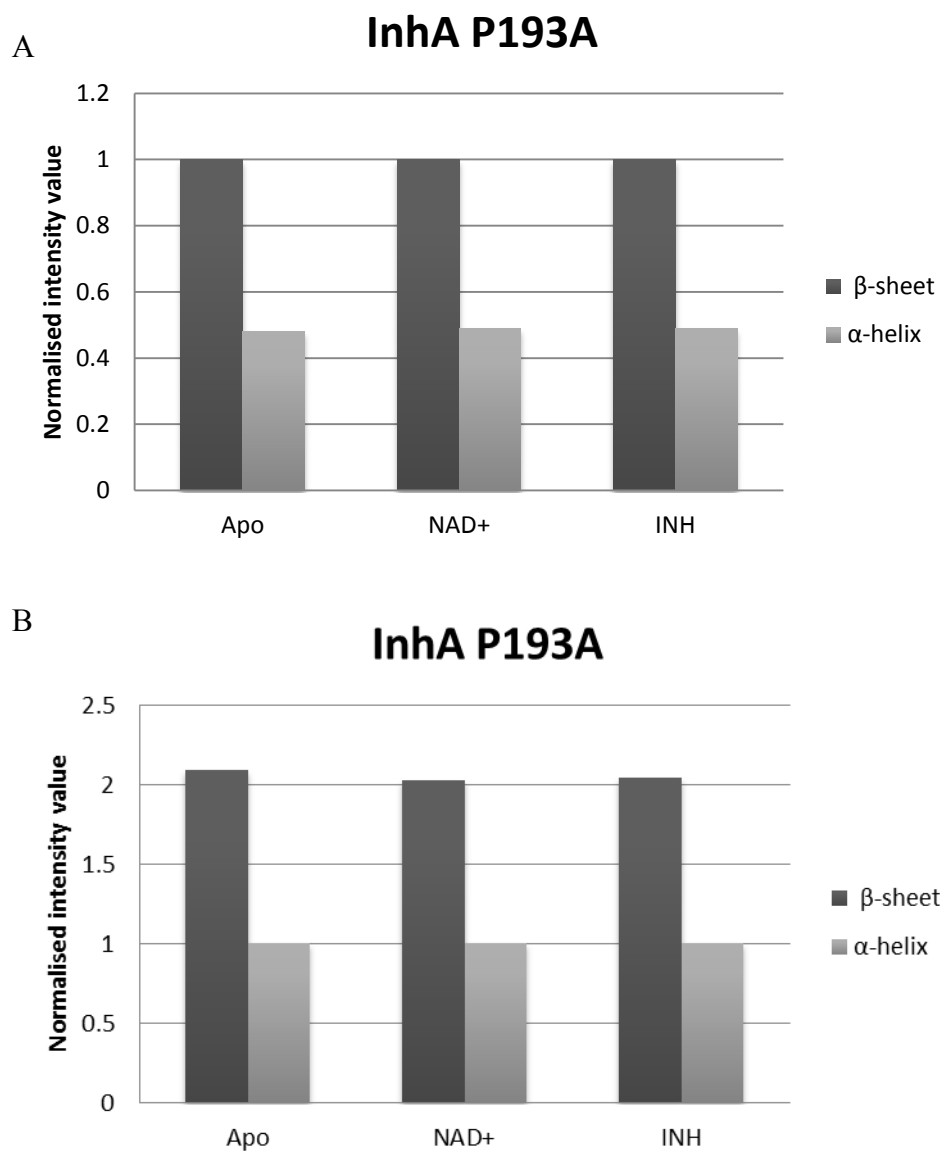


Figure 6.7 (A) Comparison of peak intensities obtained from cross sections of 2D-IR measurement of InhA P193A. Intensities were normalised to the β -sheet intensity. (B) Comparison of peak intensities normalised to the α -helix intensity. No changes in intensity are observed between the β -sheet- α -helix interaction in the *apo*, NAD+ or drug bound form of the protein.

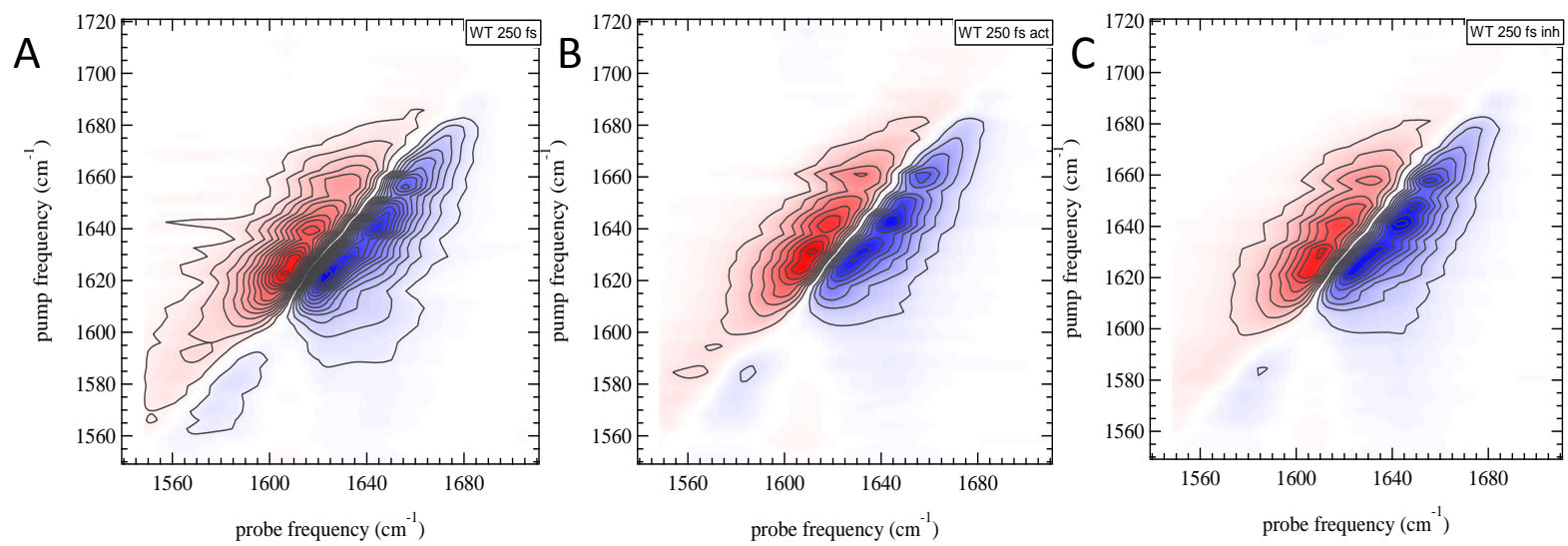


Figure 6.8 2D-IR spectra showing differences in secondary structure elements of (A) InhA WT (B) InhA WT with NAD⁺ (C) InhA WT with INH. Data were collected on ULTRA with a T_w of 250 fs.

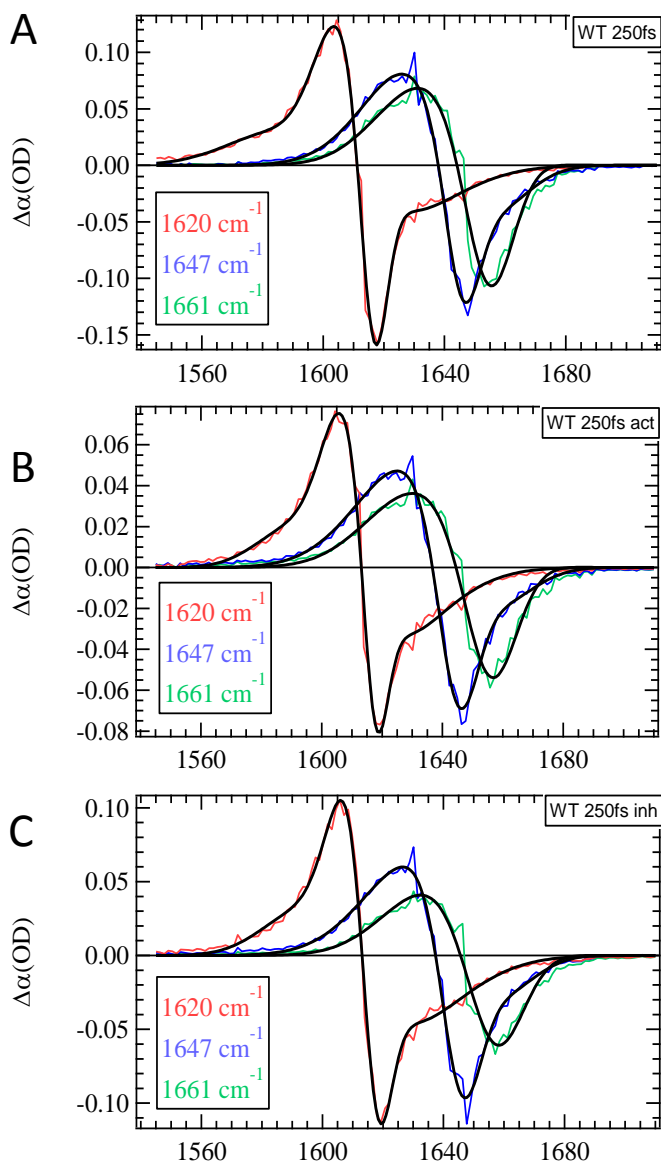


Figure 6.9 Cross section slices of 2D-IR spectra reveal differences in (A) *apo*-InhA WT compared to (B) InhA + NAD⁺ and (C) InhA with INH. Data were collected on ULTRA with $T_w = 250 \text{ fs}$.

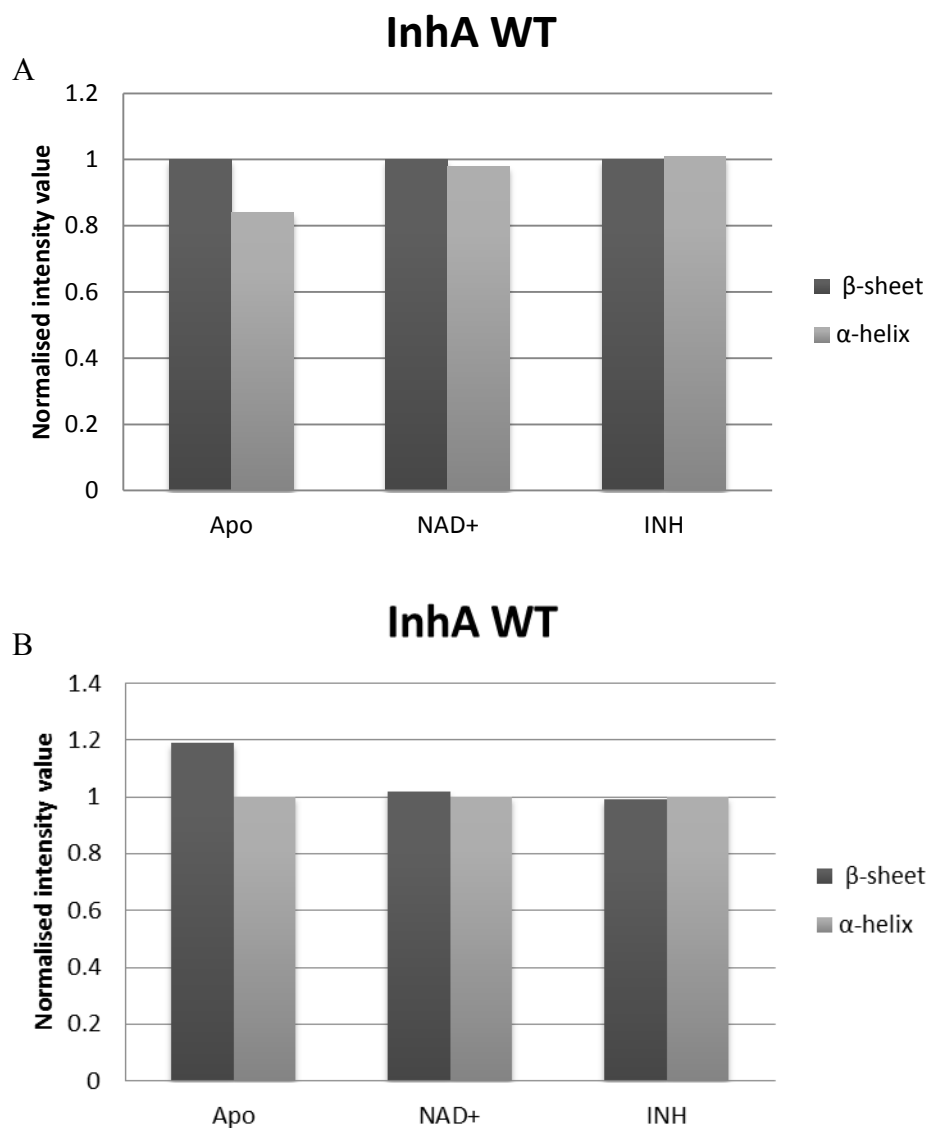


Figure 6.10 (A) Comparison of peak intensities obtained from cross sections of 2D-IR measurement of InhA WT. Changes can be observed in the interaction of between the β -sheet and α -helix α -helical secondary structure upon NAD⁺ and drug binding. Peak intensities have been normalised to the β -sheet. (B) Comparison of peak intensities normalised to the α -helix intensity.

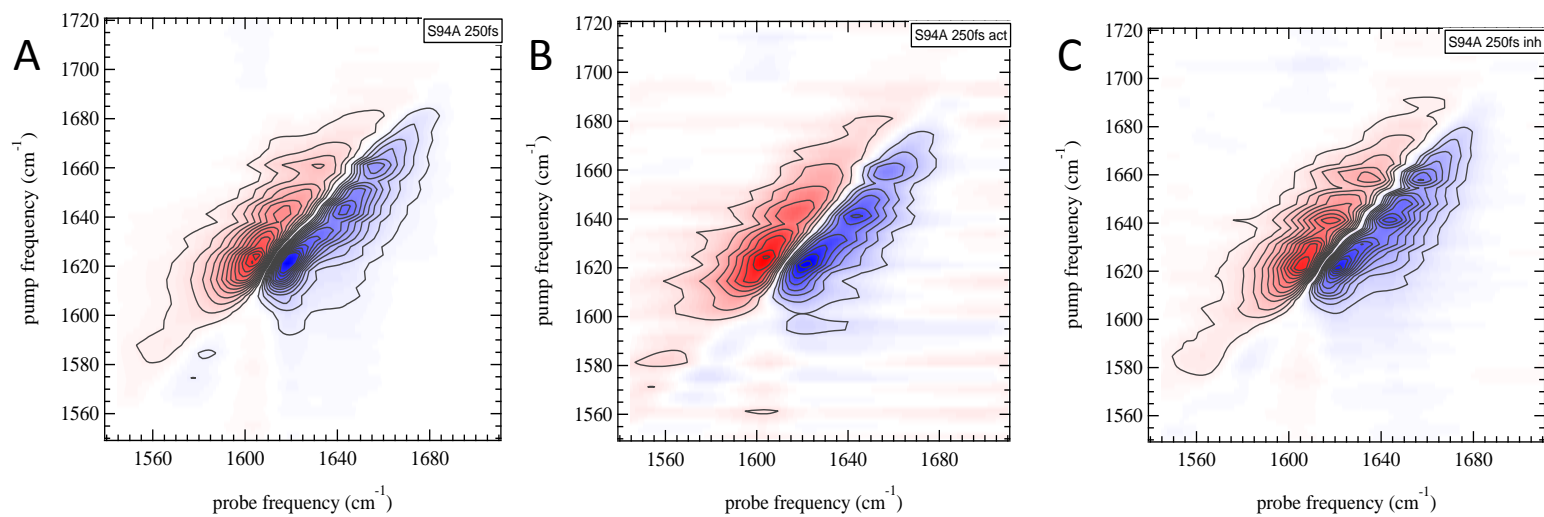


Figure 6.11 2D-IR spectra showing differences in secondary structure elements of (A) InhA S94A (B) InhA S94A with NAD+ (C) InhA with INH. Data were collected on ULTRA with a T_w of 250 fs.

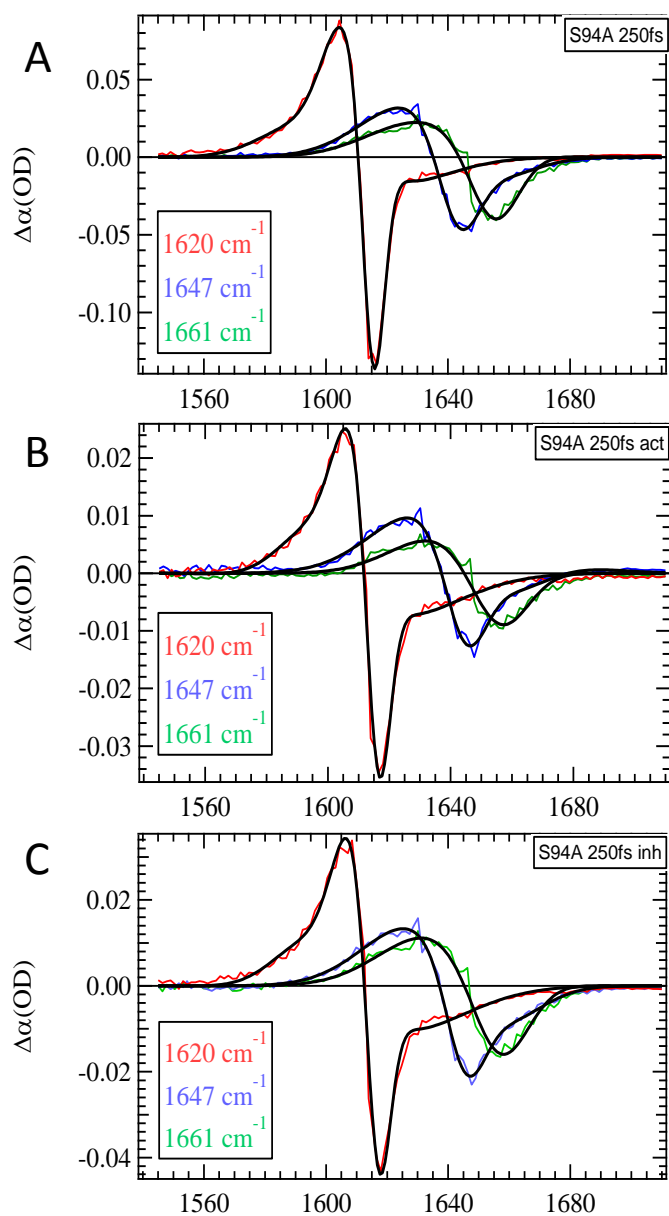


Figure 6.12 Cross section slices of 2D-IR spectra reveal differences in (A) InhA S94A (B) InhA S94A + NAD⁺ (C) InhA + INH. Data were collected on ULTRA with $T_w = 250$ fs.

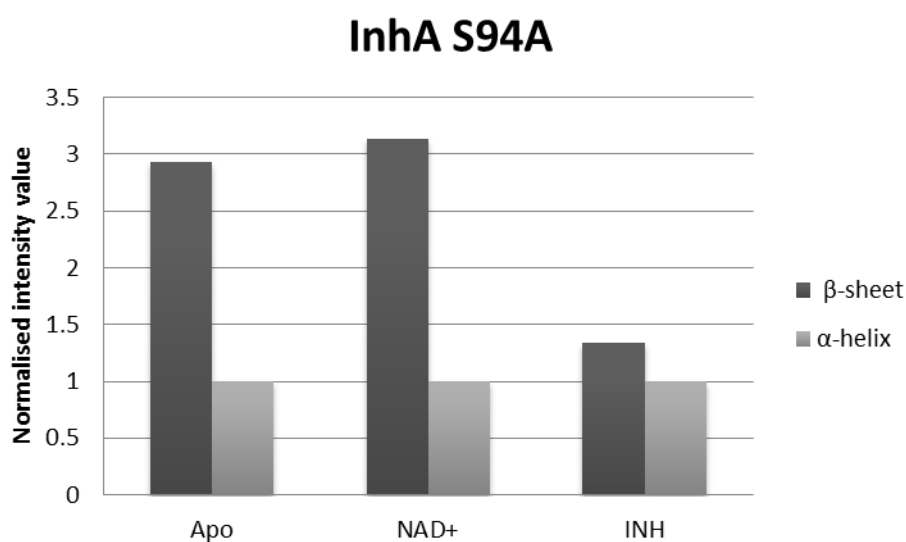
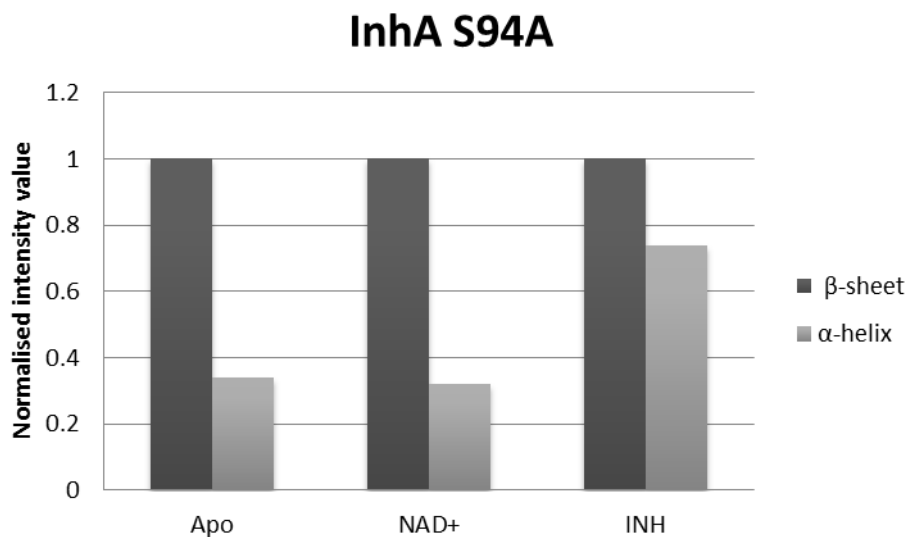


Figure 6.13 **(A)** Comparison of peak intensities obtained from cross sections of 2D-IR measurement of InhA S94A normalised to the β -sheet intensity **(B)** Comparison of peak intensities normalised to α -helix intensity. Changes can be observed between the β -sheet- α -helix interaction upon NAD⁺ and drug binding.

6.4 Preliminary 2D-IR spectroscopic analysis of HutC

Analysis of InhA and INH has provided a better understanding of conformational changes within proteins which happen in ultrafast timescales. These data have provided much insight into ultrafast binding dynamics between proteins and ligands. Only one 2D-IR measurement on GntR proteins has been carried out thus far due to time constraints. HutC data were collected on ULTRA with a waiting time of 250 femtoseconds. Analysis of the data reveals that secondary structure elements are clearly visible in the amide I region (**Figure 6.14A**). The β -sheet element is clearly reflected by the intensity at 1620 cm^{-1} in the 2D-IR spectrum shown here. The crystal structure of the Eb/O domain from the Hut repressor from *P. syringae* is shown for comparison (**Figure 6.14B**). These data suggest that clear structural elements can be isolated for this protein and suggest that it can lend itself to 2DIR analysis.

This data, along with the well-established dynamics demonstrated for the InhA systems, can be taken forward and be further applied to the study of protein-DNA and protein-ligand interactions within the GntR superfamily.

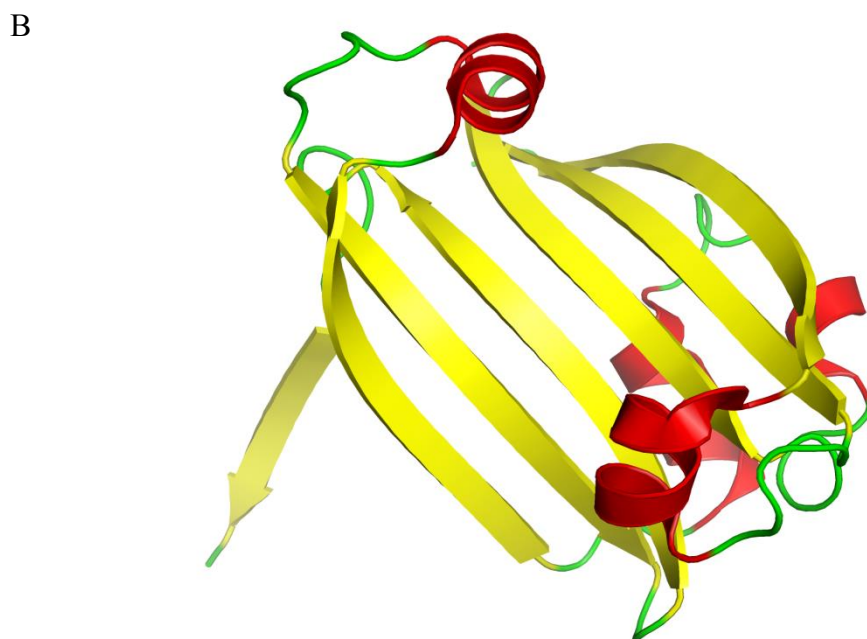
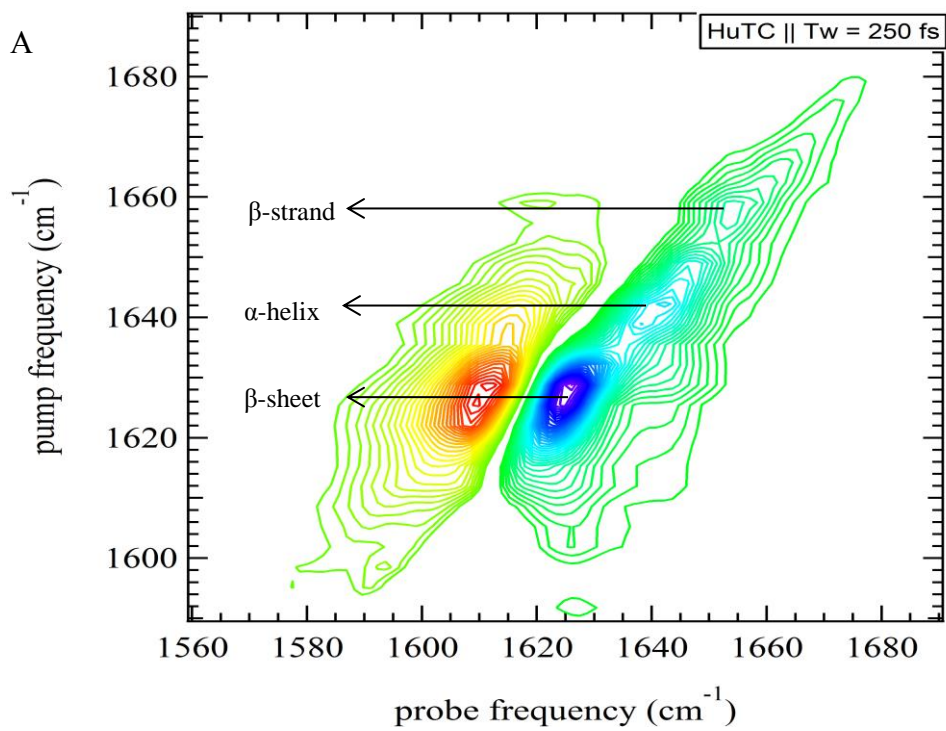


Figure 6.14 (A) 2D-IR spectra of HutC. Secondary structure elements are evident. Data were measured on ULTRA with a $T_w = 250$ fs. (B) Structural representation of the Eb/O domain of the Hut repressor of *P. syringae* (PDB code 2PKH).

Chapter 7 General discussion

This project aimed to address some of the unknowns in terms of functional and dynamical information about the GntR superfamily of transcription factors. Fulfilling the aims of the project experimentally has proved challenging at times and much time has been spent overcoming these difficulties.

Novel cloning targets were identified through sequence and phylogenetic analyses of GntR-like sequences deposited in the Pfam database and were subject to bioinformatics analyses in the C-terminal Eb/O domains. Bioinformatic analyses revealed secondary structures related to subfamilies. Gp26 is putatively a new subfamily as it clustered alone in phylogenetic analysis. Four of these targets (HutC, DevA, DevE and Gp26) were subject to structural, functional and dynamic analysis.

Difficulties were encountered in obtaining sufficient quantities of protein for crystallography. Great efforts were put into constructing and optimising plasmids that expressed sufficient quantities of soluble protein for this aspect of the project. Furthermore, much time was spent on optimisation of purification procedures to obtain protein of sufficient quality and stability for crystallisation. Once sufficient quantities of stable proteins were obtained, crystallisation proved challenging. Despite great efforts in this area only HutC and DevE have been successfully crystallised and diffraction data measured at one of the macromolecular crystallography beamlines at Diamond Light Source. Protein crystallography is notoriously difficult as there are so many factors that can affect crystallisation such as protein concentration, temperature, pH, precipitant, buffer and pressure. A list of

37 physical, chemical and biochemical factors affecting crystallisation was reviewed by (Russo Krauss *et al.*, 2013) giving an indication of the difficulties which are faced during this process.

Sample purity and homogeneity are key factors that affect the ability to grow crystals. The four targets tested in this project were all analysed by SDS-PAGE to assess purity and SEC-MALLS to assess homogeneity. All four protein samples were assessed to be of high quality and monodisperse making them suitable candidates for crystallisation.

Varying protein concentrations were used during crystallisation screening to assess the best concentration to reach a suitable supersaturation level for nucleation to occur. Nucleation was observed for all targets tested but crystals failed to grow to a suitable size and quality in the case of DevA to allow any reasonable data to be collected, whilst for Gp26 crystals were also of poor quality but initial diffraction confirmed the crystals to be protein versus salt.

HutC

HutC was included in target selection primarily as a means to validate functional experiments on other targets due to previous knowledge and as a model to take to 2D-IR work as the effector ligand was known. During this project the structure of NagR was made available in the PDB and subsequently published (Fillenberg *et al.*, 2015). As the structural data obtained here was comparable, a decision was made to focus on DevE for which structural data on this subclass of GntR remains sparse. The HutC structure determined here was obtained in the absence of the known effector molecule urocanic acid or DNA. This resulted in a structure where the N-terminal

DNA binding domains were in close proximity to the conformation observed in the structure of the NagR: 19mer DNA complex, thus providing another snapshot of the conformational states of the HutC/GntR subclass to complement the effector-bound and DNA-bound structures of NagR. Furthermore, biophysical and structural data presented in this project have revealed HutC to be a dimer in agreement with previous literature and reinforces the view that most GntR regulators are either stable dimers in solution or dimerise upon binding to their DNA operator sites (DiRusso *et al.*, 1998, Rigali *et al.*, 2002, Hoskisson & Rigali, 2009).

DNA binding studies reveal that HutC from PA14 binds to DNA with lower affinity than its homologues from *Bacillus subtilis* (NagR and YvoA). This is likely due to the differing techniques used to determine binding affinities (EMSA vs Biacore) or may reflect different physiological roles of HutC between these species.

HutC is the only protein that was subject to dynamical analysis by 2D-IR spectroscopy due to time constraints and the lack of a suitable model crystal structure for DevA, DevE and Gp26. The spectra revealed the secondary structural features of HutC, present as distinct modes indicating that it is highly likely that other GntR proteins would be suitable for analysis by the 2D-IR method. Furthermore, work carried out in the InhA/isoniazid system (see below) provides an ideal model for future work on GntR-DNA and GntR-ligand binding studies. As the signal from isoniazid has been previously shown to be negligible (Shaw *et al.*, 2015), any changes arising in the 2D-IR spectra can be attributed to changes in the protein structure upon ligand or DNA binding. 2D-IR spectroscopy has been brought into the fore during this project with pharmaceutical companies currently exploring the technique to characterise protein-drug interactions as means to supplement circular

dichroism data. This interest is directly based on the characterisation of InhA/isoniazid molecular system presented here.

DevA and DevE

Real interest lies in the DevA sub-family that currently contains proteins found in sporulating actinomycetes. This work included DevA and DevE as targets to provide structural data in order to try to elucidate effector molecules that may give further clues to functionality and the mechanisms by which these duplicated genes and their gene products affect the developmental biology of actinomycetes. It has been shown that DevA interacts with single stranded binding protein (SsbB) and topoisomerase in bacterial two hybrid assays (Vujaklija and Jakimowicz unpublished) which raises questions about possible effector molecules being implicated in DNA replication/repair or sensing chromosome integrity during sporulation. DevA controls the expression of *devB*, a putative phosphoglycolate phosphatase (Hoskisson *et al.*, 2006). Phosphoglycolates often occur at the site of single strand breaks in DNA and require removal to leave free 5'-OH on DNA prior to DNA repair (Teresa Pellicer *et al.*, 2003, Murray *et al.*, 2014). To test this hypothesis, *devA* and *devB* mutants were tested for sensitivity to bleomycin and phleomycin, both of which introduce single strand breaks in DNA. It was found that the mutants were 5-fold more sensitive to these antibiotics (Hoskisson, unpublished).

It was hoped that the structure of DevA would give an indication of what the effector could be; however despite all efforts to crystallise DevA, no structural data were obtained as crystallisation proved unsuccessful. Biophysical characterisation by gel filtration chromatography and SEC-MALLS has revealed DevA to be a dimer, in

agreement with previous SELDI mass spectrometry analysis of DevA (Hoskisson *et al.*, 2006, Hoskisson & Rigali, 2009).

DevE was successfully crystallised and two crystal forms showed diffraction to resolutions of ~ 2.7 Å. A DevE: 18mer DNA complex was also successfully isolated and characterised although as yet it is not clear if crystals of this complex have been obtained. Although, diffraction data are now available, structure solution to date has proved unsuccessful by the method of Molecular Replacement or by anomalous scattering through the production of SeMet labelled protein due to non-diffracting crystals. Efforts are continuing to solve this structure, which it is hoped, will yield valuable information with regards to potential effector molecules. It would be the first reported crystal structure of a DevA subfamily member. The problems encountered with obtaining crystals can be to some extent attributed to the flexibility between the N-terminal and C-terminal domains of GntR family of proteins. This flexibility is made possible by the relatively long linker region between the HTH and Eb/O domains.

All is not lost if the structure of DevE cannot be solved; small angle x-ray scattering (SAXS) measures scattering of x-rays at low angles (typically in the range of 0.1 - 10°) which can give information on the size and shape of macromolecules such as proteins. The advantage of SAXS is that it is a solution technique and hence proteins samples once purified can be analysed directly by the method however, the resolution is considerably lower than that of x-ray crystallography. Failing structure solution of DevE by x-ray crystallography, this would be an ideal candidate for detecting induced conformational changes in the presence and absence of DNA and would further add to our knowledge of GntR mechanism and function.

Unfortunately, scheduling this work was not possible prior to submission of the thesis due to access to equipment.

Upstream regions of *devA* and *devE* were analysed *in silico* to attempt to determine potential promoter binding regions. This work has uncovered two potential promoter sites for both DevA and DevE which are imperfect, inverted repeats. This is in contrast to previous literature suggesting that operator sequences in the GntR family are palindromic sequences (Rigali *et al.*, 2002, Rigali *et al.*, 2004). It is entirely possible that other patterns exist in promoter sites as the number of promoter regions known is comparatively low compared to the number of GntR sequences contained within the Pfam database. The unusual sequence found in the promoter region probably impacts upon the dimerisation arrangement of these proteins when bound to their operator sites. Such is the case with the MocR subfamily that appear to bind their operator sites in a head to tail fashion in a similar to the class I aminotransferase enzymes (Bramucci *et al.*, 2011). The MocR operator binding sites, however, appear as directed repeats which have large spacers that allow DNA looping which is impacted by the dimerisation arrangement of the regulator (Rigali *et al.*, 2002).

Interestingly, during this project DevE was found to bind to the 233 bp upstream region of *devA*. In contrast DevA was found not to bind to the 240 bp region upstream of *devE*. This suggests that the HTH domain has been highly conserved to the extent that DevE still has recognition for the *devA* promoter following the duplication event (Clark & Hoskisson 2011). Moreover, given that DevA did not interact with the *devE* promoter region suggests that DNA binding may have been non-specific due to the length of the DNA fragments used. Further testing is required *in vivo* with knockout mutations of *devA* and *devE* to determine if this a true

biological effect or an effect of the relatively high protein concentrations used in the EMSA skewing the balance of the protein-DNA interaction.

While this work reports the first potential promoter sequences for DevA and DevE; these cannot be considered definitive and require to be confirmed by other techniques such as DNase I foot printing. Following this, DNA binding should be re-evaluated using techniques such as ITC or SPR to gain a more accurate K_D value than those achieved by EMSA during this work. Unfortunately time constraints didn't allow this to be carried during this project.

Demonstrating the relevance of 2D-IR to protein-ligand interaction studies

Application of 2D-IR spectroscopy to study the binding of the anti-TB drug isoniazid to InhA revealed the great potential this technique has for protein ligand binding studies. Spectra obtained of isoniazid bound to WT InhA and the mutant proteins indicated that the amide I band of the protein masks the signal from the drug, and therefore no contribution is made to the signal by isoniazid because it is present in a 1:1 ratio with the protein (1 mM). Therefore the signal from isoniazid is dwarfed by that of the 269 amino acid residues of InhA. Furthermore the absorption from 1 mM isoniazid is negligible as a 70 mM solution was required to see a signal in the 2D-IR spectra. Thus, any changes in the InhA spectrum upon isoniazid binding are the result of changes to the protein caused by the ligand. This is important because it avoids the need to insert spectroscopic labels into the structure, which have previously hindered 2D-IR studies (Adamczyk, Candelaresi, Robb, *et al.*, 2012).

Comparing the spectrum of WT InhA with that of the S94A mutant revealed changes in the off-diagonal region of the spectrum. Crucially, 1D-IR spectra of the same

samples did not show any differences, highlighting the ability of 2D-IR to reveal new details (Data not shown). We have compared several point mutations surrounding the drug-binding site of InhA and the magnitude of the increase correlates well with the degree of enzymatic inhibition caused by isoniazid. This is a potentially exciting discovery but invites the question: what is the origin of this feature? Linking diagonal parts of the amide I lineshape located at ~ 1620 with those at ~ 1650 cm^{-1} (**Fig 6.4 & 6.8**) suggest there are two possible reasons for this: **1**) there is a change in coupling between residues involved in a β -sheet (which normally absorb at a frequency of 1620 cm^{-1}) and those in an α -helix (1650 cm^{-1}) as a result of drug binding or **2**) a β -sheet in the InhA structure is becoming less structured (more flexible/dynamic) upon isoniazid binding.

These data suggest that 2D-IR may have a big future in its application to drug binding studies with their target proteins, due to the sensitivity in detecting molecular flexibility/dynamics.

Concluding remarks

The importance of GntR proteins cannot be underestimated, given that the number and diversity of processes which are regulated by these transcription factors are vast. Much more work is required in this area of research given the rise in antibiotic resistance; some of these novel transcription factors have the potential to provide new antibiotic targets. Considering they have been implicated in virulence and biofilm formation they could hold great clinical significance. Rational drug design will be greatly enhanced by further understanding of GntR protein structures and





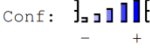

elucidation of effector molecules.

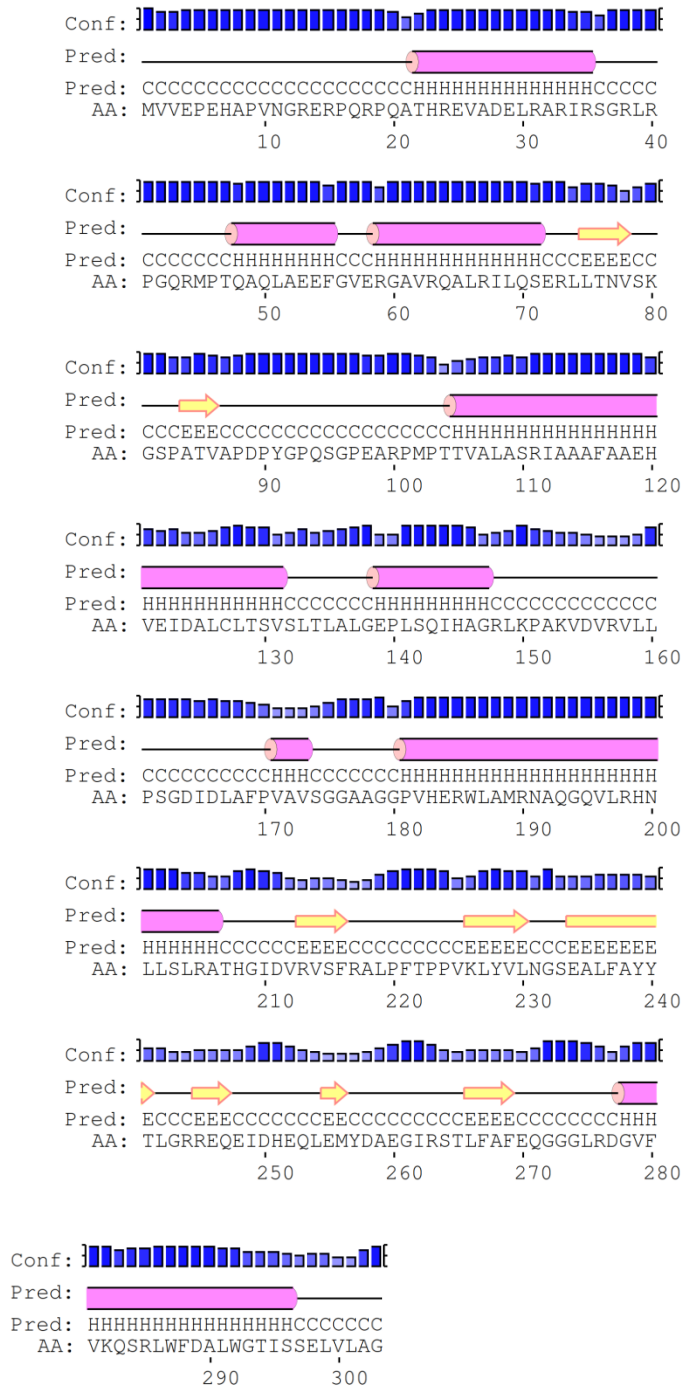
Future work in this area should include experiments to confirm the promoter sequences of DevA and DevE presented in this project. Further structural work is required for DevA, DevE and Gp26. In parallel to continuing to search for suitable crystallisation conditions for DevE and Gp26, solution SAXS experiments would be an ideal starting point to gain more insight into the conformational states of the proteins in the absence and presence of effector molecules and DNA. As for other GntR proteins, for example, PhnF from *E. coli* (Gorelik *et al.*, 2006), work can also be directed to isolating the C-terminal domains of these proteins with a view to facilitating crystallisation of this domain independent of the HTH domain and aid identification of potential effector molecules for the *S. coelicolor* DevA and DevE GntR proteins.

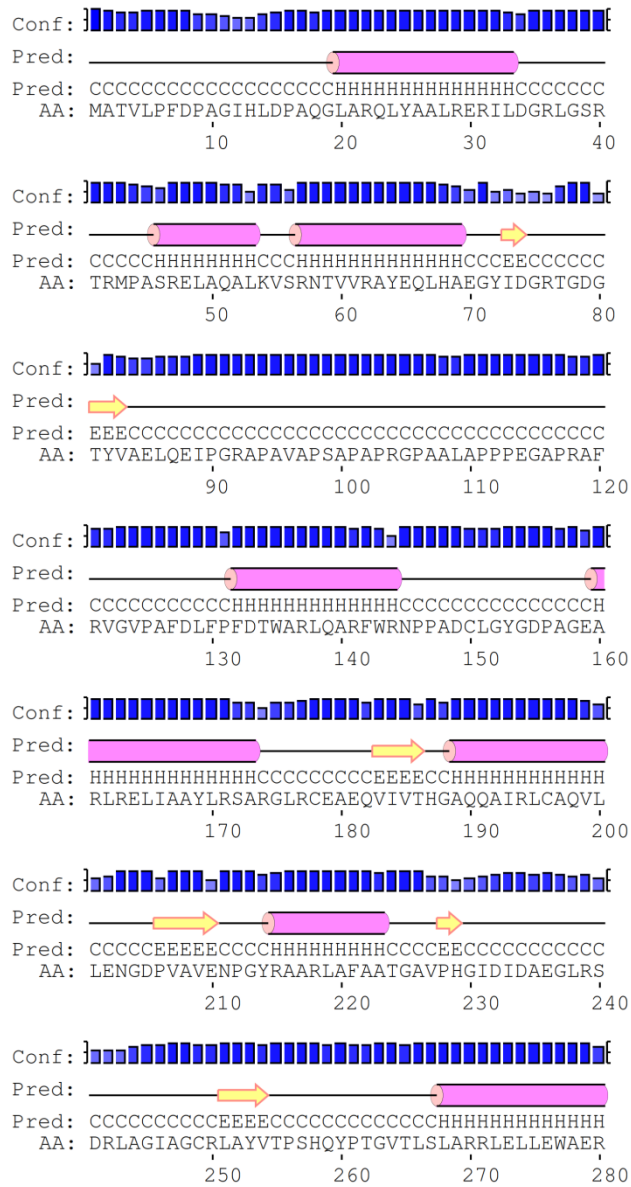
Further work on the 2D-IR spectroscopy system of molecular dynamics will provide further insights into the structure/function relationships. Measurement of HutC bound to its DNA promoter region and to its effector molecule, urocanic acid, is the next logical step for this work.

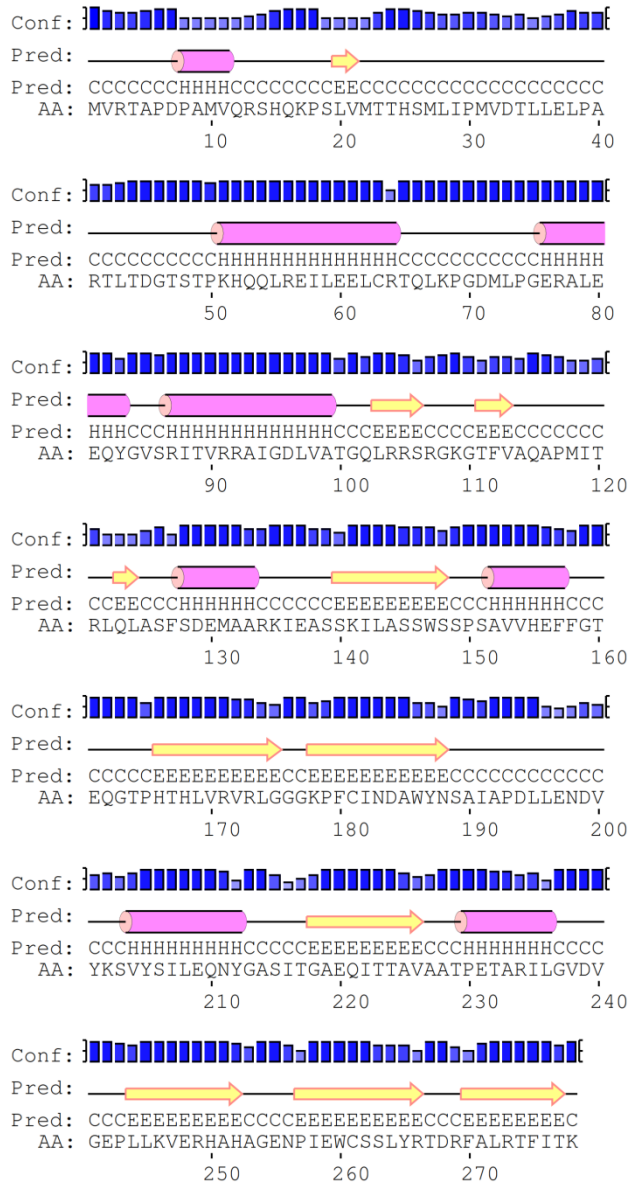
Appendix 1 Secondary structure predictions

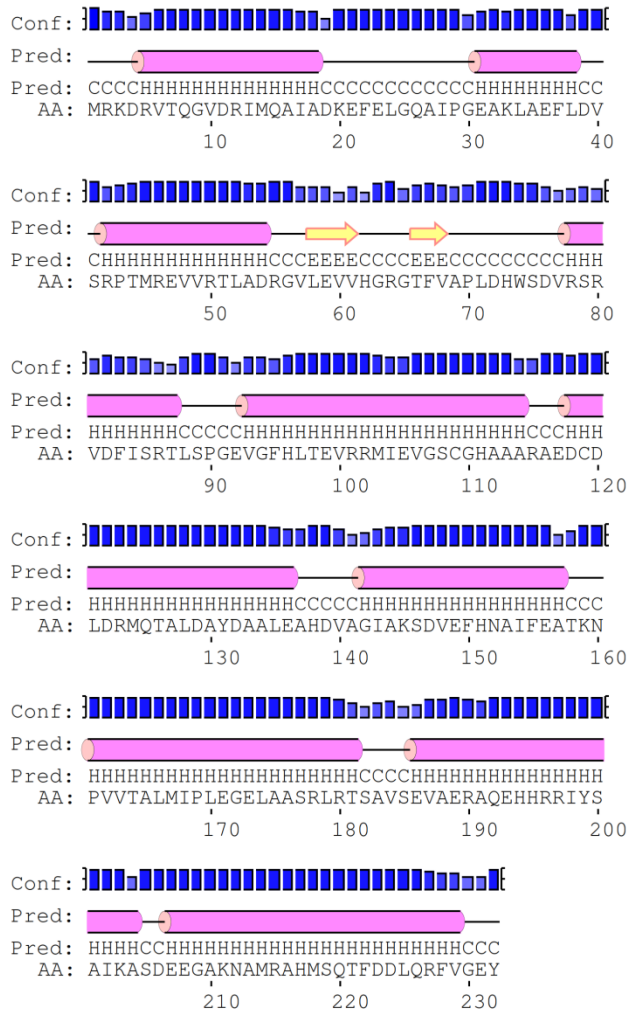
The key to the following secondary structure predictions are as follows:


Legend:	
	= helix
	= strand
	= coil
Conf: 	= confidence of prediction
Pred: 	= predicted secondary structure
AA: 	= target sequence











Conf: 


 Pred: 

 Pred: HHCCCHHHHHHHHHHHHHHHCCCEEECCCEEECCCHH

 AA: LYNVSDTTLKAIQILQEWGVVKAVRSKGFVMMDIQALK

 290 300 310 320


Conf: 


 Pred: 

 Pred: CCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCH

 AA: KIEIPPHLIACHVRRYLDTLDLLALTIEGVSAYAAEHISQ

 330 340 350 360

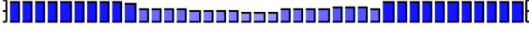
Conf: 


 Pred: 

 Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

 AA: KEIEEAMSEIKRCWNEDHRYMLTPSFLNLIKHTGDGSL

 370 380 390 400


Conf: 


 Pred: 

 Pred: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

 AA: NAIYILLRQNLGIGRSIPALRETEKTAEDYELYEQVTAL

 410 420 430 440


Conf: 

 Pred: 

 Pred: HHHHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

 AA: NQLYAGRQEDFSNGTAKAFRYIYDYVTEKCKNLGYDSAM

 450 460 470 480

Conf: 

 Pred:

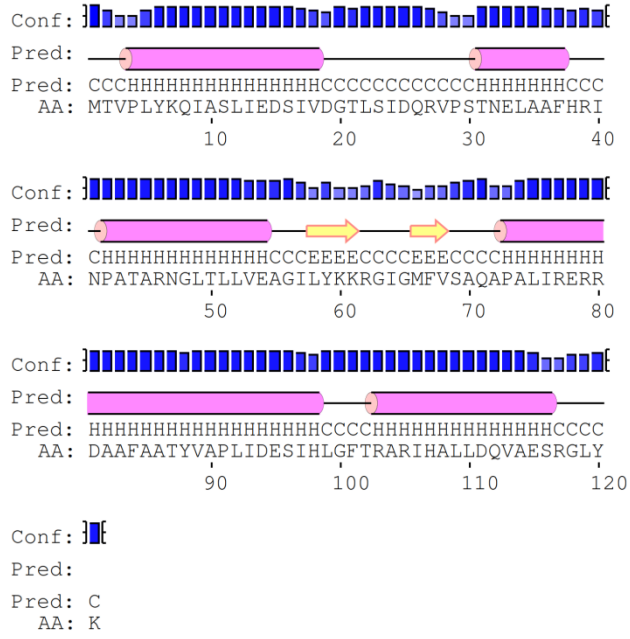
 Pred: CCCCCCCCC

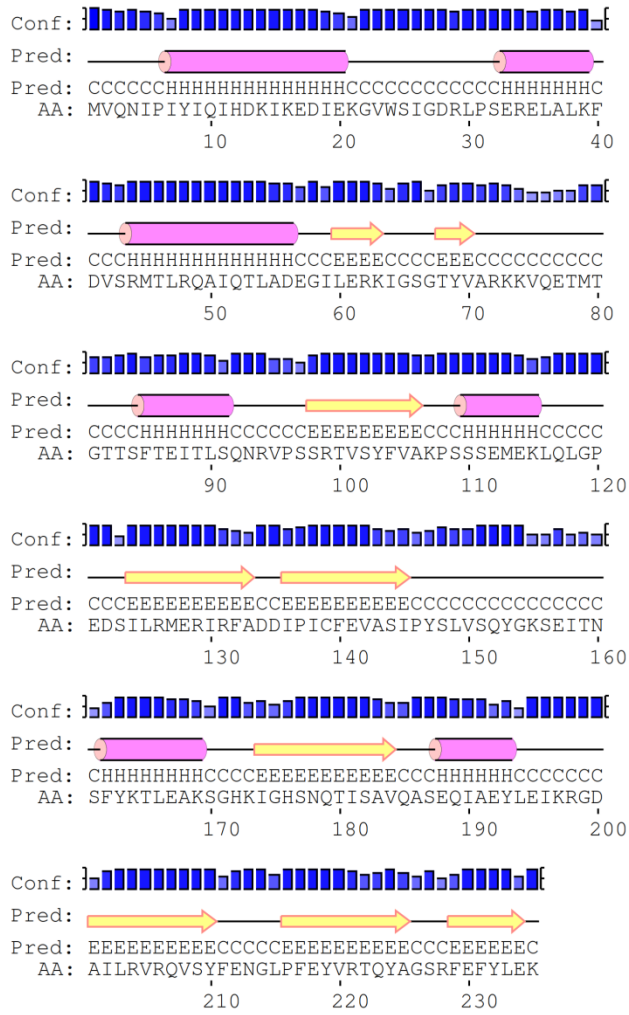
 AA: AVYDGSALWK

 490

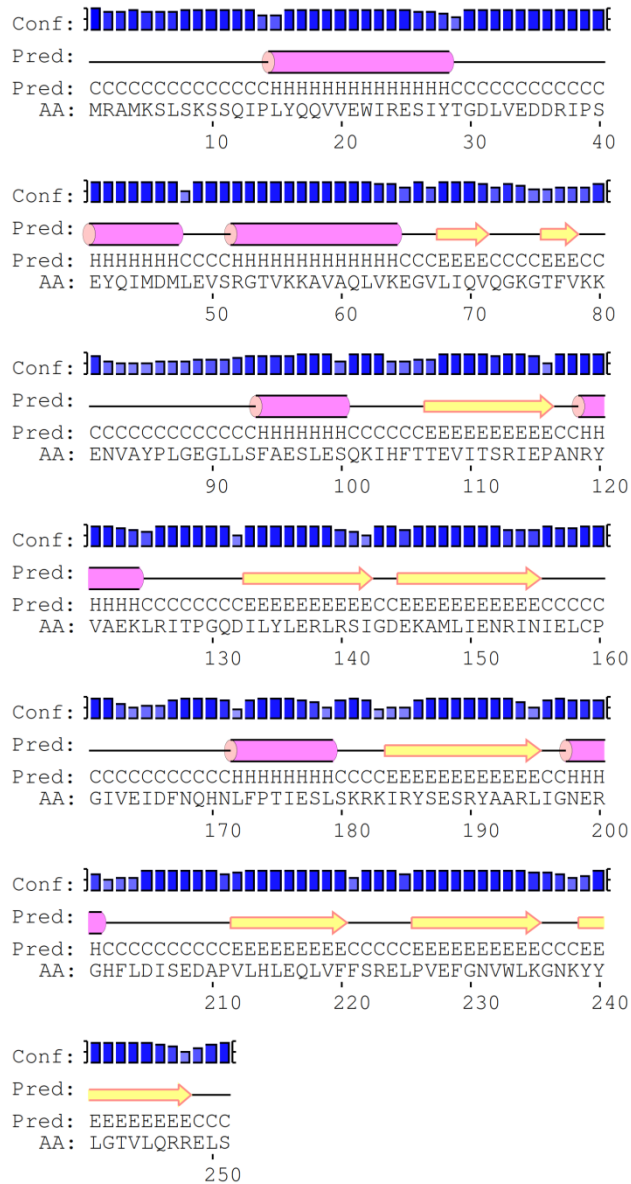
cg3261

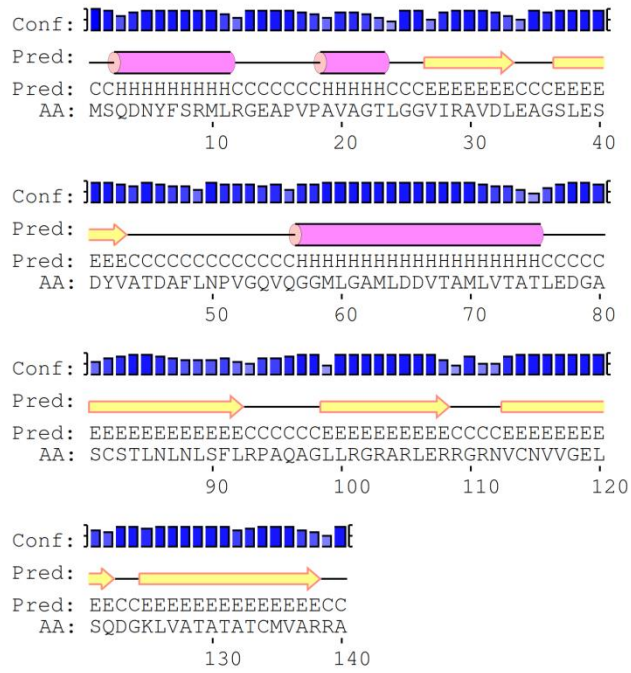
Regulatory protein, GntR Q8NLJ5





c4276 Putative regulator Q8FCM7



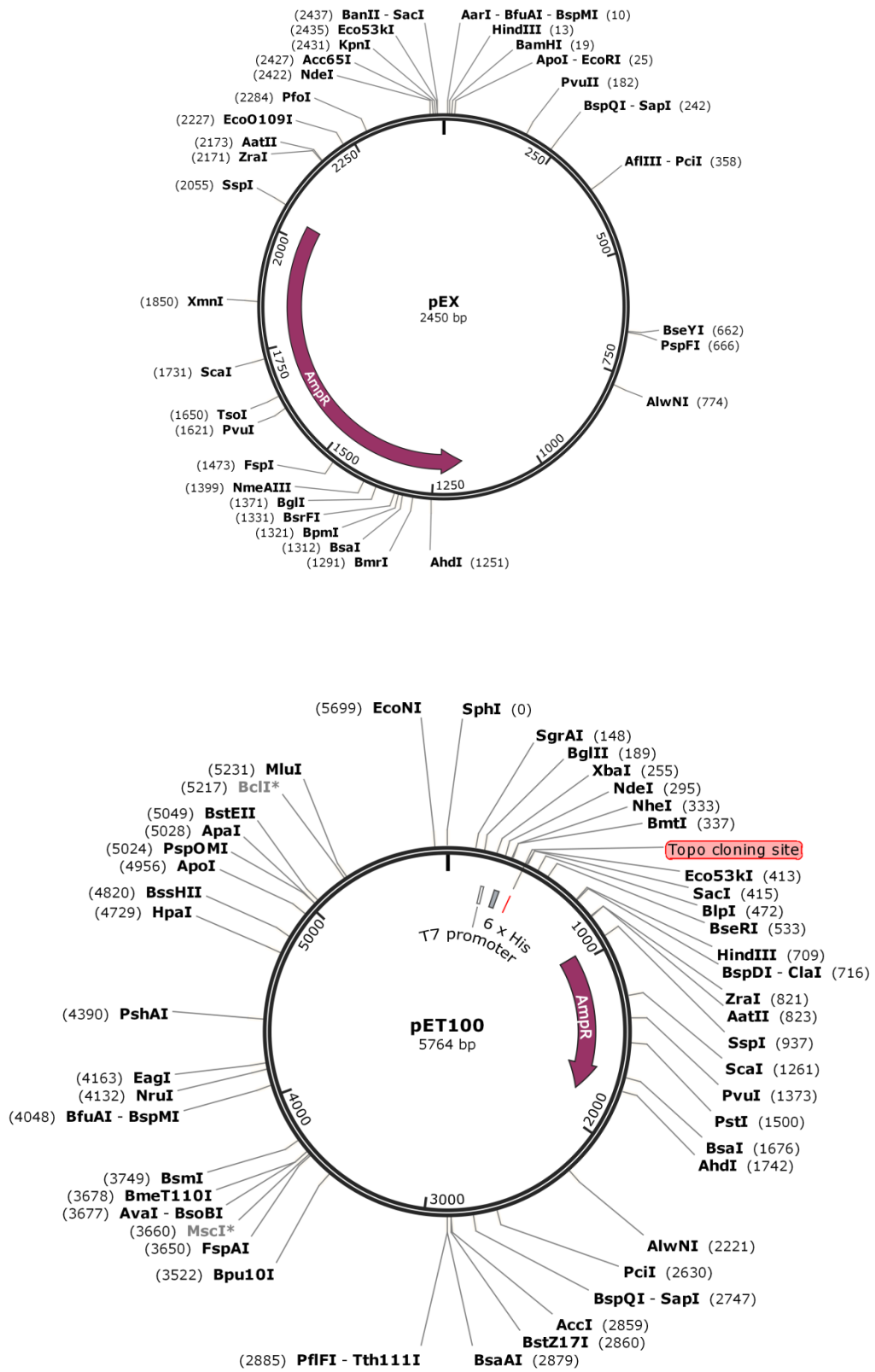


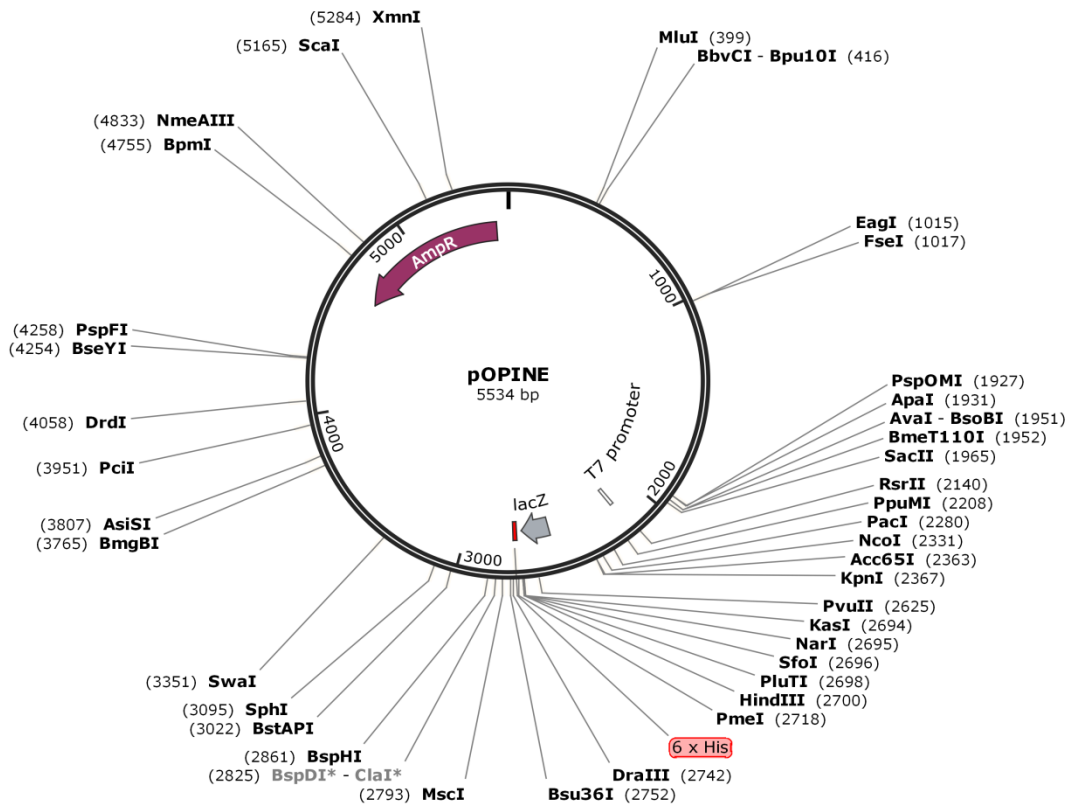
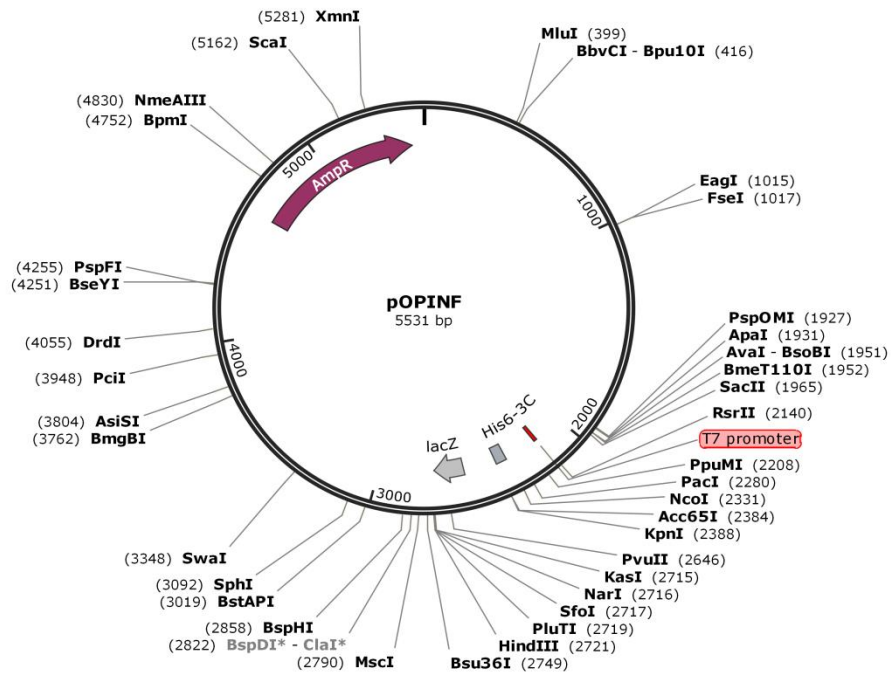
Appendix 2 Expression test results for HTP plasmids

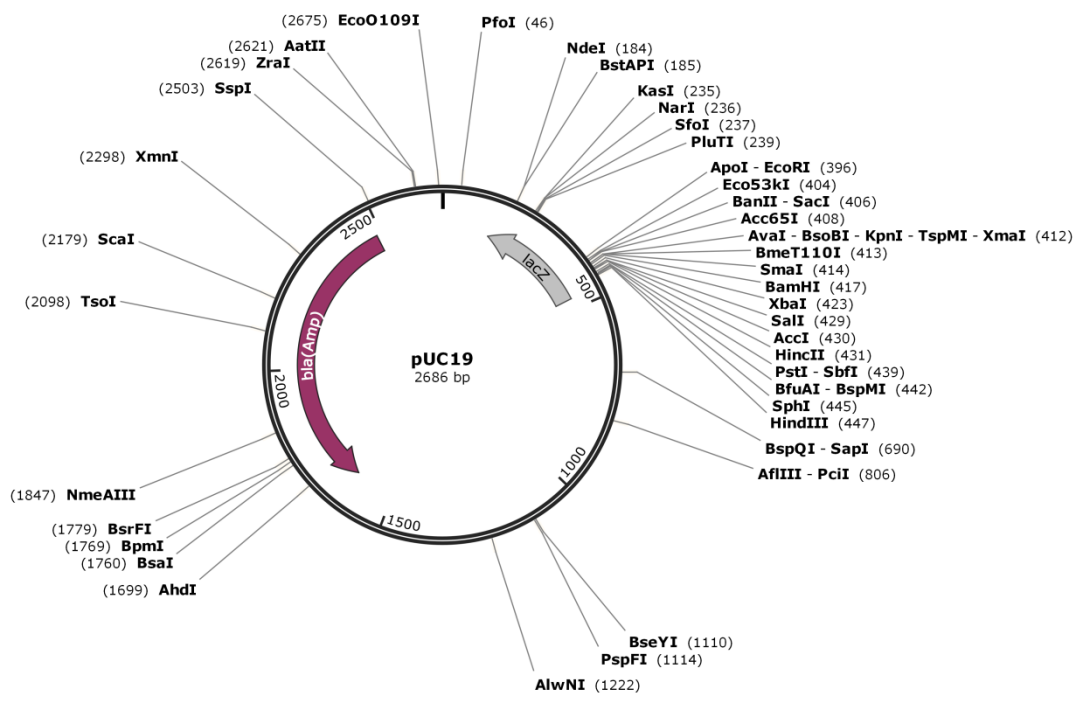
Well	Gene name	ID No.	MW	Lemo21		Rosetta 2	
				IPTG	Auto	IPTG	Auto
A01	DEVA_native	16279	32890	++	++	++	++
B01	DEVA_native	16280	31240	-	-	-	-
C01	DEVA_native	16281	31240	-	-	-	-
D01	DEVA_native	16282	44890	++	++	+	+
E01	DEVA_native	16283	43240	-	-	-	-
F01	DEVE_native	16291	34210	-	+	+	++
G01	DEVE_native	16292	32120	++	++	++	++
H01	DEVE_native	16293	29370	-	-	++	+
A02	DEVE_native	16294	32120	+	+	+++	+++
B02	DEVE_native	16295	29260	+	-	-	-
C02	DEVE_native	16296	46210	++	++	-	+
D02	STRCO Putative	16322	23650	-	-	-	+
E02	STRCO Putative	16323	22990	-	-	+	++
F02	STRCO Putative	16324	23650	-	-	-	-
G02	STRCO Putative	16325	22990	-	-	-	+
H02	HUT_native	16305	26400	-	+	-	+
A03	HUT_native	16303	26290	++	++	++	++
B03	HUT_native	16304	28380	+	+	++	++
C03	HUT_native	16306	26400	-	+	+	+++
D03	HUT_native	16307	26290	-	+	++	+++
E03	PA14_34660 GntR gene	16316	38610	-	-	-	-
F03	PA14_34660 GntR gene	16317	37070	-	-	-	-
G03	PA14_34660 GntR gene	16318	36080	-	-	-	-
H03	PA14_34660 GntR gene	16319	38610	-	-	-	-
A04	PA14_34660 GntR gene	16320	36960	++	-	+	-
B04	PA14_34660 GntR gene	16321	36080	-	-	-	-
C04	P. fluorescens	16308	59840	-	-	-	-
D04	P. fluorescens	16309	58960	-	-	-	-
E04	P. fluorescens	16310	59860	-	-	-	-
F04	P. fluorescens	16311	59840	-	-	-	-
G04	Gp26	16312	20020	-	++	-	-
H04	Gp26	16313	17930	+	++	++	++
A05	Gp26	16314	19910	++	+	++	+
B05	Gp26	16315	17820	-	-	-	-
C05	DEVA_SERp	16284	31240	+	-	+	-
D05	DEVA_SERp	16285	43240	-	-	-	-
E05	DEVA_SERp	16286	31240	-	-	-	-
F05	DEVE_SERp	16287	32120	-	+	-	+
G05	DEVE_SERp	16288	29260	-	-	-	-

H05	DEVE_SERp	16289	32120	-	+	-	+
A06	DEVE_SERp	16290	29260	-	-	-	-
B06	HUT_SERp	16297	26400	-	-	-	-
C06	HUT_SERp	16298	26180	-	-	-	-
D06	HUT_SERp	16299	26290	-	-	-	-
E06	HUT_SERp	16300	26400	-	-	-	-
F06	HUT_SERp	16301	26180	-	-	-	-
G06	HUT_SERp	16302	26290	-	-	-	-
H06	GFP positive control	N/A	27000	+++	+++	+++	++
No expression (-), Low expression (+), Good Expression (++) , Very good expression (+++)							

Appendix 3 Vector maps







Appendix 4 Thermofluor screening conditions

	1	2	3	4	5	6	7	8	9	10	11	12	
A	Citric Acid pH 4.0 No Salt	NaAc pH 4.5 No Salt	Phos. pH 5.0 No Salt	Citrate pH 5.5 No Salt	Bis-tris pH 6.0 No Salt	ADA pH 6.5 No Salt	MOPS pH 7.0 No Salt	HEPES pH 7.5 No Salt	Imidazole pH 8.0 No Salt	Tris pH 8.5 No Salt	CHES pH 9.0 No Salt	CHES pH 9.5 No Salt	No Salt
B	Citric Acid pH 4.0 100mM NaCl	NaAc pH 4.5 100mM NaCl	Phos. pH 5.0 100mM NaCl	Citrate pH 5.5 100mM NaCl	Bis-tris pH 6.0 100mM NaCl	ADA pH 6.5 100mM NaCl	MOPS pH 7.0 100mM NaCl	HEPES pH 7.5 100mM NaCl	Imidazole pH 8.0 100mM NaCl	Tris pH 8.5 100mM NaCl	CHES pH 9.0 100mM NaCl	CHES pH 9.5 100mM NaCl	100mM NaCl
C	Citric Acid pH 4.0 200mM NaCl	NaAc pH 4.5 200mM NaCl	Phos. pH 5.0 200mM NaCl	Citrate pH 5.5 200mM NaCl	Bis-tris pH 6.0 200mM NaCl	ADA pH 6.5 200mM NaCl	MOPS pH 7.0 200mM NaCl	HEPES pH 7.5 200mM NaCl	Imidazole pH 8.0 200mM NaCl	Tris pH 8.5 200mM NaCl	CHES pH 9.0 200mM NaCl	CHES pH 9.5 200mM NaCl	200mM NaCl
D	Citric Acid pH 4.0 500mM NaCl	NaAc pH 4.5 500mM NaCl	Phos. pH 5.0 500mM NaCl	Citrate pH 5.5 500mM NaCl	Bis-tris pH 6.0 500mM NaCl	ADA pH 6.5 500mM NaCl	MOPS pH 7.0 500mM NaCl	HEPES pH 7.5 500mM NaCl	Imidazole pH 8.0 500mM NaCl	Tris pH 8.5 500mM NaCl	CHES pH 9.0 500mM NaCl	CHES pH 9.5 500mM NaCl	500mM NaCl

Appendix 5 Upstream DNA fragments used in EMSAS

HutC

CGGGACGAATCTCGGCGAGAACTCCGTCCGCGGAAATCTCGAAACGGACATTTTCGCG
CCCAGCCTTCCGGCAGCAACGCACGTTTCAGCGAAAATTGCGGACATAATCAGCTCAC
ACGCAGTCTGTTTTATTTGTATATACATATACAGACGATTAAGCGGTGCGTAAACTC
ACTTCATCGACGCCTCGACAAGGAGTCTCTCCGTGACGTCTCTTCTTCCGATCGTT
CCCCGCTGG

DevA

CTGCTCGAAGGCGATGACGAAGACCGCGGAGACCAGGGCGAGGACGGGCATGGCCTT
CCCCCTTCGGGAAGATAGACTTTCGCCAACTGGGCGAAGTTGGCAACCAACTCTCC
CTATGTTGTCCC CACTTGGCTACTACCTATAAACAAGTTTCAAACAACCTCCCTATAG
GTAGGTCGAAGTTGTAGCGTTTGGTCGTGACTCAGGAGAACGTGTCCGTGAACGGCA
GCAGAA

DevE

CCCGTACTTCCACTGCACGAGCTGCTCGAACCCGATGGCCAGAGCCGCGACGACGAT
GGCGATGAGCGGCATGGTGGTCCCCCTGTCACCGTGGTGGCCAACCCTCGTGAGGCT
CGACCACGTTGCTC AAGTTGGCTGCCAACTTCACTGTA CT TATCTC CGCTTGGAAGC
GTCTCATA ACCACCTTTCGCGGAACTGCCGGAAGATGGCGAGAAGTTGTAGGGTTTG
GTCGTGGAGCCG

Consensus sequence for HutC is highlighted in yellow

Potential promoter regions for DevA and DevE are highlighted in green

Appendix 6 Tables of normalised peak intensities obtained for 2D-IR spectroscopy data

WT <i>apo</i> - form	β -sheet (cm^{-1})	α -helix (cm^{-1})	β -strand (cm^{-1})
Frequency (cm^{-1})	1618	1648	1656
Intensity (a. u.)	-0.158	-0.133	-0.107
β -sheet normalised	1.00	0.84	0.68
α -helix normalised	1.19	1.00	0.80
β -strand normalised	1.48	1.24	1.00

WT:NADH (1:1)	β -sheet (cm^{-1})	α -helix (cm^{-1})	β -strand (cm^{-1})
Frequency (cm^{-1})	1618	1648	1656
Intensity (a. u.)	-0.077	-0.075	-0.059
β -sheet normalised	1.00	0.98	0.76
α -helix normalised	1.02	1.00	0.78
β -strand normalised	1.31	1.28	1.00

WT:NADH:INH (1:1:1)	β -sheet (cm^{-1})	α -helix (cm^{-1})	β -strand (cm^{-1})
Frequency (cm^{-1})	1618	1648	1656
Intensity (a. u.)	-0.113	-0.114	-0.067
β -sheet normalised	1.00	1.01	0.59
α -helix normalised	0.99	1.00	0.59
β -strand normalised	1.69	1.70	1.00

S94A <i>apo</i> - form	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1616	1648	1656
Intensity (a. u.)	-0.136	-0.047	-0.040
β -sheet normalised	1.00	0.34	0.29
α -helix normalised	2.93	1.00	0.85
β -strand normalised	3.43	1.17	1.00

S94A:NADH (1:1)	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1616	1648	1656
Intensity (a. u.)	-0.039	-0.013	-0.009
β -sheet normalised	1.00	0.32	0.23
α -helix normalised	3.13	1.00	0.71
β -strand normalised	4.41	1.41	1.00

S94A:NADH:INH (1:1:1)	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1616	1648	1656
Intensity (a. u.)	-0.067	-0.050	-0.028
β -sheet normalised	1.00	0.74	0.42
α -helix normalised	1.34	1.00	0.56
β -strand normalised	2.40	1.79	1.00

P193A <i>apo</i> - form	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1618	1648	1656
Intensity (a. u.)	-0.295	-0.141	-0.098
β -sheet normalised	1.00	0.48	0.33
α -helix normalised	2.09	1.00	0.70
β -strand normalised	3.01	1.44	1.00

P193A:NADH (1:1)	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1618	1648	1656
Intensity (a. u.)	-0.162	-0.080	-0.053
β -sheet normalised	1.00	0.49	0.33
α -helix normalised	2.03	1.00	0.66
β -strand normalised	3.06	1.51	1.00

P193A:NADH:INH (1:1:1)	β -sheet (cm ⁻¹)	α -helix (cm ⁻¹)	β -strand (cm ⁻¹)
Frequency (cm ⁻¹)	1618	1648	1656
Intensity (a. u.)	-0.110	-0.054	-0.036
β -sheet normalised	1.00	0.49	0.33
α -helix normalised	2.04	1.00	0.67
β -strand normalised	3.06	1.50	1.00

References

- Adamczyk, K., Candelaresi, M., Kania, R., Robb, K., Bellota-Antón, C., Greetham, G. M., Pollard, M. R., Towrie, M., Parker, A. W., Hoskisson, P. A., Tucker, N. P. & Hunt, N. T. (2012). *Phys Chem Chem Phys* **14**, 7411-7419.
- Adamczyk, K., Candelaresi, M., Robb, K. A., Walsh, M., Parker, A., PA, Tucker, N. & Hunt, N. (2012). *Meas. Sci. Technol.* **23**.
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 213-221.
- Allison, S. L. & Phillips, A. T. (1990). *J Bacteriol* **172**, 5470-5476.
- Arai, H., Akahira, S., Ohishi, T. & Kudo, T. (1999). *Mol Microbiol* **33**, 1132-1140.
- Aravind, L. & Anantharaman, V. (2003). *FEMS Microbiol Lett* **222**, 17-23.
- Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). *FEMS Microbiol Rev* **29**, 231-262.
- Artursson, V., Finlay, R. D. & Jansson, J. K. (2006). *Environ Microbiol* **8**, 1-10.
- Bailey, T. L. & Elkan, C. (1994). *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36.
- Banerjee, D. & Stableforth, D. (2000). *Drugs* **60**, 1053-1064.
- Banumathi, S., Dauter, M. & Dauter, Z. (2003). *Acta Crystallogr D Biol Crystallogr* **59**, 492-498.
- Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. (2002). *Nucleic Acids Res* **30**, 276-280.

- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 271-281.
- Belitsky, B. R. (2004). *J Mol Biol* **340**, 655-664.
- Bellota-Antón, C., Munnoch, J., Robb, K., Adamczyk, K., Candelaresi, M., Parker, A. W., Dixon, R., Hutchings, M. I., Hunt, N. T. & Tucker, N. P. (2011). *Biochem Soc Trans* **39**, 1293-1298.
- Bender, R. A. (2012). *Microbiol Mol Biol Rev* **76**, 565-584.
- Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M. A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J. & Hopwood, D. A. (2002). *Nature* **417**, 141-147.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res* **28**, 235-242.
- Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. (2007). *Nucleic Acids Res* **35**, e45.
- Bradford, M. M. (1976). *Anal Biochem* **72**, 248-254.
- Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. (2013). *Nucleic Acids Res* **41**, W349-357.
- Buck, D. & Guest, J. R. (1989). *Biochem J* **260**, 737-747.
- Cachau, R. E. & Podjarny, A. D. (2005). *J Mol Recognit* **18**, 196-202.

- Cadoret, F., Soscia, C. & Voulhoux, R. (2014). *Methods Mol Biol* **1149**, 11-15.
- Campbell, J. W., Morgan-Kiss, R. M. & Cronan, J. E. (2003). *Mol Microbiol* **47**, 793-805.
- Candelaresi, M., Gumiero, A., Adamczyk, K., Robb, K., Bellota-Antón, C., Sangal, V., Munnoch, J., Greetham, G. M., Towrie, M., Hoskisson, P. A., Parker, A. W., Tucker, N. P., Walsh, M. A. & Hunt, N. T. (2013). *Org Biomol Chem* **11**, 7778-7788.
- Casali, N., White, A. M. & Riley, L. W. (2006). *J Bacteriol* **188**, 441-449.
- Clark, L. C. & Hoskisson, P. A. (2011). *PLoS One* **6**, e25049.
- Condemine, G., Berrier, C., Plumbridge, J. & Ghazi, A. (2005). *J Bacteriol* **187**, 1959-1965.
- Cook, R. J., Thomashow, L. S., Weller, D. M., Fujimoto, D., Mazzola, M., Bangera, G. & Kim, D. S. (1995). *Proc Natl Acad Sci U S A* **92**, 4197-4201.
- Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M. & Derewenda, Z. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 636-645.
- Cowtan, K. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1002-1011.
- Craig, M., Lambert, S., Jourdan, S., Tenconi, E., Colson, S., Maciejewska, M., Ongena, M., Martin, J. F., van Wezel, G. & Rigali, S. (2012). *Environ Microbiol Rep* **4**, 512-521.
- de Beer, T. A., Berka, K., Thornton, J. M. & Laskowski, R. A. (2014). *Nucleic Acids Res* **42**, D292-296.
- Derewenda, Z. (2004a). *Structure* **12**, 529-535.
- Derewenda, Z. S. (2004b). *Methods* **34**, 354-363.
- Deschamps, J. R. (2005). *AAPS J* **7**, E813-819.
- DiRusso, C. C., Metzger, A. K. & Heimert, T. L. (1993). *Mol Microbiol* **7**, 311-322.

- DiRusso, C. C., Tsvetnitsky, V., Højrup, P. & Knudsen, J. (1998). *J Biol Chem* **273**, 33652-33659.
- Doring, D. (1993). pp. pp. 245–273. New York: Plenum Press.
- Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal Biochem* **357**, 289-298.
- Favila, A., Gallo, M. & Glossman-Mitnik, D. (2007). *J Mol Model* **13**, 505-518.
- Fillenberg, S. B., Grau, F. C., Seidel, G. & Muller, Y. A. (2015). *Nucleic Acids Res* **43**, 1283-1296.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J. & Punta, M. (2014). *Nucleic Acids Res* **42**, D222-230.
- Franco, I. S., Mota, L. J., Soares, C. M. & de Sá-Nogueira, I. (2006). *J Bacteriol* **188**, 3024-3036.
- Franklin, M. C., Cheung, J., Rudolph, M. J., Burshteyn, F., Cassidy, M., Gary, E., Hillerich, B., Yao, Z. K., Carlier, P. R., Totrov, M. & Love, J. D. (2015). *Proteins*.
- Gallagher, D. T., Mayhew, M., Holden, M. J., Howard, A., Kim, K. J. & Vilker, V. L. (2001). *Proteins* **44**, 304-311.
- Gao, Y. G., Yao, M., Itou, H., Zhou, Y. & Tanaka, I. (2007). *Protein Sci* **16**, 1878-1886.
- Gebhard, S., Busby, J. N., Fritz, G., Moreland, N. J., Cook, G. M., Lott, J. S., Baker, E. N. & Money, V. A. (2014). *J Bacteriol*.
- Geerlof, A., Brown, J., Coutard, B., Egloff, M. P., Enguita, F. J., Fogg, M. J., Gilbert, R. J., Groves, M. R., Haouz, A., Nettleship, J. E., Nordlund, P., Owens, R. J., Ruff,

- M., Sainsbury, S., Svergun, D. I. & Wilmanns, M. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 1125-1136.
- Genomics, J. C. f. S. (To be published). Crystal structure of Phenylacetic acid degradation-related protein (YP_298971.1) from *Ralstonia eutropha* JMP134 at 2.20 Å resolution.
- Genomics., J. C. f. S. (To be published). Crystal structure of DNA binding protein (YP_298823.1) from *Ralstonia eutropha* JMP134 at 1.92 Å resolution.
- Giegerich, R., Meyer, F. & Schleiermacher, C. (1996). *Proc Int Conf Intell Syst Mol Biol* **4**, 68-77.
- Goldschmidt, L., Cooper, D., Derewenda, Z. & Eisenberg, D. (2007). *Protein Science* **16**, 1569-1576.
- Goldschmidt, L., Eisenberg, D. & Derewenda, Z. S. (2014). *Methods Mol Biol* **1140**, 201-209.
- Gorelik, M., Lunin, V. V., Skarina, T. & Savchenko, A. (2006). *Protein Sci* **15**, 1506-1511.
- Griffin, M. D. & Gerrard, J. A. (2012). *Adv Exp Med Biol* **747**, 74-90.
- Gui, L., Sunnarborg, A. & LaPorte, D. C. (1996). *J Bacteriol* **178**, 4704-4709.
- Hamm, P., Limm, M. & Hoschstrasser, R. M. (1998). *J. Phys. Chem. B* **102**, 6123-6138.
- Hanahan, D. (1983). *J Mol Biol* **166**, 557-580.
- Haydon, D. J. & Guest, J. R. (1991). *FEMS Microbiol Lett* **63**, 291-295.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J* **9**, 1665-1672.
- Hillerich, B. & Westpheling, J. (2006). *J Bacteriol* **188**, 7477-7487.

- Hoffman, I. D. (2012). *Methods Mol Biol* **841**, 67-91.
- Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. (1999). *Proc Natl Acad Sci U S A* **96**, 9833-9838.
- Hoskisson, P. A. & Rigali, S. (2009). *Adv Appl Microbiol* **69**, 1-22.
- Hoskisson, P. A., Rigali, S., Fowler, K., Findlay, K. C. & Buttner, M. J. (2006). *J Bacteriol* **188**, 5014-5023.
- Hu, L. & Phillips, A. T. (1988). *J Bacteriol* **170**, 4272-4279.
- Hunt, N. T. (2009). *Chem Soc Rev* **38**, 1837-1848.
- Hutchings, M. I. & Drabble, W. T. (2000). *FEMS Microbiol Lett* **187**, 115-122.
- Ilari, A. & Boffi, A. (2008). *Methods Enzymol* **436**, 187-202.
- Jain, D. & Nair, D. T. (2013). *Nucleic Acids Res* **41**, 639-647.
- Jaques, S. & McCarter, L. L. (2006). *J Bacteriol* **188**, 2625-2635.
- Jochmann, N., Götker, S. & Tauch, A. (2011). *Microbiology* **157**, 77-88.
- Jones, D. T. (1999). *J Mol Biol* **292**, 195-202.
- Jordan, S. R., Whitcombe, T. V., Berg, J. M. & Pabo, C. O. (1985). *Science* **230**, 1383-1385.
- Kendall, K. J. & Cohen, S. N. (1988). *J Bacteriol* **170**, 4634-4651.
- Kersey, P. J., Allen, J. E., Armean, I., Boddu, S., Bolt, B. J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L. J., Grabmueller, C., Humphrey, J., Kerhornou, A., Khobova, J., Aranganathan, N. K., Langridge, N., Lowy, E., McDowall, M. D., Maheswari, U., Nuhn, M., Ong, C. K., Overduin, B., Paulini, M., Pedro, H., Perry, E., Spudich, G., Tapanari, E., Walts, B., Williams, G., Tello-Ruiz, M., Stein, J., Wei, S., Ware, D., Bolser, D. M., Howe, K. L., Kulesha, E., Lawson, D., Maslen, G. & Staines, D. M. (2016). *Nucleic Acids Res* **44**, D574-580.

- Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. & Smith, M. C. (2011). *Mol Microbiol* **80**, 1450-1463.
- Klaffl, S., Brocker, M., Kalinowski, J., Eikmanns, B. J. & Bott, M. (2013). *J Bacteriol* **195**, 4283-4296.
- Klopman, G., Fercu, D. & Jacob, J. (1996). *Chemical Physics* **204**, 181-193.
- Koonin, E. V., Tatusov, R. L. & Rudd, K. E. (1995). *Proc Natl Acad Sci U S A* **92**, 11921-11925.
- Kruh, N. A., Rawat, R., Ruzsicska, B. P. & Tonge, P. J. (2007). *Protein Sci* **16**, 1617-1627.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. (2007). *Bioinformatics* **23**, 2947-2948.
- Lee, D. G., Urbach, J. M., Wu, G., Liberati, N. T., Feinbaum, R. L., Miyata, S., Diggins, L. T., He, J., Saucier, M., Déziel, E., Friedman, L., Li, L., Grills, G., Montgomery, K., Kucherlapati, R., Rahme, L. G. & Ausubel, F. M. (2006). *Genome Biol* **7**, R90.
- Lee, M. H., Scherer, M., Rigali, S. & Golden, J. W. (2003). *J Bacteriol* **185**, 4315-4325.
- Liberati, N. T., Urbach, J. M., Miyata, S., Lee, D. G., Drenkard, E., Wu, G., Villanueva, J., Wei, T. & Ausubel, F. M. (2006). *Proc Natl Acad Sci U S A* **103**, 2833-2838.
- Lipuma, J. J. (2010). *Clin Microbiol Rev* **23**, 299-323.
- Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. (2013). *Microbiol Mol Biol Rev* **77**, 112-143.

- Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. (2000). *Genome Biol* **1**, REVIEWS001.
- Magarvey, N., He, J., Aidoo, K. A. & Vining, L. C. (2001). *Microbiology* **147**, 2103-2112.
- Magasanik, B. (1976). *Prog Nucleic Acid Res Mol Biol* **17**, 99-115.
- Maloy, S. R. & Nunn, W. D. (1981). *J Bacteriol* **148**, 83-90.
- Mao, F., Dam, P., Chou, J., Olman, V. & Xu, Y. (2009). *Nucleic Acids Res* **37**, D459-463.
- Matthews, B. W. (1968). *J Mol Biol* **33**, 491-497.
- McCoy, A. J. (2007). *Acta Crystallogr D Biol Crystallogr* **63**, 32-41.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-674.
- McPherson, A. (2004). *Methods* **34**, 254-265.
- Molle, V., Gulten, G., Vilchèze, C., Veyron-Churlet, R., Zanella-Cléon, I., Sacchettini, J. C., Jacobs, W. R. & Kremer, L. (2010). *Mol Microbiol* **78**, 1591-1605.
- Mota, L. J., Sarmiento, L. M. & de Sá-Nogueira, I. (2001). *J Bacteriol* **183**, 4190-4201.
- Murray, V., Chen, J. K. & Galea, A. M. (2014). *Mutat Res* **769**, 93-99.
- Nettleship, J. E., Brown, J., Groves, M. R. & Geerlof, A. (2008). *Methods Mol Biol* **426**, 299-318.
- Nicholls, R. A., Long, F. & Murshudov, G. N. (2012). *Acta Crystallogr D Biol Crystallogr* **68**, 404-417.

- Nocek, B., Sather, A., Gu, M. & Joachimiak, A. (To be published). Structure of transcriptional regulator (GntR family member) from *Pseudomonas syringae* pv. tomato str. DC3000.
- Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982). *Nature* **298**, 718-723.
- Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. (1983). *J Mol Evol* **19**, 109-114.
- Ohlendorf, D. H. & Matthews, B. W. (1983). *Annu Rev Biophys Bioeng* **12**, 259-284.
- Ostash, B., Rebets, Y., Myronovskyy, M., Tsyplik, O., Ostash, I., Kulachkovskyy, O., Datsyuk, Y., Nakamura, T., Walker, S. & Fedorenko, V. (2011). *Microbiology* **157**, 1240-1249.
- Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P. & Salemme, F. R. (2001). *J Biomol Screen* **6**, 429-440.
- Raman, N., Black, P. N. & DiRusso, C. C. (1997). *J Biol Chem* **272**, 30645-30650.
- Resch, M., Schiltz, E., Titgemeyer, F. & Muller, Y. A. (2010). *Nucleic Acids Res* **38**, 2485-2497.
- Reuther, J., Gekeler, C., Tiffert, Y., Wohlleben, W. & Muth, G. (2006). *Mol Microbiol* **61**, 436-446.
- Rice, P., Longden, I. & Bleasby, A. (2000). *Trends Genet* **16**, 276-277.
- Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. (2002). *J Biol Chem* **277**, 12507-12515.
- Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B. & Titgemeyer, F. (2004). *Nucleic Acids Res* **32**, 3418-3426.

- Rossbach, S., Kulpa, D. A., Rossbach, U. & de Bruijn, F. J. (1994). *Mol Gen Genet* **245**, 11-24.
- Rozwarski, D. A., Grant, G. A., Barton, D. H., Jacobs, W. R. & Sacchettini, J. C. (1998). *Science* **279**, 98-102.
- Russo Krauss, I., Merlino, A., Vergara, A. & Sica, F. (2013). *Int J Mol Sci* **14**, 11643-11691.
- Salzberg, L. I., Luo, Y., Hachmann, A. B., Mascher, T. & Helmann, J. D. (2011). *J Bacteriol* **193**, 5793-5801.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. (1982). *Nature* **298**, 447-451.
- Scapin, G. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 2266-2275.
- Shaw, D. J., Adameczyk, K., Frederix, P. W., Simpson, N., Robb, K., Greetham, G. M., Towrie, M., Parker, A. W., Hoskisson, P. A. & Hunt, N. T. (2015). *J Chem Phys* **142**, 212401.
- Sieira, R., Arocena, G. M., Bukata, L., Comerci, D. J. & Ugalde, R. A. (2010). *J Bacteriol* **192**, 217-224.
- Silby, M. W., Cerdeño-Tárraga, A. M., Vernikos, G. S., Giddens, S. R., Jackson, R. W., Preston, G. M., Zhang, X. X., Moon, C. D., Gehrig, S. M., Godfrey, S. A., Knight, C. G., Malone, J. G., Robinson, Z., Spiers, A. J., Harris, S., Challis, G. L., Yaxley, A. M., Harris, D., Seeger, K., Murphy, L., Rutter, S., Squares, R., Quail, M. A., Saunders, E., Mavromatis, K., Brettin, T. S., Bentley, S. D., Hothersall, J.,

Stephens, E., Thomas, C. M., Parkhill, J., Levy, S. B., Rainey, P. B. & Thomson, N. R. (2009). *Genome Biol* **10**, R51.

Simossis, V. A. & Heringa, J. (2005). *Nucleic Acids Res* **33**, W289-294.

Simpson, N., Shaw, D. J., Frederix, P. W., Gillies, A. H., Adamczyk, K., Greetham, G. M., Towrie, M., Parker, A. W., Hoskisson, P. A. & Hunt, N. T. (2013). *J Phys Chem B* **117**, 16468-16478.

Slabinski, L., Jaroszewski, L., Rychlewski, L., Wilson, I., Lesley, S. & Godzik, A. (2007). *Bioinformatics* **23**, 3403-3405.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S. & Olson, M. V. (2000). *Nature* **406**, 959-964.

Studier, F. W. (2005). *Protein Expr Purif* **41**, 207-234.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol* **185**, 60-89.

Suhre, K. & Claverie, J. M. (2004). *Nucleic Acids Res* **32**, D273-276.

Tamura, K., Stecher, G., Peterson, D., Filipinski, A. & Kumar, S. (2013). *Mol Biol Evol* **30**, 2725-2729.

Tan, K., Skarina, T., Onopriyenko, A., Savchenko, A., Edwards, A. & Joachimiak, A. (To be published). The crystal structure of a putative transcriptional regulator GntR from *Rhodococcus* sp. RHA1.

- Tanaka, T., Komatsu, C., Kobayashi, K., Sugai, M., Kataoka, M. & Kohno, T. (To be published). Solution structure of Streptmycal repressor TraR.
- Taylor, G. L. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 325-338.
- team, E. e. (2013). *Euro Surveill* **18**.
- Teresa Pellicer, M., Felisa Nuñez, M., Aguilar, J., Badia, J. & Baldoma, L. (2003). *J Bacteriol* **185**, 5815-5821.
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J. & Adams, P. D. (2008). *Acta Crystallogr D Biol Crystallogr* **64**, 61-69.
- Truong-Bolduc, Q. C. & Hooper, D. C. (2007). *J Bacteriol* **189**, 2996-3005.
- Tucker, N. P., Ghosh, T., Bush, M., Zhang, X. & Dixon, R. (2010). *Nucleic Acids Res* **38**, 1182-1194.
- van Aalten, D. M., DiRusso, C. C. & Knudsen, J. (2001). *EMBO J* **20**, 2041-2050.
- van Aalten, D. M., DiRusso, C. C., Knudsen, J. & Wierenga, R. K. (2000). *EMBO J* **19**, 5167-5177.
- Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. & de Gier, J. W. (2008). *Proc Natl Acad Sci U S A* **105**, 14371-14376.
- Walter, T. S., Diprose, J. M., Mayo, C. J., Siebold, C., Pickford, M. G., Carter, L., Sutton, G. C., Berrow, N. S., Brown, J., Berry, I. M., Stewart-Jones, G. B., Grimes, J. M., Stammers, D. K., Esnouf, R. M., Jones, E. Y., Owens, R. J., Stuart, D. I. & Harlos, K. (2005). *Acta Crystallogr D Biol Crystallogr* **61**, 651-657.
- Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I. & Grimes, J. M. (2006). *Structure* **14**, 1617-1622.

- Watanabe, T., Inoue, R., Kimura, N. & Furukawa, K. (2000). *J Biol Chem* **275**, 31016-31023.
- Wiethaus, J., Schubert, B., Pfänder, Y., Narberhaus, F. & Masepohl, B. (2008). *J Bacteriol* **190**, 487-493.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 235-242.
- Winsor, G. L., Lam, D. K., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., Hancock, R. E. & Brinkman, F. S. (2011). *Nucleic Acids Res* **39**, D596-600.
- Winter, G., Lobley, C. M. & Prince, S. M. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1260-1273.
- Yang, C., Rodionov, D. A., Li, X., Laikova, O. N., Gelfand, M. S., Zagnitko, O. P., Romine, M. F., Obraztsova, A. Y., Nealson, K. H. & Osterman, A. L. (2006). *J Biol Chem* **281**, 29872-29885.
- Yang, Z., Thomson, R., McNeil, P. & Esnouf, R. (2005). *Bioinformatics* **21**, 3369-3376.
- Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). *Gene* **33**, 103-119.
- Yoshida, K. I., Fujita, Y. & Ehrlich, S. D. (2000). *J Bacteriol* **182**, 5454-5461.
- Zanni, M. T. & Hochstrasser, R. M. (2001). *Curr Opin Struct Biol* **11**, 516-522.
- Zhang, R., Volkart, L., Gu, M. & Joachimiak, A. (To be published). The crystal structure of a transcriptional regulator from *Oenococcus oeni*.

Zhang, R., Xu, X., Zheng, H., Savchenko, A., Edwards, A. & Joachimiak, A. (To be published). The crystal structure of the putative regulator from *Escherichia coli* CFT073.

Zhang, R., Zhou, M., Bargassa, M., Otwinowski, Z. & Joachimiak, A. (To be published). The crystal structure of the transcriptional regulator (GntR family) from *Enterococcus faecalis* V583.

Zhang, X. X. & Rainey, P. B. (2007). *Genetics* **176**, 2165-2176.

Zhang, Y., Heym, B., Allen, B., Young, D. & Cole, S. (1992). *Nature* **358**, 591-593.

Zheng, J., Kwak, K. & Fayer, M. D. (2007). *Acc Chem Res* **40**, 75-83.

Zheng, M., Cooper, D. R., Grosseohme, N. E., Yu, M., Hung, L. W., Cieslik, M., Derewenda, U., Lesley, S. A., Wilson, I. A., Giedroc, D. P. & Derewenda, Z. S. (2009). *Acta Crystallogr D Biol Crystallogr* **65**, 356-365.