Structural Characterisation of Bacterial Proteins Implicated in Resistance and Adaptation

Heather Ashwood (neé Clift)

Bachelor of Science in Biochemistry

Master of Research (MRES) in Molecular Sciences

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy

Department of Molecular Sciences

Faculty of Science and Engineering

Macquarie University, Australia

August 2018

Abstract

The emergence and rapid global spread of antibiotic resistance phenotypes amongst different bacterial lineages is threatening a return to a pre-antibiotic era. Two prime mechanisms used by bacteria for defence against antibiotics are lateral gene transfer and efflux systems. This thesis considers protein properties in solution for both mechanisms and assesses their capacities to respond to small molecule binding partners. Part I focuses on novel fold proteins recovered from integron/gene cassette systems of various bacterial isolates. Part II investigates efflux regulator proteins in *Acinetobacter baumannii*.

Part I:

Lateral gene transfer allows a dynamic gene pool to give rise to the wide phenotypic diversity and rapid evolution rates characteristic of bacteria. The integron/gene cassette system is an active player in lateral gene transfer, particularly credited with the rapid spread of multi-drug resistance phenotypes. The genes within these cassettes are remarkably diverse: 80% carry open reading frames (ORFs) with either no known homology or homology to ORFs of unknown function. Crystal structures of completely novel folds have been defined for several ORFs, derived from mobile gene cassettes in microbial isolates from Halifax Harbour and *Vibrio* spp. In this thesis the first biophysical characterisation for three of these proteins has been provided.

They are named for the bacterial hosts or locations from which they were recovered and are small oligomeric proteins of α/β and α+β fold class. Hfx1 is an extremely robust trimer which is stabilised bymetals and protamine sulfate. A series of mutant forms were generated to probe its tertiary stability and investigate hypothesised binding pockets. Each mutant, whether it targeted backbone hydrogen bonds, hydrophobic amino acids, cysteines or histidines, still assembled as a trimer, although melting temperatures varied. In addition, histidine mutant forms of Hfx1 did not reduce metal stabilisation, removing them as possible chelating residues in the hypothesised metal-binding pocket. These methods were also used for the hydrophobic dimeric protein Vch14. However, as with Hfx1, mutant forms of Vch14 still formed dimers with varying stabilities. Hfx5, a

domain-swapped dimer under crystal conditions, was found to form mixed oligomeric species of monomer and dimer in solution, so the arrangements of these quaternary structures were probed. Conditions such as low protein concentration and more destabilising buffers (using $MgCl_2$ and $CaCl_2$ as salts) resulted in a shift towards the monomeric form of Hfx5. Overall, these small bacterial proteins are highly robust and retain a propensity to oligomerise even when structurally stressed. Many of these novel-fold proteins hold the potential to form the building blocks for larger heteromeric protein structures, either in nature or *in vitro*.

Part II:

A. baumannii is a tenacious Gram-negative opportunistic human pathogen, commonly associated with hospital-acquired infections. Membrane drug efflux systems are prominent resistance factors in *A. baumannii,* likely acting as regulators for virulence, stress and biofilm formation. There are six families of efflux-pump proteins known in Gram-negative bacteria, with a large range of substrates. The structure and function for four efflux-pump regulators (AdeN, AmvR, AdeL, AceR), known to regulate the expression of major facilitator, resistance nodulation division, and proteobacterial antimicrobial compound efflux transporter family pumps have been investigated. These regulator proteins belong to the TetR and LysR-type family of transcriptional regulators.

Both AmvR and AdeL display higher melting temperatures in the presence of small molecules of varying chemistries, including polyamines. This responds to previous work demonstrating the ability of AceR to bind polyamines. In particular, both AmvR and AdeL are highly stabilised by cystamine, suggesting that their efflux systems may respond to oxidative stress. Thus, different regulators from both TetR and LysR families appear to have affinity for similar molecules. More work is required to understand the full regulatory role of these proteins in *A. baumannii*, but this study has demonstrated the viability of studying recombinant forms to address gene regulation mechanisms.

Statement of Originality

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

(Signed)_____Heather Clift_____ Date: _____ 15/07/2018_____

Acknowledgements

First, I must thank A/Prof Bridget Mabbutt, my supervisor, for supporting me throughout my research. She has been full of great ideas and has pushed me when I needed it, resulting in this completed thesis.

I also need to thank Prof. Ian Paulsen, my associate supervisor, who came up with the idea for the project that contributed to the second part of my thesis (and funded much more than half of my research). This great project, and everyone involved in it (Snow, Karl, and Varsha), kept me motivated through my candidature, and I'm glad that I could contribute. I also need to thank Varsha and Snow for helping me attain reagents and teaching me how to use the Biacore.

I would like to thank everyone that helped me out in some small way during my PhD, whether it be giving me reagents or teaching me techniques. I need to thank Bhumika, Edward, and Sophie. I also need to thank Tony for keeping the Biacore and Äktas in working condition throughout the three years. Special thanks go to Prof. Alison Rodger for rescuing the old CD from MUCAB and bringing it back to life, as well as letting me use some of her new instruments (the new CD and spectrofluorimeter).

Finally, I would like to thank Chris for always being there for me, listening to me complain, letting me bounce ideas off him, and generally letting me tell him everything I did each day. I probably would have had at least twice as many emotional breakdowns without him around. I'm glad that I will have him by my side as we go out into the wide world of employment together. I would also like to thank my family for being supportive throughout my scientific career.

Abbreviations

Abbreviations are used in this work according to the format observed by the Journal of Molecular Microbiology (Elsevier). Special terms are used as follows:

ABC	ATP-binding cassette
ADP	adenine dinucleotide phosphate
APS	ammonium persulfate
CD	circular dichroism
DBD	DNA-binding domain
DSF	differential scanning fluorimetry
EBD	effector-binding domain
EDTA	ethylenediaminetetraacetic acid
ESI	electrospray ionisation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMW	high molecular weight
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl β-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
K_{AV}	partition coefficient
LGT	lateral gene transfer
LH	linker helix
LMW	low molecular weight
LTTR	LysR-type transcriptional regulator
MATE	multidrug and toxic-compound extrusion
MDR	multidrug resistance
MES	2-(N-morpholino)ethanesulfonic acid
MFS	major facilitator superfamily
MIC	minimum inhibitory concentration
Mw	molecular weight
NADP	nicotinamide adenine dinucleotide phosphate

NEB	New England Biolabs
OD_{600}	optical density reading at 600 nm
ORF	open reading frame
PACE	proteobacterial antimicrobial compound efflux
PCR	polymerase chain reaction
PDB	Protein Data Bank
RND	resistance nodulation division
RT-PCR	real-time PCR
SEC	size exclusion chromatography
SMR	small multidrug resistance
TAE	tris-acetate-EDTA
TCEP	tris(2-carboxyethyl)phosphine
TE buffer	tris-EDTA buffer
TEMED	tetramethylethylenediamine
TetR	tetracycline repressor
TEV	Tobacco Etch Virus
TFR	tetracycline repressor family transcriptional regulators
TIM	triosephosphate isomerase
T_M	melting temperature
Tris	trisaminomethane
V_0	void volume
V_E	elution volume
WHO	World Health Organisation

Table of Contents

Abstract 1
Statement of Originality 3
Acknowledgements 4
Abbreviations 5
Table of Contents 7
CHAPTER 1 11
Overview 11
1.1. Genomic plasticity of bacteria 11
1.2. Critical efflux components in bacteria 14
1.3. Biophysical methods for monitoring protein fold states 16
1.3.1. Using differential scanning fluorimetry to measure protein stability 16
1.3.2. Size exclusion chromatography as a purification and analysis technique 18
1.4. Experimental aims 19
PART I 21
CHAPTER 2 22
Molecular Organisation of Novel-fold proteins from Bacterial Metagenomes 22
2.1. Organisation and evolution of protein oligomerisation 23
2.1.1. Homo-oligomerisation 23
2.1.2. Hetro-oligomerisation 25
2.2. Trimers of Hfx1: Alpha/beta fold structure 28
2.2.1. Oligomerisation contacts 28
2.2.2. Sequence analysis of Hfx1 31
2.2.3. Experimental plan for Hfx1 33
2.3. Domain-swapped Hfx5: Alpha + beta fold structure 34
2.3.1. Oligomerisation contacts 34
2.3.2. Sequence analysis of Hfx5 37
2.3.3. Experimental plan for Hfx5 37
2.4. Dimer of Vch3: an α+β fold 38
2.4.1. Oligomerisation contacts 38
2.4.2. Sequence analysis of Vch3 39
2.4.3. Experimental plan for Vch3 39
2.5. Dimer of Vch14: an α+β fold 41
2.5.1. Oligomerisation contacts 41

2.5.2. Sequence analysis of Vch14 41
2.5.3. Experimental plan for Vch14 45
CHAPTER 3 46
Solution Characterisation of Novel-fold Proteins 46
3.1. Production of novel-fold proteins 47
3.2. Evaluation of quaternary state and stability in solution 49
3.2.1. Hfx1 49
3.2.1.1. The structural role of cysteine residues 51
3.2.1.2. Serine and tyrosine mutant forms 54
3.2.1.3. Isoleucine mutant forms 56
3.2.1.4. Histidine mutant forms 58
3.2.2. Hfx5 58
3.2.3. Vch3 64
3.2.4. Vch14 64
3.2.4.1. Serine and isoleucine mutant forms 68
CHAPTER 4 75
Binding Studies of Novel Fold Proteins 75
4.1. Screening for potential effectors of novel-fold proteins 76
4.1.1. Screening of potential effectors for Hfx1 76
4.1.1.1. Protamine sulfate binds by electrostatic interaction 79
4.1.1.2. Metal binding capacities of Hfx1 82
4.1.2. Screening of potential effectors for Hfx5 84
4.2. Summary 86
4.3. Conclusions and future work 87
4.3.1. Bacterial adaptations through protein mimicry 87
4.3.2. Future work 88
4.3.2.1. Further analysis of Hfx1 stabilising molecules 88
4.3.2.2. Identification of ligands for Vch14 89
4.3.2.3. Using novel fold proteins as building blocks for nanomaterials 90
PART II 92
CHAPTER 5 93
Efflux Pump Regulator Proteins of *Acinetobacter baumannii* 93
5.1. Efflux systems as a mode of continued virulence and adaptation in bacteria 93
5.1.1. Factors that increase the expression of efflux systems 94
5.1.2. Families of efflux pumps 95
5.2. Molecular mechanisms of transport regulation: efflux pump transcriptional regulators. 96

5.2.1. TetR transcriptional regulators 98
5.2.2. LysR-type transcriptional regulators 100
5.3. Efflux systems in the pathogenic organism *Acinetobacter baumannii* 102
5.3.1. The RND transporter system AdeIJK 104
5.3.2. The RND transporter AdeFGH 106
5.3.3. The MFS transporter AmvA 106
5.3.4. The PACE transporter AceI 107
5.4. Experimental aims 108
CHAPTER 6 110
Solution Characterisation of *A. baumannii* Efflux-pump Regulator Proteins 110
6.1. Design of protein constructs 110
6.1.1. AdeL and AceR 110
6.1.2. AdeN and AmvR 113
6.1.3. Selection of system for heterologous expression 115
6.2. Production of proteins of the LTTR and TFR family 116
6.2.1. Amplification and cloning of genes 116
6.2.2. Expression and purification of full-length and truncated proteins 116
6.3. Evaluation of quaternary state and stability in solution 120
6.3.1. AdeN constructs with and without DBD 121
6.3.2. AmvR construct with DBD 123
6.3.3. AceR constructs with and without the DBD 125
6.3.4. AdeL regulator domain constructs 127
CHAPTER 7 131
Binding Studies of *A. baumannii* Efflux Pump Regulator Proteins 131
7.1. Screening for potential effectors of efflux pump regulators 133
7.1.1. Screening of potential effectors for AmvR 134
7.1.2. Screening of potential effectors for AdeL 139
7.1.3. Screening of potential effectors for AceR 143
7.2. Summary 145
7.3. Conclusions 147
7.3.1. The role of polyamines in bacterial species 147
7.3.2. Cystamine metabolism in bacteria 150
7.4. Future Work 151
CHAPTER 8 154
Materials and Methods 154
8.1. Materials 154

8.1.1. Reagents 154
8.1.2. Growth media and buffers 154
8.1.3. Bacterial strains and plasmids 159
8.1.4. Preservation of bacterial strains 159
8.2. Methods 159
8.2.1. Molecular biology procedures 159
8.2.1.1. Plasmid isolation 159
8.2.1.2. Plasmid digestion 160
8.2.1.3. Extraction of genomic DNA 160
8.2.1.4. Amplification of genes from genomic DNA 160
8.2.1.5. Preparation and amplification of DNA gBlocks 162
8.2.1.6. Ligation-independent cloning 162
8.2.1.7. PCR colony screening 164
8.2.1.8. Preparation of competent expression host cells 164
8.2.1.9. Site-directed mutagenesis 164
8.2.2. Recombinant protein production 166
8.2.2.1. Protein expression by autoinduction 166
8.2.2.2. Cell lysis 167
8.2.2.3. Immobilised metal affinity chromatography (IMAC) 167
8.2.2.4. Buffer exchange through dialysis 167
8.2.3. Molecular characterisation and functional assays 168
8.2.3.1. Electrophoresis 168
8.2.3.1.1. Agarose gel electrophoresis 168
8.2.3.1.2. SDS-PAGE 168
8.2.3.2. Size exclusion chromatography 169
8.2.3.3. Differential Scanning Fluorimetry 171
8.2.3.4. Circular Dichroism 172
8.2.3.5. Tryptophan fluorescence 174
8.2.3.6. Sequence analysis tools 174
References 175
Appendix 197

CHAPTER 1

Overview

Bacteria display remarkable levels of diversity in their metabolic properties, phenotypic traits, cellular architectures and range of lifestyles; diversity that at first glance appears to be beyond the evolutionary capacity of small, single-celled organisms with comparatively limited genomes (Ochman *et al.*, 2000; Fraser-Liggett, 2005). Bacteria have been found to be capable of thriving in vastly different environments, including many previously thought to be inhospitable to life (Ward *et al.*, 1998; Rothschild & Mancinelli, 2001; Hooper *et al.*, 2009).

The emergence and rapid global spread of antibiotic resistance phenotypes amongst different bacterial lineages over recent decades is threatening a return to a pre-antibiotic era (Xia *et al.*, 2016; Davies & Davies, 2010; Kumarasamy *et al.*, 2010). The Infectious Diseases Society of America has noted six multi-drug resistant bacteria that are today responsible for the majority of nosocomial infections worldwide. These species termed “ESKAPE” pathogens are *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species (Rice, 2008).

1.1. Genomic plasticity of bacteria

One mechanism exploited by bacteria for defence against antibiotics is lateral gene transfer (LGT). With ever-increasing data from whole genome sequences, it has become apparent that LGT plays a large role in generally shaping bacterial evolution (Mirkin *et al.*, 2003; Touchon *et al.*, 2009) due to replication of genetic material independent of organismal reproduction (Maczulak, 2011; Juhas, 2015). The sharing of genes across bacterial communities has led to concepts of the ‘pan-genome’ (Tettelin *et al.*, 2005) and ‘pan-plasmidome’ (Fondi & Fani, 2010) to describe shared gene pools available for microbial evolution.

Lateral movement of genes can be achieved through mechanisms of transformation, conjugation, or transduction (Jain *et al.*, 2002), and is responsible for rendering a large portion of previously

benign bacteria pathogenic. For example, *Enterococci*, once harmless commensals of the gastrointestinal tract, have emerged over the past 30 years as pathogens in hospital settings (Palmer *et al.*, 2010). Today, most *E. faecium* isolates (Rosvoll *et al.*, 2010) contain extra plasmids in addition to their core genome, which encode antibiotic resistance and virulence factors.

Bacteria also directly exploit a number of mobile genetic elements to facilitate integration of novel DNA material. Such elements include plasmids, transposons, gene cassettes, and other integrative and conjugative elements (Frost *et al.*, 2005; Skippington & Ragan, 2011; Stokes & Gillings, 2011). Integrons are defined as one of the most efficient genetic elements for the capture and expression of foreign genes (Stokes & Hall, 1989; Escudero *et al.*, 2015).

Figure 1.1 outlines the general features of an integron array, which may consist of multiple gene cassettes (500–1000 bp), each containing a single promoter-less gene (open reading frame (ORF)) and attachment sequence (known as *attC* or 59-be (Hall & Stokes, 1993)). Gene cassettes also exist independently as circular DNA (Recchia & Hall, 1995; Bennett, 1999). Multiple gene cassettes can be captured by the integron in independent recombination events, leading to large arrays, in the case of *Vibrio cholerae*, comprising over 200 gene cassettes (Rowe-Magnus *et al.*, 1999; Chen *et al.*, 2003).

With the aid of cassette PCR (Stokes *et al.*, 2001; Holmes *et al.*, 2003; Partridge *et al.*, 2009) thousands of gene cassettes have been directly recovered from metagenomic DNA, mixed culture populations and defined bacterial cell lines. ORFs within integrons can encode for a variety of proteins, differing in size and function and these include known antibiotic resistance factors. Integrons are classified (Class 1, 2 and 3) according to how they confer resistance to a wide array of antibiotics (Mazel, 2006). In marine microorganisms, integrons have been found to encode proteins such as oxidoreductases (e.g. iron/sulfur cluster-binding proteins) and alkyl transferases (Elsaied *et al.*, 2014; Elsaied *et al.*, 2007; Elsaied *et al.*, 2011; Rodriguez-Minguela *et al.*, 2009). The majority of genes within integron cassettes are remarkably diverse: 80% have been seen to carry ORFs for which no homology can be detected or that share a relationship with homologs of no annotated function (Labbate *et al.*, 2009).

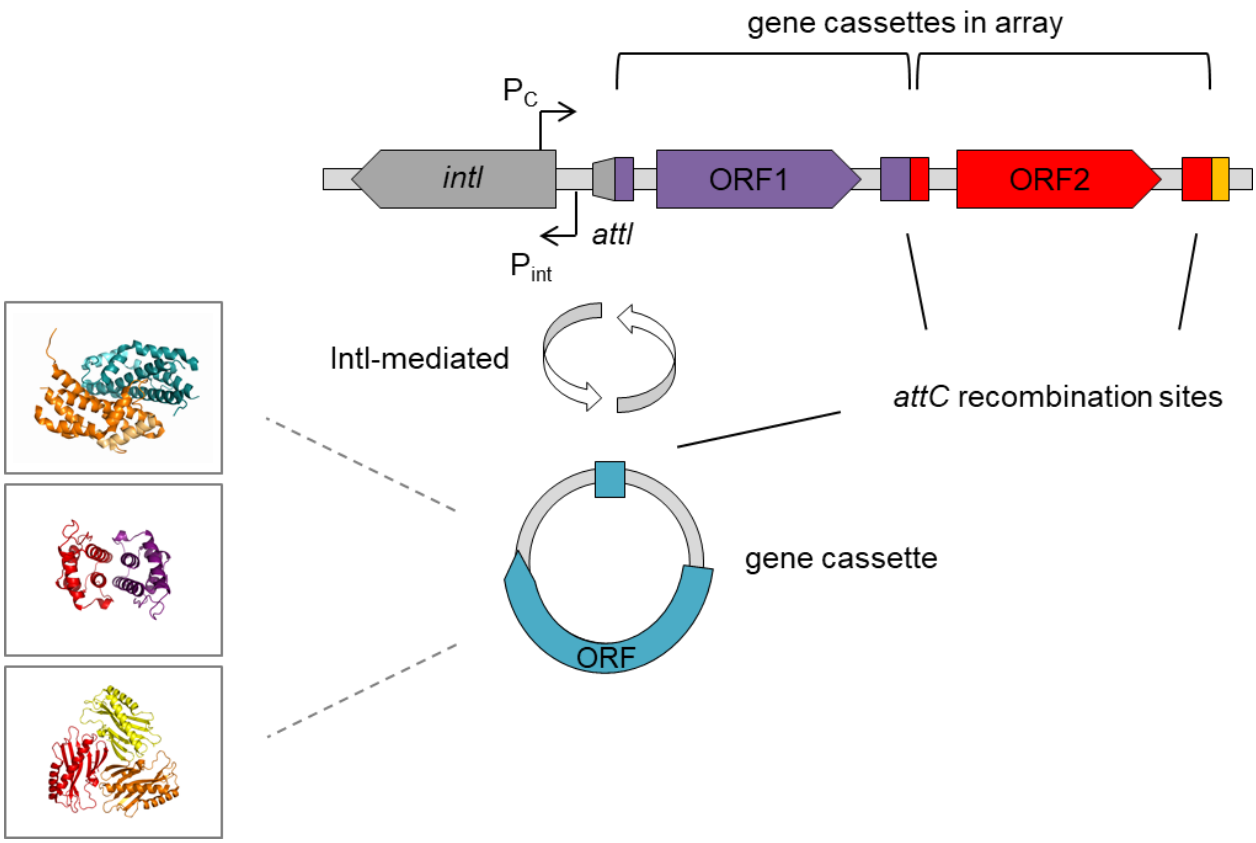


Figure 1.1 Integron gene array and insertion mechanism. Structure depicts an integron carrying two gene cassettes (ORF1 and ORF2). Cassette integration engages site-specific recombination between *attC* and *attI* sites. Adapted from (Stokes *et al.*, 2001). Gene cassettes encode small stable protein domains (Sureshan *et al.*, 2013).

The technique of cassette PCR has been incorporated into research at Macquarie University focussed on the potential impact of the integron in shaping bacterial evolutionary change through responses to fluid environmental pressures. This research used a structural genomics approach incorporating x-ray crystallography on a selection of integron cassettes to investigate evolutionary relationships hidden at the sequence level with the intention to make functional annotation from underlying structural homology identified by tertiary structure determination from known fold families. Gene cassettes were accessed from marine sediments, as well as a range of *Vibrio spp*. (Sureshan *et al.*, 2013; Robinson *et al.*, 2005; Robinson *et al.*, 2007), from which relatively small and novel genes (encoding proteins of ~70-140 amino acids) were recovered. During further

analysis, no clear annotation was evident from sequence homologues identified for these. From these gene sequences, selected due to their particular robustness for recombinant production in *E.coli*, six diffraction-quality crystals were produced yielding quality structures of soluble proteins found to possess novel folds. These new structures are diverse, encompassing all-α, α+β and α/β fold classes, with some features within these folds suggestive of ligand interaction sites, or contact points for heterogeneous regulator/activator proteins or domains (Sureshan *et al.*, 2013). Oligomerisation is common to these new integron-derived proteins. However, little else is known about the biophysical properties of this group of unknown proteins. As the first members of new structural variants in their distinct fold families, these provide new examples, each requiring delineation of biophysical properties and stability factors.

In **Part I** of this thesis the questions of protein robustness and propensity for oligomerisation have been explored for these integron gene cassette proteins. Should their role be ultimately to form single domains within more sophisticated hetero-oligomeric biomolecules, incorporating regulatory components, then they need to be resilient as a single entity. The crystal structures of these completely novel fold proteins have been defined and provide molecular templates for mutagenesis to determine integral residues and interactions critical to each protein fold.

1.2. Critical efflux components in bacteria

Phenotypic variability in multidrug resistant (MDR) bacteria is not entirely governed by genetic variation and mutations; survival rates of isogenic bacterial populations treated with low antibiotic concentrations are far higher than can be explained by these mechanisms (Adam *et al.*, 2008; George & Levy, 1983; Sanchez-Romero & Casadesus, 2014). A significant effect is instead attributed to specific efflux pump proteins of the plasma membrane which actively extrude antibiotics from growing colonies. One molecular function of these pump systems is to recognise noxious agents that have penetrated the protective cell wall of the microorganism (Amaral *et al.*, 2014; Amaral *et al.*, 2011; Pages & Amaral, 2009). By actively lowering intracellular concentrations of the relevant compound, bacteria are able to survive higher external antibiotic

concentrations. This ultimately leads to clinically-relevant levels of resistance (Nikaido, 2001; Blair *et al.*, 2014).

Efflux pumps (**Figure 1.2**) are fundamental to the physiology of bacteria by contributing to their survival and adaptation in a discrete ecological niche (Nikaido *et al.*, 1998; Prouty *et al.*, 2004). One example is the resistance nodulation division (RND) efflux pumps which are required for host virulence (Bina *et al.*, 2008; Buckley *et al.*, 2006) and are implicated in biofilm formation (Baugh *et al.*, 2012; Blair *et al.*, 2014). Other types of efflux pump actively export host-derived antimicrobial agents as well as foreign substances from the external environment (Lacroix *et al.*, 1996; Buckley *et al.*, 2006). There is evidence that efflux pumps of some species are capable of exporting virulence determinants such as adhesins, toxins or proteins responsible for colonisation and host infection (Jerse *et al.*, 2003; Burse *et al.*, 2004).

Six families of efflux-pump proteins defined for Gram-negative bacteria are outlined in **Figure 1.2**. These are classified according to the number of components (single or multiple), the number of transmembrane-spanning regions, the energy source utilised, and the type of exported substrate (Piddock, 2006). These families are discussed in more detail in Chapter 5.

These efflux systems are tightly regulated given their importance to survival of bacterial populations. Regulation is most common through transcription regulators (Romero-Rodriguez *et al.*, 2015) many of which have attracted widespread scrutiny as a consequence of their role in antibiotic resistance. About 50 families of bacterial transcription regulators have been reported to date (Rodionov, 2007). Two of these families are of specific interest for this thesis: the TetR family transcriptional regulators (TFRs) which regulate biosynthesis of antimicrobials, efflux pumps, and osmotic stress; and the LysR-type transcriptional regulator (LTTR) family which can either repress or activate transcription and are known to regulate carbon and nitrogen metabolism (Rodionov, 2007).

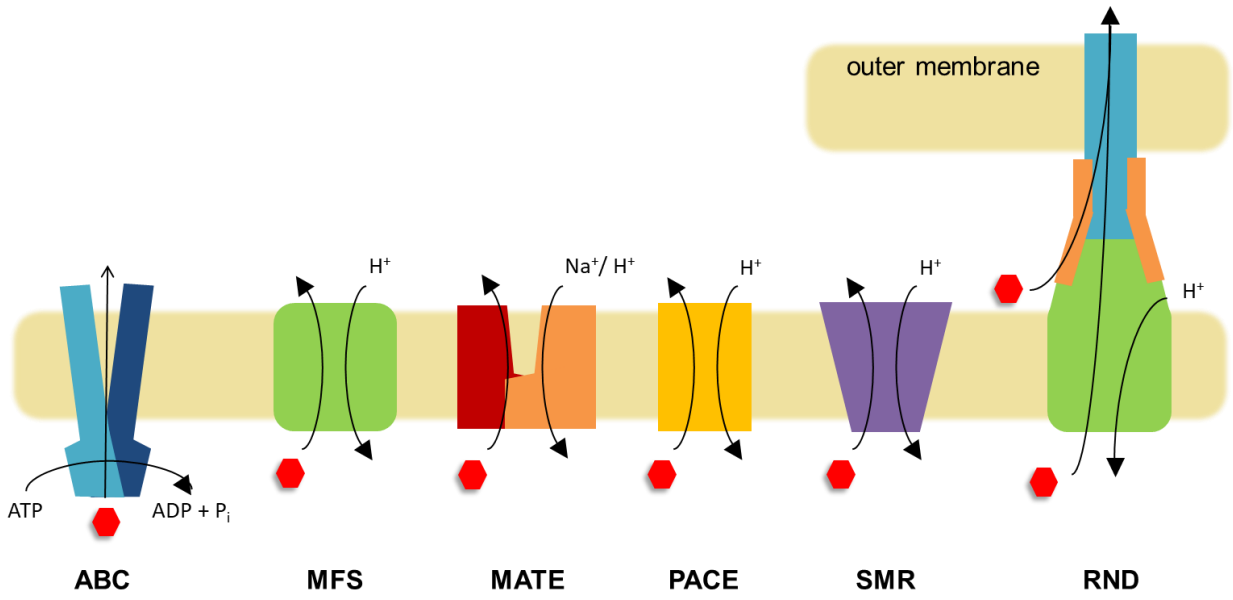


Figure 1.2 Diagram of six types of efflux pumps found in Gram-negative bacteria (adapted (Blair *et al.*, 2014)). Arrows show the movement of ions and target molecules or antimicrobials (red hexagon).

In **Part II** of this thesis, several biophysical characteristics for regulator proteins have been explored for four efflux systems in the ESKAPE pathogen *Acinetobacter baumannii*. These efflux systems are able to transport multiple antibiotic classes, and thereby confer MDR to growing colonies of this organism. The focus was to understand the specificity of substrate interactions across these regulators. *Acinetobacter baumannii*, a Gram-negative coccobacillus, now leads the World Health Organisation (WHO) list of multi-drug resistant bacteria requiring urgent research into modes of resistance acquisition (Rice, 2008).

1.3. Biophysical methods for monitoring protein fold states

1.3.1. Using differential scanning fluorimetry to measure protein stability

One of the primary techniques integrated across all aims is the technique of differential scanning fluorimtery (DSF). DSF was chosen as it is a convenient method to evaluate the thermal stability of proteins under a range of conditions, including the interaction of small molecule ligands (Niesen *et al.*, 2007). A fluorescence response is detected through an added dye reacting to alteration of

hydrophobic components of a solution. This occurs when a folded protein is heated and unfolds to expose interior hydrophobic sidechains. Several different dyes have been used to monitor the unfolding process, including Nile red, SYPRO Orange, and dapoxyl sulfonic acid. To date, SYPRO Orange (excitation 465 nm, emission 590 nm) has proven to have the most favourable properties for use in DSF experiments, primarily due to its high signal-to-noise ratio (Senisterra & Finerty, 2009).

DSF has the capacity to monitor 96 samples using relatively inexpensive materials and is completed in just over an hour using a real-time polymerase chain reaction (RT-PCR) thermocylcer. It is therefore commonly used to determine optimal buffer conditions for proteins, or to evaluate chemical compounds as potential binding ligands (Simeonov, 2013; Senisterra & Finerty, 2009). At a simplistic level, an increase to the melting temperature (T_M) observed by DSF implies an elevated stability of the protein system. This may be through improved stability brought about by interactions with a specific small molecule ligand, or through a more favourable environment of ionic strength and counter-ions, a reasoning that follows from differential scanning calorimetry (Waldron & Murphy, 2003).

Examples of applications include monitoring purified protein to establish production process reproducibility (batch to batch), as a survey method to determine the impact of mutations on protein folding, as a screening tool for optimal storage conditions, and as a medium-to-high throughput platform to screen small molecule stabilisers of protein targets for drug discovery (Simeonov, 2013). There has recently been a significant diversification of DSF applications beyond initial small molecule discovery into areas such as protein therapeutic development, formulation studies, and mechanism of action studies involving small molecule inhibitors in complex enzymatic reactions (Menzen & Friess, 2013; Lavinder *et al.*, 2009; Lea & Simeonov, 2012).

However, DSF does not work for every protein in every situation. In general, when a protein does not show a good transition curve in DSF, this is due to high initial fluorescence values, likely caused by the dye binding to hydrophobic parts of the protein that are exposed even when it is fully folded. Since the instruments have a finite dynamic range for measuring the fluorescence

signal, strong initial fluorescence can saturate the detector and prevent the observation of protein unfolding during temperature scanning (Senisterra & Finerty, 2009). For some proteins, no transition can be obtained in temperature scanning experiments, which can be explained by high protein T_M that exceed the maximum temperature limit of the instruments. For proteins with a high fluorescence background, however, employing differential static light scattering, which is an aggregation-based method, can be a solution. There is also the option of using circular dichroism (CD) to monitor protein denaturation. This is discussed more in Section 3.2.4.

1.3.2. Size exclusion chromatography as a purification and analysis technique

Size exclusion chromatography (SEC) separates macromolecules according to their hydrodynamic volume, which is defined by the Stokes radius of the tumbling form in solution (Sheehan, 2009). Stokes radius is closely related to solute mobility, factoring in not only size but also solvent effects. A smaller ion with stronger hydration, for example, may have a greater Stokes radius than a larger ion with weaker hydration. This is because the smaller ion drags a greater number of water molecules with it as it moves through the solution (Atkins & De Paula, 2006).

Size exclusion media, used in size exclusion columns, consists of porous polymer beads with clearly defined pores of specific dimensions. In mobile phase, particles with smaller hydrodynamic volumes have a longer path length (Striegel, 2009) and therefore can be separated from species with larger hydrodynamic volumes. For this reason it is an appropriate method to separate protein species of different sizes. An elution profile should be nearly symmetric for a sample consisting of homogenous and monodisperse particles. Any peak asymmetry should not go unnoticed and can be indicative of particle-column interaction, multimerisation or heterogeneity.

When using a set of standards (proteins with known mass and dimensions) a SEC column can be calibrated such that the elution time corresponds to the mass of the particle, therefore allowing for the determination of approximate molecular weight. However, this technique assumes the protein can be approximated by a simple sphere whose radius scales linearly with mass. For asymmetric or elongated particles, a calibrated SEC column will give erroneous mass estimates,

since it is essentially the largest dimension of the particle that determines how the particle will travel through the column (Rambo, 2017). For research that requires precise measurements of protein size, added scattering techniques such as multi-angle light scattering (MALS (Ogawa & Hirokawa, 2018)) or small-angle X-ray scattering (SAXS (Glatter & Kratky, 1982)) should be used.

1.4. Experimental aims

It has now been outlined that LGT and efflux systems are two mechanisms used by bacteria to adapt to rapidly changing environments. LGT utilises integrons, which harbour a large repository of diverse cassette-associated genes, as key players in bacterial adaptation and evolution. The cassette metagenome is enriched in novel and highly unique bacterial genes, and there is a need to enhance the functional annotation of this mobile gene pool. In response to this need, previous studies have identified the structures of six proteins encoded by such bacterial genes, revealing them to be novel in fold (Sureshan *et al.*, 2013). To further this research, one of the aims of this thesis was to understand the intra-molecular forces which enable these 3D folds. This was done by exploring the enthalpic contributions to folded and unfolded forms of these new proteins by monitoring steps of unfolding of native sequence protein, as well as panels of selected mutants designed to perturb hydrophobic and hydrogen-bonding interactions presumed to stabilise the folded state.

The specific aims of this study regarding these novel-fold proteins were therefore to:

- Prepare solution forms of each protein in order to determine their structural integrity and possible binding partners. This was done using recombinant forms that were easy to purify.
- Probe each protein for possible ligands in solution using the high-throughput method of DSF paired with a 96-well small molecule cocktail screen.
- Study each protein in its native form, and then design and prepare single amino acid mutants for amenable targets to test overall protein stability and answer specific questions of disulphide interactions and ligand binding.

Membrane drug efflux systems are prominent resistance factors in bacteria, acting as regulators for virulence, stress and biofilm formation. *A. baumannii*, a major hospital-acquired opportunistic human pathogen, carries intrinsic resistance genes in its core genome, including a large number of putative drug efflux pumps. Multidrug efflux pumps are frequently encoded in close proximity to a regulatory protein that controls expression of the pump gene in response to its substrates. *A. baumannii* encodes six families of efflux-pump proteins, some with identified regulator proteins such as the resistance nodulation division AdeABC/AdeRS system (Magnet *et al.*, 2001; Marchand *et al.*, 2004), and others with only hypothesised regulators. Although these efflux pump regulator proteins have well-conserved structures, they tend to display low sequence identity between family members, possibly because of evolving distinct effector recognition. For this reason, it has become important to understand the binding specificities of these proteins, which could in turn infer functionality of bacterial efflux systems as a whole. Therefore, another aim of this thesis was to express a selection of regulator proteins in solution for the first time using recombinant means. These proteins were then to be screened against an assortment of chemically diverse small molecules, using changes in melting temperature as an indicator of molecular interaction.

The specific aims of this study regarding these regulator proteins were therefore to:

- Design protein variants to maximise solubility and stability in solution.
- Test each protein to identify oligomerisation states and stability in solution using SEC and DSF.
- Determine ligand affinity for structurally sound regulators using a panel of small molecules in conjunction with measuring changes in melting temperature.

PART I

CHAPTER 2

Molecular Organisation of Novel-fold proteins from Bacterial Metagenomes

As previously discussed in Chapter 1, the technique of cassette-PCR (Stokes *et al.*, 2001) allows sequence-independent recovery of gene cassettes from metagenomic DNA, mixed cultured populations, and defined bacterial cell lines (Stokes *et al.*, 2001; Boucher *et al.*, 2006; Robinson *et al.*, 2008). It utilises PCR primers targeted to integron recombination sites to directly recover cassette-associated genes harboured within the integron array. The technique can thus amplify multiple gene cassettes from any microbial colony without prior knowledge of their genomes. As a result of revealing protein-encoding gene sequences hidden from conventional genomic endeavours, a notable proportion appear to be novel, with sequences that cannot be readily matched to other known proteins (Sureshan *et al.*, 2013). From the first structural surveys of gene cassette proteins (Robinson *et al.*, 2005; Robinson *et al.*, 2007) marked features were found to be relatively short sequence lengths and organisation into homo-oligomers in solution.

Four proteins, regarded as new structural family representatives, recovered from gene cassettes identified in environmental bacterial species have been studied in this work. They are Hfx_cass1, Hfx_cass5, Vch_cass3, and Vch_cass14, henceforth shortened to Hfx1, Hfx5, Vch3, and Vch14. The Hfx proteins were found within the genomes of soil bacteria collected from Halifax Harbour, Canada, whereas Vch proteins are from *V. cholerae* strains recovered from a brackish marine environment (Sureshan *et al.*, 2013). The four integron cassette proteins that are studied here form dimeric (Hfx5, Vch3, Vch14) and trimeric (Hfx1) structures in solution. However, as each is very structurally different, and therefore oligomerises in unique ways, this chapter will start with an introduction to concepts in protein oligomerisation, followed by detailed analyses of the structures and sequences of each of these four proteins. It then finishes with experimental plans that were carried out in subsequent chapters.

2.1. Organisation and evolution of protein oligomerisation

As is seen with the novel-fold proteins mentioned above, generally proteins are biologically active as oligomers containing some features of symmetry (Goodsell & Olson, 2000; Abraham *et al.*, 2009). The advantages of oligomerisation include the possibility of allosteric control, higher local concentration of active sites, larger binding surfaces, new active sites at subunit interfaces, and economic ways to produce large protein interaction networks and molecular machines (Liu & Eisenberg, 2002). Symmetrical structures often result from the homomeric association of elements that are not themselves symmetrical (Monod *et al.*, 1965). The reasons for this pervasive symmetry remain speculative: the symmetrical state could be the lowest-energy state (Blundell & Srinivasan, 1996), or symmetrical organisation may allow simple oligomerisation into a defined number of elements to avoid aggregation (Goodsell & Olson, 2000). However, it is common for functional proteins to oligomerise either as homomers or heteromers.

2.1.1. Homo-oligomerisation

Homo-oligomerisation is the association of multiple identical components, or in the case of proteins, chains. *Escherichia coli* proteins show an average oligomerisation state of ~4 and only a minority of proteins are found in monomeric form (Levy *et al.*, 2008). In general, the single most frequent complex state of a protein is a homodimer with a one symmetry rotation axis (60–70 % of all known complexes). Homotetramers are less frequent (15–20 %), while homotrimers, homohexamers and homo-octamers are even rarer (Goodsell & Olson, 2000; Levy *et al.*, 2006). It can be seen in **Figure 2.1** that while Vch14 takes the form of a homodimer in the crystal lattice, a representative for the most common quaternary organisation seen in proteins, the other three novel-fold proteins take on more complex organisations. Hfx1, for example is a homotrimer, whereas Hfx5 appears to be a candidate for a domain-swapped dimer.

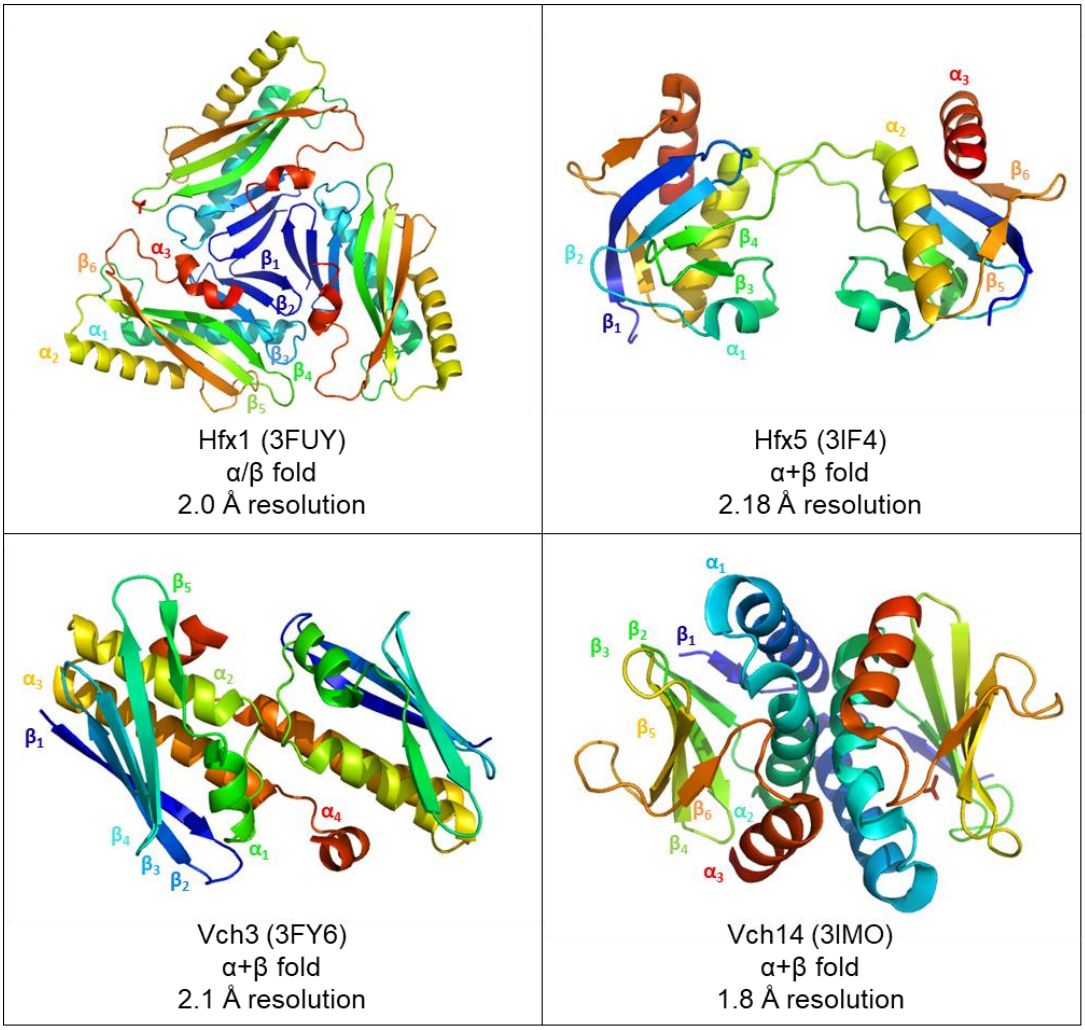


Figure 2.1 3D structures of the four novel-fold proteins analysed in this thesis. Each chain is represented in rainbow format (blue = N-terminal, red = C-terminal end) with secondary features labelled. PDB code, fold type and crystal structure resolution are listed for each protein.

One mechanism that allows for homo-oligomerisation is domain swapping. 3D domain swapping is a mechanism for two or more protein molecules to form a dimer or higher oligomer by exchanging an identical structural element or domain (Liu & Eisenberg, 2002). If both the monomer and the dimer of a molecule exist in stable forms, in which the dimer adopts a domain-swapped conformation and the monomer adopts a closed conformation, then this protein is considered to be a true example of 3D domain swapping. However, there are other examples where proteins exchange domains and no longer can form monomers. These are considered to be `quasidomain swapped' or a candidate for 3D domain swapping (Schlunegger *et al.*, 1997).

2.1.2. Hetro-oligomerisation

Hetero-oligomerisation is the organisation of multiple proteins or chains that are not themselves similar in sequence or structure. New hetero-oligomeric assemblies are most often created by gene duplications of a pre-existing homomer (Pereira-Leal *et al.*, 2007). The resulting oligomeric paralogs initially coassemble because both have the same sequence (and hence structure and interfaces) as their ancestor (Kaltenegger & Ober, 2015). This coassembly can easily become entrenched if evolution of the two resulting duplicates is functionally constrained to maintain the interaction (Diss *et al.*, 2017; Finnigan *et al.*, 2012), implying that heteromerisation should be the most likely fate of oligomeric paralogs. However, this is not the case, and studies have shown that hetero-oligomerisation acts as a constraint on the functional divergence of oligomeric paralogs (Baker *et al.*, 2013). Relieving this constraint is therefore a key step in the evolutionary trajectories of oligomeric proteins toward evolving new functions.

Despite this, protein–protein interactions within oligomeric organisations cause constraints which result in interface positions being more conserved than other surface positions, whether homomeric or heteromeric interactions (Elcock & McCammon, 2001; Teichmann, 2002; Valdar & Thornton, 2001). Moreover, these evolutionary pressures vary according to the nature of the interaction: residues involved in permanent interactions evolve more slowly than those required for transient interactions, which in turn evolve more slowly than non-interacting surface residues (Caffrey *et al.*, 2004; Mintseris & Weng, 2005; Teichmann, 2002).

Hetero-oligomerisation capacity has partly been addressed by recent outputs from the research group led by David Baker in Seattle, Washington (King *et al.*, 2014). Through a well validated computational method for modelling protein tertiary folds, this team has designed protein nanomaterials incorporating mixtures of protein domains co-assembled to a target symmetric architecture. The building blocks used for their designs were small and robust oligomerising proteins, one of which is the Macquarie-defined structure of Hfx1. This was combined with another trimeric protein to form a 3D tetrahedral cage structure as depicted in **Figure 2.2**.

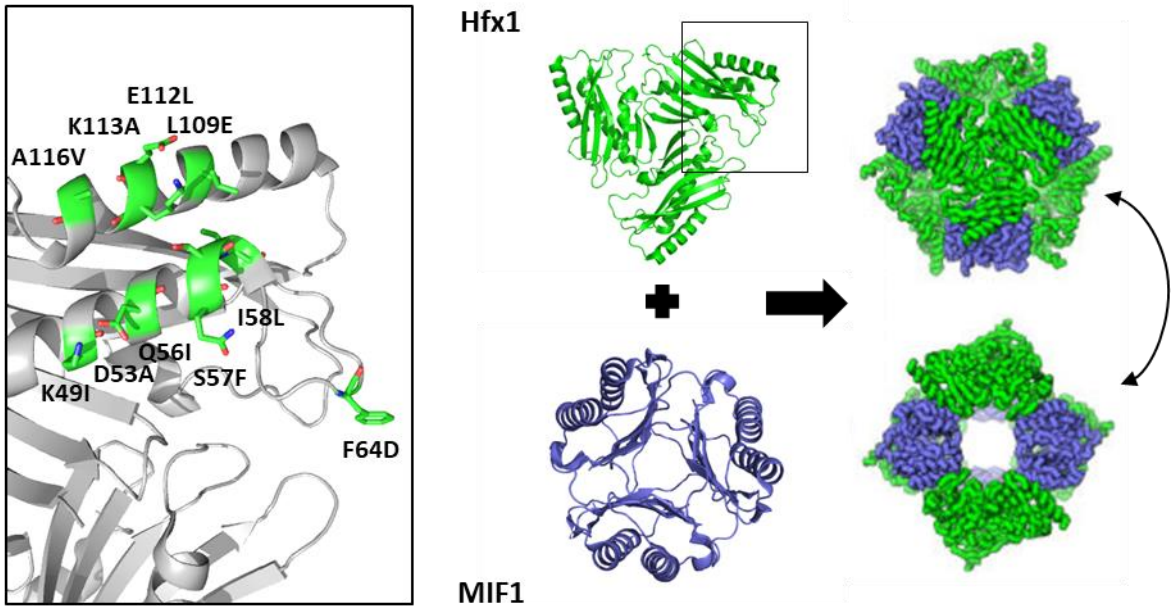


Figure 2.2 Cage nanostructure self-assembled from proteins Hfx1 and MIF1. Left: Close-up of the designed interface of Hfx1 with mutant residues labelled (King *et al.*, 2014). Right: The Hfx1 trimer (Sureshan *et al.*, 2013) and MIF1 trimer (Richardson *et al.*, 2009), when co-expressed in *E. coli* form the cage structure on the right (PDB 4NWR).

In the design of Hfx1 for its interaction with MIF1, interface areas between each protein needed to be mutated to promote self-assembly into this 24-mer structure. Residues were chosen in structured areas along helices α_1 and α_2 and mutated to hydrophobic amino acids in order to create a hydrophobic interface when self-assembly occurred (**Figure 2.2**). Although the substitution of these residues with hydrophobic amino acids resulted in insoluble protein, when each component was co-expressed with its partner, the resultant structure was soluble.

This is a fascinating design application that uses the inherent stability of small proteins to its advantage and addresses the evolutionary pressures of such proteins to form homo-oligomers and hetero-oligomers. Although the changes made to Hfx1 and MIF1 to promote their protein-protein interaction were performed all at once, these changes usually occur naturally over periods of millions of years. Evolution of proteins is slow, because most amino acid substitutions are forbidden at any given time owing to their deleterious effects on protein structure, function or expression (Povolotskaya & Kondrashov, 2010). However, almost all positions stand a chance to

undergo substitutions over an evolutionary timescale, following changes in other positions. As proteins evolve in rugged fitness landscapes, the tolerated substitutions at one moment depend heavily on whatever mutations have occurred previously, owing to possible compensations (Povolotskaya & Kondrashov, 2010). This concept is termed epistasis, or non-additive interactions between mutations (Andreani & Guerois, 2014).

The proteins studied here have most likely undergone these evolutionary pressures, possibly starting out as gene duplications, and slowly evolving to become structurally and sequentially novel. In order to elucidate their possible functions, each protein has been presented below (Hfx1, Hfx5, Vch3, and Vch14) and their 3D structures and sequences analysed so as to inform the experimental approaches taken in later chapters, where oligomerisation in solution and fold stability have been studied through mutation and melting temperature determination. PISA, an online tool that accounts for chemical thermodynamics involved in protein oligomerisation interfaces, was used to better understand protein oligomerisation interfaces. Inference of protein assemblies from crystalline state is often seen as a bioinformatical problem. In the framework of informatics-based approaches, macromolecular interfaces, found in crystals, are classified into “biologically relevant” and “insignificant” (crystal packing), according to a certain scoring system (Ponstingl *et al.*, 2000). The score may depend on the interface area, residue/atom composition and contacts, hydropathy index, charge distribution, topological complementarity, and other parameters (Krissinel & Henrick, 2007). PISA (Krissinel & Henrick, 2007) was developed to help distinguish between biologically relevant and insignificant protein interfaces and was used in this work to further analyse oligomeric contacts of the novel-fold proteins. The tool lists the most thermodynamically favourable oligomeric structures along with their parameters (interface area, ΔG, and the ΔG p-value) and all interface contacts including hydrogen bonds, salt bridges, disulphide bonds, and hydrophobic residues inaccessible to solvent. From this it was possible to better understand relative contributions to protein fold at the side-chain level for Hfx1, Hfx5, Vch3, and Vch14.

In addition to the analysis of oligomeric contacts of these proteins, sequences were studied in conjunction with structures to identify key residues involved in protomer stabilisation, oligomeric

stabilisation, and possible ligand binding sites. This was helped by comparison of homologous protein sequences where conserved motifs were identified in regions of the structure that may be evolutionarily important to protein function.

2.2. Trimers of Hfx1: Alpha/beta fold structure

The first protein to be analysed in such a way was Hfx1. As depicted in **Figure 2.3** and **2.4**, the trimer of Hfx1 forms a distinctive flattened shape (Sureshan *et al.*, 2013). Each protomer subunit contains a mixed six-strand central sheet flanked on one side by two extended α-helices, and on the alternate face by a 3_{10}-helix (helix 3, **Figure 2.3 A**). All three helices pack against the central beta sheet via a mix of hydrogen bonds and hydrophobic contacts. Hydrophobic residues inaccessible to solvent (found using PISA (Krissinel & Henrick, 2007)) are indicated in **Figure 2.3 D** and form a hydrophobic core that helps to stabilise each protomer.

2.2.1. Oligomerisation contacts

The Hfx1 crystal structure indicates the three component chains to engage at three symmetrical interfaces of ~700 $Å^2$. Residues involved in stabilising these interfaces are depicted in **Figure 2.4 C**. The B-factor representation of Hfx1 (**Figure 2.4 A**) shows that the extremities (β_5-α_2) have far more flexibility than the rest of the protein, even compared to the exposed areas on the top and bottom of the flattened structure. Hydrogen bonds and salt bridge contacts measured by PISA analysis contribute an average ΔG of -8.1 kcal/mol. A reduction in G (Gibbs free energy) is thermodynamically favourable and a necessary condition for the spontaneity of processes at constant pressure and temperature, which allows for the oligomerisation of protein chains (Perrot, 1998). Critical contacts are listed in **Table 2.1** and include:

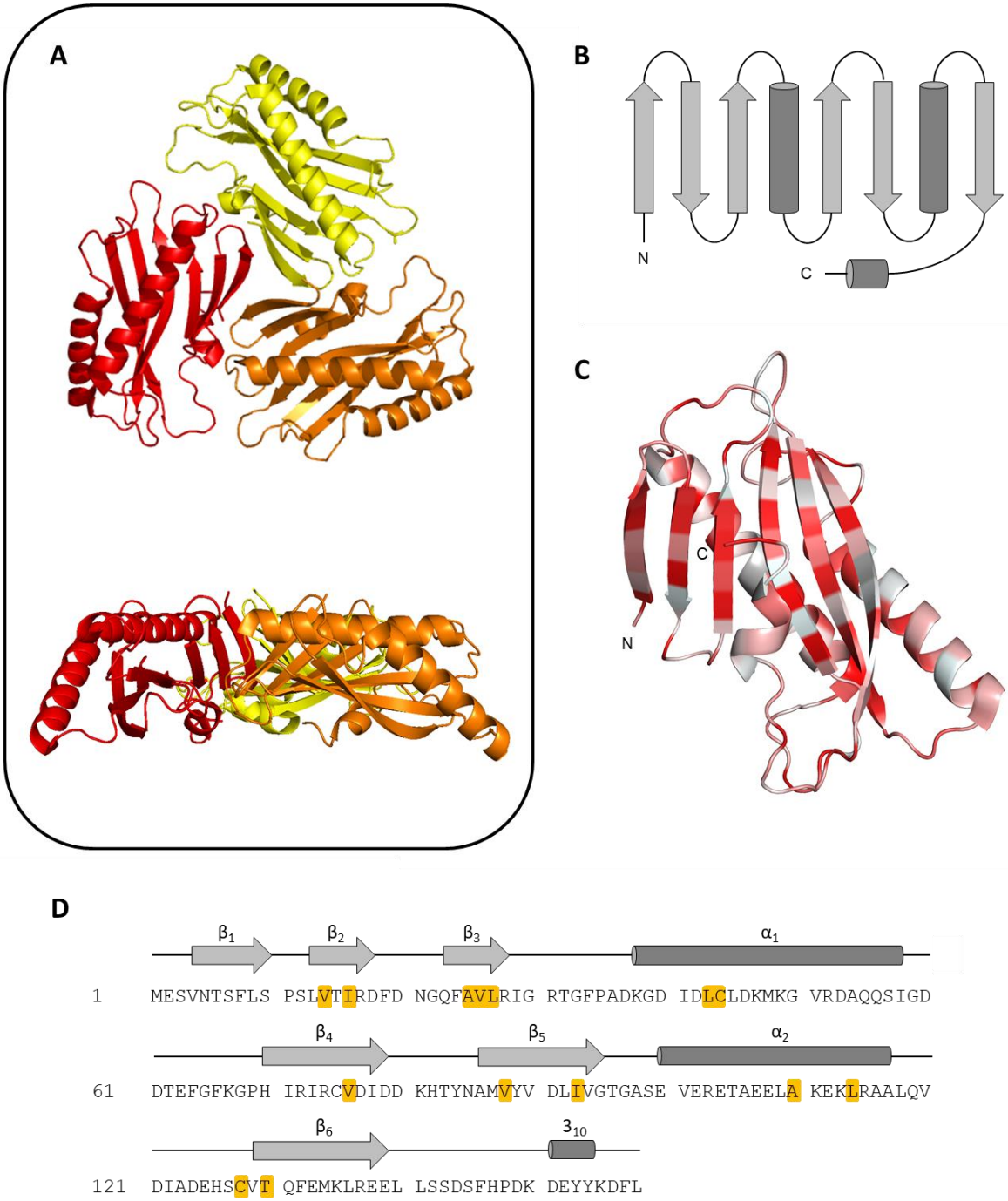


Figure 2.3 Crystal structure of Hfx1 (PDB 3FUY (Sureshan *et al.*, 2013)). (**A**) Ribbon views of trimer. (**B**) Topology map of monomer. (**C**) Ribbon structure coloured by hydrophobicity (red → hydrophobic, white → hydrophilic). (**D**) Hfx1 sequence mapped to secondary structure. Residues inaccessible to solvent are in orange (Scientific, 2006).

i) salt bridges (between His-82'/Asp-144 at 2.9 Å and Lys-38'/Asp-20 at 4.5 Å, depicted in red **Figure 2.3 C**) which engage two adjacent loops (β_4'-β_5' and β_6-helix 3) and the N-terminus of α_1' and the β_2-β_3 loop.

ii) hydrogen bonds between residues on β_1 and β_2 and hydrophobic loops, and the C-terminal 3_{10}-helix (Ser-12' - Asp-151 - Gly-33').

The interfaces of Hfx1 are not only stabilised by hydrogen bonds, but also include hydrophobic contacts across the trimer that engage loop residues between β_3-α_1 (e.g. Thr-32 – Ala-36) and those on β_6-3_{10} loop (Phe-146).

Table 2.1 Bond partners across A/C interface of Hfx1 trimer [a]

	chain A	**chain C**
hydrogen bond [b]	Asn-5 [$N_{\delta 2}$]	Thr-6 [O]
	Ser-7 [O_{γ}]	Phe-8 [N]
	Thr-15 [$O_{\gamma 1}$]	Phe-8 [O]
	Arg-17 [$N_{\eta 2}$]	Ala-36 [O]
	Arg-17 [N_{ε}]	Ala-36 [O]
	Asp-151 [$O_{\delta 2}$]	Ser-12 [O_{γ}]
	Asp-151 [N]	Gly-33 [O]
salt bridge	Asp-20 [$O_{\delta 1}$]	Lys-38 [N_{ζ}]
	Asp-144 [$O_{\delta 2}$]	His-82 [$N_{\delta 1}$]

[a] As recorded by PISA (Krissinel & Henrick, 2007)
[b] Listed hydrogen bond contacts are 2.7-3.4 Å

2.2.2. Sequence analysis of Hfx1

The proposed active site of Hfx1 is based solely on structure, as there are no significant sequence homologs currently known for the protein (Sureshan *et al.*, 2013). A search for sequence homologs of Hfx1 conducted in May 2018 revealed a hypothetical protein (A2X20_ 08820 from *Bacteroidetes bacterium* GWE2_40_15) at 28% sequence identity to be the only sequence of some similarity. However, when aligning these two using tools such as NCBI Protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and TCoffee (http://tcoffee.crg.cat/apps/tcoffee/do:expresso), there were no identifiable regions of similarity. In the structure of Hfx1, two splayed strands, β_4 and β_5, create a narrow polar cavity, occupied by water molecules in all three subunits (Sureshan *et al.*, 2013). Surrounded by pronounced acidic clusters, largely found on loops, this region has been proposed to have the appearance of a functional binding site (Figure 2.3 A (Sureshan *et al.*, 2013). The β-bulge secondary feature between β_4 and the adjacent β_5 strand, has long been associated with active sites of proteins (Richardson, 1981).

The structure of Hfx1 contains four histidine residues in every protomer (His-70, His-82, His-126, and His-147). His-70 is located at the base of strand β_4, His-82 on the β_4-β_5 loop, His-126 on the α_2-β_6 loop, and His-147 on the β_6-α_3 loop (**Figure 3.5 A**). These histidine residues surround the area hypothesised to be the ligand binding pocket (Sureshan *et al.*, 2013). In addition, His-82 is involved in one of the salt bridges between subunits, possibly playing a role in oligomerisation.

One intriguing aspect about the Hfx1 trimer is that there are three cysteines residues per protomer (Cys-44, Cys-75, and Cys-128). Cys-44 is located along α_1 and is completely inaccessible to solvent, Cys-75 is on β_4 facing helix 3, and Cys-128 is on the α_2-β_6 loop, also inaccessible to solvent. The crystal structure shows none to be located close enough to any other to form disulfide bonds (the closest at 6.4 Å), either within each protomer or with adjacent protomers in the crystal lattice (**Figure 2.4 B**).

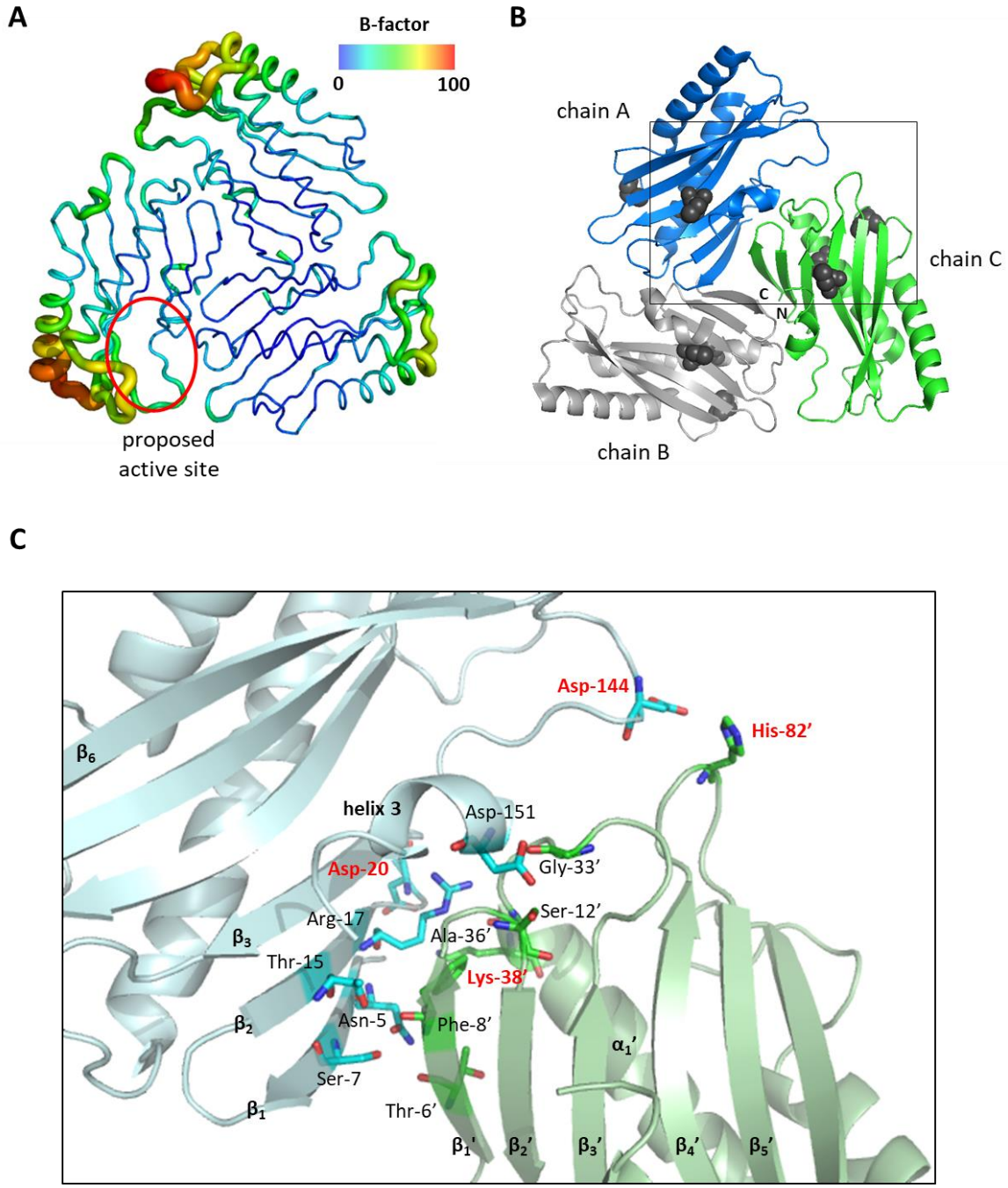


Figure 2.4 Oligomerisation interfaces of Hfx1. (**A**) Putty view of the trimer, based on B-factor values with proposed active site circled in red. (**B**) Hfx1 with each protomer distinguished by colour and cysteine residues represented as black spheres. Boxed section is magnified below. (**C**) Details of residues involved in hydrogen bonds and salt bridges between protomers. Chain C residues are labelled [']. Residues engaging in salt bridges are labelled in red as per **Table 2.1**.

2.2.3. Experimental plan for Hfx1

One of the aims of this work is to probe the tertiary structure of Hfx1 using a number of single residue mutants to test the stability of the protomer and subsequent trimer. As exemplified in the comprehensive work by the Fersht group on barnase (a 12 kDa protein of α+β fold class (Fersht, 1993)), mutations can be prepared to delete side chains that make simple and defined interactions, yet do not cause any disruption of structure other than removal of those interactions. Residues targeted for these changes are generally involved in hydrophobic interactions and hydrogen bonds. As shown in **Figure 2.3 D**, there are fourteen residues that contribute to protein stability through inter-chain hydrophobic contacts. However, there are only four that are inaccessible to solvent and are therefore hypothesised to play the largest role in protomer stabilisation. In addition to hydrophobic interactions are hydrogen bonds, which when added together within a protomer contribute a large stabilising force. The Fersht group suggests, however, that only residues that form hydrogen bonds with the backbone be targeted for mutation, as they are not likely to wholly destabilise the protein structure. Residues identified within the structure of Hfx1 that target these individual bonding contacts for disruption are as follows:

I) isoleucine residues residing in the hydrophobic core (Ile-16, Ile-29, Ile-78, and Ile-93) can be mutated to valine to destabilise hydrophobic interactions.

II) tyrosine and serine residues (Tyr-154, Ser-10, Ser-57, Ser-99, and Ser-143) can be modified to phenylalanine and alanine, respectively, to remove specific hydrogen bonds to the peptide backbone

Plans to mutate these identified residues were made in order to test their effects on the stability of HfX1. In addition to these small impact mutations, single mutants of Cys-44, Cys-75, and Cys-128 to a serine were engineered in order to disrupt any possible intramolecular contacts and to see effects on the stability of Hfx1 under different redox conditions. It must be noted that although these mutations were planned to test the intramolecular stability of Hfx1, all mutants were also analysed to determine if these changes had an effect on protein oligomerisation.

Finally, Hfx1 was exposed to a range of small molecules in order to probe its function. Given the possible effector molecules that arose from these screens, further active site mutants were also created to test their role in binding these ligands, such as the histidine residues surrounding the proposed active site. All four His residues (His-70, His-82, His-126, and His-147) were targeted for mutation to leucines in order to disrupt any possible contacts and test their role in ligand binding and protein oligomerisation.

2.3. Domain-swapped Hfx5: Alpha + beta fold structure

The crystal structure of Hfx5 (**Figure 2.5 A**) reveals a symmetrical domain-swapped dimer, each chain contributing to a pair of compact α+β domains (Sureshan *et al.*, 2013). One face of each domain contains a five-stranded β-sheet (topology β_6'-β_1-β_2-β_4-β_3), which incorporates a strand from the paired chain. This central sheet is overlayed by helices α_2' and α_3' and strand β_5'. An extended loop between strands β_2 and β_3, which includes a short 3_{10}-helix, contributes to the inter-domain contact. A long unstructured Pro-rich linker located between β_4 and α_2, connecting domains should also be noted. As observed in other domain-swapped proteins, an extended conformation in the domain-swapped dimer is common for linker regions (Rousseau *et al.*, 2012).

2.3.1. Oligomerisation contacts

The atomic interactions between the Hfx5 sheet and its overlaid features include many hydrogen-bond and hydrophobic contacts spanning an area of over 2645 $Å^2$ resulting in a highly stable interface (measured by PISA as ΔG of -48.2 kcal/mol, **Figure 2.5**). Residues involved in hydrogen bonding and salt bridges are listed in **Table 2.2**. Regions of slightly elevated mobility (distinguished by higher B-factors) includes the Pro-containing inter-domain linker, the C-terminal helix (α_3) and the β-turn between strands β_5 and β_6 (**Figure 2.5 B**).

It should be noted that within the crystal lattice, Hfx5 is grouped as a tetramer (**Figure 2.5 A**). This seems, however, to be a product of crystal packing, with very few hydrogen bonds or other

contacts across the dimer pairings at helix 1 (i.e. between chain A and C). In fact, PISA identified only eight stabilising hydrogen bonds between the two dimers (average ΔG of -2.1 kcal/mol, **Table 2.2**).

Technically, in order to be classed as a domain swapping protein, both monomer and domain-swapped forms need to be experimentally observed (Mascarenhas & Gosavi, 2017). In the case of Hfx5, the crystal structure defines the domain-swapped form, but evidence for a monomer has yet to been found, although it can be considered a 'candidate' for domain swapping. The structure of subunits within domain-swapped oligomers is identical to that of the monomers, with the exception of the region that connects the exchanging domain with the rest of the protein (the inter-domain linker). Originally the proline composition of hinges was seen to be critical, but although proline residues contribute to domain-swapping in some proteins (Rousseau *et al.*, 2001; Miller *et al.*, 2010) this is not a universal effect (Barrientos *et al.*, 2002; Cho *et al.*, 2005).

Table 2.2 Bond partners across A/B interface of Hfx5 dimer and across A/C interface of tetramer [a]

	chain A	**chain B**	**chain A**	**chain B**
	Lys-2 [O]	Lys-72 [N]	Mse-34 [O]	Arg-51 [$N_{\eta}1$]
	Glu-4 [O]	Val-74 [N]	Gly-70 [O]	Lys-2 [N]
	Val-6 [O]	Thr-79 [$O_{\gamma}1$]	Lys-72 [O]	Glu-4 [N]
	Ala-7 [O]	Arg-80 [N]	Val-74 [O]	Val-6 [N]
hydrogen	Ile-9 [O]	Val-82 [N]	Arg-80 [O]	Ile-9 [N]
bonds [b]	Ile-11 [O]	Asn-83 [$N_{\delta}2$]	Val-82 [O]	Ile-11 [N]
	Gly-15 [O]	Tyr-55 [O_{η}]	**chain A**	**chain C**
	Tyr-28 [O]	Gln-69 [$N_{\varepsilon}2$]	Ser-24 [O_{γ}]	Arg-31 [$N_{\eta}2$]
	Glu-32 [$O_{\varepsilon}2$]	Gln-69 [$N_{\varepsilon}2$]	Gln-25 [$N_{\varepsilon}2$]	Glu-32 [$O_{\varepsilon}1$]
	Mse-34 [O]	Gln-61 [$N_{\varepsilon}2$]	Glu-32 [$O_{\varepsilon}2$]	Tyr-30 [O_{η}]
salt bridges [c]	Glu-35 [$O_{\varepsilon}1$]	Arg-51 [$N_{\eta}1$]		

[a] As recorded by PISA (Krissinel & Henrick, 2007).
[b] All hydrogen bonds are doubled across the dimer but are not listed here twice. Hydrogen bond contacts are between 2.6 and 3.6 Å.
[c] Salt bridge contact is 3.1 Å.

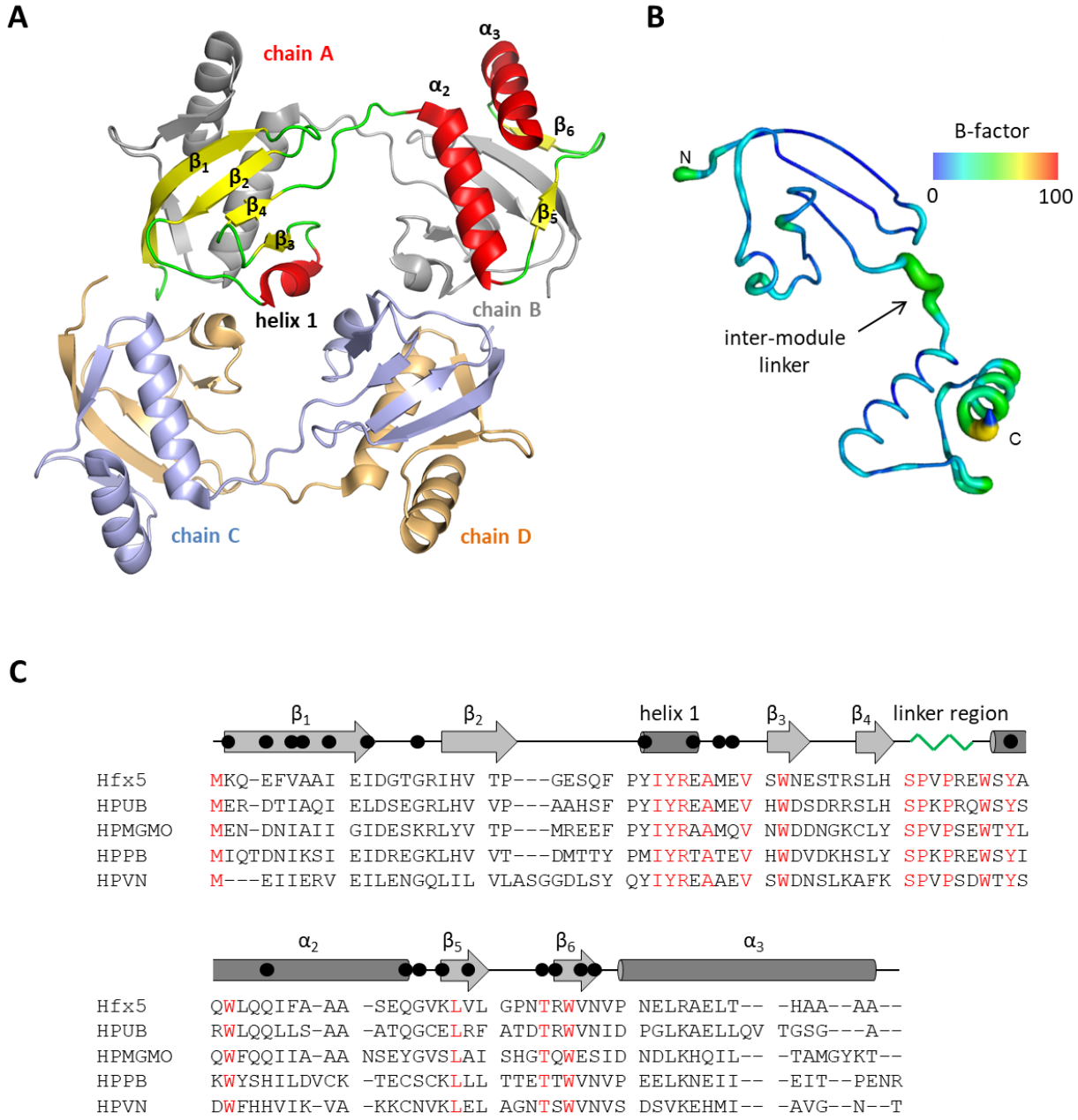


Figure 2.5 Crystal structure of Hfx5 (PDB 3IF4 (Sureshan *et al.*, 2013)). (**A**) The Hfx5 dimer with chain A coloured and labelled by secondary structure. (**B**) Putty view depicting B-factor with inter-module linker indicated. (**C**) Sequence alignment (performed 2018) of Hfx5 and its closest homologs (HPUB: uncultured bacterium, AAK28610.1, 55% sequence identity; HPMGMO: MGMO_201c00090 *Methyloglobulus morosus* KoM1, ESS66229.1, 47% identity; HPPB: *Prolixibacter bellariivorans*, WP_025865846.1, 43% identity; HPVN: *Vibrio nigripulchritudo*, WP_045961100.1, 37% identity) with fully conserved residues highlighted red. 21 residues involved in oligomerisation between the two chains of the Hfx5 structure as identified by PISA are indicated (●).

2.3.2. Sequence analysis of Hfx5

A small group of sequence homologs were found for Hfx5 in bacterial species from fresh and salt water environments. Alignment of these sequences (**Figure 2.5 C**) immediately revealed strong preservation of residues comprising the 3_{10}-helix (YIYR, helix 1) and the inter-domain linker. This suggests that these features are integral to the structure for this protein family. The consensus sequence (SPxPx_2W) at residues 47-53 of the Hfx5 sequence may serve as a relatively mobile 'hinge' allowing for inter-domain movement.

Each protomer of Hfx5 contains three phenylalanine, three tyrosine, and four tryptophan residues (all tryptophan residues are fully conserved across the small family of homologs, **Figure 3.9 B**), comprising 10% of the total residues in the protein. However, in the crystal structure of the dimer, none of these residues seem to be involved in pi stacking as they are all facing away from each other. There is the possibility of Trp-53/Trp-58 and Tyr-28/Tyr-30 pi stacking following a slight structural rearrangement. This could be the case in the change from the Hfx5 dimer to the monomer.

2.3.3. Experimental plan for Hfx5

The main aim in studying Hfx5 was to understand the structural arrangement of the protein. Domain-swapping is of biotechnological interest because inhibiting domain-swapping can reduce disease-causing fibrillar protein aggregation in humans (Mascarenhas & Gosavi, 2017). To achieve such inhibition, it is important to understand both the energetics that stabilise domain-swapped structures and the protein dynamics that enable the swapping. These analyses were carried out using SEC, circular dichroism (CD) and tryptophan fluorescence. Another aim was to screen Hfx5 for possible ligands, in order to predict its functional role in the environment.

2.4. Dimer of Vch3: an α+β fold

The crystal structure of Vch3 (Sureshan *et al.*, 2013), depicted in **Figure 2.6 A**, reveals a dimer in which each protomer adopts a relatively simple two-layered α+β fold. The N-terminal portion (residues 1-61) of each chain forms an anti-parallel β-sheet of five strands in meander topology. This curves around an antiparallel pair of helices (α_2 and α_3).

2.4.1. Oligomerisation contacts

Within the dimer, helices α_2 and α_2' align essentially end-to-end, creating a distinctive central core with residues from helices α_3 and α_3'. As detailed in **Table 2.3**, the interface region (approximately 1230 $Å^2$) contains a significant number of aromatic side chains which contribute hydrophobic and hydrogen-bonding stability to the dimer resulting in an average ΔG of -24.7 kcal/mol (**Figure 2.6 C**).

Table 2.3 Bond partners and hydrophobic contacts across A/B interface of Vch3 dimer [a]

hydrogen bonds [b]	
chain A	**chain B**
Pro-65 [O]	Glu-67 [N]
Asn-71 [O_δ1]	Tyr-117 [O_η]
Glu-93 [O_ε1]	Ser-111 [O_γ]
Glu-93 [O]	Trp-114 [N_ε1]
Trp-100 [N_ε1]	Val-104 [O]

[a] As recorded by PISA (Krissinel & Henrick, 2007).
[b] Hydrogen bond contacts are between 2.3 and 3.9 Å.

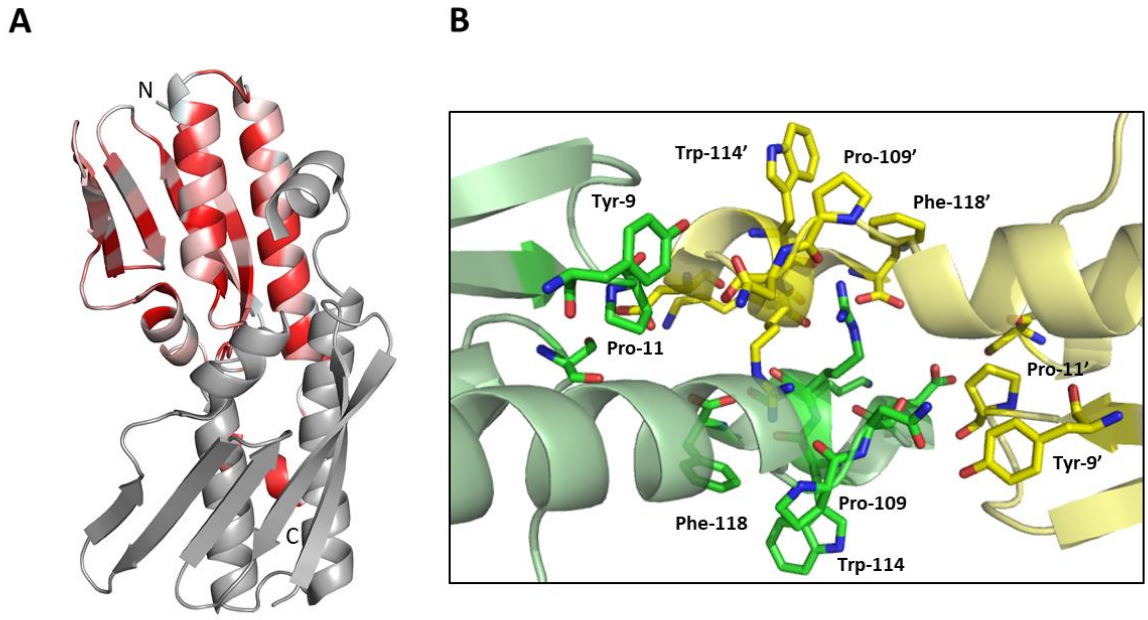
Calculations with PISA (Krissinel & Henrick, 2007) suggested the dimeric Vch3 structure to be the predominant species in solution (interface area of 1200 $Å^2$, ΔG of -25 kcal/mol). However, there is also a smaller tetrameric interface within the crystal lattice (helix α_4 and the β_1'-β_2' loop of adjacent dimers, **Figure 3.10 B**) which PISA did not see as a strong contributor to oligomerisation (interface area of 580 $Å^2$, ΔG of 5 kcal/mol).

2.4.2. Sequence analysis of Vch3

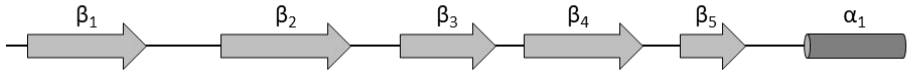
When first crystallised only two sequence homologs of Vch3 were known. However, as of May 2018, there were many sequence relatives with moderate identities (> 45%). Some of these hypothetical proteins from environmental bacterial strains are shown in **Figure 2.6 C**. Alignments showed conserved residues span the whole sequence, with strongest preservation of residues at interface components. This includes residues comprising the β_1-β_2 turn as well as C-terminal residues across helices α_3 and α_4, particularly component hydrophobic sidechains (e.g. Vch3 residues Phe-113, Trp-114, Phe-118). There is also a strong consensus sequence at the C-terminus of Vch3 comprising WxNxV/ITxAKPxT/SxFWx_2YF, which denotes a key amphipathic helix signature necessary to maintain a hydrophobic core.

2.4.3. Experimental plan for Vch3

The aims of research with Vch3 were to express the protein as a stable oligomer and measure the quaternary structure in solution. It was also planned to probe the stability of Vch3 by making single residue mutations designed to disrupt bonds within the protomer. Vch3 was also earmarked to be screened against possible effector molecules, which will help to discover its function.



C




```
VCH3        1  MTEVNLNIYS PRWGRHETYI VELHKDYMEI SMGAVTIKAT YSENQDPEWS EETLQDIMNN 60
HPPM30-35   1  -TTVDLEIFS PRWGNEDTYT VQLNRDFMEI KMSPRVSRAT WIENQDPEWS GESIQAIMNN 59
HPBU        1  MPTIELQIFS PRWGHDDTYE IELERDHMEI TRGANTARAD WQDNADPAWS GWTVEGMMGN 60
HPBU2       1  ---VTFELFS PRWGHGDTYT VELERDHIEI SMQVRTARAE HRENRDPEWS GEAIEDILSN 57
HPBC        1  MPIIELMIFS PRWGHDDTYE IELEHDHMEI TRGANTARAD WQDYADPAWS GWTVEGMMGN 60
HPSMLSB     1  MRIVELKIYS PRWGHHDIYE IELAKDKMTI THNISSAICT WRDNLDPVWS GNNLEDILRN 60
```


```
VCH3        61 DSVYPPEITQ NLFQHAWLEW RKGALDNDEV TRELELVAQW VNKVTEAKPN SDFWRKYF 118
HPPM30-35   60 DNIYPPKITQ DLFEHVWKEW RNGDINDQQA EAELQEIAKW INAVTRAKPD TEFWNKYF 117
HPBU        61 DSIHPPAVTQ RMFQRVWTAW RGGEVDDAQA EAELQALADW INAVTRAKPR TDFWRAYF 118
HPBU2       58 DSIAAPDDLQ GLFEYVWTSW RNGELSDEQV VEEIRALATW INAITAAKPR TDFWRRYF 115
HPBC        61 DSIHPPAVTQ RMFQKVWTAW RGGEINTAQA EAELQALADW INAVTRAKPR TNFWSAYF 118
HPSMLSB     61 DAIYPPAILN DLLEHVWEAW RNGYLKDESV DQELHAVEEW LNTITEAKPK TEFWERYF 118
```

Figure 2.6 Crystal structure of Vch3 (PDB 3FY6 (Sureshan *et al.*, 2013)). (**A**) Vch3 dimer coloured by hydrophobicity (red → hydrophobic, white → hydrophilic). (**B**) The dimer interface with contributing aromatic residues labelled. (**C**) Sequence alignment between Vch3 and its closest homologs (HPPM30-35: *Pseudomonas* sp. M30-35, WP_087515181.1, 55% sequence identity; HPBU: *Burkholderia ubonensis*, WP_059723787.1, 48% identity; HPBU2: *Burkholderia ubonensis*, WP_060011828.1, 48% identity; HPBC: WI95_09460 *Burkholderia contaminans*, AOL04123.1, 45% identity; and HPSMLSB: JU82_08550 *Sulfuricurvum* sp. MLSB, KFN39076.1, 47% identity) with fully conserved residues highlighted red. Residues involved in interface hydrophobic contacts as identified by PISA are indicated (●).

2.5. Dimer of Vch14: an α+β fold

Vch14 (Sureshan *et al.*, 2013) forms a dimer in which each protomer incorporates a single antiparallel sheet of six strands overlaid by a second face of three helices (**Figure 2.7 A**). The two protomers stack orthogonally via these helix faces, allowing each component helix extensive hydrophobic contact with the paired chain (residues indicated in **Figure 2.7 C**). Helix α_2 is completely buried and hydrophobic, whereas helices α_1 and α_3 are amphipathic.

2.5.1. Oligomerisation contacts

Interactions between the chains are extensive, involving approximately 28% of residues (interface area 1100 $Å^2$). Some additional H-bonding (listed in **Table 2.4**) as well as two salt bridges from Lys-111 (to Glu-30' and Asp-23') contribute ΔG -17 kcal/mol. A notable feature of the Vch14 dimer is the highly positively-charged surface presented by each exposed β-sheet face of the dimer (indicated in **Figure 2.7**).

2.5.2. Sequence analysis of Vch14

Vch14 defines the structure for a small protein family from soil-dwelling bacteria that share moderate-to-strong sequence identities (42-84%, **Figure 2.7 C**). Conserved residues were found throughout the Vch14 structure comprising a large portion of hydrophobic residues located within helices α_1 and α_2. This is likely in order to preserve hydrophobic interactions between helices that maintain the dimer, which would suggest conservation of structure across the family. In addition, there were many conserved residues within β-strands that are involved in hydrogen bonds with solvent surrounding the protein (e.g. Ala-2, Asp-60, Trp-63, Tyr-75). These contacts must be necessary to counteract the solvation effects of having an openly hydrophobic binding pocket.

Table 2.4 Bond partners across B/D interface of Vch14 [a]

	chain B	chain D
hydrogen bonds [b]	Lys-6 [O]	Val-8 [N]
	Asp-46 [O]	Asn-9 [$N_{\delta 2}$]
	Leu-41 [O]	Ser-12 [O_{γ}]
	Trp-42 [O]	Ser-16 [O_{γ}]
	Lys-44 [N_{ζ}]	Asn-9 [$O_{\delta 1}$]
salt bridges	Lys-111 [N_{ζ}]	Asp-23 [$O_{\delta 1}$]
	Lys-111 [N_{ζ}]	Glu-30 [$O_{\varepsilon 1}$ & $O_{\varepsilon 2}$]

[a] As recorded by PISA (Krissinel & Henrick, 2007).
[b] Hydrogen bond contacts are between 2.7 and 3.7 Å.

The structure of Vch14 was examined for interactions involving specific serine and tyrosine hydrogen bonds to the peptide backbone and isoleucine residues residing in the hydrophobic core which provide stability to the protomer. Four residues were identified based on these criteria: Ser-16, Ser-12, Ser-79, and Ile-66, the only isoleucine residue internal to the protein structure. By studying the conserved residues between Vch14 family members (**Figure 2.7**), it was seen that Ser-12 and Ser-16 are not well conserved and are almost completely unique to Vch14 with the exception of HP6 (sequence identity of 84%). This is in contradiction to their direct involvement in oligomerisation (**Table 2.4**).

Internal to each Vch14 monomer lies a deep binding pocket lined by residues of helices α_1 and α_2 and the central four strands of the β-sheet (namely β_2-β_5) as depicted in **Figure 2.8**. In the crystal form, electron density consistent with a linear organic molecule (~15 Å in length) was observed in this site (**Figure 2.8 B**). The pocket is extensively lined with hydrophobic side chains (Ile-15, 40, 49, 66, and 82, Val-18, 29, and 63, **Figure 2.8 C**), a large portion of which are fully conserved for the Vch14 sequence relatives (**Figure 2.8 D**). At the entrance to the pocket, a distinct cluster of polar residues has been noted (Arg-21, Lys-59, His-25 and 77, Asp-60, Tyr-75, and Ser-74 (Sureshan *et al.*, 2013)).

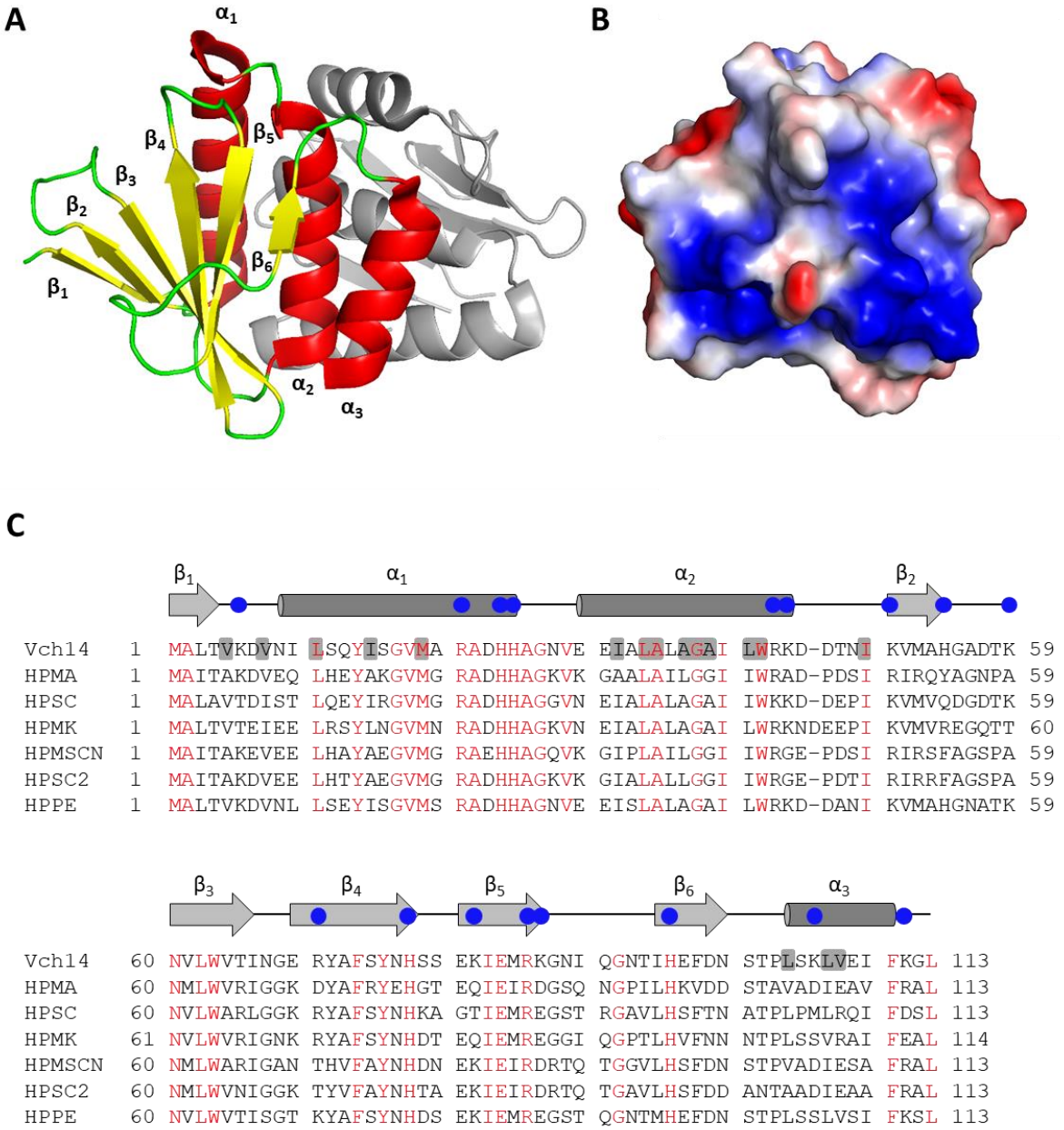


Figure 2.7 Crystal structure of Vch14 (PDB 3IMO (Sureshan *et al.*, 2013)). (**A**) Vch14 dimer coloured and labelled by secondary structure. (**B**) Solvent-facing β-sheet surface of component monomer coloured by charge (blue → positive, red → negative). (**C**) Multiple sequence alignment against Vch14 with its closest homologs (HPMA: *Methylocapsa aurea*, WP_036263122.1, 45% sequence identity; HPSC: *Sorangium cellulosum*, WP_012233093.1, 61% identity; HPMK: A1355_09625 *Methylomonas koyamae*, WP_082885599.1, 63% identity; HPMSCN: ABS59_03570 *Methylobacterium* sp. SCN 67-24, ODT55409.1, 42% identity; HPSC2: *Sphingobium chlorophenolicum*, WP_037454892.1, 45% identity; HPPE: *Pseudoalteromonas elyakovii*, WP_039493850.1, 84% identity). Fully conserved residues are in red with hydrophobics located at the interface boxed in grey. Positively charged residues are indicated (●).

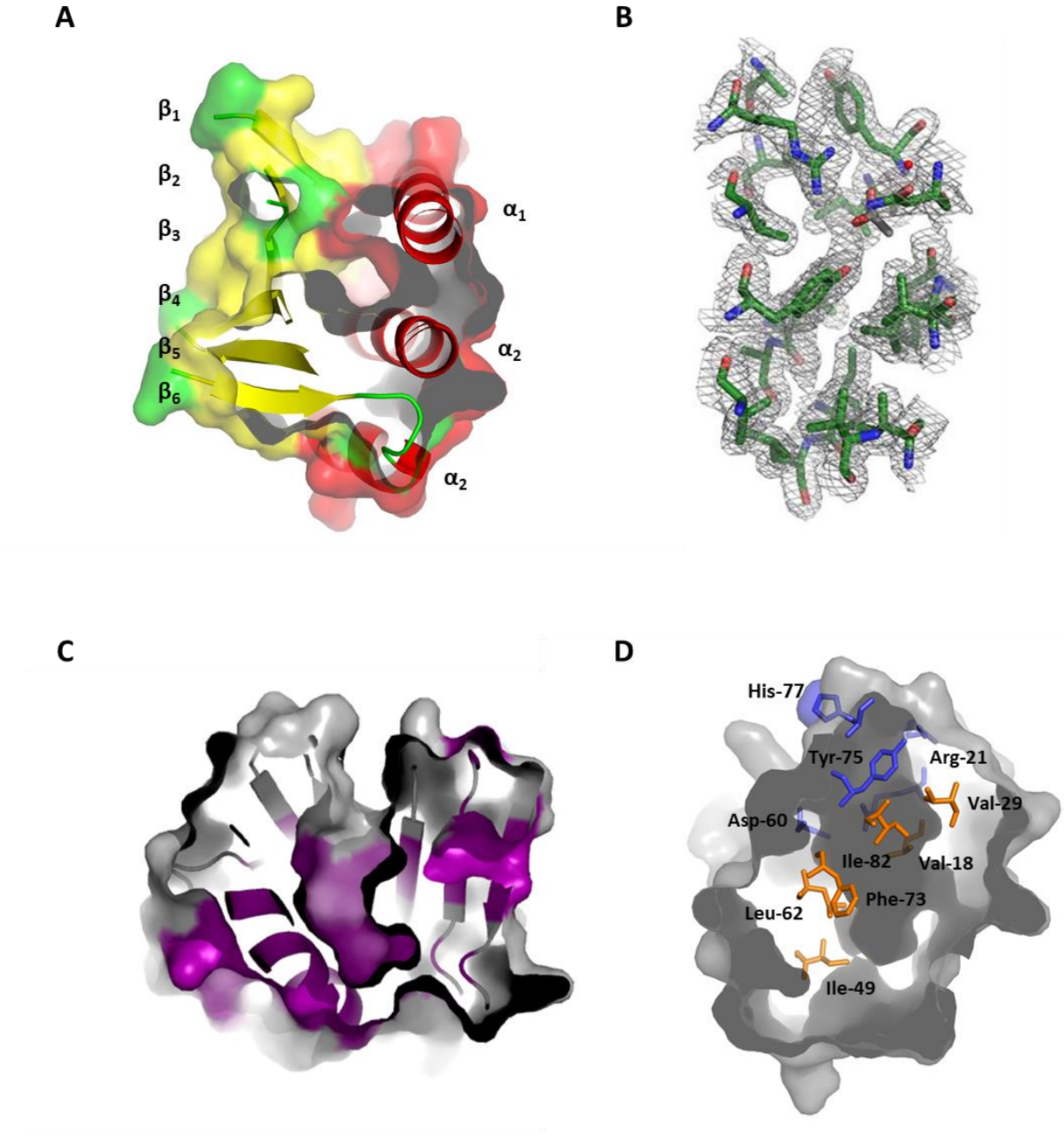


Figure 2.8 Visualisation of the binding cleft of Vch14. (**A**) Top view of binding pocket in ribbon representation with surface and coloured by secondary fold type. (**B**) Selected portion of the Vch14 1.8 Å $2F_o$-F_c map (contoured at a level of 1σ) showing residues that line the extended cavity and an area of electron density attributed to an unidentified small molecule (Deshpande, 2010). (**C**) Side view of the Vch14 binding cleft with hydrophobic residues coloured in purple. (**D**) Fully conserved residues surrounding the binding pocket. Hydrophobic residues are in orange, and polar in blue.

2.5.3. Experimental plan for Vch14

For Vch14, single residue mutants were designed to test stability, focusing mainly on hydrophobic residues, as they play a major role within the active site and the dimer interface. There was also interest in discovering the identity of the linear organic molecule found with X-ray diffraction by performing a compound screen which could also lead to the discovery of other classes of molecule that stabilise the structure of the protein. As stated in Chapter 1, there are cases where DSF will not detect changes to protein stability where the protein has large sections of exposed hydrophobic residues. So, although an attempt was made to measure changes to Vch14 melting temperatures with DSF, CD was also used, which does not rely on hydrophobic contacts.

CHAPTER 3

Solution Characterisation of Novel-fold Proteins

The focus of this research was to explore the solution state forms of the four novel proteins, Hfx1, Hfx5, Vch3, and Vch14, described in Chapter 2, with an emphasis on structure and stability. Recombinantly expressing each protein will allow me to provide the first biophysical measurements of these protein folds. Informed by sequence analysis performed in Chapter 2, chemical parameters of the four proteins being studied are listed in **Table 3.1**.

In this chapter the production of each protein has been outlined, after which quaternary state and stability in solution were evaluated using SEC, DSF, CD, and tryptophan fluorescence. After this initial evaluation, proteins that expressed well and were stable in solution were mutated on a single-residue basis using information acquired through structural analysis in Chapter 2. For Hfx1, these mutations targeted structurally important residues including specific cysteine, isoleucine, serine and tyrosine amino acids. In addition, histidine mutant forms were prepared as they were speculated to be important to ligand binding. Vch14 mutant forms only included changes to isoleucine and serine residues.

Table 3.1 Overview of physicochemical properties of proteins under study

protein (PDB ID)	# amino acids	Mw (Da)[a]	pI	GRAVY[b]	# cysteines	oligomer[c]	fold class
Hfx1 (3FUY)	178	20006.3	5.27	-0.511	3	trimer	α/β
Hfx5 (3IF4)	118	13229.8	6.54	-0.485	0	tetramer	α+β
Vch3 (3FY6)	138	16278.0	5.48	-0.846	0	dimer	α+β
Vch14 (3IMO)	133	14717.6	7.23	-0.336	0	dimer	α+β

[a] Including 20 residue His_6-tag added by pET15b vector.
[b] Grand average of hydropathy (GRAVY) value for protein sequences is defined by the sum of hydropathy values of all amino acids divided by the protein length.
[c] Protein oligomerisation as predicted by PISA and seen in the protein crystal structure (Krissinel & Henrick, 2007; Sureshan *et al.*, 2013).

3.1. Production of novel-fold proteins

Required genes were synthesised commercially as gBlocks, amplified using appropriate primers (Chapter 8, **Table 8.6**), and cloned into pET-15b vectors (**Figure 3.1 A**). This vector encodes an N-terminal His_6-tag chosen so as to replicate the protein samples generated in Toronto for crystal structure determination (Sureshan *et al.*, 2013). **Figure 3.1 B** shows the gel obtained after colony PCR to validate correct insertion prior to further amplification in Stellar competent *E. coli* cells. Vectors were then extracted from Stellar cells and transformed into the cell line BL21(DE3)pLysS *E. coli* for expression. In addition to original constructs, selected mutant constructs were also designed and produced for Hfx1 and Vch14 as discussed in Chapter 2 and 8.

Transformants were propagated within small-scale cultures, and these used to inoculate auto-induction media (500 ml) as described in Section 8.2.2.1. Cell pellets recovered following growth were lysed, and soluble protein products recovered by immobilised metal affinity chromatography (IMAC (Section 8.2.2)) with high imidazole and desalted as in Section 8.2.2. Protein product purity can be seen by SDS-PAGE (**Figure 3.2**). Each protein was able to be expressed with a high yield and purity. Extra bands on the Coomassie-stained gel are due to His-rich *E.coli* proteins, with one particular common contaminant (SlyD, 20.8 kDa) repeated across the gel for each recombinant target (Robichon *et al.*, 2011; Andersen *et al.*, 2013).

During the dialysis step into HEPES buffer (pH 7.5, 50 mM HEPES, 200 mM NaCl, 5% glycerol), Vch14 precipitated out of solution. As Vch14 was calculated to have a higher pI value (approx. 7.2) than the other proteins (**Table 3.1**), it was purified again and dialysed into glycine buffer (pH 9.0, 50 mM glycine, 200 mM NaCl, 5% glycerol). This resolved the solubility challenge for Vch14. After dialysis, protein concentrations were still high (Hfx1 = 35 mg/ml, Hfx5 = 13 mg/ml, Vch3 = 8 mg/ml, and Vch14 = 9 mg/ml) and have been expressed in mg/ml throughout Part I. This is the final concentration per 500 mL of culture after purification using IMAC and overnight dialysis.

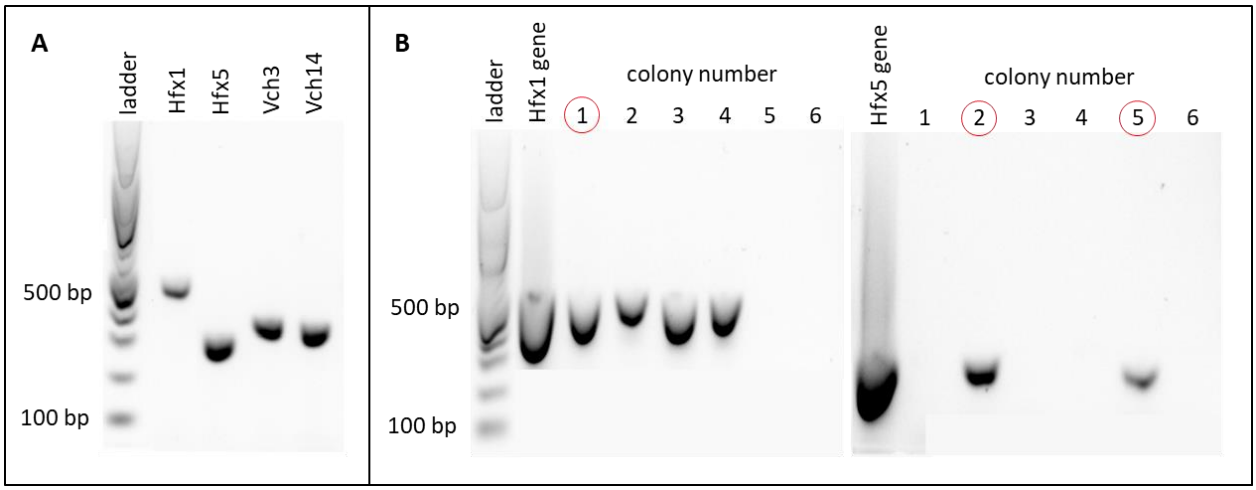


Figure 3.1 Agarose gel electrophoresis (1.2%) of PCR products. (**A**) Amplification of genes from gBlock DNA. (**B**) Colony PCR results for Hfx1 and Hfx5. Gene amplification was performed using T7 primers. Circled colony numbers were sent for sequencing.

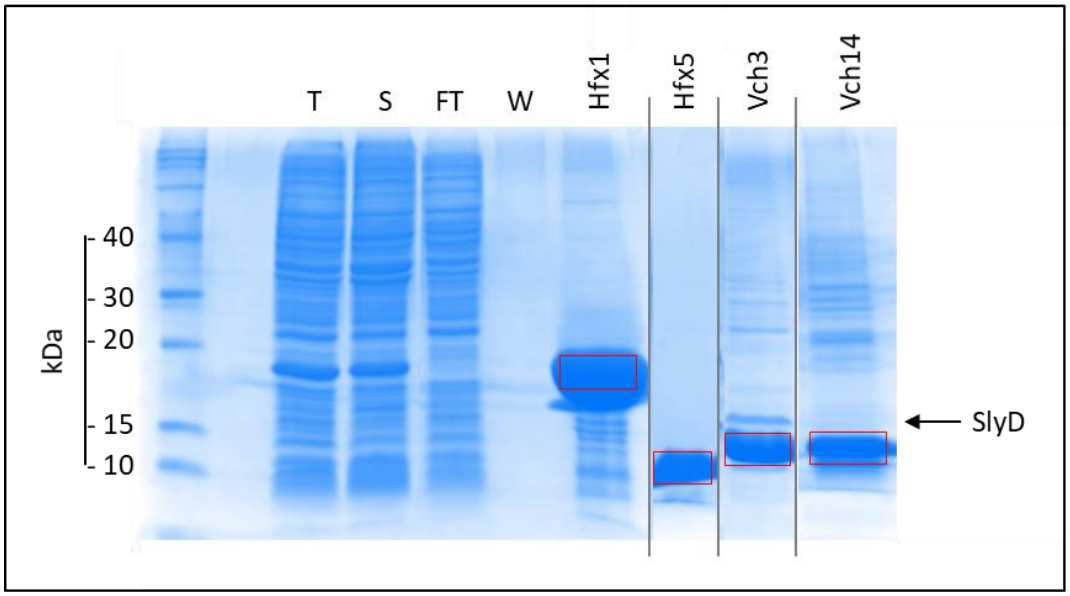


Figure 3.2 SDS-PAGE analysis of expressed novel-fold proteins. Total cell lysate (T), soluble fraction after centrifugation (S), IMAC column flow through (FT), column wash after 45 min (W), and pure protein were collected and run on a reducing SDS-PAGE gel (15% separating and 5% stacking gel) in tris/glycine buffer at 100 V (10 min) and then at 150 V (1 h). Target proteins are boxed in red. Contaminant *E. coli* proteins are indicated by an arrow.

3.2. Evaluation of quaternary state and stability in solution

3.2.1. Hfx1

The crystal structures determined for Vch3, Hfx5, Vch14, and Hfx1 revealed all to form oligomers as summarised in **Table 3.1** (Sureshan *et al.*, 2013). Analytical SEC was employed to determine the oligomeric state in solution for each novel-fold protein to confirm preservation of chain interactions in both solution and crystal forms.

Hfx1 eluted cleanly, with no discernible contamination presenting as extra peaks, as a single, symmetrical peak (K_{AV} 0.45), which corresponds to a native mass of 80 kDa (**Figure 3.3 B**). Given the protomer mass of Hfx1 is 20 kDa, the presumed trimer appeared to be eluting earlier than theoretically predicted (K_{AV} 0.48). Consideration of the crystal structure for Hfx1 (**Figure 3.3 A**) shows markedly anisotropic dimensions for the complex (flattened triangular form, 75 Å x 75 Å, 25 Å thickness) presumably creating an expanded Stokes radius for this species. It is therefore taken that Hfx1 is a trimer in solution, as defined in the crystal structure.

In addition to SEC, thermal melt curves were recorded, first to optimise stabilising buffers for each protein, and then to probe overall protein stability. Six biochemical buffers were tested in the range of pH 4.0 to pH 9.0, as well as with variation in salt concentration (50-300 mM NaCl). DSF responses for Hfx1 showed that in buffers at pH values > 4.0, samples displayed a smooth unfolding transition (**Figure 3.3 C**). Derivatives showed a slight preference for buffer pH values between 5.0 and 7.0, with a T_M value of 53 °C recorded. This is a mid-range value for a stable protein of 179 residues (Rees & Robertson, 2001).

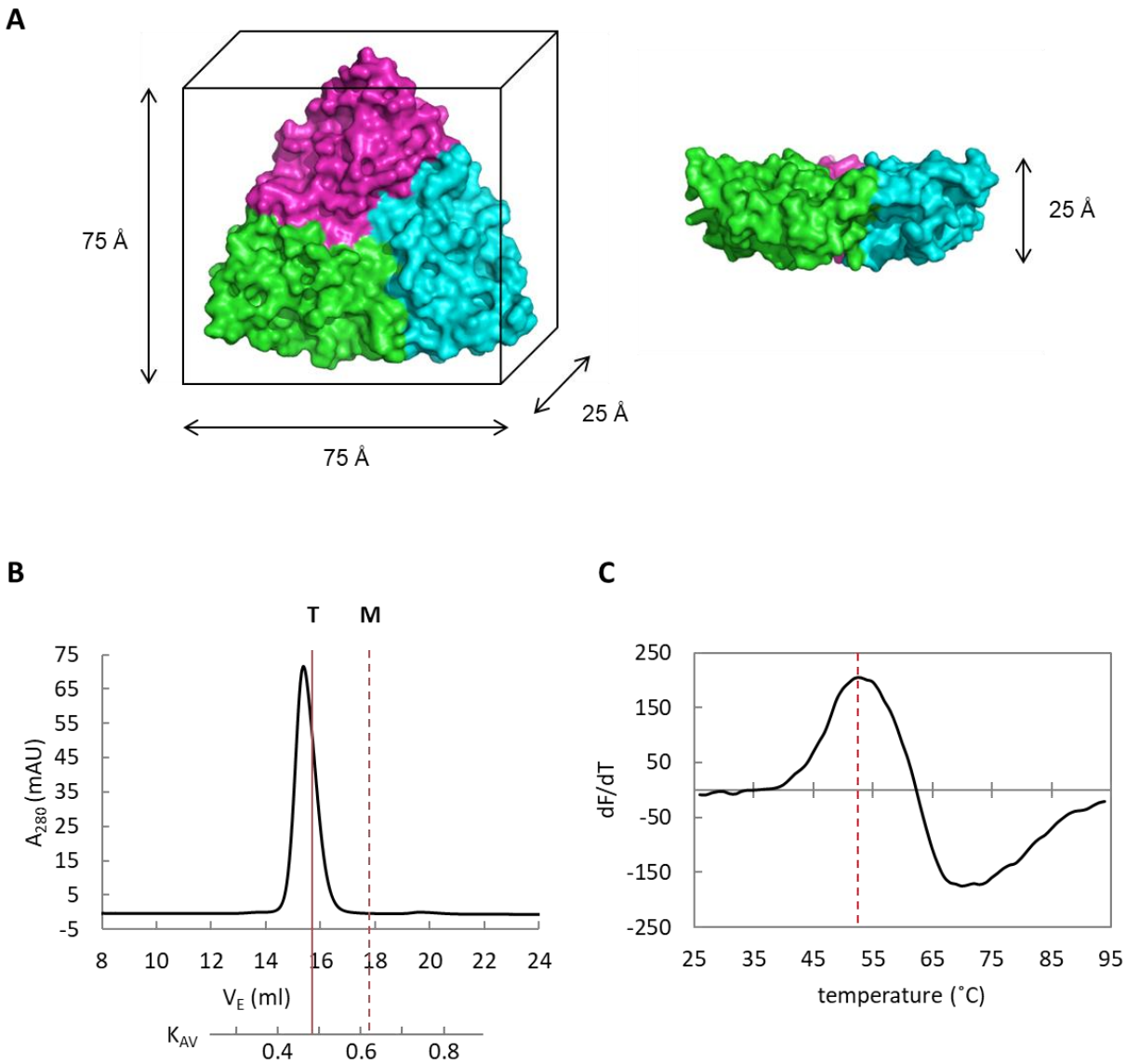


Figure 3.3 Preliminary analysis of Hfx1. (**A**) Hfx1 is 75 Å high and wide, with a thickness of 25 Å, defining a flattened triangular shape. (**B**) Analytical SEC (V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml in HEPES buffer, pH 7.5 with NaCl. Elution positions are indicated for the theoretical trimer (T) and monomer (M). (**C**) DSF 1st derivative responses (measured on an Mx3005P rt-qPCR machine, Strategene) for Hfx1 at pH 7.0 (0.5 M HEPES, 200 mM NaCl) with T_M indicated (dotted red line).

3.2.1.1. The structural role of cysteine residues

Each chain of Hfx1 contains three cysteine residues (Cys-44, Cys-75, and Cys-128), none of which appear surface exposed, and therefore lacking the potential to form disulfide bonds between subunits (**Figure 3.4 A**). In addition, they are all too far apart in the crystal structure to form disulfide bonds within each protomer. However, protein crystal structures are not always identical to those in solution. A paper comparing 109 protein crystal structures to NMR structures of nearly identical proteins found that: (1) hydrophobic amino acids are more similar in crystal and NMR structures than hydrophilic amino acids; (2) beta strands on average match better between NMR and crystal structures than helices and loops; (3) residues at the surface of globular proteins with good matching of their backbone (Cα atoms) have differently oriented side chains as their atomic positions are ill-defined given the intrinsic flexibility of molecular moieties protruding towards the solvent (Sikic *et al.*, 2010). As seen in **Figure 3.4** and previously in **Figure 2.2**, Cys-44 is part of helix 1, Cys-75 is within β_4 and Cys-128 is on the loop between α_2 and β_6. Therefore, it is possible that Cys-44 and Cys-128 are in more flexible regions than the crystal structure identifies. In addition, unpaired cysteines can play a role in redox functions, which can result in structural changes. For these reasons, cysteine chemistry was tested further. To test their propensity for disulfide bonding, SEC analysis was conducted with 1 mM TCEP in HEPES buffer (pH 7.5, 50 mM HEPES, 200 mM NaCl), which confirmed retention of a trimer species eluting at K_{AV} 0.45. This establishes that the Hfx1 trimer does not contain any inter-chain TCEP-accessible disulfide bridges.

To further probe the function of these residues, each cysteine was mutated separately to a serine as in Section 8.2.1.9, and all mutant forms expressed in equally high yields (~35 mg/ml). Fold integrity was then monitored by analytical SEC. Although C44S eluted as a trimer (K_{AV} 0.46), the other two mutants were seen to elute as monomers (K_{AV} 0.57, **Figure 3.4 B**). This change in oligomerisation of Hfx1 must have been in response to the Cys-Ser mutations, resulting in a structural deviation. Although cysteine and serine are structurally similar, they are not chemically similar. Cysteine is hydrophobic whereas serine is hydrophilic, and serine is capable of hydrogen bonding to the main chain. The C75S and C128S mutations, therefore, seem to have changed

the structure of Hfx1 enough to prevent oligomerisation. In order to see if the cysteine mutations had resulted in a rearrangement of secondary structure, or changes to overall percentages of α-helices and β-strands, a circular dichroism (CD) spectrum of each mutant was measured. There was, however, no noticeable change in structure between the native Hfx1 protein and the cysteine mutants.

To test this structural change more fully, analytical SEC of Hfx1 under low and high salt conditions (50 mM and 500 mM NaCl respectively) and in high acidity buffer (pH 5.0) was performed to see if electrostatic forces play a part in oligomerisation. It was thought that by mutating the cysteines, a change to hydrogen bonding had been made or the pI of a region of the protein had changed. Analytical SEC showed Hfx1 to elute as a trimer at pH 5.0 and high salt (K_{AV} 0.45), however, at low salt Hfx1 eluted as a hexamer (K_{AV} 0.37, 145 kDa). This hexameric species can be attributed to the salting in effect, whereby a protein is less soluble at low salt concentrations and therefore more prone to aggregation. As the salt concentration increases (in this case from 50 mM to 300 mM NaCl), the protein becomes more soluble and transitions from a hexamer to a trimer (Arakawa & Timasheff, 1984).

Thermal melt curves were also measured using DSF for each cysteine mutant to attain its relative stability to native Hfx1 (**Figure 3.4 C**). Whereas C44S showed an increase in T_M of +8 ˚C (61 ˚C), C75S was unchanged, and C128S had a ΔT_M of -6 ˚C (47 ˚C). All curves looked characteristic of a single unfolding event and did not vary depending on whether the Hfx1 variant was seen as a monomer or a trimer. This shows that monomeric forms of Hfx1 are no less stable than trimeric Hfx1 species.

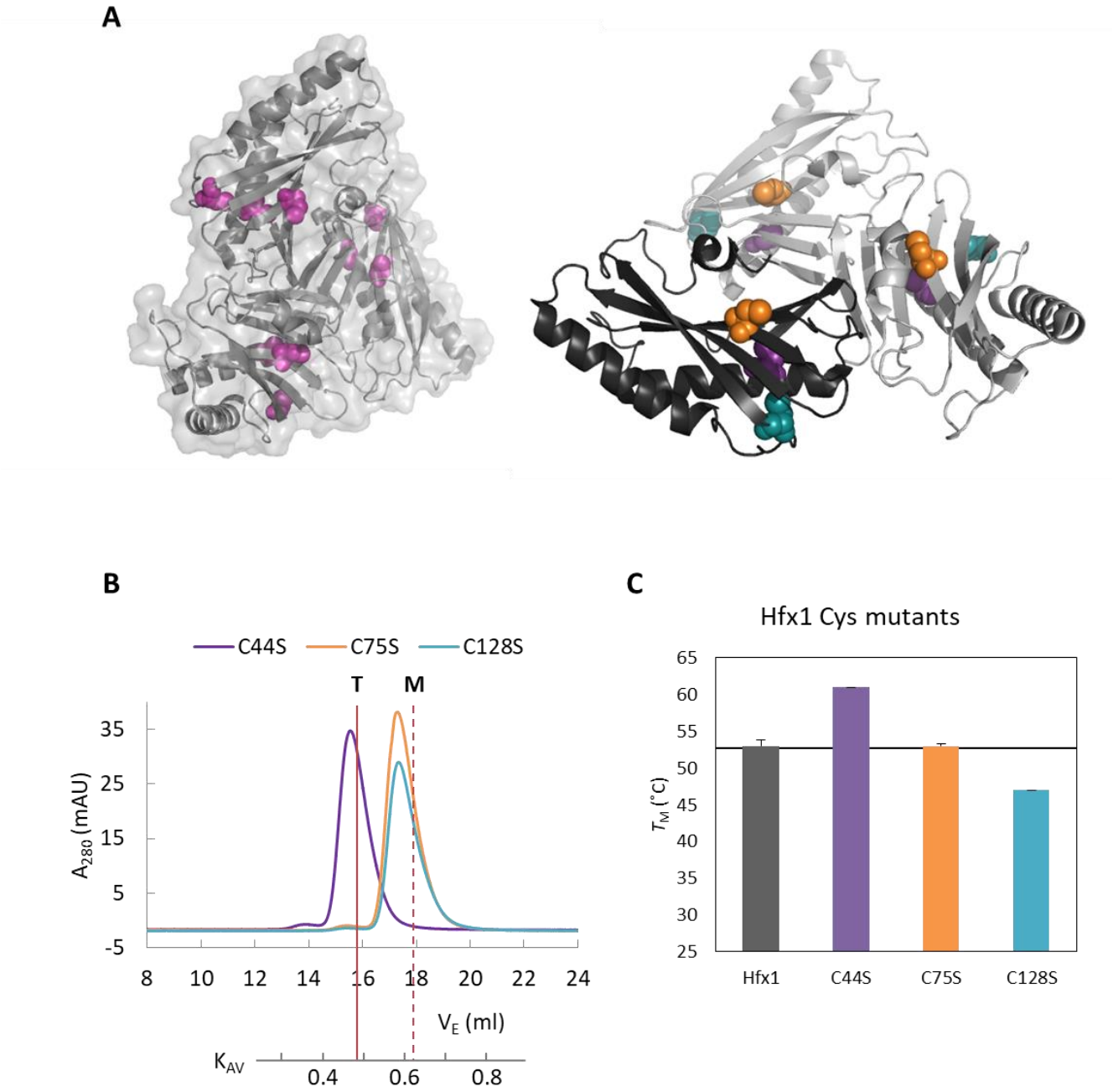


Figure 3.4 Cysteine mutants of Hfx1. (**A**) Trimeric structure of Hfx1 in ribbon form with a surface (left) and each cysteine residue represented as spheres and colour coded (Cys-44 purple, Cys-75 orange, and Cys-128 cyan, right). (**B**) Analytical SEC (V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml in HEPES buffer, pH 7.5 with NaCl. Elution positions are indicated for the theoretical trimer (T) and monomer (M). (**C**) Melting temperatures for Hfx1 and each mutant measured using DSF. Error bars represent SEM (n = 3).

3.2.1.2. Serine and tyrosine mutant forms

As previously stated in Chapter 2, serine and tyrosine residues that form hydrogen bonds with the backbone of Hfx1 were chosen for mutation to probe their structural stabilising bonds (Ser-10 hydrogen bonds form as follows: $O_\gamma \rightarrow N$ position 12 and $O_\gamma \rightarrow N$ position 11; Ser-57: $O_\gamma \rightarrow O$ position 53; Ser-99: $O \rightarrow N$, $O_\gamma \rightarrow N$, and $N \rightarrow O_{\varepsilon 1}$ position 102; Ser-143: $O_\gamma \rightarrow O_\gamma$ and $O_\gamma \rightarrow N$ position 145; and Tyr-154: $O_\eta \rightarrow O$ position 73 and $O_\eta \rightarrow O$ position 24, **Figure 3.5 A**). All substitutions were made with residue size and shape in mind. **Figure 3.5 C** shows the structure of each type of amino acid mutated within Hfx1 (and subsequently Vch14) and its relevant replacement to demonstrate that there should be little steric hindrance to neighbouring residues within the protein.

During preparation of these mutant forms, the S57A mutant strain of XL1-Blue supercompetent cells did not yield colonies, suggesting unsuccessful introduction of the mutated DNA into the host cell. However, all other serine mutants expressed well (in the range of 30-60 mg/ml after affinity chromatography) except for S99A which had a relatively low expression of 8 mg/ml. SEC traces confirmed that each mutant eluted cleanly as a trimer (K_{AV} 0.44). S99A showed a small shoulder to the right of the main peak which is most likely due to its low expression and instability resulting in partially unfolded versions of protein eluting later. DSF thermal melts showed S99A to be largely degraded, displaying no characteristic melting curve, and with a very low fluorescent response. Ser-99 is found at the N-terminal end of helix α_3 on the outside edge of the Hfx1 trimer (**Figure 3.5 A**). It is part of a β-α-β motif commonly seen in the 4-OT superfamily of enzymes (Poelarends *et al.*, 2008). The serine residue at the top of the α-helix is seen in most members of this family, suggesting that it is important to protein folding and stability. It is possible that the change of serine to alanine led to partial folding of the protein or even misfolding by disrupting the protein folding pathway. This could be down to a direct impact on stabilisation of the α_3 helix or it could be more complicated. As Ser-99 has no contact with other chains of Hfx1, the change in quaternary structure must be directly related to breaking the hydrogen bond between this residue and the peptide backbone (position 102).

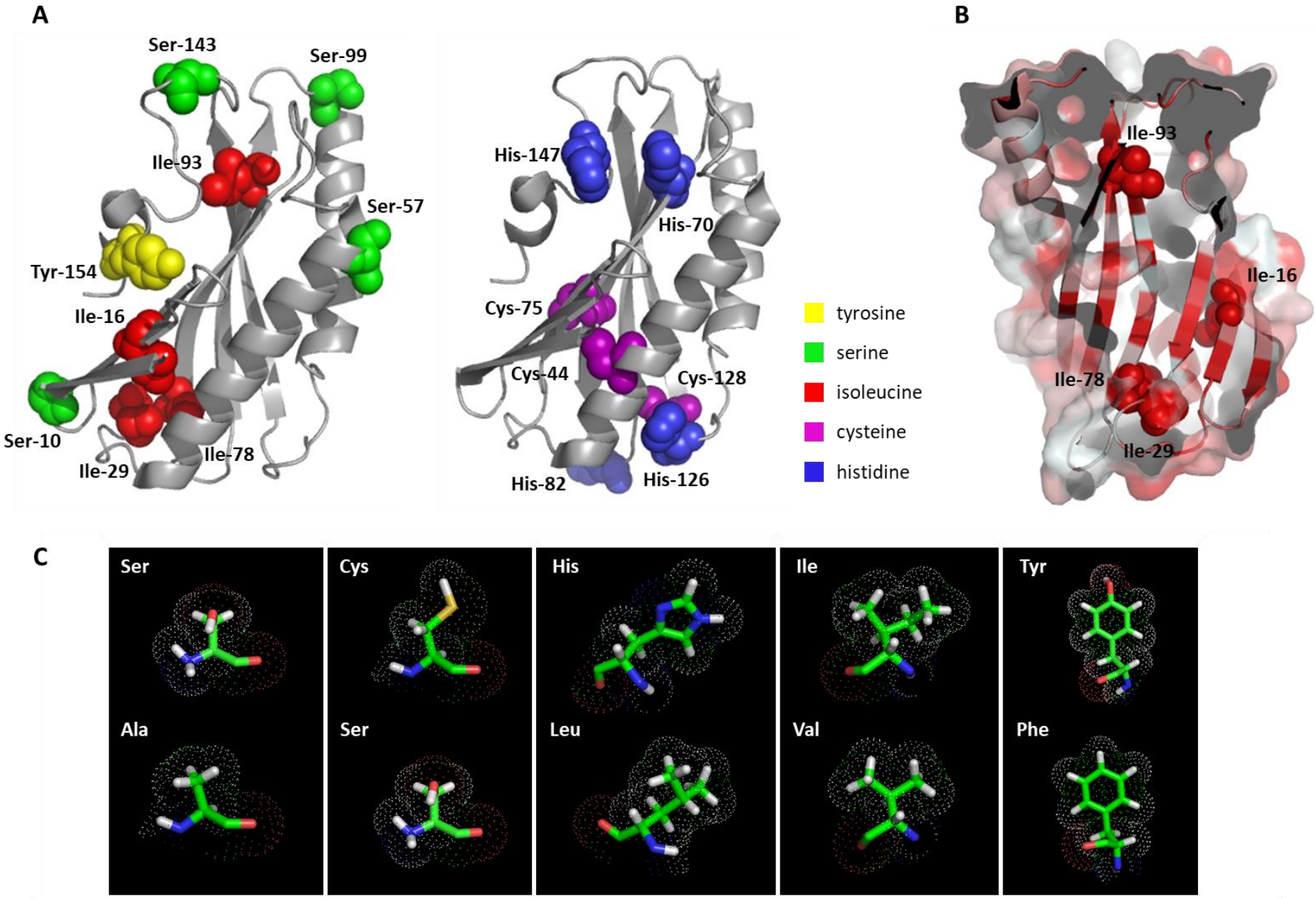

A
Ser-143
Ser-99
Ile-93
Ser-57
Tyr-154
Ile-16
Ser-10
Ile-29
Ile-78
His-147
His-70
Cys-75
Cys-44
Cys-128
His-82
His-126
tyrosine
serine
isoleucine
cysteine
histidine
B
Ile-93
Ile-16
Ile-78
Ile-29
C
Ser
Cys
His
Ile
Tyr
Ala
Ser
Leu
Val
Phe

Figure 3.5 Hfx1 target residues for mutation. (**A**) Ribbon representation with targeted mutant residues labelled. Each residue is colour coded according to type. (**B**) Hfx1 coloured by hydrophobicity (red → hydrophobic, white → hydrophilic) with isoleucines targeted for mutation shown as spheres and labelled. (**C**) Amino acid substitutions are shown in stick representation with dots.

Thermal melt curves were measured for each mutant and native Hfx1 to determine if there was a change in overall protein stability, affecting T_M. Although changing single serine and tyrosine residues had little effect on quaternary structure, it did have a noticeable effect on protein stability. Melting temperatures of all mutants were different than native Hfx1 with a ΔT_M between +3 ˚C and -4 ˚C (T_M 56-49 ˚C, **Figure 3.6 A**).

Mutant forms were also analysed using CD to determine if a change in secondary structure was responsible for the changes in overall stability seen with DSF. However, CD spectra showed no change in secondary structure between mutants and this therefore cannot be attributed to changes in melting temperature.

3.2.1.3. Isoleucine mutant forms

Out of the nine isoleucines that Hfx1 contains, only four (Ile-16, 29, 78, and 93) are completely inaccessible to solvent, placing their position in the hydrophobic core (**Figure 3.5 B**). However, only two of those mutants were successfully produced as soluble protein (I78V and I93V) suggesting that Ile-16 and Ile-29 are essential to protein structure. As with the serine and tyrosine mutants, all eluted as trimers during analytical SEC (K_{AV} 0.44). Each mutant was also measured using DSF and CD, with T_M decreasing 2-4 ˚C (T_M 49-51 ˚C), and with no noticeable changes in the secondary structure.

A

mutation	location	ΔT_M (˚C)	SEM	mutation	location	ΔT_M (˚C)	SEM
hydrogen bonds				cysteines			
Y154F	α_3	+3	0	C44S	middle α_1	+8	0
S143A	β_6-α_3 loop	-3	0	C75S	middle β_4	0	0.33
S99A	start of α_2	-4	0	C128S	α_2-β_6 loop	-6	0
S10A	end of β_1	-4	0.33	active site			
hydrophobic core				H70L	start of β_4	-2	0.33
I78V	end of β_4	-2	0.33	H82L	β_4-β_5 loop	-3	0.33
I93V	β_5	-4	0.58	H126L	α_2-β_6 loop	-5	0
				H147L	β_6-α_3 loop	-14	0.33

B

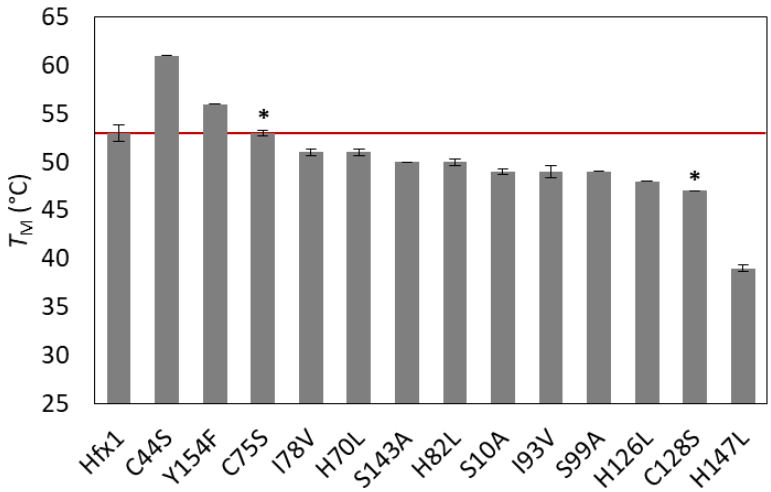


Figure 3.6 Hfx1 mutant properties attained using DSF. (**A**) Table of selected mutants, their locations in the structure of Hfx1, and the change in T_M from the native protein (Hfx1 T_M 53 ˚C) with calculated SEM (n = 3). (**B**) Bar graph of the T_M of Hfx1 and each mutant. Monomeric constructs are starred. SEM error bars for each measurement are shown.

3.2.1.4. Histidine mutant forms

The structure of Hfx1 contains four histidine residues in every protomer (His-70, His-82, His-126, and His-147) that surround the area hypothesised to be the ligand binding pocket (**Figure 3.5 A**). In order to determine their ligand-binding potential and to see if these histidine residues also play a role in structural stability, each one was mutated to a leucine (**Table 8.7**). All mutants were produced successfully as recombinant protein, with each one yielding over 20 mg/ml after affinity chromatography. However, there was a variation in stability, with three of the mutants decreasing protein T_M by between 2 and 5 °C (changes in T_M greater than 2 °C are considered significant), and H147L being the least stable mutant with a ΔT_M of -14 ˚C (39 ˚C, **Figure 3.6**).

Despite these changes in protein stability, and the role of His-82 in oligomerisation, the mutations had no effect on oligomerisation of Hfx1, with all eluting from the analytical SEC column as trimers (K_{AV} 0.46). For His-82, this demonstrates that although the salt bridge helps to stabilise the trimeric interface, it is not the major contributor.

3.2.2. Hfx5

Purified samples of Hfx5 were subjected to biophysical investigation in solution under the same buffer conditions as Hfx1. Hfx5 was the only novel fold protein to be seen as two oligomeric species in solution (K_{AV} 0.57 and 0.64) with elution peaks corresponding to a size of 30 and 15 kDa, respectively (**Figure 3.7 A**). As the tagged form of Hfx5 has a monomeric molecular mass of 13 kDa, these are consistent with a mixture of monomer and dimer. These observations confirm that Hfx5 is a true domain-swapping protein (Mascarenhas & Gosavi, 2017).

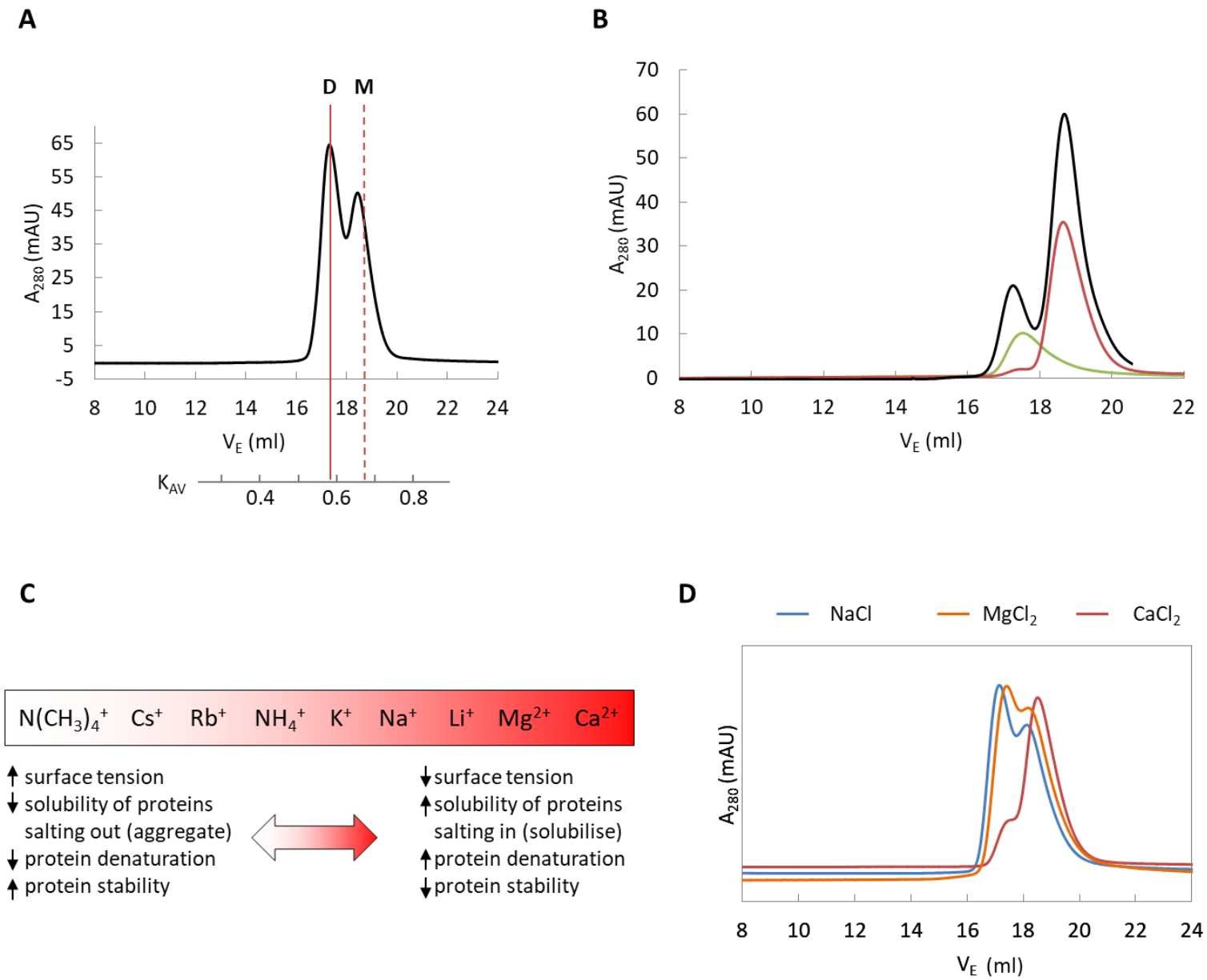

A
D
M
65
55
45
35
25
15
5
-5
A_{280} (mAU)
8
10
12
14
16
18
20
22
24
V_E (ml)
K_{AV}
0.4
0.6
0.8
B
70
60
50
40
30
20
10
0
A_{280} (mAU)
8
10
12
14
16
18
20
22
V_E (ml)
C
$N(CH_3)_4^+$
Cs^+
Rb^+
NH_4^+
K^+
Na^+
Li^+
Mg^{2+}
Ca^{2+}
↑ surface tension
↓ solubility of proteins
salting out (aggregate)
↓ protein denaturation
↑ protein stability
↓ surface tension
↑ solubility of proteins
salting in (solubilise)
↑ protein denaturation
↓ protein stability
D
NaCl
$MgCl_2$
$CaCl_2$
A_{280} (mAU)
8
10
12
14
16
18
20
22
24
V_E (ml)

Figure 3.7 Hfx5 oligomerisation states. (**A**) Analytical SEC trace (HEPES buffer, pH 7.5 with NaCl, V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml. Elution positions are indicated for the theoretical dimer (D) and monomer (M). (**B**) Three SEC traces (120 ml Hiload 16/60 Superdex 200 prep grade column) at 1 ml/min of Hfx5. The monomer (red) and dimer (green) were separated from mixed species of purified protein (black) and rerun. (**C**) Hofmeister series. (**D**) Analytical SEC for Hfx5 dialysed and run in HEPES buffer (pH 7.5) containing 200 mM NaCl, $MgCl_2$, and $CaCl_2$.

To determine if the dispersal of species is concentration dependent, the monomer and dimer were separated from each other using preparative SEC. It can be seen in **Figure 3.7 B** that after separation there was little inter-conversion between species, with only a small shoulder seen for the monomer, and a broadening of the dimer peak, which could also be attributed to the separation method. However, it must be noted that between analytical and preparative SEC, the ratios of monomer to dimer changed. In **Figure 3.7 A**, there is slightly more dimer than monomer, however in panel **B**, there is almost twice as much monomer than dimer. As both columns use Superdex 200 media, the only difference in conditions between the two columns would be concentration of the protein. Because the preparative grade column is five times the volume of the analytical grade one, the concentration of Hfx5 would be less on the larger column. Using this logic, Hfx5 favours the monomeric state when it is in a state of low concentration.

In addition to these tests, buffers were made using different chloride salts from different positions on the Hofmeister series (**Figure 3.7 C**) to see if this would affect protein oligomerisation. The species to the left of K^+ are referred to as kosmotropes, while those to its right are called chaotropes. These terms originally referred to a cation's ability to alter the hydrogen bonding network of water (Collins & Washabaugh, 1985). The kosmotropes, which were believed to be 'water structure makers', are strongly hydrated and have stabilising and salting-out effects on proteins and macromolecules. On the other hand, chaotropes ('water structure breakers') are known to destabilise folded proteins and give rise to salting-in behaviour (Vlachy *et al.*, 2009).

Four variations on HEPES buffer (pH 7.5, 50 mM HEPES, 5% glycerol) were prepared, containing 200 mM KCl, NaCl, $CaCl_2$, and $MgCl_2$. Hfx5 was purified, spilt into four aliquots, and then dialysed into each buffer overnight. Initial observations after dialysis revealed a small amount of protein

precipitation in the $MgCl_2$ and $CaCl_2$ buffers, however, all protein had completely aggregated in the KCl buffer. This shows that the change in salt cation even from sodium to potassium can have a significant effect on protein solubility. The oligomeric states of the other three Hfx5 samples were compared using analytical SEC (**Figure 3.7 D**). As the salts move to the right along the Hofmeister series, Hfx5 shifts in population from more dimer to more monomer, with the $CaCl_2$ buffer at approximately the same ratio as seen when preparative SEC was performed. These results suggest that the shift in dimer to monomer is based on protein solubility. So, when Hfx5 is present at lower concentrations, it must be more soluble.

In order to see if there was a difference in stability between the two species of Hfx5, each was probed separately using DSF and CD to measure melting temperature. However, as can be seen in **Figure 3.8**, there was not a significant difference between the two, with both the monomer and dimer displaying approximately the same T_M of 62 ˚C.

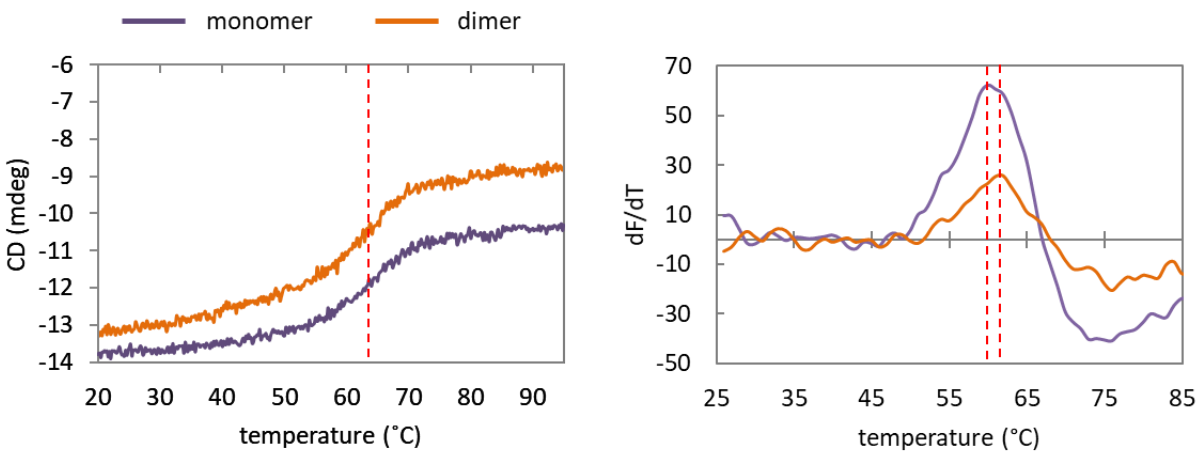


Figure 3.8 CD and DSF melting curves of Hfx5 monomer and dimer. Left: CD melting curves at 215 nm in potassium phosphate buffer (pH 7.8, 50 mM potassium phosphate, 50 mM NaF) with approximate T_M indicated (dotted red line). Right: DSF 1st derivative melt curve of Hfx5 monomer and dimer in HEPES buffer (pH 7.5, 50 mM HEPES, 200 mM NaCl) with T_M indicated.

As discussed in Section 2.2, sequence similarity searches revealed that Hfx5 is the first structure defining a small family of six proteins of unknown function. From sequence alignments it is evident that the inter-module linker (between β_4 and α_2) is highly conserved across the group of proteins (Sureshan, 2012). This highly conserved inter-module linker displays some elevated flexibility. Given the conserved features and the experimental results discussed above (i.e. unchanged T_M between the two species), it is most likely that the Hfx5 monomer engages the same interface as the dimer, but from a single chain, bending at the inter-module linker. This arrangement is commonly seen in other domain-swapping proteins (Rousseau *et al.*, 2012). Given this proposition, there could be a slight change in structure between the species. To investigate this, CD and tryptophan fluorescence were used to detect changes to secondary structure.

The circular dichroism spectra of the monomer and dimer seen in **Figure 3.9 A** shows a noticeable difference in secondary fold. Between the two species there is a change in signal around 230 nm. This region is known to correlate to the occurrence of pi stacking between aromatic residues (phenylalanine, tyrosine, tryptophan, (Andrushchenko *et al.*, 2006)). Depending on the orientation of these residues, the polarised signal can be either positive or negative (Woody, 1994). It is clear from the CD spectra that the monomer displays a change in aromatic pi stacking, which is consistent with a structural rearrangement. As each protomer of Hfx5 contains seven aromatic residues (**Figure 3.9 B**), with the possibility of Trp-53/Trp-58 and Tyr-28/Tyr-30 pi stacking given even a slight structural rearrangement. This could be the case in the change from the Hfx5 dimer to the monomer.

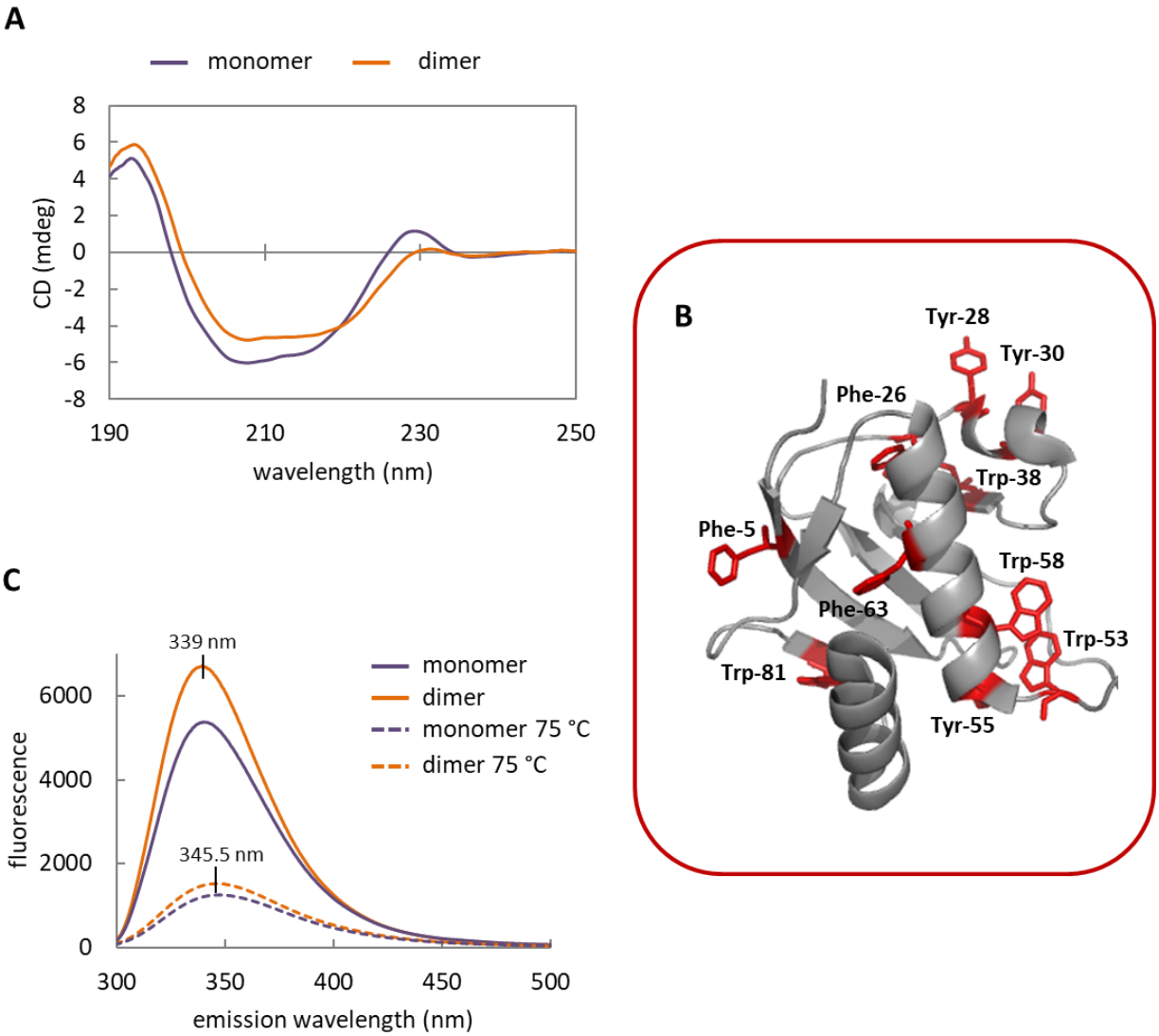


Figure 3.9 Secondary structural analysis of Hfx5 monomer and dimer. (**A**) CD spectra of both forms of Hfx5 (0.1 mg/ml) measured in potassium phosphate buffer (pH 7.8, 50 mM potassium phosphate, 50 mM NaF). (**B**) Locations of aromatic residues in the proposed monomeric structure of Hfx5. (**C**) Tryptophan fluorescence at 280 nm for monomer and dimer (50 µM) at room temperature (25 °C) and 75 °C. Peak apexes are labelled.

In order to investigate the pi stacking further, tryptophan fluorescence was used to see if there was a change in tryptophan positioning between the monomer and dimer of Hfx5. Samples were measured at room temperature (25 °C) and at 10 °C past the melting temperature (75 °C, **Figure 3.9 C**). Experiments were designed this way because if there was an emission difference at room temperature, it should dissipate once the protein is denatured, as the amino acid sequence is the same. However, no detectable shift in emission wavelength was seen, corresponding to no change in surface tryptophan positioning. With these results and those from CD, there seems to be a slight rearrangement in structure between the Hfx5 monomer and dimer, but not enough to produce a wavelength shift using tryptophan fluorescence or to affect melting temperature.

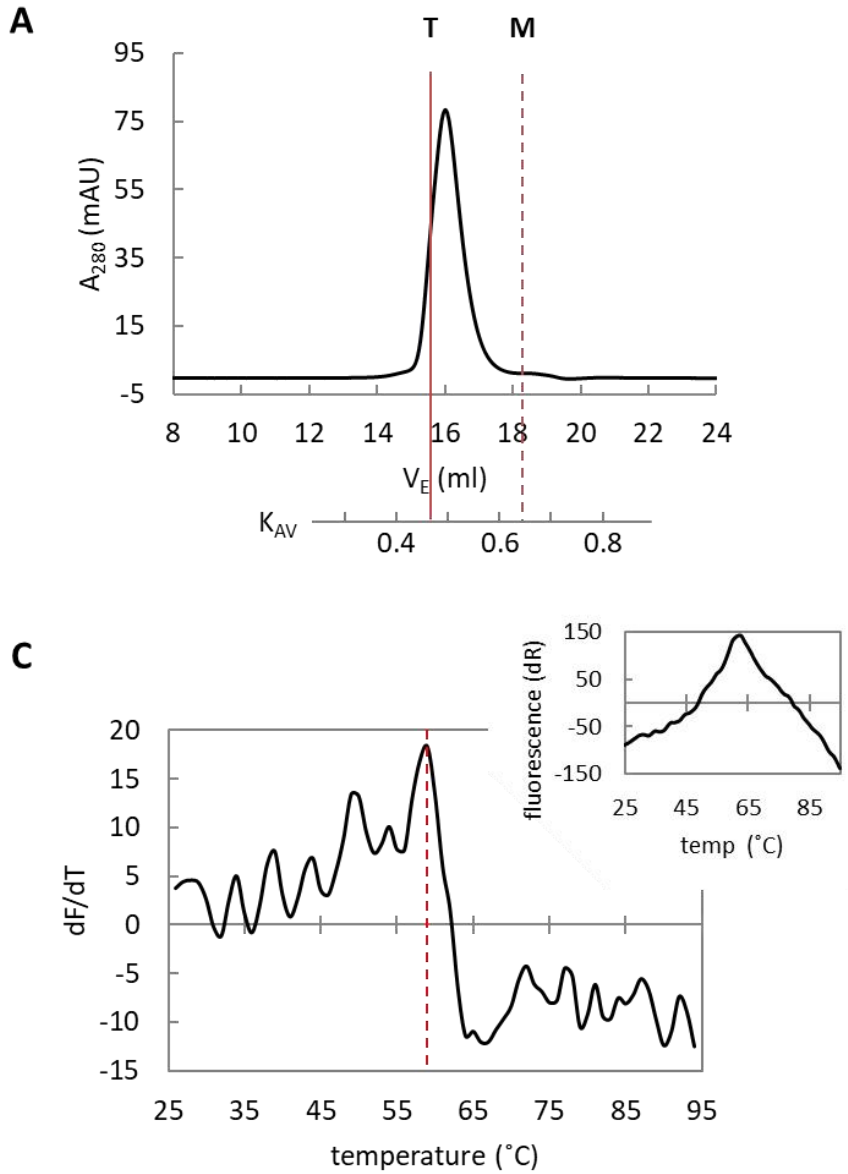
3.2.3. Vch3

Vch3 eluted cleanly as a single species (K_{AV} 0.49), corresponding to a mass of approximately 60 kDa (**Figure 3.10 A**). As a monomer of Vch3 is 16.3 kDa, this can be matched approximately with a tetramer in solution. As discussed in Section 2.3, there is a tetrameric interface within the crystal lattice (helix α_4 and the β_1'-β_2' loop of adjacent dimers, **Figure 3.10 B**), which must be favourable in solution given the chromatography results.

Thermal melt curves for Vch3 were far from optimal with a relatively low fluorescence. Vch3 is seen to have the same T_M in any buffer from pH 5.0 to 9.0 (T_M 54 ˚C, **Figure 3.10 C**). Poor quality of the melt curves is most likely due to a low final concentration of protein in the assay, which in turn can be attributed to low stability and the propensity for aggregation.

3.2.4. Vch14

Purified samples of Vch14 were analysed with SEC to establish the native mass in HEPES buffer containing salt. Vch14 eluted cleanly as a dimer (K_{AV} 0.56) corresponding to a native mass of 30 kDa (**Figure 3.11 A**).


A
T
M
95
75
55
35
15
-5
A_{280} (mAU)
8
10
12
14
16
18
20
22
24
V_E (ml)
K_{AV}
0.4
0.6
0.8
C
20
15
10
5
0
-5
-10
-15
dF/dT
25
35
45
55
65
75
85
95
temperature (˚C)
150
50
-50
-150
fluorescence (dR)
25
45
65
85
temp (˚C)

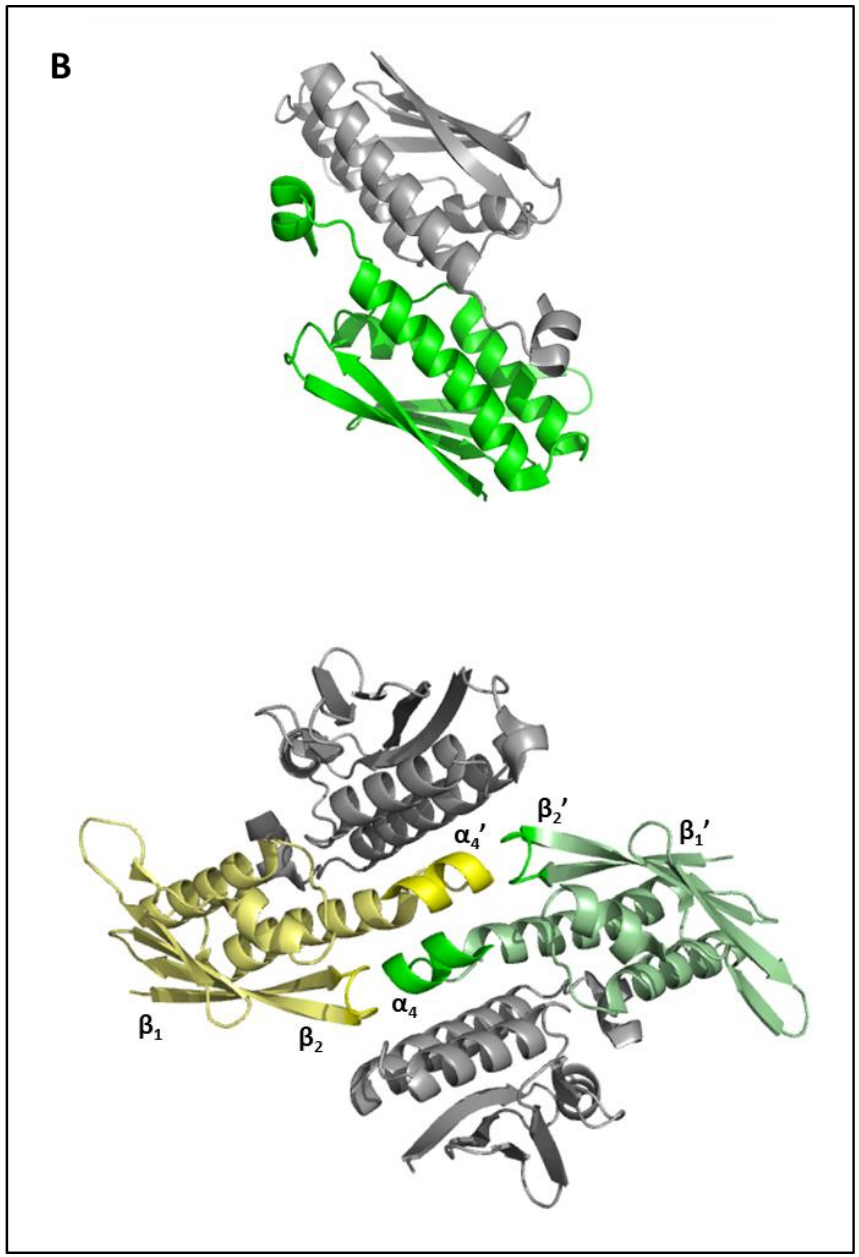

B
β_2'
β_1'
α_4'
α_4
β_1
β_2

Figure 3.10 Preliminary analysis of Vch3. (**A**) Analytical SEC trace (HEPES buffer, pH 7.5 with NaCl, V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml. Elution positions are indicated for the theoretical tetramer (T) and monomer (M). (**B**) Top: Vch3 dimer with chain A and C coloured in grey and green respectively. Bottom: Vch3 tetramer with highlighted tetrameric interfaces between chain D (yellow) and chain C (green). All residues labelled with ['] are from chain C. (**C**) Raw and 1st derivative DSF response curves (measured on an Mx3005P rt-qPCR machine, Strategene) for Vch3 at pH 8.0 (50 mM HEPES, 300 mM NaCl) with T_M indicated (dotted red line).

As was seen during dialysis, Vch14 is most stable in buffers of pH 9.0, which prevent precipitation of the protein. The T_M of the protein was not able to be confirmed by DSF due to the hydrophobic nature of Vch14. Since it has a large, open, hydrophobic pocket, it is not a suitable candidate for T_M determination using DSF (Niesen *et al.*, 2007). As can be seen in **Figure 3.11 B**, Sypro Orange immediately bound to the residues in this pocket, leading to a high initial fluorescence. An alternative method which could have been used avoids the use of environmentally sensitive dyes and instead uses a sulfhydro specific probe (7-diethylamino-3-(40-maleimidylphenyl)-4-methylcoumarin), which only fluoresces when bonded to free SH-groups that are exposed during protein unfolding (Alexandrov *et al.*, 2008). While useful in many instances, the method is limited to proteins with buried SH groups, and is sensitive to the presence of reducing agents such as dithiothreitol (DTT) and b-mercaptoethanol (bME). For this reason, a thermal melt was performed using CD.

CD measures the T_M through the loss of secondary structure alone and does not require a reporting molecule. Based on the CD spectrum of Vch14 at set temperatures (intervals of 5 ˚C from 20-95 ˚C, **Figure 3.11 C**), the CD signal was monitored at 208 nm at a ramp of 1 ˚C/min in order to determine the melting temperature. As can be seen in **Figure 3.11 D**, the T_M of Vch14 is approximately 63 ˚C.

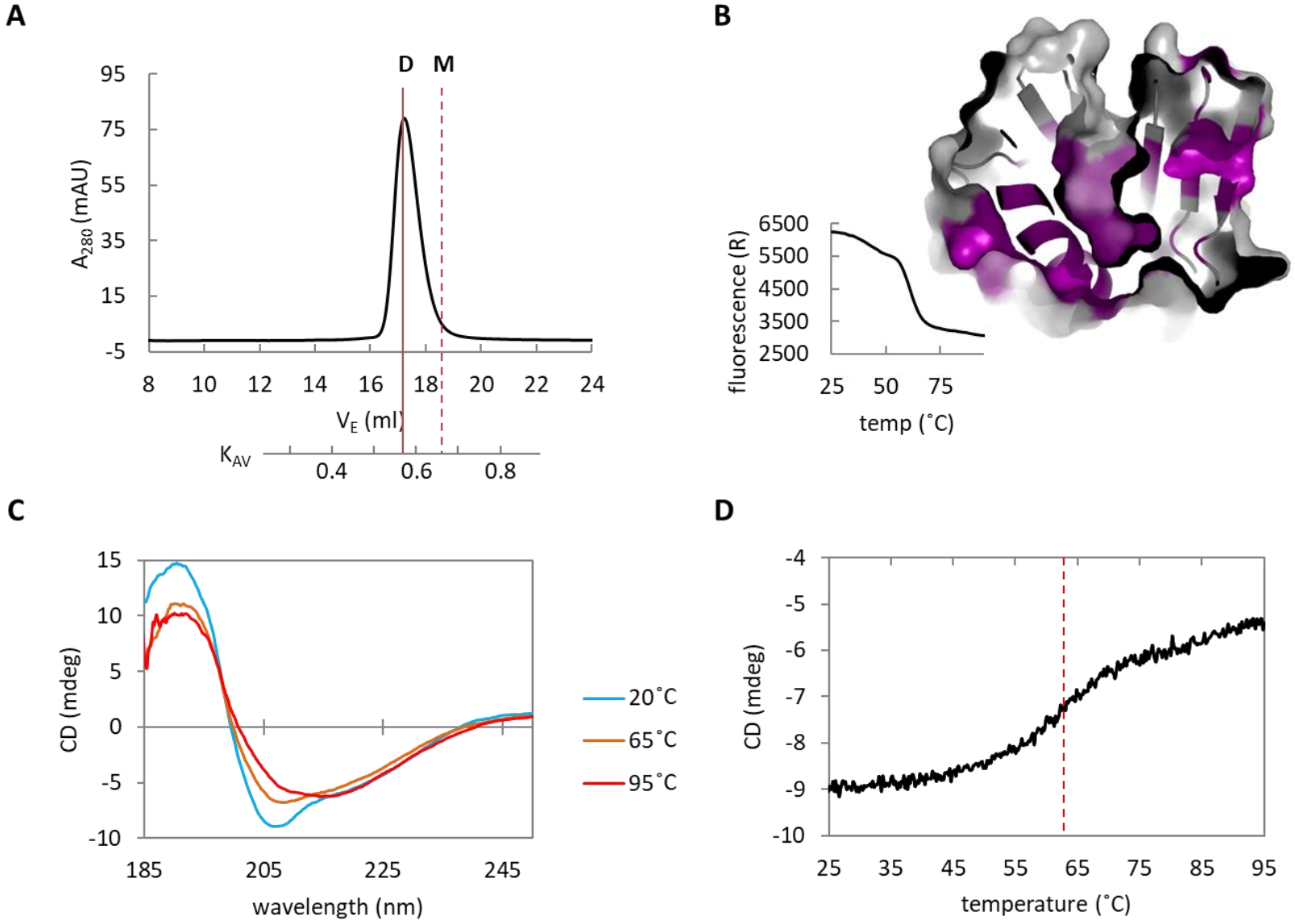
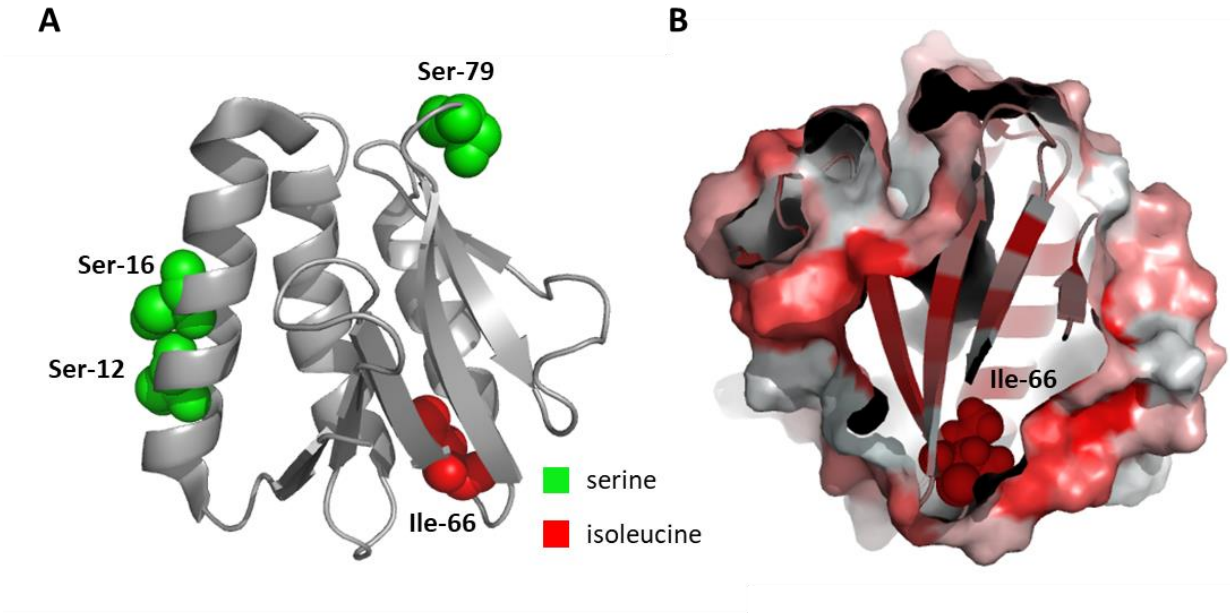

A
D
M
95
75
55
35
15
-5
A_{280} (mAU)
8
10
12
14
16
18
20
22
24
V_E (ml)
K_{AV}
0.4
0.6
0.8
B
6500
5500
4500
3500
2500
fluorescence (R)
25
50
75
temp (˚C)
C
15
10
5
0
-5
-10
CD (mdeg)
185
205
225
245
wavelength (nm)
20˚C
65˚C
95˚C
D
-4
-5
-6
-7
-8
-9
-10
CD (mdeg)
25
35
45
55
65
75
85
95
temperature (˚C)

Figure 3.11 Preliminary analysis of Vch14. (**A**) Analytical SEC (V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml in glycine buffer, pH 9.0 with NaCl. Elution positions are indicated for the theoretical dimer (D) and monomer (M). (**B**) Uncorrected DSF thermal melt for Vch14, and side-on view of Vch14 structure where hydrophobic residues are highlighted in purple. (**C**) CD spectrum of Vch14 (0.1 mg/ml) in sodium borate buffer (pH 9.0, 50 mM sodium borate, 50 mM NaF) at three different temperature points, 20, 65, and 95 ˚C. (**D**) Thermal melt curve of Vch14 (208 nm).

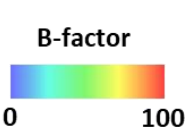
3.2.4.1. Serine and isoleucine mutant forms

Four residues were chosen within Vch14 for mutation (**Figure 3.12 A**). Three serine residues form hydrogen bonds as follows: Ser-16, N→O and O_γ→O position 12; Ser-12, O_γ→O and N→O position 8; and Ser-79, N→$O_{\delta 1}$ and O_γ→O position 76. In addition to these serine residues, one solvent inaccessible isoleucine residue was also chosen for mutation, Ile-66. All Vch14 mutants expressed as soluble proteins with slight variations in expression levels (3-15 mg/ml) and all were present in solution as dimers (K_{AV} 0.56). Overall, it seems that the change in these selected residues did not disrupt the hydrogen bonds and hydrophobic interactions enough to affect the overall quaternary structure of Vch14.

As there was no change in quaternary structure, CD was used to determine if there was a change in secondary structure or melting temperature between the native Vch14 protein and its mutants. The results from CD analysis revealed there was little change in structure between native Vch14 and each mutant, however, there was a noticeable change in melting temperature between mutant constructs (**Figure 3.12**). As with native Vch14, melting curves were measured at 208 nm, as this was the region of greatest change in secondary structure. Although there was little stability change for S12A and S16A, there was a large decrease in stability for S79A and I66V (ΔT_M of -7 to -10 ˚C). It can be seen in **Figure 3.12** that although the interface helices of the Vch14 dimer are highly ordered, the loops connecting β-strands are less so. The location of both Ser-79 and Ile-66 are located on these loops whereas Ser-12 and Ser-16 are found on the interfacing helices.



C



mutation	location	ΔT_M (˚C)
Hydrogen bonds		
S16A	α_1	+1
S12A	α_1	+1
S79A	β_4-β_5 loop	-7
Hydrophobic core		
I66V	end of β_3	-10

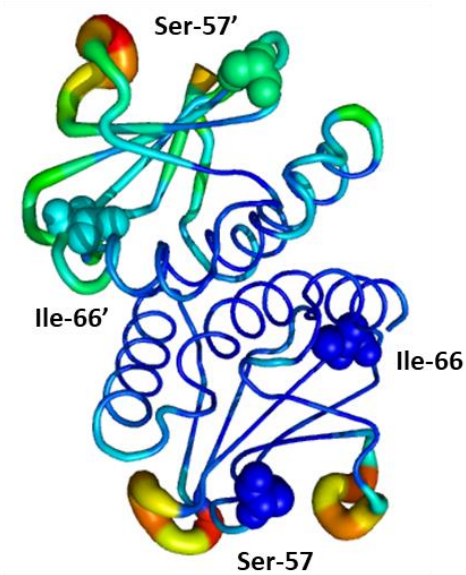


Figure 3.12 Vch14 mutant forms. (**A**) Ribbon representation of Vch14 monomer with residues targeted for mutation as spheres. (**B**) Vch14 coloured by hydrophobicity (red → hydrophobic, white → hydrophilic) with Ile-66 coloured red as spheres. (**C**) Table of Vch14 mutants, their positions within the protein structure, and their changes in T_M. Right: Positions of the least stable mutants, S79A and I66V, located on a B-factor representation of the Vch14 dimer.

It is seen that mutation of residues within loop regions destabilises the protein more than those within well-defined tertiary structures. In addition, it is notable that Ser-79 and Ile-66 are found surrounding the proposed active site of Vch14 and that Ile-66 (the only internal isoleucine residue) is almost fully conserved across Vch14 homologs. The mutation of these residues may have led to a destabilisation of the hydrophobic pocket, which could lead to a lower melting temperature.

3.3. Summary

Out of the four novel-fold proteins focused on in this work, three seem to be very robust, displaying high melting temperatures, and maintained structural stability under mutational stress. Although Vch3 was not as amenable to recombinant production as the other three proteins, it was interesting to see that it forms a tetramer in solution (**Figure 3.13**). Although this form of oligomerisation was seen in the crystal structure, it was only thought to be an artefact of crystal packing with no strong contributing forces to a viable tetrameric interface. However, it is possible that the tetramer would be stabilised through an interaction between helix α_4 and the β_1'-β_2' loop of adjacent dimers, as this was the interaction seen in the crystal lattice, or that Vch3 oligomerises as a different tetrameric structure in solution, for which the arrangement is not known.

The remainder of proteins studied oligomerised as expected, even in the case of Hfx5, seen as both a monomer and dimer in solution (**Figure 3.13**). As Hfx5 is a domain-swapped dimer, it is expected that it would form both quaternary structures, but until now, it was only ever seen as a dimer in the crystal structure. Given the flexibility and conservation of its inter-module linker, the Hfx5 monomer would use the flexibility of the linker to bend back on itself and use the same electrostatic contacts as the dimer, which is seen in other domain-swapping proteins (Rousseau *et al.*, 2012). The structure of subunits within domain-swapped oligomers is identical to that of the monomers, with the exception of the inter-domain linker. Both the monomeric and dimeric forms of Hfx5 were analysed for rearrangement using tryptophan fluorescence and CD. However, the only difference between the two species was that the monomer of Hfx5 displayed a peak at 230 nm on a CD spectrum that the dimer did not. Other studies show that this peak is associated with pi-pi stacking, which alludes to a slight change in structure between the dimer and monomer to

allow for an increase in pi stacking in the monomer (Andrushchenko *et al.*, 2006). However, this change in structure cannot be too pronounced, as there was no shift in tryptophan fluorescence between the two species.

In order to probe the nature of the different oligomers further, Hfx5 was exposed to high and low salt environments as well as a range of salts positioned at different intervals on the Hofmeister series. Results showed that salt concentration (between 50 and 300 mM NaCl) did not shift the populations of monomer and dimer, but protein concentration and a change in salt elemental composition did. When Hfx5 was present at low concentrations (during preparative SEC), the monomer to dimer ratio would shift from about 40:60 (as seen during analytical SEC) to 70:30. A similar shift from dimer to monomer was also seen in buffers incorporating salts holding positions further to right of the Hofmeister series (decreased surface tension and protein stability, with increased solubility and denaturation). Both results point to the conclusion that Hfx5 prefers to be in monomeric form when solubility is high (low protein concentration or appropriate salts) but becomes more dimeric as solubility decreases.

The protein Vch14 was seen as a dimer in solution with a relatively stable melting temperature of 63 °C. It is unique amongst the proteins studied, as it is highly hydrophobic, and therefore not amenable to measurement using DSF. For this reason, the melting temperature of the dimeric native protein and its mutant forms was measured using CD. Four mutant forms of Vch14 were created and expressed as protein products, three Ser→Ala mutations to remove hydrogen bonds with the backbone, and one Ile→Val mutant to disrupt the hydrophobic core. The two serine mutations, located within α_1, had no effect on protein stability or oligomerisation, however, S79A and I66V decreased protein stability by 7-10 °C. Both of these residues are located in a more flexible region of the protein, resulting in a greater effect on protein stability when they are disrupted. In addition, Ile-66 is more highly conserved within the homologous Vch14 family, and most likely more integral to protein stability and function.

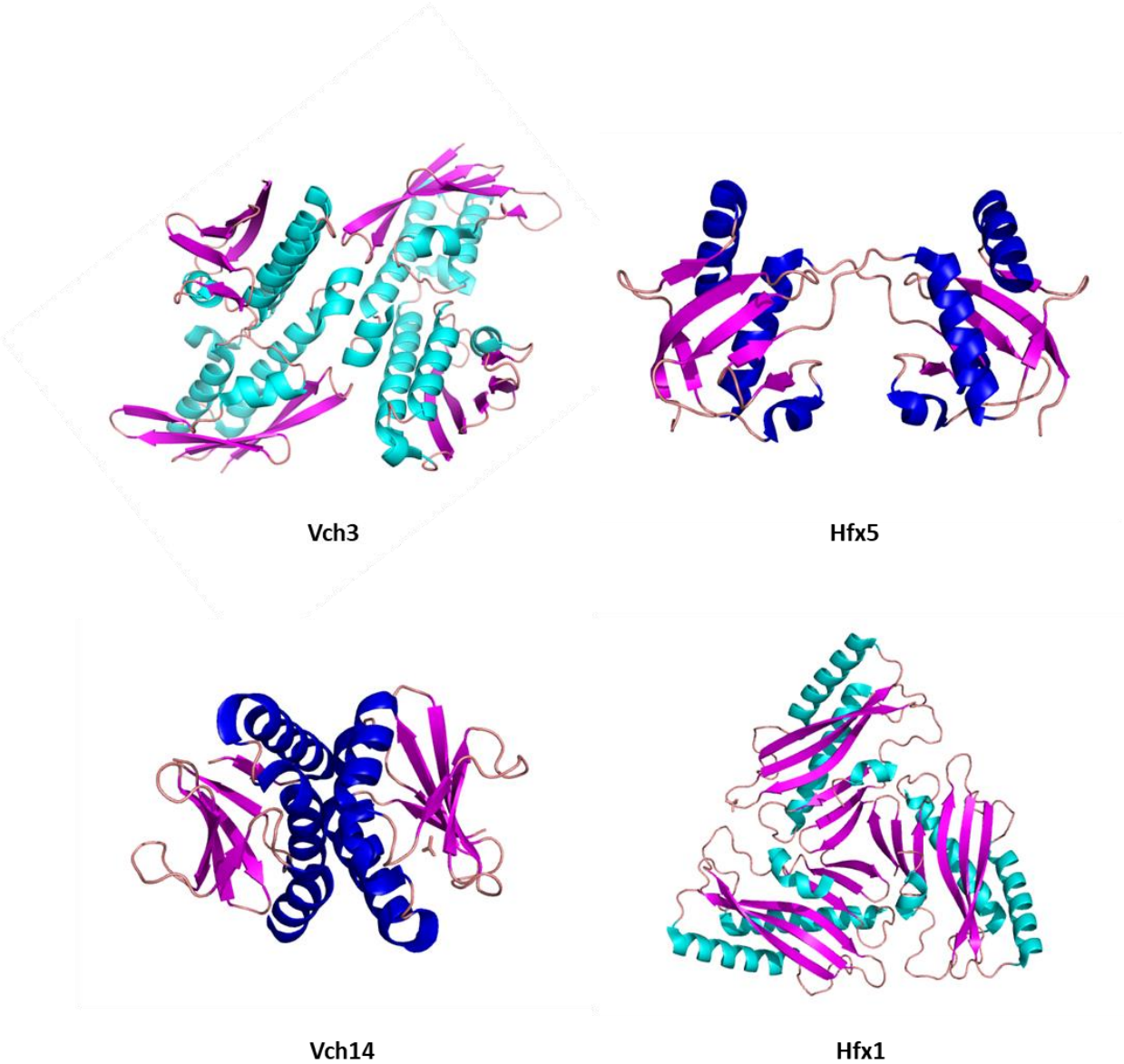


Figure 3.13 Crystal structures of the four novel fold proteins characterised in this work. Structures are coloured by secondary folds with helices in blue or cyan, β-sheets in purple, and loops in pink.

Hfx1 was the most amenable to experimentation as it was very stable and expressed in high yields of around 40 mg/ml after affinity chromatography. Hfx1 was further analysed for structural stability through the design and implementation of single-site mutagenesis. Overall, residues were targeted for mutation based on: I) stabilising bonds formed with the protein backbone, II) maintenance of the stability of the hydrophobic core, III) presumption of roles in active-site chemistry, and IV) pure curiosity as to their function within the protein structure. Based on these reasons, 13 mutant forms of Hfx1 were expressed recombinantly, including mutations in tyrosine, serine, isoleucine, cysteine, and histidine residues. Out of all mutant forms expressed as protein products, H147L was the least stable. In **Figure 3.14**, each residue targeted for mutation has been located on the structure of Hfx1 to see if there is a trend between stability and location within the protein. Coloured according to ΔT_M (green = increase, yellow to red = decrease), there is not a strong trend, but it does seem that residues located within the structure of the protomer have less effect on stability than those on the periphery. Most mutations did not perturb the quaternary structure of Hfx1 except for the change of Cys-75 and Cys-128 to serine residues. These two mutant forms resulted in monomeric protein species in solution. The mechanism of this change in structure is still unknown.

The next step in this research is to identify possible binding partners for each protein. Given results from work in this section, the constructs most amenable to these binding assays are Hfx1 and Hfx5. Each of these proteins has been shown to oligomerise in solution and, determined by their thermal melt curves, all are intact and folded, and therefore capable of being functional. In the case of Hfx1, the wild type and some mutant forms will be compared to detect losses in binding capacity.

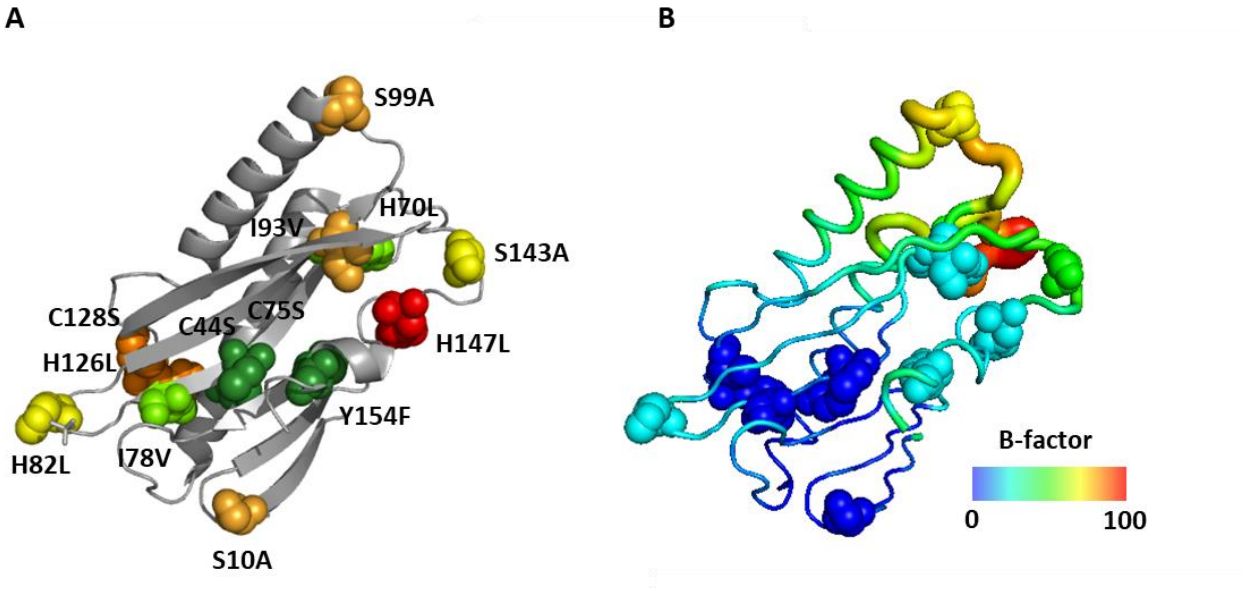


Figure 3.14 Positions of target mutant residues within an Hfx1 protomer. (**A**) Residues are coloured according to effect on melting temperature (green = no change or increase, red = decrease). (**B**) Residues located on b-factor representation of Hfx1 and coloured by flexibility (red = highly flexible, blue = not flexible).

CHAPTER 4

Binding Studies of Novel Fold Proteins

Three-dimensional protein structures that are intrinsically linked to cellular integrity or biochemical function are highly conserved over evolutionary time. Even when sequence similarities are no longer detectable, it can be possible to infer a functional classification based on homologous relationships between protein tertiary structures (Brenner, 2001; Rigden, 2006; Bateman *et al.*, 2010). It has been determined that 66% of proteins having a similar fold also have a similar function (Koppensteiner *et al.*, 2000). Knowledge of tertiary structure therefore provides an important means of inference of molecular function in several different ways. But how do you infer function when a protein has a novel fold, and the organism from which is hales is unknown? In this Chapter Hfx1 and Hfx5 were earmarked for functional characterisation.

Proteins can take on a number of roles within organisms, including transportation across the cell membrane, membrane makeup, protein-protein interactions, small molecule transportation, and enzymatic functions. It was determined that a good starting point for functional characterisation of Hfx1 and Hfx5 was small molecule screening due to the availability of screens containing chemically diverse ligands.

Ligand binding has the capacity to stabilise protein structure in a manner dependent on concentration and binding affinity. Therefore, ligands can be identified and characterised by their effect on protein stability elicited by binding (Schon *et al.*, 2013). For example, the shift in protein denaturation temperature (T_M shift), measured using fluorescence, has become a popular approach to identify potential ligands. The technique used to measure this shift was DSF coupled with the addition of the Silver Bullets/Silver Bullets Bio Screens provided by Hampton Research (Silver Bullets Bio conditions will be denoted by an asterisk, Section 8.2.3.3). These screens contain a chemically diverse portfolio of small molecules selected for their ability to establish stabilising, intermolecular, hydrogen bonding, hydrophobic and electrostatic interactions which have the potential to promote protein stability (McPherson & Cudney, 2006; Larson *et al.*, 2007a;

Larson *et al.*, 2007b). These screens were selected for their ease of use, as they come in a 96-well format, making them suitable for use with the RT-PCR machine required for DSF.

4.1. Screening for potential effectors of novel-fold proteins

4.1.1. Screening of potential effectors for Hfx1

Previous work by Dr. V. Sureshan (Sureshan *et al.*, 2013) discovered that the inter-subunit binding cleft of Hfx1 does not immediately match any known binding sites of previously characterised protein structures. However, a search against a database of non-redundant cognate binding sites using Isocleft (Najmanovich *et al.*, 2008) identified some shared features between the Hfx1 cleft and the binding sites of several enzymes implicated in binding mono- or di-nucleotide cofactors (e.g. NADP(H), myristoyl coA, 3-hydroxy 3-methylglutaryl coA, S-adenosylmethionine, ADP, coA, etc).

In order to interrogate these findings, a commercial screen of chemical cocktails including small molecules ranging from amino acids and peptides, to metals and salts, and nucleotides and carbohydrates (**Table S.1 and S.2**) was combined with Hfx1 and changes to stability (T_M of 53 °C) were monitored using DSF (**Table 4.1**). As the small molecule formulations are different between the screening plates used, they provided a diverse range of molecules for testing. Each protein was combined with a hydrophobic fluorescent dye and the mixture transferred to a 96-well plate containing the small molecule cocktails. The plate was heated over a range of 25-95 ˚C in a real-time qPCR machine and the change of fluorescence intensity monitored by the instrument. Derivatives of the curves produced were then used to calculate the transition midpoint which corresponds to the melting temperature (Section 8.2.3.3).

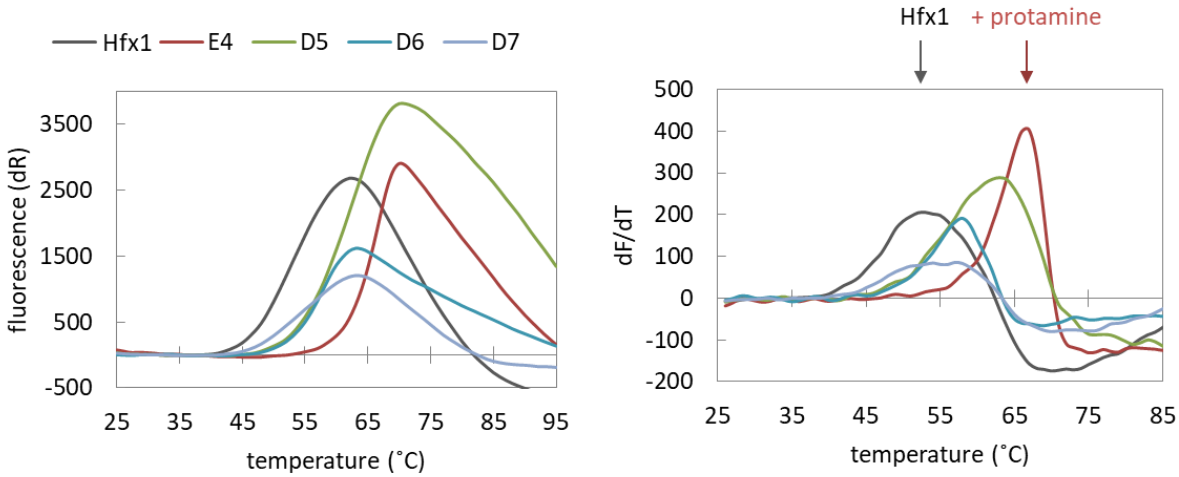
Table 4.1 DSF ligand screening results for Hfx1

ΔT_M (°C)	SEM (+/-)	cocktail ingredients	well number [a]
+14	0.33	protamine sulfate	E4
+10	0.33	calcium chloride, magnesium chloride, manganese(II) chloride, zinc chloride	D5
+7	0.33	protamine sulfate, g-strophanthin, benzamidine, D-fructose 1,6-diphosphate, oxamic acid	*A11
+5	0.88	nickel(II) chloride, copper(II) chloride, cobalt(II) chloride, molybdenum(III) chloride	*D11
	1.33	sodium 1-pentanesulfonate, 3,5-dinitrosalicylic acid, 3-aminosalicylic acid, salicylamide	A8
	0	cadmium chloride, cobalt(II) chloride, copper(II) chloride, nickel(II) chloride	D6
	1.20	3,5-dinitrosalicylic acid, 4-aminobenzoic acid, benzamidine hydrochloride, hexamminecobalt(III) chloride, mellitic acid	H3
+4	1.33	3,5-dinitrosalicylic acid, 3-indolebutyric acid, naphthalene-1,3,6-trisulfonic acid, trans-1,2-cyclohexanedicarboxylic acid	D7
	2.19	tacsimate pH 7.0	G11

[a] Well numbers with asterisks are from the Silver Bullets Bio screen. All others are from the Silver Bullets screen.

A selection of thermal responses is provided in panel **A** of **Figure 4.1**. The melting temperature, T_M, is more accurately determined by the derivative of these curves (Simeonov, 2013), shown on the right. Out of the 192 conditions tested, nine wells increased the T_M of Hfx1 by ≥ 4 °C (E4, D5, *A11, *D11, A8, D6, H3, D7, and G11). For example, wells containing protamine sulfate were seen to increase the stability of Hfx1 by as much as 14 ˚C (T_M 67 ˚C). Other stabilising conditions included Group II and transition metals, 3,5-dinitrosalicylic acid and tacsimate (**Table 4.1**).

A


Hfx1
E4
D5
D6
D7
fluorescence (dR)
3500
2500
1500
500
-500
25 35 45 55 65 75 85 95
temperature (˚C)
Hfx1
+ protamine
dF/dT
500
400
300
200
100
0
-100
-200
25 35 45 55 65 75 85
temperature (˚C)

B

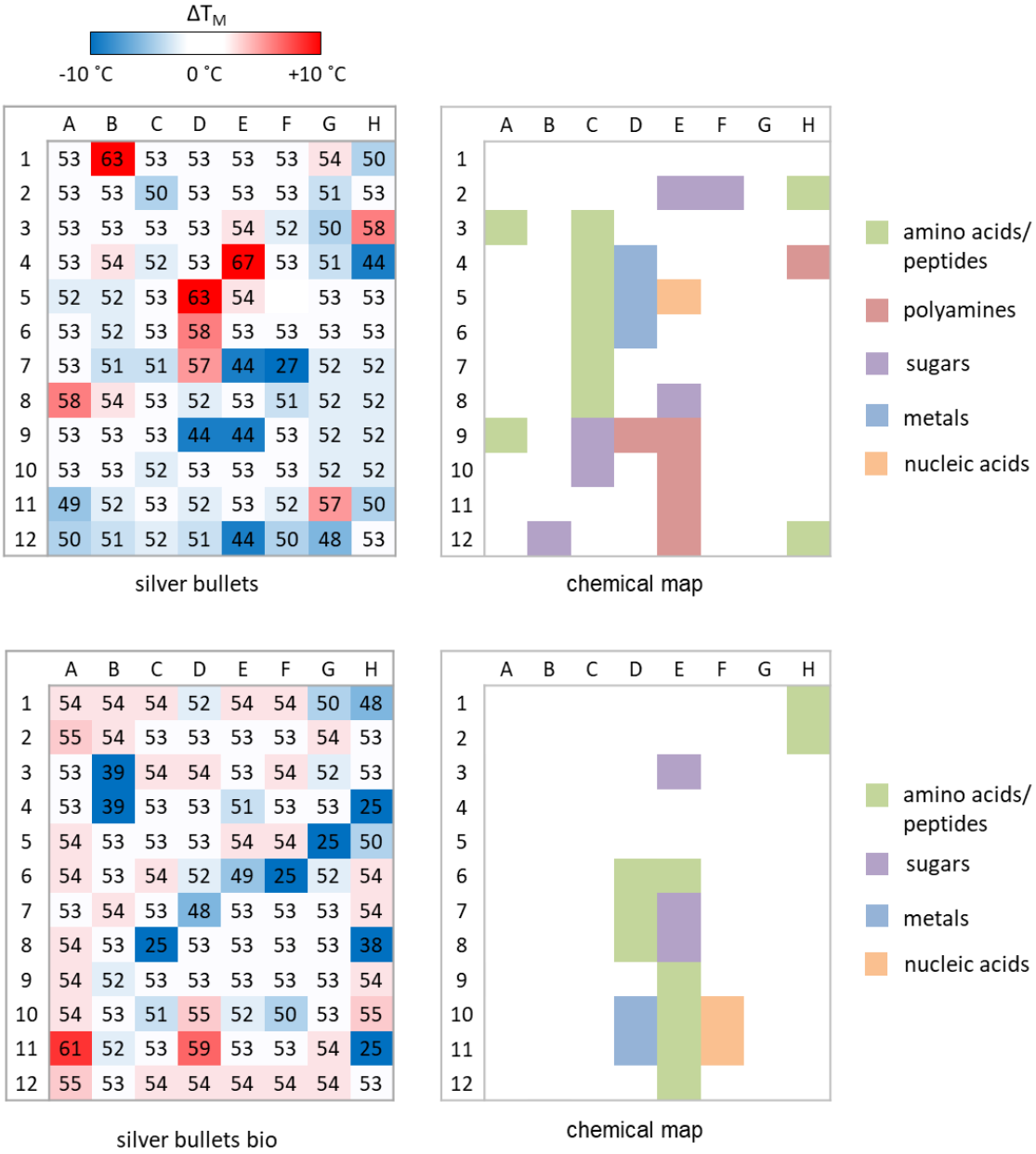

ΔT_M
-10 ˚C
0 ˚C
+10 ˚C
A B C D E F G H
1 53 63 53 53 53 53 54 50
2 53 53 50 53 53 53 51 53
3 53 53 53 53 54 52 50 58
4 53 54 52 53 67 53 51 44
5 52 52 53 63 54 53 53
6 53 52 53 58 53 53 53 53
7 53 51 51 57 44 27 52 52
8 58 54 53 52 53 51 52 52
9 53 53 53 44 44 53 52 52
10 53 53 52 53 53 53 52 52
11 49 52 53 52 53 52 57 50
12 50 51 52 51 44 50 48 53
silver bullets
chemical map
amino acids/ peptides
polyamines
sugars
metals
nucleic acids
A B C D E F G H
1 54 54 54 52 54 54 50 48
2 55 54 53 53 53 53 54 53
3 53 39 54 54 53 54 52 53
4 53 39 53 53 51 53 53 25
5 54 53 53 53 54 54 25 50
6 54 53 54 52 49 25 52 54
7 53 54 53 48 53 53 53 54
8 54 53 25 53 53 53 53 38
9 54 52 53 53 53 53 53 54
10 54 53 51 55 52 50 53 55
11 61 52 53 59 53 53 54 25
12 55 53 54 54 54 54 54 53
silver bullets bio
chemical map
amino acids/ peptides
sugars
metals
nucleic acids

Figure 4.1 Hfx1 melting temperatures in the presence of molecular cocktails. (**A**) Raw and 1st derivative melt curves for select conditions producing a ΔT_M > +4 °C from the silver bullets screen. T_M points are indicated on the derivative curve for Hfx1 combined with protamine sulfate and in apo form. (**B**) Left: T_M values across 96 chemical cocktails, coloured according to change in T_M (red for + ΔT_M, blue for - ΔT_M). T_M values (°C) are listed for each condition. Right: Location of cocktails containing common molecules in experimental grid. All molecules are listed in **Tables S.1 and S.2**.

There were also many conditions that destabilised Hfx1 (approximately 13). When looking at common molecules in these conditions, there are two noticeable trends. The first is that conditions containing cystamine dihydrochloride resulted in a ΔT_M of -10 to -15 °C (D9, E9, E12, H4, *B3, *B4, and *H8, **Figure 4.1 B**). The second is that riboflavin and flavin adenine dinucleotide disodium salt hydrate completely denatured the protein at room temperature (T_M 25 °C). A summary map across all wells of the commercial screens utilised is presented in **Figure 4.1** and is shaded to indicate discrete changes in T_M where values increased or decreased. This is combined with a categorisation of chemical classes in each screen.

As stated before, 3,5-dinitrosalicylic acid was proposed as a stabilising molecule, as it is present in three cocktails that increased the T_M of Hfx1 (A8, H3, and D7). However, it is also contained in three other conditions that resulted in no increase in T_M (wells A1, A4, and B4). It is possible that there are masking effects from other compounds present in these cocktails, or that compounds found in wells A8, H3, and D7 reacted with 3,5-dinitrosalicylic acid in such a way to increase stability of Hfx1. Unfortunately, there are no other common compounds within the three favourable conditions, shedding no light on the increased stability of Hfx1 under these conditions.

4.1.1.1. Protamine sulfate binds by electrostatic interaction

During ligand screening, protamine sulfate was seen to increase the T_M of Hfx1 by as much as 14 ˚C, qualifying it as a possible ligand. This was not seen for any other protein studied in this thesis. Protamine sulfate (or salmine) is a 32 amino acid, arginine-rich peptide that is most

commonly known to bind sperm DNA during cell division (see below for sequence, (Roque *et al.*, 2011; Moir & Dixon, 1988)).

```
         10         20         30
PRRRRSSSRP VRRRRRPRVS RRRRRRGGRR RR
```

As arginine is positively charged at neutral pH (Hfx1 is maintained in pH 7.5 buffer), the interaction between this small peptide and Hfx1 most likely occurs at negatively charged regions on the surface, either around the hypothesised active site or at the trimer interface in the centre of the protein (**Figure 4.2**). Protamine has previously been crystallised in complex with other proteins in the PDB, with **Figure 4.2** showing insulin NPH and trypsin as examples (PDB 2OMG and 2J9N). In both cases protamine is bound to the negative regions between subunits of each protein (Norrman *et al.*, 2007; Viola *et al.*, 2007). Electron density is not high enough to resolve the structure of the bound protamine in complex with insulin NPH, but the binding region is known, and therefore it is denoted in **Figure 4.2**. Although Hfx1 does not have open spaces between subunits for protamine to bind, as with insulin and trypsin, there are many negatively charged regions that could act as possible binding sites.

The role of protamine in bacterial cells is unknown and may even be non-existent. However, there have been some studies into its effect within these cells. Transcriptional termination due to the high binding affinity of arginine-rich protamines to the negatively charged DNA backbone was investigated in *E. coli* cells. It was found that protamine expression significantly attenuated cell proliferation when compared with control cells (Gunther *et al.*, 2015). The antibacterial properties of protamine were also seen in a study focussing on its bactericidal effects on twelve strains of *Viridans streptococci* (Kim *et al.*, 2015). However, the opposite effect was seen when researching the marine bacterium *Pseudoalteromonas* sp. Results showed that protamine is only active as an antimicrobial in seawater absent of divalent cations. In the presence of the divalent cations Mg^{2+} and Ca^{2+}, protamine enhances the growth of bacterium and produced chains rather than individual cells (Pustam *et al.*, 2014). So, it seems that there are many possible roles for protamine in bacteria, and following on from that, many possible reasons that Hfx1 may interact with protamine within a bacterial cell.

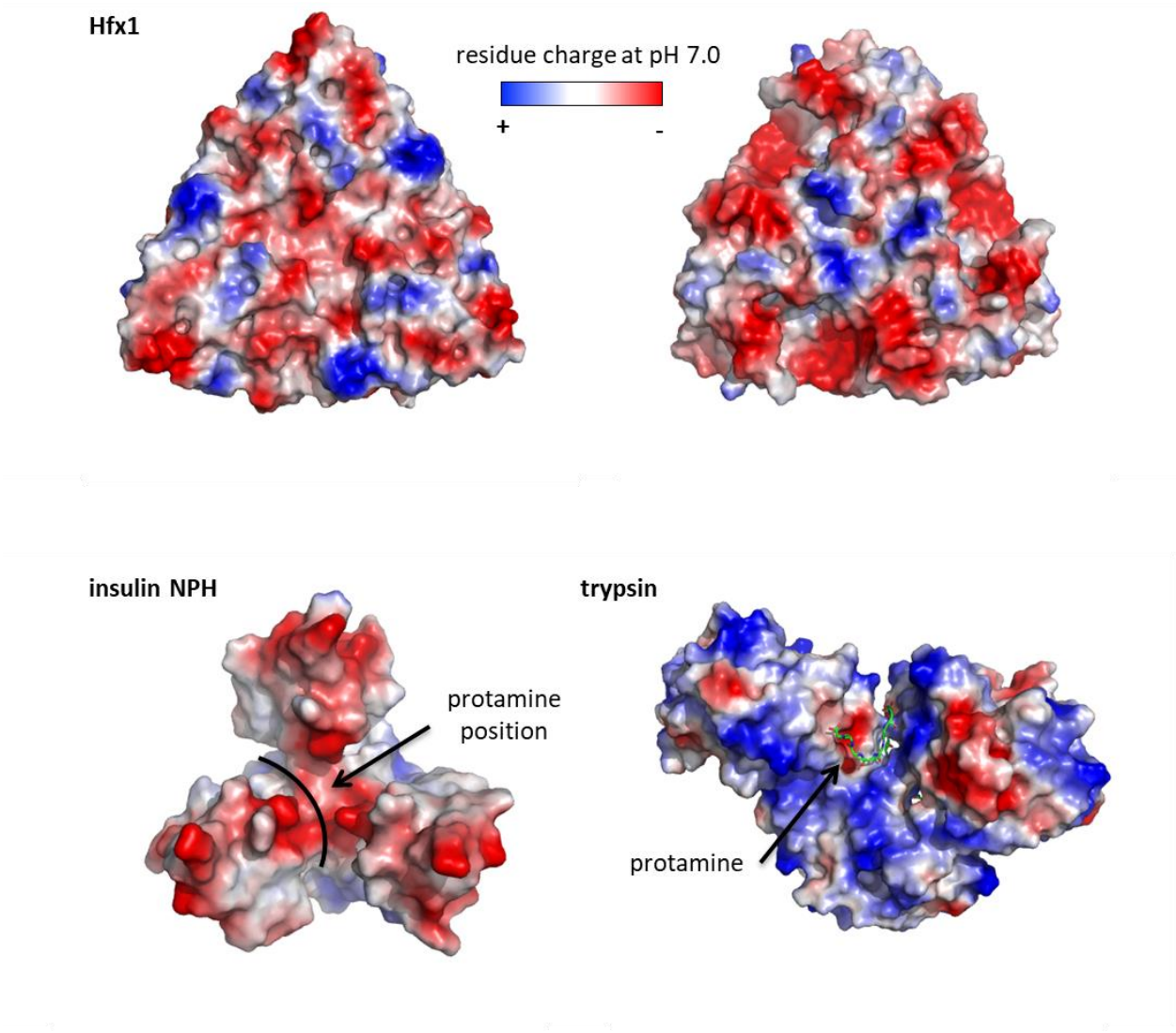


Figure 4.2 Surface charges of Hfx1 and protamine bound insulin NPH and trypsin. Surface patches coloured red are negative and blue are positive. The approximate position of protamine is indicated on the structure of insulin. The backbone of protamine can be seen in green bound to the structure of trypsin.

4.1.1.2. Metal binding capacities of Hfx1

After it was found that three of the screens that gave rise to an increase in melting temperature contained metals, the structure of Hfx1 was examined and it was found that each monomer contains four histidine residues (His-70, His-82, His-126, and His-147). His-70 is located at the base of β_4, His-82 on the β_4-β_5 loop, His-126 on the α_2-β_6 loop, and His-147 on the β_6-α_3 loop. These histidine residues surround the area hypothesised to be the ligand binding pocket and could play a part in metal binding (**Figure 4.3**).

There have been studies into preferred coordination geometries of metals in known protein structures (Dokmanic *et al.*, 2008). This information could aid in the discovery of which metals are more likely to bind to Hfx1. Metals such as cobalt, nickel, and cadmium can form an octahedral arrangement, copper forms a square planar coordination, and zinc and cadmium prefer tetrahedral coordination (Rulisek & Vondrasek, 1998).

Metals preferring either square planar or tetrahedral coordination would be more likely ligands for Hfx1 compared to octahedral metals as there are only four His residues available for electron donation. In addition, their metalloprotein complexes belong to the most abundant arrangements seen in protein structures, and these metals are major pollutants of the environment, increasing their potential for uptake by organisms (Rulisek & Vondrasek, 1998). Although the histidines in Hfx1 are all too far apart for them to simultaneously be involved in metal binding, it is possible for other ions in solution to assist in metal coordination. In order to further investigate the roles of these histidine residues in metal binding, each one was separately mutated to a leucine (**Table 8.7**). All mutants were successfully expressed as soluble proteins, with each one yielding over 20 mg/ml of protein. However, there was variation in stability, with H147L being the least stable mutant (T_M 39 ˚C, **Figure 4.3**). The mutations had no effect on Hfx1 oligomerisation, with all mutants eluting from an analytical SEC column as trimers (K_{AV} 0.46).

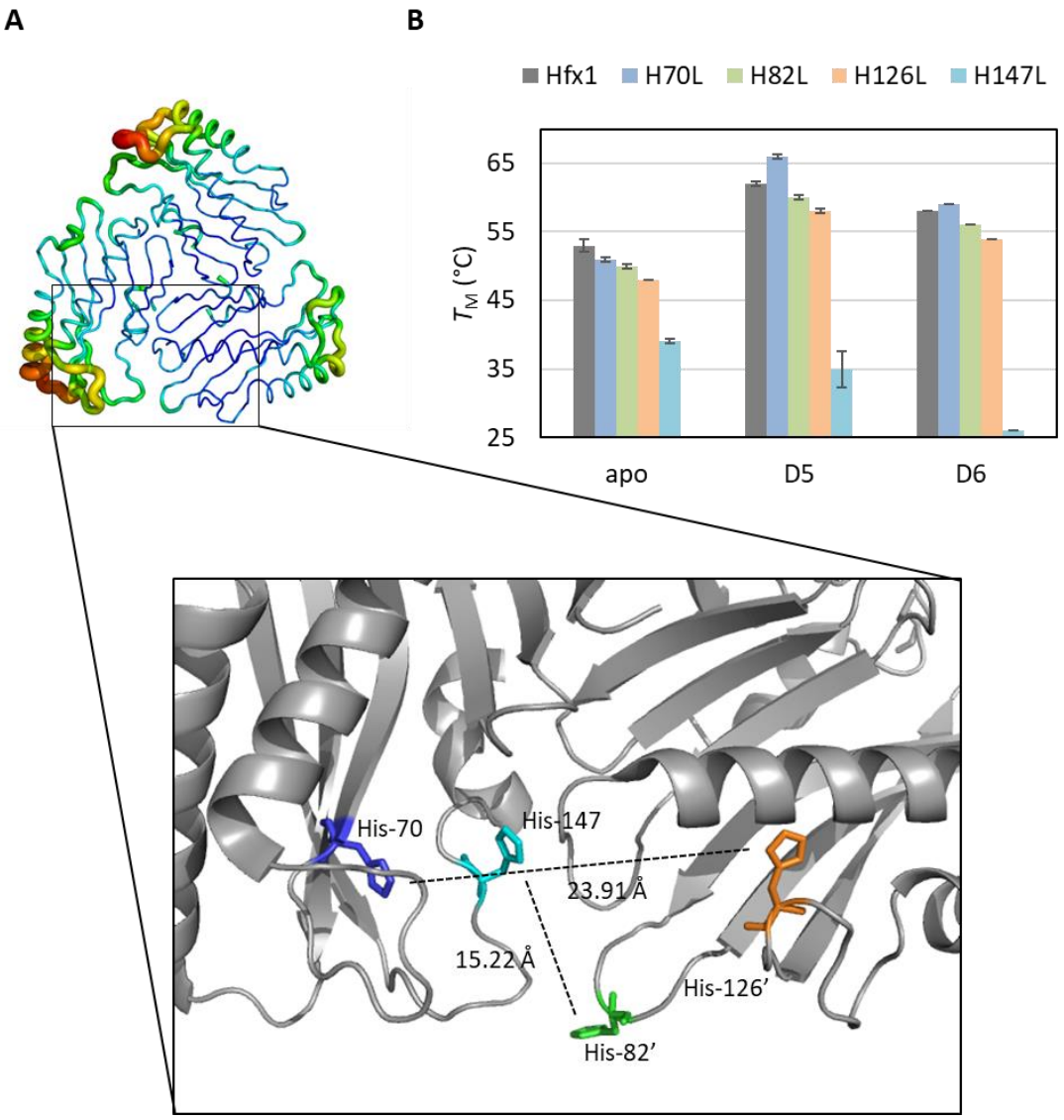


Figure 4.3 Histidine residues of Hfx1 chosen for mutagenesis. (**A**) B-factor of Hfx1 trimer with a close-up of the hypothetical binding pocket with His residues shown. Distances between residues are shown by a dotted line. (**B**) DSF melting temperatures of native Hfx1 and each His mutant in apo form and in the presence of metals from conditions D5 and D6 of the Silver Bullets screen. SEM error bars are indicated (n = 3).

As conditions D5 and D6 in the Silver Bullets screen contain Group II and transition metals and result in an increase in T_M of native Hfx1, the same conditions were used to test the change in T_M of the mutants using DSF. It can be seen from the graph in **Figure 4.3** that the relative T_M remains the same across conditions for three of the histidine residues (His-70, His-82, and His-126) when compared to native Hfx1. This suggests that they do not individually participate in metal binding and are not integral to the overall stability of Hfx1. If one of the residues did participate in metal binding, the T_M would remain at 53 °C or drop to a lower value when mutated to leucine. This indeed could be the case for the H147L mutant. As seen in **Figure 4.3**, H147L decreases in thermal stability when exposed to conditions D5 and D6. As all screens were tested simultaneously with DSF, loss in stability is not a time related issue. This suggests that His-147 is necessary for stabilisation of Hfx1, and that in the absence of the amino acid, metals continue to destabilise the protein.

4.1.2. Screening of potential effectors for Hfx5

As Hfx5 purified as both a dimer and a monomer, it was important to screen each species separately for binding potential. Without separation, different responses by each species would be seen as noise in the signal and could add additional complexity to characterisation. For this reason, the monomer was separated from the dimer using preparative SEC, as described in Section 3.2.2, and tested each sample separately against a panel of small molecule cocktails. These analyses were carried out within the same time period when Hfx5 monomer and dimer were seen to be distinct in solution after separation, and in the same buffer. For these reasons they were presumed to be separate oligomeric species. Between the two species there was a noticeable difference in response to the cocktail screen. It can be seen in **Figure 4.4** that although there was little difference in T_M between the two species under identical buffer conditions, in conditions with additives, such as sugars, nucleic acids, and amino acids, the monomer form of Hfx5 was less stable. The Hfx5 monomer was less stable in 32 conditions when compared to the dimer (e.g. wells H5 and D12). The reverse is only true for approximately 4 conditions, including wells A9, G5, A7, and D11.

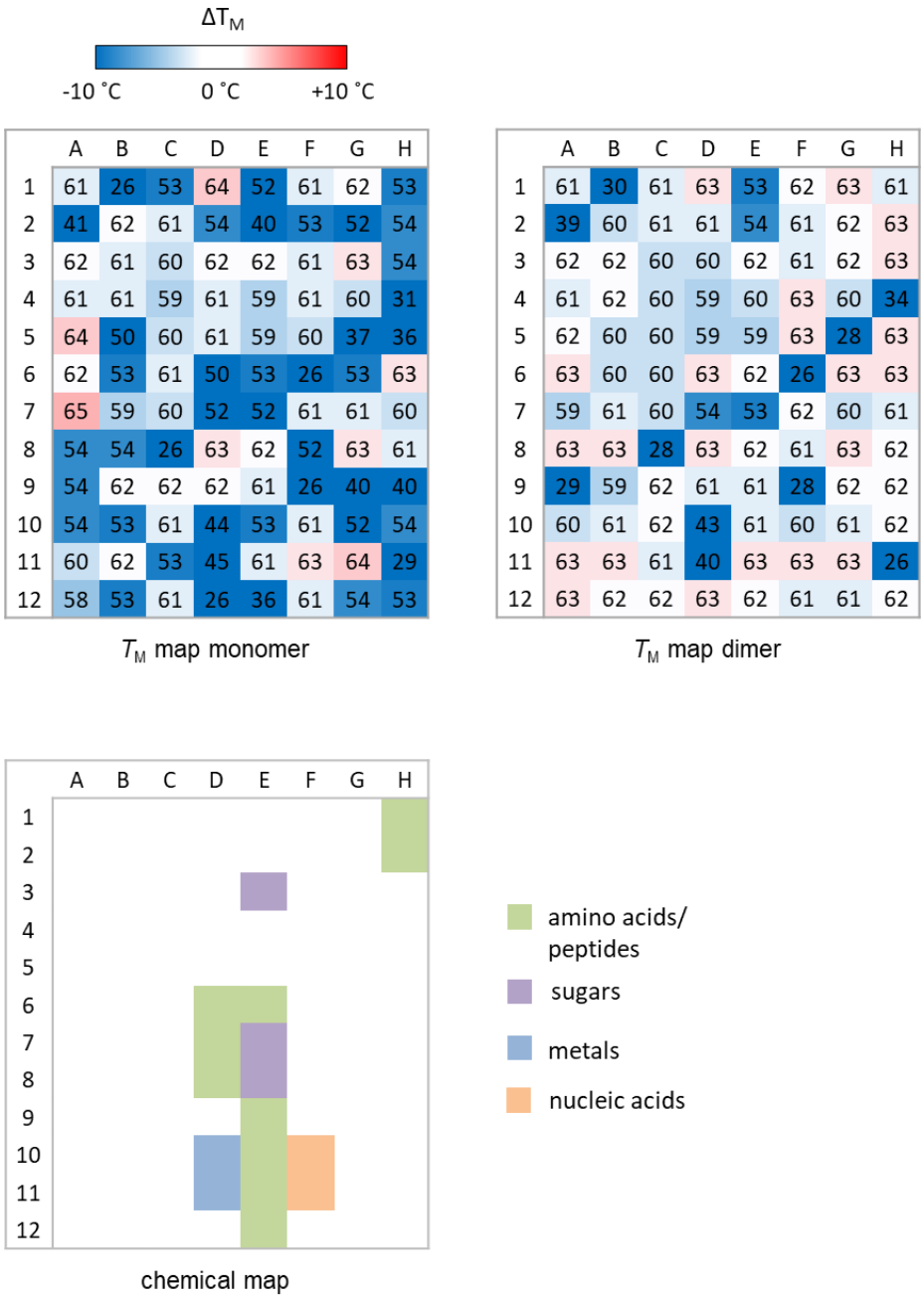


Figure 4.4 Hfx5 melting temperatures in the presence of molecular cocktails. Top: T_M values across 96 chemical cocktails, coloured according to change in T_M (red for + ΔT_M, blue for - ΔT_M) for both monomer and dimer. T_M values (°C) are listed for each condition. Bottom: Location of cocktails containing common molecules in experimental grid.

Unfortunately, over all 96 conditions there was no noticeable increase in protein stability (nothing above +2 ˚C). There were many conditions that decreased protein stability, but the variation in molecules was so wide, that no trend could be discerned.

4.2. Summary

Previously, searches against a database of cognate binding sites (Najmanovich *et al.*, 2008) identified some features of the proposed active site (within the β-bulge between β_3 and β_4 strands) common to enzymes utilising nucleotide-based cofactors (e.g. adenosine and/or nicotinamide moieties (Sureshan *et al.*, 2013)). However, Hfx1 displays none of the known sequence motifs for binding these cofactors. There are many conditions within the screens that contain mono- or di-nucleotide cofactors, but not a single condition increased protein stability. In addition, some sub-fold similarity was detected to the zinc transporter CzrB (PDB 3BYP (Cherezov *et al.*, 2008)) from *Thermus thermophilus*. In CzrB, the domain presents a cluster of zinc-binding residues for metal chelation and controls a dimerisation event critical to function. However, these active site residues are not replicated in the equivalent strands (β_3-β_6) of Hfx1, to which there appears to be no functional relationship. Therefore, it is unfounded to think that the metal affinity displayed by Hfx1 occurs in this region.

Ligand screening pointed to two molecular groups that increased the fold stability of Hfx1. Protamine sulfate most likely interacts with negative patches within the proposed active site of Hfx1, as the molecule itself is very positively charged at neutral pH (ligand screening was carried out at pH 7.5). However, it is still unclear where the Group II and transition metals would bind. Originally, it was proposed that they would bind to a selection of four histidine residues within a cleft between chains. Each histidine residue was therefore mutated to a leucine to determine if any of the residues were responsible for metal binding, however, the only mutation that resulted in a binding difference was H147L, which almost completely destabilised Hfx1.

In the previous chapter, Hfx5 was found to be able to associate as both monomer and dimer in solution. Both species had the same melting temperature after size exclusion separation and

seemed similar in structure. However, this was contrasted by the ligand screening results, which showed a noticeable difference in stability between the monomer and dimer form of Hfx5, not originally seen with DSF. When combined with small molecule cocktails, the monomer was less stable under a third of conditions tested. There were no conditions that resulted in a noticeable increase in protein stability.

4.3. Conclusions and future work

4.3.1. Bacterial adaptations through protein mimicry

The pressures of survival have resulted in a fascinating spectrum of adaptations in organisms. Different organisms have evolved sophisticated methods to exploit the surrounding environment and each other. An important mechanism that frequently reoccurs in this process of adaptation is that of mimicry (Stebbins & Galan, 2001). Studies have revealed that many bacterial pathogens mimic the function of host proteins to manipulate host physiology and cellular functions for the microbe's benefit (Guan & Dixon, 1990; Hardt *et al.*, 1998; Hamburger *et al.*, 1999). Mimicry through convergent evolution involves taking materials (genes and the proteins that they encode) already available to the pathogen and then sculpting them to perform a new function (Stebbins & Galan, 2001). Proteins resulting from this approach usually have a distinct three-dimensional architecture from that of the molecule it mimics, but would typically have evolved to imitate the chemical groups on the surface of its functional homolog.

An ideal example of this is seen in sugar kinases. Each of the three families of sugar kinases (hexokinase, ribokinase, and galactokinase families) appears to have a distinct three-dimensional fold, since conserved sequence patterns are strikingly different for the three families (Bork *et al.*, 1993). Yet each catalyses chemically equivalent reactions on similar or identical substrates. The enzymatic function of sugar phosphorylation appears to have evolved independently on the three distinct structural frameworks, by convergent evolution. These are examples of independent Darwinian adaptation of a structure to the same substrate at different evolutionary times (Bork *et al.*, 1993).

This subset of novel-fold proteins, Hfx1, Hfx5, Vch3, and Vch14, are a possible result of convergent evolution within bacterial populations. Although they possess unique folds, they may perform similar functions to other proteins in their environmental niche or resemble them on the surface. In most cases, clusters of genes that are physically close together in the genome are often conserved as a unit through evolution, and tend to encode proteins that interact or are part of the same operon (Gabaldon & Huynen, 2004). Thus, chromosomal proximity can be used to predict functional similarity between proteins, at least in prokaryotes. However, as this method is not possible for the proteins studied here, due to the mode of their recovery from gene cassettes in environmental samples, other methods were required to infer function. However, the work in this thesis has only just scratched the surface in understanding the role of these proteins.

4.3.2. Future work

4.3.2.1. Further analysis of Hfx1 stabilising molecules

T_M shifts cannot be readily transformed into binding affinities, and the ligand affinity obtained at denaturation temperatures does not necessarily coincide with the affinity at physiological temperature (Schon *et al.*, 2013). For this reason, it will be necessary to obtain kinetics and thermodynamic data for each compound seen to increase protein stability using DSF. One method that can be used for obtaining the kinetics of ligand binding is surface plasmon resonance (SPR). SPR uses an optical method to detect changing of mass brought about by the binding of analyte molecules to any receptor tethered on a thin surface (Homola, 2008). SPR can measure binding in real time and, if used proficiently, can provide accurate measurements of binding rates and affinity (Day *et al.*, 2002). One benefit of using SPR is that it requires only µM concentrations of protein. However, there is always the risk of nonspecific binding of analyte to the chip surface. Isothermal titration calorimetry (ITC) is a method that does not suffer from non-specific binding, but requires much more protein sample for accurate thermodynamic data. Isothermal titration calorimetry can monitor the formation or dissociation of molecular complexes by measuring the energy needed to offset the heat of reaction (Ghai *et al.*, 2012). In short, a ligand is injected in aliquots into the sample solution containing protein, and the heat either released (exothermic

reaction) or taken up (endothermic reaction) is measured by determining the power required to maintain a constant temperature with respect to a reference solution (Ladbury, 2010). As the titration progresses and the binding sites become saturated, the reaction heats tend to zero (Perozzo *et al.*, 2004). ITC is also able to determine binding stoichiometries, which is useful when analysing the binding of proteins to small molecules that have the capacity to associate at multiple sites within the protein or on the surface. Both SPR and ITC would be ideal for understanding the modes of stabilisation brought about by the addition of protamine and divalent metals to solution containing Hfx1, distinguishing between specific and non-specific binding, and should be used in tandem.

Detailed knowledge of ligand binding sites is limited to proteins with known structures in which the ligand is also present. Many methods have used residue conservation and surface clefts for prediction of binding sites (Wass *et al.*, 2011), but predictions are still only theoretical. In addition to the methods mentioned above, it would be highly valuable to attempt co-crystallisation of Hfx1 with the single metals and protamine sulfate seen in each stabilising cocktail. If crystallisation were successful, the binding site for each of the molecules would be known, and binding contacts could be studied in detail.

4.3.2.2. Identification of ligands for Vch14

Identifying ligands for co-crystallisation experiments in structural genomics, however, requires a different strategy than for drug discovery. The latter is characterised by a high degree of knowledge about the protein target, its biochemical mechanism, and substrates. The structural genomics effort is characterised by a significantly reduced amount of information about the protein. In some circumstances, such as the one presented in this thesis, a newly determined structure will represent the first three-dimensional model of the protein. Any additional protein-ligand interaction data that is generated can provide valuable context for increasing the biological impact of the structure. There are two main ways to approach the acquisition of ligand interaction data, *in silico*, and *in vitro*.

Computational approaches to determine protein-ligand interactions are becoming more feasible as technology becomes more advance. Using a variety of publicly available software tools, it is possible to computationally model, predict, and evaluate how different ligands interact with a given protein. This can be achieved using molecular docking programs such as DOCK 6 (University of California, San Francisco (Lang *et al.*, 2009)) and AUTODOCK (The Scripps Research Institute (Morris *et al.*, 2009)). As a first step for determining binding partners for Vch14, this would be an ideal approach as it is seen to be efficient, minimising experimental efforts, reduces costs, and improves the success of ligand identification for protein targets (Binkowski *et al.*, 2014). After this first step, identified ligand targets could be tested with methods such as ITC or SPR, or further ligand screening could be carried out experimentally.

4.3.2.3. Using novel fold proteins as building blocks for nanomaterials

Overall, Hfx5, Vch14, and Hfx1 are very stable bacterial proteins that oligomerise and are relatively small. For this reason, they would be ideal as building blocks in larger macromolecular structures. Such synthetic structures have the potential to address important issues related to drug delivery, such as reducing drug toxicity, protection from drug degradation/sequestration, increasing circulation times, targeting, and increasing bioavailability (Farokhzad & Langer, 2009; Maham *et al.*, 2009; Wagner *et al.*, 2006). Conventional materials investigated for drug delivery include synthetic polymeric and liposomal nanoparticles (Wagner *et al.*, 2006). These, however, may have limitations such as wide size distributions, difficulty in site-specific functionalisation, low drug loading, and instability. Protein cages represent a class of nanomaterial that may address many of these concerns (Yildiz *et al.*, 2011; Smith *et al.*, 2013).

Protein cage complexes are hollow structures comprised of self-assembling protein subunits that produce nanocapsules with a nearly monodisperse size distribution. The individual asymmetrical subunits may comprise a single protein, as with pyruvate dehydrogenase E2 (Ren *et al.*, 2011), or multiple proteins, such as with the structure comprising Hfx1 and MIF1 seen in **Figure 2.1** (King *et al.*, 2014). Typical sizes of protein nanocages range from 10-100 nm, and they display repetitive

symmetrical features, both of which are ideal structural features for receptor-mediated endocytosis (Molino & Wang, 2014). The proteins studied here, with a focus on Hfx1, are ideal for creation of these nanoparticles, especially if they can be engineered to self-assemble. The work discussed in Chapter 2 regarding Hfx1 and MIF1 is proof of concept. With changes to the residues in structured areas along helices α1 and α2 to hydrophobic amino acids, a hydrophobic interface was created where self-assembly could occur.

The data collected in this research is a first step to providing a functional identity to the four novel-fold proteins recovered from unknown soil-dwelling bacteria. Although it is important to see which genes are being transferred between bacteria through LGT, and the discovery of proteins of novel fold is exciting, this alone does not explain why these elements are being traded between species. It is only through functional knowledge that this information becomes relevant. It is now known that Hfx1 has affinity for metals and protamine sulfate and is extremely robust under a number of conditions and mutations. Hfx5 is a fully functional domain-swapping protein that tends toward the monomer form when concentrations are low. Although Vch3 was thought to be a dimer, it presents as a tetramer in solution. All of these parameters can now be added to the knowledge that each of these proteins folds in a way that has never been seen before. Although there is still more data to be acquired on these, and other proteins whose structures are their only known factor, this is a step in the right direction.

PART II

CHAPTER 5

Efflux Pump Regulator Proteins of *Acinetobacter baumannii*

5.1. Efflux systems as a mode of continued virulence and adaptation in bacteria

Efflux pumps are grouped as either primary or secondary transporters. All engage conformational changes in order to transport compounds against a concentration gradient. Primary transporters utilise the hydrolysis of ATP as their main source of energy for this process (Marshall & Piddock, 1997; Lorca *et al.*, 2007), whereas secondary transporters use the proton motive force (PMF (Amaral *et al.*, 2014)). The latter is the electrochemical gradient due to the greater concentration of hydronium ions (H_3O^+) in the periplasm over that in the cytoplasm proximal to the inner membrane (Pages & Amaral, 2009).

Efflux pumps utilise single or multiple protein components. Efflux pumps of the resistance nodulation division (RND) family, expressed in Gram-negative bacteria, are organised as tripartite systems, as depicted in **Figure 5.1**. They comprise a transporter (or efflux) protein, embedded in the inner (cytoplasmic) membrane (Aires & Nikaido, 2005), an accessory (or membrane-fusion) protein located in the periplasmic space, and an outer membrane protein (or outer membrane channel (Piddock, 2006)). Substrates from within the phospholipid bilayer of the inner membrane or from the cytoplasm are captured by the transporter protein and then moved to the extracellular space through the outer-membrane protein channel (Koronakis *et al.*, 2004). The role of the periplasmic accessory protein is to mediate between transporter and outer membrane proteins (Eswaran *et al.*, 2004). In some cases, a single outer-membrane protein (e.g. TolC and OprM of *P. aeruginosa*) can function promiscuously as the channel for different efflux family pumps, or multiple pumps of the same family (Piddock, 2006; Borges-Walmsley *et al.*, 2003).

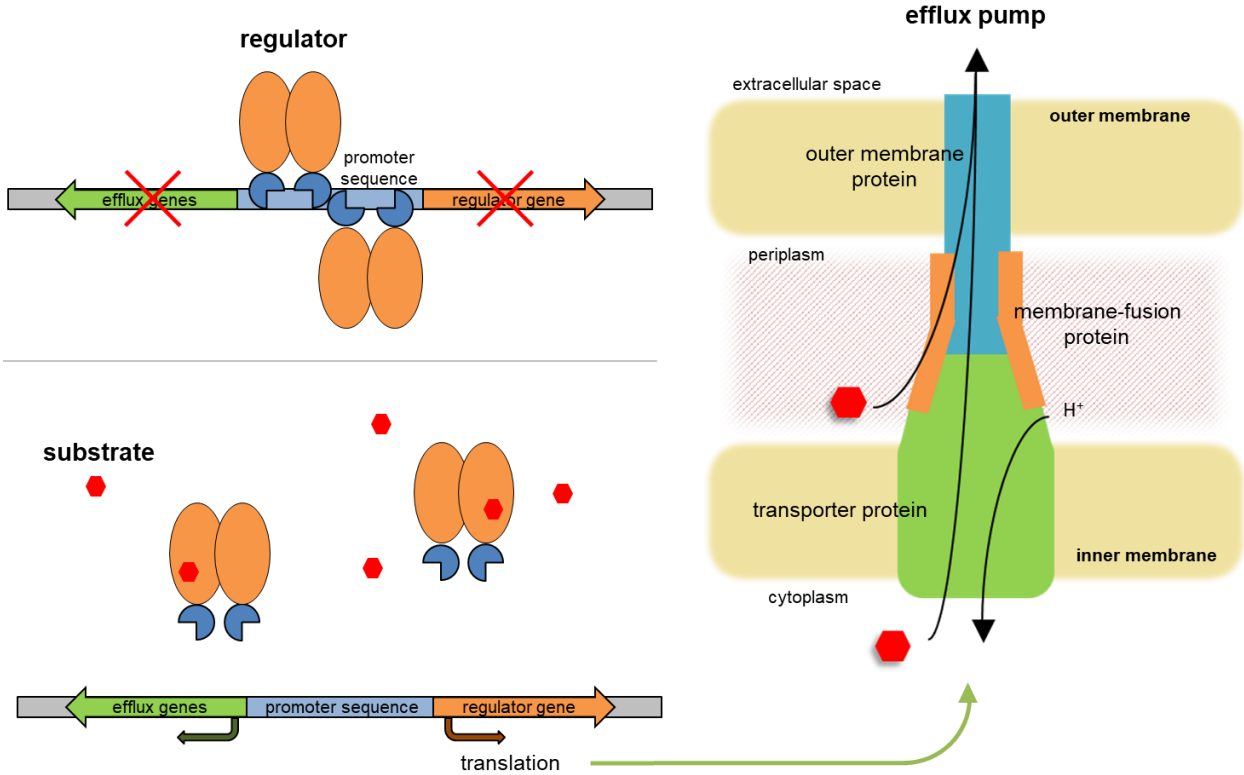


Figure 5.1 Genetic control of resistance nodulation division efflux pump expression by regulator proteins. The substrate-free regulator (orange/blue) engages the intergenic DNA region. When bound to its substrate (red hexagon), the regulator-ligand complex can no longer bind DNA, allowing translation of efflux pump genes. A schematic of the pump is coloured to indicate a tripartite organisation of transporter (green), membrane-fusion (orange), and outer membrane (cyan) proteins within bacterial membranes, showing the export of substrate by proton motive force.

5.1.1. Factors that increase the expression of efflux systems

In some cases, basal levels of efflux can cause intrinsic resistance to antimicrobial compounds (Nikaido, 2001). Alternatively, resistance to antimicrobial agents can result from a constitutive increase in gene expression of the efflux-pump transporter protein (Cox & Wright, 2013), detected through increased minimum inhibitory concentration (MIC, the lowest threshold concentration that inhibits growth (Blair *et al.*, 2014)). Permanent upregulation of efflux-pump expression may occur through mutation of key components: the local repressor gene, a global transcriptional activator,

the promoter region for the efflux-pump gene, or insertion elements upstream of the efflux-pump gene (Piddock, 2006; Paulsen, 2003). The MIC values in cases of such mutations are higher (typically 2–8-fold) than the MICs for the susceptible strain, so sufficient to render the bacterium resistant to recommended breakpoint concentrations (Poole, 2004).

Adaptive resistance of bacteria may also arise through alterations in gene and/or protein expression (Barclay *et al.*, 1992). The resulting temporary increase in the ability of the organism to survive antibiotic insult usually reverts upon removal of the inducing condition (Fernandez & Hancock, 2012). Adaptive resistance in bacteria builds over multiple generations, either when exposed to increasing concentrations of antibiotics, or when the antibiotic is removed (Adam *et al.*, 2008; George & Levy, 1983; Toprak *et al.*, 2011).

5.1.2. Families of efflux pumps

A single organism can express efflux pumps from several families or more than one type of the same family.

- The ATP-binding cassette (ABC) superfamily often consist of multiple subunits; sometimes two transmembrane proteins, or two membrane-associated ATPases (Jones & George, 2004). The ATPase subunits utilise ATP binding and hydrolysis to energise the translocation process. Substrates of the ABC efflux superfamily include amphipathic compounds, bile salts, linear and cyclic peptides, steroids, detergents, fluorescent dyes, ionophores and lipids (Dassa & Bouige, 2001; Higgins, 2001).
- The major facilitator superfamily (MFS) function in the uptake of sugars and other diverse compounds (including metabolites, oligosaccharides, amino acids and oxyanions (Marger & Saier, 1993)). The MFS transporter is a transmembrane system built around two homologous domains each organised as 6-helix bundles (Foster *et al.*, 1983; Maiden *et al.*, 1987). There is some evidence of a functional role for this dimeric organisation (Sun *et al.*, 2014a).

- The multidrug and toxic-compound extrusion (MATE) family function as exporters of cationic drugs through H^+ or Na^+ exchange (Omote *et al.*, 2006). They possess 12 transmembrane segments, but are distinct from the MFS family (Morita *et al.*, 1998).
- The small multidrug resistance (SMR) family confer resistance to a variety of quaternary ammonium compounds and other lipophilic cations (Paulsen *et al.*, 1996; Bay *et al.*, 2008). They are composed of four transmembrane α-helices, 100–140 amino acids in size.
- The resistance nodulation division (RND) family have been implicated in host virulence (Bina *et al.*, 2008; Chan & Chua, 2005; Buckley *et al.*, 2006) and biofilm formation (Baugh *et al.*, 2012; Blair *et al.*, 2014). The best characterised RND system is the *Escherichia coli* AcrAB-TolC system made up of an inner membrane transporter (AcrB), an outer membrane protein channel (TolC) and a periplasmic adaptor protein (AcrA, (Du *et al.*, 2014)). Both oxidative and nitrosative stress can induce the expression of RND pumps, suggesting these pumps to be part of the bacterial defence system against reactive oxygen and nitrogen species (Hirakata *et al.*, 2002).
- The proteobacterial antimicrobial compound efflux (PACE) transporter family has recently been defined after its founding member, AceI, was seen to confer resistance to chlorhexidine in *Acinetobacter baumannii* (Farrugia *et al.*, 2013; Hassan *et al.*, 2015). Genes encoding proteins homologous to AceI are evident in the genomes of many bacteria, including pathogens, and are particularly common among proteobacteria (Hassan *et al.*, 2018). These genes are thought to encode an inner membrane protein of four transmembrane helices (Hassan *et al.*, 2018). The PACE family possibly respond to a relatively limited range of substrates (Hassan *et al.*, 2013; Hassan *et al.*, 2015).

5.2. Molecular mechanisms of transport regulation: efflux pump transcriptional regulators

Regulators of gene transcription are capable of either repression or activation following binding at specific promoters (Romero-Rodriguez *et al.*, 2015). Of the large number of bacterial transcription regulators, two families are of particular relevance to efflux systems: the TetR family

(TFR) and the LysR-type transcriptional regulator (LTTR) family. The former are responsible for regulating biosynthesis of antibiotics, efflux pumps, and osmotic stress, and the latter for repressing or activating carbon and nitrogen metabolism (Rodionov, 2007). Both TFR and LTTR protein families incorporate a DNA-binding domain (DBD) with different helix-turn-helix motifs, as depicted in **Figure 5.2**. The two types of regulators can be reliably distinguished through sequence similarities of this N-terminal segment (Yu *et al.*, 2010; Ramos *et al.*, 2005). The effector-binding domain (EBD) varies broadly in sequence and structure, consistent with its ability to interact with a wide variety of ligand chemistries and structures.

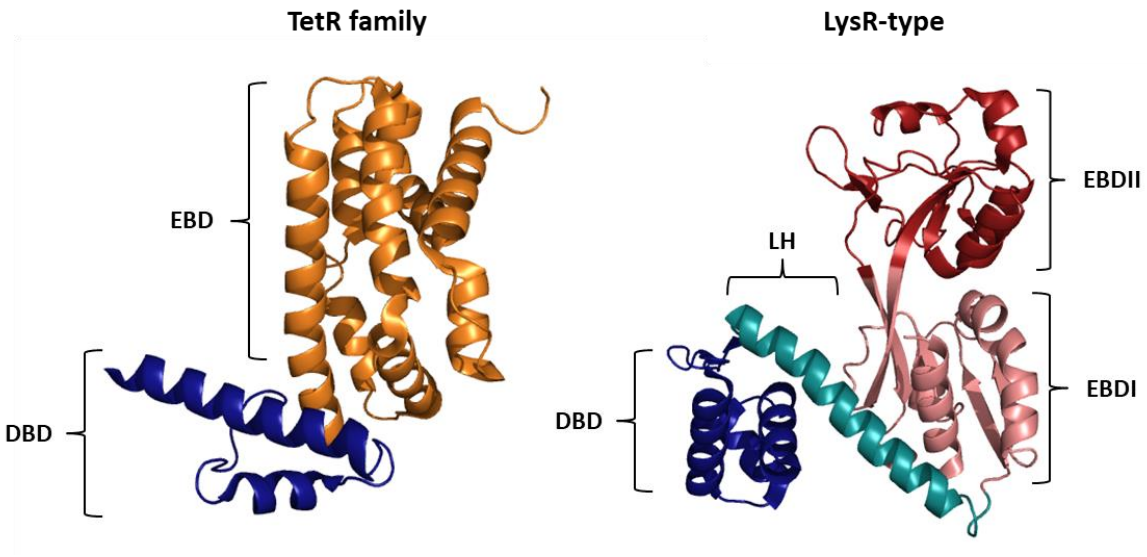


Figure 5.2 Similarity of domain organisation for TFR and LTTR regulators. Domains indicated are DNA-binding domain (DBD, blue), effector-binding domain (EBD, orange), linker helix (LH, cyan), EBDI (pink), and EBDII (red). The representative TetR family (TFR) structure is CmeR (Gu *et al.*, 2007), and the LysR-type (LTTR) structure is CrgA (Sainsbury *et al.*, 2009).

5.2.1. TetR transcriptional regulators

Most bacterial genomes encode at least one TFR (Ramos *et al.*, 2005; Cuthbertson & Nodwell, 2013), but they are not found in pathogens with genomes under 2 Mbp. Notably, the *Actinobacteria* sp., along with other soil-dwelling strains, encode a high number of TFRs (Cuthbertson & Nodwell, 2013). TFRs can be either global or local regulators, sometimes acting in concert with several intermediate regulatory genes (Ramos *et al.*, 2005). Although the binding mechanism for TFRs is conserved, the placements of regulator and target genes are not. Three types of gene orientation have been outlined by **Figure 5.3** (Cuthbertson & Nodwell, 2013).

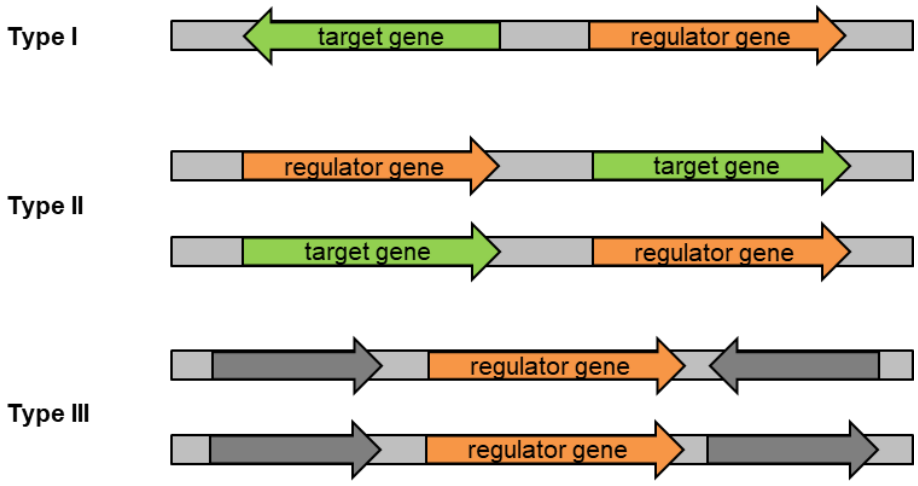


Figure 5.3 Types of repressor and transcriptional identities. Type I is transcribed divergently; Type II is cotranscribed; Type III is constitutively expressed and shares no relationship with adjacent genes (Cuthbertson & Nodwell, 2013).

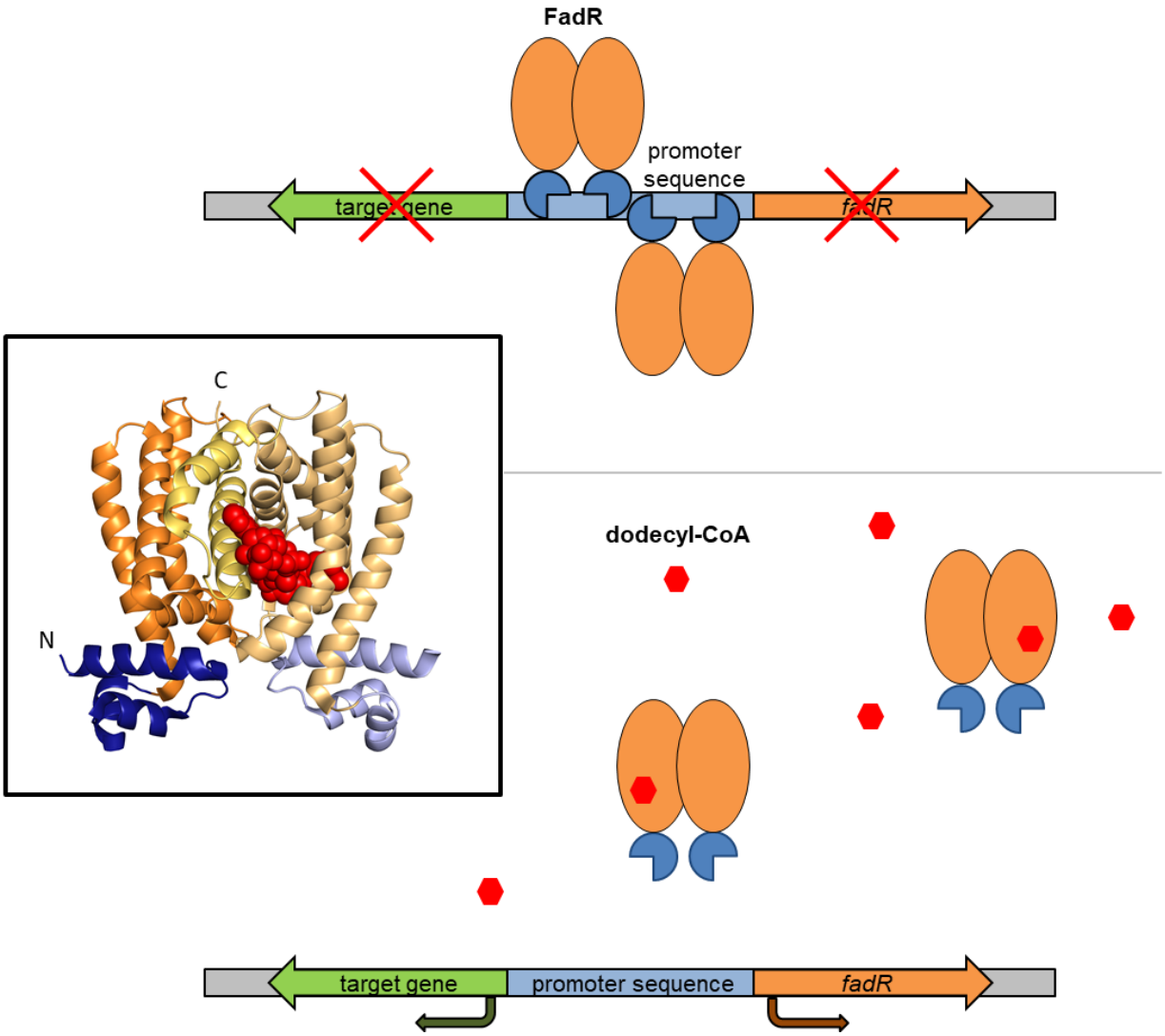


Figure 5.4 Regulatory mechanism of TFR repressor FadR. Substrate-free FadR (orange/blue) engages the intergenic DNA region. The FadR dimer is boxed with DBD and EBD domains indicated (blue and orange, respectively). The dimerisation interface is highlighted (yellow, α_8 and α_9) with the bound ligand, dodecyl-CoA, in red (⬢) (Fujihashi *et al.*, 2014).

A typical example of a TFR regulator is FadR from *Bacillus subtilis* (**Figure 5.4**). FadR represses five *fad* operons including 15 genes, most involved in β-oxidation of fatty acids (Fujihashi *et al.*, 2014). The tertiary structure of the EBD of FadR is helical, with two of the six helices engaged in the dimerisation interface (highlighted yellow in **Figure 5.4** (Yu *et al.*, 2010)). In the absence of any sequestered ligand, FadR dimers bind to palindromic nucleotide repeat sequences via the DBD in the intergenic region (**Figure 5.4** (Cuthbertson & Nodwell, 2013)). The binding of FadR to these regions is specifically inhibited by long-chain acyl-CoAs, which subsequently causes derepression of the *fad* operons (Matsuoka *et al.*, 2007). This ligand binding occurs at the dimer interface of the regulator protein, inducing a series of structural changes which increase the gap between the DNA recognition helices of the component monomers. This organisation is no longer compatible with DNA binding (Orth *et al.*, 1998; Yu *et al.*, 2010).

5.2.2. LysR-type transcriptional regulators

Nearly 20% of all bacterial transcriptional regulators appear to be of the LTTR family (Pareja *et al.*, 2006). Strains of *Acinetobacter*, *Agrobacterium*, *Escherichia*, *Pseudomonas*, and *Sinorhizobium* have genomes encoding 40-120 family members (Ezezika *et al.*, 2007), functioning as global activators or as repressors of single or operonic genes. LTTRs are observed to be divergently transcribed or located elsewhere on the bacterial chromosome (i.e. Types I and II (Heroven & Dersch, 2006; Hernandez-Lucas *et al.*, 2008)). LTTRs have well conserved structures but tend to display low sequence identity between family members as they have evolved to recognise an extremely diverse set of molecules (Russell *et al.*, 2004; Maddocks & Oyston, 2008). The N-terminal DBD has a characteristic winged helix-turn-helix DNA-binding motif followed by a linker helix (LH (Schell, 1993; Aravind *et al.*, 2005)). The EBD is made up of two subdomains of α/β fold (EBDI and EBDII), held together by hinge-like β-strands (**Figure 5.2** (Ezezika *et al.*, 2007; Choi *et al.*, 2001)). Generally, it is between these two subdomains that LTTRs bind their effector molecules. On binding, a conformational change alters the position of the LTTR on the promoter DNA, thereby enabling activation of transcription (Bundy *et al.*, 2002).

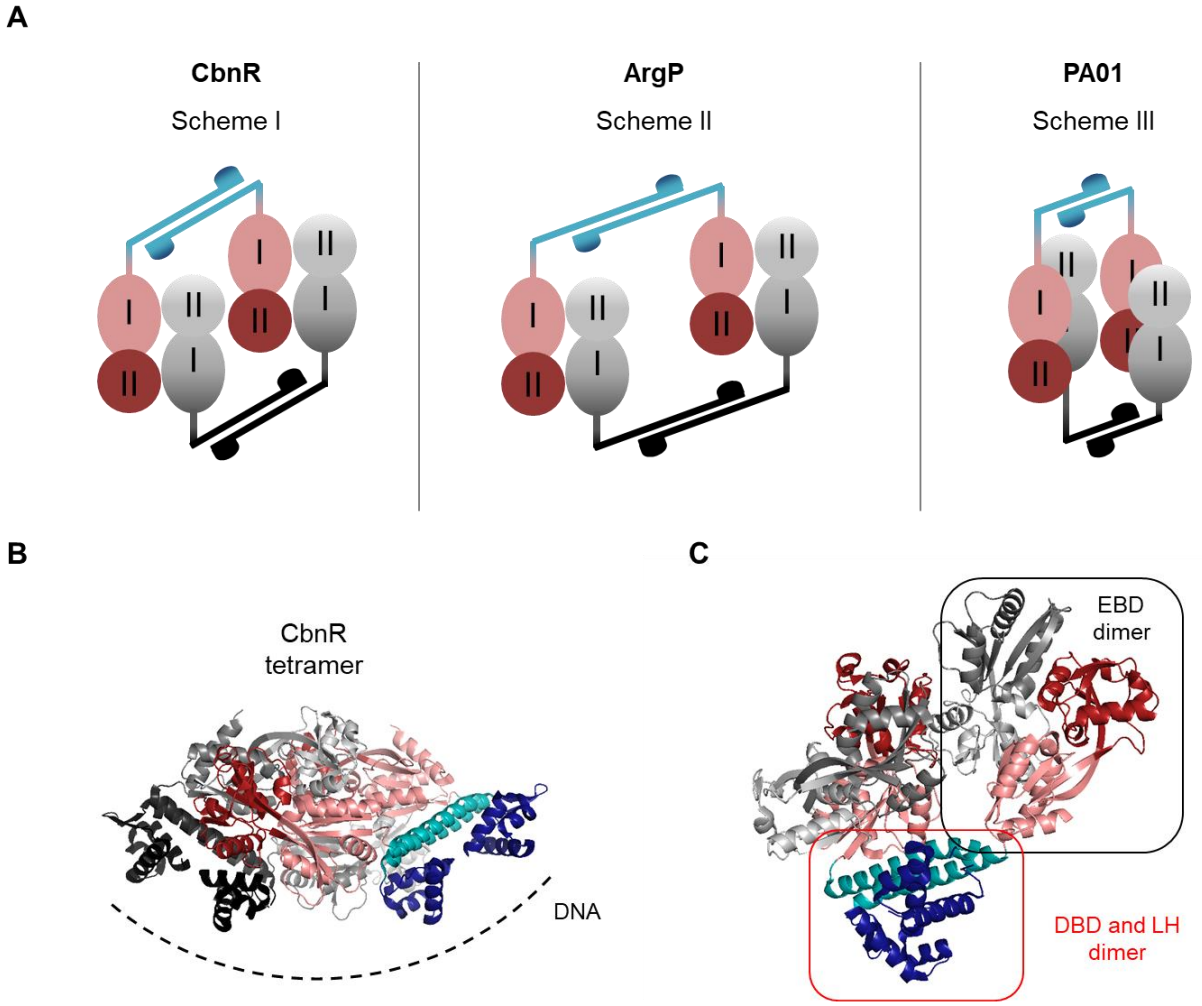


Figure 5.5 Three proposed models for LTTR oligomerisation (adapted from (Ruangprasert *et al.*, 2010)). (**A**) Representative proteins are CbnR (PDB 1IXC (Muraoka *et al.*, 2003)), ArgP (PDB 3ISP (Zhou *et al.*, 2010)) and PA01 (PDB 3FZV) with domains coloured as DBD (blue), LH (cyan), EBDI (pink), and EBDII (red). EBD dimers are between coloured and greyscale chains. (**B**) CbnR crystal structure shows the tetrameric arrangement needed to bind DNA (dotted line) and (**C**) the dimeric interfaces of an LTTR DBD and EBD (Muraoka *et al.*, 2003).

Unlike TFRs, LTTRs are active as tetramers, i.e. as a dimer of dimers (**Figure 5.5**). Three different oligomerisation schemes have been proposed for these regulators (Ruangprasert *et al.*, 2010). All evoke dimerisation of the EBD, but the proposed schemes vary as to the tetrameric interfaces in the presence of DNA (**Figure 5.5** (Ruangprasert *et al.*, 2010)). The specific contribution of stabilising forces from the DBD or the EBD for forming the LTTR tetramer in the presence of both ligand and promoter DNA remains under debate (Tropel & van der Meer, 2004).

Details are still emerging as to the complexity of LTTR mechanisms, with two distinct modes of regulation currently proposed (Picossi *et al.*, 2007). Regulators may bind to a primary site (effector-independent) involved in negative autoregulation, and a secondary proximal site (effector-dependent) to activate transcription. A distinct group of regulators is proposed to bind an additional binding site distinct from the secondary site for transcriptional activation as well as to the primary site (effector-independent). The effectors required for these two groups are thought to be different (Maddocks & Oyston, 2008). It is possible that some LTTRs, such as GtlC (*Bacillus subtilis*), perform both types of transcriptional regulation (Picossi *et al.*, 2007).

5.3. Efflux systems in the pathogenic organism *Acinetobacter baumannii*

The wide multidrug resistance of *Acinetobacter baumannii* has recently been attributed to the overexpression of efflux systems (Coyne *et al.*, 2011). RND efflux systems are the most prevalent in *A. baumannii* genomes and have a role in host colonisation and persistence of infection. These RND pumps have been demonstrated to actively export host-derived antimicrobials (Join-Lambert *et al.*, 2001; Hirakata *et al.*, 2002).

Efflux systems currently characterised in *A. baumannii* are summarised in **Table 5.1**, grouped according to their associated transporter. Four of these are reviewed in detail below: two RND transporters, AdeIJK and AdeFGH; the MFS pump, AmvA; and a newly defined PACE element, AceI. Such efflux systems are obviously important when addressing the pathogenicity of *A. baumannii*.

Table 5.1 Known efflux pumps in *A. baumannii* and their identified regulator proteins

transporter family [a]	transporter	possible substrates	regulators	regulator family [b]	references
RND	AdeABC	aminoglycosides, fluoroquinolones, erythromycin, tetracycline, trimethoprim, chloramphenicol	AdeR AdeS	RR HK	(Magnet *et al.*, 2001; Marchand *et al.*, 2004)
	AdeIJK	β-lactams, chloramphenicol, tetracycline, erythromycin, lincosamide, fluoroquinolones, fusidic acid, rifampicin, trimethoprim, novobiocin, clindamycin, acridine, safranin, pyronine, and SDS	AdeN	TFR	(Damier-Piolle *et al.*, 2008; Rosenfeld *et al.*, 2012)
	AdeFGH	chloramphenicol, ciprofloxacin, trimethoprim, fluoroquinolones, tetracycline-tigecycline, clindamycin, sulfamethoxazole, SDS, and dyes such as ethidium bromide, safranin O, and acridine orange	AdeL	LTTR	(Coyne *et al.*, 2010b)
MFS	AmvA	dyes, disinfectants, detergents, erythromycin	AmvR	TFR	(Rajamohan *et al.*, 2010b; Hassan *et al.*, 2016)
	CraA	chloramphenicol	unknown		(Roca *et al.*, 2009)
MATE	AbeM	aminoglycosides, fluoroquinolones, chloramphenicol, trimethoprim, ethidium bromide, dyes	unknown		(Su *et al.*, 2005)
SMR	AbeS	chloramphenicol, fluoroquinolones, erythromycin, novobiocin, dyes, detergents	unknown		(Srinivasan *et al.*, 2009)
PACE	AceI	chlorhexidine	AceR	LTTR	(Farrugia *et al.*, 2013)

[a] resistance nodulation division (**RND**), major facilitator superfamily (**MFS**), multidrug and toxic-compound extrusion (**MATE**), small multidrug resistance (**SMR**), proteobacterial antimicrobial compound efflux (**PACE**).
[b] response regulator (**RR**), histidine kinase (**HK**), TetR family regulator (**TFR**), LysR-type transcriptional regulator (**LTTR**).

5.3.1. The RND transporter system AdeIJK

In 2007, a team at the Institut Pasteur identified and characterised three genes from *A. baumannii* BM4454; *adeI*, *adeJ*, and *adeK*, proposed to encode an RND efflux system (Damier-Piolle *et al.*, 2008). This system cotranscribes three proteins: a membrane fusion protein, AdeI; a transporter protein, AdeJ; and an outer membrane protein, AdeK. The three genes occur within a distribution of *A. baumannii* strains. As with other RND transporters, AdeIJK contributes to resistance of chemically important antibiotics (listed, **Table 5.1**). Recent transcriptomic experiments and quantitative real-time reverse transcription PCR have established that levels of *adeIJK* overexpression are low compared to those of other RND transporters, suggesting AdeIJK to be tightly regulated (Coyne *et al.*, 2010b).

During initial studies of AdeIJK, no regulatory genes were identified in the vicinity of the efflux genes (Damier-Piolle *et al.*, 2008). However, following whole-genome sequencing, *adeN* was identified in a susceptible strain of *A. baumannii* (BM4587A) in which a premature stop codon resulted in deletion of the 7 C-terminal residues of the protein, rendering it inactive (Rosenfeld *et al.*, 2012). The regulatory element *adeN* is located 813 kbp from *adeIJK* in *A. baumannii* strain ATCC17978, as indicated in panel **A** of **Figure 5.6**.

AdeN is ascribed to the TetR family of transcriptional regulators and acts to repress expression of the relevant *adeIJK* operon (Rosenfeld *et al.*, 2012). Unlike most other TFRs, it is constitutively expressed (i.e. a Type III mechanism, **Figure 5.3**) and does not regulate its own expression (Ramos *et al.*, 2005). At this point the possibility that AdeN could be involved in the regulation of other genes cannot be excluded, nor the notion that more than one regulator protein controls expression of AdeIJK.

For this reason, more details concerning this transport system are needed, incorporating structural analysis and binding assays of the proteins *in vitro*. In addition, it would be useful to know where AdeN specifically binds within the *A. baumannii* genome to better understand its regulatory roles in the cell, which could be as extensive as FadR (discussed above).

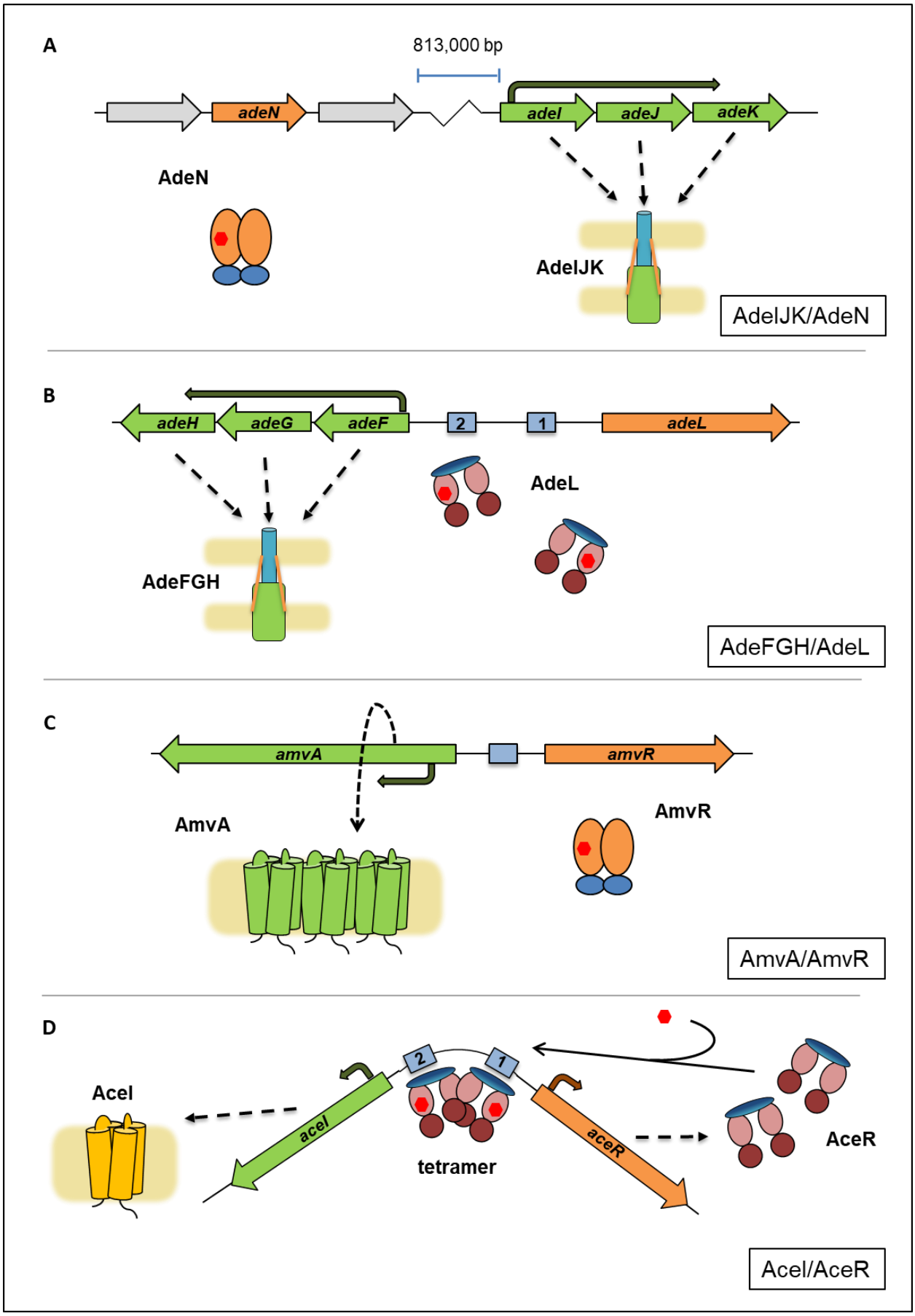

A
813,000 bp
adeN
adeI
adeJ
adeK
AdeN
AdeIJK
AdeIJK/AdeN
B
adeH
adeG
adeF
2
1
adeL
AdeL
AdeFGH
AdeFGH/AdeL
C
amvA
amvR
AmvA
AmvR
AmvA/AmvR
D
AceI
acel
2
1
tetramer
aceR
AceR
AceI/AceR

Figure 5.6 Mechanisms of transcriptional regulation in the presence of efflux specific substrates. (**A**) *adeN* is a constitutively expressed transcriptional repressor of *adeIJK*. When AdeN binds to a substrate (⬢), expression of the RND AdeIJK pump is allowed (Rosenfeld *et al.*, 2012). (**B**) *adeL* is a divergently transcribed repressor of *adeFGH*. Expression of the RND AdeFGH efflux pump is allowed when AdeL is bound to a substrate, as it disengages from the promoter region (blue rectangles (Brzoska *et al.*, 2013)). (**C**) *amvR* is a divergently transcribed repressor of *amvA*. When AmvR binds to a substrate, expression of the MFS transporter, AmvA, is allowed (Brzoska *et al.*, 2013). (**D**) *aceR* is a divergently transcribed activator of the *aceI* efflux gene. When AceR binds to its substrate, it in turn binds in tetrameric form to the intergenic promoter region and induces expression of both AceI and AceR proteins (Qi Liu, 2017).

5.3.2. The RND transporter AdeFGH

By studying *A. baumannii* mutants selected in the presence of chloramphenicol and norfloxacin, Perichon and co-workers recently identified a new RND efflux system, named AdeFGH (Coyne *et al.*, 2010a). Inactivation of *adeFGH* in one mutant strain restored antibiotic susceptibility, indicating AdeFGH to be responsible (Coyne *et al.*, 2010b). An open reading frame, encoding an LTTR named *adeL*, was located upstream from the *adeFGH* operon, and found transcribed in the opposite direction (an example of Type I organisation, **Figure 5.6 B**).

Mutations in *adeL* were found in three *adeFGH*-overexpressing mutants, suggesting them to be responsible for overexpression of AdeFGH (Coyne *et al.*, 2010b). Subsequent inactivation of *adeL* resulted in overexpression of the transporter genes, indicating it to be the transcriptional repressor of the AdeFGH efflux pump (Brzoska *et al.*, 2013). It remains to be seen if this is the only regulator for this system. Strains overexpressing *adeFGH* with no mutation of the cognate regulator gene, *adeL*, or its promoter region have been identified, indicating additional control by unknown regulators (Yoon *et al.*, 2013).

5.3.3. The MFS transporter AmvA

Transcription of the *amvA* gene has been found to be elevated in *A. baumannii* clinical isolates that exhibit very high MICs towards carbapenems, cephalosporins, aminoglycosides and

fluoroquinolones (Rajamohan *et al.*, 2010a). AmvA is a member of the MFS family of efflux pumps and exhibits an ability to export a wide range of compounds (**Table 5.1**). The same gene was found to be inactivated in an MDR clinical isolate of *A. baumannii*, and its role in conferring resistance to broad-spectrum compounds subsequently confirmed (Rajamohan *et al.*, 2010a).

Paulsen's team at Macquarie University (Sydney) have since identified the regulator for AmvA. The gene *amvR*, with the organisation of a TFR, was found adjacent to and divergently transcribed from *amvA* in the *A. baumannii* BAL062 chromosome (**Figure 5.6 C** (Hassan *et al.*, 2016)). This is consistent with a Type I transcriptional mechanism. The level of *amvA* expression in a transposon insertion strain was 6-fold higher than the parental strain, indicating AmvR to control expression of *amvA* (Brzoska *et al.*, 2013). AmvR acts to repress *amvA* expression (Hassan *et al.*, 2016).

Given the large range of possible substrates for the AmvA/AmvR system, it is important to examine the diversity of chemical compounds to which the system can respond, including the regulatory protein. Although the MFS efflux pump is relatively simple due to its single component, it remains relevant to study the soluble regulator protein it incorporates.

5.3.4. The PACE transporter AceI

Alongside the recent discovery of the PACE transporter family in *A. baumannii*, outlined above (Section 5.1.2), the gene for transporter AceI was observed to be highly overexpressed under chlorhexidine stress (Farrugia *et al.*, 2013). This transporter appears to be widespread, with orthologs of the *aceI* gene strongly conserved across a broad range of proteobacteria. AceI is encoded adjacent to a divergently-transcribed LTTR, and deletion mutants of the regulator gene created in *A. baylyi* strains suggest the LTTR protein functions as an activator. Expression of *aceI* (and its orthologs in *E. coli*) are repeatedly linked to resistance to chlorhexidine (Farrugia *et al.*, 2013).

It has also been shown that purified AceR protein binds directly to the *aceI-aceR* intergenic region in the presence of chlorhexidine (**Figure 5.6 D**). Electrophoretic mobility shift assays and DNase

I footprinting assays demonstrated chlorhexidine-stimulated binding of AceR with two sites upstream of the putative *aceI* promoter (Qi Liu, 2017). Together, these results indicate that AceR is a locally encoded activator of *aceI* expression that positively autoregulates. The AceI/AceR regulatory system is reasonably well defined in regards to regulator binding capacities. However, it is unlikely that AceI only exports one substrate. Regulatory mechanisms of orthologous genes appear to respond to a broad range of compounds (including acriflavine, proflavine, benzalkonium and dequalinium), suggesting that AceI has this increased binding capacity as well. By testing the affinity of AceR to a wide array of chemically and structurally different compounds, it may be possible to discover new substrates for the AceI/AceR efflux system.

5.4. Experimental aims

Although there has been considerable research into gene expression and regulation of each of the distinct efflux systems AdeIJK, AdeFGH, AmvA, and AceI, there has been little examination of the regulator proteins with which they are linked. Of the eight efflux pumps characterised in *A. baumannii*, only five (AdeABC, AdeIJK, AdeFGH, AmvA and AceI) have their specific regulators identified, and these are only very basically characterised. There has been some investigation on the binding of AceR to intergenic regions of DNA and binding of possible ligands such as chlorhexidine (Qi Liu, 2017), but a comprehensive study of multiple regulators to this component has yet to be performed. Molecular knowledge of the regulator proteins AdeN, AmvR, AdeL, and AceR is updated in this thesis.

The role of regulator proteins in bacterial efflux, addressed by myself and others in this field of research, has never been explored in *A. baumannii* and may provide graphic insight into the currently unexplained mechanisms underpinning the rapid rise of this pathogen. In this thesis, each regulator protein was produced for the first time in recombinant form, so allowing *in vitro* study as to biophysical characteristics such as oligomerisation and stability in solution. Following this, structurally sound regulator proteins were screened for affinity towards a panel of small, suitable ligands. This method of separate *in vitro* study is common for analysing binding

mechanisms of regulator proteins. For example, FadR, from *Bacillus subtilis*, is inhibited from binding intergenic DNA in the presence of long chain acyl-CoAs with 14-20 carbon atoms. These *in vitro* findings were supported by the *in vivo* observations that the knockout of acyl-CoA dehydrogenation resulted in FadR inactivation, due to the accumulation of long chain acyl-CoAs in the cells (Matsuoka *et al.*, 2007).

CHAPTER 6

Solution Characterisation of *A. baumannii* Efflux-pump Regulator Proteins

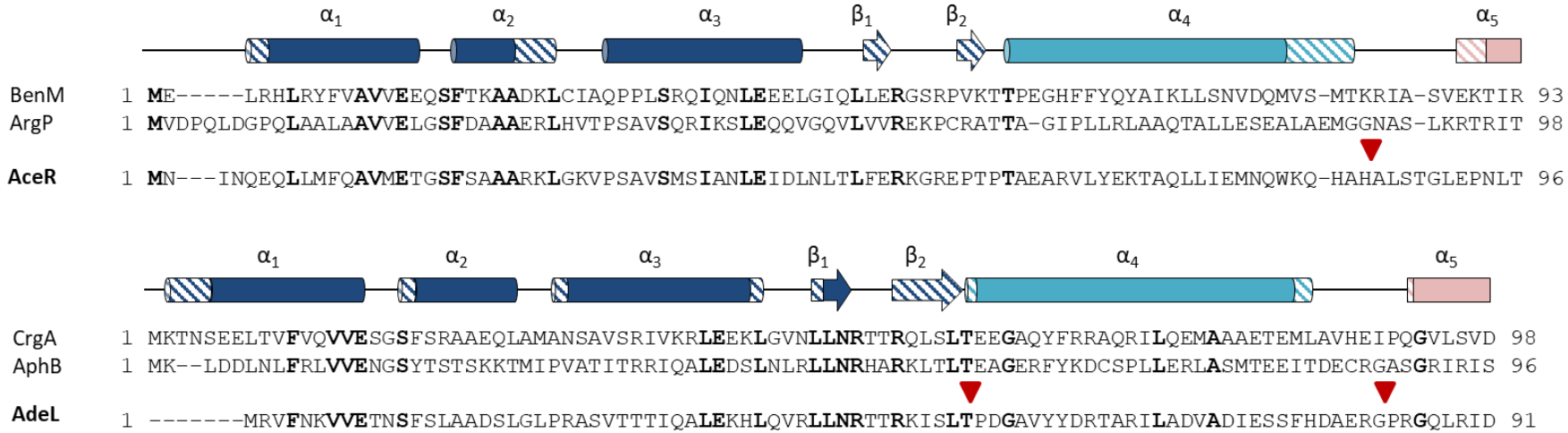
Four efflux regulator proteins were examined in this thesis in order to provide a first biophysical characterisation and to better identify their ability to interact with multiple effector molecules. To achieve this, regulators AdeN, AmvR, AdeL, and AceR were produced by recombinant means via heterologous expression in *E. coli*, as previously utilised for other TFR and LTTR proteins with great success (Ezezika *et al.*, 2007). The production of pure protein *in vitro* aimed to clearly determine solution-state oligomerisation and ligand binding affinities of four previously uncharacterised efflux regulators.

6.1. Design of protein constructs

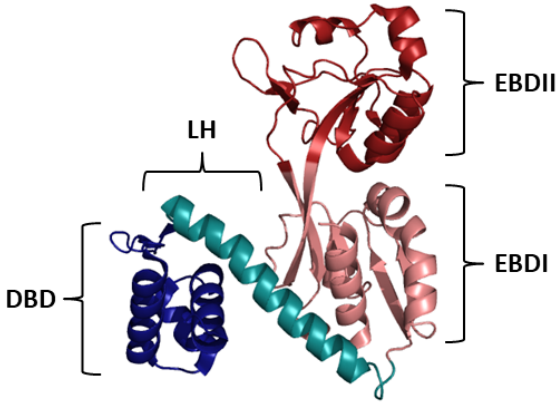
6.1.1. AdeL and AceR

Historically, functional and structural studies of LTTR proteins have been impeded by aggregation and low solubility. As LTTR proteins are composed of distinct domains, however, it becomes possible to separate integral domains and still have a functioning and folded protein entity. Successful preparation of LTTR systems *in vitro* has involved removal of the N-terminal sequence encoding the DBD, as well as the use of high-salt buffers (Ezezika *et al.*, 2007). In addition, any C-terminal histidine tag has been shown not to interfere with protein function *in vivo*. Of 85 structures classified as LysR-type regulators in the PDB (as of December 2017), only 17 are full length (Berman *et al.*, 2000). The remaining structures define isolated EBD (63 examples) or, to a lesser extent, DBD truncations (5 examples).

A


α1 α2 α3 β1 β2 α4 α5
BenM 1 ME-----LRHLRYFVAVVEEQSFTKAADKLCIAQPPLSRQIQNLEEELGIQLLERGSRPVKTTPEGHFFYQYAIKLLSNVDQMVS-MTKRIA-SVEKTIR 93
ArgP 1 MVDPQLDGPQLAALAAVVELGSFDAAAERLHVTPSAVSQRIKSLEQQVGQVLVVREKPCRATTA-GIPLLRLAAQTALLESEALAEMGGNAS-LKRTRIT 98
AceR 1 MN---INQEQLLMFQAVMETGSFSAAARKLGKVPSAVSMSIANLEIDLNLTLFERKGREPTPTAEARVLYEKTAQLLIEMNQWKQ-HAHALSTGLEPNLT 96
α1 α2 α3 β1 β2 α4 α5
CrgA 1 MKTNSEELTVFVQVVESGSFSRAAEQLAMANSAVSRIVKRLEEKLGVNLLNRTTRQLSLTEEGAQYFRRAQRILQEMAAAETEMLAVHEIPQGVLSVD 98
AphB 1 MK--LDDLNLFRLVVENGSYTSTSKKTMIPVATITRRIQALEDSLNLRLLNRHARKLTLTEAGERFYKDCSPLLERLASMTEEITDECRGASGRIRIS 96
AdeL 1 -------MRVFNKVVETNSFSLAADSLGLPRASVTTTIQALEKHLQVRLLNRTTRKISLTPDGAVYYDRTARILADVADIESSFHDAERGPRGQLRID 91

B **LTTRs**


EBDII
LH
EBDI
DBD

C

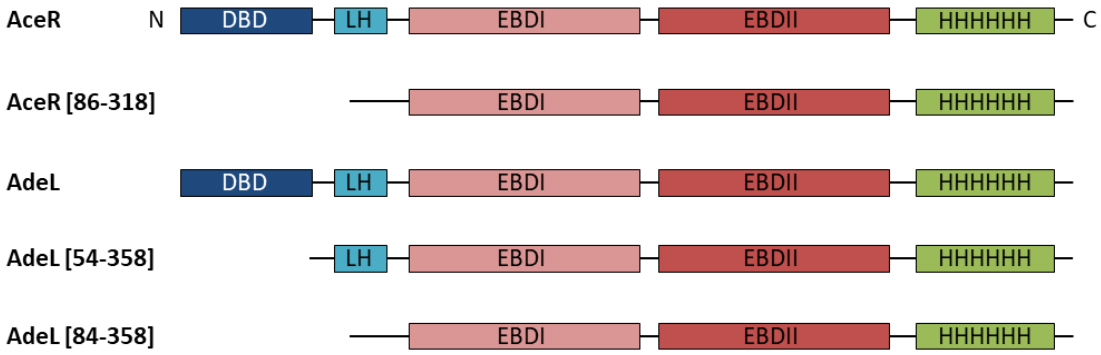

AceR N DBD LH EBDI EBDII HHHHHH C
AceR [86-318] EBDI EBDII HHHHHH
AdeL DBD LH EBDI EBDII HHHHHH
AdeL [54-358] LH EBDI EBDII HHHHHH
AdeL [84-358] EBDI EBDII HHHHHH

Figure 6.1 Identification of DNA-binding domains for LTTR proteins AceR, AdeL, and homologs. All protein domains depicted are coloured as follows: DBD (dark blue), LH (light blue), EBDI (pink), and EBDII (red). (**A**) Sequence alignments include BenM (*A. baylyi*. ADP1, PDB 3K1N (Craven *et al.*, 2009)), ArgP (*M. tuberculosis*, PDB 3ISP (Zhou *et al.*, 2010)), CrgA (*N. meningitidis*, PDB 3HHG (Sainsbury *et al.*, 2009)) and AphB (*V. cholerae*, PDB 3SZP (Taylor *et al.*, 2012)). Alignments are based on output by Expresso (Armougom *et al.*, 2006), with fully conserved residues across each sub-group highlighted in bold. Red triangles indicate truncation points engineered for AceR and AdeL. Hashed tertiary features indicate variation of secondary structure between homologs. (**B**) Cartoon of CrgA (PDB 3HHG) labelled according to the general organisation of LTTR proteins. (**C**) Representation of all AdeL and AceR full-length and truncated constructs engineered for this work, with His_6-tags indicated (green).

Accordingly, for successful production of AdeL and AceR for this study, it was important to first identify likely domain boundaries within each primary sequence. This design process was guided by several well-defined crystal structures available for LTTR proteins closely related in sequence. For AceR, a sequence search across all PDB depositions identified homologs BenM (*A. baylyi* (Craven *et al.*, 2009)) and ArgP (*Mycobacterium tuberculosis* (Zhou *et al.*, 2010)) as suitable design templates (with 27 and 25% sequence identity, respectively). For AdeL, a similar search identified CrgA (*Neisseria meningitidis* (Sainsbury *et al.*, 2009)) and AphB (*Vibrio cholerae* (Taylor *et al.*, 2012)) as close homologs (with 34-33% identity). Sequence alignments of the relevant N-terminal sequence segments across these two protein groups are depicted in **Figure 6.1 A**.

As outlined in **Figure 6.1 B**, which depicts the crystal structure of CrgA, the three helices of the DBD are generally separated from the succeeding linker helix (LH) by a loop element incorporating short strand segments. As the LH is again separated from the subsequent EBD by another (generally shorter) loop, this LH feature serves as a useful truncation region with limited perturbation to tertiary structure of the DBD or EBD. From the sequence of AceR, Ala-86 was therefore chosen as the truncation point, as it likely resides in this inter-domain region (**Figure 6.1 C**). The preceding sequence to this site is consistent with an amphipathic helix of a likely LH element. This would create a construct AceR [86-318] in which both DBD and LH segments are absent. Other designs for truncated forms of AceR did not prove viable and are therefore not discussed. For AdeL, two truncation points were located in the primary sequence before and after

the LH region: Pro-54 and Pro-84 (**Figure 6.1 A**). By creating two constructs, AdeL [54-358] and AdeL [84-358], the impact of the LH on the behaviour of the EBD portion of AdeL was able to be investigated. The four variants of LTTR proteins utilised in this work are depicted in **Figure 6.1 C**.

6.1.2. AdeN and AmvR

Regulator proteins AdeN and AmvR are members of the TFR family, which tend to express more readily than LTTR proteins. This is reflected in the larger number of deposited crystal structures: 246 depositions in the PDB include TetR regulator in their description (as of December 2017 (Berman *et al.*, 2000)).

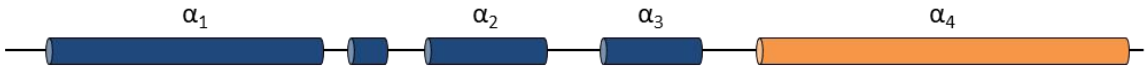
With a two-domain architecture for TFR proteins lacking any linking segment (**Figure 6.2**), variants of AdeN and AmvR, without the DBD sequence portion were less straightforward to design. A sequence search was carried out for AdeN and AmvR against deposited PDB structures, with resulting sequence homologs outlined in **Figure 6.2 A**. For AdeN, structures for CmeR (*Campylobacter jejuni* (Gu *et al.*, 2007)) and MLR_4833 (*Mesorhizobium japonicum*), with sequence identity 25-26%, served as design templates. Although there is high variation at the N-terminus of helix 1, strong homology across an amphipathic segment of the AdeN sequence suggests helices of the DBD to end at the motif [F-G-N-K-D-G]. Due to this suggestion of a turn feature, the truncation point chosen, Gly-65, was anticipated to precede the EBD, creating the AdeN [65-238] variant.

For AmvR, one full-length homolog with sequence identity > 25% occurs in the PDB, a putative transcriptional regulator from *Saccharomonospora viridis* (PTR_SV, 26% identity). However, this specific homolog is of immediate interest, as there is relatively strong preservation of sequence at the boundary between the DBD and EBD. From the sequence organisation, it appears the Pro-Ser sequence, present in AmvR as Phe-Ser-Ser serves as a potential interface segment, and therefore, an appropriate truncation point (**Figure 6.2 C**). Thus, the construct AmvR [51-213] was prepared, designed to contain a full sequence encoding only the EBD of this TFR. **Figure 6.2 C** summarises the four TFR protein constructs utilised in this project.

A


```
CmeR     1 MNSNRTPSQ----------KVLARQEKIKAVALELFLTKGYQETSLSDIIKLSGGSYSNIYDGFKSKEGLFF 62
MLR_4833 1 GMKID-GET----------RSARKDREIIQAATAAFISKGYDGTSMEEIATKAGASKQTVYKHFTDKETLFG 61

AdeN     1 MHDPVLESHHLVCEKPQTRRGIERRLALLLSATELFLEKGYDAVSLDDIVNHAGGSKTSIYKYFGNKDGLFT 72
```

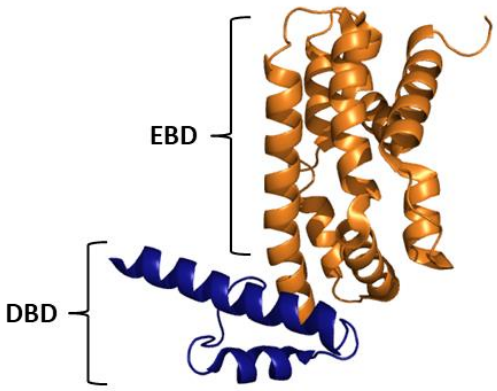



```
PTR_SV  93 PQSEARRRILETAWRLIARRGYHNVRIHDIASELGTSNATIHYHFPSKKDILLEALRRNVKLAFDRQVAELH 164

AmvR     6 -RDQRREMILQAAMQIALTEGFTAMTVRRIATEAQTSTGQVHHHFSSASHLKAEAFLKLME-----QLDEIE 71
```

B

TFRs



C

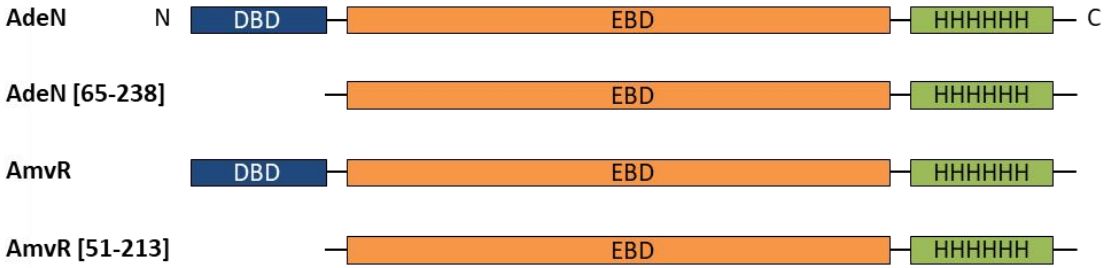


Figure 6.2 Identification of DNA-binding domains for TFR proteins AdeN, AmvR and homologs. All protein domains depicted are coloured as follows: DBD (dark blue) and EBD (orange). (**A**) Sequence alignments include CmeR (*C. jejuni*, PDB 2QCO (Gu *et al.*, 2007)), MLR_4833 (*M. japonicum*, PDB 3BHQ), and a putative transcriptional regulator (PTR_SV, *S. viridis*, PDB 4ICH). Alignments are based on output by Expresso (Armougom *et al.*, 2006), with fully conserved residues across each sub-group highlighted in bold. Red triangles indicate truncation points engineered for AdeN and AmvR. Hashed tertiary features indicate variation of secondary structure between homologs. (**B**) Cartoon of CmeR (PDB 2QCO) labelled according to the organisation of TFR proteins. (**C**) Representation of all AdeN and AmvR full-length and truncated constructs engineered for this work, with His_6-tags indicated (green).

6.1.3. Selection of system for heterologous expression

Members of both LTTR and TFR protein families are known to undergo conformational rearrangement upon binding to cognate DNA and effector ligands, with their DBDs particularly mobile (Ezezika *et al.*, 2007; Yu *et al.*, 2010). For this reason, it was preferable not to add an affinity tag to the relatively mobile N-terminal portions of the constructs, and they were instead cloned into a vector coding for a C-terminal His_6-tag. In this way, resulting recombinant products should have retained unimpeded capacity to bind DNA. This mirrors the C-terminal tagging previously utilised in recombinant production of these families (e.g. BenM, and CrgA (Craven *et al.*, 2009)).

The plasmid $pTTQ18_{RGSH6}$ MCS (Stark, 1987) was therefore utilised to provide this affinity purification sequence feature (see Section 8.1.3). In addition, a Tobacco Etch Virus (TEV) cleavage site (ENLYFQ^S (Kim *et al.*, 2011)) was included between the regulator gene and message for the affinity tag in order to facilitate enzymatic excision of the C-terminal His_6 sequence, if necessary. As $pTTQ18_{RGSH6}$ contains a tac promoter, *E. coli* BL21 competent cells were chosen for protein expression. This widely-used strain does not express the T7 RNA polymerase and is therefore compatible with this expression vector (see url: www.neb.com/products/c2530-bl21-competent-e-coli).

6.2. Production of proteins of the LTTR and TFR family

6.2.1. Amplification and cloning of genes

All genes required (two forms of AdeN, two forms of AmvR, three forms of AdeL, and one truncated form of AceR) were amplified from *A. baumannii* chromosomal DNA by PCR using primers outlined in **Table 8.5** (see Section 8.2.1.3). Full-length AceR was a kind gift of Qi Liu, Macquarie University. **Figure 6.3 A** shows the successfully amplified genes, visualised on an agarose gel. Sizes corresponded to 675 bp for AdeN, 600 bp for AmvR, and shorter forms for truncated variants. All were confirmed by sequencing. Gene products were not initially recovered by this PCR step for AceR (full-length) and the three AdeL constructs. To rectify this, commercial gBlock sequences were purchased for amplification of AdeL and its truncated forms (AdeL [84-358] and AdeL [54-358]). These amplified genes are indicated in **Figure 6.3 A**. Full-length AdeL still proved difficult to amplify. Reasons for failed PCR of the AceR and AdeL genes could be due to forward primer design, as they contained the lowest GC content out of the panel (35 and 38% respectively). Further study of these two constructs was not continued.

Eight genes were cloned into pTTQ18 vectors and checked for insertion using colony PCR. Primers designed to specifically target inserted genes were used to determine if the construct contained these DNA fragments of interest. Some of the amplified fragments recovered are depicted in **Figure 6.3 B**, and all results indicated correctly transformed colonies. Vectors were extracted and transformed into BL21 *E. coli* for expression.

6.2.2. Expression and purification of full-length and truncated proteins

Transformants were propagated within small-scale cultures, and these used to inoculate auto-induction media (500 ml) as described in Section 8.2.2.1. Cell pellets recovered following growth were lysed, and soluble protein products recovered by IMAC (Section 8.2.2). Recovered protein material is visualised by SDS-PAGE in **Figure 6.4** and allows comparison of expression levels. All protein products were evident at the expected size within the range of 19-35 kDa (see chemical properties of designed protein products tabulated in **Table 6.1**).

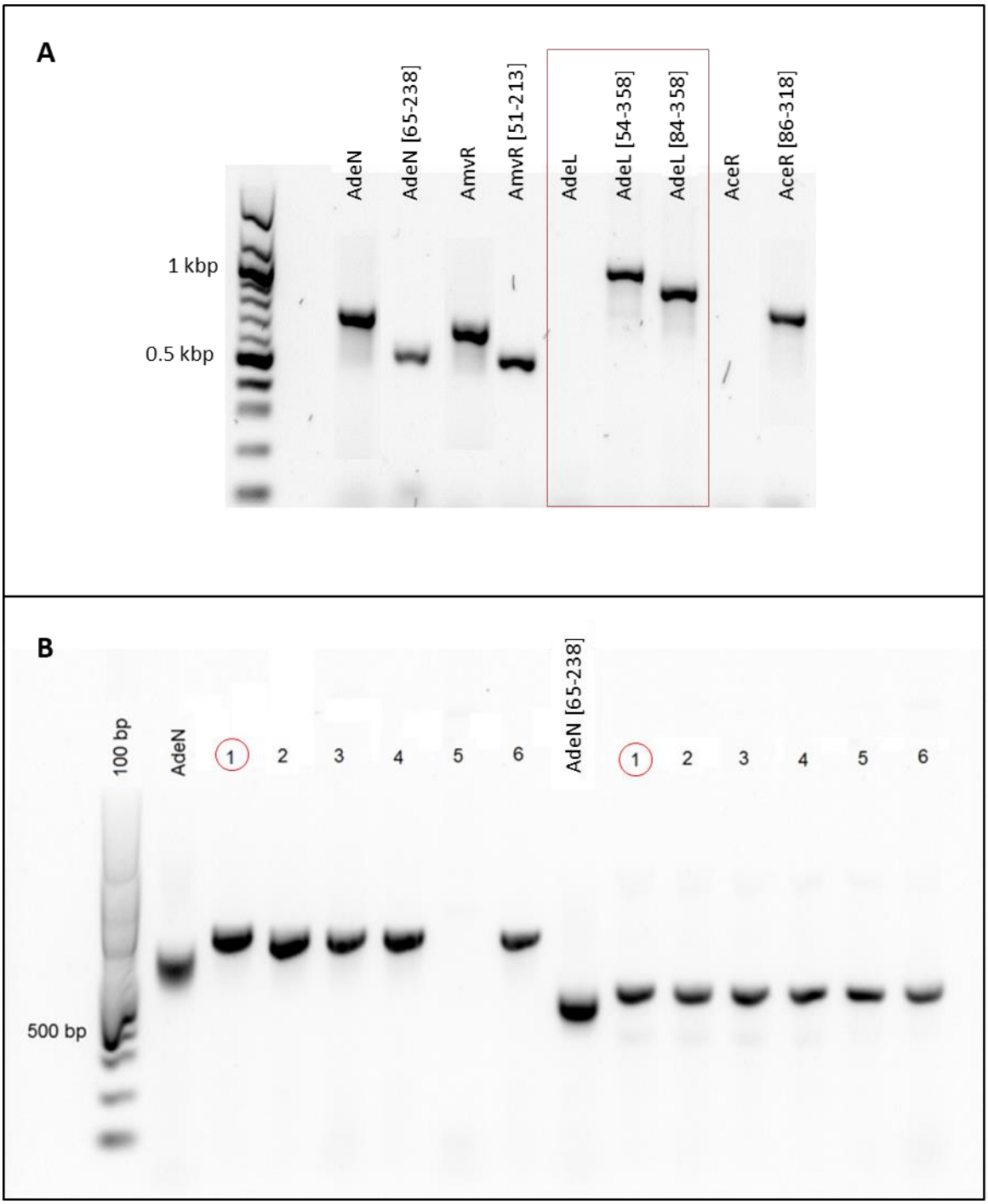


Figure 6.3 Agarose gel electrophoresis (1.2%) identifying gene products. (**A**) Amplification of genes derived from *A. baumannii* genomic DNA (ATCC 17978) or (boxed in red) from commercially synthesised gBlocks. (**B**) Example genes excised by colony PCR for six selected AdeN and AdeN [65-238] constructs. Lanes with gene name above contain DNA from genomic amplification. Circles indicate colonies sent for sequence confirmation.

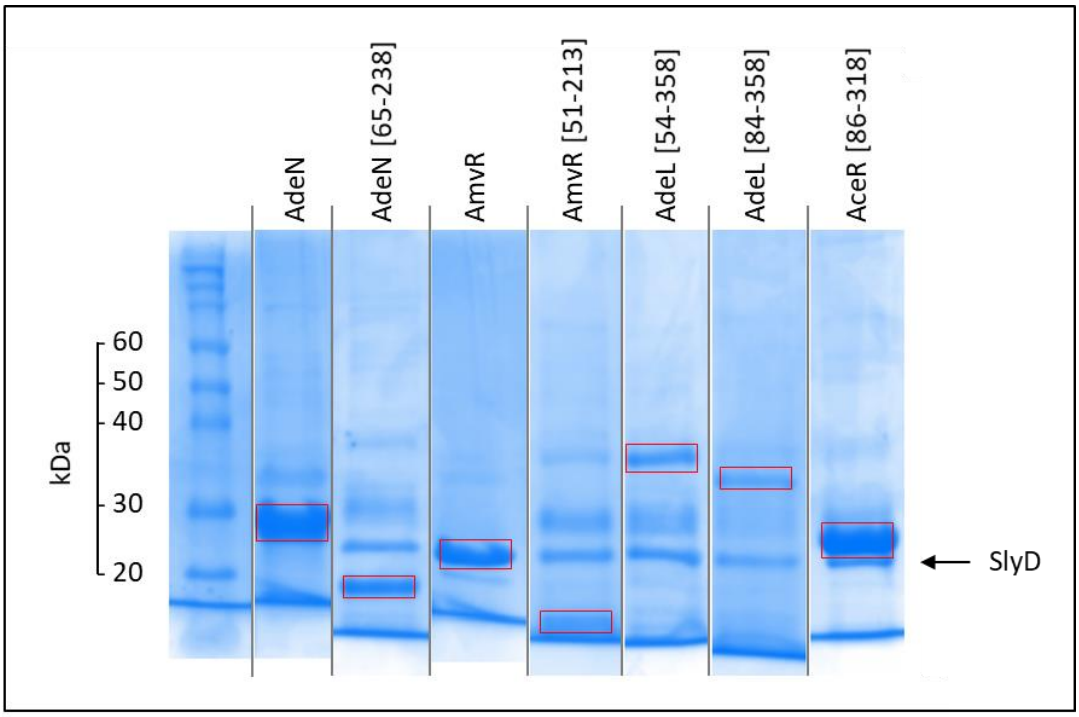


Figure 6.4 Reducing SDS-PAGE of IMAC purified recombinant LTTR and TFR products from separate preparations. Target protein mass values are boxed in red. Arrow indicates host cell protein SlyD (21 kDa) evident across all samples. Gel stained with Coomassie dye. M_R scale determined from commercial standards is indicated to the left.

Overall, the recovery of required products was generally good, with bands corresponding to a few contaminant proteins appearing more strongly in samples of lower yielding product. These likely are due to His-rich *E. coli* host cell proteins such as the 21 kDa protein SlyD (Robichon *et al.*, 2011; Andersen *et al.*, 2013). Recombinant AdeN and AceR [86-318], seen as bands at 28 and 24 kDa, respectively, were obtained at highest yield, measured as 6.5 mg/ml and 5 mg/ml. Relatively low yield was evident for AmvR [51-213], appearing as a minor component at 20 kDa within the recovered IMAC eluent. This particular production was not at high enough yield to be useful, so the material was not pursued in this study. However, seven regulator proteins remained viable, including full length AmvR.

Table 6.1 Physico-chemical properties of regulator proteins [a]

	protein construct	repressor (-) or activator (+)	# amino acids	# cysteines	molecular mass (kDa) [b]	pI [b]	observed native mass (kDa) [c]	T_M (°C) [d]
TFR family	AdeN	–	238	8	27.5	6.33	45	nd
	AdeN [65-238]		175	7	20.4	6.28	40	nd
	AmvR	–	213	3	24.6	6.05	50	52
	AmvR [51-213]		164	3	19.0	5.69	nd	nd
LTTR family	AdeL	–	358	6	40.1	6.73	nd	nd
	AdeL [54-358]		306	6	34.3	6.05	40, 70	50
	AdeL [84-358]		276	6	31.0	6.49	50	50 & 62
	AceR	+	318	3	36.6	6.01	75	53
	AceR [86-318]		234	3	27.1	5.89	55	64

[a] Including 21 residue His_6-tag and TEV site added by primers and pTTQ18 vector. "nd" indicates values not determined
[b] Molecular mass and pI calculated using Protparam algorithm.
[c] From SEC, pH 7.5 (50 mM HEPES, 300 mM NaCl, 10% glycerol)
[d] T_M determined by DSF in this work (Section 7.3).

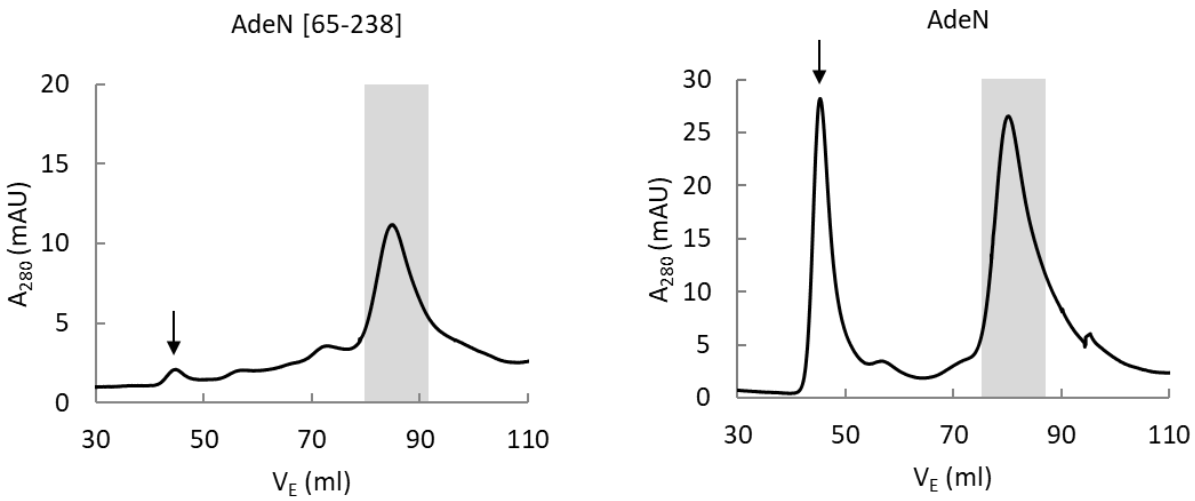


Figure 6.5 Elution trace for preparative SEC carried out on Superdex 200, pH 7.5 for preparations of recombinant AdeN and AdeN [65-238]. Elution volumes (V_E) are indicated for column (120 ml) operating at 1 ml/min in HEPES buffer (50 mM) with 300 mM NaCl and 10% glycerol. Column void volume (V_0) is arrowed. Shaded panels indicate protein fractions recovered and combined for further analysis.

SEC was used as a polishing step for each protein preparation, for buffer exchange and to remove any aggregates following IMAC elution. The four truncated variants tended to elute as a single peak from each preparative SEC column, as seen for AdeN [65-238] in **Figure 6.5**, which eluted at 80-90 ml (consistent with a kDa range 30 to 40 indicating a dimer). For samples of AdeN and AmvR, a large proportion of material eluted in the void, indicative of aggregates exceeding 2000 kDa (**Figure 6.5**). This was most likely due to instability caused by the presence of the N-terminal DBD in these recombinant products. However, even after removal of aggregates, full-length proteins AdeN and AmvR were recovered at higher concentrations.

6.3. Evaluation of quaternary state and stability in solution

The oligomerisation state of efflux pump regulator proteins is integral to the bioactivity of these molecules. TFRs are active as dimers, and generally form dimers in solution (Yu *et al.*, 2010),

with a dimerisation interface engaging paired helical elements of each EBD (see Section 5.4.1). LTTRs exhibit more complex behaviour, occurring as dimers when inactive, yet acting as tetramers to bind both DNA and effector molecules (Ruangprasert *et al.*, 2010). Across known LTTR structures, the tetrameric interfaces appear to vary (see Section 5.4.2), despite similar EBD dimerisation interfaces. For these reasons, it was expected that each recombinant regulator protein in this study, including those in which the DBD is absent, would dimerise in solution. Analytical SEC was therefore utilised to assess the quaternary solution-state organisation of the group of seven prepared recombinant proteins.

6.3.1. AdeN constructs with and without DBD

Purified samples of the two variants of AdeN were subjected to SEC to establish the native mass of each species in solution. In relatively high salt (300 mM NaCl) and pH 7.5 buffer (to promote a residual negative charge), single protein species were seen to elute in the SEC traces. These are shown in **Figure 6.6** and establish K_{AV} = 0.52 for AdeN and K_{AV} = 0.54 for AdeN [65-238]. From a size standard curve (Section 8.2.3.2), these measurements correspond to 45 kDa for full-length AdeN, and 40 kDa for its truncated variant. These sizes indicated both proteins to occur in solution as dimers (see **Table 6.1** for sequence properties). Slight variation from theoretical dimeric mass values (55 and 40 kDa, respectively) were due to alteration of Stoke's radius by conformational rearrangements resulting from the presence and absence of the DBD. These results suggested full-length recombinant AdeN may have folded in a particularly compact manner, with a smaller molecular volume than expected. AdeN [65-238] was certainly seen to be folded in solution, with quaternary structure present, despite the DBD truncation.

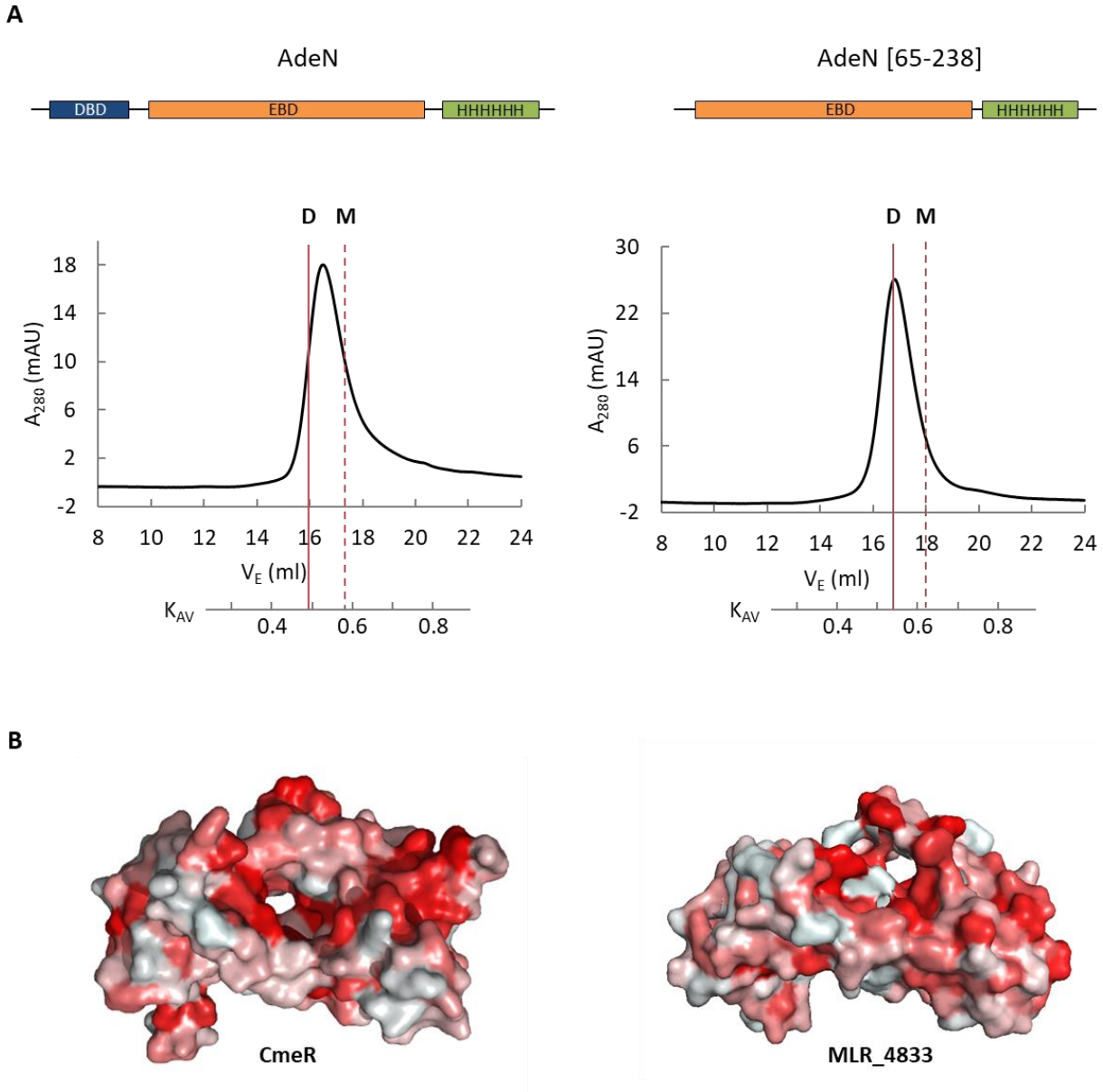


Figure 6.6 Preliminary analysis of AdeN variants. (**A**) Analytical SEC traces (HEPES buffer, pH 7.5 with NaCl, V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml. Elution positions are indicated for the theoretical dimer (D) and monomer (M). (**B**) Crystal structures of single chains for the closest homologs for AdeN (CmeR PDB 2QCO (Gu *et al.*, 2007), MLR_4833 PDB 3BHQ) with surface hydrophobicity indicated (hydrophobic residues → red, hydrophilic residues → white).

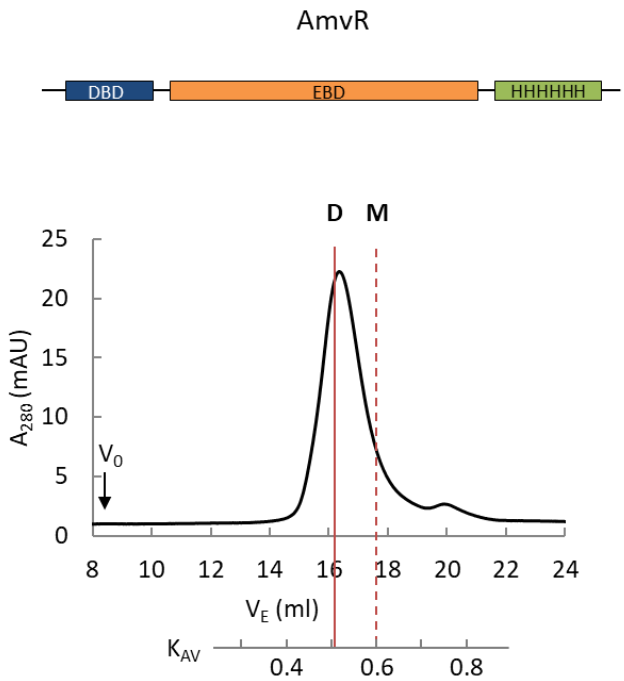
An attempt was made to monitor the thermal melting of these AdeN oligomers. However, T_M could not be measured by the technique of DSF due to an intrinsically high initial fluorescence signal in the presence of Sypro Orange. This points to a protein of high hydrophobicity or one with hydrophobic patches on the surface. An examination of the known crystal structures for TFR dimers CmeR and MLR_4833 (**Figure 6.6 B** (Gu *et al.*, 2007)), close homologs of AdeN, shows large open and hydrophobic cavities in the EBD whether the DBD is present or not. It is probable that AdeN variants share this feature. An alternative method used for measuring thermal melting of proteins is CD. However, this method was not amenable to these protein variants as it required the use of low salt buffers (< 50 mM). As previously mentioned, the required regulator proteins were only stable in solutions containing salt concentration above 200 mM.

6.3.2. AmvR construct with DBD

Purified samples of AmvR were similarly evaluated to establish native mass in solution for this full-length TFR repressor protein. Using HEPES buffer at a pH of 7.5 (i.e. generating a residual negative charge for the protein), a single species was seen by analytical SEC at K_{AV} = 0.51 (**Figure 7.6 A**). This behaviour corresponds to a 50 kDa solution species, indicating recombinant AmvR to be organised as a dimer.

The integrity of this AmvR species was additionally probed by monitoring the thermal melting temperature of the pure sample with DSF. Well-formed melting curves not only gave a distinct T_M for AmvR, but clearly showed this protein sample to be folded and stable in solution. DSF allows a wide range of solution conditions to be screened (see Section 8.2.3.3). Thus, DSF responses for AmvR are displayed for pH values in **Figure 6.7 B**. Samples of the AmvR dimer were found to be most stable at pH 8.0, with a T_M value elevated slightly to 52 °C over readings taken at acidic pH (T_M = 48 °C). Thus, an increase in overall negative charge seems to stabilise the protein.

A



B

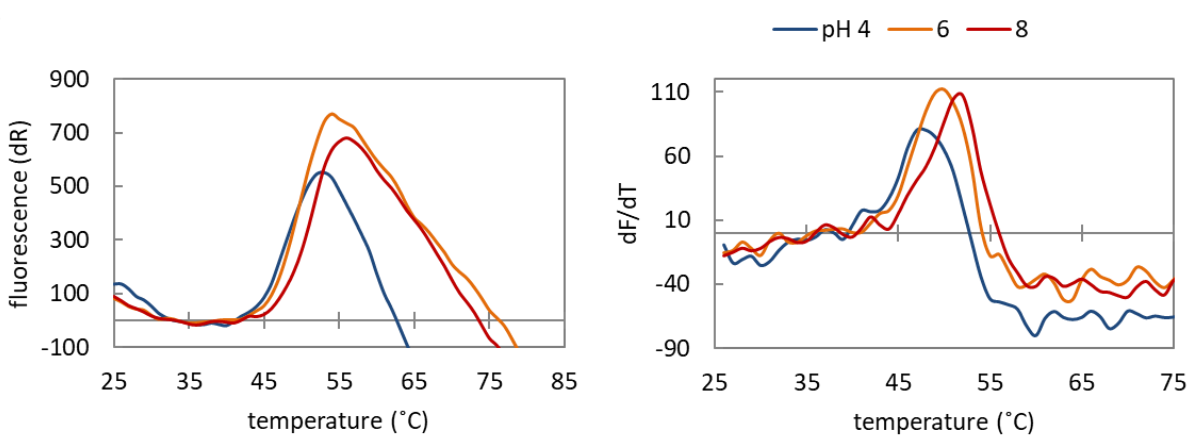


Figure 6.7 Analytical SEC trace for AmvR on Superdex 200, and thermal melt curves monitored by DSF. (**A**) Analytical SEC (V_T 24 ml), V_0 = 8.34 ml in HEPES buffer, pH 7.5 with NaCl. Predicted elution positions are indicated for the theoretical dimer (D) and monomer (M) forms. (**B**) DSF responses during temperature gradient for AmvR in acetate (0.5 M, pH 4.0), MES (0.5 M, pH 6.0), and HEPES (0.5 M, pH 8.0) buffers with 300 mM NaCl. Melting curves (left) and 1st derivative (right), are used to determine the melting temperature (T_M).

Also tested was a range of salt concentrations between 50-300 mM NaCl, with no ensuing change to the T_M value. Thus, the AmvR dimer does not appear sensitive to the ionic strength of the solution. Conditions have now been established which promote protein stability, allowing for further screening of possible ligands and eventual crystallisation.

6.3.3. AceR constructs with and without the DBD

Recombinant samples of AceR, both full-length and truncated forms, were subjected to biophysical investigation at pH 7.5 with 300 mM NaCl, so ensuring a negatively charged protein. On SEC analytical traces in HEPES buffer, single species were seen at a K_{AV} = 0.45 (for AceR) and K_{AV} = 0.49 (for AceR [86-318], **Figure 6.8 A**). This distinctly indicated both proteins to be present in solution as dimeric forms due to derived native mass values of 75 and 55 kDa, respectively. Both recombinant forms were folded proteins with quaternary structure in solution, indicating that the EBD had integrity in the truncated variant AceR [86-318].

Derivatives of the DSF responses recorded for samples of both AceR constructs are shown in **Figure 6.8 B**. The observed T_M values of 53 °C for AceR at pH 7.0 and 64 °C for AceR [86-318] at pH 6.0 were relatively high and confirm both were folded and stable in solution. However, the difference between the two values (11 °C) clearly suggests that truncation of the DBD from this LTTR protein resulted in a large increase in stability. This result validated the truncation made in designing the AceR variant, which formed a very stable dimeric organisation. It also confirmed that the DBD and LH contribute significantly to an instability of AceR, as observed with similar LTTR proteins such as BenM and CatM (Ezezika *et al.*, 2007).

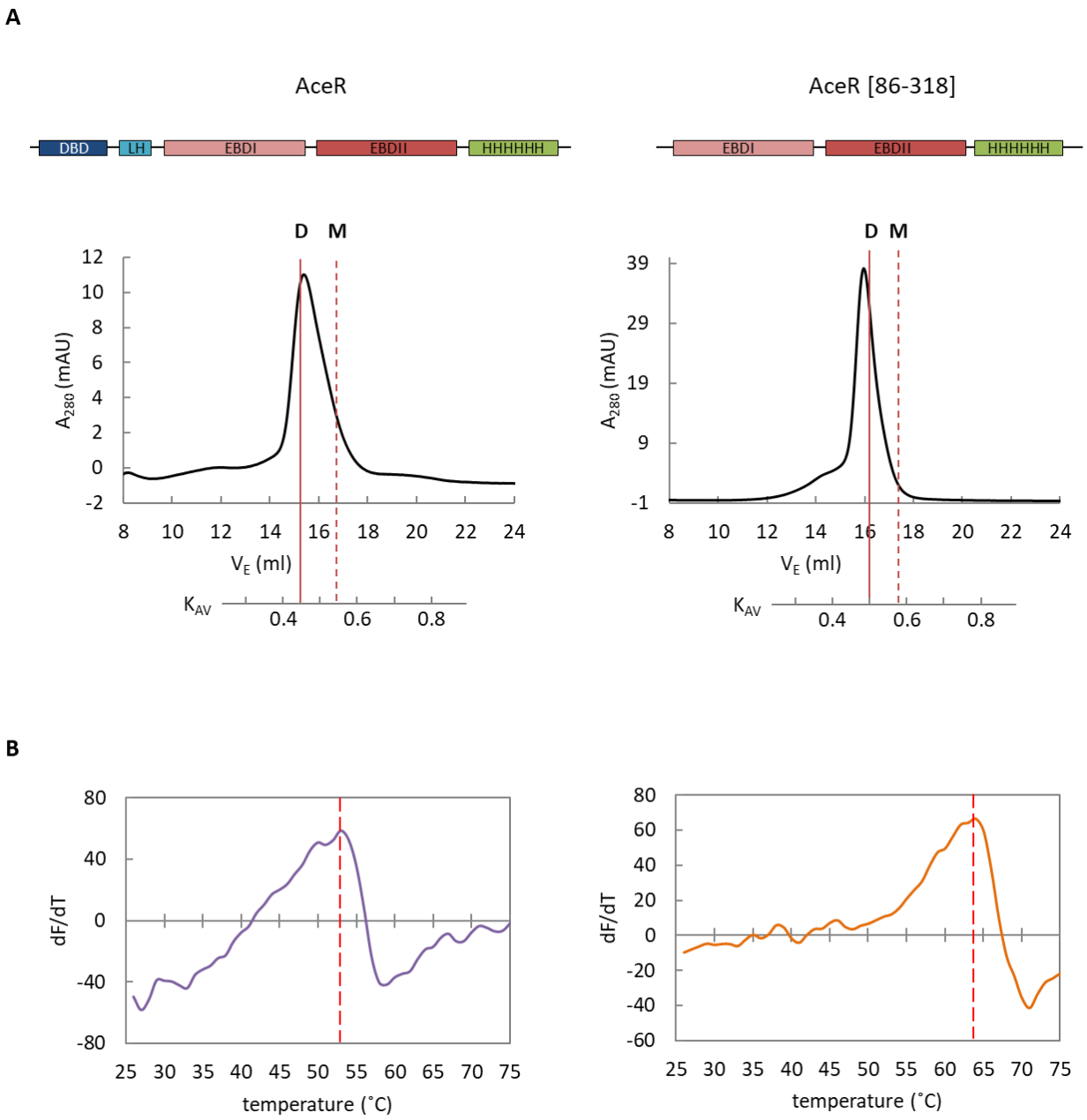


Figure 6.8 Analytical SEC traces and thermal melt curves for AceR variants. (**A**) Analytical SEC (V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml in HEPES buffer, pH 7.5 with NaCl. Elution positions are indicated for the theoretical dimer (D) and monomer (M). (**B**) DSF 1st derivative responses for AceR at pH 7.0 (0.5 M HEPES, 300 mM NaCl) and AceR [86-318] at pH 6.0 (0.5 M MES, 300 mM NaCl) with T_M indicated (dotted red line).

6.3.4. AdeL regulator domain constructs

Purified samples of two AdeL truncated forms ([54-358] and [84-358]) were available for characterisation after successful purification. The longer of the two contained both the LH and the EBD portions which define the LTTR protein family. When subjected to SEC, a single species (K_{AV} = 0.50) was seen to elute for AdeL [84-358] in HEPES buffer containing salt. However, a mixture of two separate species was seen for AdeL [54-358], its components eluting at K_{AV} = 0.53 and K_{AV} = 0.46 (**Figure 6.9 A**). These measurements corresponded to native mass values of 50 kDa for AdeL [84-358], and 40 and 70 kDa for the two components of AdeL [54-358]. This clearly indicated that while AdeL [84-358] was a dimer in solution, AdeL [54-358], designed to include the LH portion, existed as a mixture of monomeric and dimeric species. The mixed oligomeric state of AdeL [54-358] suggested that the LH segment acts as a steric hindrance to dimerisation via the retained EBD or that the truncation design was flawed, most likely resulting in premature termination of secondary structures, destabilising the protein. In most full-length LTTR structures, such as CrgA, the LH associates to a greater degree with the DBD than the EBD (see **Figure 5.5**). Thus, in the absence of the DBD, the LH segment obscures dimerisation. These results indicate that while AdeL [84-358] was folded into a stable quaternary structure, in AdeL [54-358], the dimeric interface had been perturbed. The latter species was not, however, completely unfolded.

Thermal melt curves were investigated for both protein constructs across a range of pH and salt concentrations. DSF responses for AdeL [54-358] were not ideal, as the overall fluorescent signal was diminished under low protein concentration of prepared samples. However, a T_M value of 50 °C at pH 7.0 was indicated for the oligomeric mixture tested (**Figure 6.9 B**).

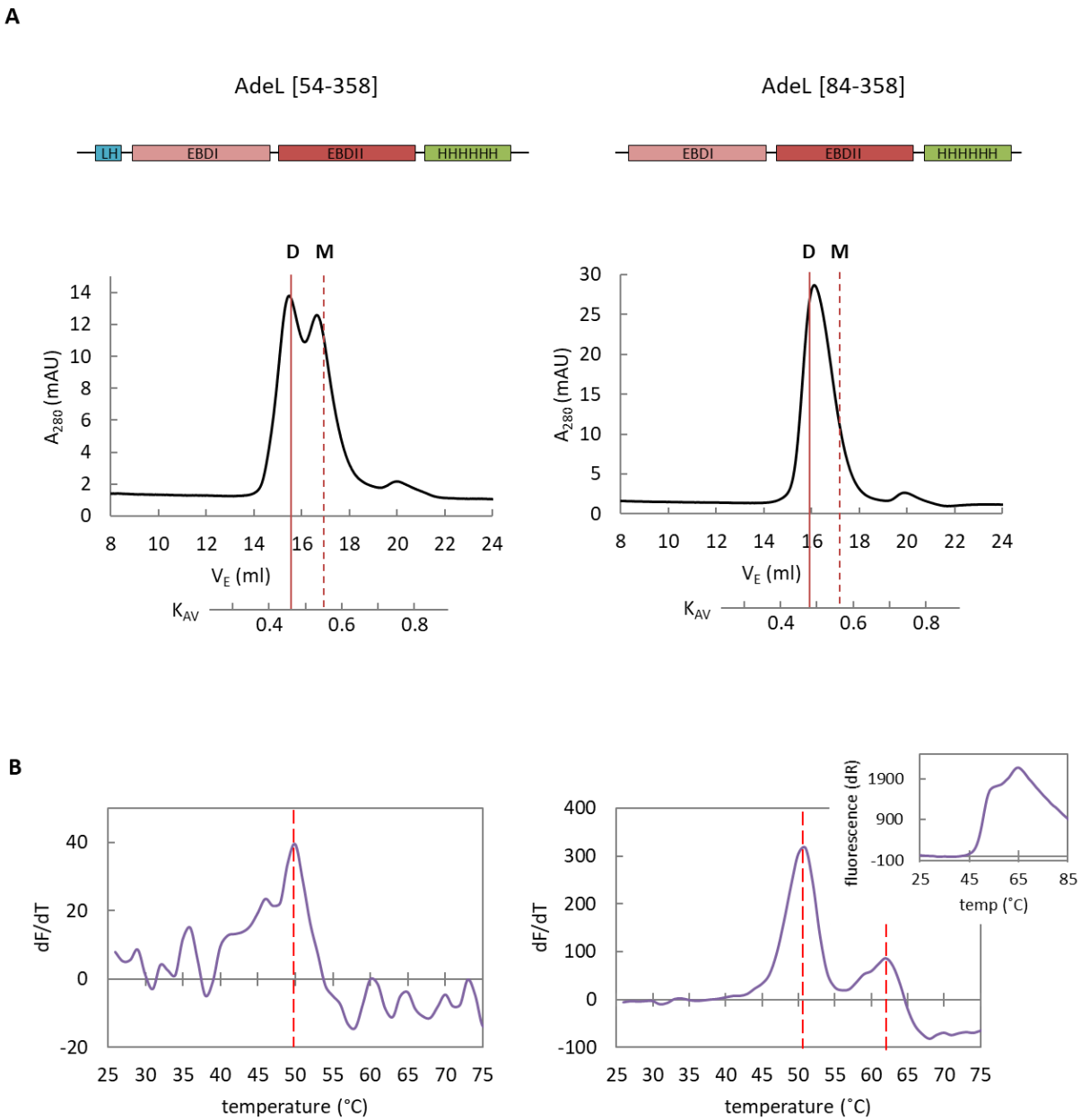


Figure 6.9 Analytical SEC traces and thermal melt curves for AdeL variants. (**A**) Analytical SEC (V_T 24 ml, Superdex 200) at 0.5 ml/min, V_0 = 8.34 ml in HEPES buffer, pH 7.5 with NaCl. Elution positions are indicated for the theoretical dimer (D) and monomer (M). (**B**) DSF 1st derivative responses for AdeL [54-358] and AdeL [84-358] at pH 7.0 (0.5 M HEPES, 300 mM NaCl) with T_M indicated (dotted red line). The raw melt curve for AdeL [84-358] is also shown for comparison.

In contrast to all other regulator proteins investigated in this work by DSF, AdeL [84-358] samples showed a two-state melting curve (**Figure 6.9 B**). Following an initial unfolding event at 50 ˚C, a retained species with some stability then dissipated at 62 ˚C. Given there is a suggestion that the [LH + EBD] dimer collapsed at 50 °C (see AdeL [54-358] above), then the lingering thermostable entity signified some tight interaction to be possible between EBDs only in the absence of the LH component. As the EBD construct AdeL [84-358] was easier to purify at good yield than AdeL [54-358], and appeared relatively stable in solution, this construct was pursued for ligand screening.

6.1. Summary

Parameters measured for seven versions of regulator proteins are summarised in **Table 6.1**. Certain proteins proved more amenable to study than others. Of the TFR representatives, AdeN and AmvR, the full-length constructs containing both DBD and EBD components were more stable and expressed in higher yields. AdeN and AmvR were seen to be dimers in solution, and a clear melting step (T_M of 52 °C) was determined for AmvR. This T_M was determined in basic pH conditions, which imparted an overall negative charge on AmvR. As efflux regulator proteins bind DNA, in the absence of this ligand they would contain a disproportionate amount of exposed positively charged residues in the DBD. By placing the protein in conditions that result in a more negatively charged surface, it may have compensated for the imbalance.

It is always important to mind the chemistry of protein structures when expressing variations. Perturbation of disulphide bonds is especially important in truncations. However, as TFR and LTTR families maintain a high order of flexibility in their DBDs, it was unlikely that a crucial disulphide bond would be disrupted upon removal of the domain. AdeN contains seven cysteine residues within the full-length structure, one of which is located within the DBD. Although removal of the DBD resulted in a lower yield of the protein variant AdeN [65-238], both this and the full-length form were still folded as dimers in solution, showing that Cys-13 did not form a crucial bond with the EBD. No other regulator proteins contained cysteine residues within the DBD, and therefore removal of this domain would not affect oxidative chemistry of the variants.

Expression of LTTR proteins was limited to EBD constructs. The comparison of AceR and AceR [86-318] thermal melt curves highlighted the advantages to removing the DBD in this family, with a pronounced increase in stability from the full-length to truncated forms. This change in protein stability was also seen between constructs containing the LH portion and EBD only. The variant AdeL [54-358], retaining the LH, was less stable and had lower yield than AdeL [84-358]. These results highlight the strong dimerisation ability of EBD components within the LTTR family.

Protein forms that have solution folds have now been provided, allowing identification of possible binding partners for each protein. The regulators that appear to be most amenable to binding assays are AmvR, AdeL [84-358], and AceR [86-318]. Given each dimer has a stable fold, they are therefore capable of being functional.

CHAPTER 7

Binding Studies of *A. baumannii* Efflux Pump Regulator Proteins

The multicomponent efflux systems of bacteria have likely evolved to actively extrude chemical compounds required for molecular signalling needs across a variety of habitats, or to remove toxic metabolic products (Blanco *et al.*, 2016). Expression of efflux transporters is tightly controlled, requiring regulator proteins with capacity to respond to a wide range of compounds (Sun *et al.*, 2014b; Damier-Piolle *et al.*, 2008). The AdeFGH transporter of *A. baumannii*, for example, engages over ten different compounds, including chloramphenicol, ciprofloxacin, trimethoprim, fluoroquinolones, tetracycline-tigecycline, clindamycin, sulfamethoxazole, SDS, and dyes (Coyne *et al.*, 2010b).

As outlined in Chapter 5, some efflux transporters are triggered by more than one specific substrate (Rosenfeld *et al.*, 2012). The organisation of LTTRs can, for instance, accommodate multiple active sites. An example of this, BenM, has been crystallised with two ligands (*cis*,*cis*-muconate and benzoate) occupying different sites within the effector-binding domain (**Figure 7.1**, (Ezezika *et al.*, 2007)). When benzoate binds in the secondary binding site, discrete electrostatic interactions (engaging Glu162 and Arg146) enhance the negative charge of muconate bound to the adjacent primary site. The redistribution of electrostatic potential draws both effector-binding sub-domains (EBDI and EBDII) more closely towards muconate. Therefore, with both effectors, a unique conformation capable of high level transcriptional activation is stabilised (Ezezika *et al.*, 2007). This explains previous observations that benzoate or muconate alone can activate BenM-mediated transcription, yet together yield a level of transcriptional activation higher than the sum of their individual effects (Bundy *et al.*, 2002; Collier *et al.*, 1998).

BenM, along with a close homolog CatM (59% sequence identity), controls a complex regulon for aromatic compound degradation in *Acinetobacter baylyi* ADP1 in which benzoate is converted to tricarboxylic acid cycle intermediates (**Figure 7.1**, (Romero-Arroyo *et al.*, 1995; Collier *et al.*, 1998)). Benzoate consumption requires transcriptional activation by BenM and CatM at four loci

where the relative importance of each regulator varies (numbered 1-4 in **Figure 7.1 B** (Bundy *et al.*, 2002)). This system represents how complicated regulation can be in a cell, possibly requiring multiple regulator proteins or effector molecules. For this reason, it is important to try and understand every aspect, for which some research groups start at the transcriptional level, and some at the protein level.

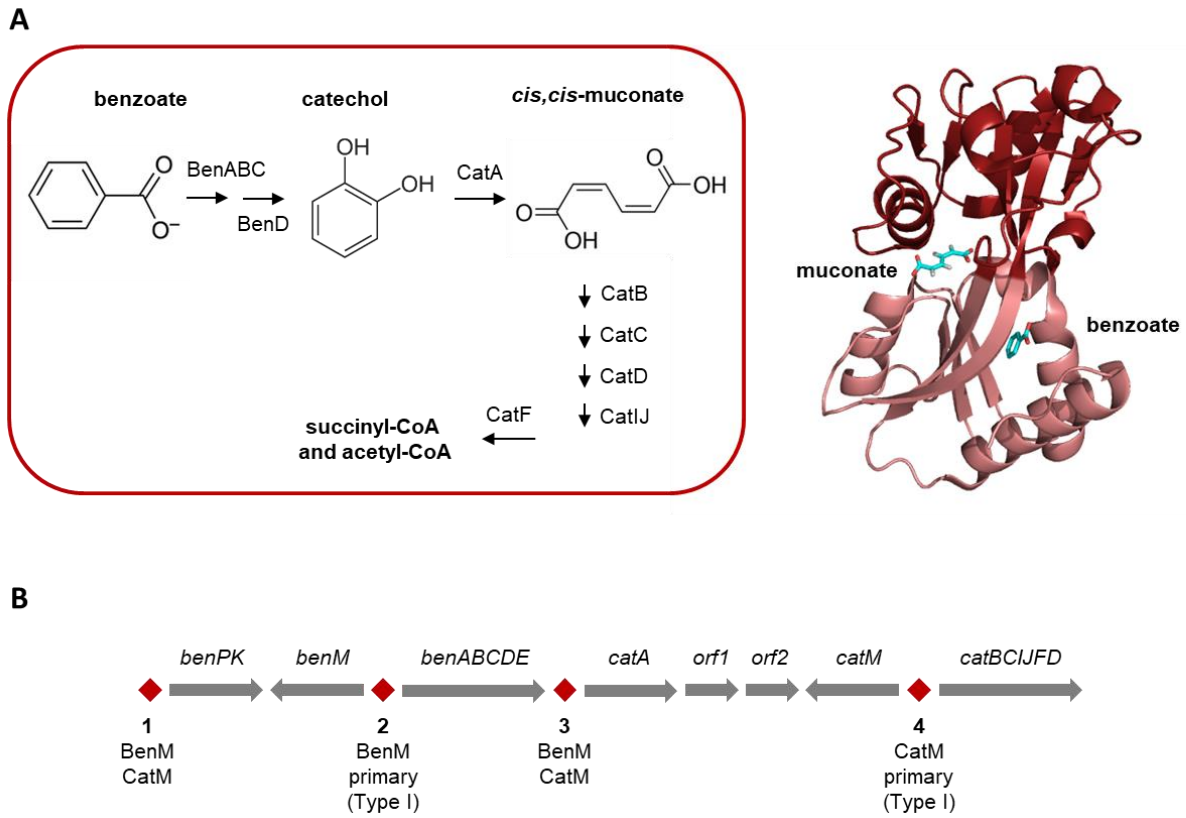


Figure 7.1 Regulatory role of BenM in *A. baylyi*. (**A**) Pathway for the conversion of benzoate to tricarboxylic acid cycle intermediates. Right: Crystal structure of BenM indicating location of bound substrates muconate and benzoate (PDB: 2F7A, (Ezezika *et al.*, 2007)). Ribbon view is coloured to illustrate EBDI (light pink) and EBDII (crimson) sub-domains. (**B**) Roles of BenM and CatM in controlling expression from multiple promoters (numbered 1-4) in a supraoperonic cluster of chromosomal genes (Bundy *et al.*, 2002).

Table 7.1 Summary of regulator protein characterisation in this work

	regulator	domains present [a]	solution oligomerisation state (Section 6.3)	T_M (°C) (Section 6.3)	suitable for ligand screening?
TFR family	AdeN	DBD & EBD	2	–	–
	AdeN [65-238]	EBD	2	–	–
	AmvR	DBD & EBD	2	52	✓
LTTR family	AdeL [54-358]	LH & EBD	1 + 2	50	–
	AdeL [84-358]	EBD	2	51, 62	✓
	AceR	DBD, LH, & EBD	2	53	–
	AceR [86-318]	EBD	2	64	✓

[a] DBD: DNA-binding domain, LH: linker helix, EBD: effector-binding domain

7.1. Screening for potential effectors of efflux pump regulators

The preparation of several distinct regulator proteins by recombinant means (described in Chapter 6) allowed for specific effector chemistry for AmvR, AdeL and AceR to be explored. **Table 7.1** outlines the constructs I prepared.

As outlined in previous chapters, the method of DSF can respond to thermodynamic events of binding and allows for rapid screening of compounds as leads to chemistries of ligands to a protein partner. Cocktails of small molecular weight compounds ranging from amino acids and peptides, to metals and salts, and nucleotides and carbohydrates were therefore screened against prepared protein materials. A full listing of chemical screens utilised is given in **Table S.1**. Each protein was combined with a hydrophobic fluorescent dye and the mixture transferred to a 96-well plate containing the small molecule cocktails. The plate was heated over a range of 25-95 ˚C in a real-time qPCR machine and the change of fluorescence intensity monitored by the instrument. Derivatives of the curves produced were then used to calculate the transition midpoint which corresponds to the melting temperature (Section 8.2.3.3).

7.1.1. Screening of potential effectors for AmvR

It is known that the AmvA transporter is responsive to dyes, disinfectants, detergents, and the antibiotic erythromycin (Rajamohan *et al.*, 2010b), however, it is unknown which of these might act as effector molecules for the regulator to which it is coupled, namely, AmvR. The dimer of AmvR was screened with additives, and the effects noted relative to its T_M value of 52 °C.

Figure 7.2 summarises some of the thermal melt data collected, with a selection of thermal responses for mixtures provided in panel A. Derivative curves are included for accuracy. Compounds present in four cocktails (D3, D4, D5, and D9) were found to increase thermal stability of AmvR. The thermal response obtained for cocktail D4, for instance, yielded a T_M value of ~55 °C, i.e. an increase of 3 °C. This specific cocktail contains trivalent metal chlorides as well as benzamidine, and salicin (an alcoholic β-glucoside). Notably, a marked elevation of T_M by 8 ˚C was seen for mixture D5, which includes divalent metals (two transition metals, Mn and Zn, and two alkaline earth metals, Ca and Mg). This increase in T_M was not observed for cocktail D6, which also contains divalent ions of transition metals (ΔT_M of -5 °C). It can thus be surmised that Ca^{+2} and Mg^{+2} ions increased the thermal stability of AmvR.

A summary map across all wells of the commercial screen utilised is presented in **Figure 7.2**, shaded to indicate discrete changes in T_M, where values increased or decreased. This is combined with a generalised categorisation of chemical classes in each screen. **Table 7.2** summarises responses obtained where a screened group of compounds resulted in changes of $T_M \geq 2$ °C. Compounds in well F7, barbituric acid, betaine, phloroglucinol, resorcinol, and tetrahydroxy-1,4-benzoquinone, severely destabilised the protein, a common effect observed when added to many proteins samples screened across this study.

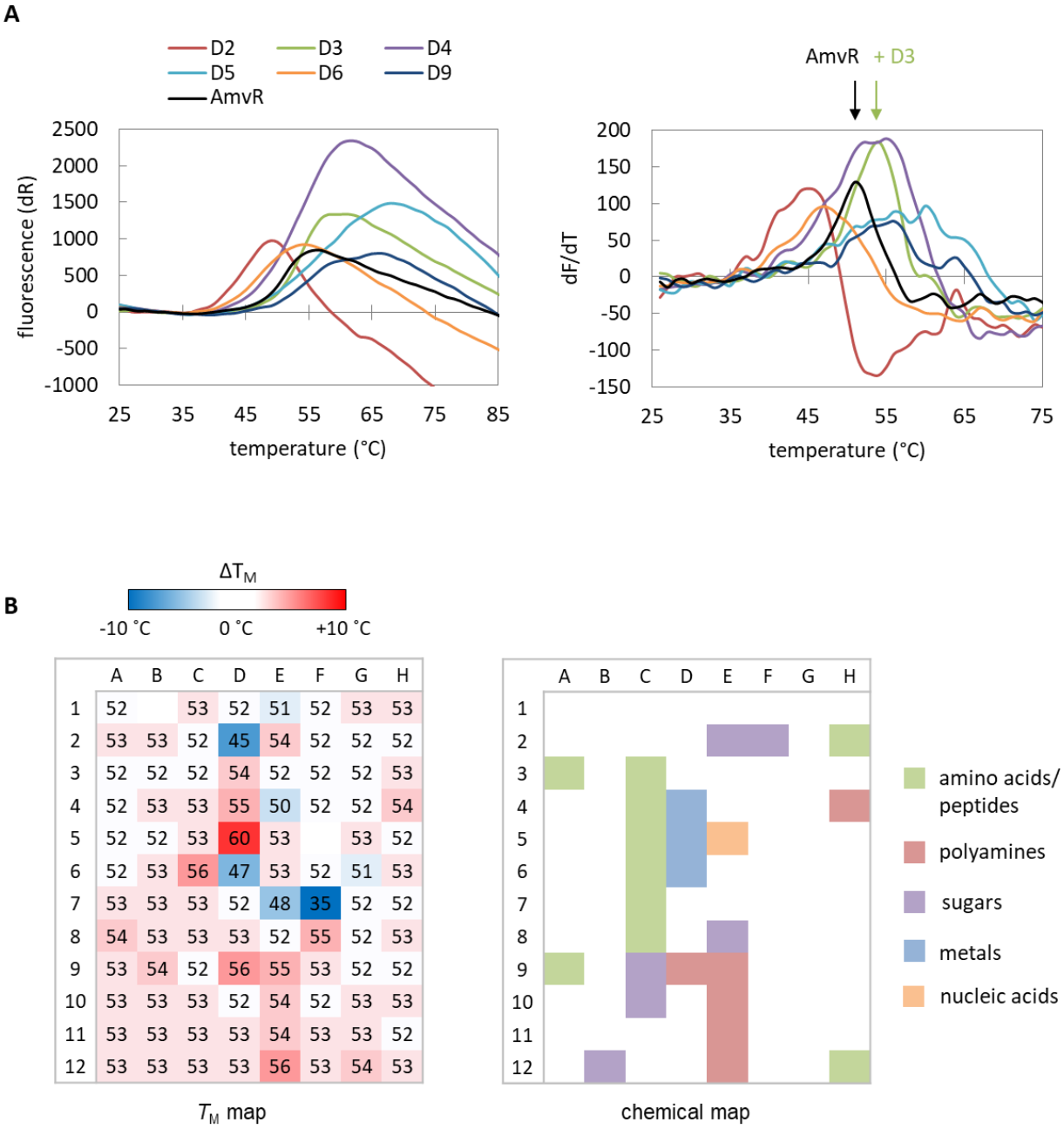


Figure 7.2 Thermal behaviour of AmvR in the presence of molecular cocktails. (**A**) Raw and 1st derivative melt curves for selected screens with $\Delta T_M > 2$ °C (full compound listing is given in **Table S.1**). T_M points are indicated on the derivative response for AmvR combined with cocktail D3 and in apo form. (**B**) Left: Heat map for T_M values across 96 cocktails, coloured according to change in T_M (red for + ΔT_M, blue for - ΔT_M). T_M values (°C) are listed for each condition. Right: Mapping of chemical functional groups screened in this experimental grid.

Table 7.2 Elevated T_M values detected via DSF for small molecule mixtures with AmvR (EBD + DBD)

ΔT_M (°C)	cocktail ingredients	well number [a]
+8	calcium chloride, magnesium chloride, manganese(II) chloride, zinc chloride	D5
+4	Gly-Phe, Gly-Tyr, Leu-Gly-Gly	C6
	1,2-diaminocyclohexane sulfate, 4-nitrobenzoic acid, cystamine dihydrochloride, spermine	D9
	putrescine, 1,8-diaminooctane, cadaverine, cystamine dihydrochloride, spermidine	E12
+3	putrescine, cystamine dihydrochloride, diloxanide furoate, sarcosine, spermine	E9
	gadolinium(III) chloride, samarium(III) chloride, benzamidine hydrochloride, salicin	D4
	1,6-hexanediol, methylphosphonic acid, Gly-Gly, myo-inositol, phloroglucinol	F8
+2	sodium 1-pentanesulfonate monohydrate, 3,5-dinitrosalicylic acid, 3-aminosalicylic acid, salicylamide	A8
	hexamminecobalt(III) chloride, salicylamide, sulfanilamide, vanillic acid	B9
	MES monohydrate, PIPES, hexamminecobalt(III) chloride	D3
	dextran sulfate, dextranase, α-amylase	E2
	1,2-diaminocyclohexane sulfate, 1,8-diaminooctane, cadaverine, spermine	E10
	1,2-diaminocyclohexane sulfate, diloxanide furoate, fumaric acid, spermine, sulfaguanidine	E11
	benzenephosphonic acid, gallic acid monohydrate, melatonin, N-(2-carboxyethyl)-iminodiacetic acid, trimellitic acid	G12
	putrescine, 1,8-diaminooctane, cadaverine, cystamine dihydrochloride, spermidine, spermine	H4

[a] Silver Bullets, Hampton Research, www.hamptonresearch.com/product_detail.aspx?cid=30&sid=179&pid=562 (McPherson & Cudney, 2006)

Among the compounds tested, polyamines proved a common class of potential ligands, as seen across the grid of elevated T_M values (**Figure 7.2 B**). All wells containing such compounds showed an increased T_M (> 2 ˚C). **Figure 7.3 A** outlines the distribution of specific polyamines within screens D9, E9, E10, E11, E12, and H4. The molecule spermine was a component in five out of six cocktails screened, and cystamine was an ingredient in four. Thus, AmvR appears to be responsive to these polyamines.

In order to investigate specificity of binding, polyamine components from these cocktails were individually added to preparations of AmvR and thermal melts recorded. Example thermal responses in the presence of putrescine and cystamine are depicted in **Figure 7.3** (panel **B**)

showing T_M values of 53 °C and 63 °C, respectively. The polyamines tested range in charge (between +2 and +4) and size (4-10 methyl groups in length), but no correlations to these parameters were observed. Four of the polyamines tested increased the T_M of AmvR by > 2 °C, with cystamine producing the largest change in T_M (+11 ˚C). This large increase in protein stability appeared to be masked when cystamine was combined with other compounds. This suggests that cystamine might be altered by the cocktail chemistry itself. Cystamine is redox sensitive, and in the presence of amines, may be reduced to cysteamine, with a lessened affinity for AmvR.

AmvR contains two cysteine residues (Cys-136 and Cys-161). A consideration of the structure of its closest homolog (PTR_SV from *Saccharomonospora viridis*), indicates these likely to be too far apart to form disulfide bonds within the protomer. However, as Cys-161 likely resides within helix α_8 (known to act in conjunction with α_9 to form the dimer interface), there is the potential for it to form a disulfide bond with its counterpart (Cys-161') within the dimer. This in turn could form the basis for a cysteine-based redox switch within AmvR. These residues thus could also play a part in the interaction of AmvR with cystamine and the overall function of the protein.

In summary, full-length AmvR is stabilised by a range of small molecules, with calcium and magnesium producing the highest ΔT_M amongst compounds tested (+8 ˚C). These metals play an important role in many biological mechanisms within an organism, such as the cleavage and regulation of DNA, and their intracellular concentrations are therefore tightly controlled. As AmvR is a transcriptional regulator, it is likely that either Mg^{2+} or Ca^{2+} (or both) are required for DNA binding and protein activity (Bellamy *et al.*, 2009). It is also possible that AmvR controls the export of these metals from the cell. An example of a transcriptional family that regulates intracellular metal concentrations is the ArsR–SmtB family. ArsR–SmtB family members possess a highly conserved DNA recognition HTH motif and bind as homodimers to their operator/promoter region, repressing the expression of operons, in the absence of metal ions, associated with metal ion sequestration or efflux in both Gram-negative and Gram-positive bacteria, while derepressing the operons in the presence of toxic concentrations of heavy metal ions, allowing survival of bacterial cells in challenging environments (Saha *et al.*, 2017).

A

		inclusion in commercial screen					
linear formula (pH 7.5)	**amine**	D9	E9	E10	E11	E12	H4
$NH_3^+(CH_2)_3NH_2^+(CH_2)_4NH_2^+(CH_2)_3NH_3^+$	spermine	✓	✓	✓	✓		✓
$NH_3^+CH_2CH_2SSCH_2CH_2NH_3^+$	cystamine	✓	✓			✓	✓
$NH_3^+(CH_2)_4NH_3^+$	putrescine		✓			✓	✓
$NH_3^+(CH_2)_8NH_3^+$	1,8-diaminooctane			✓		✓	✓
$NH_3^+(CH_2)_5NH_3^+$	cadaverine			✓		✓	✓
$NH_3^+(CH_2)_3NH_2^+(CH_2)_4NH_3^+$	spermidine					✓	✓

B

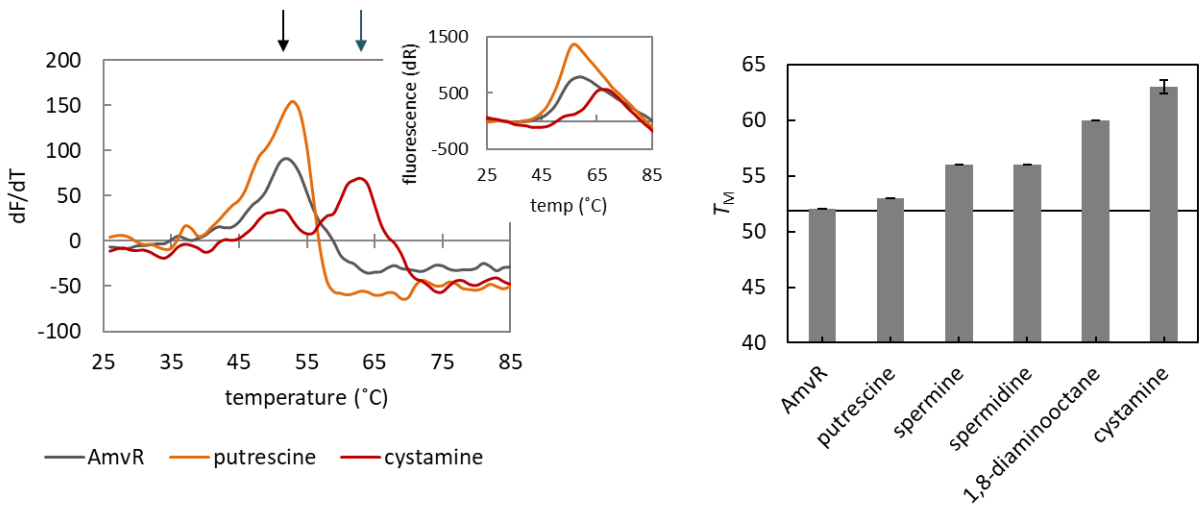


Figure 7.3 AmvR melting temperatures in the presence of polyamines. (**A**) Occurrence of designated polyamines in cocktail mixtures from Hampton Research (McPherson & Cudney, 2006) screen. Each structural formula is displayed at pH 7.5. (**B**) Left: Raw and derivative thermal melt curves for apo AmvR (gray) with and without putrescine (orange) and cystamine (red). Derivative curves are shown. Right: Determined T_M values for AmvR when combined with single polyamines. SEM error bars are shown (n = 3).

Additionally, all cocktails containing polyamines increased the T_M of AmvR by at least 2 ˚C. Amine functional groups are highly represented in the classes of molecules that have been shown to react to the AmvA pump within *A. baumannii*. Ethidium bromide, chlorhexidine, acriflavine, methyl viologen and 4,6-diamidino-2-phenylindole have all produced transcriptional changes of AmvA (Hassan *et al.*, 2016; Rajamohan *et al.*, 2010b). Each of these compounds contain at least two amine groups each and half contain a reactive nitrogen group. When the polyamines were tested individually for stabilising capacities, cystamine produced a significant effect. Cystamine contains two amine groups along with an oxidation-dependent disulfide group. Although disulfide bonds are not seen in any of the above compounds, it is possible that AmvR could contain unpaired cysteine residues that have the potential to react with cystamine.

Alongside the work carried out in this thesis, some transcriptomics data has been acquired, by other researchers at Macquarie University, for each regulator system studied here under dosing of four polyamines. Transcriptional changes to efflux genes were monitored in the presence of high concentrations of spermine, spermidine, cadaverine, and putrescine (private communication, Q. Liu and I. Paulsen). Only two of these polyamines came close to inducing a 2-fold change in expression of the target efflux systems: increased expression was seen for component genes of the *amvR/amvA* system under spermidine and spermine. These transcriptional changes are only hinted at in the binding results acquired in this work. AmvR interacts with both spermine and spermidine more strongly than putrescine under the kinetic regime offered by the DSF experimental protocol.

7.1.2. Screening of potential effectors for AdeL

While it is known that AdeFGH/AdeL efflux can confer host resistance to the antibiotics chloramphenicol, ciprofloxacin and trimethoprim (Coyne *et al.*, 2010b), little is known concerning the underlying molecular mechanism for this system. Recombinant preparation of AdeL [84-358], which contains the EBD portion of the regulator, provided a useful module for initial probing of

protein activity. The dimer of AdeL [84-358] was screened with additives, and the effects noted relative to its two-state melting curve, with T_M values of 50 and 62 °C.

The thermal melt data collected for small molecule mixtures with AdeL [84-358] is summarised in **Figure 7.4**. A selection of thermal responses is provided in panel A with the derivative of these curves shown on the right, allowing determination of the T_M. What immediately became apparent from the measured thermal melt curves was the transition of AdeL [84-358] from a two-state quaternary unfolding to a single event of collapse in the presence of particular compounds. **Figure 7.4 A** illustrates this change, showing that when AdeL [84-358] was combined with small molecules from well D9, the T_M was represented by a single value of 55 °C. Three screened conditions increased the T_M for AdeL [84-358] and six sets decrease the stability of the protein (**Figure 7.4 B**). Destabilising compounds were similar to those found for AmvR (wells D2, E7, and F7) containing a mix of buffer components, protein degrading enzymes, and small molecules. Notably, AdeL [84-358] was lowered in thermostability by sets containing primarily divalent transition metals with the exception of calcium and magnesium (wells D5 and D6).

Cocktails containing polyamines were seen to stabilise AdeL [84-358], with four out of the six conditions containing polyamines (wells D9, E9, E12, and H4) increasing the T_M of AdeL [84-358] by 5 °C. The molecules specific to these four cocktails are putrescine and cystamine, so these individual polyamines were tested for their separate effects on AdeL. It was found that cystamine increased the T_M by 4 ˚C, proving it to be the most stabilising of all compounds tested (**Figure 7.5**). This agrees with the findings from the cocktail screen, as the only conditions that increased the stability of AdeL [84-358] contained cystamine (see **Table 7.3**).

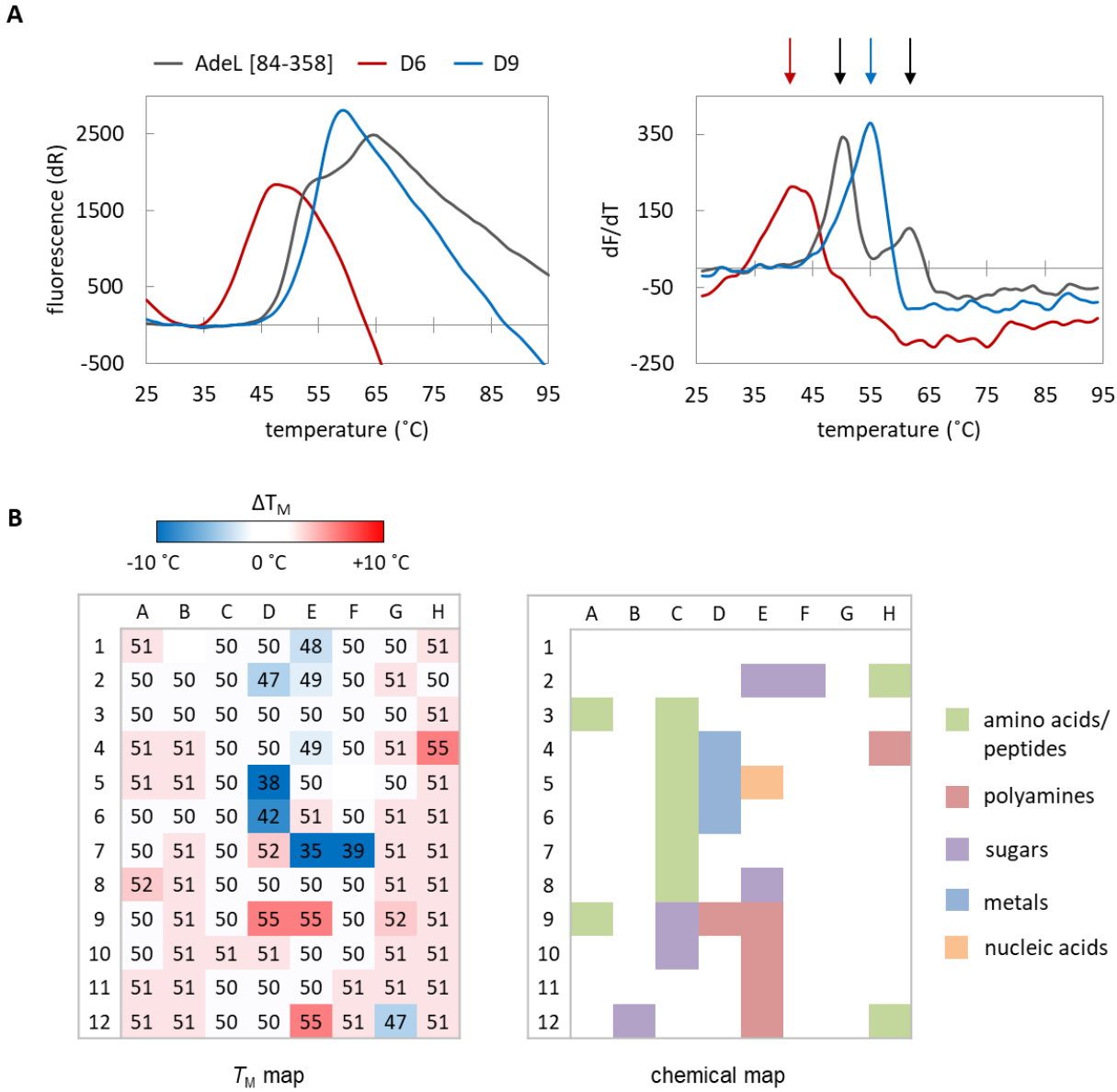


Figure 7.4 AdeL [84-358] melting temperatures in the presence of molecular cocktails. (**A**) Raw and 1st derivative melt curves for select small molecule mixtures producing a decrease or increase in ΔT_M. T_M points are indicated on the derivative curve for AdeL [84-358] combined with molecules in well D6 and D9 and in apo form. (**B**) Left: T_M values across 96 chemical cocktails, coloured according to change in T_M (red for $+\Delta T_M$, blue for $-\Delta T_M$). T_M values (°C) are listed for each condition. Right: Location of cocktails containing common molecules in experimental grid.

As cystamine is responsive to redox conditions, it is important to understand the redox potential of AdeL on its own. AdeL contains five cysteine residues, all found within the EBD. When the sequence of AdeL is compared against the structure of its closest homolog, CrgA from *Neisseria meningitidis*, the first two residues (Cys-135 and Cys-156) are most likely found in separate EBD subdomains on strands β_5 and β_6. These strands, however, are not near the dimer interface of AdeL, and therefore not capable of forming disulfide bonds between protomers. The other three cysteines (Cys-292, Cys-306, and Cys-310) are located after the C-terminal alpha helix (α_{14}), therefore likely to reside in the unstructured region of the protein.

Screening of ligands for AdeL again shows that the protein was most strongly stabilised by cystamine. It is tempting to say that AdeL (and the previously tested AmvR) only responds to redox chemistries present in these wells. It is noted, however, that other wells (C1 and H11) contain L-cystine, a similarly oxidised compound, which elicits no increase in protein stability.

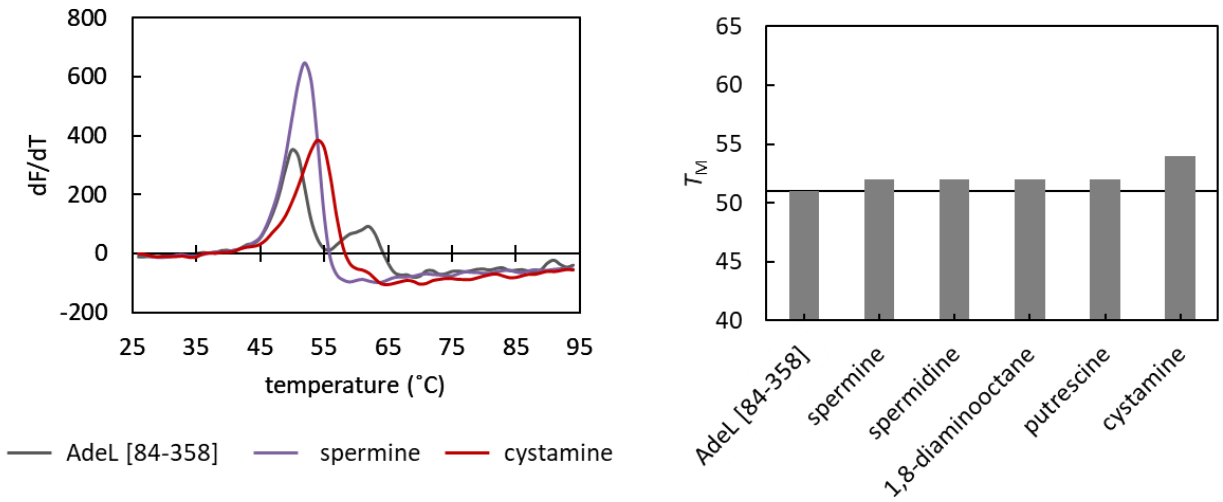


Figure 7.5 Polyamine thermal stabilisation of AdeL [84-358]. Left: Derivative melt curves for apo protein (gray) with and without spermine (purple) and cystamine (red). Right: Determined T_M values for AdeL [84-358] when combined with single polyamines.

Table 7.3 T_M values detected via DSF for small molecule screening of AdeL [84-358]

ΔT_M (°C)	cocktail ingredients	well number
+5	1,2-diaminocyclohexane sulfate, 4-nitrobenzoic acid, cystamine dihydrochloride, spermine	D9
	putrescine, cystamine dihydrochloride, diloxanide furoate, sarcosine, spermine	E9
	putrescine, 1,8-diaminooctane, cadaverine, cystamine dihydrochloride, spermidine	E12
	putrescine, 1,8-diaminooctane, cadaverine, cystamine dihydrochloride, spermidine, spermine	H4

As previously mentioned, transcriptomics data has been acquired by other researchers for this regulator system (private communication, Q. Liu and I. Paulsen). An increase in expression was seen for component genes of the *adeL/adeFGH* system in the presence of spermidine. In this work, it was found that AdeL [84-358] does not bind spermidine with a higher specificity than spermine or putrescine.

It should also be noted that within *A. baumannii*, transcription for the *adeFGH* gene is increased under dosing of a huge range of compounds (Coyne *et al.*, 2010b). These include chloramphenicol, ciprofloxacin, trimethoprim, fluoroquinolones, tetracycline-tigecycline, clindamycin, sulfamethoxazole, SDS, and dyes such as ethidium bromide, safranin O, and acridine orange. This transcriptional response could not solely be the outcome of interactions of AdeL with all these compounds, as the chemical structures of these compounds are all very different, rather the engagement of these compounds with other regulatory components.

7.1.3. Screening of potential effectors for AceR

Chlorhexidine has been identified as a substrate for the AceI/AceR efflux system (Qi Liu, 2017). Susceptibility and induction of *aceI* gene expression was determined by MIC and quantitative real-time PCR, respectively, in *A. baumannii* WT and Δ*aceR* mutant strains, showing that mutant strains showed lower resistance to chlorhexidine than the parental strain (Qi Liu, 2017). In

response to these findings, recombinant AceR was prepared as both a full-length protein and as a DBD-truncated form, AceR [86–299]. The binding interaction of the purified AceR protein and its putative operator region was investigated by electrophoretic mobility shift assays and DNase I footprinting assays demonstrating chlorhexidine-stimulated binding of AceR with two sites upstream of the putative *aceI* promoter (Qi Liu, 2017). The binding of AceR [86-299] with its putative ligand chlorhexidine was examined using surface plasmon resonance and tryptophan fluorescence quenching assays. Results suggested that the purified EBD of AceR was able to bind chlorhexidine with high affinity (Qi Liu, 2017).

Although both full-length AceR and its truncated construct (EBD only) were successfully expressed in this work, only the construct AceR [86-318] was stable at the significant concentration (4 mg/ml) to allow ligand screening by the technique of DSF. This truncated form, which incorporates the EBD domain, proved to be relatively stable (T_M of 67 °C) as a dimer in solution.

Given the previous discovery of chlorhexidine as a ligand for AceR, its affinity for the truncated form of the protein, AceR [86-318], was tested using DSF. Four concentrations of the compound were chosen for testing, ranging from 0-50 µM, for which melting curves can be seen in **Figure 7.6 A**. AceR [86-318] did not display any increases in T_M in the presence of chlorhexidine. However, there was a change to the shape of the melting curve, suggesting that some interaction occurred between the protein and chlorhexidine, even if the nature of this interaction is still unknown.

A selection of thermal responses to the small molecule screens is provided in panel **A** of **Figure 7.6**, with the derivative of these curves shown on the right. For AceR [86-318], the majority of screens elicited little change in T_M, with no cocktails increasing protein stability and nine decreasing the T_M. Divalent transition and Group II metals (D5 = ΔT_M -17 °C and D6 = ΔT_M -23 °C), wells containing cystamine (D9, E9, E12, and H4 with an average ΔT_M of -6 °C), and the compounds in well F7 were found to considerably decrease the T_M value of AceR [86-318]. Although there was one condition that resulted in a ΔT_M of +2 °C, it was unfolded, and the resultant T_M calculation was due to the Mx3005P software trying to correct the curve. It was inferred from

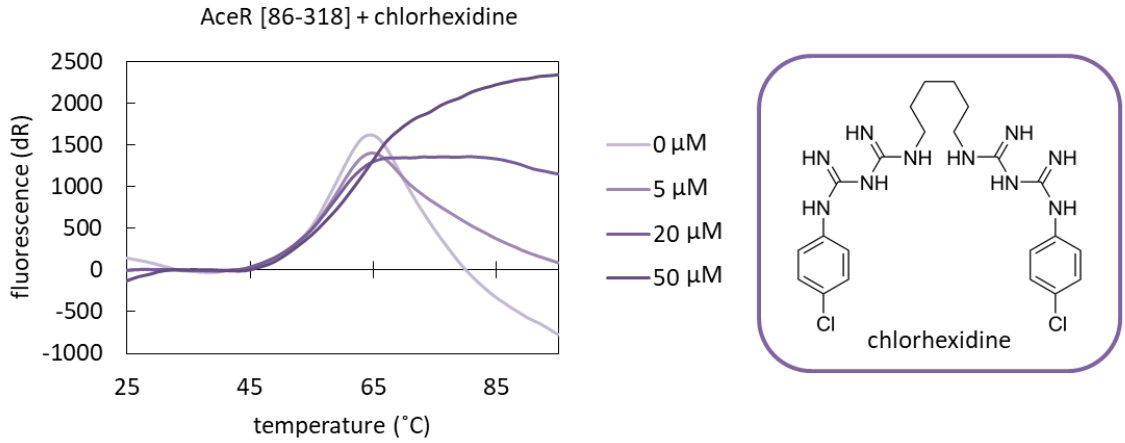
this that AceR [86-318] did not bind with adequate strength to any molecules trialled. It may be that the protein already bound strongly to a component sequestered during the purification process and therefore could not further accommodate an additional binding partner. **Figure 7.6** summarises the thermal melt data collected for mixtures with AceR [86-318].

Transcriptomics data has been acquired by another research group at Macquarie University for the *aceR*/*aceI* efflux system showing that addition of spermidine, cadaverine, and putrescine each gave rise to a 6-fold change in expression for the *aceI* gene (private communication, Q. Liu and I. Paulsen). As AceR is an activator that positively self regulates, it is expected that the 6-fold increase in expression of *aceI* would also apply to *aceR*. However, this is not reflected in the transcriptomics data, with expression levels of *aceR* remaining normal under the same conditions. These results could point towards a second regulator protein within the efflux system that can also control *aceI* expression.

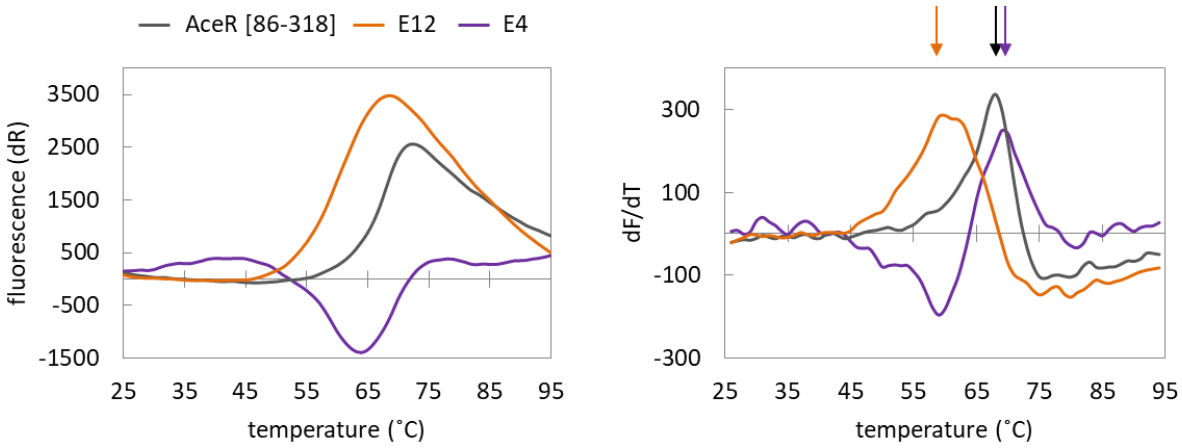
7.2. Summary

Integral membrane drug efflux proteins are prominent drug resistance factors in *A. baumannii*. As for other Gram-negative pathogens, drug efflux proteins of *A. baumannii* can be expected to have significant roles in virulence, but their specific role in this organism's pathogenicity has not been explored. Relative to genome size, *A. baumannii* appears to have one of the largest repertoires of drug efflux pumps of any pathogen (Antunes *et al.*, 2014), and at least eight highly conserved drug efflux systems have been characterised to date (Hassan *et al.*, 2015). Many of these efflux systems are multidrug transporters and their overexpression can shift the clinical breakpoints and epidemiological cut-off values from susceptible to resistant for a swathe of structurally and mechanistically dissimilar antibiotics.

A


AceR [86-318] + chlorhexidine
fluorescence (dR)
2500
2000
1500
1000
500
0
-500
-1000
25
45
65
85
temperature (˚C)
0 µM
5 µM
20 µM
50 µM
HN
NH
Cl
chlorhexidine

B


AceR [86-318]
E12
E4
fluorescence (dR)
3500
2500
1500
500
-500
-1500
25 35 45 55 65 75 85 95
temperature (˚C)
dF/dT
300
100
-100
-300
25 35 45 55 65 75 85 95
temperature (˚C)

C

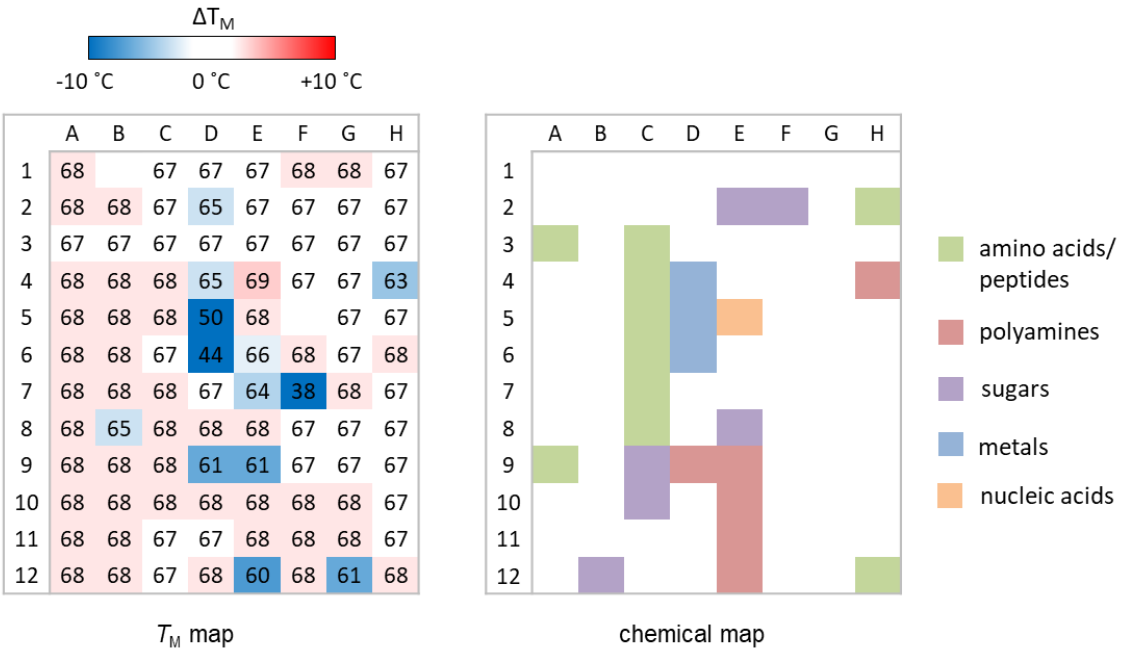

ΔT_M
-10 ˚C
0 ˚C
+10 ˚C
A B C D E F G H
1 68 67 67 67 68 68 67
2 68 68 67 65 67 67 67 67
3 67 67 67 67 67 67 67 67
4 68 68 68 65 69 67 67 63
5 68 68 68 50 68 67 67
6 68 68 67 44 66 68 67 68
7 68 68 68 67 64 38 68 67
8 68 65 68 68 68 67 67 67
9 68 68 68 61 61 67 67 67
10 68 68 68 68 68 68 68 67
11 68 68 67 67 68 68 68 67
12 68 68 67 68 60 68 61 68
T_M map
chemical map
amino acids/ peptides
polyamines
sugars
metals
nucleic acids

Figure 7.6 AceR [86-318] melting temperatures in the presence of chlorhexidine and molecular cocktails. (**A**) Raw melt curves in the presence of 0-50 µM chlorhexidine. The structure of chlorhexidine is seen on the right. (**B**) Raw and 1st derivative melt curves for select cocktails producing a change in ΔT_M. T_M points are indicated on the derivative curve for AdeL [84-358] combined with molecules in well E4 and E12 and in apo form. (**C**) Left: T_M values across 96 chemical cocktails, coloured according to change in T_M (red for $+\Delta T_M$, blue for $-\Delta T_M$). T_M values (°C) are listed for each condition. Right: Location of cocktails containing common molecules in experimental grid.

For this work, AdeN, AmvR, AdeL, and AceR were engineered as full-length and truncated forms to explore levels of solubility and expression. Successfully produced proteins formed dimers in solution, except for AdeL [54-358], which was present as a mixed species of monomer and dimer. This was attributed to the LH segment preventing proper folding, and this inherent instability was supported by melting curves of the protein. With suitable quantities of material, AmvR, AdeL [84-358], and AceR [86-318], were screened to distinguish their possible ligand chemistry.

Although no potential effector molecules were found for AceR [86-318] (possibly due to a tightly bound molecule sequestered during production), both AmvR and AdeL [84-358] responded to polyamines, with cystamine proving to be the tightest binder.

7.3. Conclusions

7.3.1. The role of polyamines in bacterial species

There has been relatively little effort to systematically address the function of polyamines, particularly in bacteria. Most of the major routes for polyamine biosynthesis appear to have been identified, but for the most part, the regulation and function of polyamine biosynthesis in bacteria are unknown. It is clear that polyamines are essential for growth in some bacterial species and influence biofilm formation in others (Lee *et al.*, 2009; Burrell *et al.*, 2010). Furthermore, polyamines are prominent in many natural products produced in bacteria, particularly siderophores (Burrell *et al.*, 2012).

The common feature of diverse polyamines found in eukaryotes, bacteria, and archaea is that they are all derived from amino acids and carry a large positive charge at physiological pH. In bacteria, putrescine, cadaverine, spermidine and spermine are the predominant polyamines (Hamana & Matsuzaki, 1992). These intracellular components are regulated by concerted biosynthesis and uptake mechanisms, as well as by degradation and efflux processes (Tabor & Tabor, 1985). Synthesis of bacterial spermidine, putrescine and cadaverine usually depends on the decarboxylation of precursor amino acids or other intermediates (Di Martino *et al.*, 2013), illustrated in **Figure 7.7**. Diamines are mainly produced biosynthetically, but in *E. coli*, they are produced as a response to acid stress, through specific acid-induced decarboxylation of arginine, ornithine, and lysine and subsequent export of agmatine, putrescine, and cadaverine (Kanjee *et al.*, 2011a; Kanjee *et al.*, 2011b). Together with Mg^{2+}, polyamines account for the majority of intracellular cationic charges, and are essential for normal cell growth and viability (Canellakis *et al.*, 1979).

Although the effect of polyamines on cell growth is believed to occur at the level of translation, it has been shown that polyamines can also regulate gene expression at the transcriptional level *in vitro* by affecting the binding of regulatory proteins to DNA (Igarashi & Kashiwagi, 2006). The polyamine modulon in *E. coli* has been defined as a group of genes whose expression is regulated by polyamines and comprises several transcription factors whose translation is affected by polyamines (Igarashi & Kashiwagi, 2006). Through experimentation it was found that protein synthesis from mRNAs having a weak Shine-Dalgarno sequence is enhanced by polyamines. An example is FecI, a transcriptional regulator involved in the expression of the iron uptake operon (fecABCDE (Visca *et al.*, 2002)). The level of FecI protein is significantly higher in cells cultured in the presence of putrescine. However, the level of FecI mRNA in these cells is ~ 70% of the level in the absence of putrescine. This disparity between mRNA and protein levels is due to a weak Shine-Dalgarno-like sequence in the FecI mRNA, which is enhanced in the presence of polyamines (Yoshida *et al.*, 2004). Another example is the *Burkholderia pseudomallei* BpeAB-OprB resistance-nodulation-division (RND) family pump which is known to participate in the efflux of aminoglycosides, macrolide antibiotics, and acylhomoserine lactones and has been implicated

in spermidine homeostasis in *B. pseudomallei* (Chan & Chua, 2010). This could also be a possible secondary function of the efflux systems studied in this work.

Both AmvR and AdeL show some specificity to the polyamines putrescine and spermidine. However, it is important to remember that DSF only provides leads as to the chemical functional groups that can promote stability within a protein. It is therefore likely that AmvR and AdeL are responding to chemical changes brought about by the amine groups on these polyamines.

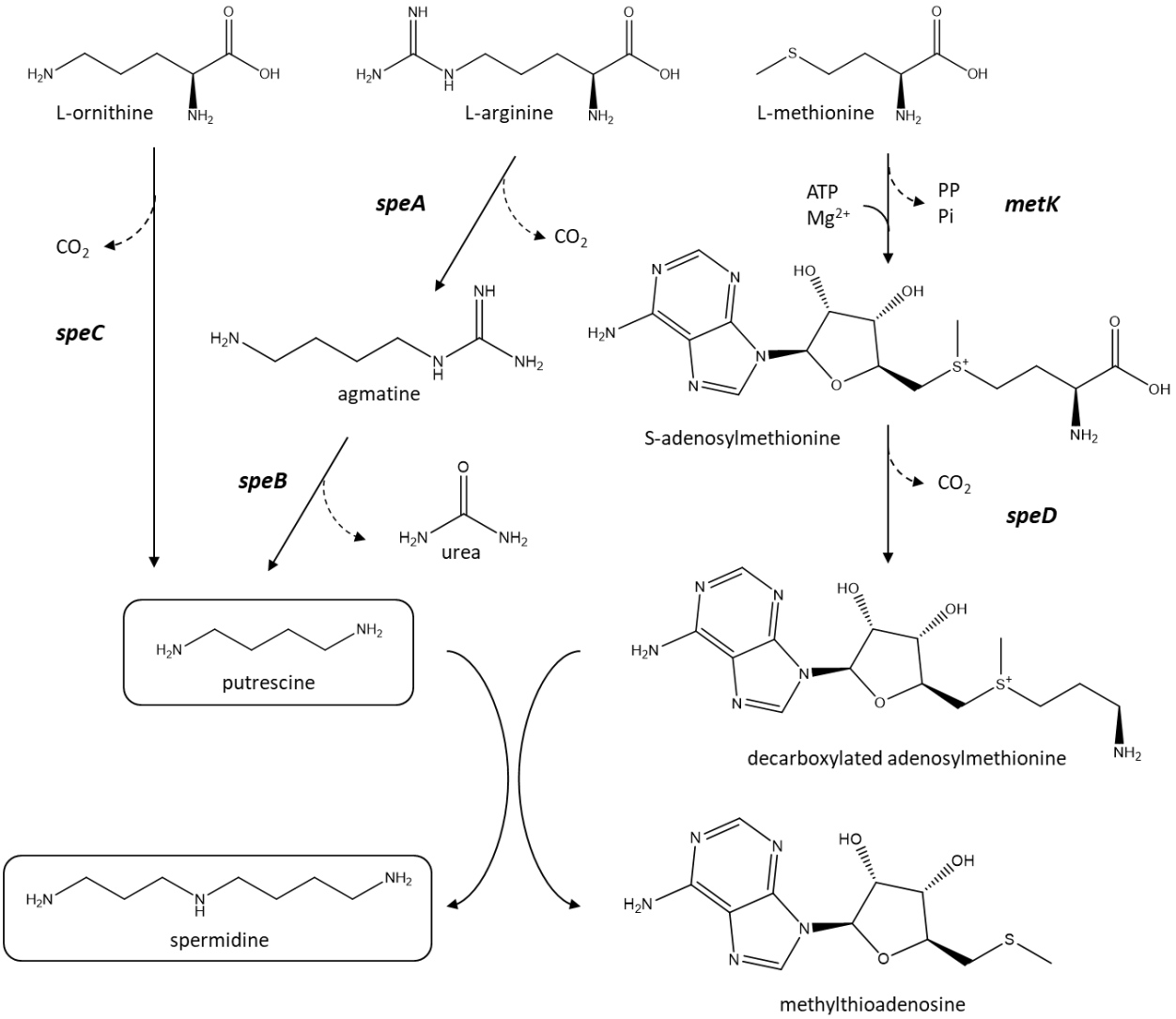


Figure 7.7 Pathway for the biosynthesis of putrescine (1,4-diaminobutane) and spermidine in *E. coli*. Adapted from (Tabor & Tabor, 1985).

7.3.2. Cystamine metabolism in bacteria

There has been very little study on cystamine metabolism in bacteria. Generally, it seems that cystamine is used solely as a means of acquiring cysteamine, a sulfhydryl, which has the ability to protect against oxidative stress (Thomas, 1984).

Sulfhydryl compounds have specific functions which include providing a reducing environment that protects protein sulfhydryls against oxidation and other forms of chemical modification (Singhal *et al.*, 2015; Giles *et al.*, 2017). Protection afforded by such compounds may be especially important to oral pathogens, such as *Streptococcus mutans* (Hamada & Slade, 1980), facultative anaerobes that depend primarily on glycolysis for their energy metabolism (Hogg, 2005), but take up O_2 at rates comparable to those of aerobic organisms (Anders *et al.*, 1970). A portion of the O_2 taken up by *streptococci* is converted to potentially toxic electrophilic agents, including O_2 and H_2O_2 (DiGuiseppi & Fridovich, 1982). Incubation of *S. mutans* cells with glucose and certain disulfide or sulfhydryl compounds results in a dramatic increase in cell sulfhydryl content and increased resistance to at least one oxidising agent (Thomas *et al.*, 1983).

As *A. baumannii* can cause pneumonia in immunocompromised patients, the bacteria would be present in the same sort of environment as *S. mutans* and therefore may need the same mechanisms for handling oxidative stress. Although my results show that both AmvR and AdeL bind cystamine, they would also have the potential to bind cysteamine. This would result in deregulation of their constitutive efflux pumps, which could then export the sulfhydryls into either the periplasm or the environment surrounding the cell to counteract oxidative conditions.

As the binding affinity of AmvR and AdeL for more common polyamines is lower than their affinity for cystamine, it was hypothesise that cystamine or cysteamine are possible ligands of both the AmvA and AdeFGH efflux systems. However, there may be an underlying redox mechanism that is affecting protein stability when cystamine is present. As previously stated, both AmvR and AdeL contain an odd number of cysteines and therefore are vulnerable to oxidative and reductive conditions. There are only a couple of examples of TFR and LTTR regulators being directly involved in preventing oxidative stress in bacteria and only one uses a cysteine-based redox

switch. The most well-documented case is that of the LTTR protein OxyR, the master peroxide sensor in Gram-negative bacteria that regulates the transcription of defence genes in response to a low level of cellular H_2O_2 (Jo *et al.*, 2015). This protein contains an intramolecular disulfide bond that upon oxidation, facilitates a large movement of the DBD. The other is TetR from a horizontally transferred gene cluster in *Acidovorax avenae* which contributes to cell survival under oxidative stress (Liu *et al.*, 2014). When this repressor gene is deleted, oxidative stress resistance was enhanced in the bacterium.

There are, however, no records of transcriptional regulators in the TFR or LTTR families that employ intermolecular disulfide bonds in their regulatory roles. Cysteine-mediated redox switches are common among peroxiredoxins. Peroxiredoxins are a superfamily of ubiquitous cysteine peroxidases that decompose cellular hydroperoxides including hydrogen peroxide (H_2O_2 (Yewdall *et al.*, 2018)). A particular subclass of peroxiredoxins consist of homodimers which can assemble into decameric or dodecameric rings depending on the redox status of the peroxidatic cysteine residue (Karplus, 2015). This peroxidatic cysteine residues is housed in a highly conserved active site. As there is little conservation of cysteine residues across TFR and LTTR proteins, it is unlikely that AmvR and AdeL contain an intermolecular disulfide bond that is paramount to oligomerisation.

7.4. Future Work

This body of work is a good starting point for further examination of the efflux systems studied here. The first step in expanding the knowledge gained here would be to further test the binding capacity of each regulator, in conjunction with its proposed ligand, to intergenic DNA. As mentioned above, this has already been achieved with AceR in the presence of chlorhexidine (Qi Liu, 2017). Both electrophoretic mobility shift assays and DNase I footprinting would be useful to show if AmvR and AdeL bind to intergenic DNA found between the *amvR/adel and amvA/adeFGH* genes, respectively, in the presence of cystamine.

The next step would be to understand the redox potential of AmvR and AdeL. This could be achieved through a number of techniques. The first would be to purify each protein under reducing conditions and then to perform SEC under these same conditions. If there was a redox-mediated oligomerisation event, changing these conditions could lead either to monomeric or tetrameric species in solution, depending on the mechanism. Another option would be to mutate cysteine residues to serines in order to remove functional sidechain groups. This could either be carried out using single-residue mutants, by mutating pairs, or even going as far as to remove all cysteine residues from the proteins.

It will also be important to look at the structure and function of each regulator protein more closely. As hinted at earlier, it might be appropriate to screen a wider group of compounds than the relatively limited pool used here. Such screens are available either commercially or through drug-screening enterprises. A wider range of compounds may dispel the obvious bias toward polyamines seen during the testing of AmvR and AdeL.

Despite the valuable data acquired in this work, it is still ideal to have a protein structure when determining the specificity of ligands and their role in the functionality of the protein. High resolution structures are able to provide exact chemical compositions of proposed active sites. In the case of AdeN, this would give great insight as to what sort of molecule could bind in lieu of thermal melt data. It has already been hypothesised that AdeN would most likely bind a hydrophobic molecule, based on the large percentage of surface exposed hydrophobic residues and the active site structure of its closest homolog CmeR, but knowing the active site chemistry would confirm this. An additional advantage provided by crystallisation is the possibility of co-crystallising the ligand within the structure, giving exact placement and contacts with sidechains of the active site.

Protein structures also open the possibility of discovering the unknown, such as secondary binding sites (e.g. the one previously discussed within BenM (Ezezika *et al.*, 2007)) or molecules from growth media that have bound to the protein without prior knowledge (e.g. the cofactor NADP within the extended short-chained dehydrogenase/reductase WbjB (Shah *et al.*, 2018)). Being able to visualise a protein structure would be especially useful for AceR, as it is hypothesised that

it has sequestered a small molecule from the bacterial growth media. Protein structures can also reveal movement upon binding of ligands. As this movement is known to be imperative to the regulatory actions of TFRs and LTTRs (Yu *et al.*, 2010), being able to see it and measure the change when a proposed ligand is bound would help solidify it as an effector molecule.

Another technique that would be useful for determining if AceR binds to a molecule present in culture medium is mass spectrometry. In parallel to the use of mass spectrometry in proteomics for primary structure elucidation, the technique has become a complementary tool in structural biology for the investigation of secondary, tertiary, and quaternary structures of protein complexes, and their interactions with DNA, RNA, ligands, and cofactors (Miranker *et al.*, 1993; Mirza *et al.*, 1993). In particular, electrospray ionisation (ESI) is well suited to detect and investigate non-covalent complexes by transferring whole intact assemblies into the vacuum inside the mass spectrometer (Heck & Van Den Heuvel, 2004).

The data collected in this research suggests that efflux systems are not only used by bacteria in the extrusion of antibiotics. Efflux pumps are fundamental to the physiology of bacteria. The RND efflux pumps of Gram-negative bacteria are required for virulence in their specific host (Bina *et al.*, 2008; Chan & Chua, 2005; Buckley *et al.*, 2006; Jerse *et al.*, 2003), exporting determinants such as adhesins, toxins or other proteins that are important for the colonisation and infection of human and animal cells (Jerse *et al.*, 2003; Burse *et al.*, 2004). They also contribute to biofilm formation (Baugh *et al.*, 2012; Kvist *et al.*, 2008; Blair *et al.*, 2014). Some, such as the RND efflux pump ZrpADBC from *Serratia* sp. ATCC 39006 (Gristwood *et al.*, 2008), have the ability to recognise toxic compounds derived from bacterial metabolism, and hence perform excretory functions as well (Li & Nikaido, 2009; Nikaido, 2011). These findings have led to the knowledge that the physiological role of these systems is evasion of such naturally produced molecules, thereby allowing the bacterium to survive in its ecological niche (Nikaido *et al.*, 1998; Prouty *et al.*, 2004).

CHAPTER 8

Materials and Methods

8.1. Materials

8.1.1. Reagents

All chemical reagents used were of analytical grade or higher and most obtained through mainstream suppliers (see **Table 8.1**). Primers and genes were sourced through Integrated DNA Technologies (Baulkham Hills, NSW), and restriction enzymes through Promega (Alexandria, NSW) or New England Biolabs (Arundel, Qld). DNA sequencing was processed by Macrogen (Seoul, South Korea). Kits for DNA extraction (QIAprep Spin Miniprep) and purification (QIAquick PCR Purification) were supplied by QIAGEN. Purified water from a MilliQ system (Millipore) was used for preparation of all buffers and reagents. All buffers and water were degassed and filtered before use in chromatography steps.

8.1.2. Growth media and buffers

All growth media were prepared with components as defined in **Table 8.2** and autoclaved before use. Buffers and growth media were stored at 4 ˚C, except for SOC media which was kept at -20 ˚C.

Table 8.1 Chemical reagents used in this work

reagents	supplier	reagents (cont.)	supplier	reagents (cont.)	supplier
acetic acid, glacial	*VWR*	glycine	*Astral*	spermidine	*Sigma-Aldrich*
acrylamide/Bis solution	*Bio-rad*	HEPES* free acid	*Astral*	spermine	*Sigma-Aldrich*
agar, bacteriological	*Amresco*	hydrochloric acid	*Panreac*	SYPRO orange	*Invitrogen*
agarose	*Panreac*	isopropyl alcohol	*Amresco*	TCEP*	*Sigma-Aldrich*
APS	*Astral*	D-lactose monohydrate	*Astral*	TEMED*	*Bio-rad*
ammonium chloride	*Amresco*	magnesium chloride hexahydrate	*Amresco*	Tris* HCl	*Amresco*
ammonium sulfate	*Amresco*	magnesium sulfate heptahydrate	*Astral*	tryptone	*Amyl Media*
boric acid	*Univar*	MES* buffer	*Astral*	yeast extract	*Amresco*
Bromophenol Blue	*Progen*	nickel(II) chloride hexahydrate	*Sigma-Aldrich*		
cadaverine	*Sigma-Aldrich*	nickel(II) sulfate hexahydrate	*Sigma-Aldrich*		
calcium chloride	*Amresco*	potassium chloride	*Astral*		
Coomassie brilliant blue	*Amresco*	potassium dihydrogen phosphate	*VWR*		
cystamine dihydrochloride	*Sigma-Aldrich*	potassium phosphate monobasic	*Amresco*		
DTT	*BDH*	putrescine	*Sigma-Aldrich*		
1,8-diaminooctane	*Sigma-Aldrich*	SDS	*Amresco*		
EDTA* disodium salt	*Amresco*	sodium chloride	*VWR*		
ethanol, absolute	*VWR*	sodium hydroxide, pellets	*Panreac*		
D-glucose	*Amresco*	sodium phosphate dibasic	*Astral*		
glycerol	*VWR*	sodium sulfate decahydrate	*Sigma-Aldrich*		

* APS (ammonium persulfate), EDTA (ethylenediaminetetraacetic acid), HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), MES (2-(N-morpholino)ethanesulfonic acid), TCEP (tris(2-carboxyethyl)phosphine), TEMED (tetramethylethylenediamine), Tris (trisaminomethane).

Table 8.2 Composition of growth media

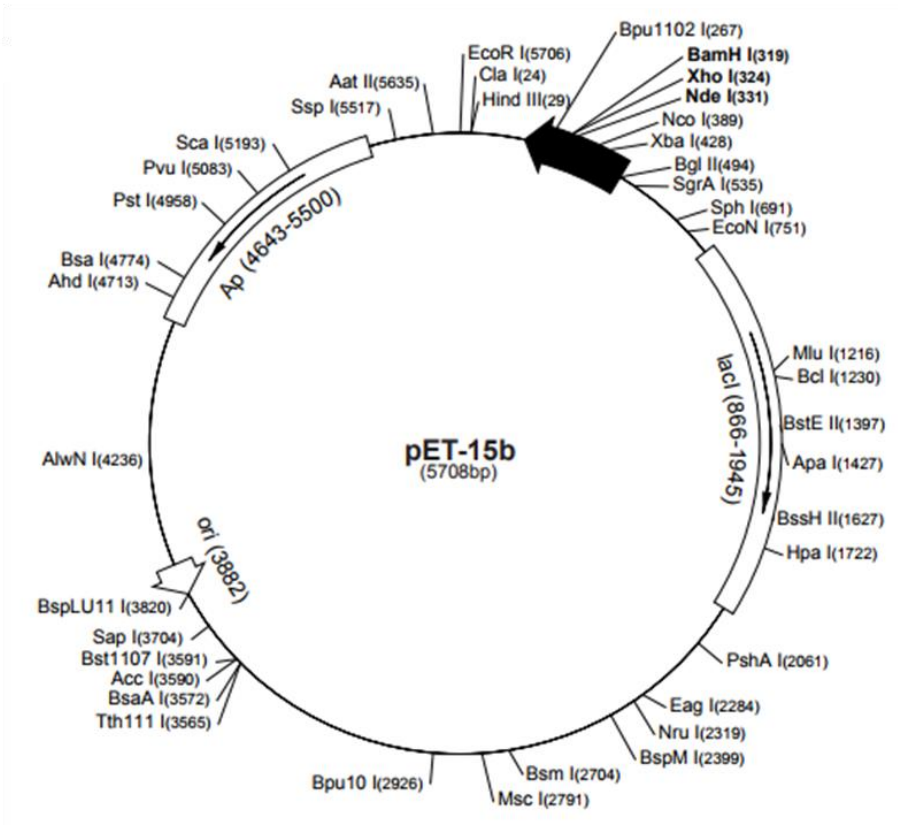
growth media (1 L) [a]	composition	reference
SOC medium	20 g tryptone, 5 g yeast extract, 0.58 g NaCl, 0.18 g KCl, 0.95 g $MgCl_2$, 1.20 g $MgSO_4$, 20% glucose [b]	(Sambrook & Russell, 2001)
Luria Bertani (LB) broth	10 g tryptone, 5 g yeast extract, 10 g NaCl	(Sambrook & Russell, 2001)
5052 media (50x)	250 g glycerol, 25 g glucose, 100 g α-lactose	(Studier, 2005)
NPS media (20x)	66 g $(NH_4)_2SO_4$, 136 g KH_2PO_4, 142 g Na_2HPO_4 (pH 6.75)	(Studier, 2005)
ZY media	10 g tryptone, 5 g yeast extract	(Studier, 2005)
ZYP-rich media	1 ml 1 M $MgSO_4$, 20 ml 50x 5052 media, 50 ml 20x NPS, 1 L ZY media	(Studier, 2005)
LB agar	10 g tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacteriological agar	(Sambrook & Russell, 2001)
M9 salts	15 g KH_2PO_4, 64 g $Na_2HPO_4 \cdot 7H_2O$, 2.5 g NaCl, 5 g NH_4Cl	(Sambrook & Russell, 2001)

[a] For antibiotic selectivity, ampicillin (100 mg/ml, Amresco) and chloramphenicol (25 mg/ml, Amresco) were added to listed components when required.
[b] Added after autoclaving.

Table 8.3 Competent cell lines used in this study

strain	genotype	source
BL21	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*	New England Biolabs
BL21(DE3)pLysS	*F^- ompT $hsdS_B(r_B^- m_B^-)$ gal dcm (DE3) pLysS (Cam^R)*	Novagen
dH5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
Stellar	F–, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ (lacZYA - argF) U169, Δ (mrr - hsdRMS - mcrBC), ΔmcrA, λ–	Clonetech
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB $lacI^q$ ZΔM15* Tn*10* (Tet^r)]	Agilent

A



T7 promoter → lac operator — *Xba* I — rbs

```
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
```

Nco I — His_6 tag — *Nde* I *Xho* I *Bam*H I

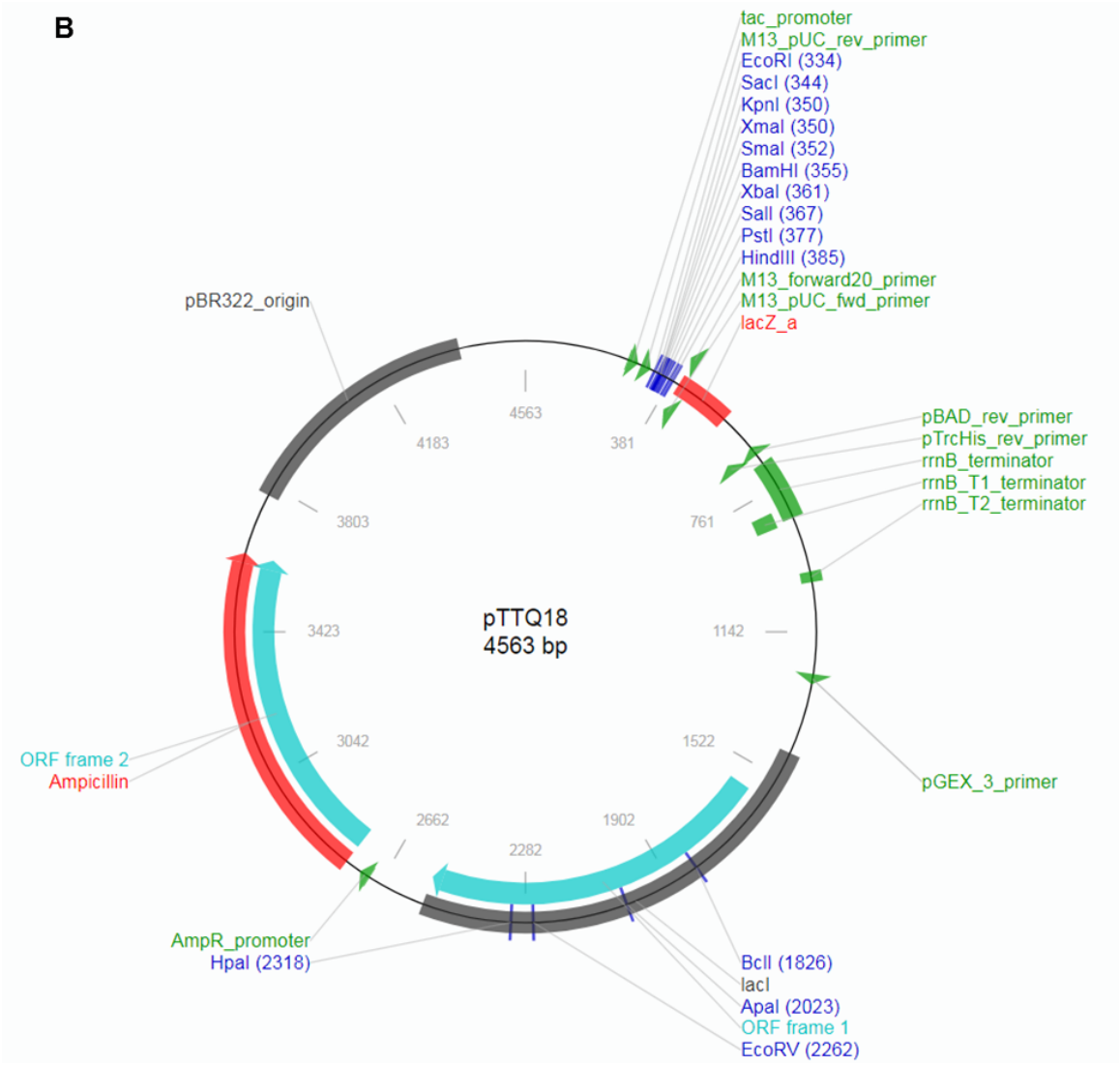
```
TATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCGA
      MetGlySerSerHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerHisMetLeuGluAspProAlaAlaAsnLysAlaArg
```

thrombin

```
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
LysGluAlaGluLeuAlaAlaAlaThrAlaGluGlnEnd
```

pET-15b cloning/expression region

Figure 8.1 Plasmid map of pET-15b and $pTTQ18_{RGSH6}$ vectors. (**A**) Circular plasmid map of pET-15b (Novagen (Merck, 2017b)) with a highlight of the cloning site below. (**B**) Circular plasmid map of $pTTQ18_{RGSH6}$ (Addgene (Stark, 1987; Saidijam *et al.*, 2006)) with a highlight of the cloning site seen below. Primers used for sequencing are labelled.



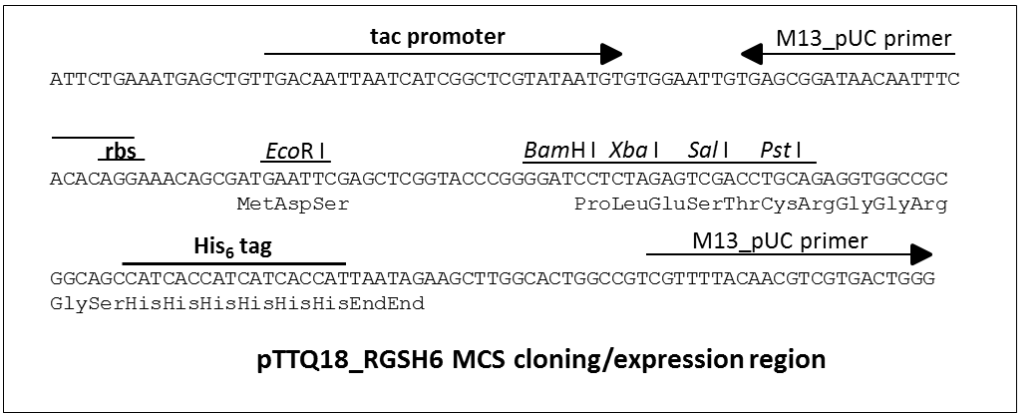


Figure 8.1 continued.

8.1.3. Bacterial strains and plasmids

DNA propagation was achieved using Stellar™ (Clontech (Clontech, 2017)) and DH5α (New England Biolabs (Biolabs, 2017b)) competent cells (**Table 8.3**). BL21(DE3)pLysS competent cells (Novagen) were used for protein expression where genes of interest were inserted in a pET-15b plasmid (**Figure 8.1 A**). This cell line allows inducible protein expression under the control of a T7 promoter (Merck, 2017a). For protein expression of genes located in a $pTTQ18_{RGSH6}$ plasmid (**Figure 8.1 B** (Stark, 1987; Saidijam *et al.*, 2006)), BL21 competent cells were used, due to their ability to control expression of a tac promoter (Biolabs, 2017a). For propagation of site-directed mutant genes, XL1-Blue supercompetent cells were used according to manufacturer's instructions (Agilent (Technologies, 2015)).

8.1.4. Preservation of bacterial strains

All bacterial strains were preserved as glycerol stocks (-80 ˚C). These stocks were prepared by shaking cells overnight (37 ˚C, 250 rpm) in LB broth (5 ml) with ampicillin (2.5 µl). Cultures were then spun down (2200 x *g*, 10 min, 4 ˚C), and resuspended in M9 salts (750 µl) and 50% glycerol (750 µl).

8.2. Methods

8.2.1. Molecular biology procedures

8.2.1.1. Plasmid isolation

Vectors pET-15b and $pTTQ18_{RGSH6}$ were isolated from laboratory glycerol stocks (Stellar and dH5α competent cells, Section 8.1.3) using a commercial kit (QIAprep, QIAGEN) according to manufacturer's instructions. Isolated plasmids were stored in supplied Tris buffer (10 mM, pH 8.0) and kept at -20 ˚C.

8.2.1.2. Plasmid digestion

The pET-15b plasmid was digested at 37 ˚C with NdeI and BamHI restriction enzymes. To digest 30 µl of vector, the following reagent mix was used: BSA (1 µl), 10x NEB buffer #3 (10 µl), NdeI (4 µl), and water (53 µl). After 3 h, BamHI (2 µl) was added and the mixture left for another 2 h. Digested vector was purified by kit (QIAquick, QIAGEN) according to manufacturer's instructions, and visualised by electrophoresis (Section 8.2.3.1). The $pTTQ18_{RGSH6}$ plasmid was digested at 37 ˚C with EcoRI and PstI restriction enzymes (Promega). To digest 30 µl of vector the following reagents were used: BSA (1 µl), Buffer H (10 µl), EcoRI (4 µl), and water (53 µl). After 3 h, PstI (4 µl) was added and the mixture left for another 2 h. Digested vector was purified and visualised as mentioned previously.

8.2.1.3. Extraction of genomic DNA

Genomic DNA was obtained from *Acinetobacter baumannii* ATCC 17978 (kind gift of Karl Hassan, GenBank accession no. NC009085). Fresh bacterial colonies were grown on non-selective agar plates from glycerol stocks. Four colonies were picked, added to LB broth, and incubated at 37 ˚C overnight with shaking (250 rpm). Chromosomal DNA (100 µl) was extracted from each culture following the protocol for Gram-negative bacteria using a commercial kit (DNeasy blood and tissue kit, QIAGEN).

8.2.1.4. Amplification of genes from genomic DNA

Primers for gene amplification from genomic DNA, incorporating a TEV cleavage site sequence, were designed using Oligocalc and Serial Cloner (**Table 8.5** (Kibbe, 2007)). These were diluted to 50 µM in Tris buffer (10 mM, pH 8.0). Target genes were amplified using standard PCR reagents (see **Table 8.4**) and the following thermal cycle (94 ˚C (5 min), 35 x [94 ˚C (30 s), 55 ˚C (30 s), and 72 ˚C (90 s)], final extension 72 ˚C (10 min)). Products were purified by kit (QIAquick, QIAGEN) and visualised by electrophoresis (Section 8.2.3.1).

Table 8.4 Reagent mixtures used for polymerase chain reaction (PCR) procedures

components	gene amplification	colony PCR
commercial buffer	10x KOD buffer 1 (5µl)	10x Taq buffer (5 µl)
dNTPs (10 mM)	1 µl	1 µl
$MgCl_2$ (25 mM)	3 µl	2 µl
forward primer	1 µl (0.3 µM)	1 µl
reverse primer	1 µl (0.3 µM)	1 µl
target DNA	1 µl (200 ng)	1 µl
polymerase	KOD (0.4 µl)	Taq (0.4 µl)
sterile water	37.6 µl	38.6 µl

Table 8.5 Primers for gene amplification from *A. baumannii* ATCC 17978

gene product [a]	primer sequence [b]
AdeN (A1S_1979)	F- ttcacacaggaaacagcgATGCATGATCCAGTCCTTGAG R- cggccacctctgcagccggattggaagtacaggttctcGACTTTATGATGCCC
AdeN G65	F- ttcacacaggaaacagcgATGGGTAATAAAGATGGCTTATTTACT
AmvR (A1S_2058)	F- ttcacacaggaaacagcgATGGCCTATCTTAATCGCGAT R- cggccacctctgcagccggattggaagtacaggttctcTAATAATTCTAGGCG
AmvR S51	F- ttcacacaggaaacagcgATGTCGGCATCTCATTTAAAAGCT
AceR (A1S_2064)	F- ttcacacaggaaacagcgATGAATATTAATCAAGAACAACTTCTCATG R- cggccacctctgcagccggattggaagtacaggttctcCGGAGCTGGTTGCAT
AceR A86	F- ttcacacaggaaacagcgATGGCATTAAGTACTGGGCTG
AdeL (A1S_2303)	F- ttcacacaggaaacagcgATGAGAGTATTCAACAAAGTTGTTGAA R- cggccacctctgcagccggattggaagtacaggttctcAGTTTTGAGCGTATA
AdeL P54	F- ttcacacaggaaacagcgATGCCGGATGGCGCCGTA
AdeL P84	F- ttcacacaggaaacagcgATGCCAAGAGGTCAGCTT

[a] Gene identifiers in *A. baumannii* ATCC 17978.
[b] Lower case letters are vector specific and upper case are gene specific.

Table 8.6 Primers for gene insertion into pET-15b plasmid

gene product [a]	primer sequence [b]
Hfx1 (CAP47818.1)	F- gcgcggcagccatATGGAGAGCGTGAATAC R- gttagcagccggatccTTACAGAAAATCTTTATAATATTCATCTTTGTCC
Hfx5 (3IF4_A)	F- gcgcggcagccatATGAAACAAGAATTCGTTGCG R- gttagcagccggatccTTATGCGGCGGCC
Vch3 (WP_095490559.1)	F- gcgcggcagccatATGACCGAGGTTAACC R- gttagcagccggatccTTAAAAATATTTGCGCCAAAAGTCTG
Vch14 (CBB93061.1)	F- gcgcggcagccatATGGCACTGACAGTAAAG R- gttagcagccggatccTTACAGGCCTTTAAAAATTTCGAC

[a] GenBank accession ID (Sureshan *et al.*, 2013).
[b] Lower case letters are vector specific and upper case are gene specific.

8.2.1.5. Preparation and amplification of DNA gBlocks

Full length gene sequences were ordered as gBlocks (200 ng) with required primers designed using Oligocalc (**Table 8.6** (Kibbe, 2007)). Primers were diluted 1:10 into Tris buffer (10 mM, pH 8.0) to create working stocks. gBlock DNA was diluted and amplified according to supplier's recommendations. PCR amplification utilised standard reagents (**Table 8.4**), and the following thermal cycle (95 ˚C (2 min), 12 x [95 ˚C (20 s), 55 ˚C (10 s), and 70 ˚C (15 s)], final extension 70 ˚C (5 min)). Products were visualised by electrophoresis (Section 8.2.3.1) on agarose gel and DNA concentrations determined with a Nanodrop spectrophotometer (Thermo Scientific).

8.2.1.6. Ligation-independent cloning

I) Novel-fold genes

Cloning reactions were carried out using a 2:1 molar ratio of insert:vector, with insert concentrations calculated based on the supplied vector quantity (100 ng) as recommended by

the manufacturer. Mixtures were diluted in water to a total of 5 µl, added to supplied pellets (1 pellet per gene sequence, Clontech), and incubated at 37 ˚C (15 min). Following heating at 50 ˚C (15 min), the mixture was placed on ice in preparation for further steps. Tris-EDTA (TE) buffer (40 µl) was added to each reaction mixture, and then 2.5 µl of the mixture was added to thawed aliquots of competent cells (25 µl, Stellar). Components were then cooled on ice (30 min), heat shocked (42 ˚C, 45 s), and again cooled (0 °C, 2 min). SOC (450 µl, room temperature) was added to each mixture and cells were recovered by shaking at 37 ˚C (1 h, 250 rpm). Cells were plated on ampicillin agar plates at two separate quantities (150 µl and 300 µl) and incubated overnight at 37 ˚C.

II) Efflux regulator genes

Infusion cloning reactions were carried out with a 5:1 molar ratio of insert:vector. An excess of insert gene over recommended quantities (see above) was required for successful cloning. Insert concentrations were calculated based on 100 ng of vector, as recommended by the manufacturer (BioTool, 2017). Mixtures of insert, vector, commercial fusion enzyme (1 µl, BioTool), and 5x fusion buffer (2 µl, BioTool) were diluted with water to a total volume of 10 µl. The mixture was incubated at 37 ˚C (30 min), and then transferred to ice. This ligation mixture (2.5 µl) was added to aliquots of competent cells (25 µl, Stellar), mixed, cooled (0 °C, 30 min), heat shocked (42 ˚C, 60 s), and then quenched on ice (2 min). SOC (500 µl, room temperature) was added to each mixture and cells were recovered at 37 ˚C (1.5 h) with shaking (250 rpm). Aliquots of cells (100 µl) were plated on ampicillin agar plates and incubated overnight at 37 ˚C.

Genes encoding the efflux regulator AceR were previously cloned into dH5α competent cells (kind gift of Qi Liu, Macquarie University).

8.2.1.7. PCR colony screening

To confirm correct insertion of genes, colony PCR (Bergkessel & Guthrie, 2013) was carried out using T7 primers or M13 primers for pET-15b and pTTQ18$_{RGSH6}$, respectively. Relevant enzymes and reagents (**Table 8.4**) were mixed with a scraping of cells, and the following thermal cycle performed (94 ˚C (3 min), 30 x [94 ˚C (30 s), 55 ˚C (30 s), and 72 ˚C (30 s)], final extension 72 ˚C (5 min)). Products were purified by kit (QIAquick, QIAGEN), visualised on agarose gel (Section 8.2.3.1), and subsequently verified by sequencing.

8.2.1.8. Preparation of competent expression host cells

Aliquots (50 µl) of competent cells (BL21(DE3)pLysS or BL21) were combined with 2 µl of purified plasmid (pET-15b or pTTQ18$_{RGSH6}$, respectively) and cooled on ice (30 min). Following heat shock (42 ˚C, 30 s) and cooling on ice (2 min), SOC medium (250 µl, room temperature) was added and cells allowed to recover (37 ˚C, 1.5 h) with shaking (250 rpm). Cells were incubated overnight at 37 ˚C on agar plates containing either ampicillin, or ampicillin + chloramphenicol (Section 8.1.2). Three colonies were picked from each culture, replated on fresh selective agar plates, and incubated overnight, as before. Plasmid extraction was performed for each culture group using a commercial kit, and the purified vector sequenced to confirm integrity of the incorporated gene.

8.2.1.9. Site-directed mutagenesis

Primers for single-site mutagenesis were designed with the software tool OligoCalc (Kibbe, 2007) to have a melting temperature of 78 ˚C and a GC content close to 40%. Designed sequences for Hfx1 and Vch14 are listed in **Table 8.7**.

Table 8.7 Primer design for generation of single, site-specific mutants of Hfx1 and Vch14

target mutation	primer sequence [a]
Hfx1	
S10A	F - ACGTCCTTCCTC**GCG**CCATCCCTCGTAACCATTCGGG R - CCCGAATGGTTACGAGGGATGGCGCGAGGAAGGACGT
S57A	F - GGTGTGCGCGATGCTCAACAG**GCG**ATTGGCGATG R - CATCGCCAATCGCCTGTTGAGCATCGCGCACACC
S99A	F - CTGATCGTTGGAACCGGAGCT**GCG**GAAGTCGAAC R - GTTCGACTTCCGCAGCTCCGGTTCCAACGATCAG
S143A	F - GAGGAGTTACTGAGC**GCG**GATTCTTTCCACCCGG R - CCGGGTGGAAAGAATCCGCGCTCAGTAACTCCTC
I16V	F - CCCTCGTAACC**GTG**CGGGACTTTGACAACGG R - CCGTTGTCAAAGTCCCGCACGGTTACGAGGG
I29V	F - GGTCTTGCGT**GTG**GGACGTACCGGCTTTCCGGCCG R - CGGCCGGAAAGCCGGTACGTCCCACACGCAAGACC
I78V	F - CCGGTGCGTGGAC**GTG**GATGATAAACATACCTATAATGCC R - GGCATTATAGGTATGTTTATCATCCACGTCCACGCACCGG
I93V	F - GCCATGGTATACGTTGATCTG**GTG**GTTGGAACCGGAGCTAGT R - ACTAGCTCCGGTTCCAACCACCAGATCAACGTATACCATGGC
Y154F	F - CTGAGCTCAGATTCTTTCCACCCGGACAAAGATGAATAT**TTC**AAAGATTTTCT R - AGAAAATCTTTGAAATATTCATCTTTGTCCGGGTGGAAAGAATCTGAGCTCAG
C44S	F - GGCGATATTGACCTG**TCT**CTGGACAAAATGAAAGGTGTGCGCG R - CGCGCACACCTTTCATTTTGTCCAGAGACAGGTCAATATCGCC
C75S	F - AGGTCCGCACATTCGTATCCGG**TCT**GTGGACATTGATG R - CATCAATGTCCACAGACCGGATACGAATGTGCGGACCT
C128S	F - GGTGGATATTGCAGACGAACACAGC**TCT**GTGACGCAGTTTGAAATG R - CATTTCAAACTGCGTCACAGAGCTGTGTTCGTCTGCAATATCCACC
H70L	F - CGGCTTCAAAGGTCCG**CTG**ATTCGTATCCGGTGCG R - CGCACCGGATACGAATCAGCGGACCTTTGAAGCCG
H82L	F - CCGGTGCGTGGACATTGATGATAAA**CTG**ACCTATAATGCCATGG R - CCATGGCATTATAGGTCAGTTTATCATCAATGTCCACGCACCGG
H126L	F - GCAGGTGGATATTGCAGACGAA**CTG**AGCTGTGTGACGCAG R - CTGCGTCACACAGCTCAGTTCGTCTGCAATATCCACCTGC
H147L	F - GGAGTTACTGAGCTCAGATTCTTTC**CTG**CCGGACAAAGATGAATATT R - AATATTCATCTTTGTCCGGCAGGAAAGAATCTGAGCTCAGTAACTCC
Vch14	
S12A	F - GGACGTTAATATCCTG**GCG**CAGTATATCTCAGGCGTGATGGCG R - CGCCATCACGCCTGAGATATACTGCGCCAGGATATTAACGTCC
S16A	F - CCCAGTATATC**GCG**GGCGTGATGGCGCGTGCGGACCACCACG R - CGTGGTGGTCCGCACGCGCCATCACGCCCGCGATATACTGGG
S79A	F – GCTATGCGTTCAGTTATAATCACTCA**GCG**GAAAAAATCGAAATGCG R - CGCATTTCGATTTTTTCCGCTGAGTGATTATAACTGAACGCATAGC
I66V	F - CGTTCTGTGGGTGACA**GTG**AATGGTGAGCGCTATGCG R - CGCATAGCGCTCACCATTCACTGTCACCCACAGAACG

[a] Bold blocked out letters indicate alteration sites.

Supplied oligonucleotides were diluted (1:10) to make working stocks (30 µl). PCR was performed with 125 ng of forward and reverse primers, pET-15b plasmid (containing target genes, 100 ng)

Supplied oligonucleotides were diluted (1:10) to make working stocks (30 µl). PCR was performed with 125 ng of forward and reverse primers, pET-15b plasmid (containing target genes, 100 ng) and kit reagents (10x Pfu buffer (5 µl), dNTPs (1 µl), Pfu Ultra (1 µl), and water (to final volume 50 µl), Quikchange II, Agilent). The following thermal cycle was used (95 ˚C (1 min), 16 x [95 ˚C (1 min), 55 ˚C (1 min), and 68 ˚C (7 min)]).

To each PCR product, Dnp1 enzyme was added (1 µl, Agilent) and left to incubate (37 ˚C, 1 h). Transformation of Dnp-treated DNA (1.5 µl) into thawed competent XL1-Blue cells (50 µl, Agilent) was initiated by 30 min at 0 ˚C. Samples were then heat shocked (30 s, 42 ˚C) and cooled (2 min, 0 ˚C). Following addition of SOC medium (500 µl, room temperature), samples were incubated (2 h, 37 ˚C) with shaking (250 rpm). Samples (150 µl and 350 µl) were applied to ampicillin-selective LB agar plates to distinguish positive clones.

8.2.2. Recombinant protein production

8.2.2.1. Protein expression by autoinduction

Autoinduction was chosen for expression over IPTG (isopropyl β-D-thiogalactopyranoside) induction due to its tendency to yield higher concentrations of recombinant product (Studier, 2005). This method relies on medium components that are metabolised differentially than IPTG to promote high-density growth and automatic induction of protein expression from *lac*-based promoters.

BL21(DE3)pLysS or BL21 competent cells were grown in LB broth (5 ml, 37 ˚C, overnight) containing respectively ampicillin + chloramphenicol, or ampicillin only, to create starter cultures. These were then added to ZYP-rich media (500 ml, **Table 8.2**) in baffled flasks (2 L) containing selective antibiotics (Section 8.1.2). Cultures were grown at 25 ˚C to an optical density reading at 600 nm (OD_{600}) of 1.2 – 1.3 (approx. 24 h) and pelleted by centrifugation (5,000 *g*, 30 min). Recovered cells were resuspended in 30 ml HEPES buffer (pH 7.5, 50 mM, with 200 mM NaCl,

5% glycerol, 5 mM imidazole) and protease inhibitor cocktail (275 µl, Sigma Aldrich) added prior to storage (-80 ˚C).

8.2.2.2. Cell lysis

Frozen cell aliquots (35 ml) were thawed in the presence of lysozyme (275 µl, 1 mg/ml) and DNase I (20 µl, 5 mg/ml). Lysis was fully achieved through sonication (S-2500 Branson digital sonifier), with 60% amplitude and a duty cycle (10 s on, 10 s off) over 60 s on ice. After centrifugation (11,000 *g*, 40 min), and filtration (0.2 µm syringe filter on ice), protein solutions were retained for purification.

8.2.2.3. Immobilised metal affinity chromatography (IMAC)

Prepacked columns (1 ml) of Ni-Sepharose media (His trap, GE Healthcare) were used for immobilised metal affinity chromatography (IMAC). The media was washed with water (10 cv) and 10 cv of Buffer A with 5 mM imidazole (50 mM HEPES buffer, pH 7.5 with 200 mM NaCl, 5% glycerol). Cell lysate (35 ml) was loaded under pressure using a bench-top peristaltic pump. The media (generally overloaded with protein, > 40 mg/ml) was next washed with a 50 cv quantity of 50 mM imidazole in Buffer A to remove unbound proteins. The columns were attached to a liquid chromatography system (Äkta Start, GE Healthcare) for elution of His_6-tagged proteins at 0.5 ml/min with 500 mM imidazole in Buffer A. Fractions yielding absorbance readings at 280 nm (A_{280}) indicating protein content were pooled.

8.2.2.4. Buffer exchange through dialysis

For more robust proteins, eluted samples from IMAC procedures were combined with EDTA (final concentration 10 µM) and dialysed overnight in dialysis tubing (CelluSep T1, MW cut-off: 3,500). Buffers used were either Buffer A or sodium borate buffer (50 mM, pH 9.0 with 200 mM NaCl, 5%

glycerol). Following dialysis, protein samples (50 µl aliquots) were snap-frozen in liquid nitrogen for storage at -80 ˚C.

8.2.3. Molecular characterisation and functional assays

8.2.3.1. Electrophoresis

8.2.3.1.1. Agarose gel electrophoresis

All DNA samples were visualised on agarose gel containing GelRed™ (0.5 µl, Biotium). Gels of 0.8% or 1.2% agarose were used to separate larger or smaller DNA fragments, respectively (Lewis, 2011). Agarose gels (1.2%, 50 ml) were prepared by combining ingredients (0.6 g agarose, 50 ml 1x Tris-acetate-EDTA (TAE) buffer, **Table 8.8**) followed by heating in a microwave (40 s). The solution was poured and allowed to set (30 min) prior to loading of molecular weight markers (1 µl) and samples (loading dye (1 µl) and sample (2 µl)). Gels were run at 100 mV and visualised by UV light (Gel Doc EZ, Bio-Rad).

8.2.3.1.2. SDS-PAGE

SDS polyacrylamide electrophoresis gels (Brunelle & Green, 2014) were constructed as hand-poured 15% separating and 5% stacking gels and run in tris/glycine buffer (BioRad, 2017). Reagents are itemised in **Table 8.8**. Samples (3 µl) were boiled with 2x loading dye (8 µl) for 15 min prior to loading. Electrophoresis was carried out in Tris-HCl buffer (25 mM, pH 8.0) with glycine (250 mM) and SDS (10% (w/v)) at power settings of 100 V (10 min) and then at 150 V (1 h). Gels were fixed in an ethanol (50% (v/v)) and acetic acid (10% (v/v)) mix for 10 min, stained with Coomassie solution (5 min), then gently shaken with acetic acid (10% (v/v)) for 2 h. A commercial molecular size ladder (Benchmark, Invitrogen) allowed for estimation of protein size.

8.2.3.2. Size exclusion chromatography

Size exclusion chromatography (SEC) separates macromolecules according to their hydrodynamic volume, which is defined by the Stokes radius of the tumbling form in solution (Sheehan, 2009). Size exclusion media consists of porous polymer beads with clearly defined pores of specific dimensions. In mobile phase, particles with smaller hydrodynamic volumes have a longer path length (Striegel, 2009) and therefore can be separated from species with larger hydrodynamic volumes.

SEC was used in this work both for the separation of mixed oligomeric protein species and also as a means of buffer exchange (Porath & Flodin, 1959). For the latter, procedures were performed using a 120 ml column of Superdex 200 preparation grade media (Hiload 16/60, GE Healthcare), operating at 1 ml/min on an Äkta Pure system (GE Healthcare). Superdex 200 media separates over the size range of 10-600 kDa. Following equilibration with HEPES buffer (50 mM, pH 7.5 with 300 mM NaCl, 10% glycerol), sample was applied. Pooled fractions (1 ml) were concentrated by spin concentrator (Amicon Ultra, 10 kDa or 3 kDa cutoff, Millipore).

Table 8.8 Solutions used in this work for gel electrophoresis

buffer/solution	composition
agarose gel	
50 x TAE buffer (500 ml)	121 g Tris-HCl, 28.55 ml acetic acid, 50 ml EDTA (0.5 M, pH 8.0)
SDS-PAGE	
15% separating gel	1.2 ml glycerol (50% v/v), 1.25 ml tris (1.5 M, pH 8.8), 1.25 ml acrylamide (40%), 50 µl SDS (10% w/v), 30 µl APS (10% w/v), 6 µl TEMED
5% stacking gel	1.22 ml water, 0.5 ml tris (0.5 M, pH 6.8), 0.26 ml acrylamide (40%), 20 µl SDS (10% w/v), 10 µl APS (10% w/v), 2 µl TEMED, 10 µl Bromophenol Blue (10 mg/ml in water)
2 x loading dye	4% (w/v) SDS, 20% (v/v) glycerol, 200 mM DTT, 100 mM Tris buffer (pH 6.8), 0.2% (w/v) bromophenol blue
Coomassie solution	0.25% (w/v) Coomassie Brilliant Blue, 10% (v/v) ethanol, 10% (v/v) acetic acid

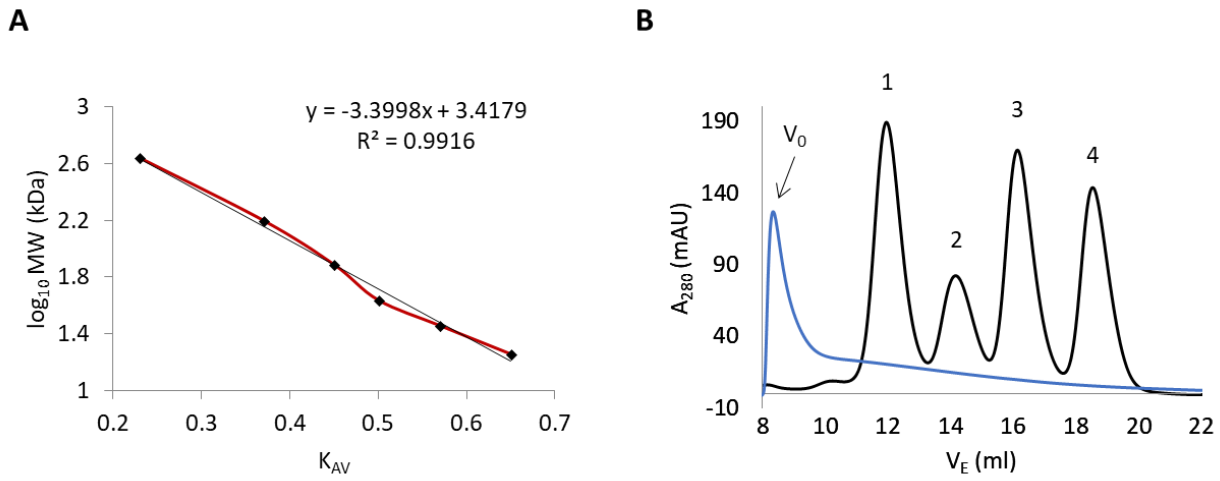


Figure 8.3 Elution behaviour for protein standards on Superdex 200 column (10 x 300 mm) eluting with HEPES buffer (pH 7.5, 50 mM HEPES, 200 mM NaCl, 5% glycerol) at 0.5 ml/min. (**A**) K_{AV} values for commercial molecular standards (ferritin (440 kDa), aldolase (158 kDa), ovalbumin (43 kDa), RNase I (13.7 kDa), conalbumin (75 kDa), and carbonic anhydrase (26 kDa)) are graphed against native mass in solution. V_0 of 8.34 ml was determined using Blue Dextran. Linear regression parameters are indicated. (**B**) Traces of V_E for size standards on the Äkta Pure (GE Healthcare). The elution peak for dextran is indicated as V_0. Elution peaks for ferritin (12.0 ml), aldolase (14.1 ml), ovalbumin (16.1 ml), and Rnase I (18.5 ml) are numbered 1-4 respectively.

For evaluation of protein mass in solution, termed the native mass, Superdex 200 matrix (GE Healthcare) was utilised in a 300 x 10 mm format. This analytical column (cv 24 ml) was equilibrated (1 h) at 0.5 ml/min on an LC system (Äkta Pure, GE Healthcare). Generally, samples were analysed in either HEPES (50 mM, pH 7.5) or glycine buffers (50 mM, pH 9.0) with the inclusion of 200 mM NaCl and 5% glycerol. Calibration of this column was carried out with commercial standards as shown in **Figure 8.3** (HMW and LMW calibration sets, GE Healthcare). The column void volume (V_0) was empirically determined to be 8.34 ml with dextran dye (GE Healthcare). The V_E of proteins was converted to partition coefficient (K_{AV}), based on the relationship: $K_{AV} = V_E – V_0 / V_T – V_0$.

8.2.3.3. Differential Scanning Fluorimetry

As discussed in Section 1.3.1, DSF is a relatively quick method to identify the characteristic T_M of a protein sample or mixture. A fluorescence response is detected through an added dye reacting to alteration of hydrophobic components of a solution (**Figure 8.4**). This occurs when a folded protein is heated and unfolds to expose interior hydrophobic sidechains.

For this work, a commercial fluorescent dye (100x SYPRO Orange, Invitrogen) was diluted into either HEPES (50 mM, pH 7.5) or glycine (50 mM, pH 9.0) buffers in the presence of 200 mM NaCl and 5% glycerol. Protein samples were added to the dye mix. The mixture was then diluted further into screening conditions at a 1:10 ratio, so resulting in a final protein concentration of 1 mg/ml. Aliquots (20 µl) were transferred to a 96-well plate in triplicate, and solutions gently mixed by plate centrifugation (1000 rpm, 1 min). The plate was heated at 1 ˚C/min over 25-95 ˚C in a real-time qPCR machine (Mx3005P, Strategene). The change of fluorescence intensity (dR, automatically baseline corrected) at 610 nm wavelength was monitored by the instrument. Once in Excel format (Microsoft), derivatives were used to calculate the transition midpoint which corresponds to the T_M. Based on the melting curve results of multiple proteins, a change of ± 2 ˚C was seen to be significant for this work.

Buffer conditions used in screening included sodium acetate buffer (0.5 M, 0.5 M acetic acid, pH 4.0 and 5.0), MES buffer (0.5 M, pH 6.0), HEPES buffer (0.5 M, pH 7.0 and 8.0), and glycine buffer (0.5 M, pH 9.0). Salt (NaCl) concentrations ranged from 50-300 mM. Small molecule additives were provided by Hampton Research in the form of 96-well Silver Bullets and Silver Bullets Bio Screens (Silver Bullets Bio conditions are denoted by an asterisk throughout this work). These chemical cocktails include small molecules ranging from amino acids and peptides, to metals and salts, and nucleotides and carbohydrates (a full chemical listing can be found in the Appendix, **Table S.1 and S.2**).

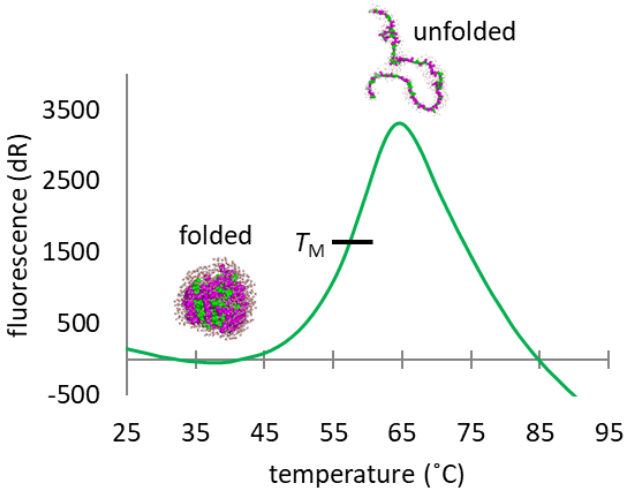


Figure 8.4 Clear DSF melting curve obtained for an AceR sample (see Chapter 6) in HEPES buffer (50 mM, pH 7.5 with 300 mM NaCl, 10% glycerol). Clear transition is seen between low temperature (folded) and high temperature (unfolded) states. The halfway point is denoted by the melting temperature, T_M.

8.2.3.4. Circular Dichroism

Circular dichroism (CD) is recognised as a valuable technique for evaluating protein secondary structure under solution conditions. The most widely used applications of protein CD are to determine whether an expressed, purified protein is folded, or if a mutation affects its conformation or stability (Greenfield, 2006). CD is reported either in units of ΔE (the difference in absorbance by an asymmetric molecule of vectors of equal length, E_R and E_L), or in degrees ellipticity (i.e. the angle whose tangent is the ratio of the minor to the major axis of the ellipse (Fasman, 1996)). Different structural elements of a protein tertiary structure yield characteristic CD spectra (**Figure 8.5**). For example, α-helical proteins tend to display negative bands at 222 nm and 208 nm and a positive band at 193 nm (Holzwarth & Doty, 1965). Proteins with well-defined antiparallel β-pleated sheets have negative bands at 218 nm and positive bands at 195 nm (Greenfield & Fasman, 1969). In contrast, disordered proteins have very low ellipticity above 210 nm and negative bands near 195 nm (Venyaminov *et al.*, 1993).

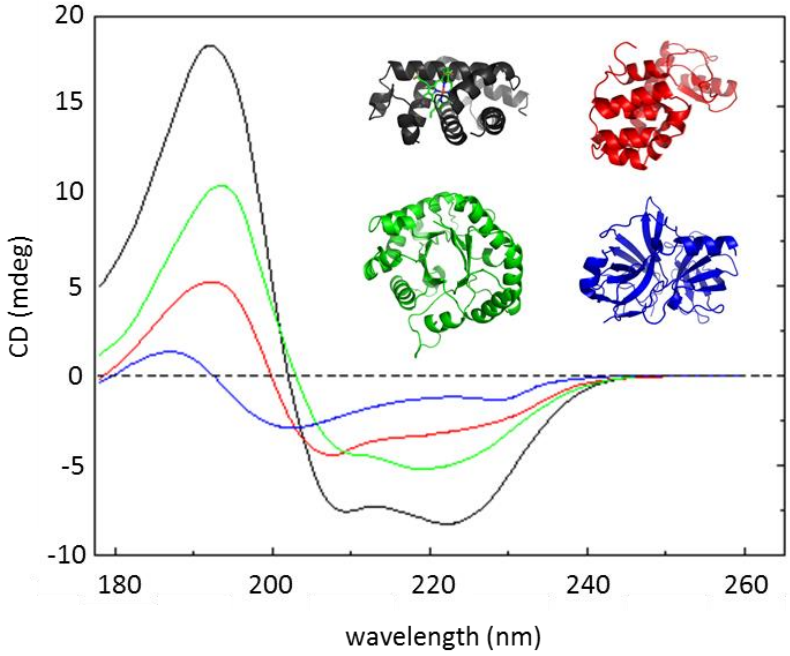


Figure 8.5 CD responses for proteins of increasing β-strand content. Myoglobin is an all-α protein (black, (Vojtechovsky *et al.*, 1999)), lysozyme has two β-strands (α+β, red, (Shoichet *et al.*, 1995)), triosephasphate isomerase (TIM) displays the classic TIM barrel fold (α/β, green, (Alvarez *et al.*, 1998)), and chymotrypsin is all-β (blue, (Kashima *et al.*, 1998)).

For this study, CD spectra were measured in the far UV (300-180 nm) on both Jasco J-810 and J-1500 spectropolarimeters with a 1 mm path length cuvette. Proteins were prepared in either phosphate (50 mM, pH 7.5) or borate (50 mM, pH 9.0) buffers, depending on protein pI value, with 50 mM NaF, and then diluted to 0.05 and 0.1 mg/ml. Eight accumulations were measured using data pitch 0.1 nm, scanning speed 100 nm/min, response 1 s, and a bandwidth of 1 nm. Final spectra were attained by subtracting response for a buffer blank.

Thermal melts were carried out on the Jasco J-1500 instrument from 20-95 ˚C with a 1 ˚C/min gradient (0.2 ˚C interval). At selected temperatures, full spectra (250-180 nm) were measured using 0.2 nm data pitch, 50 nm/min scanning speed, and a bandwidth of 2 nm over 4 sec.

8.2.3.5. Tryptophan fluorescence

Proteins display intrinsic fluorescence predominantly due to tryptophan sidechains, considering the remaining contributions from phenylalanine and tyrosine have very low quantum yield, and are often quenched (Chen & Barkley, 1998). Fluorescence arising only from tryptophan sidechains can be selectively measured by exciting at 295 nm, since at this wavelength there is no absorption by tyrosine (Ghisaidoobe & Chung, 2014). Changes in emission spectra from tryptophan may be seen in response to protein conformational transitions, subunit association, ligand binding, or denaturation, all of which can affect the local environment surrounding the indole ring (Teale & Weber, 1957).

Tryptophan fluorescence was carried out using a Jasco FP-8500 spectrofluorimeter on samples of interest (50 µM) at 25 and 75 °C to record emission spectra of folded and unfolded states. Excitation was carried out at 280, 285, and 290 nm with the following parameters: 5 nm bandwidth, 0.2 s response, low sensitivity, 0.5 nm data interval, and the subsequent emission levels were recorded from 270-600 nm with 100 nm/min scan speed. The apices of the resulting curves were then examined for shifts along the emission spectrum.

8.2.3.6. Sequence analysis tools

The following is a list of sequence analysis tools used throughout this work and their URLs. Standard settings were used for all analysis tools.

- BLAST https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
- Expresso http://tcoffee.crg.cat/apps/tcoffee/do:expresso
- OligoCalc http://biotools.nubic.northwestern.edu/OligoCalc.html
- Serial Cloner http://serialbasics.free.fr/Serial_Cloner.html
- Ugene http://ugene.net/

References

Abraham, A.L., J. Pothier & E.P. Rocha, (2009) Alternative to homo-oligomerisation: the creation of local symmetry in proteins by internal amplification. *Journal of molecular biology* **394**: 522-534.

Adam, M., B. Murali, N.O. Glenn & S.S. Potter, (2008) Epigenetic inheritance based evolution of antibiotic resistance in bacteria. *BMC evolutionary biology* **8**: 52.

Aires, J.R. & H. Nikaido, (2005) Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of *Escherichia coli*. *Journal of bacteriology* **187**: 1923-1929.

Alexandrov, A.I., M. Mileni, E.Y. Chien, M.A. Hanson & R.C. Stevens, (2008) Microscale fluorescent thermal stability assay for membrane proteins. *Structure* **16**: 351-359.

Alvarez, M., J.P. Zeelen, V. Mainfroid, F. Rentier-Delrue, J.A. Martial, L. Wyns, R.K. Wierenga & D. Maes, (1998) Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus.* Kinetic and structural properties. *The Journal of biological chemistry* **273**: 2199-2206.

Amaral, L., S. Fanning & J.M. Pages, (2011) Efflux pumps of gram-negative bacteria: genetic responses to stress and the modulation of their activity by pH, inhibitors, and phenothiazines. *Adv Enzymol Relat Areas Mol Biol* **77**: 61-108.

Amaral, L., A. Martins, G. Spengler & J. Molnar, (2014) Efflux pumps of Gram-negative bacteria: what they do, how they do it, with what and how to deal with them. *Front Pharmacol* **4**: 168.

Anders, R.F., D.M. Hogg & G.R. Jago, (1970) Formation of hydrogen peroxide by group *N streptococci* and its effect on their growth and metabolism. *Appl Microbiol* **19**: 608-612.

Andersen, K.R., N.C. Leksa & T.U. Schwartz, (2013) Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **81**: 1857-1861.

Andreani, J. & R. Guerois, (2014) Evolution of protein interactions: from interactomes to interfaces. *Arch Biochem Biophys* **554**: 65-75.

Andrushchenko, V.V., H.J. Vogel & E.J. Prenner, (2006) Solvent-dependent structure of two tryptophan-rich antimicrobial peptides and their analogs studied by FTIR and CD spectroscopy. *Biochimica et biophysica acta* **1758**: 1596-1608.

Antunes, L.C., P. Visca & K.J. Towner, (2014) *Acinetobacter baumannii*: evolution of a global pathogen. *Pathog Dis* **71**: 292-301.

Arakawa, T. & S.N. Timasheff, (1984) Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding. *Biochemistry* **23**: 5912-5923.

Aravind, L., V. Anantharaman, S. Balaji, M.M. Babu & L.M. Iyer, (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS microbiology reviews* **29**: 231-262.

Armougom, F., S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas & C. Notredame, (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic acids research* **34**: W604-608.

Atkins, P.W. & J. De Paula, (2006) *Atkins' Physical chemistry,* p. xxx, 1064 p. Oxford University Press, Oxford ; New York.

Baker, C.R., V. Hanson-Smith & A.D. Johnson, (2013) Following gene duplication, paralog interference constrains transcriptional circuit evolution. *Science* **342**: 104-108.

Barclay, M.L., E.J. Begg & S.T. Chambers, (1992) Adaptive resistance following single doses of gentamicin in a dynamic in vitro model. *Antimicrobial agents and chemotherapy* **36**: 1951-1957.

Barrientos, L.G., J.M. Louis, I. Botos, T. Mori, Z. Han, B.R. O'Keefe, M.R. Boyd, A. Wlodawer & A.M. Gronenborn, (2002) The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures. *Structure* **10**: 673-686.

Bateman, A., P. Coggill & R.D. Finn, (2010) DUFs: families in search of function. *Acta crystallographica. Section F, Structural biology and crystallization communications* **66**: 1148-1152.

Baugh, S., A.S. Ekanayaka, L.J. Piddock & M.A. Webber, (2012) Loss of or inhibition of all multidrug resistance efflux pumps of *Salmonella enterica* serovar *Typhimurium* results in impaired ability to form a biofilm. *The Journal of antimicrobial chemotherapy* **67**: 2409-2417.

Bay, D.C., K.L. Rommens & R.J. Turner, (2008) Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochimica et biophysica acta* **1778**: 1814-1838.

Bellamy, S.R., Y.S. Kovacheva, I.H. Zulkipli & S.E. Halford, (2009) Differences between Ca2+ and Mg2+ in DNA binding and release by the SfiI restriction endonuclease: implications for DNA looping. *Nucleic acids research* **37**: 5443-5453.

Bennett, P.M., (1999) Integrons and gene cassettes: a genetic construction kit for bacteria. *The Journal of antimicrobial chemotherapy* **43**: 1-4.

Bergkessel, M. & C. Guthrie, (2013) Colony PCR. *Methods in enzymology* **529**: 299-309.

Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov & P.E. Bourne, (2000) The Protein Data Bank. *Nucleic acids research* **28**: 235-242.

Bina, X.R., C.L. Lavine, M.A. Miller & J.E. Bina, (2008) The AcrAB RND efflux system from the live vaccine strain of *Francisella tularensis* is a multiple drug efflux system that is required for virulence in mice. *FEMS microbiology letters* **279**: 226-233.

Binkowski, T.A., W. Jiang, B. Roux, W.F. Anderson & A. Joachimiak, (2014) Virtual high-throughput ligand screening. *Methods in molecular biology* **1140**: 251-261.

Biolabs, N.E., (2017a) BL21 Competent *E. coli*. In: Product Information. pp.

Biolabs, N.E., (2017b) NEB® 5-alpha Competent *E. coli* (High Efficiency). In: Product information. pp.

BioRad, (2017) Buffer Systems and Gel Chemistries. In., pp.

BioTool, (2017) Quick-Fusion Cloning Kit.

Blair, J.M., G.E. Richmond & L.J. Piddock, (2014) Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future microbiology* **9**: 1165-1177.

Blanco, P., S. Hernando-Amado, J.A. Reales-Calderon, F. Corona, F. Lira, M. Alcalde-Rico, A. Bernardini, M.B. Sanchez & J.L. Martinez, (2016) Bacterial Multidrug Efflux Pumps: Much More Than Antibiotic Resistance Determinants. *Microorganisms* **4**.

Blundell, T.L. & N. Srinivasan, (1996) Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 14243-14248.

Borges-Walmsley, M.I., K.S. McKeegan & A.R. Walmsley, (2003) Structure and function of efflux pumps that confer resistance to drugs. *The Biochemical journal* **376**: 313-338.

Bork, P., C. Sander & A. Valencia, (1993) Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein science : a publication of the Protein Society* **2**: 31-40.

Boucher, Y., C.L. Nesbo, M.J. Joss, A. Robinson, B.C. Mabbutt, M.R. Gillings, W.F. Doolittle & H.W. Stokes, (2006) Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC evolutionary biology* **6**: 3.

Brenner, S.E., (2001) A tour of structural genomics. *Nature reviews. Genetics* **2**: 801-809.

Brunelle, J.L. & R. Green, (2014) One-dimensional SDS-polyacrylamide gel electrophoresis (1D SDS-PAGE). *Methods in enzymology* **541**: 151-159.

Brzoska, A.J., K.A. Hassan, E.J. de Leon, I.T. Paulsen & P.J. Lewis, (2013) Single-step selection of drug resistant *Acinetobacter baylyi* ADP1 mutants reveals a functional redundancy in the recruitment of multidrug efflux systems. *PloS one* **8**: e56090.

Buckley, A.M., M.A. Webber, S. Cooles, L.P. Randall, R.M. La Ragione, M.J. Woodward & L.J. Piddock, (2006) The AcrAB-TolC efflux system of *Salmonella enterica* serovar *Typhimurium* plays a role in pathogenesis. *Cellular microbiology* **8**: 847-856.

Bundy, B.M., L.S. Collier, T.R. Hoover & E.L. Neidle, (2002) Synergistic transcriptional activation by one regulatory protein in response to two metabolites. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 7693-7698.

Burrell, M., C.C. Hanfrey, L.N. Kinch, K.A. Elliott & A.J. Michael, (2012) Evolution of a novel lysine decarboxylase in siderophore biosynthesis. *Molecular microbiology* **86**: 485-499.

Burrell, M., C.C. Hanfrey, E.J. Murray, N.R. Stanley-Wall & A.J. Michael, (2010) Evolution and multiplicity of arginine decarboxylases in polyamine biosynthesis and essential role in *Bacillus subtilis* biofilm formation. *The Journal of biological chemistry* **285**: 39224-39238.

Burse, A., H. Weingart & M.S. Ullrich, (2004) The phytoalexin-inducible multidrug efflux pump AcrAB contributes to virulence in the fire blight pathogen, *Erwinia amylovora*. *Molecular plant-microbe interactions : MPMI* **17**: 43-54.

Caffrey, D.R., S. Somaroo, J.D. Hughes, J. Mintseris & E.S. Huang, (2004) Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein science : a publication of the Protein Society* **13**: 190-202.

Canellakis, E.S., D. Viceps-Madore, D.A. Kyriakidis & J.S. Heller, (1979) The regulation and function of ornithine decarboxylase and of the polyamines. *Curr Top Cell Regul* **15**: 155-202.

Chan, Y.Y. & K.L. Chua, (2005) The *Burkholderia pseudomallei* BpeAB-OprB efflux pump: expression and impact on quorum sensing and virulence. *Journal of bacteriology* **187**: 4707-4719.

Chan, Y.Y. & K.L. Chua, (2010) Growth-related changes in intracellular spermidine and its effect on efflux pump expression and quorum sensing in *Burkholderia pseudomallei*. *Microbiology* **156**: 1144-1154.

Chen, C.Y., K.M. Wu, Y.C. Chang, C.H. Chang, H.C. Tsai, T.L. Liao, Y.M. Liu, H.J. Chen, A.B. Shen, J.C. Li, T.L. Su, C.P. Shao, C.T. Lee, L.I. Hor & S.F. Tsai, (2003) Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. *Genome research* **13**: 2577-2587.

Chen, Y. & M.D. Barkley, (1998) Toward understanding tryptophan fluorescence in proteins. *Biochemistry* **37**: 9976-9982.

Cherezov, V., N. Hofer, D.M. Szebenyi, O. Kolaj, J.G. Wall, R. Gillilan, V. Srinivasan, C.P. Jaroniec & M. Caffrey, (2008) Insights into the mode of action of a putative zinc transporter CzrB in *Thermus thermophilus*. *Structure* **16**: 1378-1388.

Cho, S.S., Y. Levy, J.N. Onuchic & P.G. Wolynes, (2005) Overcoming residual frustration in domain-swapping: the roles of disulfide bonds in dimerization and aggregation. *Phys Biol* **2**: S44-55.

Choi, H., S. Kim, P. Mukhopadhyay, S. Cho, J. Woo, G. Storz & S.E. Ryu, (2001) Structural basis of the redox switch in the OxyR transcription factor. *Cell* **105**: 103-113.

Clontech, (2017) Chemically Competent Cells—High Efficiency Competent Cells for Cloning. In: Stellar™ Competent Cells Protocol-At-A-Glance. pp.

Collier, L.S., G.L. Gaines, 3rd & E.L. Neidle, (1998) Regulation of benzoate degradation in *Acinetobacter* sp. strain ADP1 by BenM, a LysR-type transcriptional activator. *Journal of bacteriology* **180**: 2493-2501.

Collins, K.D. & M.W. Washabaugh, (1985) The Hofmeister effect and the behaviour of water at interfaces. *Q Rev Biophys* **18**: 323-422.

Cox, G. & G.D. Wright, (2013) Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *International journal of medical microbiology : IJMM* **303**: 287-292.

Coyne, S., P. Courvalin & B. Perichon, (2011) Efflux-mediated antibiotic resistance in *Acinetobacter spp*. *Antimicrobial agents and chemotherapy* **55**: 947-953.

Coyne, S., G. Guigon, P. Courvalin & B. Perichon, (2010a) Screening and quantification of the expression of antibiotic resistance genes in *Acinetobacter baumannii* with a microarray. *Antimicrobial agents and chemotherapy* **54**: 333-340.

Coyne, S., N. Rosenfeld, T. Lambert, P. Courvalin & B. Perichon, (2010b) Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in *Acinetobacter baumannii*. *Antimicrobial agents and chemotherapy* **54**: 4389-4393.

Craven, S.H., O.C. Ezezika, S. Haddad, R.A. Hall, C. Momany & E.L. Neidle, (2009) Inducer responses of BenM, a LysR-type transcriptional regulator from *Acinetobacter baylyi* ADP1. *Molecular microbiology* **72**: 881-894.

Cuthbertson, L. & J.R. Nodwell, (2013) The TetR family of regulators. *Microbiology and molecular biology reviews : MMBR* **77**: 440-475.

Damier-Piolle, L., S. Magnet, S. Bremont, T. Lambert & P. Courvalin, (2008) AdeIJK, a resistance-nodulation-cell division pump effluxing multiple antibiotics in *Acinetobacter baumannii*. *Antimicrobial agents and chemotherapy* **52**: 557-562.

Dassa, E. & P. Bouige, (2001) The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms. *Res Microbiol* **152**: 211-229.

Davies, J. & D. Davies, (2010) Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR* **74**: 417-433.

Day, Y.S., C.L. Baird, R.L. Rich & D.G. Myszka, (2002) Direct comparison of binding equilibrium, thermodynamic, and rate constants determined by surface- and solution-based biophysical methods. *Protein science : a publication of the Protein Society* **11**: 1017-1025.

Deshpande, C., (2010) Solving protein structures from the mobile genome of *Vibrio*. In: Department of Chemistry and Biomolecular Sciences. Sydney, Australia: Macquarie University, pp.

Di Martino, M.L., R. Campilongo, M. Casalino, G. Micheli, B. Colonna & G. Prosseda, (2013) Polyamines: emerging players in bacteria-host interactions. *International journal of medical microbiology : IJMM* **303**: 484-491.

DiGuiseppi, J. & I. Fridovich, (1982) Oxygen toxicity in *Streptococcus sanguis*. The relative importance of superoxide and hydroxyl radicals. *The Journal of biological chemistry* **257**: 4046-4051.

Diss, G., I. Gagnon-Arsenault, A.M. Dion-Cote, H. Vignaud, D.I. Ascencio, C.M. Berger & C.R. Landry, (2017) Gene duplication can impart fragility, not robustness, in the yeast protein interaction network. *Science* **355**: 630-634.

Dokmanic, I., M. Sikic & S. Tomic, (2008) Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. *Acta crystallographica. Section D, Biological crystallography* **64**: 257-263.

Du, D., Z. Wang, N.R. James, J.E. Voss, E. Klimont, T. Ohene-Agyei, H. Venter, W. Chiu & B.F. Luisi, (2014) Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**: 512-515.

Elcock, A.H. & J.A. McCammon, (2001) Identification of protein oligomerization states by analysis of interface conservation. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 2990-2994.

Elsaied, H., H.W. Stokes, K. Kitamura, Y. Kurusu, Y. Kamagata & A. Maruyama, (2011) Marine integrons containing novel integrase genes, attachment sites, attI, and associated gene cassettes in polluted sediments from Suez and Tokyo Bays. *ISME J* **5**: 1162-1177.

Elsaied, H., H.W. Stokes, T. Nakamura, K. Kitamura, H. Fuse & A. Maruyama, (2007) Novel and diverse integron integrase genes and integron-like gene cassettes are prevalent in deep-sea hydrothermal vents. *Environmental microbiology* **9**: 2298-2312.

Elsaied, H., H.W. Stokes, H. Yoshioka, Y. Mitani & A. Maruyama, (2014) Novel integrons and gene cassettes from a Cascadian submarine gas-hydrate-bearing core. *FEMS Microbiol Ecol* **87**: 343-356.

Escudero, J.A., C. Loot, A. Nivina & D. Mazel, (2015) The Integron: Adaptation On Demand. *Microbiol Spectr* **3**: MDNA3-0019-2014.

Eswaran, J., E. Koronakis, M.K. Higgins, C. Hughes & V. Koronakis, (2004) Three's company: component structures bring a closer view of tripartite drug efflux pumps. *Curr Opin Struct Biol* **14**: 741-747.

Ezezika, O.C., S. Haddad, T.J. Clark, E.L. Neidle & C. Momany, (2007) Distinct effector-binding sites enable synergistic transcriptional activation by BenM, a LysR-type regulator. *Journal of molecular biology* **367**: 616-629.

Farokhzad, O.C. & R. Langer, (2009) Impact of nanotechnology on drug delivery. *ACS Nano* **3**: 16-20.

Farrugia, D.N., L.D. Elbourne, K.A. Hassan, B.A. Eijkelkamp, S.G. Tetu, M.H. Brown, B.S. Shah, A.Y. Peleg, B.C. Mabbutt & I.T. Paulsen, (2013) The complete genome and phenome of a community-acquired *Acinetobacter baumannii*. *PloS one* **8**: e58628.

Fasman, G.D., (1996) *Circular dichroism and the conformational analysis of biomolecules,* p. ix, 738 p. Plenum Press, New York.

Fernandez, L. & R.E. Hancock, (2012) Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance. *Clinical microbiology reviews* **25**: 661-681.

Fersht, A.R., (1993) The sixth Datta Lecture. Protein folding and stability: the pathway of folding of barnase. *FEBS letters* **325**: 5-16.

Finnigan, G.C., V. Hanson-Smith, T.H. Stevens & J.W. Thornton, (2012) Evolution of increased complexity in a molecular machine. *Nature* **481**: 360-364.

Fondi, M. & R. Fani, (2010) The horizontal flow of the plasmid resistome: clues from inter-generic similarity networks. *Environmental microbiology* **12**: 3228-3242.

Foster, D.L., M. Boublik & H.R. Kaback, (1983) Structure of the lac carrier protein of *Escherichia coli*. *The Journal of biological chemistry* **258**: 31-34.

Fraser-Liggett, C.M., (2005) Insights on biology and evolution from microbial genome sequencing. *Genome research* **15**: 1603-1610.

Frost, L.S., R. Leplae, A.O. Summers & A. Toussaint, (2005) Mobile genetic elements: the agents of open source evolution. *Nature reviews. Microbiology* **3**: 722-732.

Fujihashi, M., T. Nakatani, K. Hirooka, H. Matsuoka, Y. Fujita & K. Miki, (2014) Structural characterization of a ligand-bound form of *Bacillus subtilis* FadR involved in the regulation of fatty acid degradation. *Proteins* **82**: 1301-1310.

Gabaldon, T. & M.A. Huynen, (2004) Prediction of protein function and pathways in the genome era. *Cellular and molecular life sciences : CMLS* **61**: 930-944.

George, A.M. & S.B. Levy, (1983) Amplifiable resistance to tetracycline, chloramphenicol, and other antibiotics in *Escherichia coli*: involvement of a non-plasmid-determined efflux of tetracycline. *Journal of bacteriology* **155**: 531-540.

Ghai, R., R.J. Falconer & B.M. Collins, (2012) Applications of isothermal titration calorimetry in pure and applied research--survey of the literature from 2010. *Journal of molecular recognition : JMR* **25**: 32-52.

Ghisaidoobe, A.B. & S.J. Chung, (2014) Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Forster resonance energy transfer techniques. *Int J Mol Sci* **15**: 22518-22538.

Giles, G.I., M.J. Nasim, W. Ali & C. Jacob, (2017) The Reactive Sulfur Species Concept: 15 Years On. *Antioxidants (Basel)* **6**.

Glatter, O. & O. Kratky, (1982) *Small angle x-ray scattering,* p. 515 p. Academic Press, London ; New York.

Goodsell, D.S. & A.J. Olson, (2000) Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure* **29**: 105-153.

Greenfield, N. & G.D. Fasman, (1969) Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**: 4108-4116.

Greenfield, N.J., (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **1**: 2876-2890.

Gristwood, T., P.C. Fineran, L. Everson & G.P. Salmond, (2008) PigZ, a TetR/AcrR family repressor, modulates secondary metabolism via the expression of a putative four-component resistance-nodulation-cell-division efflux pump, ZrpADBC, in Serratia sp. ATCC 39006. *Molecular microbiology* **69**: 418-435.

Gu, R., C.C. Su, F. Shi, M. Li, G. McDermott, Q. Zhang & E.W. Yu, (2007) Crystal structure of the transcriptional regulator CmeR from *Campylobacter jejuni*. *Journal of molecular biology* **372**: 583-593.

Guan, K.L. & J.E. Dixon, (1990) Protein tyrosine phosphatase activity of an essential virulence determinant in Yersinia. *Science* **249**: 553-556.

Gunther, K., A. Paradowska-Dogan, B. Barmann, H. Klein, C. von Eichel-Streiber, R. Hartley, W. Weidner, R. Behr & K. Steger, (2015) Expression of sperm-specific protamines impairs bacterial and eukaryotic cell proliferation. *Histochem Cell Biol* **143**: 599-609.

Hall, R.M. & H.W. Stokes, (1993) Integrons: novel DNA elements which capture genes by site-specific recombination. *Genetica* **90**: 115-132.

Hamada, S. & H.D. Slade, (1980) Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiological reviews* **44**: 331-384.

Hamana, K. & S. Matsuzaki, (1992) Polyamines as a chemotaxonomic marker in bacterial systematics. *Critical reviews in microbiology* **18**: 261-283.

Hamburger, Z.A., M.S. Brown, R.R. Isberg & P.J. Bjorkman, (1999) Crystal structure of invasin: a bacterial integrin-binding protein. *Science* **286**: 291-295.

Hardt, W.D., L.M. Chen, K.E. Schuebel, X.R. Bustelo & J.E. Galan, (1998) *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**: 815-826.

Hassan, K.A., A.K. Cain, T. Huang, Q. Liu, L.D. Elbourne, C.J. Boinett, A.J. Brzoska, L. Li, M. Ostrowski, N.T. Nhu, H. Nhu Tdo, S. Baker, J. Parkhill & I.T. Paulsen, (2016) Fluorescence-Based Flow Sorting in Parallel with Transposon Insertion Site Sequencing Identifies Multidrug Efflux Systems in *Acinetobacter baumannii*. *mBio* **7**.

Hassan, K.A., S.M. Jackson, A. Penesyan, S.G. Patching, S.G. Tetu, B.A. Eijkelkamp, M.H. Brown, P.J. Henderson & I.T. Paulsen, (2013) Transcriptomic and biochemical analyses identify a family of chlorhexidine efflux proteins. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 20254-20259.

Hassan, K.A., Q. Liu, L.D.H. Elbourne, I. Ahmad, D. Sharples, V. Naidu, C.L. Chan, L. Li, S.P.D. Harborne, A. Pokhrel, V.L.G. Postis, A. Goldman, P.J.F. Henderson & I.T. Paulsen, (2018) Pacing across the membrane: the novel PACE family of efflux pumps is widespread in Gram-negative pathogens. *Res Microbiol*.

Hassan, K.A., Q. Liu, P.J. Henderson & I.T. Paulsen, (2015) Homologs of the *Acinetobacter baumannii* AceI transporter represent a new family of bacterial multidrug efflux systems. *mBio* **6**.

Heck, A.J. & R.H. Van Den Heuvel, (2004) Investigation of intact protein complexes by mass spectrometry. *Mass Spectrom Rev* **23**: 368-389.

Hernandez-Lucas, I., A.L. Gallego-Hernandez, S. Encarnacion, M. Fernandez-Mora, A.G. Martinez-Batallar, H. Salgado, R. Oropeza & E. Calva, (2008) The LysR-type transcriptional regulator LeuO controls expression of several genes in *Salmonella enterica* serovar *Typhi*. *Journal of bacteriology* **190**: 1658-1670.

Heroven, A.K. & P. Dersch, (2006) RovM, a novel LysR-type regulator of the virulence activator gene rovA, controls cell invasion, virulence and motility of *Yersinia pseudotuberculosis*. *Molecular microbiology* **62**: 1469-1483.

Higgins, C.F., (2001) ABC transporters: physiology, structure and mechanism--an overview. *Res Microbiol* **152**: 205-210.

Hirakata, Y., R. Srikumar, K. Poole, N. Gotoh, T. Suematsu, S. Kohno, S. Kamihira, R.E. Hancock & D.P. Speert, (2002) Multidrug efflux systems play an important role in the invasiveness of *Pseudomonas aeruginosa*. *J Exp Med* **196**: 109-118.

Hogg, S., (2005) *Essential microbiology,* p. viii, 468 p. John Wiley and Sons, West Sussex ; Hoboken, NJ.

Holmes, A.J., M.R. Gillings, B.S. Nield, B.C. Mabbutt, K.M. Nevalainen & H.W. Stokes, (2003) The gene cassette metagenome is a basic resource for bacterial genome evolution. *Environmental microbiology* **5**: 383-394.

Holzwarth, G. & P. Doty, (1965) The Ultraviolet Circular Dichroism of Polypeptides. *Journal of the American Chemical Society* **87**: 218-228.

Homola, J., (2008) Surface plasmon resonance sensors for detection of chemical and biological species. *Chemical reviews* **108**: 462-493.

Hooper, S.D., K. Mavromatis & N.C. Kyrpides, (2009) Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome biology* **10**: R45.

Igarashi, K. & K. Kashiwagi, (2006) Polyamine Modulon in *Escherichia coli*: genes involved in the stimulation of cell growth by polyamines. *J Biochem* **139**: 11-16.

Jain, R., M.C. Rivera, J.E. Moore & J.A. Lake, (2002) Horizontal gene transfer in microbial genome evolution. *Theor Popul Biol* **61**: 489-495.

Jerse, A.E., N.D. Sharma, A.N. Simms, E.T. Crow, L.A. Snyder & W.M. Shafer, (2003) A gonococcal efflux pump system enhances bacterial survival in a female mouse model of genital tract infection. *Infection and immunity* **71**: 5576-5582.

Jo, I., I.Y. Chung, H.W. Bae, J.S. Kim, S. Song, Y.H. Cho & N.C. Ha, (2015) Structural details of the OxyR peroxide-sensing mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 6443-6448.

Join-Lambert, O.F., M. Michea-Hamzehpour, T. Kohler, F. Chau, F. Faurisson, S. Dautrey, C. Vissuzaine, C. Carbon & J. Pechere, (2001) Differential selection of multidrug efflux mutants by trovafloxacin and ciprofloxacin in an experimental model of *Pseudomonas aeruginosa* acute pneumonia in rats. *Antimicrobial agents and chemotherapy* **45**: 571-576.

Jones, P.M. & A.M. George, (2004) The ABC transporter structure and mechanism: perspectives on recent research. *Cellular and molecular life sciences : CMLS* **61**: 682-699.

Juhas, M., (2015) Horizontal gene transfer in human pathogens. *Crit Rev Microbiol* **41**: 101-108.

Kaltenegger, E. & D. Ober, (2015) Paralogue Interference Affects the Dynamics after Gene Duplication. *Trends Plant Sci* **20**: 814-821.

Kanjee, U., I. Gutsche, E. Alexopoulos, B. Zhao, M. El Bakkouri, G. Thibault, K. Liu, S. Ramachandran, J. Snider, E.F. Pai & W.A. Houry, (2011a) Linkage between the

bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *The EMBO journal* **30**: 931-944.

Kanjee, U., I. Gutsche, S. Ramachandran & W.A. Houry, (2011b) The enzymatic activities of the *Escherichia coli* basic aliphatic amino acid decarboxylases exhibit a pH zone of inhibition. *Biochemistry* **50**: 9388-9398.

Karplus, P.A., (2015) A primer on peroxiredoxin biochemistry. *Free Radic Biol Med* **80**: 183-190.

Kashima, A., Y. Inoue, S. Sugio, I. Maeda, T. Nose & Y. Shimohigashi, (1998) X-ray crystal structure of a dipeptide-chymotrypsin complex in an inhibitory interaction. *European journal of biochemistry* **255**: 12-23.

Kibbe, W.A., (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic acids research* **35**: W43-46.

Kim, Y., G. Babnigg, R. Jedrzejczak, W.H. Eschenfeldt, H. Li, N. Maltseva, C. Hatzos-Skintges, M. Gu, M. Makowska-Grzyska, R. Wu, H. An, G. Chhor & A. Joachimiak, (2011) High-throughput protein purification and quality assessment for crystallization. *Methods* **55**: 12-28.

Kim, Y.H., S.M. Kim & S.Y. Lee, (2015) Antimicrobial Activity of Protamine against Oral Microorganisms. *Biocontrol Sci* **20**: 275-280.

King, N.P., J.B. Bale, W. Sheffler, D.E. McNamara, S. Gonen, T. Gonen, T.O. Yeates & D. Baker, (2014) Accurate design of co-assembling multi-component protein nanomaterials. *Nature* **510**: 103-108.

Koppensteiner, W.A., P. Lackner, M. Wiederstein & M.J. Sippl, (2000) Characterization of novel proteins based on known protein structures. *Journal of molecular biology* **296**: 1139-1152.

Koronakis, V., J. Eswaran & C. Hughes, (2004) Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu Rev Biochem* **73**: 467-489.

Krissinel, E. & K. Henrick, (2007) Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**: 774-797.

Kumarasamy, K.K., M.A. Toleman, T.R. Walsh, J. Bagaria, F. Butt, R. Balakrishnan, U. Chaudhary, M. Doumith, C.G. Giske, S. Irfan, P. Krishnan, A.V. Kumar, S. Maharjan, S. Mushtaq, T. Noorie, D.L. Paterson, A. Pearson, C. Perry, R. Pike, B. Rao, U. Ray, J.B. Sarma, M. Sharma, E. Sheridan, M.A. Thirunarayan, J. Turton, S. Upadhyay, M. Warner, W. Welfare, D.M. Livermore & N. Woodford, (2010) Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet. Infectious diseases* **10**: 597-602.

Kvist, M., V. Hancock & P. Klemm, (2008) Inactivation of efflux pumps abolishes bacterial biofilm formation. *Applied and environmental microbiology* **74**: 7376-7382.

Labbate, M., R.J. Case & H.W. Stokes, (2009) The integron/gene cassette system: an active player in bacterial adaptation. *Methods in molecular biology* **532**: 103-125.

Lacroix, F.J., A. Cloeckaert, O. Grepinet, C. Pinault, M.Y. Popoff, H. Waxin & P. Pardon, (1996) *Salmonella typhimurium* acrB-like gene: identification and role in resistance to biliary salts and detergents and in murine infection. *FEMS microbiology letters* **135**: 161-167.

Ladbury, J.E., (2010) Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochemical Society transactions* **38**: 888-893.

Lang, P.T., S.R. Brozell, S. Mukherjee, E.F. Pettersen, E.C. Meng, V. Thomas, R.C. Rizzo, D.A. Case, T.L. James & I.D. Kuntz, (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *Rna* **15**: 1219-1230.

Larson, S.B., J.S. Day, R. Cudney & A. McPherson, (2007a) A new crystal form of bovine pancreatic RNase A in complex with 2'-deoxyguanosine-5'-monophosphate. *Acta crystallographica. Section F, Structural biology and crystallization communications* **63**: 728-733.

Larson, S.B., J.S. Day, R. Cudney & A. McPherson, (2007b) A novel strategy for the crystallization of proteins: X-ray diffraction validation. *Acta crystallographica. Section D, Biological crystallography* **63**: 310-318.

Lavinder, J.J., S.B. Hari, B.J. Sullivan & T.J. Magliery, (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J Am Chem Soc* **131**: 3794-3795.

Lea, W.A. & A. Simeonov, (2012) Differential scanning fluorometry signatures as indicators of enzyme inhibitor mode of action: case study of glutathione S-transferase. *PloS one* **7**: e36219.

Lee, J., V. Sperandio, D.E. Frantz, J. Longgood, A. Camilli, M.A. Phillips & A.J. Michael, (2009) An alternative polyamine biosynthetic pathway is widespread in bacteria and essential for biofilm formation in *Vibrio cholerae*. *The Journal of biological chemistry* **284**: 9899-9907.

Levy, E.D., E. Boeri Erba, C.V. Robinson & S.A. Teichmann, (2008) Assembly reflects evolution of protein complexes. *Nature* **453**: 1262-1265.

Levy, E.D., J.B. Pereira-Leal, C. Chothia & S.A. Teichmann, (2006) 3D complex: a structural classification of protein complexes. *PLoS computational biology* **2**: e155.

Lewis, M., (2011) Agarose gel electrophoresis (basic method). *Biological Protocols*.

Li, X.Z. & H. Nikaido, (2009) Efflux-mediated drug resistance in bacteria: an update. *Drugs* **69**: 1555-1623.

Liu, H., C.L. Yang, M.Y. Ge, M. Ibrahim, B. Li, W.J. Zhao, G.Y. Chen, B. Zhu & G.L. Xie, (2014) Regulatory role of tetR gene in a novel gene cluster of Acidovorax avenae subsp. avenae RS-1 under oxidative stress. *Frontiers in microbiology* **5**: 547.

Liu, Y. & D. Eisenberg, (2002) 3D domain swapping: as domains continue to swap. *Protein science : a publication of the Protein Society* **11**: 1285-1299.

Lorca, G.L., R.D. Barabote, V. Zlotopolski, C. Tran, B. Winnen, R.N. Hvorup, A.J. Stonestrom, E. Nguyen, L.W. Huang, D.S. Kim & M.H. Saier, Jr., (2007) Transport capabilities of

eleven gram-positive bacteria: comparative genomic analyses. *Biochimica et biophysica acta* **1768**: 1342-1366.

Maczulak, A.E., (2011) *Encyclopedia of microbiology,* p. xii, 858 p. Facts On File, New York, NY.

Maddocks, S.E. & P.C. Oyston, (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**: 3609-3623.

Magnet, S., P. Courvalin & T. Lambert, (2001) Resistance-nodulation-cell division-type efflux pump involved in aminoglycoside resistance in *Acinetobacter baumannii* strain BM4454. *Antimicrobial agents and chemotherapy* **45**: 3375-3380.

Maham, A., Z. Tang, H. Wu, J. Wang & Y. Lin, (2009) Protein-based nanomedicine platforms for drug delivery. *Small* **5**: 1706-1721.

Maiden, M.C., E.O. Davis, S.A. Baldwin, D.C. Moore & P.J. Henderson, (1987) Mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**: 641-643.

Marchand, I., L. Damier-Piolle, P. Courvalin & T. Lambert, (2004) Expression of the RND-type efflux pump AdeABC in *Acinetobacter baumannii* is regulated by the AdeRS two-component system. *Antimicrobial agents and chemotherapy* **48**: 3298-3304.

Marger, M.D. & M.H. Saier, Jr., (1993) A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem Sci* **18**: 13-20.

Marshall, N.J. & L.J. Piddock, (1997) Antibacterial efflux systems. *Microbiologia* **13**: 285-300.

Mascarenhas, N.M. & S. Gosavi, (2017) Understanding protein domain-swapping using structure-based models of protein folding. *Prog Biophys Mol Biol* **128**: 113-120.

Matsuoka, H., K. Hirooka & Y. Fujita, (2007) Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *The Journal of biological chemistry* **282**: 5180-5194.

Mazel, D., (2006) Integrons: agents of bacterial evolution. *Nature reviews. Microbiology* **4**: 608-620.

McPherson, A. & B. Cudney, (2006) Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol* **156**: 387-406.

Menzen, T. & W. Friess, (2013) High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. *J Pharm Sci* **102**: 415-428.

Merck, (2017a) BL21(DE3)pLysS Competent Cells - Novagen. In: Overview. pp.

Merck, (2017b) pET-15b DNA - Novagen. In: Supporting Documentation. pp.

Miller, K.H., J.R. Karr & S. Marqusee, (2010) A hinge region cis-proline in ribonuclease A acts as a conformational gatekeeper for C-terminal domain swapping. *Journal of molecular biology* **400**: 567-578.

Mintseris, J. & Z. Weng, (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 10930-10935.

Miranker, A., C.V. Robinson, S.E. Radford, R.T. Aplin & C.M. Dobson, (1993) Detection of transient protein folding populations by mass spectrometry. *Science* **262**: 896-900.

Mirkin, B.G., T.I. Fenner, M.Y. Galperin & E.V. Koonin, (2003) Algorithms for computing parsimonious evolutionary scenarios for genome evolution, the last universal common ancestor and dominance of horizontal gene transfer in the evolution of prokaryotes. *BMC evolutionary biology* **3**: 2.

Mirza, U.A., S.L. Cohen & B.T. Chait, (1993) Heat-induced conformational changes in proteins studied by electrospray ionization mass spectrometry. *Anal Chem* **65**: 1-6.

Moir, R.D. & G.H. Dixon, (1988) Characterization of a protamine gene from the chum salmon (*Oncorhynchus keta*). *Journal of molecular evolution* **27**: 8-16.

Molino, N.M. & S.W. Wang, (2014) Caged protein nanoparticles for drug delivery. *Curr Opin Biotechnol* **28**: 75-82.

Monod, J., J. Wyman & J.P. Changeux, (1965) On the Nature of Allosteric Transitions: A Plausible Model. *Journal of molecular biology* **12**: 88-118.

Morita, Y., K. Kodama, S. Shiota, T. Mine, A. Kataoka, T. Mizushima & T. Tsuchiya, (1998) NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrobial agents and chemotherapy* **42**: 1778-1782.

Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell & A.J. Olson, (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**: 2785-2791.

Muraoka, S., R. Okumura, N. Ogawa, T. Nonaka, K. Miyashita & T. Senda, (2003) Crystal structure of a full-length LysR-type transcriptional regulator, CbnR: unusual combination of two subunit forms and molecular bases for causing and changing DNA bend. *Journal of molecular biology* **328**: 555-566.

Najmanovich, R., N. Kurbatova & J. Thornton, (2008) Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics* **24**: i105-111.

Niesen, F.H., H. Berglund & M. Vedadi, (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature protocols* **2**: 2212-2221.

Nikaido, H., (2001) Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria. *Seminars in cell & developmental biology* **12**: 215-223.

Nikaido, H., (2011) Structure and mechanism of RND-type multidrug efflux pumps. *Adv Enzymol Relat Areas Mol Biol* **77**: 1-60.

Nikaido, H., M. Basina, V. Nguyen & E.Y. Rosenberg, (1998) Multidrug efflux pump AcrAB of *Salmonella typhimurium* excretes only those beta-lactam antibiotics containing lipophilic side chains. *Journal of bacteriology* **180**: 4686-4692.

Norrman, M., F. Hubalek & G. Schluckebier, (2007) Structural characterization of insulin NPH formulations. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences* **30**: 414-423.

Ochman, H., J.G. Lawrence & E.A. Groisman, (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299-304.

Ogawa, T. & N. Hirokawa, (2018) Multiple analyses of protein dynamics in solution. *Biophys Rev* **10**: 299-306.

Omote, H., M. Hiasa, T. Matsumoto, M. Otsuka & Y. Moriyama, (2006) The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends Pharmacol Sci* **27**: 587-593.

Orth, P., F. Cordes, D. Schnappinger, W. Hillen, W. Saenger & W. Hinrichs, (1998) Conformational changes of the Tet repressor induced by tetracycline trapping. *Journal of molecular biology* **279**: 439-447.

Pages, J.M. & L. Amaral, (2009) Mechanisms of drug efflux and strategies to combat them: challenging the efflux pump of Gram-negative bacteria. *Biochimica et biophysica acta* **1794**: 826-833.

Palmer, K.L., V.N. Kos & M.S. Gilmore, (2010) Horizontal gene transfer and the genomics of enterococcal antibiotic resistance. *Current opinion in microbiology* **13**: 632-639.

Pareja, E., P. Pareja-Tobes, M. Manrique, E. Pareja-Tobes, J. Bonal & R. Tobes, (2006) ExtraTrain: a database of Extragenic regions and Transcriptional information in prokaryotic organisms. *BMC Microbiol* **6**: 29.

Partridge, S.R., G. Tsafnat, E. Coiera & J.R. Iredell, (2009) Gene cassettes and cassette arrays in mobile resistance integrons. *FEMS microbiology reviews* **33**: 757-784.

Paulsen, I.T., (2003) Multidrug efflux pumps and resistance: regulation and evolution. *Current opinion in microbiology* **6**: 446-451.

Paulsen, I.T., R.A. Skurray, R. Tam, M.H. Saier, Jr., R.J. Turner, J.H. Weiner, E.B. Goldberg & L.L. Grinius, (1996) The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. *Molecular microbiology* **19**: 1167-1175.

Pereira-Leal, J.B., E.D. Levy, C. Kamp & S.A. Teichmann, (2007) Evolution of protein complexes by duplication of homomeric interactions. *Genome Biol* **8**: R51.

Perozzo, R., G. Folkers & L. Scapozza, (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *Journal of receptor and signal transduction research* **24**: 1-52.

Perrot, P., (1998) *A to Z of thermodynamics,* p. vi, 329 p. Oxford University Press, Oxford ; New York.

Picossi, S., B.R. Belitsky & A.L. Sonenshein, (2007) Molecular mechanism of the regulation of *Bacillus subtilis* gltAB expression by GltC. *Journal of molecular biology* **365**: 1298-1313.

Piddock, L.J., (2006) Multidrug-resistance efflux pumps - not just for resistance. *Nature reviews. Microbiology* **4**: 629-636.

Poelarends, G.J., V.P. Veetil & C.P. Whitman, (2008) The chemical versatility of the beta-alpha-beta fold: catalytic promiscuity and divergent evolution in the tautomerase superfamily. *Cellular and molecular life sciences : CMLS* **65**: 3606-3618.

Ponstingl, H., K. Henrick & J.M. Thornton, (2000) Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins* **41**: 47-57.

Poole, K., (2004) Efflux-mediated multiresistance in Gram-negative bacteria. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* **10**: 12-26.

Porath, J. & P. Flodin, (1959) Gel filtration: a method for desalting and group separation. *Nature* **183**: 1657-1659.

Povolotskaya, I.S. & F.A. Kondrashov, (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* **465**: 922-926.

Prouty, A.M., I.E. Brodsky, S. Falkow & J.S. Gunn, (2004) Bile-salt-mediated induction of antimicrobial and bile resistance in *Salmonella typhimurium*. *Microbiology* **150**: 775-783.

Pustam, A., C. Smith, C. Deering, K.M. Grosicki, T.Y. Leng, S. Lin, J. Yang, D. Pink, T. Gill, L. Graham, D. Derksen, C. Bishop, M.E. Demont, R.C. Wyeth & T. Smith-Palmer, (2014) Interactions of protamine with the marine bacterium, *Pseudoalteromonas* sp. NCIMB 2021. *Lett Appl Microbiol* **58**: 225-230.

Qi Liu, K.A.H., Heather E. Ashwood, Hasinika K.A. Hewawasam Gamage, Liping Li, Bridget C. Mabbutt, Ian T. Paulsen, (2017) Regulation of the *aceI* multidrug efflux pump gene in *Acinetobacter baumannii*. *Journal of Antimicrobial Chemotherapy*.

Rajamohan, G., V.B. Srinivasan & W.A. Gebreyes, (2010a) Molecular and functional characterization of a novel efflux pump, AmvA, mediating antimicrobial and disinfectant resistance in *Acinetobacter baumannii*. *The Journal of antimicrobial chemotherapy* **65**: 1919-1925.

Rajamohan, G., V.B. Srinivasan & W.A. Gebreyes, (2010b) Novel role of *Acinetobacter baumannii* RND efflux transporters in mediating decreased susceptibility to biocides. *The Journal of antimicrobial chemotherapy* **65**: 228-232.

Rambo, R.P., (2017) Considerations for Sample Preparation Using Size-Exclusion Chromatography for Home and Synchrotron Sources. *Adv Exp Med Biol* **1009**: 31-45.

Ramos, J.L., M. Martinez-Bueno, A.J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M.T. Gallegos, R. Brennan & R. Tobes, (2005) The TetR family of transcriptional repressors. *Microbiology and molecular biology reviews : MMBR* **69**: 326-356.

Recchia, G.D. & R.M. Hall, (1995) Gene cassettes: a new class of mobile element. *Microbiology* **141 (Pt 12)**: 3015-3027.

Rees, D.C. & A.D. Robertson, (2001) Some thermodynamic implications for the thermostability of proteins. *Protein science : a publication of the Protein Society* **10**: 1187-1194.

Ren, D., F. Kratz & S.W. Wang, (2011) Protein nanocapsules containing doxorubicin as a pH-responsive delivery system. *Small* **7**: 1051-1060.

Rice, L.B., (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *The Journal of infectious diseases* **197**: 1079-1081.

Richardson, J.M., L.S. Morrison, N.D. Bland, S. Bruce, G.H. Coombs, J.C. Mottram & M.D. Walkinshaw, (2009) Structures of *Leishmania major* orthologues of macrophage migration inhibitory factor. *Biochemical and biophysical research communications* **380**: 442-448.

Richardson, J.S., (1981) The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**: 167-339.

Rigden, D.J., (2006) Understanding the cell in terms of structure and function: insights from structural genomics. *Current opinion in biotechnology* **17**: 457-464.

Robichon, C., J. Luo, T.B. Causey, J.S. Benner & J.C. Samuelson, (2011) Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Applied and environmental microbiology* **77**: 4634-4646.

Robinson, A., A.P. Guilfoyle, S.J. Harrop, Y. Boucher, H.W. Stokes, P.M. Curmi & B.C. Mabbutt, (2007) A putative house-cleaning enzyme encoded within an integron array: 1.8 A crystal structure defines a new MazG subtype. *Molecular microbiology* **66**: 610-621.

Robinson, A., A.P. Guilfoyle, V. Sureshan, M. Howell, S.J. Harrop, Y. Boucher, H.W. Stokes, P.M. Curmi & B.C. Mabbutt, (2008) Structural genomics of the bacterial mobile metagenome: an overview. *Methods in molecular biology* **426**: 589-595.

Robinson, A., P.S. Wu, S.J. Harrop, P.M. Schaeffer, Z. Dosztanyi, M.R. Gillings, A.J. Holmes, K.M. Nevalainen, H.W. Stokes, G. Otting, N.E. Dixon, P.M. Curmi & B.C. Mabbutt, (2005) Integron-associated mobile gene cassettes code for folded proteins: the structure of Bal32a, a new member of the adaptable alpha+beta barrel family. *Journal of molecular biology* **346**: 1229-1241.

Roca, I., S. Marti, P. Espinal, P. Martinez, I. Gibert & J. Vila, (2009) CraA, a major facilitator superfamily efflux pump associated with chloramphenicol resistance in *Acinetobacter baumannii*. *Antimicrobial agents and chemotherapy* **53**: 4013-4014.

Rodionov, D.A., (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chemical reviews* **107**: 3467-3497.

Rodriguez-Minguela, C.M., J.H. Apajalahti, B. Chai, J.R. Cole & J.M. Tiedje, (2009) Worldwide prevalence of class 2 integrases outside the clinical setting is associated with human impact. *Applied and environmental microbiology* **75**: 5100-5110.

Romero-Arroyo, C.E., M.A. Schell, G.L. Gaines, 3rd & E.L. Neidle, (1995) catM encodes a LysR-type transcriptional activator regulating catechol degradation in *Acinetobacter calcoaceticus*. *Journal of bacteriology* **177**: 5891-5898.

Romero-Rodriguez, A., I. Robledo-Casados & S. Sanchez, (2015) An overview on transcriptional regulators in *Streptomyces*. *Biochimica et biophysica acta* **1849**: 1017-1039.

Roque, A., I. Ponte & P. Suau, (2011) Secondary structure of protamine in sperm nuclei: an infrared spectroscopy study. *BMC structural biology* **11**: 14.

Rosenfeld, N., C. Bouchier, P. Courvalin & B. Perichon, (2012) Expression of the resistance-nodulation-cell division pump AdeIJK in *Acinetobacter baumannii* is regulated by AdeN, a TetR-type regulator. *Antimicrobial agents and chemotherapy* **56**: 2504-2510.

Rosvoll, T.C., T. Pedersen, H. Sletvold, P.J. Johnsen, J.E. Sollid, G.S. Simonsen, L.B. Jensen, K.M. Nielsen & A. Sundsfjord, (2010) PCR-based plasmid typing in *Enterococcus faecium* strains reveals widely distributed pRE25-, pRUM-, pIP501- and pHTbeta-related replicons associated with glycopeptide resistance and stabilizing toxin-antitoxin systems. *FEMS immunology and medical microbiology* **58**: 254-268.

Rothschild, L.J. & R.L. Mancinelli, (2001) Life in extreme environments. *Nature* **409**: 1092-1101.

Rousseau, F., J. Schymkowitz & L.S. Itzhaki, (2012) Implications of 3D domain swapping for protein folding, misfolding and function. *Adv Exp Med Biol* **747**: 137-152.

Rousseau, F., J.W. Schymkowitz, H.R. Wilkinson & L.S. Itzhaki, (2001) Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 5596-5601.

Rowe-Magnus, D.A., A.M. Guerout & D. Mazel, (1999) Super-integrons. *Research in microbiology* **150**: 641-651.

Ruangprasert, A., S.H. Craven, E.L. Neidle & C. Momany, (2010) Full-length structures of BenM and two variants reveal different oligomerization schemes for LysR-type transcriptional regulators. *Journal of molecular biology* **404**: 568-586.

Rulisek, L. & J. Vondrasek, (1998) Coordination geometries of selected transition metal ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+) in metalloproteins. *Journal of inorganic biochemistry* **71**: 115-127.

Russell, D.A., G.A. Byrne, E.P. O'Connell, C.A. Boland & W.G. Meijer, (2004) The LysR-type transcriptional regulator VirR is required for expression of the virulence gene vapA of *Rhodococcus equi* ATCC 33701. *Journal of bacteriology* **186**: 5576-5584.

Saha, R.P., S. Samanta, S. Patra, D. Sarkar, A. Saha & M.K. Singh, (2017) Metal homeostasis in bacteria: the role of ArsR-SmtB family of transcriptional repressors in combating varying metal concentrations in the environment. *Biometals* **30**: 459-503.

Saidijam, M., G. Benedetti, Q. Ren, Z. Xu, C.J. Hoyle, S.L. Palmer, A. Ward, K.E. Bettaney, G. Szakonyi, J. Meuller, S. Morrison, M.K. Pos, P. Butaye, K. Walravens, K. Langton, R.B. Herbert, R.A. Skurray, I.T. Paulsen, J. O'Reilly, N.G. Rutherford, M.H. Brown, R.M. Bill & P.J. Henderson, (2006) Microbial drug efflux proteins of the major facilitator superfamily. *Current drug targets* **7**: 793-811.

Sainsbury, S., L.A. Lane, J. Ren, R.J. Gilbert, N.J. Saunders, C.V. Robinson, D.I. Stuart & R.J. Owens, (2009) The structure of CrgA from *Neisseria meningitidis* reveals a new octameric assembly state for LysR transcriptional regulators. *Nucleic acids research* **37**: 4545-4558.

Sambrook, J. & D.W. Russell, (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sanchez-Romero, M.A. & J. Casadesus, (2014) Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 355-360.

Schell, M.A., (1993) Molecular biology of the LysR family of transcriptional regulators. *Annual review of microbiology* **47**: 597-626.

Schlunegger, M.P., M.J. Bennett & D. Eisenberg, (1997) Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem* **50**: 61-122.

Schon, A., R.K. Brown, B.M. Hutchins & E. Freire, (2013) Ligand binding analysis and screening by chemical denaturation shift. *Analytical biochemistry* **443**: 52-57.

Scientific, D., (2006) The PyMOL Molecular Graphics System. In., pp.

Senisterra, G.A. & P.J. Finerty, Jr., (2009) High throughput methods of assessing protein stability and aggregation. *Molecular bioSystems* **5**: 217-223.

Shah, B.S., H.E. Ashwood, S.J. Harrop, D.N. Farrugia, I.T. Paulsen & B.C. Mabbutt, (2018) Crystal structure of a UDP-GlcNAc epimerase for surface polysaccharide biosynthesis in *Acinetobacter baumannii*. *PloS one* **13**: e0191610.

Sheehan, D., (2009) *Physical biochemistry : principles and applications,* p. xiv, 407 p., 408 p. of plates. Wiley-Blackwell, Chichester, UK ; Hoboken, NJ.

Shoichet, B.K., W.A. Baase, R. Kuroki & B.W. Matthews, (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 452-456.

Sikic, K., S. Tomic & O. Carugo, (2010) Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. *Open Biochem J* **4**: 83-95.

Simeonov, A., (2013) Recent developments in the use of differential scanning fluorometry in protein and small molecule discovery and characterization. *Expert opinion on drug discovery* **8**: 1071-1082.

Singhal, S.S., S.P. Singh, P. Singhal, D. Horne, J. Singhal & S. Awasthi, (2015) Antioxidant role of glutathione S-transferases: 4-Hydroxynonenal, a key molecule in stress-mediated signaling. *Toxicol Appl Pharmacol* **289**: 361-370.

Skippington, E. & M.A. Ragan, (2011) Lateral genetic transfer and the construction of genetic exchange communities. *FEMS microbiology reviews* **35**: 707-735.

Smith, M.T., A.K. Hawes & B.C. Bundy, (2013) Reengineering viruses and virus-like particles through chemical functionalization strategies. *Curr Opin Biotechnol* **24**: 620-626.

Srinivasan, V.B., G. Rajamohan & W.A. Gebreyes, (2009) Role of AbeS, a novel efflux pump of the SMR family of transporters, in resistance to antimicrobial agents in *Acinetobacter baumannii*. *Antimicrobial agents and chemotherapy* **53**: 5312-5316.

Stark, M.J., (1987) Multicopy expression vectors carrying the lac repressor gene for regulated high-level expression of genes in *Escherichia coli*. *Gene* **51**: 255-267.

Stebbins, C.E. & J.E. Galan, (2001) Structural mimicry in bacterial virulence. *Nature* **412**: 701-705.

Stokes, H.W. & M.R. Gillings, (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS microbiology reviews* **35**: 790-819.

Stokes, H.W. & R.M. Hall, (1989) A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular microbiology* **3**: 1669-1683.

Stokes, H.W., A.J. Holmes, B.S. Nield, M.P. Holley, K.M. Nevalainen, B.C. Mabbutt & M.R. Gillings, (2001) Gene cassette PCR: sequence-independent recovery of entire genes from environmental DNA. *Applied and environmental microbiology* **67**: 5240-5246.

Striegel, A.M., (2009) *Modern size-exclusion liquid chromatography : practice of gel permeation and gel filtration chromatography,* p. xvi, 494 p. Wiley, Hoboken, N.J.

Studier, F.W., (2005) Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**: 207-234.

Su, X.Z., J. Chen, T. Mizushima, T. Kuroda & T. Tsuchiya, (2005) AbeM, an H+-coupled *Acinetobacter baumannii* multidrug efflux pump belonging to the MATE family of transporters. *Antimicrobial agents and chemotherapy* **49**: 4362-4364.

Sun, J., J.R. Bankston, J. Payandeh, T.R. Hinds, W.N. Zagotta & N. Zheng, (2014a) Crystal structure of the plant dual-affinity nitrate transporter NRT1.1. *Nature* **507**: 73-77.

Sun, J., Z. Deng & A. Yan, (2014b) Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochem Biophys Res Commun* **453**: 254-267.

Sureshan, V., (2012) Protein structures from the environmental mobile cassette metagenome. In: Department of Chemistry and Biomolecular Sciences. Sydney,Australia: Macquarie University, pp.

Sureshan, V., C.N. Deshpande, Y. Boucher, J.E. Koenig, G. Midwest Center for Structural, H.W. Stokes, S.J. Harrop, P.M. Curmi & B.C. Mabbutt, (2013) Integron gene cassettes: a repository of novel protein folds with distinct interaction sites. *PloS one* **8**: e52934.

Tabor, C.W. & H. Tabor, (1985) Polyamines in microorganisms. *Microbiological reviews* **49**: 81-99.

Taylor, J.L., R.S. De Silva, G. Kovacikova, W. Lin, R.K. Taylor, K. Skorupski & F.J. Kull, (2012) The crystal structure of AphB, a virulence gene activator from *Vibrio cholerae*, reveals residues that influence its response to oxygen and pH. *Molecular microbiology* **83**: 457-470.

Teale, F.W. & G. Weber, (1957) Ultraviolet fluorescence of the aromatic amino acids. *Biochem J* **65**: 476-482.

Technologies, A., (2015) QuikChange II Site-Directed Mutagenesis Kit. In: Instruction Manual. pp.

Teichmann, S.A., (2002) The constraints protein-protein interactions place on sequence divergence. *Journal of molecular biology* **324**: 399-407.

Tettelin, H., V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S.V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. Deboy, T.M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S.C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J. O'Connor, S. Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M.R. Wessels, R. Rappuoli & C.M. Fraser, (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America* **102**: 13950-13955.

Thomas, E.L., (1984) Disulfide reduction and sulfhydryl uptake by *Streptococcus mutans*. *Journal of bacteriology* **157**: 240-246.

Thomas, E.L., K.A. Pera, K.W. Smith & A.K. Chwang, (1983) Inhibition of *Streptococcus mutans* by the lactoperoxidase antimicrobial system. *Infect Immun* **39**: 767-778.

Toprak, E., A. Veres, J.B. Michel, R. Chait, D.L. Hartl & R. Kishony, (2011) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature genetics* **44**: 101-105.

Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M.E. Karoui, E. Frapy, L. Garry, J.M. Ghigo, A.M. Gilles, J. Johnson, C. Le Bouguenec, M. Lescat, S. Mangenot, V. Martinez-Jehanne, I. Matic, X. Nassif, S. Oztas, M.A. Petit, C. Pichon, Z. Rouy, C.S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Medigue, E.P. Rocha & E. Denamur, (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS genetics* **5**: e1000344.

Tropel, D. & J.R. van der Meer, (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiology and molecular biology reviews : MMBR* **68**: 474-500, table of contents.

Valdar, W.S. & J.M. Thornton, (2001) Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins* **42**: 108-124.

Venyaminov, S., I.A. Baikalov, Z.M. Shen, C.S. Wu & J.T. Yang, (1993) Circular dichroic analysis of denatured proteins: inclusion of denatured proteins in the reference set. *Analytical biochemistry* **214**: 17-24.

Viola, R., P. Carman, J. Walsh, E. Miller, M. Benning, D. Frankel, A. McPherson, B. Cudney & B. Rupp, (2007) Operator-assisted harvesting of protein crystals using a universal micromanipulation robot. *Journal of applied crystallography* **40**: 539-545.

Visca, P., L. Leoni, M.J. Wilson & I.L. Lamont, (2002) Iron transport and regulation, cell signalling and genomics: lessons from Escherichia coli and Pseudomonas. *Molecular microbiology* **45**: 1177-1190.

Vlachy, N., B. Jagoda-Cwiklik, R. Vacha, D. Touraud, P. Jungwirth & W. Kunz, (2009) Hofmeister series and specific interactions of charged headgroups with aqueous ions. *Adv Colloid Interface Sci* **146**: 42-47.

Vojtechovsky, J., K. Chu, J. Berendzen, R.M. Sweet & I. Schlichting, (1999) Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophysical journal* **77**: 2153-2174.

Wagner, V., A. Dullaart, A.K. Bock & A. Zweck, (2006) The emerging nanomedicine landscape. *Nat Biotechnol* **24**: 1211-1217.

Waldron, T.T. & K.P. Murphy, (2003) Stabilization of proteins by ligand binding: application to drug screening and determination of unfolding energetics. *Biochemistry* **42**: 5058-5064.

Ward, D.M., M.J. Ferris, S.C. Nold & M.M. Bateson, (1998) A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and molecular biology reviews : MMBR* **62**: 1353-1370.

Wass, M.N., A. David & M.J. Sternberg, (2011) Challenges for the prediction of macromolecular interactions. *Current opinion in structural biology* **21**: 382-390.

Woody, R.W., (1994) Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins. *European biophysics journal : EBJ* **23**: 253-262.

Xia, J., J. Gao & W. Tang, (2016) Nosocomial infection and its molecular mechanisms of antibiotic resistance. *Biosci Trends* **10**: 14-21.

Yewdall, N.A., A.V. Peskin, M.B. Hampton, D.C. Goldstone, F.G. Pearce & J.A. Gerrard, (2018) Quaternary structure influences the peroxidase activity of peroxiredoxin 3. *Biochem Biophys Res Commun* **497**: 558-563.

Yildiz, I., S. Shukla & N.F. Steinmetz, (2011) Applications of viral nanoparticles in medicine. *Curr Opin Biotechnol* **22**: 901-908.

Yoon, E.J., P. Courvalin & C. Grillot-Courvalin, (2013) RND-type efflux pumps in multidrug-resistant clinical isolates of *Acinetobacter baumannii*: major role for AdeABC overexpression and AdeRS mutations. *Antimicrobial agents and chemotherapy* **57**: 2989-2995.

Yoshida, M., K. Kashiwagi, A. Shigemasa, S. Taniguchi, K. Yamamoto, H. Makinoshima, A. Ishihama & K. Igarashi, (2004) A unifying model for the role of polyamines in bacterial cell growth, the polyamine modulon. *The Journal of biological chemistry* **279**: 46008-46013.

Yu, Z., S.E. Reichheld, A. Savchenko, J. Parkinson & A.R. Davidson, (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. *Journal of molecular biology* **400**: 847-864.

Zhou, X., Z. Lou, S. Fu, A. Yang, H. Shen, Z. Li, Y. Feng, M. Bartlam, H. Wang & Z. Rao, (2010) Crystal structure of ArgP from *Mycobacterium tuberculosis* confirms two distinct conformations of full-length LysR transcriptional regulators and reveals its function in DNA binding and transcriptional regulation. *Journal of molecular biology* **396**: 1012-1024.

Appendix

Table S.1 Silver Bullets screen contents (Hampton Research). All screens in 0.02 M HEPES buffer pH 6.8 (https://www.hamptonresearch.com/default.aspx).

well number	screen contents
A1	0.33% w/v 1,5-naphthalenedisulfonic acid disodium salt, 0.33% w/v 2,5-pyridinedicarboxylic acid, 0.33% w/v 3,5-dinitrosalicylic acid
A2	0.25% w/v benzidine, 0.25% w/v nicotinamide, 0.25% w/v pyromellitic acid, 0.25% w/v sulfaguanidine
A3	0.25% w/v Gly-Gly, 0.25% w/v Gly-Gly-Gly, 0.25% w/v Gly-Gly-Gly-Gly, 0.25% w/v pentaglycine
A4	0.25% w/v 3,5-dinitrosalicylic acid, 0.25% w/v 4-aminobenzoic acid, 0.25% w/v salicylic acid, 0.25% w/v trimesic acid
A5	0.33% w/v 4-nitrobenzoic acid, 0.33% w/v 5-sulfosalicylic acid dihydrate, 0.33% w/v naphthalene-1,3,6-trisulfonic acid trisodium salt hydrate
A6	0.33% w/v 2,6-naphthalenedisulfonic acid disodium salt, 0.33% w/v 2,7-naphthalenedisulfonic acid disodium salt, 0.33% w/v anthraquinone-2,6-disulfonic acid disodium salt
A7	0.33% w/v 1,5-naphthalenedisulfonic acid disodium salt, 0.33% w/v naphthalene-1,3,6-trisulfonic acid trisodium salt hydrate, 0.33% w/v PIPES
A8	0.25% w/v sodium 1-pentanesulfonate monohydrate, 0.25% w/v 3,5-dinitrosalicylic acid, 0.25% w/v 3-aminosalicylic acid, 0.25% w/v salicylamide
A9	0.16% w/v L-histidine, 0.16% w/v L-isoleucine, 0.16% w/v L-leucine, 0.16% w/v L-phenylalanine, 0.16% w/v L-tryptophan, 0.16% w/v L-tyrosine
A10	0.2% w/v D-(+)-trehalose dihydrate, 0.2% w/v guanidine hydrochloride, 0.2% w/v phenol, 0.2% w/v trimethylamine N-oxide dihydrate, 0.2% w/v urea
A11	0.33% w/v 2,5-pyridinedicarboxylic acid, 0.33% w/v 4-nitrobenzoic acid, 0.33% w/v mellitic acid
A12	0.25% w/v benzidine, 0.25% w/v phenylglyoxal monohydrate, 0.25% w/v sulfaguanidine, 0.25% w/v sulfanilamide
B1	0.33% w/v anthrone, 0.33% w/v Congo Red, 0.33% w/v N-(2-acetamido)-2-aminoethanesulfonic acid
B2	0.33% w/v 1,3,5-pentanetricarboxylic acid, 0.33% w/v 5-sulfosalicylic acid dihydrate, 0.33% w/v trimesic acid
B3	0.25% w/v 5-sulfoisophthalic acid monosodium salt, 0.25% w/v cystathionine, 0.25% w/v dithioerythritol, 0.25% w/v L-citrulline
B4	0.33% w/v 3,5-dinitrosalicylic acid, 0.33% w/v 3-aminobenzenesulfonic acid, 0.33% w/v 5-sulfosalicylic acid dihydrate
B5	0.33% w/v 2,7-naphthalenedisulfonic acid disodium salt, 0.33% w/v azelaic acid, 0.33% w/v trans-cinnamic acid
B6	0.33% w/v 2,6-naphthalenedisulfonic acid disodium salt, 0.33% w/v 2-aminobenzenesulfonic acid, 0.33% w/v m-benzenedisulfonic acid disodium salt
B7	0.33% w/v 1,4-cyclohexanedicarboxylic acid, 0.33% w/v 2,2'-thiodiglycolic acid, 0.33% w/v 5-sulfoisophthalic acid monosodium salt
B8	0.33% w/v 3-aminobenzoic acid, 0.33% w/v 3-aminosalicylic acid, 0.33% w/v salicylic acid
B9	0.25% w/v hexamminecobalt(III) chloride, 0.25% w/v salicylamide, 0.25% w/v sulfanilamide, 0.25% w/v vanillic acid
B10	0.25% w/v p-coumaric acid, 0.25% w/v phenylurea, 0.25% w/v poly(3-hydroxybutyric acid), 0.25% w/v sulfaguanidine

B11	0.25% w/v 1,2-diaminocyclohexane sulfate, 0.25% w/v 1,4-cyclohexanedicarboxylic acid, 0.25% w/v methylenediphosphonic acid, 0.25% w/v sulfanilic acid
B12	0.25% w/v D-fructose 1,6-bisphosphate trisodium salt hydrate, 0.25% w/v D-glucose 6-phosphate, 0.25% w/v L-O-phosphoserine, 0.25% w/v O-phospho-L-tyrosine
C1	0.25% w/v benzamidine hydrochloride, 0.25% w/v L-carnitine hydrochloride, 0.25% w/v L-cystine, 0.25% w/v L-ornithine hydrochloride
C2	0.33% w/v caffeine, 0.33% w/v dithioerythritol, 0.33% w/v L-methionine
C3	0.25% w/v Ala-Ala, 0.25% w/v Ala-Gly, 0.25% w/v Gly-Gly-Gly-Gly, 0.25% w/v Leu-Gly-Gly
C4	0.2% w/v aspartame, 0.2% w/v Gly-Asp, 0.2% w/v Gly-Ser, 0.2% w/v Ser-Tyr, 0.2% w/v Tyr-Phe
C5	0.16% w/v Ala-Ala, 0.16% w/v aspartame, 0.16% w/v Gly-Tyr, 0.16% w/v Leu-Gly-Gly, 0.16% w/v Ser-Glu, 0.16% w/v Tyr-Ala
C6	0.33% w/v Gly-Phe, 0.33% w/v Gly-Tyr, 0.33% w/v Leu-Gly-Gly
C7	0.16% w/v Ala-Ala, 0.16% w/v Gly-Asp, 0.16% w/v Gly-Gly, 0.16% w/v Gly-Phe, 0.16% w/v Gly-Ser, 0.16% w/v Ser-Tyr
C8	0.05% w/v glycine, 0.05% w/v L-(-)-threonine, 0.05% w/v L-(+)-lysine, 0.05% w/v L-alanine, 0.05% w/v L-arginine, 0.05% w/v L-asparagine monohydrate, 0.05% w/v L-aspartic acid, 0.05% w/v L-glutamic acid, 0.05% w/v L-glutamine, 0.05% w/v L-histidine, 0.05% w/v L-isoleucine, 0.05% w/v L-leucine, 0.05% w/v L-methionine, 0.05% w/v L-phenylalanine, 0.05% w/v L-proline, 0.05% w/v L-serine, 0.05% w/v L-tryptophan, 0.05% w/v L-tyrosine, 0.05% w/v L-valine
C9	0.2% w/v D-(+)-maltose monohydrate, 0.2% w/v D-(+)-melibiose monohydrate, 0.2% w/v D-(+)-raffinose pentahydrate, 0.2% w/v D-(+)-trehalose dihydrate, 0.2% w/v stachyose hydrate
C10	0.16% w/v β-cyclodextrin, 0.16% w/v D-(+)-cellobiose, 0.16% w/v D-(+)-maltotriose, 0.16% w/v D-(+)-melezitose hydrate, 0.16% w/v D-(+)-raffinose pentahydrate, 0.16% w/v stachyose hydrate
C11	0.16% w/v azelaic acid, 0.16% w/v m-benzenedisulfonic acid disodium salt, 0.16% w/v mellitic acid, 0.16% w/v pimelic acid, 0.16% w/v pyromellitic acid, 0.16% w/v trans-cinnamic acid
C12	0.25% w/v 5-sulfoisophthalic acid monosodium salt, 0.25% w/v anthraquinone-2,6-disulfonic acid disodium salt, 0.25% w/v N-(2-acetamido)-2-aminoethanesulfonic acid, 0.25% w/v tetrahydroxy-1,4-benzoquinone hydrate
D1	0.25% w/v 1,3,5-pentanetricarboxylic acid, 0.25% w/v 5-sulfosalicylic acid dihydrate, 0.25% w/v o-sulfobenzoic acid monoammonium salt, 0.25% w/v sodium 4-aminosalicylate dihydrate
D2	0.06 M CHAPS, 0.06 M HEPES, 0.06 M Tris, 0.25% w/v hexamminecobalt(III) chloride
D3	0.06 M MES monohydrate, 0.06 M PIPES, 0.33% w/v hexamminecobalt(III) chloride
D4	0.005 M gadolinium(III) chloride hexahydrate, 0.005 M samarium(III) chloride hexahydrate, 0.05 M benzamidine hydrochloride, 0.25% w/v salicin
D5	0.004 M calcium chloride dihydrate, 0.004 M magnesium chloride hexahydrate, 0.004 M manganese(II) chloride tetrahydrate, 0.004 M zinc chloride
D6	0.004 M cadmium chloride hydrate, 0.004 M cobalt(II) chloride hexahydrate, 0.004 M copper(II) chloride dihydrate, 0.004 M nickel(II) chloride hexahydrate
D7	0.25% w/v 3,5-dinitrosalicylic acid, 0.25% w/v 3-indolebutyric acid, 0.25% w/v naphthalene-1,3,6-trisulfonic acid trisodium salt hydrate, 0.25% w/v trans-1,2-cyclohexanedicarboxylic acid
D8	0.2% w/v betaine anhydrous, 0.2% w/v L-glutamic acid, 0.2% w/v L-proline, 0.2% w/v taurine, 0.2% w/v trimethylamine N-oxide dihydrate
D9	0.25% w/v 1,2-diaminocyclohexane sulfate, 0.25% w/v 4-nitrobenzoic acid, 0.25% w/v cystamine dihydrochloride, 0.25% w/v spermine

D10	0.25% w/v 1,5-naphthalenedisulfonic acid disodium salt, 0.25% w/v 2,7-naphthalenedisulfonic acid disodium salt, 0.25% w/v 5-sulfoisophthalic acid monosodium salt, 0.25% w/v sulfanilic acid
D11	0.25% w/v 2,6-naphthalenedisulfonic acid disodium salt, 0.25% w/v 4-aminobenzoic acid, 0.25% w/v 5-sulfosalicylic acid dihydrate, 0.25% w/v naphthalene-1,3,6-trisulfonic acid trisodium salt hydrate
D12	0.2% w/v rhenium(IV) oxide, 0.2% w/v sodium bromide, 0.2% w/v sodium nitrate, 0.2% w/v sodium phosphate dibasic dihydrate, 0.2% w/v sodium tetraborate decahydrate
E1	0.2% w/v caffeine, 0.2% w/v cytosine, 0.2% w/v gallic acid monohydrate, 0.2% w/v nicotinamide, 0.2% w/v sodium pyrophosphate tetrabasic decahydrate
E2	1% w/v dextran sulfate sodium salt, 0.005% w/v dextranase, 0.005% w/v α-amylase
E3	1% w/v tryptone
E4	1% w/v protamine sulfate
E5	0.005% w/v deoxyribonuclease I, 0.5% w/v deoxyribonucleic acid, 0.005% w/v ribonuclease A, 0.5% w/v ribonucleic acid
E6	0.5% w/v casein, 0.5% w/v hemoglobin, 0.005% w/v pepsin, 0.005% w/v protease, 0.005% w/v proteinase K, 0.005% w/v trypsin
E7	1% w/v ovalbumin, 0.005% w/v pepsin, 0.005% w/v proteinase K, 0.005% w/v trypsin
E8	0.2% w/v D-sorbitol, 0.2% w/v glycerol, 0.2% w/v glycine, 0.2% w/v myo-inositol, 0.2% w/v sarcosine
E9	0.2% w/v 1,4-diaminobutane, 0.2% w/v cystamine dihydrochloride, 0.2% w/v diloxanide furoate, 0.2% w/v sarcosine, 0.2% w/v spermine
E10	0.25% w/v 1,2-diaminocyclohexane sulfate, 0.25% w/v 1,8-diaminooctane, 0.25% w/v cadaverine, 0.25% w/v spermine
E11	0.2% w/v 1,2-diaminocyclohexane sulfate, 0.2% w/v diloxanide furoate, 0.2% w/v fumaric acid, 0.2% w/v spermine, 0.2% w/v sulfaguanidine
E12	0.2% w/v 1,4-diaminobutane, 0.2% w/v 1,8-diaminooctane, 0.2% w/v cadaverine, 0.2% w/v cystamine dihydrochloride, 0.2% w/v spermidine
F1	0.25% w/v methylenediphosphonic acid, 0.25% w/v phytic acid sodium salt hydrate, 0.25% w/v sodium pyrophosphate tetrabasic decahydrate, 0.25% w/v sodium triphosphate pentabasic
F2	0.2% w/v D-fructose 1,6-bisphosphate trisodium salt hydrate, 0.2% w/v glycerol phosphate disodium salt hydrate, 0.2% w/v L-O-phosphoserine, 0.2% w/v O-phospho-L-tyrosine, 0.2% w/v phytic acid sodium salt hydrate
F3	0.16% w/v 4-aminobutyric acid, 0.16% w/v 6-aminohexanoic acid, 0.16% w/v L-(+)-lysine, 0.16% w/v L-ornithine hydrochloride, 0.16% w/v taurine, 0.16% w/v β-alanine
F4	0.2% w/v L-arginine, 0.2% w/v L-canavanine, 0.2% w/v L-carnitine hydrochloride, 0.2% w/v L-citrulline, 0.2% w/v taurine
F5	0.2% w/v 1,2,3-heptanetriol, 0.2% w/v 1,3-propanediol, 0.2% w/v 1,6-hexanediol, 0.2% w/v Gly-Gly, 0.2% w/v resorcinol
F6	0.2% w/v (+/-)-2-methyl-2,4-pentanediol, 0.2% w/v 1,2,3-heptanetriol, 0.2% w/v Diethylenetriaminepentakis(methylphosphonic acid), 0.2% w/v D-sorbitol, 0.2% w/v glycerol
F7	0.2% w/v barbituric acid, 0.2% w/v betaine anhydrous, 0.2% w/v phloroglucinol, 0.2% w/v resorcinol, 0.2% w/v tetrahydroxy-1,4-benzoquinone hydrate
F8	0.2% w/v 1,6-hexanediol, 0.2% w/v Diethylenetriaminepentakis(methylphosphonic acid), 0.2% w/v Gly-Gly, 0.2% w/v myo-inositol, 0.2% w/v phloroglucinol
F9	0.2% w/v 6-aminohexanoic acid, 0.2% w/v benzamidine hydrochloride, 0.2% w/v Congo Red, 0.2% w/v nicotinamide, 0.2% w/v salicin
F10	0.2% w/v anthrone, 0.2% w/v benzidine, 0.2% w/v N-(2-acetamido)-2-aminoethanesulfonic acid, 0.2% w/v phenylurea, 0.2% w/v β-alanine

F11	0.25% w/v sodium 1-pentanesulfonate monohydrate, 0.25% w/v 4-aminobutyric acid, 0.25% w/v cytosine, 0.25% w/v salicylamide
F12	0.11% w/v dodecanedioic acid, 0.11% w/v fumaric acid, 0.11% w/v glutaric acid, 0.11% w/v hexadecanedioic acid, 0.11% w/v maleic acid, 0.11% w/v oxamic acid, 0.11% w/v pimelic acid, 0.11% w/v sebacic acid, 0.11% w/v suberic acid
G1	0.16% w/v 5-sulfosalicylic acid dihydrate, 0.16% w/v dodecanedioic acid, 0.16% w/v hippuric acid, 0.16% w/v mellitic acid, 0.16% w/v oxalacetic acid, 0.16% w/v suberic acid
G2	0.2% w/v 2,2'-thiodiglycolic acid, 0.2% w/v adipic acid, 0.2% w/v benzoic acid, 0.2% w/v oxalic acid, 0.2% w/v terephthalic acid
G3	0.25% w/v 2,2'-thiodiglycolic acid, 0.25% w/v azelaic acid, 0.25% w/v mellitic acid, 0.25% w/v trans-aconitic acid
G4	0.16% w/v 3-indolebutyric acid, 0.16% w/v hexadecanedioic acid, 0.16% w/v oxamic acid, 0.16% w/v pyromellitic acid, 0.16% w/v sebacic acid, 0.16% w/v suberic acid
G5	0.25% w/v 1,3,5-pentanetricarboxylic acid, 0.25% w/v 4-hydroxyphenylacetic acid, 0.25% w/v benzoic acid, 0.25% w/v poly(3-hydroxybutyric acid)
G6	0.16% w/v glutaric acid, 0.16% w/v mellitic acid, 0.16% w/v oxalic acid, 0.16% w/v pimelic acid, 0.16% w/v sebacic acid, 0.16% w/v trans-cinnamic acid
G7	0.2% w/v 4-aminobenzoic acid, 0.2% w/v azelaic acid, 0.2% w/v o-sulfobenzoic acid monoammonium salt, 0.2% w/v p-coumaric acid, 0.2% w/v sodium 4-aminosalicylate dihydrate
G8	0.16% w/v 3-aminobenzenesulfonic acid, 0.16% w/v 3-aminobenzoic acid, 0.16% w/v hippuric acid, 0.16% w/v oxalacetic acid, 0.16% w/v salicylic acid, 0.16% w/v trimesic acid
G9	0.2% w/v 2-aminobenzenesulfonic acid, 0.2% w/v 3-indolebutyric acid, 0.2% w/v 4-hydroxyphenylacetic acid, 0.2% w/v barbituric acid, 0.2% w/v terephthalic acid
G10	0.2% w/v 1,4-cyclohexanedicarboxylic acid, 0.2% w/v 2,5-pyridinedicarboxylic acid, 0.2% w/v glutaric acid, 0.2% w/v trans-1,2-cyclohexanedicarboxylic acid, 0.2% w/v trans-aconitic acid
G11	10% v/v tacsimate pH 7.0
G12	0.2% w/v benzenephosphonic acid, 0.2% w/v gallic acid monohydrate, 0.2% w/v melatonin, 0.2% w/v N-(2-carboxyethyl)-iminodiacetic acid, 0.2% w/v trimellitic acid
H1	0.2% w/v D-(−)-3-phosphoglyceric acid disodium salt, 0.2% w/v maleic acid, 0.2% w/v 1,3-propanediol, 0.2% w/v glycerol phosphate disodium salt hydrate, 0.2% w/v Gly-Gly
H2	0.2% w/v Ala-Ala, 0.2% w/v Ala-Gly, 0.2% w/v Gly-Asp, 0.2% w/v Gly-Phe, 0.2% w/v Ser-Glu
H3	0.2% w/v 3,5-dinitrosalicylic acid, 0.2% w/v 4-aminobenzoic acid, 0.2% w/v benzamidine hydrochloride, 0.2% w/v hexamminecobalt(III) chloride, 0.2% w/v mellitic acid
H4	0.16% w/v 1,4-diaminobutane, 0.16% w/v 1,8-diaminooctane, 0.16% w/v cadaverine, 0.16% w/v cystamine dihydrochloride, 0.16% w/v spermidine, 0.16% w/v spermine
H5	0.16% w/v 4-aminobutyric acid, 0.16% w/v 6-aminohexanoic acid, 0.16% w/v oxamic acid, 0.16% w/v sulfanilic acid, 0.16% w/v trimesic acid, 0.16% w/v β-alanine
H6	0.16% w/v D-3-phosphoglyceric acid disodium salt, 0.16% w/v D-fructose 1,6-bisphosphate trisodium salt hydrate, 0.16% w/v D-glucose 6-phosphate, 0.16% w/v L-O-phosphoserine, 0.16% w/v O-phospho-L-tyrosine, 0.16% w/v phytic acid sodium salt hydrate
H7	0.0625% w/v 1,3,5-pentanetricarboxylic acid, 0.0625% w/v azelaic acid, 0.0625% w/v dodecanedioic acid, 0.0625% w/v glutaric acid, 0.0625% w/v hexadecanedioic acid, 0.0625% w/v pimelic acid, 0.0625% w/v sebacic acid, 0.0625% w/v suberic acid
H8	0.16% w/v 1,5-naphthalenedisulfonic acid disodium salt, 0.16% w/v 2,6-naphthalenedisulfonic acid disodium salt, 0.16% w/v 2,7-naphthalenedisulfonic acid disodium salt, 0.16% w/v 4-nitrobenzoic acid, 0.16% w/v m-benzenedisulfonic acid disodium salt, 0.16% w/v naphthalene-1,3,6-trisulfonic acid trisodium salt hydrate

H9	0.2% w/v 2,5-pyridinedicarboxylic acid, 0.2% w/v pyromellitic acid, 0.2% w/v salicylic acid, 0.2% w/v trans-1,2-cyclohexanedicarboxylic acid, 0.2% w/v trans-cinnamic acid
H10	0.16% w/v 3-aminobenzenesulfonic acid, 0.16% w/v 5-sulfosalicylic acid dihydrate, 0.16% w/v p-coumaric acid, 0.16% w/v PIPES, 0.16% w/v terephthalic acid, 0.16% w/v vanillic acid
H11	0.07% w/v barbituric acid, 0.07% w/v benzidine, 0.07% w/v cystathionine, 0.07% w/v L-canavanine, 0.07% w/v L-carnitine hydrochloride, 0.07% w/v L-cystine, 0.07% w/v mellitic acid
H12	0.16% w/v aspartame, 0.16% w/v Gly-Gly-Gly, 0.16% w/v Leu-Gly-Gly, 0.16% w/v pentaglycine, 0.16% w/v Tyr-Ala, 0.16% w/v Tyr-Phe

Table S.2 Silver Bullets Bio screen contents (Hampton Research). All screens in 0.02 M HEPES buffer pH 6.8.

well number	screen contents
A1	0.16% w/v L-citrulline, 0.16% w/v L-ornithine hydrochloride, 0.16% w/v urea, 0.16% w/v oxalic acid, 0.16% w/v kanamycin monosulfate, 0.16% w/v L-arginine
A2	0.16% w/v L-carnitine hydrochloride, 0.16% w/v tannic acid, 0.16% w/v aspartame, 0.16% w/v caffeine, 0.16% w/v p-coumaric acid, 0.16% w/v 4-hydroxy-L-proline
A3	0.16% w/v disodium beta-glycerophosphate tetrahydrate, 0.16% w/v trans-cinnamic acid, 0.16% w/v O-phospho-L-tyrosine, 0.16% w/v betaine anhydrous, 0.16% w/v maltotriose hydrate, 0.16% w/v cytidine
A4	0.16% w/v L-canavanine, 0.16% w/v O-phospho-L-serine, 0.16% w/v taurine, 0.16% w/v quinine, 0.16% w/v sodium glyoxylate monohydrate, 0.16% w/v cholic acid
A5	0.08% w/v ellipticine, 0.20% w/v gibberellin A_3, 0.20% w/v trans-cinnamic acid, 0.20% w/v phenol, 0.20% w/v succinic acid disodium salt hexahydrate
A6	0.16% w/v sulfanilamide, 0.16% w/v D-(+)-cellobiose, 0.16% w/v D-glyceric acid calcium salt dihydrate, 0.16% w/v 6-phosphogluconic acid trisodium salt, 0.16% w/v Nα-benzoyl-L-arginine ethyl ester hydrochloride, 0.16% w/v D-(+) galactosamine hydrochloride
A7	0.16% w/v benzamidine hydrochloride, 0.16% w/v acarbose, 0.16% w/v ethidium bromide, 0.16% w/v cholesterol, 0.16% w/v sarcosine, 0.16% w/v L-homoserine
A8	0.20% w/v sodium gluconate, 0.20% w/v acarbose, 0.20% w/v D-(+)-maltose monohydrate, 0.20% w/v salicylic acid, 0.20% w/v D-(+)-melibiose monohydrate
A9	0.20% w/v g-strophanthin, 0.20% w/v D-sorbitol, 0.20% w/v 1,4-diaminobutane, 0.20% w/v D-(+)-trehalose dihydrate, 0.20% w/v L-α-phosphatidylcholine
A10	0.16% w/v sucrose, 0.16% w/v cadaverine, 0.16% w/v L-glutamic acid, 0.16% w/v L-arginine, 0.16% w/v oxalic acid, 0.16% w/v tetrahydrofolic acid
A11	0.20% w/v protamine sulfate salt, 0.20% w/v g-strophanthin, 0.20% w/v benzamidine hydrochloride, 0.20% w/v D-fructose 1,6-diphosphate trisodium salt octahydrate, 0.20% w/v oxamic acid
A12	0.20% w/v 2'-deoxyadenosine 5'-triphosphate disodium salt, 0.20% w/v benzidine, 0.20% w/v L-carnitine hydrochloride, 0.20% w/v sulfanilamide, 0.20% w/v cytosine
B1	0.20% w/v 2'-deoxyguanosine 5'-monophosphate sodium salt hydrate, 0.20% w/v N-acetyl-D-galactosamine, 0.20% w/v O-phospho-L-serine, 0.20% w/v 4-aminobenzoic acid, 0.20% w/v uridine
B2	0.16% w/v thymidine 5'-triphosphate sodium salt, 0.16% w/v L-canavanine, 0.16% w/v D-sorbitol, 0.16% w/v salicylic acid, 0.16% w/v 2'-deoxyguanosine hydrate, 0.16% w/v D-(+)-raffinose pentahydrate

B3	0.16% w/v cytidine 5’-triphosphate disodium salt, 0.16% w/v stachyose hydrate, 0.16% w/v cystathionine, 0.16% w/v cystamine dihydrochloride, 0.16% w/v trans-cinnamic acid, 0.16% w/v inosine 5’-monophosphate disodium salt
B4	0.20% w/v uridine 5’-diphosphate sodium salt, 0.20% w/v quinine, 0.20% w/v p-coumaric acid, 0.20% w/v D-(+)-melibiose monohydrate, 0.20% w/v cystamine dihydrochloride
B5	0.20% w/v nicotinic acid, 0.20% w/v inosine 5’-monophosphate disodium salt, 0.20% w/v gibberellin A_3, 0.20% w/v O-phospho-L-tyrosine, 0.20% w/v caffeine
B6	0.20% w/v adenosine 3’,5’-cyclic monophosphate sodium salt monohydrate, 0.20% w/v cadaverine, 0.20% w/v D-(+)-melezitose hydrate, 0.20% w/v aspartame, 0.20% w/v guanosine
B7	0.16% w/v thymidine, 0.16% w/v adenosine 3’,5’-cyclic monophosphate sodium salt monohydrate, 0.16% w/v sarcosine, 0.16% w/v 4-aminobenzoic acid, 0.16% w/v acarbose, 0.16% w/v inosine
B8	0.08% w/v ellipticine, 0.20% w/v protamine sulfate salt, 0.20% w/v D-(+)-trehalose dihydrate, 0.20% w/v 6-phosphogluconic acid trisodium salt, 0.20% w/v D-(+)-glucose
B9	0.20% w/v penicillin G sodium salt, 0.20% w/v L-arginine, 0.20% w/v D-(+)-cellobiose, 0.20% w/v pyridoxamine dihydrochloride, 0.20% w/v betaine anhydrous
B10	0.20% w/v nicotinic acid, 0.20% w/v adenosine, 0.20% w/v L-tyrosine, 0.20% w/v xanthine, 0.20% w/v L-lactic acid
B11	0.20% w/v 2’-deoxyadenosine monohydrate, 0.20% w/v N-acetylneuraminic acid, 0.20% w/v N-p-tosyl-L-phenylalanine chloromethyl ketone, 0.20% w/v α-D-glucose 1-phosphate disodium salt hydrate, 0.20% w/v fumaric acid
B12	0.25% w/v thymidine 5’-triphosphate sodium salt, 0.25% w/v α-ketoglutaric acid disodium salt, 0.25% w/v 2-nitrophenyl β-D-galactopyranoside, 0.25% w/v cis-aconitic acid
C1	0.20% w/v 2’-deoxyguanosine 5’-monophosphate sodium salt hydrate, 0.20% w/v ethanolamine, 0.20% w/v theophylline, 0.20% w/v isopropyl 1-thio-β-D-galactopyranoside, 0.20% w/v oxalacetic acid
C2	0.20% w/v 2’-deoxyadenosine 5’-triphosphate disodium salt, 0.20% w/v D-glucose 6-phosphate sodium salt, 0.20% w/v acetylsalicylic acid, 0.20% w/v choline base solution, 0.20% w/v nicotinamide
C3	0.20% w/v uridine 5’-diphosphate sodium salt, 0.20% w/v L-carnosine, 0.20% w/v L-thyroxine, 0.20% w/v sodium phenyl phosphate dibasic dihydrate, 0.20% w/v glutaric acid
C4	0.20% w/v cytosine, 0.20% w/v vitamin B12, 0.20% w/v pepstatin A, 0.20% w/v shikimic acid, 0.20% w/v N-acetyl-L-glutamic acid
C5	0.20% w/v 2’-deoxycytidine 5’-triphosphate disodium salt, 0.20% w/v hydrocortisone, 0.20% w/v hemin, 0.20% w/v O-phospho-DL-threonine, 0.20% w/v L-phenylalanine
C6	0.04% w/v cortisone, 0.04% w/v (±)-epinephrine, 0.04% w/v protoporphyrin disodium salt, 0.04% w/v pyridoxine, 0.04% w/v thymidine 5’-monophosphate disodium salt hydrate
C7	0.20% w/v adenosine, 0.20% w/v pepstatin A, 0.20% w/v (±)-epinephrine, 0.20% w/v sodium phenyl phosphate dibasic dihydrate, 0.20% w/v inosine 5’-triphosphate trisodium salt
C8	0.20% w/v 2’-deoxyadenosine monohydrate, 0.20% w/v O-phospho-DL-threonine, 0.20% w/v flavin adenine dinucleotide disodium salt hydrate, 0.20% w/v DL-xylose, 0.20% w/v nalidixic acid
C9	0.20% w/v guanosine, 0.20% w/v D-glucose 6-phosphate sodium salt, 0.20% w/v 2,3-pyridinedicarboxylic acid, 0.20% w/v L-tyrosine, 0.20% w/v adenosine 5’-diphosphate sodium salt
C10	0.25% w/v 2’-deoxyguanosine hydrate, 0.25% w/v cortisone, 0.25% w/v α-ketoglutaric acid disodium salt, 0.25% w/v pyruvic acid
C11	0.20% w/v thymidine, 0.20% w/v trans-dehydroandrosterone, 0.20% w/v xanthine, 0.20% w/v shikimic acid, 0.20% w/v histamine dihydrochloride

C12	0.20% w/v uridine, 0.20% w/v hydrocortisone, 0.20% w/v N-acetylneuraminic acid, 0.20% w/v 2-nitrophenyl β-D-galactopyranoside, 0.20% w/v phospho(enol)pyruvic acid monosodium salt hydrate
D1	0.20% w/v cytosine, 0.20% w/v pyridoxine, 0.20% w/v hemin, 0.20% w/v L-carnosine, 0.20% w/v cytidine 5'-diphosphocholine sodium salt dihydrate
D2	0.20% w/v cytosine, 0.20% w/v α-D-glucose 1-phosphate disodium salt hydrate, 0.20% w/v trans-dehydroandrosterone, 0.20% w/v isopropyl 1-thio-β-D-galactopyranoside, 0.20% w/v D-(−)-ribose
D3	0.20% w/v thiamine pyrophosphate, 0.20% w/v D-glucosamic acid, 0.20% w/v choline base solution, 0.20% w/v theophylline, 0.20% w/v ethanolamine
D4	0.20% w/v acetylsalicylic acid, 0.20% w/v vitamin B12, 0.20% w/v 2,3-pyridinedicarboxylic acid, 0.20% w/v ethanolamine, 0.20% w/v thiamine hydrochloride
D5	0.16% w/v L-lactic acid, 0.16% w/v L-aspartic acid, 0.16% w/v L-thyroxine, 0.16% w/v pyridoxine, 0.16% w/v L-ascorbic acid, 0.16% w/v phytic acid sodium salt hydrate
D6	0.16% w/v thiamine pyrophosphate, 0.16% w/v L-aspartic acid, 0.16% w/v L-glutamic acid, 0.16% w/v L-(+)-lysine, 0.16% w/v L-arginine, 0.16% w/v L-histidine
D7	0.125% w/v L-methionine, 0.125% w/v L-phenylalanine, 0.125% w/v L-leucine, 0.125% w/v L-isoleucine, 0.125% w/v L-valine, 0.125% w/v L-alanine, 0.125% w/v L-tryptophan, 0.125% w/v L-proline
D8	0.14% w/v glycine, 0.14% w/v L-asparagine monohydrate, 0.14% w/v L-glutamine, 0.14% w/v L-tyrosine, 0.14% w/v L-serine, 0.14% w/v L-(-)-threonine, 0.14% w/v maleic acid
D9	0.14% w/v sodium bromide, 0.14% w/v sodium fluoride, 0.14% w/v sodium carbonate, 0.14% w/v sodium tetraborate decahydrate, 0.14% w/v sodium phosphate dibasic dihydrate, 0.14% w/v sodium nitrate, 0.14% w/v rhenium(IV) oxide
D10	0.002 M calcium chloride dihydrate, 0.002 M cadmium chloride hydrate, 0.002 M magnesium chloride hexahydrate, 0.002 M manganese(II) chloride tetrahydrate, 0.002 M zinc chloride
D11	0.002 M nickel(II) chloride hexahydrate, 0.002 M copper(II) chloride dihydrate, 0.002 M cobalt(II) chloride hexahydrate, 0.002 M molybdenum(III) chloride
D12	0.16% w/v L-homoserine, 0.16% w/v 4-hydroxy-L-proline, 0.16% w/v argininosuccinic acid disodium salt hydrate, 0.16% w/v cytidine, 0.16% w/v inosine, 0.16% w/v guanine
E1	0.16% w/v thiamine monophosphate chloride dihydrate, 0.16% w/v acetylsalicylic acid, 0.16% w/v cholic acid, 0.16% w/v 1,2,3-heptanetriol, 0.16% w/v vanillin, 0.16% w/v N-acetyl-D-mannosamine
E2	0.20% w/v adenosine 5'-(β,γ-imido)triphosphate tetralithium salt hydrate, 0.20% w/v protoporphyrin disodium salt, 0.20% w/v coproporphyrin I dihydrochloride, 0.20% w/v Tris, 0.20% w/v DL-glyceraldehyde
E3	0.16% w/v D-glucosamic acid, 0.16% w/v D-(+)-galactosamine hydrochloride, 0.16% w/v D-(+)-glucosamine hydrochloride, 0.16% w/v sodium gluconate, 0.16% w/v D-mannosamine hydrochloride, 0.16% w/v D-(+)-mannose
E4	0.16% w/v D-glucosamine 6-phosphate sodium salt, 1.00% w/v tryptone, 0.16% w/v isomaltose
E5	0.04% w/v coproporphyrin I dihydrochloride, 0.04% w/v vanillin, 0.04% w/v D-mannosamine hydrochloride, 0.04% w/v D-(+)-fucose, 0.04% w/v guanine
E6	0.05% w/v L-citrulline, 0.05% w/v glycine, 0.05% w/v L-(-)-threonine, 0.05% w/v L-(+)-lysine, 0.05% w/v L-alanine, 0.05% w/v L-arginine, 0.05% w/v L-asparagine monohydrate, 0.05% w/v L-aspartic acid, 0.05% w/v L-glutamic acid, 0.05% w/v L-glutamine, 0.05% w/v L-histidine, 0.05% w/v L-isoleucine, 0.05% w/v L-leucine, 0.05% w/v L-methionine, 0.05% w/v L-phenylalanine, 0.05% w/v L-proline, 0.05% w/v L-serine, 0.05% w/v L-tryptophan, 0.05% w/v L-tyrosine, 0.05% w/v L-valine
E7	0.16% w/v L-(+)-arabinose, 0.16% w/v DL-glyceraldehyde, 0.16% w/v D-glucose 6-phosphate sodium salt, 0.16% w/v D-(+)-fucose, 0.16% w/v D-(+)-mannose,

	0.16% w/v isomaltose
E8	0.16% w/v N-acetylneuraminic acid, 0.16% w/v N-acetylmuramic acid, 0.16% w/v N-acetyl-D-mannosamine, 0.16% w/v N-acetyl-D-glucosamine, 0.16% w/v Tris, 0.16% w/v L-(+)-arabinose
E9	0.16% w/v Tyr-Phe, 0.16% w/v Leu-Leu, 0.16% w/v Ala-Leu, 0.16% w/v Val-Ser, 0.16% w/v Trp-Gly hydrochloride, 0.16% w/v Met-Gly
E10	0.16% w/v Tyr-Phe, 0.16% w/v Tyr-Leu, 0.16% w/v Z-Val-Phe, 0.16% w/v Gly-Gly-Gly, 0.16% w/v Ala-Ala, 0.16% w/v Trp-Gly hydrochloride
E11	0.20% w/v Leu-Gly-Gly, 0.20% w/v Leu-Leu, 0.20% w/v Met-Ala-Ser, 0.20% w/v Ala-Ala-Ala, 0.20% w/v Gly-Gly-Gly, 0.20% w/v Trp-Gly hydrochloride
E12	0.16% w/v Gly-Gly-Gly-Gly, 0.16% w/v Leu-Gly-Gly, 0.16% w/v Met-Ala-Ser, 0.16% w/v Gly-Gly-Gly-Gly-Gly, 0.16% w/v Ala-Leu, 0.16% w/v Tyr-Leu
F1	0.25% w/v β-nicotinamide adenine dinucleotide phosphate tetrasodium salt, 0.25% w/v adenosine 5’-triphosphate disodium salt hydrate, 0.25% w/v N-acetyl-D-galactosamine, 0.25% w/v gentamicin sulfate salt hydrate
F2	0.20% w/v tryptone, 0.20% w/v inosine 5’-monophosphate disodium salt, 0.20% w/v D-lactose monohydrate, 0.20% w/v L-citrulline, 0.20% w/v 2’-deoxycytidine 5’-monophosphate sodium salt
F3	0.20% w/v sodium pyrophosphate tetrabasic decahydrate, 0.20% w/v D-(−)-ribose, 0.20% w/v phytic acid sodium salt hydrate, 0.20% w/v adenosine 3’,5’-cyclic monophosphate sodium salt monohydrate, 0.04% w/v acetyl coenzyme A sodium salt
F4	0.25% w/v S-(5’-adenosyl)-L-methionine chloride, 0.25% w/v disodium beta-glycerophosphate tetrahydrate, 0.25% w/v stachyose hydrate, 0.25% w/v cytidine 5’-diphosphocholine sodium salt dihydrate
F5	0.20% w/v biotin, 0.20% w/v phospho(enol)pyruvic acid monosodium salt hydrate, 0.20% w/v sodium pyruvate, 0.20% w/v D-(+)-melezitose hydrate, 0.20% w/v coenzyme B12
F6	0.20% w/v flavin adenine dinucleotide disodium salt hydrate, 0.20% w/v D-(−)-3-phosphoglyceric acid disodium salt, 0.20% w/v spermine, 0.20% w/v adenosine-5’-diphosphoglucose disodium salt, 0.20% w/v L-tyrosine
F7	0.20% w/v thymine, 0.20% w/v sodium pyrophosphate tetrabasic decahydrate, 0.20% w/v D-glyceric acid calcium salt dihydrate, 0.20% w/v β-cyclodextrin, 0.20% w/v myo-inositol
F8	0.20% w/v β-nicotinamide adenine dinucleotide hydrate, 0.20% w/v D-fructose 1,6-bisphosphate trisodium salt octahydrate, 0.20% w/v spermidine, 0.20% w/v adenosine 5’-diphosphoribose sodium salt, 0.20% w/v histamine dihydrochloride
F9	0.20% w/v pyridoxal 5-phosphate monohydrate, 0.20% w/v pyruvic acid, 0.20% w/v β-cyclodextrin, 0.20% w/v (±)-epinephrine, 0.20% w/v myo-inositol 2-monophosphate bis(cyclohexylammonium) salt
F10	0.20% w/v adenosine 5’-diphosphate sodium salt, 0.20% w/v adenosine 5’-monophosphate sodium salt, 0.20% w/v uridine 5’-monophosphate disodium salt hydrate, 0.20% w/v cytidine 5’-diphosphate sodium salt hydrate, 0.20% w/v cytidine 5’-monophosphate disodium salt
F11	0.20% w/v pyridoxal hydrochloride, 0.20% w/v 2’-deoxyadenosine 5’-monophosphate, 0.20% w/v guanosine 5’-diphosphate sodium salt, 0.20% w/v nalidixic acid, 0.20% w/v uridine 5’-diphospho-N-acetylglucosamine sodium salt
F12	0.25% w/v guanosine 5’-triphosphate sodium salt hydrate, 0.25% w/v thymidine 5’-monophosphate disodium salt hydrate, 0.25% w/v L-tryptophan, 0.25% w/v inosine 5’-triphosphate trisodium salt
G1	0.20% w/v adenosine 5’-triphosphate disodium salt hydrate, 0.20% w/v 2’-deoxycytidine 5’-monophosphate sodium salt, 0.20% w/v β-estradiol, 0.20% w/v D-(+)-galactose, 0.20% w/v 2’-deoxyguanosine hydrate
G2	10% v/v tacsimate pH 7.0

G3	0.20% w/v DL-α-lipoic acid, 0.20% w/v creatine monohydrate, 0.20% w/v L-glutathione reduced, 0.20% w/v D-pantothenic acid hemicalcium salt, 0.20% w/v γ-aminobutyric acid
G4	0.20% w/v pyridoxal hydrochloride, 0.20% w/v nicotinamide, 0.20% w/v batyl alcohol, 0.20% w/v glutaric acid, 0.20% w/v N-acetyl-L-glutamic acid
G5	0.20% w/v riboflavin 5'-monophosphate sodium salt dihydrate, 0.20% w/v maleic acid, 0.20% w/v pyridoxamine dihydrochloride, 0.20% w/v acetylcholine chloride, 0.20% w/v uridine 5'-diphospho-N-acetylglucosamine sodium salt
G6	0.20% w/v fumaric acid, 0.20% w/v cis-aconitic acid, 0.20% w/v DL-isocitric acid trisodium salt, 0.20% w/v oxalacetic acid, 0.20% w/v sodium pyruvate
G7	0.04% w/v β-nicotinamide adenine dinucleotide phosphate sodium salt hydrate, 0.04% w/v tetrahydrofolic acid, 0.04% w/v L-ascorbic acid, 0.04% w/v D-(+)-glucose, 0.04% w/v folic acid, 0.04% w/v D-pantothenic acid hemicalcium salt
G8	0.16% w/v riboflavin, 0.16% w/v L-α-phosphatidylcholine, 0.16% w/v D-(+)-raffinose pentahydrate, 0.16% w/v D-(−)-3-phosphoglyceric acid disodium salt, 0.16% w/v γ-aminobutyric acid, 0.16% w/v thionicotinamide adenine dinucleotide
G9	0.16% w/v 2'-deoxyguanosine 5'-monophosphate sodium salt hydrate, 0.16% w/v cholesterol, 0.16% w/v thymine, 0.16% w/v oxamic acid, 0.16% w/v D-glucosamine 6-phosphate sodium salt, 0.16% w/v guanosine 3',5'-cyclic monophosphate sodium salt
G10	0.20% w/v thiamine monophosphate chloride dihydrate, 0.20% w/v creatine phosphate disodium salt tetrahydrate, 0.20% w/v Nα-benzoyl-L-arginine ethyl ester hydrochloride, 0.20% w/v tetracycline hydrochloride, 0.20% w/v succinic acid disodium salt hexahydrate
G11	0.20% w/v uridine 5'-triphosphate trisodium salt hydrate, 0.20% w/v kanamycin monosulfate, 0.20% w/v maltotriose hydrate, 0.20% w/v leupeptin hydrochloride, 0.20% w/v sodium glyoxylate monohydrate
G12	0.16% w/v 5-phospho-D-ribose 1-diphosphate pentasodium salt, 0.16% w/v maleic acid, 0.16% w/v N-acetyl-D-glucosamine, 0.16% w/v L-(−)-malic acid sodium salt, 0.16% w/v γ-aminobutyric acid, 0.16% w/v acetylcholine chloride
H1	0.05% w/v glycine, 0.05% w/v L-(-)-threonine, 0.05% w/v L-(+)-lysine, 0.05% w/v L-alanine, 0.05% w/v L-arginine, 0.05% w/v L-asparagine monohydrate, 0.05% w/v L-aspartic acid, 0.05% w/v L-glutamic acid, 0.05% w/v L-glutamine, 0.05% w/v L-histidine, 0.05% w/v L-isoleucine, 0.05% w/v L-leucine, 0.05% w/v L-methionine, 0.05% w/v L-phenylalanine, 0.05% w/v L-proline, 0.05% w/v L-serine, 0.05% w/v L-tryptophan, 0.05% w/v L-tyrosine, 0.05% w/v L-valine
H2	0.14% w/v Gly-Gly-Gly-Gly-Gly, 0.14% w/v Leu-Gly-Gly, 0.14% w/v Ala-Ala-Ala, 0.14% w/v Met-Ala-Ser, 0.14% w/v Val-Ser, 0.14% w/v Ala-Leu, 0.14% w/v Z-Val-Phe
H3	0.16% w/v adenosine 5'-triphosphate disodium salt hydrate, 0.16% w/v pyridoxal 5-phosphate monohydrate, 0.16% w/v creatine monohydrate, 0.16% w/v thymine, 0.16% w/v L-(−)-malic acid sodium salt, 0.16% w/v spermine
H4	0.16% w/v adenosine 5'-triphosphate disodium salt hydrate, 0.16% w/v riboflavin 5'-monophosphate sodium salt dihydrate, 0.16% w/v DL-α-lipoic acid, 0.16% w/v acetylsalicylic acid, 0.16% w/v L-ornithine hydrochloride, 0.16% w/v adenosine
H5	0.16% w/v creatine phosphate disodium salt tetrahydrate, 0.16% w/v β-nicotinamide adenine dinucleotide phosphate sodium salt hydrate, 0.16% w/v guanosine 5'-triphosphate sodium salt hydrate, 0.16% w/v uridine, 0.16% w/v D-pantothenic acid hemicalcium salt, 0.16% w/v N-acetyl-D-glucosamine
H6	0.14% w/v thiamine monophosphate chloride dihydrate, 0.14% w/v β-nicotinamide adenine dinucleotide hydrate, 0.14% w/v guanosine, 0.14% w/v thymidine, 0.14% w/v D-(+)-galactose, 0.14% w/v biotin, 0.14% w/v D-fructose 1,6-bisphosphate trisodium salt octahydrate
H7	0.16% w/v 2'-deoxyguanosine 5'-monophosphate sodium salt hydrate, 0.16% w/v thymidine 5'-triphosphate sodium salt, 0.16% w/v 4-aminobenzoic acid, 0.16% w/v cytosine, 0.16% w/v benzamidine hydrochloride, 0.16% w/v β-nicotinamide adenine dinucleotide, reduced dipotassium salt

H8	0.16% w/v β-nicotinamide adenine dinucleotide, reduced dipotassium salt, 0.16% w/v guanidine hydrochloride, 0.16% w/v cystamine dihydrochloride, 0.16% w/v aspartame, 0.16% w/v caffeine, 0.16% w/v cytidine 5'-monophosphate disodium salt
H9	0.20% w/v uridine 5'-triphosphate trisodium salt hydrate, 0.20% w/v adenosine 5'-monophosphate sodium salt, 0.20% w/v N-acetylneuraminic acid, 0.20% w/v L-tyrosine, 0.20% w/v L-thyroxine
H10	0.16% w/v pyridoxamine dihydrochloride, 0.16% w/v guanosine 5'-diphosphate sodium salt, 0.16% w/v cytidine 5'-triphosphate disodium salt, 0.16% w/v D-glucose 6-phosphate sodium salt, 0.16% w/v O-phospho-L-tyrosine, 0.16% w/v spermidine
H11	0.16% w/v L-carnitine hydrochloride, 0.16% w/v β-nicotinamide adenine dinucleotide phosphate tetrasodium salt, 0.16% w/v 2'-deoxyadenosine 5'-triphosphate disodium salt, 0.16% w/v choline base solution, 0.16% w/v α-D-glucose 1-phosphate disodium salt hydrate, 0.16% w/v riboflavin
H12	0.16% w/v uridine 5'-monophosphate disodium salt hydrate, 0.16% w/v D-(+)-maltose monohydrate, 0.16% w/v D-sorbitol, 0.16% w/v O-phospho-L-serine, 0.16% w/v benzidine, 0.16% w/v sodium phenyl phosphate dibasic dihydrate