# Characterisation of membrane transport proteins of NCS1 and PACE families using biochemical and biophysical techniques

IRSHAD AHMAD

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Biomedical Sciences

July 2017

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

**Chapter 3 – Section 3.5**

The work in Section 3.5 of this thesis is solely the candidate's own work and part of it (Figure 3.6) has been submitted for publication:

Ekaterina Ivanova, Scott M. Jackson, Antonio N. Calabrese, David Sharples, **Irshad Ahmad,** Oliver Beckstein, Sheena E. Radford, Alexander D. Cameron, Alison E. Ashcroft, Peter J.F. Henderson (2017). Kinetic basis of the alternating access mechanism of the Mhp1 membrane transport protein determined by mass spectrometry, fluorimetry, and mutagenesis.

## Acknowledgements

I am highly grateful and wish to express my deepest sense of gratitude to my honourable supervisor Professor Peter Henderson for introducing me to the inspiring world of membrane transport proteins. His continuous encouragement, great patience, critical inputs and golden advice enabled me to develop a better understanding of membrane transport proteins and their analysis.

I would like to thank my co-supervisor Professor Alison Ashcroft for her help and support. I also wish to thank Dr Antonio Calabrese and Dr Vincent Postis for their continuous generous help. I am thankful to past and present members of Henderson group for their valuable discussion and practical help.

I offer special thanks to Dr Karl Hassan for providing the strains of PACE family and for help and advice in the laboratory. I would also like to mention Dr Anna Polyakova, who taught me a lot different techniques in my first weeks of experiments in Leeds.

Thanks to fermentation technician Mr. David Sharples and CD facility manager Mr. Ghulam Nasir Khan who were always there for any kind of help– science and life.

Last, but certainly not least, a special thanks goes to my sincere wife, Nadia Irshad, who has been a constant source of support in many different ways throughout these years of study. I believe that without her constant support I could not have finished this research.

I also would like to thank my family especially my mother, and elder brother Dr Mushtaq and younger brother Dr Naeem for their continuous love and support that gave me the strength, courage and confidence to complete this project.

I also express my gratitude to my loving son Ayaan Nawaz.

I would also like to acknowledge the Australian NHMRC and Leverhulme Trust for funding this research.

# Abstract

The aim of this study was to work towards the structural and functional characterisation of NCS1 and PACE family transporters. The genes encoding proteins of these two families were cloned into the plasmid pTTQ18 under the control of the *tac* promoter with the fusion of a hexa-histidine tag at the C-terminus of each protein to facilitate overexpression and purification. Significant expression levels of these proteins were achieved after induction with isopropyl-β-D-thiogalactoside. Efficient solubilisation of proteins was obtained with 1% detergent n-dodecyl-β-D-maltoside from inner membranes of *E. coli* BL21(DE3) cells. For purification an immobilised metal affinity chromatography procedure was undertaken by exploiting the engineered C-terminal $His_6$-tag and the purity of proteins was up to 89%. The protein yields obtained ranged from 0.9 mg/litre to 1.6 mg/litre.

After purification far UV Circular Dichroism (CD) was used to confirm that the proteins were correctly folded and had retained their alpha helical secondary structure. The thermal stability of these proteins was also examined by far UV CD spectroscopy. Variable melting points were observed with different proteins; however all proteins showed melting temperatures above 30 ºC. In view of the CD measurements it was reasonable to conclude that all proteins are quite stable for performing biophysical assays using a temperature range of 18-25 ºC. After thermal denaturation the ability to refold was not observed with any protein.

Site directed mutagenesis was employed to generate three mutants of the conserved residue (Asp229) in the prototypical NCS1 family protein Mhp1 to a variety of amino acids (Asp229Glu, Asp229Asn and Asp229Ala). These mutants revealed the importance of a carboxyl group at this position because the Asn and Ala mutants showed significantly reduced ligand binding affinity measured by spectrophotofluorimetry. Substrates were identified for two novel NCS1 family proteins, AAN69889 and VPA1242, which transported allantoin and cytosine, respectively, as determined by using radioactive uptake assays in energised whole cells of *E. coli.*

The PACE family proteins are multidrug efflux transporters. Out of 24 PACE family proteins screened, seven were expressed at high levels. Spectrophotofluorimetry measurements identified acriflavine and chlorhexidine as substrates of three of these proteins, Fbal_3166, PSPTO_3587 and AceI, whereas $^{3}$H-spermidine efflux was observed for Fbal_3166, PSPTO_3587, AceI, Tmarg_opt and PFL_4558 using uptake assays in energised whole cells of *E. coli*.

Taken all together the work in this thesis lays the foundation for a range of future analyses on the proteins of both the NCS1 and PACE families.

# List of abbreviations

| | |
|---|---|
| APS | Ammonium persulphate |
| ATP | Adenosine 5’ triphosphate |
| BH | Benzylhydantoin |
| BME | β-Mercaptoethanol |
| BSA | Bovine serum albumin |
| CD | Circular dichroism |
| DDM | n-dodecyl β-D-maltoside |
| $DH_2O$ | Deionised water |
| DMSO | Dimethyl sulphoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | 2-Deoxy nucleoside-5-triposphate |
| EB | Elution buffer |
| EDTA | Ethylene diamine tetraacetic acid |
| HAWP | Hydrophobic acetate white plain (filter) |
| HRP | Horse radish peroxidase |
| IMAC | Immobilized metal affinity chromatography |
| LB | Luria broth culture medium |
| NCS1 | Nucleobase cation symporter-1 |
| NTA | Nitrilo-triacetic acid |
| PACE | proteobacterial antimicrobial compound efflux |
| PCR | Polymerase chain reaction |
| PDB | Protein Data Bank |
| SDS | Sodium dodecyl sulphate |
| TCA | Trichloroacetic acid |

# CONTENTS

# Index of Figures, Tables and Appendices

## Index of Figures

## Index of Tables

## Appendices

# Chapter 1

# General introduction

# 1 Introduction

A semi-permeable cell membrane that separates intracellular compartments from the external environment is found in all living prokaryotic or eukaryotic cells. Membranes not only maintain the physical integrity of the cell but also provide suitable environment for the proteins that effect and control the movement of particles into and out of the cell (Alberts et al., 2002*)*. The movement of particles, including organic and inorganic solutes and ions, is important for acquiring essential nutrients, generation of electrochemical gradients, energy transduction and expulsion of waste or harmful compounds including antibiotics (Singer and Nicolson, 1972; Engel and Gaub, 2008; Dobson et al., 2015). Most functions of the cell membrane are carried out by proteins that are highly abundant in all biological systems. Membrane proteins are estimated to represent approximately 20-30% of all the genes in most sequenced genomes and up to 16% of these are transport proteins (Ren and Paulsen, 2005; Gao and Cross, 2006). Membrane proteins are highly important in human health and disease. Dysfunction in membrane proteins has been associated with numerous human diseases including heart failure, stroke, cystic fibrosis, diabetes, obesity, cancer, depression and many others (Morais et al., 2014). Membrane proteins also provide one of the mechanisms that bacteria have evolved for resistance to antibiotics in the form of drug efflux proteins (Alvarez-Ortega et al., 2013; Martins et al., 2013; Blanco et al., 2016). These proteins also constitute more than 50% of the molecular targets for current drugs (Overington et al., 2006; Arinaminpathy et al., 2009; Rawson et al., 2016) and remain the principal targets for drug discovery (Wirmer-Bartoschek and Bartoschek, 2012; Yin and Flynn, 2016; Renaud et al., 2016).

Despite the abundance and importance of membrane proteins, a relatively small number of high resolution structures of membrane proteins have been determined since they constitute less than 3% of structures deposited in the Protein Data Bank (PDB) (Gautier, 2014; Rawson et al., 2016). A number of challenges are involved in the structural and functional characterisation of membrane proteins including cloning, expression, purification and a range of application of molecular, biochemical, biophysical and computational techniques. It is usually more difficult to overcome these challenges and apply experimental techniques to eukaryotic membrane proteins. Therefore, homologous proteins from bacteria are often used as

model systems. Bacterial membrane proteins are themselves also important to study as drug targets, especially for the development of new antibiotics.

## 1.1 Bacterial membranes

Bacteria are differentiated into two major groups: Gram positive and Gram negative, based on the structural differences of the cell envelope. Gram-positive bacteria have a thick layer of peptidoglycan but lack an outer membrane whereas Gram-negative bacteria have a thin layer of peptidoglycan but are surrounded by an additional outer membrane (Figure 1.1). Due to this difference, the cell can be distinguished by the use of a technique known as the Gram stain, discovered by Danish scientist Hans Christaina Gram in 1884.

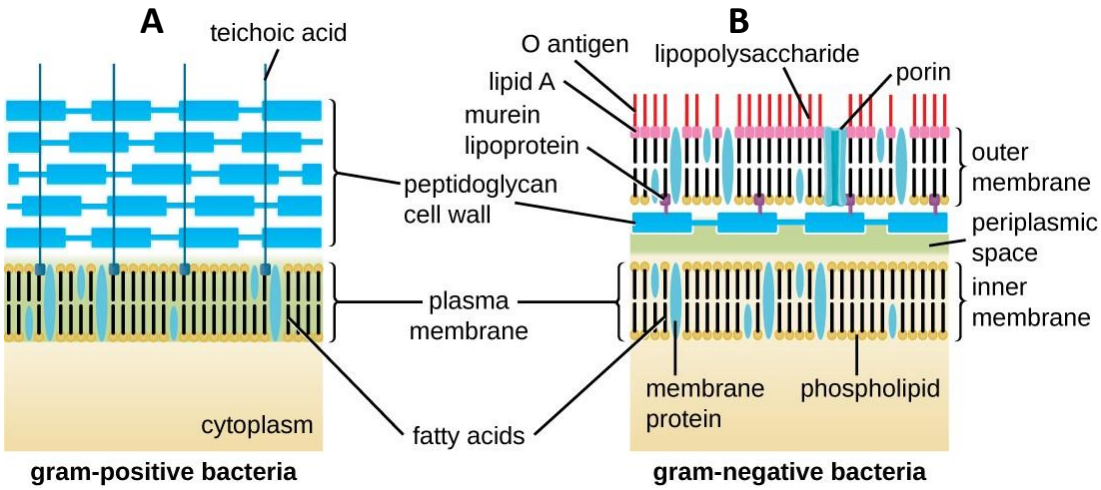



**Figure 1.1 A comparison of the cell walls of Gram-negative and Gram-positive bacteria.** Cell walls of Gram-positive bacteria (A) have a thick layer of peptidoglycan but lack an outer membrane whereas cell walls of Gram-negative bacteria (B) contain a thin layer of peptidoglycan, the inner membrane and an outer membrane. This image has been adopted from http://Franciscosp/Wikimedia Commons)

The Gram stain is the key step in the preliminary identification of a bacterial organism. Gram positive bacteria retain the crystal violet dye within the thick peptidoglycan layer, resulting in a purple colour whereas stain from Gram negative bacteria is easily removed and results in a pink colour. The region in Gram-negative bacteria between the inner and outer membrane is the periplasm where a single thin

layer of peptidoglycan resides. The outer membrane in Gram negative bacteria serves as an additional layer protecting the cell from harmful agents such as antibodies, proteases, antibiotics, toxins, osmotic pressure and phages (Cowan et al., 1992).

## 1.2 Membrane structure

The basic structure of a membrane is primarily comprised of a lipid bilayer and proteins (Figure 1.2).

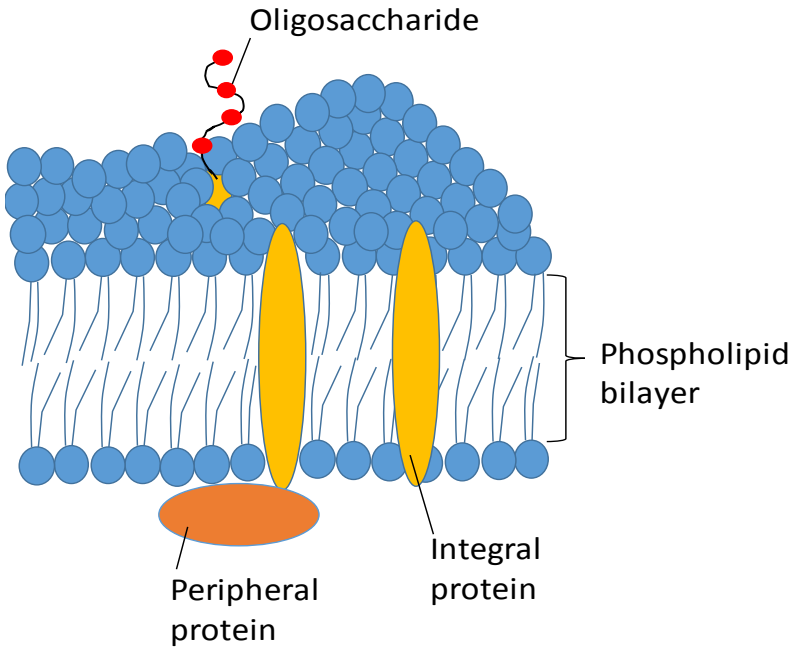



**Figure 1.2 A simplified model of cell membrane lipids bilayer and proteins.** A cartoon of a lipid bilayer crowded with proteins. Integral proteins span the lipid bilayer and have portions of the protein sticking out on both faces of the membrane. Peripheral proteins are attached to the membrane indirectly, via protein-protein interactions.

### 1.2.1 Membrane lipids

Lipids are amphipathic molecules that are universal components of all cell membranes. As biological molecules, lipids are largely composed of hydrocarbons having low water solubility and high nonpolar solvents solubility. The bilayer acts as an effective barrier to more polar molecules. Biological membranes generally consist of three different kinds of lipids: glycerophospholipids, glycolipids and cholesterol. Usually, glycerophospholipids, also known as phospholipids, are ubiquitous in

cellular organisms and are composed of two fatty acid chains, glycerol, and a phosphate group. Glycerol is an organic compound with three-carbon atoms, serving as the backbone of these lipids. The first and second carbons of the glycerol backbone are occupied by fatty acids whereas phosphate group is attached to the third carbon atom. Characteristically, phospholipids are amphipathic in nature by having a hydrophilic (water-loving) phosphate end and hydrophobic (water-fearing) lipid end. Because of the amphipathic nature, phospholipids spontaneously self-assemble into a bilayer, with hydrophobic tails away from the aqueous environment facing inward and the hydrophilic heads towards the aqueous solution facing outside (Figure 1.3). Phospholipids are not only the key structural components of the cell membrane but are also involved in metabolism and cell signalling (Berridge and Irvine., 1989).

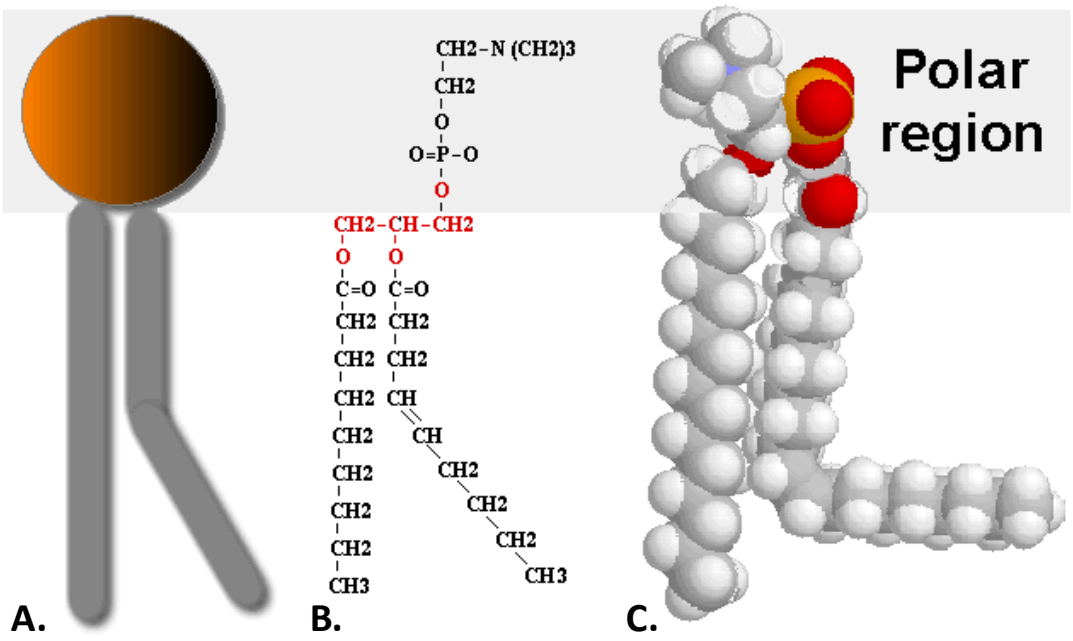



**Figure 1.3 A typical structure of glycerophospholipid in the membrane bilayer.** This example is phosphatidylcholine: (A) as an image (B) by structural formula; (C) as a space-filling model. A kink into the fatty acid chain is introduced by cis configuration double bond. This figure is adopted from https://biofoundations.org

In addition to glycerophospholipds, two other major classes of lipids found in a typical biomembrane are known as sphingolipids and sterols. Sphingolipids contain sphingosine rather than a glycerol backbone and are not generally found in bacteria and fungi except some anaerobic bacteria (Olsen and Jantzen, 2001). The conformation and the charge distribution of sphingolipids are quite similar to glycerophospholipids but they are chemically different (Brown and London, 1998; Voet et al., 2008). Another class of lipids found in biological membranes are sterols which are present in the membranes of most eukaryotic cells. They play an important role to maintain membranes in a fluid state, which is suitable for proper functions. As sphingolipids and sterols are not commonly found in bacterial membranes they will not be discussed further.

Usually, the lipid composition of the inner membrane in *E. coli* is 5–10% cardiolipin, 20–25% phosphatidylglycerol and 70–80% phosphatidylethanolamine. (Figure 1.4). There may be some variations due to strains or growth conditions (Dowhan, 1997; Wikstrom et al., 2009). Phospholipids are the main components of the bacterial inner membrane and usually contain saturated or unsatured fatty acids hydrocarbon chains of either 16 or 18 carbon atoms. In *E. coli* membranes the most common fatty acids found are palmitic acid, palmitoleic acid and cis-vaccenic acids (Kadner et al., 1996). Moreover, almost all Gram-negative bacteria possess an additional layer of lipopolysaccharides that make up the outer monolayer of the outer membrane whereas the phospholipids make up the inner monolayer (Raetz, 1996; Nikaido, 2003).

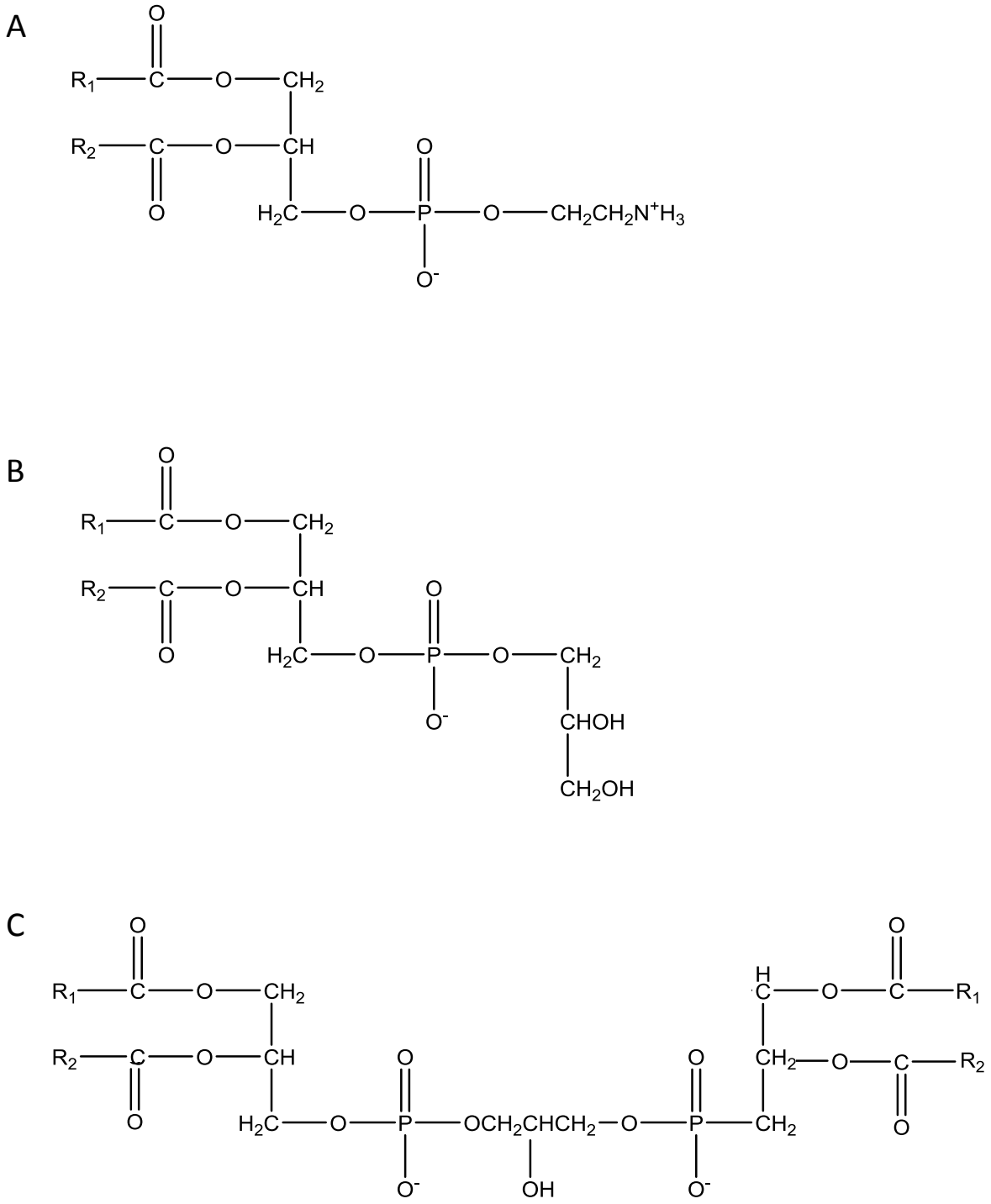



**Figure 1.4 Chemical structures of major phosphoglyceride components of the bacterial inner membrane.** Three different kinds of lipids are shown: (A) phosphatidylethanolamine (B) phosphatidylglycerol (C) diphosphatidylglycerol (cardiolipin). Saturated and unsaturated fatty acids are represented by R1 and R2.

### 1.2.2 Membrane proteins

Generally, any protein that is located or associated with a biological membrane is known as a membrane protein. Membrane proteins are diverse structurally as well as functionally. The various specialised functions of a membrane are largely determined by the included membrane proteins. The major constituents of the lipid bilayer are membrane proteins that are roughly equal in *E. coli* to the phospholipid by weight (Kadner, 1996). For membrane proteins, a suitable environment is provided by membrane lipids to perform a range of functions including roles as transporters, receptors, channels, signal transduction, energy transduction, and enzymes (Müller et al., 2008; Padan, 2002; Cournia et al., 2015). The genomic sequencing projects have revealed that around 33% of entire proteins in a cell are found in membranes and approximately 33% of these are transport proteins (Ren et al., 2007; Pielak and Tian, 2012; Aboulwafa and Saier, 2013). These proteins are categorised into two main classes on the basis of their attachment to the membrane as peripheral (extrinsic) membrane proteins and integral (intrinsic) membrane proteins.

**1.2.2.1 Peripheral (extrinsic) membrane proteins.** Usually these proteins are attached very weakly to the peripheral regions of the lipid bilayer of biological membrane or to the integral membrane proteins (Figure 1.2). Changing the pH or increasing the salt concentration can easily dissociate these proteins from the membrane. Peripheral membrane proteins are water soluble proteins that have roles in the electron transport chain, function as peripheral enzymes, and can regulate ion channels and membrane receptors (Johnson and Cornell, 1999; Takida and Wedegaertner, 2004; Cafiso, 2005).

**1.2.2.2 Integral membrane proteins.** These proteins have strong interactions with the membrane (Figure 1.2) compared to peripheral membrane proteins and can only be dissociated by disrupting the membrane using detergents or nonpolar solvents. Based on their relationship with the bilayer, these proteins are further classified into integral monotopic proteins and integral polytopic proteins. Integral monotopic proteins do not necessarily span the whole way across the membrane and are permanently attached from one side to the cell membrane. Integral polytopic proteins

span across the membrane more than once and are also known as transmembrane proteins. The secondary structures of membrane proteins are either alpha helices or beta sheets (Xiong, 2006). Usually, the alpha-helical proteins are present in the inner membranes but sometimes in the outer membranes of bacterial cells and having righthand-coiled or spiral conformation (helix). The majority of the transmembrane proteins are alpha-helical. Beta barrels are cylinder-like channel proteins, mostly found only in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts. Sometimes these proteins are found in the cell walls of a few Gram-positive bacteria

### 1.3 Prokaryotic membrane transport mechanisms

Transport processes are enormously important for all organisms in order to remain alive. Membrane proteins are around one-third of all proteins present in a cell while about one-third of these are transport proteins reflecting the importance of membrane transport. Some specialized mammalian cells allocate up to two–thirds of their total metabolic energy utilisation to membrane transport processes (Albert et al., 2008; Aboulwafa and Saier, 2013). All living cells are surrounded by a phospholipid bilayer constituting a semi-permeable membrane, through which selectively small molecules can diffuse at a slow rate. This membrane acts as a hydrophobic barrier between the interior and exterior of the cell. So to acquire the essential nutrients and excrete waste products, which are water soluble, membrane transport proteins play a pivotal role to translocate ions and metabolites across the cell membranes.

Various transport processes have been recognised to be similar in different types of microrganisms and in higher organisms including man (Henderson, 1998; Aboulwafa and Saier, 2013). Membrane proteins have crucial roles in a number of biological activities. They are largely responsible for transport of essential substrates across the membrane, cell bioenergetics, signal transduction, nutrient uptake, metabolite excretion and drug efflux (Padan et al., 2009; Natale et al., 2008; Rettner and Saier, 2010; LeVine et al., 2016). There are at least five distinct mechanisms of transport in bacteria.

### 1.3.1 Simple diffusion

Diffusion is the simplest mechanism of passing substances across the membrane from areas of higher concentration to areas of lower concentration. Simple diffusion process always follows a concentration gradient as long as there is no restriction of, for example, size or charge of the molecule (Figure 1.5). Neutral, lipophilic and typically small molecules for example oxygen, carbon dioxide, glycerol and ammonia are able to pass freely across the membranes (Henderson, 2013). For instance, during cellular respiration, oxygen molecules are diffusing into the cell because oxygen concentration is always higher outside the cell. Similarly, carbon dioxide concentration is always higher inside the cell so it diffuses out. The diffusion rate across the membrane of a particular substance is directly related to the concentration gradient, surface area of the membrane and solubility in the lipid bilayer and is inversely proportional to the molecular weight of the substance and thickness of the membrane.

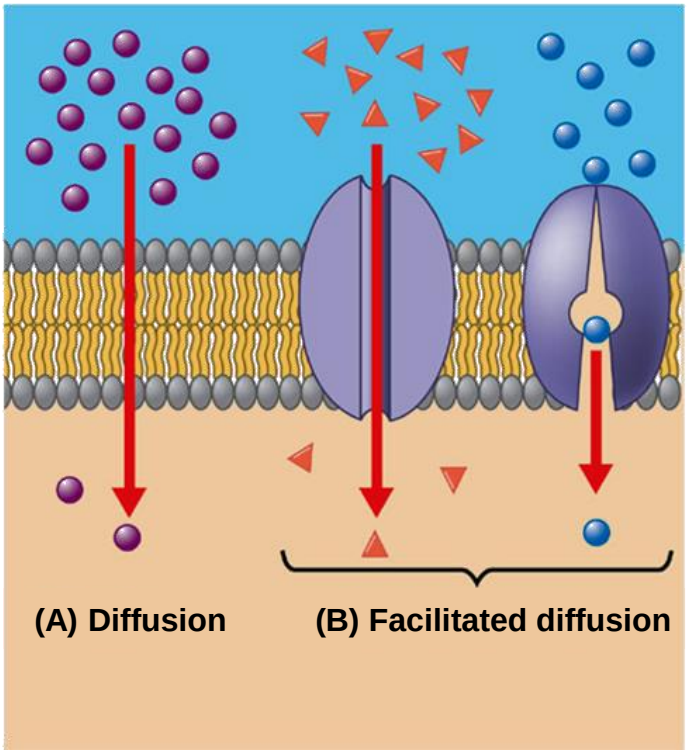



**Figure 1.5 Diagram showing simple and facilitated diffusion**. Small molecules like $O_2$ and $CO_2$ can simply diffuse across a membrane without any help (A), while larger molecules like glucose and amino acids cannot diffuse directly through the phospholipid bilayer and move through carrier proteins or channel proteins (B).

### 1.3.2 Facilitated diffusion

Many molecules cannot enter or exit the cell because of the hydrophobic nature of the fatty acid tails of the phospholipids, their large size or their charge. Facilitated diffusion does not require energy and the movements of molecules are always facilitated by special transport proteins implanted in the cell membrane (Figure 1.5). The direction of transport in this type of diffusion is always down a concentration gradient across the membrane. Glucose is a water-soluble six-carbon sugar, difficult to be transported across the cell membrane through simple diffusion (Thorens, 1993). Therefore, such molecules require specific transport proteins to diffuse across the cell membranes down the concentration gradient (Carruthers, 1990).

### 1.3.3 Group translocation

In prokaryotes one of the distinct methods of transport mechanism is group translocation, also known as the phosphotransferase system, which was first discovered by Kundig et al., (1964) in *E. coli* for the transport of sugars across the membrane. A variety of sugars including glucose, fructose, mannose and xylitol are transported, using this type of transport mechanism. The best understood group translocation system of *E. coli* is the phosphoenolpyruvate-dependent phosphotransferase system (PTS). From phosphoenolpyruvate a high energy phosphate group is transferred to glucose by a series of enzymes. The final enzyme phosphorylates the glucose as glucose-6-phospate and transports it across the membrane; the cytoplasmic membrane becomes impermeable to sugar phosphates, generating a concentration gradient that allows further import of the sugars. The phosphoenolpyruvate (PEP) energised this type of transport, where the phosphoryl moiety is transffered from PEP and transferred to numerous cytoplasmic proteins before attachment to the sugar (Erni, 1992).

### 1.3.4 Active transport

Movement of substances across the membranes against a concentration gradient is recognised as active transport. Energy is required for such types of transport to overcome a concentration gradient, this also involves specific membrane proteins that directly may use source of energy such as ATP. On the basis of energy usage,

active transport systems are broadly categorised into primary or secondary active transport.

**1.3.4.1 Primary active transport.** Such transport directly uses energy in the form of ATP to pass materials across a membrane against their concentration gradient (Figure 1.6). All groups of this type of transport contain one or more ATP binding sites. The human sodium-potassium pump ($Na^+/K^+$ - ATPase), is the best example of this type of transport that helps in maintaining the high level of extracellular sodium ions and intracellular potassium ions. This protein uses the energy released from hydrolysis of one adenosine triphosphate (ATP) molecule to pump three sodium ions to the outside of the cell and two potassium ions into the cell, against their concentration gradients.

According to the TCDB, there are five classes of primary active transporters. These are P-P-bond hydrolysis-driven transporters, decarboxylation-driven transporters, methyltransfer-driven transporters, oxidoreduction-driven transporters and light absorption-driven transporters.

P-P-bond hydrolysis-driven transporters drive the active transport utilising ATP or hydrolyse the diphosphate bond of inorganic pyrophosphate or another nucleoside triphosphate. The substrate is not phosphorylated whereas the transporter may or may not be transiently phosphorylated. Decarboxylation-driven transporters are currently thought to be restricted to prokaryotes and drive solute (e.g. ion) uptake or extrusion by decarboxylation of a cytoplasmic substrate. Methyltransfer-driven transporters are thought to be restricted to archaea. A single characterised multisubunit protein family falls into this subclass, the $Na^+$-transporting methyltetrahydromethanopterin: coenzyme M methyltransferase. Oxidoreduction-driven transporters drive transport of a solute (e.g. an ion) energised by the exothermic flow of electrons from a reduced substrate to an oxidised substrate. These transporters are found in prokaryotes, chloroplasts, mitochondria and other eukaryotic organelles. Facultative anaerobes largely dependent on fermentation, have been shown to possess functional electron transfer chains. Light absorption-driven transporters utilise light energy to drive transport of a solute (e.g. an ion).

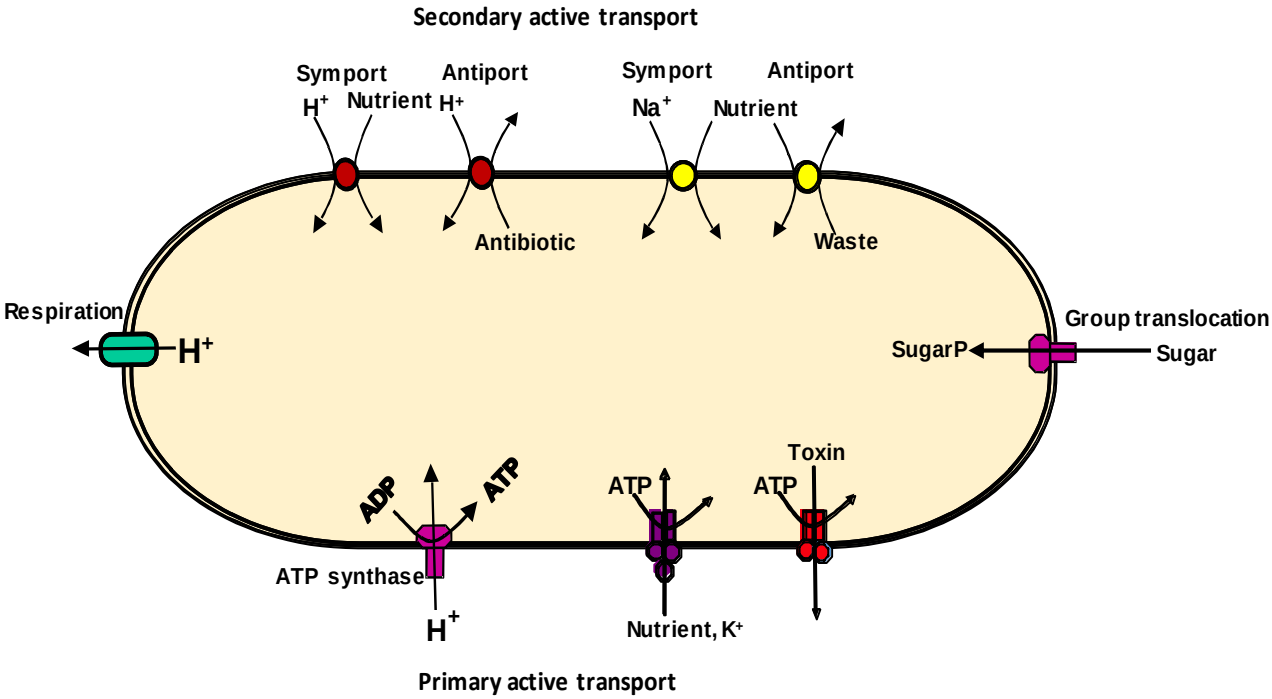



**Figure 1.6 Different types of transport mechanisms in bacteria**. The oval shape represents the cytoplasmic membrane of a bacterium. A transmembrane electrochemical gradient of protons is generated by respiration or photosynthesis, shown on the left hand side. Two symporters and two antiporters, secondary active transport systems are shown along the top. Primary active transporters are shown along the bottom. A phosphotransferase system is shown on the right hand side. Diagram is modified from Saidijam et al., (2005).

**1.3.4.2 Secondary active transport.** This is also recognised as co-transport or coupled transport, it is a type of active transport across a membrane that does not directly utilise ATP, instead it relies on the electrochemical potential difference and couples the movement of an ion (typically $Na^+$, $H^+$ or $K^+$) down its electrochemical gradient to facilitate the uphill movement of another molecule or ion against a concentration gradient (Figure 1.6). Secondary active transporters are categorised into symporters or antiporters depending on the movement of substances across the membrane. The protein is called a symporter, when two different molecules move in the same direction across the membrane, while if they move in opposite directions, the protein is known as an antiporter. For example, the concentration of extracellular sodium is ~145mM and the intracellular sodium concentration is about 15mM, maintained by the $Na^+/K^+$/ATPase. The $Na^+$/glucose symporter (SGLT1), found in

the proximal tubules of kidney, transports two sodium ions and one glucose molecule across the membrane into the cell (Harada and Inagaki, 2012). Similarly, the $Na^{+}/Ca_{2}^{+}$ antiporter (NCX), found in cardiac muscle cells and other cells of the body, transports three sodium ions into the cell in exchange for one calcium ion transported out of the cell (Palty et al., 2012).

## 1.4 General overview of protein structure

A firm knowledge of protein structure is essential in order to understand, modify and ultimately exploit proteins for useful desirable purposes. Proteins are complex macromolecules consisting of amino acids joined together by peptide bonds. Twenty different amino acids are known that constitute the majority of proteins. Each amino acid comprises a central alpha carbon (Cα) to which the carboxyl group ($-COO^{-}$) and an amino group ($-NH^{3+}$) are covalently linked. Additionally, a hydrogen and a variable side chain R are linked to the same alpha carbon. A typical amino acid structure is shown in Figure 1.7

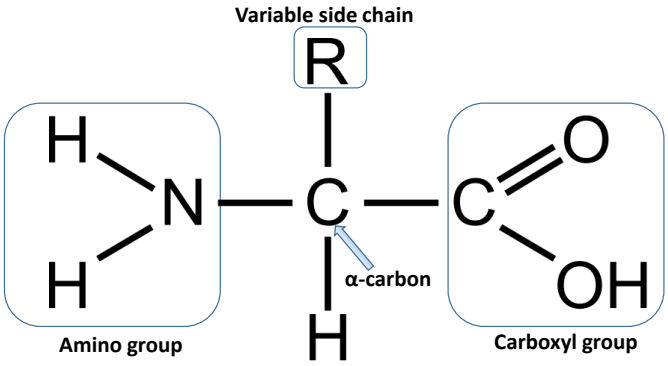



**Figure 1.7 General amino acid structure.** Alpha carbon of all amino acids is bonded to a hydrogen atom, amino group ($NH_2$) and carboxyl group (COOH). The variable side chain makes amino acids different from each other.

Each amino acid has a different R group that differentiates one amino acid from another and dictates its fundamental distinctive properties. All twenty amino acids have different shapes, size, ionic charges and relative hydrophobicity (Figure 1.8). Amino acids are the structural blocks of proteins which are built so precisely that the change of even one amino acid can sometimes disrupt the structure of the whole

molecule and lose the function. The amino acid sequence determines the three-dimensional structure of a protein and the three-dimensional shape of a protein molecule is so important to its function that four levels of structure are used to describe a protein.

- **Primary protein structure** is the simplest level of protein structure and is simply the linear sequence of amino acids in a polypeptide chain. These amino acids are held together by peptide bonds, which are formed by the carboxylic group of one amino acid reacting with the amino group of adjust amino acid resulting in a condensation reaction, removing water.

- **Secondary protein structure** refers to the polypeptide chain twisted/arranged into characteristic helical or pleated sheet. The α-helices and the β-pleated sheets are the most common types of secondary structures. These structures are stabilized by hydrogen bonding, which forms between the amino (N-H) of one amino acid and the carbonyl (C=O) of another.

- **Tertiary protein structure** refers to the overall three-dimensional structure of an entire protein molecule. The polypeptide chain of the protein bends and folds in such a manner as to achieve maximum stability. This is partly due to a large number of non-covalent interactions, especially between the side chains of the amino acids, but also to effects by which interactions between water and hydrophobic R groups are minimised.

- **Quaternary protein structure** comprises the non-covalent interactions that bind multiple polypeptides into a single, larger complex. Some proteins are made up of a single polypeptide chain and have three levels of structure while some proteins are made up of multiple polypeptide chains, often known as protein subunits. When these protein subunits interact with each other and arrange themselves to form a larger complex, it is known as quaternary protein structure.

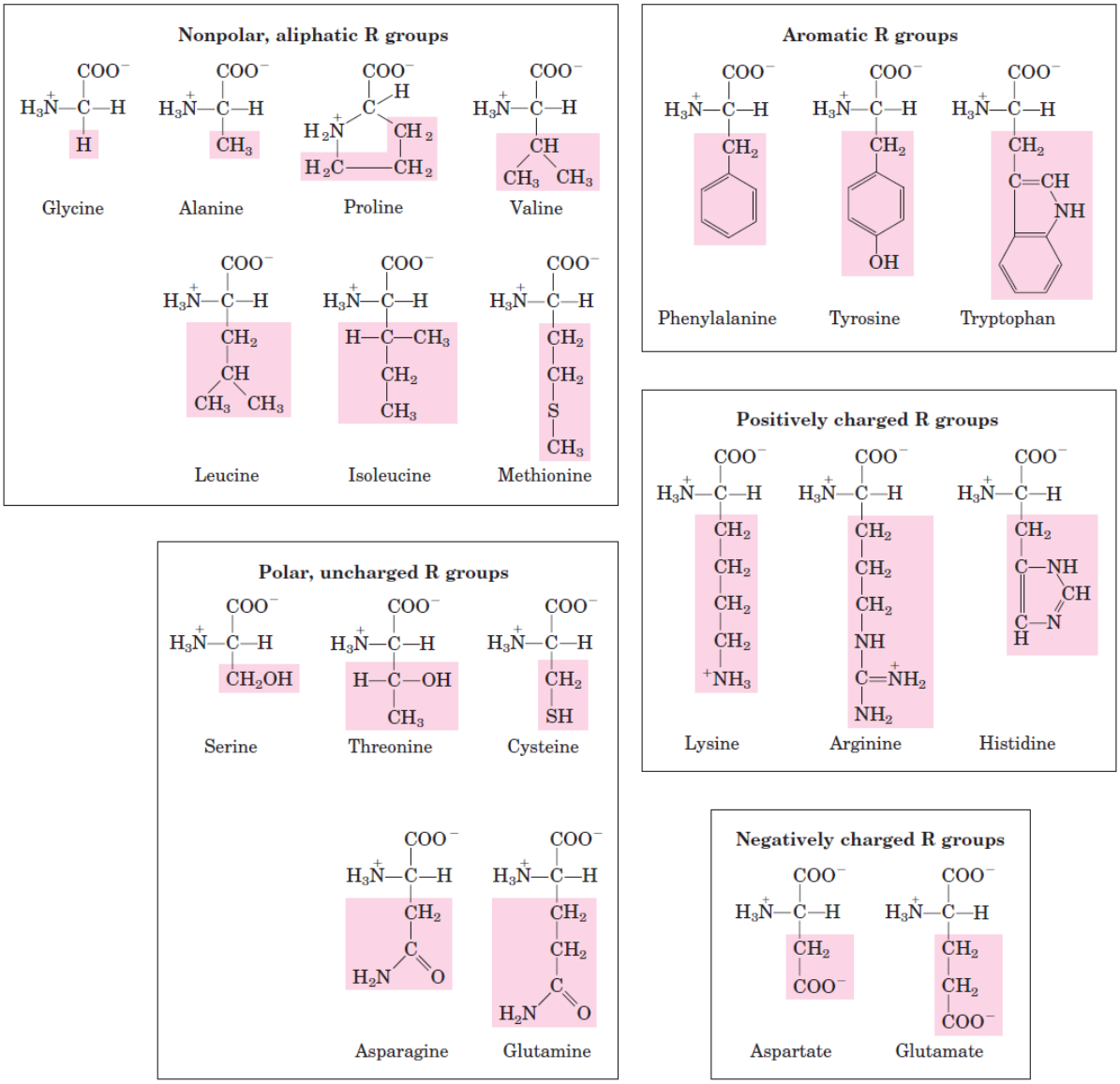



**Figure 1.8 Structure of the 20 commonly occurring amino acids.** The amino acids are grouped according to their side chain properties with the side chains highlighted in pink. Diagram was taken from (Nelson and Cox, 2008).

## 1.5 Current challenges and approaches for studies on membrane proteins

Structural elucidation of membrane proteins started much later than for soluble proteins. This delay was mainly due to their amphipathic nature, limited availability of crystallization studies and many other factors such as natural low abundance and instability. Despite much advancement in studying membrane proteins, there is still no universal and simple protocol for membrane proteins extraction from their native

membrane environment, while maintaining their normal structures and functions. There are a number of reasons that make membrane proteins extremely difficult to study.

a) The hydrophobic nature of membrane proteins is one of the main hurdles for cells to express them efficiently.
b) Lipid bilayer environment stabilise the membrane proteins and may have specific lipid requirements.
c) Membrane proteins are typically implanted into the membrane during or after expression. Due to the limited volume of membrane, organisms are generally very sensitive to any changes in membrane properties.
d) Recombinant proteins do not usually express well in heterologous expression systems.
e) During the production of heterologous proteins in bacterial hosts, membrane protein overexpression may lead to the formation of inclusion bodies and adopts non-native conformations. In order to acquire functionally active protein, appropriate strategies for solubilisation and refolding is required (Laage and Langosch, 2001).
f) Membrane proteins are usually less stable and partially functional outside their native environment and exhibit a tendency to aggregate (Iwata, 2003)
g) Detergents are needed for purifications of membrane proteins and are generally more difficult and challenging to achieve stable and functional proteins.

The standard flow chart of studying membrane proteins from targeting gene of interest to the characterisation of proteins is shown in Figure 1.9.

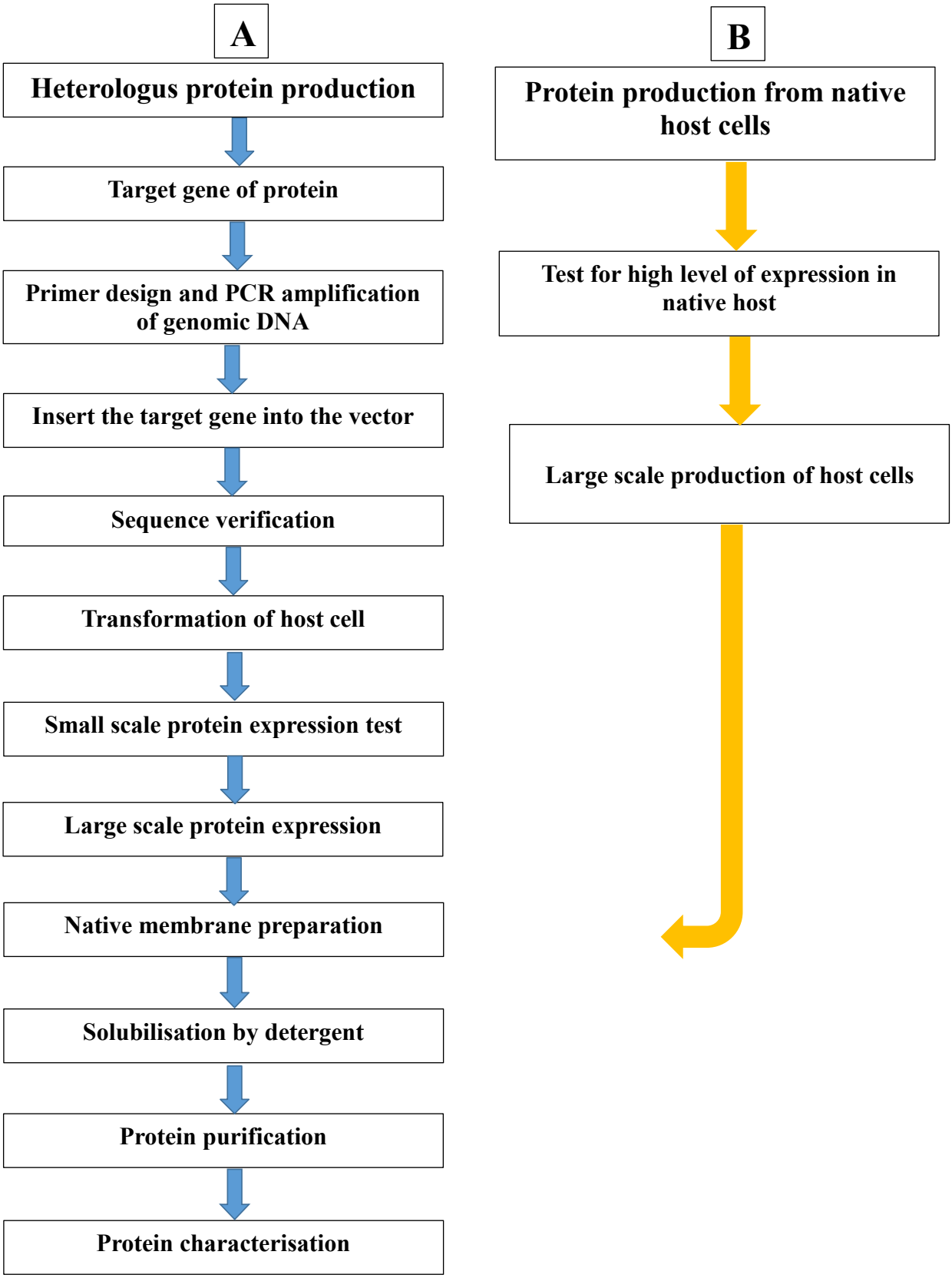



**Figure 1.9** **The standard flow chart presenting the methods of studying membrane proteins. (A) Heterologous proteins production (B) Proteins production from native host cells.**

### 1.5.1 Overexpression of membrane proteins

Theoretically, recombinant protein production is fairly straightforward, beginning with choosing the gene of interest, cloning it in desirable vector, transforming it into the choice host, inducing it for overexpression and finally the protein purification and characterization. But in practice, a number of obstacles can be encountered. Poor growth of the host, protein instability or toxicity, inclusion body formation, unsuitability of temperature, pH, salt concentration etc. and even not expressing any protein at all are some of the complications often found down the pipeline. There is no need to overexpress the protein, if the protein is expressing at high level in native host. For example, few membrane proteins such as bacterial and mammalian rhodopsins, ATPases, aquaporins, photosynthetic complexes, respiratory complexes, reaction centers and light-harvesting proteins are naturally abundant in their native membranes.

Continuous progress has been made in the field of recombinant protein expression and purification over the past few decades. Overexpression of protein is the first essential step for achieving the structural and functional characterisation of nearly all proteins. Membrane protein overexpression can still be a matter of "trial and error". A number of structural genomics programmes have been initiated with prokaryotic membrane proteins (Daley et al., 2005; Gao and Cross, 2006: Lundstrom, 2007). Whilst highly sophisticated techniques have enabled rapid screening of expression, the expression success rate for membrane proteins is ~ 30 - 44 % compared to 70% for soluble proteins (Christendat et al., 2000). Success in membrane protein overexpression is dependent on choosing the appropriate type of organism, expression plasmid, host strain or cell type and cell culture conditions. The importance of all these factors is well established and considered helpful in the overexpression and stabilisation of membrane proteins.

**1.5.1.1 Expression hosts.** The different types of expression hosts available to produce heterologous proteins include bacteria, filamentous fungi, yeast, unicellular algae, insect and mammalian cells. Each species has different advantages and disadvantages and their choice may depend on the protein of interest (Demain and Vaishnay, 2009; Rosano and Ceccarelli, 2014). In this project, our choice organism

was *E. coli*. The advantages are well recognised of using *E. coli* as a host: (i) fast growing organism (Sezonov et al., 2007); (ii) high cell density easily obtained for high production of protein (Shiloach and Fass, 2005); (iii) inexpensive rich complex media can be used for growth (Sivashanmugam et al., 2009); (iv) well-characterised genetics, physiology and metabolism has ensured the availability of a large number of cloning and expression strains with particular advantages (Andersen et al., 2013); (v) transformation with foreign DNA can be fast and easy (Pope and Kent, 1996).

**1.5.1.2 Expression vector.** Vector is one of the key factors in the expression of target proteins in *E. coli.* There are huge numbers of expression vectors available but the most commonly used ones are the result of various arrangements of promoters, replicons, selections markers, multiple cloning sites, and affinity tags (Figure 1.10). To choose a suitable vector, good understanding of these features must be available and their usefulness carefully evaluated according to the desirable needs. High copy number is the most important parameter of vector that can yield more recombinant protein in the cell. Although occasionally a high plasmid number may drop the bacterial growth rate by imposing a metabolic burden and could cause plasmid instability and decrease the protein synthesis (Birnbaum and Bailey, 1991: Marisch et al., 2013). A variety of expression vectors carrying different promoters are commercially available and for an ideal expression system, the vector must have a tightly regulated strong promoter with a low basal expression level (Makrides, 1996; Wang et al., 2003). In some cases, the use of a very strong promoter has been reported lethal for the expression host and tends to result in inclusion body formation (Lilie et al., 1998; Weickert et al., 1996). The $lacI^{Q}$ is the mutated repressor of lacI gene that control the basal expression and achieve higher expression levels (about 10-fold) than lacI (Calos, 1978). For recombinant protein production, the lac promoter and its derivative are weaker and consequently not very beneficial (Deuschle et al., 1986; Makoff and Oxer, 1991). Combination of the different promoters has proven more advantages than others. For example, the tac promoter comprised of −35 region of the trp (tryptophan) promoter and −10 region of the lac recognition site for the repressor, is about ten times stronger than the standard lacUV5 promoter (de Boer et al., 1983). To select the anticipated cells carrying the desirable plasmid and prevent the growth of plasmid-free cells, generally a resistance

marker is used in the plasmid. Antibiotic resistance genes are usually used in the *E. coli* system for this purpose. A multiple cloning site which contains many (up to ~20) restriction sites is one of the key features of a vector which allows foreign DNA to be inserted. To detect the recombinant protein expression and purify from crude extracts a variety of affinity tags have recently been developed. Small peptide tags are less likely to affect properties of recombinant protein; however, in some cases they might have adverse effects on the biological activity or structure of the protein (Rosano and Ceccarelli, 2014).

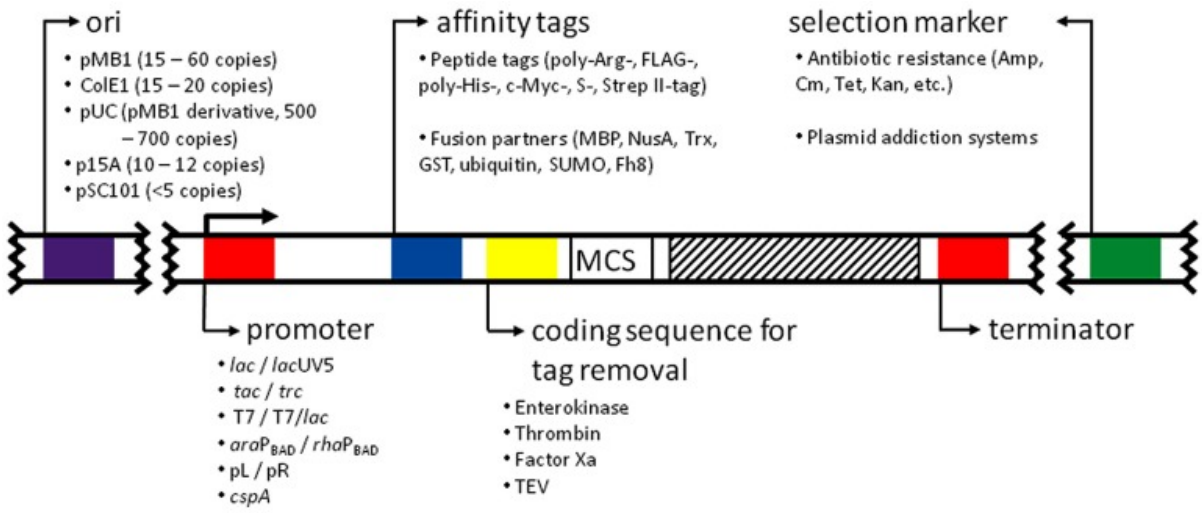



**Figure 1.10 Features of expression vector.** The figure illustrates the main features of an ideal expression vectors. All of them are described in the text. This diagram was adapted from Rosano and Ceccarelli, (2014).

**1.5.1.3 Optimise expression conditions.** Membrane protein expression can be enhanced by the optimization of cell culture conditions. Various factors during cell culture have prominent effects on the expression level of membrane proteins. These include the various types of culture medium, cell density at the time of induction, concentration of inducer, the temperature profile, particularly the post-induction temperature and time. Expression level can also be improved by using different *E. coli* strains (Wang et al., 2003). Optimisation of cell culture conditions has increased expression level by 3 to 5 fold (Auer et al., 2001; Gräslund et al., 2008). Induction is usually carried out during the mid-log phase of bacterial growth, and cells are

harvested prior to reaching the stationary phase. For some proteins overexpression has been achieved by the addition of IPTG at early log phase (Donovan et al., 1996). High level expression of membrane proteins is lethal to cells and may lead to the formation of inclusion bodies, which render them non-functional. Unfortunately, every protein is different and consequently, it is very challenging to create a suitable system for expression and achieve high quality optimum expression and yield of membrane protein.

### 1.5.2 Purification of membrane proteins

The first step in purification of membrane proteins is to extract proteins from their cell membrane using an appropriate detergent. Some membrane proteins are more stable in specific detergents for long periods of time, while removal of other membrane proteins from their native membrane environment can have profound effects on protein stability, structure and function (Seddon et al., 2004; Midgett and Madden, 2007; Postis et al., 2008). Unfortunately, in most cases proteins lose their normal function by leaving their native environment resulting in aggregation, unfolding and protein degradation (Zhou and Bowie, 2000; Rosenbusch, 2001; Booth, 2003; Postis et al., 2008; Bill et al., 2011). A variety of detergents are commercially available, with many shapes and forms widely used to solubilise membrane proteins (Luckey, 2014). Detergents are amphipathic molecules comprised of a hydrophilic headgroup connected to a hydrophobic chain (or tail). Based on their structure, they are classified into four major types. Detergents containing a headgroup of a negative charge are known as anionic detergents while positively charged headgroups are cationic detergents. Zwitterionic detergents possess both negative and positive charges while those detergents having no charge at all are classified as non-ionic (Seddon et al., 2004). Although ionic detergents are good at solubilising membrane proteins but some such as sodium dodecyl sulfate (SDS), usually denature proteins rendering them useless for structural studies (Johansson et al., 2009). Therefore, nonionic detergents are commonly used for solubilisation of membrane proteins; these rarely affect the proteins structural features and are usually considered mild and comparatively non-denaturing. Nonionic detergents such as n-dodecyl-β-d-maltoside (DDM) are most commonly used for many membrane proteins solubilisation, as they often yield

stable proteins with retention of functional properties in aqueous solution (Starling et al., 1995; Misquitta and Caffrey, 2003).

One of the important properties of a detergent is to form micelles. The concentration of detergent at which micelles form is called the critical micelle concentration (CMC). Solubility of membrane proteins by detergent is mostly related to the CMC. Detergent at low concentration (below the CMC), penetrates the lipid bilayer but cannot form micelles. At the higher concentration (above the CMC), detergent molecules replace the lipids and result in mixed micelles (Figure 1.11) (Menger et al., 1998; Garavito and Miller, 2001; Luckey, 2014). All purification buffers must contain detergent at a concentration significantly (10 – 20 times) greater than the CMC in order to avoid membrane proteins precipitation. Although detergents have allowed the purifications of many important membrane proteins and have led to hundreds of membrane proteins structures, they present several issues by removing the proteins from their native membrane environment. To get around this problem researchers have developed methods to reconstitute into proteoliposomes and nanodiscs. However, isolation of the protein still requires initial detergent solubilisation. As an alternative, it has been demonstrated recently that membrane proteins can be directly isolated into a lipid nanodisc using co-polymer compounds such as styrene-maleic acid. These methods provide several advantages. For example, the protein always in the native lipid environment, the protein can be concentrated without concentrating detergent micelles, the proteins typically show high stability and the use of co-polymers has been good for EM technologies. However, there are some disadvantages. For example, the buffers used for solubilisation are restricted since the styrene-maleic acid may depolymerise at low pH and high salt concentrations are required. The bulk of lipid is not good for crystallography and may not be good for NMR. Moreover, temperature and pH have also profound effects on the stability of purified membrane proteins. Most of the purified membrane proteins are more stable at 4°C and prefer approximately neutral pH (6–8). Moreover, additions of protease inhibitors and glycerol are also considered important factors that can stabilise purified membrane proteins (Hunte et al., 2003).

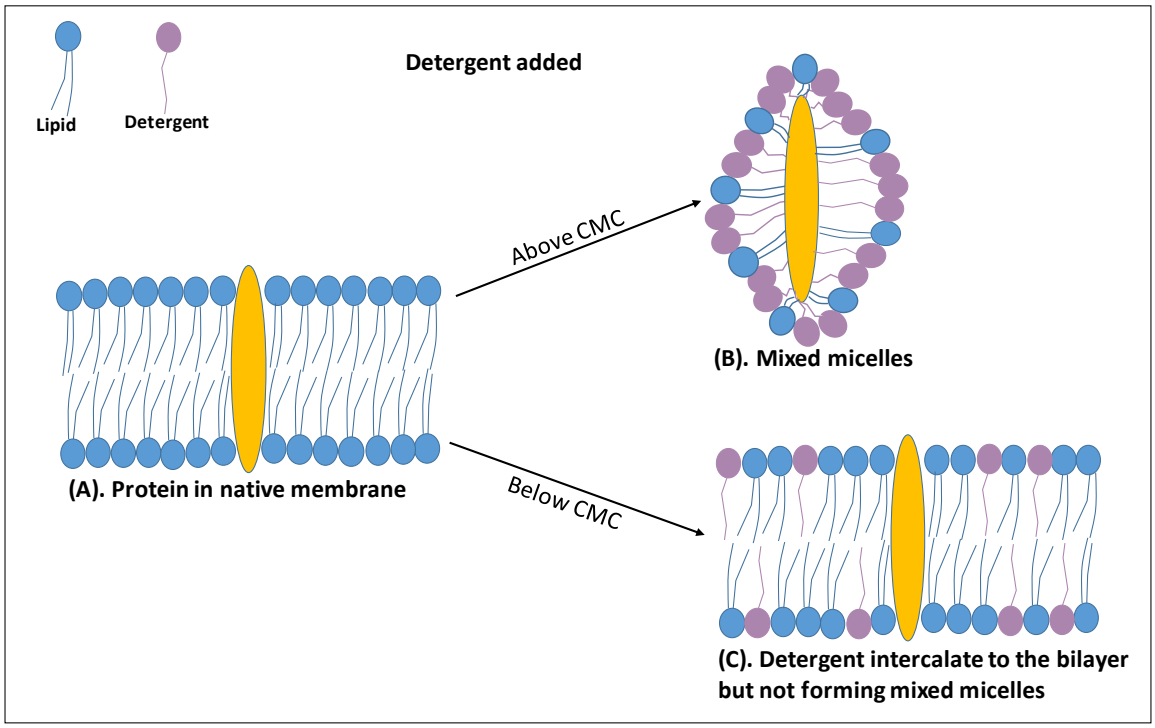



**Figure 1.11 Solubilisation of membrane integral proteins by mild detergents.** (A) Protein in native membrane. (B) Addition of detergent higher than CMC, the lipids adjoining the membrane protein are stripped away and replaced by detergent resulting in the formation of mixed micelles containing the lipids, protein and detergent molecules. (C) Addition of detergent lower than CMC, detergent molecules intercalate in the bilayer without forming mixed micelles.

Following solubilisation, membrane proteins of interest can be purified prior to further biophysical and biochemical characterisation experiments. A number of chromatographic methods are used to purify membrane proteins such as ion-exchange, gel filtration and affinity columns. Ion-exchange and gel-filtration methods are most commonly used for purification of soluble proteins as with membrane proteins the presence of detergent limits their effectiveness. Immobilized metal-affinity chromatography (IMAC) is a popular purification technique that exploit the affinity tags which are fused to the desired protein (Jarvik and Telmer, 1998). This method can be used to purify polyhistidine-tagged proteins very rapidly and efficiently.

In addition to His-tag there are range of alternative affinity tags that can be used for protein purification. For example, Glutathione S-Transferase, Maltose Binding Protein, Calmodulin Binding Peptide, Intein-Chitin Binding Domain, Streptavidin/Biotin-based tags, His-Patch ThioFusion and Tandem Affinity Purification etc. Although longer and shorter affinity tags have been used effectively, generally affinity tags comprising six polyhistidine residues are most commonly used in IMAC. In order to minimize the possible effect of affinity tags on protein function, a six histidine tag is an appropriate choice for efficient purification. Technically, the affinity tags may affect the activity of protein but the polyhistidine affinity tag is relatively small size and rarely interferes with activity. Therefore, the affinity tag removal is usually not recommended following protein purification (Crowe et al., 1994). If affinity tag removal is essential, then a protease cleavage site can be inserted between the protein and the tag (Waugh, 2011). A range of commercially available immobilized metal matrices are used in IMAC for protein purification. Iminodiacetic acid (IDA) and nitrilotriacetic acid (NTA) are the two most commonly used resins, both have two available valencies for interaction with the histidine residues but IDA has three while NTA has four valencies for coordination with metal ions (Porath et al., 1975). NTA is therefore a stronger coordinator of metal ions than IDA and has significant effect on the yield and quality of purified protein (Crowe et al., 1994). Currently, polyhistidine affinity-tagged proteins purification is carried out by commercially available matrices such as nickel-nitrilotriacetic acid ($Ni^{2+}$-NTA) (Hochuli et al., 1987) and $Co^{2+}$–carboxylmethylaspartate ($Co^{2+}$-CMA) (Chaga et al., 1999), which are attached to a solid support resin. These matrices have four coordination sites with metal ions while two coordination sites of the transition metal interact with histidine residues in the affinity tag. A molecular illustration of the interactions between the $Ni^{2+}$-NTA and polyhistidine affinity tag is shown in Figure 1.12.

With IMAC up to 95 % purified protein of high yield can be achieved in a single purification step (Hochuli et al., 1989). Purification using polyhistidine tags has been successful in a number of expression systems, including mammalian cells, (Bornhorst and Falke, 2000) baculovirus-infected insect cells, (Kuusinen et al., 1995), *E. coli*, (Van Dyke et al., 1992) and *Saccharomyces cerevisiae*, (Kaslow and Shiloach, 1994).

**Features of the $His_6$-tag, $Ni^{+2}$ and NTA-resin**

- $His_6$-tag is short length amino acid and in most cases does not affect structure, function and stability of protein.
- Proteins can be purified under native or denaturing conditions due to high binding affinity to Ni-NTA.
- Due to high specificity contaminating proteins are easily removed.
- Nickel has a high binding affinity to protein without leaching of metal ions.
- NTA-resin has high selective binding and a binding capacity of 5-10mg/ml.
- NTA has four chelating coordination sites with nickel ion whereas two coordination sites of the nickel interact with histidine residues in the affinity tag.
- The quite stable affinity resin can be reused up to 5 times.

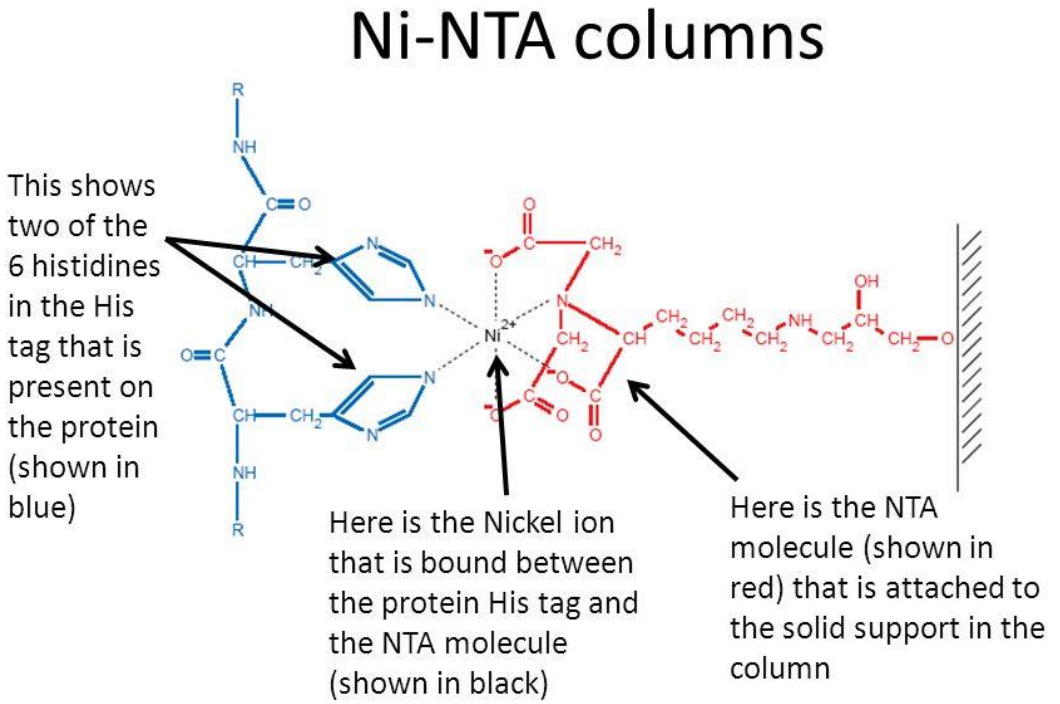



**Figure 1.12 Interaction between Ni-NTA and a His-tagged protein.** For protein purification, NTA is the most commonly used IMAC affinity resin (shown in red). Two imidazole rings of the 6 histidines tag (shown in blue) interacting with the $Ni^{2+}$ (shown in black).This figure was adopted from figures in the Qiagen expressions manual (Qiagen, 2003).

## 1.6 Classification of membrane transport proteins

The Transporter Classification Database (TCDB) provides a comprehensive classification system for membrane transport proteins where more than 600 transporter families have been organised with free access from the web (http://www.tcdb.org/) (Saier et al., 2006; Saier et al., 2013). Transport proteins from all known organisms are classified into families and superfamilies based on their amino acid sequences similarities and their functions. To give each protein a transporter classification number, the system is based on five components which are V, W, X, Y and Z. Firstly, each transporter is assigned a number (V) which represents the molecular mode of transporter class (i.e. primary transport/ secondary transport etc). W refers to the source of energy to drive transport. X, Y and Z refer to the protein amino acid sequence, family/subfamily and substrate specificity.

### 1.6.1 Families of nucleobases and nucleosides transport proteins

There are five distinct superfamilies of nucleobase and nucleoside transport proteins that have been identified in bacteria, fungi, protozoa, algae, plants and mammals (Cabrita et al., 2002; Cecchetto et al., 2003). These include the purine related transporter family (PRT) now also named as nucleobase cation symporter family 1 (NCS1). The nucleobase-ascorbate transporter family (NAT) also called nucleobase cation symporter family 2 (NCS2) (Karatza et al., 2006); the purine related permeases (PUPs) family (de Koning and Diallinas, 2000) the equilibrative nucleoside transporter family (ENT) and the concentrative nucleoside transporter (CNT) (Diallinas & Gournas, 2008).

**1.6.1.1 The NCS1 family transporters.** Proteins of this family function as transporters for nucleobases including nucleosides, allantoin, hydantoin, thiamine, uric acid and other related compounds. Presently the NCS1 family includes over 2000 sequenced proteins derived from bacteria, yeast, archaea, fungi and plants (Saier et al., 2009; Hamari et al., 2009; Witz et al., 2014; Krypotou et al., 2015; Ma et al., 2016; Sioupouli et al., 2017). Transporters of this family are predicted to have 12 putative transmembrane spanning alpha-helices (Saier et al., 2009; Witz et al., 2014) and function using a symport mechanism driven by a proton or sodium

gradient (Krypotou et al., 2015). Generally, all transporters of the NCS1 family function as vital components of salvage pathways for nucleobases and related metabolites (Weyand et al., 2008). The NCS1 family proteins are usually 419–635 amino acid residues in length and show no sequence similarity to the NCS2 family. The first structural model for the NCS-1 family is the sodium-coupled hydantoin transport protein, Mhp1 from *Microbacterium liquefaciens* (Suzuki and Henderson, 2006) for which crystal structures have been determined in three different conformations i.e outward-facing open, occluded with substrate and inward-facing open (Weyand et al., 2008; Shimamura et al., 2010; Simmons et al., 2014). The common transport mechanism catalysed by NCS1 family proteins is simplified as;

Nucleobase or Vitamin or Hydantoin (out) + $H^+$ or $Na^+$ (out) ⟶

Nucleobase or Vitamin or Hydantoin (in) + $H^+$ or $Na^+$ (in)

**1.6.1.2 The NCS2/ NAT family.** Members of this family are found in bacteria, archaea, fungi, plants, metazoans and mammals including human. This is the most widespread and conserved family of nucleobase transporters (de Koning and Diallinas, 2000; Frillingos, 2012). Historically this family was determined from characterization of the UapA and UapC (uric acid-xanthine permeases) of *Aspergillus nidulans* and the UraA and PyrP (uracil permeases) of *Escherchia coli* and *Bacillus subtilis* (Diallinas & Gournas, 2008; Gournas at al., 2008). The NCS2 family proteins are usually 400–650 amino acid residues in length and possess mostly 10-14 transmembrane helices (Gournas et al., 2008; Lu et al., 2011). The UapA is the best characterised and widely studied transporter of the NAT family and is therefore used as a model for understanding this family (de Koning and Diallinas, 2000; Leung et al., 2010).

**1.6.1.3 The PUP family.** The members of this family are typically hydrophobic membrane proteins found only in plants. The PUP family transporters do not share any amino acid sequence similarities with other nucleobase transporters. They are $H^+$ symporters and exhibit a broad substrate spectrum including cytosine, adenine, uracil, uric acid, vitamins, alkaloids and purine related metabolites like allantoin, xanthine, cytokinin and caffeine (Gillissen et al., 2000; Desimone et al., 2002;

Diallinas and Gournas, 2008; Girke et al., 2014). Several homologous proteins have been identified in many plant species like Arabidopsis, Oryza sativa and Nicotiana tabacum (Hildreth et al., 2011; Jelesko, 2012; Goodstein et al., 2012).

**1.6.1.4 The ENT and CNT families.** ENT and CNT are two widely expressed nucleoside transporter families that mediate the transport of nucleosides and their analogues across cell membranes (Hyde et al., 2001; Young et al., 2013; Cabrita et al., 2002; Li et al., 2003). Concerning the transport mechanism, ENTs generally transport nucleosides down their concentration gradients while CNTs catalyse transport against their concentration gradients using either sodium or proton symporters (Cabrita et al., 2002; Li et al., 2003; King et al., 2006). The ENT family transporters are present in a wide range of eukaryotic organisms and their homologous proteins have been sequenced from protozoa, yeast and nematodes.

### 1.6.2 Role of nucleobases in organism

Transport of various metabolites and ions into and out of cells is essential in order to maintain their homeostatic conditions. Transport of nucleobase is important in cells for nucleic acid synthesis, coenzyme function (FAD, $NAD^+$, $NADP^+$), and as a source of nitrogen in many organisms (Gournas et al., 2008; Diallinas & Gournas, 2008). From the medicinal point of view, a large number of nucleobase and nucleoside analogs have been developed and are widely used in the treatment of many diseases. Basically, these analogs are chemically modified by being similar enough to the usual nucleobases and nucleosides required for DNA replication and once incorporated into the growing DNA strands, obstruct the DNA strands extension and stop the cell growth and division. These nucleobase analogues include anticancer drugs, antivirals, antibiotics, antitumor agents, immunosuppressants, and drugs for parasitic diseases (Pantazopoulou and Diallinas, 2007; Diallinas and Gournas, 2008). Research on nucleobase transporters largely strives to advance understanding of nucleobase specificity and their roles in medicine, agriculture and pharmacology (de Koning and Diallinas, 2000).

**1.6.2.1 Nomenclature of nucleobases, nucleosides and nucleotides.** Nucleobases are biological molecules containing nitrogen heterocycles and are the basic building blocks of genetic material in all living organisms. In genetics, they are often known as nitrogenous bases or simply called bases due to their ability to form base pairs in the structure of DNA and RNA. Nitrogenous bases either belong to the double-ring class of molecules called purines (adenine and guanine) or to the single six-sided di-nitrogenous ring of molecules called pyrimidines (thymine, cytosine and uracil), abbreviated as A, G, T, C and U, respectively (Figure 1.13).

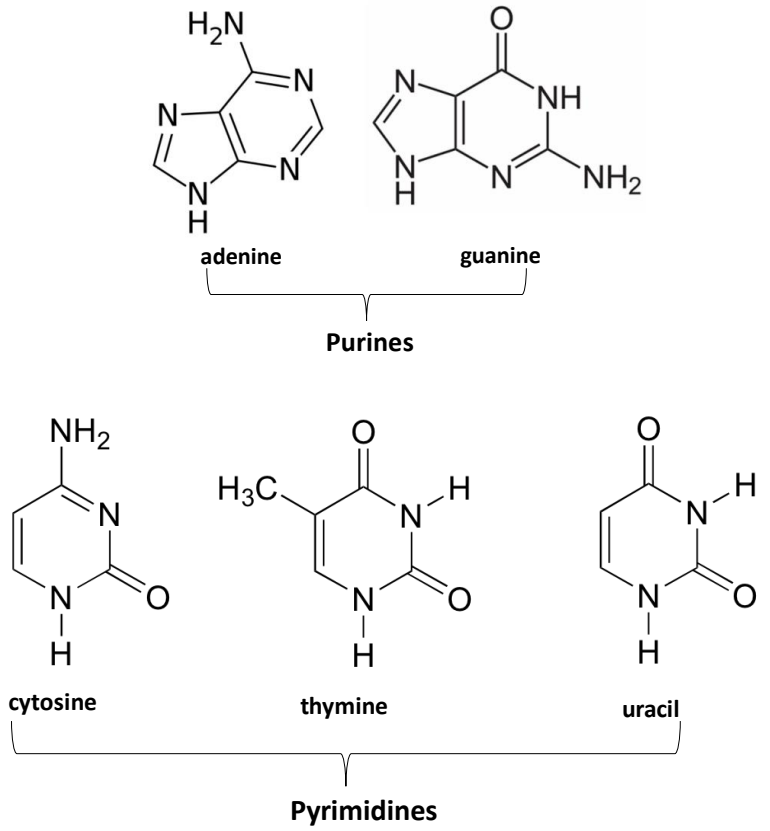



**Figure 1.13 Structures of purines and pyrimidines.** Structures were drawn using ChemDraw software.

When a nitrogenous base is covalently linked to a five-carbon sugar (ribose or deoxyribose), without any phosphate groups; they are known as nucleosides but addition of one to three phosphate groups to the nucleoside results in nucleotide (Figure 1.14).

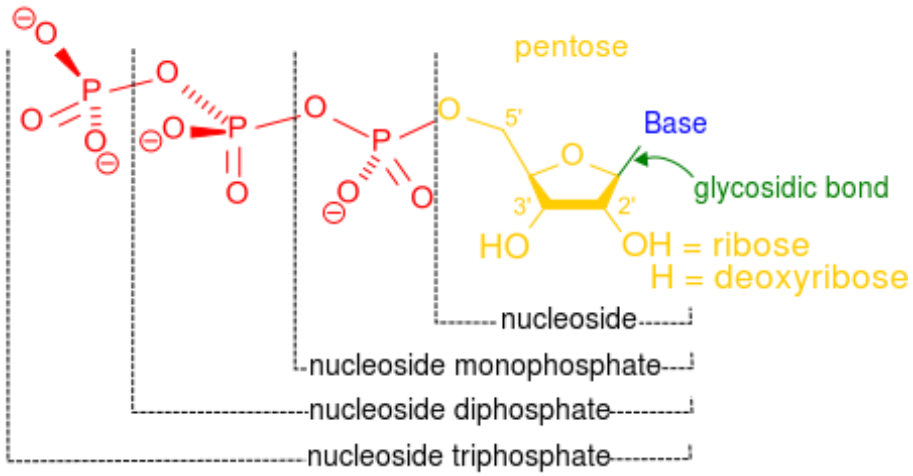



**Figure 1.14 Structure of a nucleoside and various forms of nucleotides.**

## 1.7 Antimicrobial

The term antimicrobial is a combination of the Greek words anti (against), mikros (little) and bios (life) and includes all substances (natural or synthetic) that kill or inhibit the growth of all types of microorganisms with no or little damage to the host. Antimicrobials are classified according to the microbes they act mainly against such as bacteria (antibacterial), fungi (antifungal), viruses (antiviral), and protozoa (antiprotozoal). The antibacterials are the largest and well known class of antimicrobials that inhibit or kill the growth of bacteria. Antibacterials can be categorised in several different ways, including spectrum of activity, effect on bacteria and mode of action.

### 1.7.1 Spectrum of activity

An antibiotic may be recognised as "broad-spectrum" or "narrow-spectrum" depending on the range of susceptible bacterial species. Broad-spectrum antibiotics are effective against wide range of Gram-negative and Gram-positive bacteria. Examples of broad-spectrum antibiotics are the aminoglycosides, the third and fourth generation cephalosporins, the quinolones and some synthetic penicillins. Narrow-spectrum antibiotics have limited activity and inhibit only the growth of specific bacteria, either Gram-negative or Gram-positive. Examples of narrow-spectrum antibiotics are penicillin, bacitracin and glycopeptides that are only effective against Gram-positive bacteria, whereas polymixins are only effective against Gram-negative bacteria.

### 1.7.2 Effect on bacteria

Based on their mechanism of action, antibiotics are generally divided into two classes: bactericidal and bacteriostatic. Bactericidal antibiotics kill the bacteria and their actions are irreversible while bacteriostatic antibiotics inhibit or delay the bacterial growth and their actions are reversible. Examples of bactericidal drugs are quinolones, aminoglycosides, cephalosporins, and penicillins. Bacteriostatic examples are macrolides, tetracyclines and sulfonamides. Some antibiotics have both bactericidal and bacteriostatic actions depending on the dose, period of exposure and the nature of the bacteria. For example, aminoglycosides, metronidazole and fluoroquinolones exert concentration-dependent killing characteristics, as the drug concentration increases, their rate of killing also increases.

### 1.7.3 Modes of action

Different antibacterial agents have different modes of action, which generally fall within one of five mechanisms (Figure 1.15).

**1.7.3.1 Inhibitors of cell wall synthesis.** The most common target for antibiotics is the inhibition of cell wall synthesis. Animals and human cells do not have cell walls, but this structure is found in bacteria. Drugs that target cell walls can therefore selectively kill or inhibit growth of bacterial organisms. Examples of such drugs include penicllins, bacitracin, cephalosporins and vancomycin.

**1.7.3.2 Inhibitors of cell membrane function.** Some antibiotics disrupt the cell membranes, which could result in leakage of solutes essential for cell survival. The actions of this class of antibiotics are often poorly selective and can be toxic during systemic use in mammals, because cell membranes are found in both prokaryotic and eukaryotic cells. Therefore, clinical usage is limited to topical applications. Examples include polymixin B and colistin.

**1.7.3.3 Inhibitors of protein synthesis.** Several types of antibacterial agents target bacterial protein synthesis by binding to a target site or sites (30S or 50S subunits) on the ribosomes and inhibit protein synthesis, consequently leading to bacterial cell death or inhibition of growth. Examples include aminoglycosides, macrolides, streptogramins, tetracyclines and chloramphenicol.

**1.7.3.4 Inhibitors of nucleic acid synthesis.** Some antibiotics interrupted the process of DNA or RNA synthesis by binding to the components which are involved in their production. This causes interference with the normal cellular processes of bacterial multiplication and cause cell death. Examples include metronidazole, quinolones, and rifampicin.

**1.7.3.5 Inhibitors of other metabolic processes.** Some antibiotics have an effect on the selective processes of bacterial cell, which are essential for their survival. For example, sulfonamides and trimethoprim inhibit synthesis of the folic acid pathway, which is essential for bacterial DNA synthesis. Trimethoprim inhibits dihydrofolate reductase and sulfonamides bind to dihydropteroate synthase. These two enzymes are important for the folic acid production, which is a vitamins, synthesized by bacteria, but not humans. Bacteria must synthesize folic acid from p-aminobenzoic acid because folic acid cannot enter bacterial cells by active transport or diffusion.

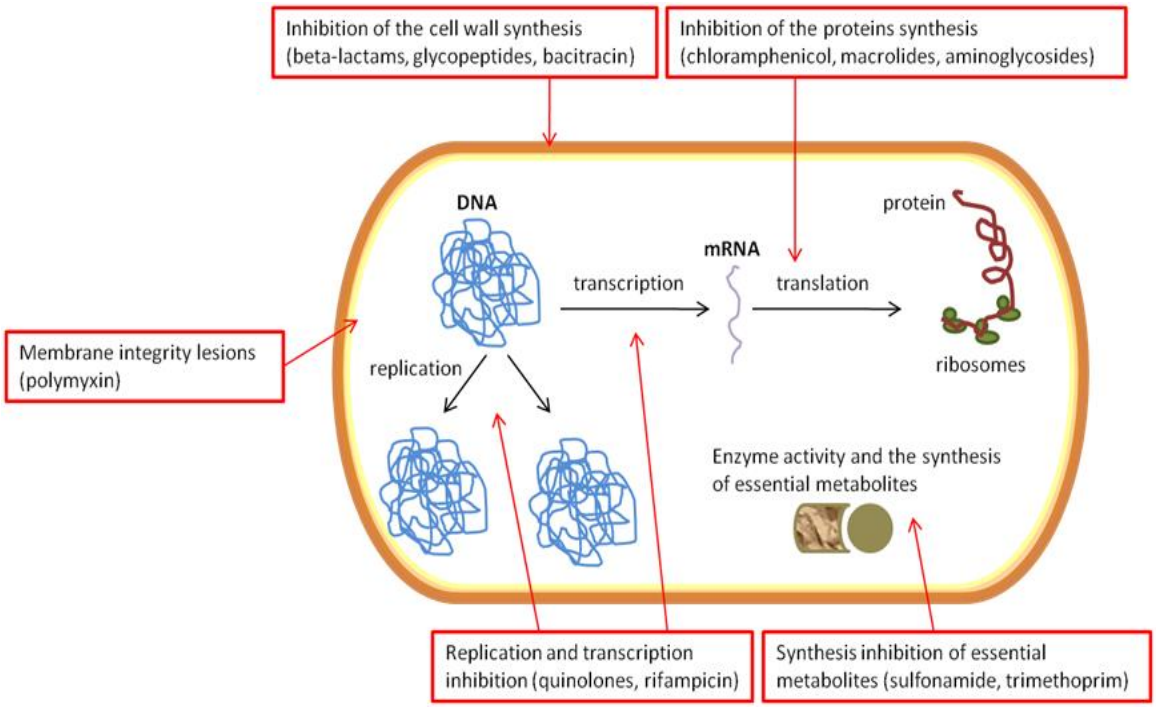



**Figure 1.15 Major targets of common antibacterial agents.** Antimicrobial agents function in one of the five mechanisms. Inhibiting cell wall synthesis, injuring the plasma membrane, inhibiting protein synthesis, inhibiting nucleic acid synthesis or inhibiting synthesis of essential metabolites. This diagram was taken from http://www.intechopen.com.

## 1.8 Antibiotic resistance

When an antibiotic loses its ability to stop or kill bacterial growth it is termed as antibiotic resistance; in other words, the bacteria are known to be "resistant" when they continue to grow in the presence of therapeutic levels of an antibiotic. There is a rapid emergence of resistant bacteria occurring worldwide, which is endangering the efficacy of antibiotics that have transformed medicine and saved millions of lives (Sengupta et al., 2013; Golkar et al., 2014; Wright et al., 2014; Lekshmi et al., 2016; Chaudhary, 2016). The developing crisis in antibiotic resistance has been attributed to their over misuse and acquisition of genetic elements, as well as a lack of new drug development (Magee et al., 1999; Bartlett et al., 2013; Viswanathan, 2014; Read and Woods, 2014; Ventola, 2015). The mechanisms of antibiotic resistance are classified as either intrinsic or acquired (Crumplin and Odell, 1986; Martinez and Baquero, 2000; Hegstad et al., 2010).

### 1.8.1 Intrinsic or natural resistance

Intrinsic resistance is the innate ability of a bacterial species carrying genes that create a resistance phenotype. Different species and strains exhibit different ranges of antibiotic response phenotypes. Since the beginning of the antibiotic era, the handiness of genomewide mutagenesis methods and quick bacterial genome sequencing has exposed many intrinsic gene functions in bacteria that may lead to resistance phenotypes in clinical practice. For instance, gene amplification is a common genetic route that can enhance antibiotic resistance to the sulphonamides (Kashmiri and Hotchkiss, 1975) and trimethoprim (Brochet et al., 2008). These studies can help to predict future problems that may emerge due to antibiotics selection pressure associated with the clinical environment (Davies and Dorothy, 2010).

### 1.8.2 Acquired resistance

This occurs when a microorganism gains resistance to the activity of a particular antimicrobial agent to which it was susceptible previously. Acquired resistance occurs as a result of mutation of chromosomal genes or acquisition of plasmid DNA, integrons or transposable elements. Bacteria transfer genes horizontally by means of

three different mechanisms, which are transformation, transduction and conjugation (Figure 1.16).

Natural transformation is one of the major mechanisms in the evolution of microbes. It is the process by which bacterial cells take up a naked piece of DNA from the surrounding atmosphere and incorporate it into their genomes (Johnston et al., 2014). Consequently, the entering DNA carrying a resistance gene is replicated with the recipient cell, so conferring resistance to an antibiotic to which bacteria were previously sensitive.

Transduction is another mechanism of gene transfer by which foreign DNA is introduced into a bacterial cell by a bacteriophage (Davis, 1990). The donor DNA is packed into the capsid of bacteriophage and during a subsequent infection is inserted into the recipient bacterial cell. The transfer of resistance genes is limited to strictly related species, as a high level of specificity is required for the absorption of bacteriophage (Courvalin, 1996). The information about transduction is limited by comparing with transformation and conjugation. Transduction has traditionally been underestimated as a significant mechanism of horizontal gene transfer in naturally occurring environments. Numerous studies have shown that antibiotic resistance genes are found in phage particles in nature (Muniesa et al., 2013). Recently, transduction has been considered a very significant mechanism in the transfer of antibiotic resistance among potential pathogens as validated during in vitro experiments from *Enterococcus gallinarum* to *Enterococcus faecalis* (Vidana, 2015). Further in vitro experiments with *Salmonella* transducing bacteriophages confirmed that tetracycline, chloramphenicol, and ampicillin resistance genes could be transduced (Schmieger and Schicklmaier, 1999).

Conjugation is the process by which bacterial cells exchange genetic material through direct contact whereas transformation and transduction mechanisms do not involve cell to cell contact (Holmes and Jobling, 1996). This is the best studied mechanism of horizontal gene transfer and is mediated by plasmids encoding their own transfer from donor to recipient (Arutyunov and Frost, 2013). Conjugative plasmids are self-transmissible and carry genes that code for all the functions needed for conjugation and promote its own transfer. Many conjugative plasmids can facilitate the transfer of nonconjugative ones through cointegration, mobilisation or retrotransfer (Mercier et al. 2007; Sota and Top, 2008).

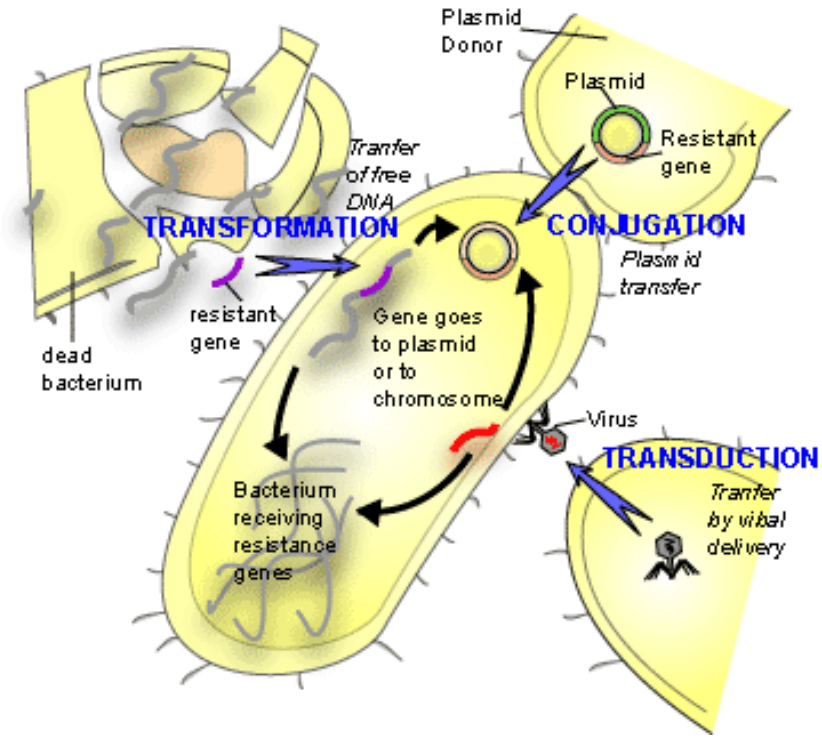



**Figure 1.16 Genetic pathways used by bacteria to enhance antibiotic resistance.** The figure illustrates three possible mechanisms (conjugation, transduction and transformation) of gene transfer in bacteria. This diagram was taken from http://ww.textbookofbacteriology.net.

## 1.9 Molecular mechanism of antibiotic resistance

Bacteria use a range of defence mechanisms to protect themselves from lethal effects of antibiotics (Figure 1.17). They can be classified into four basic types (McManus, 1997).

### 1.9.1 Reduced permeability or uptake

Most of the antibiotics target intracellular biological processes, and therefore, must penetrate through the bacterial cell envelope. Gram-negative bacteria are naturally more resistant to many antibiotics than Gram-positive bacteria, which is largely due to the presence of an outer membrane in Gram-negative bacteria that serves as a tough barrier for permeation (Delcour, 2009). Bacteria use this type of mechanism and reduce permeability or uptake of antibiotics by structural alteration of the membrane. For example, Gram negative bacteria induce some change to the

complement of porin channels and prevent penetration of antibiotic into the cell (Nikaido, 2001; Delcour, 2009; Shah et al., 2014; Zgurskaya et al., 2015; Blair et al., 2015).

### 1.9.2 Enzymatic modification or inactivation

Direct destruction or modification of antibiotic is another resistance mechanism frequently used by bacteria (Davies, 1994). The enzyme β-lactamase is a typical example of this mechanism that hydrolyses the β-lactam ring of penicillins. Many other enzymes mediate some structural changes to the antibiotics by transferring their functional group such as phosphoryl, acyl, ribosyl, or thiol groups. Due to the irreversible structural modification, the antibiotics are unable to bind to the target sites (Wright, 2005).

### 1.9.3 Alteration of the target site

Bacteria use numerous mechanisms to inactivate the antibiotics by altering the molecular target sites, for example DNA gyrase, RNA polymerase, prokaryotic ribosome, protein synthesis inhibitors and targets of antimetabolite drugs, such as the sulfonamides and related drugs (Fabrega et al., 2009; Spratt, 1994; Drapeau et al., 2010; Widdowson and Klugman, 1998; Roberts, 1996). Bacteria usually develop alternative metabolic pathways that bypass the reaction inhibited by antibiotics. Such resistance may arise as a result of mutational events or acquisition of new genes carried by plasmids or transposons that modify the normal target and render it unable to bind with antibiotics (Lambert, 2005).

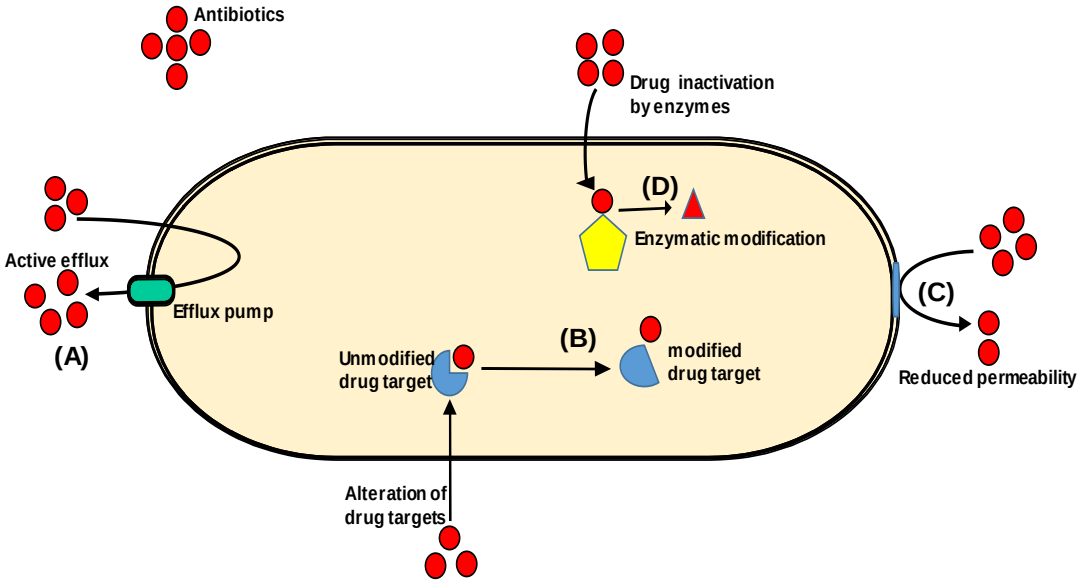



**Figure 1.17 Four major mechanisms of antibiotic resistance.** An oval shape represents a bacterium with various mechanisms illustrated for resistance to antimicrobial agents. (A) active efflux of drugs from the bacterial cell, (B) drug target modification, (C) reduced drug permeability by membrane modification, and (D) drug inactivation by enzymes. This figure was adapted from Kumar and Varela, (2012).

### 1.9.4 Drug efflux

Active efflux is one of the most important mechanisms of antibiotic resistance (Webber and Piddock, 2003). Efflux pumps are proteinaceous transporters that can extrude unwanted toxic substances from inside bacterial cells into the outside environment. They are found in Gram-negative as well as Gram-positive bacteria (Van-Bambeke et al., 2000). Efflux pumps not only play a key role in drug resistance but also perform many other functions in bacteria like biofilm formation, cell to cell communication, and heavy metal extrusion; they also have a role in bacterial pathogenicity and virulence (Piddock, 2006, Li and Nikaido, 2009; Sun et al., 2014; Blanco et al., 2016). They are active transporters and use energy to perform their function. The primary active transporters utilise ATP as a source of energy, whilst secondary active transporters use electrochemical potential difference created by proton pumping or sodium ions.

Currently there are six major families of bacterial efflux transporters (Figure 1.18). These families are classified on the basis of their amino acid sequence similarity, substrate specificity and the energy source used to export their substrates: the adenosine triphosphate (ATP)-binding cassette (ABC) superfamily (Lubelski et al., 2007); the major facilitator superfamily (MFS) (Pao et al., 1998; Reddy et al., 2012); the small multidrug resistance (SMR) family (Chung et al., 2001); the resistance-nodulation-division (RND) family (Tseng et al., 1999; Seeger et al., 2008; Nikaido and Takatsuka, 2009); the multidrug and toxic compound extrusion (MATE) family (Kuroda and Tsuchiya, 2009); and the Proteobacterial Antimicrobial Compound Efflux (PACE) family (Hassan et al., 2015). Only the ABC superfamily are primary transporters utilizing ATP hydrolysis to expel substrates, while the other families are secondary transporters using the proton or sodium gradient as the energy source.

**1.9.4.1 The ATP-binding cassette (ABC) superfamily.** The ABC family is one of the largest and possibly one of the oldest superfamilies of membrane-bound proteins that utilise the free energy of ATP-hydrolysis and translocate a broad range of substrates including amino acids, lipids, sugars, and antimicrobial compounds (Jones and George, 2004). ABC transporters play various physiological roles in the cell e.g. uptake of nutrients, removal of waste/toxic materials and export of many cellular components such as cell wall polysaccharides (Rees et al., 2009). Significant amino acid sequence homology is shown by members of this family. All member proteins have two ATP Binding Cassette regions (occasionally referred to as a nucleotide binding domain where ATP binds) and most have 12 transmembrane alpha-helices (Higgins, 1992). Proteins of this family are widespread in all species from microbes to man and play various physiological roles in the cell, e.g. uptake of nutrients, removal of waste/toxic materials and export of many cellular components such as cell wall polysaccharides (Dassa and Bouige, 2001; Rees et al., 2009). In *E. coli* about 5% of the entire genome codes for the ABC transporters (Higgins, 2001). In microorganisms many ABC transporters contribute to antibiotic and antifungal resistance, whereas in man many are associated with several genetic diseases including tangier disease, cystic fibrosis, obstetric cholestases, and multidrug resistance of cancers.

**1.9.4.2 The major facilitator superfamily (MFS).** The MFS is a well characterised widespread superfamily of secondary active transport proteins family (Griffith et al., 1992; Marger and Saier, 1993; Wong et al., 2012). Members of this family are universally found in the membranes of all types of living organisms, catalysing transport of a diverse range of substrates including sugars, amino acids, metabolic intermediates and drugs (Marger and Saier, 1993; Paulsen et al., 1998). Structurally, most of the MFS transporters consist of 400-600 amino acids with either 12 or 14 transmembrane spanning helices (Henderson and Maiden, 1990; Saier et al., 1999; Chitsaz and Brown, 2017). These proteins act as symporters, uniporters or antiporters, exploiting the power of electrochemical gradients to drive transport (Paulsen et al., 1996; Paulsen et al., 1998). The best example of a well characterised protein of this family is EmrD from *E. coli*, which is a proton-dependent secondary transporter that exports amphipathic compounds, such as carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), across the membrane whose crystal structure has been determined (Yin et al., 2006).

**1.9.4.3 The resistance-nodulation-division (RND) superfamily.** Proteins of this family exist mainly in Gram-negative bacteria but can also be found in Gram-positive bacteria, eukaryotes and archaea (Goldberg et al., 1999). These proteins are located in the cytoplasmic membrane and possess large periplasmic domains, involved in maintaining homeostasis of the cell and export of macromolecules and toxic materials across the inner and outer membranes (Saier et al., 1994; Mao et al., 2002; Murakami and Yamaguchi, 2003; Elkins and Nikaido, 2003). The RND family transporters are made of large polypeptide chains and range in size from 700 to over 1300 amino acid residues in length (Saier et al., 1994). Like MFS transporters, RND transporters have 12 putative transmembrane alpha-helices, with two large loops between TMS 1 and 2 and 7 and 8 (Saier et al., 1994; Tseng et al., 1999). The well-studied example of the RND transporter is AcrB from *E. coli*, for which the crystal structure has been solved (Pos and Diederichs, 2002; Murakami et al., 2002; Elkins and Nikaido, 2003; Pos, 2009). AcrB is the transporter protein and TolC is the outer membrane protein whereas AcrA is the membrane fusion protein (Nikaido and Takatsuka, 2009). It is believed that AcrB and AcrA proteins are in constant contact, while the protein TolC is only recruited as soon as a ligand is bound to the inner

membrane protein (Anes et al., 2015). Other examples of this family include MexB and MexD from *Pseudomonas aeruginosa* (Li et al., 1995) and MtrD from *Neisseria gonorrhoeae* (Bolla et al., 2014).

**1.9.4.4 The small multidrug resistance (SMR) family.** Proteins of this family are small membrane transport proteins, ranging in size from 100 to 150 amino acid residues with only four putative transmembrane helices (Grinius and Goldberg, 1994; Bay et al, 2008; Chitsaz and Brown, 2017). They are highly hydrophobic proteins and are widely found in bacteria and archaea (Paulsen et al., 1996; Bay et al., 2008). An electrochemical proton gradient is the driving force in these transporters for drug efflux, which confers resistance to a variety of antimicrobial agents (Grinius and Goldberg, 1994; Chitsaz and Brown, 2017). Proteins of the SMR family demonstrate unique regions of conservation with other efflux family transporters (Bay and Turner, 2009). Site-directed mutagenic studies have revealed that the negatively charged glutamate (Glu14) of SMR members is a highly conserved residue that is essential for their activity and is directly involved in drug and proton binding (Gutman et al., 2003; Koteiche et al., 2003). The most well characterized is EmrE from *E. coli*, the prototype of all SMR proteins, that transports a range of toxic substrates specifically positively charged aromatic compounds such as acriflavine, ethidium, benzalkonium, methyl viologen, tetraphenylphosphonium, erythromycin, tetracycline, and sulfadiazine (Paulsen et al., 1993; Schuldiner et al., 1997; Heir et al., 1999; Bay et al., 2007; Chitsaz and Brown, 2017).

**1.9.4.5 The multidrug and toxic compound extrusion (MATE) family.** Proteins of this family are found in bacteria, eukaryotes and archaea and function as basic transporters of metabolic and xenobiotic organic cations (Omote et al., 2006; Moriyama et al., 2008; Kuroda and Tsuchiya, 2009). MATE proteins consist of 400-700 amino acid residues with 12 putative transmembrane alpha-helices (Hvorup et al., 2003). Proteins of this family function as drug/sodium or drug/proton antiporters (Brown et al., 1999; Omote et al., 2006; Kuroda and Tsuchiya, 2009). NorM from *Vibrio parahaemolyticus* is a prototype of this family, which functions as drug/sodium antiporter and is the first example of a $Na^+$-coupled multidrug efflux

transporter (Morita et al., 1998; Morita et al., 2000). The multidrug efflux transporter NorM and its *E. coli* homologue YdhG were initially classified as MFS pumps because they possessed 12 putative transmembrane alpha-helices (Morita et al., 1998). Later on, they were reclassified as members of a new family of multidrug resistant proteins (MATE), as there was little sequence homology of these pumps with other multidrugs resistant efflux pumps (Brown et al., 1999; Morita et al., 2000).

**1.9.4.6 The proteobacterial antimicrobial compound efflux (PACE) family.** In the prokaryotic kingdom PACE is the sixth family of multidrug efflux systems found in proteobacterial lineages (Hassan et al., 2015). Transporters of this family are small membrane proteins consisting of 140-200 amino acid residues with four putative transmembrane alpha-helices. The amino acid sequence of the proteins of this family are related to the *Acinetobacter baumannii* protein (AceI) protein, which is a prototype for this novel family of multidrug efflux pumps (Hassan et al., 2013; Hassan et al., 2015). AceI was first time recognised during a transcriptomic study in *A. baumannii* from being highly overexpressed in response to chlorhexidine shock treatment (Hassan et al., 2013). The study also revealed that AceI uses the mechanism of active efflux to confer resistance to chlorhexidine (Hassan et al., 2013). Computational analyses and functional assays have demonstrated that PACE family proteins are found in the genome of various proteobacterial species. Currently, the substrates for these transporters have been identified as synthetic biocides including chlorhexidine, acriflavine, proflavine, benzalkonium and dequalinium. Only during the last few decades have most of these synthetic compounds come into use, so it seems very unlikely that these biocides would be the physiological substrates for these transport proteins. In light of their current revelation, PACE family proteins are the least well characterized and generally very little is known about their structures and functions.

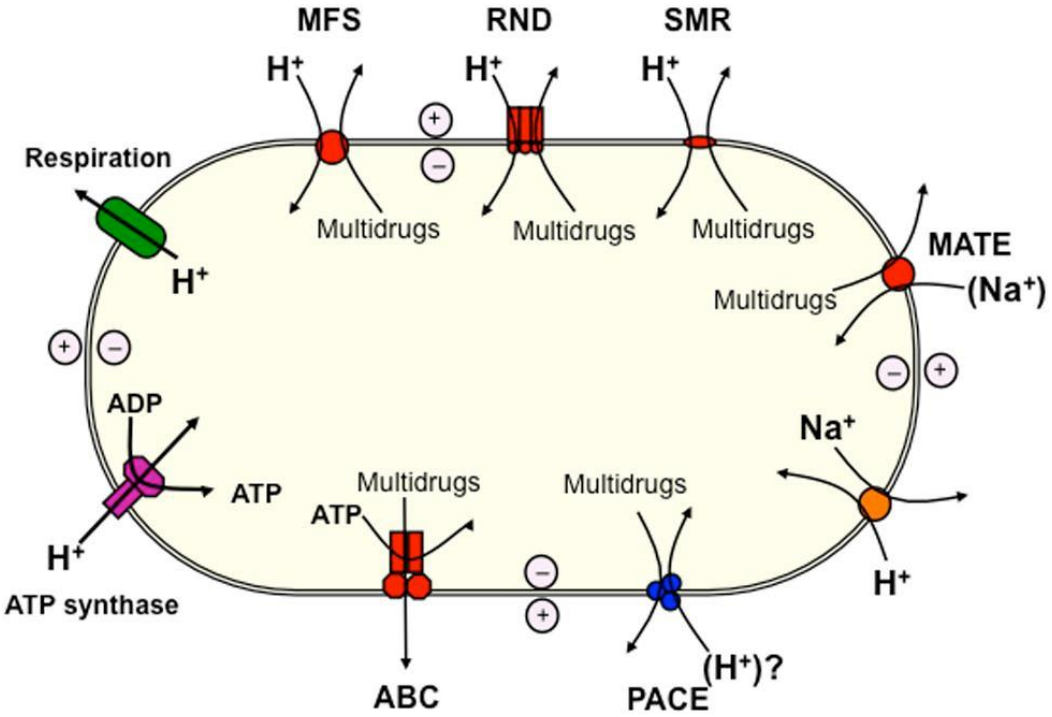



**Figure 1.18 Schematic diagram representing the main types of multidrug efflux systems in bacteria.** The oval shape represents the bacterial cell illustrating the six major families of efflux transporters: the adenosine triphosphate (ATP)-binding cassette (ABC) superfamily, the major facilitator superfamily (MFS), the resistance-nodulation-division (RND) family, the small multidrug resistance (SMR) family, the multidrug and toxic compound extrusion (MATE) family and the proteobacterial antimicrobial compound efflux (PACE) family. This diagram was adapted from Hassan et al., (2015).

## 1.10 Aims and objectives of this research

The overall aim of this study is to increase our knowledge and understanding of the molecular mechanisms of transport proteins by which desirable nutrients are imported and toxic materials are exported across biological membranes. Although high levels of expression of many bacterial membrane proteins have been achieved, a sizeable proportion of these proteins have been difficult to overexpress using existing methodologies. Thus, one of the objectives of this study is to implement strategies for high level expression and purification of membrane proteins for structural and functional studies. In particular, the *Pseudomonas putida* allantoin transport protein (AAN69889) and *Vibrio parahaemolyticus* cytosine transporter (VPA1242) will be characterized further using a range of biochemical and biophysical techniques as representative examples of NCS1 family transporters.

Another part of this project is focused on the structural and functional characterisation of PACE family proteins that are multidrug efflux transporters. Because of their recent discovery, this family of proteins is the least well characterised and very little is known about their structure and function. The *Acinetobacter* chlorhexidine efflux protein (AceI) from *A. baumannii* is a prototype for a novel family of multidrug efflux pumps. The multidrug efflux transporters are particularly important to study as these proteins may help to find a possible convenient target for the development of novel antibacterials.

Therefore, this project was designed to focus on the structural and functional characterisation of NCS1 and PACE family proteins. The functional informations obtained will be combined with structural data to illuminate transport mechanisms, ligand binding, potential target for new antibacterial compounds, proteins structure and thermal stability. This study will also increase our understanding about mechanism of drug efflux proteins and other homologous proteins of therapeutic and/or biotechnological importance.

# Chapter-2

# Materials and Methods

## 2.1 Sources of materials

### 2.1.1 Bacteriological media and antibiotics

- Bacteriological tryptone, yeast extract and agar were obtained from Sigma-Aldrich, United Kingdom (UK).
- Tetracycline and chloramphenicol were from Sigma Chemicals, St Louis, United States of America (USA).
- Carbenicillin disodium was from Melford Laboratories Ltd., UK.

### 2.1.2 Reagents for bacterial gene cloning

- Restriction endonucleases, T4 DNA ligase, PCR product purification kit and DNA gel loading dye (6x) were purchased from New England Biolabs (NEB), UK.
- pfu Turbo DNA polymerase were from Agilent Technologies LDA, UK.
- dNTP Mix from Promega, USA.
- 1kb DNA ladder and DNA SYBR safe were from Invitrogen Ltd., USA.
- DNA purification kits were purchased from QIAgen Ltd., UK.
- Agarose for electrophoresis was from Life Technologies, UK.

### 2.1.3 Membrane preparations andprotein purification materials

- Imidazole, lysozyme, β-mercaptoethanol, ethylenediaminetetraacetate (EDTA), sucrose, tris base ultrapure, NaCl, glycerol and guanidine hydrochloride were purchased from Sigma-Aldrich., UK.
- Nickel-nitriloacetic acid (Ni-NTA) resin was from QIAGEN Ltd., UK.
- n-dodecyl-beta-D-maltoside (DDM) was from Melford laboratories, UK.
- Econo-pac 10DG desalting column was from BioRad Laboratories, UK.
- Vivaspin-20 100,000 and 30,000 MWCO and Vivaspin-6 100,000 and 30,000 MWCO tubes were from Sartorius Stedim Ltd., UK.

### 2.1.4 Filters and membranes

- Cellulose nitrate membrane filters (0.45 µm, 0.22 µm) were purchased from Whatman International Ltd., UK.
- Immobilion-P membrane PVDF (0.45 µm), Millex-GP syringe filter (0.22 µm) and polyethersulfone membrane were from Millipore Ltd., UK.

### 2.1.5 Protein molecular weight markers

- High range rainbow molecular weight marker RPN756E (12-225 kDa) was purchased from GE Healthcare Ltd., UK.
- Sigma SDS7 protein molecular weight marker was from Sigma-Aldrich, UK.

### 2.1.6 SDS-PAGE and western blotting

- Sodium dodecyl sulphate (SDS), ammonium persulphate, bovine serum albumin (BSA), brilliant blue R, glycine, β-mercaptoethanol, TEMED, TRIS and Tween 20 were purchased from Sigma-Aldrich, UK.
- Supersignal west picochemiluminescent substrate and HisProbe HRP was from Fisher Thermo Science Pierce Ltd., UK.
- Bis-acrylamide (2%) and acrylamide (40%) were from BioRad Laboratories, UK.
- Ethanol, methanol, isopropanol and glacial acetic acid were from Fisher Scientific Ltd., UK.

### 2.1.7 Circular dichroism, spectrophotofluorimetry, and uptake assay

- Choline chloride and dimethyl sulfoxide (DMSO) were purchased from Fisher Scientific Ltd., UK.
- $Na_2HPO_4$ and NaH2PO4 were from Melford Laboratories, UK.
- MOPS (3-[N-Morpholino] propane sulphonic acid), was from Calbiochem Ltd, UK.
- KCl, $K_2HPO_4$, $KH_2PO_4$ were from Jencons Scientific Ltd., UK.
- Scintillation fluid emulsifier safe was from Perkin Elmer, UK.
- Scintillation vials were from Sarstedt Stedim Ltd, UK.
- Vials squat with pp screw cap were from VWR International, UK.
- Chlorhexidine, spermidine, cadaverine, putrescine and acriflavine were from Sigma-Aldrich, UK.
- L-5-Benzyl hydantoin and D-5-benzyl hydantoin were synthesized by the Department of Chemistry, University of Leeds, UK.
- L-5-(4-brombutyl) hydantoin, L-5-(4-hydroxybutyl) hydantoin and L-5-(neo-pentyl) hydantoin were from Toronto Research Chemicals (TRC), Canada.
- All radiolabeled compounds were purchased from Perkin Elmer, UK.

## 2.2 General microbiology

### 2.2.1 Bacterial strains used in this research

The *E. coli* bacterial strains used in this research are detailed in Table 2.1

**Table 2.1 Bacterial strains used for expression.**

| Strain | Genotype | Source |
|---|---|---|
| BL21 (DE3) | *F- ompT hsdS$_B$ (r$^-_B$ m$^-_B$) gal dcm (DE3)* | Novagen™ |
| BL21 star (DE3) | *F- ompT hsdSB (rB - mB -) gal dcm rne131 (DE3)* | Invitrogen™ |
| BL21 gold (DE3) | *E. coli B F- ompT hsdS(rB - mB - ) dcm+ Tetr gal λ(DE3) endA Hte* | Stratagene™ |
| *C41 (DE3)* | *F$^-$ ompT gal dcm hsdS$_B$(r$_B^-$ m$_B^-$)(DE3)pLysS* | Lucigen™ |
| *C43 (DE3)* | *F$^-$ ompT gal dcm hsdS$_B$(r$_B^-$ m$_B^-$) (DE3) pLysS (Cm$^r$)* | Lucigen™ |
| OmniMAX | *Tet$^r$ Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F´ proAB lacI$^q$ZΔM15 Tn10 (Tet$^r$) Amy Cam$^r$]* | Stratagene™ |

### 2.2.2 Bacterial growth and storage media ingredients

Various compositions of media were used for the growth and storage of *E. coli* and expression studies and these are detailed in Table 2.2. Solutions were prepared using MilliQ water and sterilised by autoclaving at 15 Psi for 20 minutes at 121 ºC.

### 2.2.3 Preparation and use of antibiotic stock solutions

Stock solution of the antibiotics carbenicillin (100 mg/ml) and tetracycline (15 mg/ml) were prepared in MilliQ water and 70% ethanol, respectively, and were filter sterilised using 0.22 µm Millex-GP membrane filters and stored at -20 ºC. Final concentrations of carbenicillin and tetracycline in media were 100 µg/ml and 15 µg/ml, respectively.

**Table 2.2 Growth and storage media components**

| Media and buffers for bacterial growth and storage | Ingredients |
|---|---|
| Luria broth (LB) | 10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract, pH 7.4 |
| 2 x Tryptone / Yeast extract (2TY) | 10 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl, pH 7.4 |
| Minimal medium | 0.2% w/v casamino acids, 2 mM $MgSO_4.7H_2O$, 20 mM glycerol, 0.4 mM $CaCl_2.2H_2O$, 20 mM $NH_4Cl$, 1 x M9 salts (6 g/L $Na_2HPO_4$, 3 g/L $KH_2PO_4$, 0.5g/L NaCl) |
| Super optimal broth (SOB) | 20 g/L tryptone, 0.5 g/L NaC, 5 g/L yeast extract, 20 mM $MgSO_4$, pH 7.4 |
| Transformation buffer | 10 mM PIPES, 55 mM $MnCl_2$, 15 mM $CaCl_2.H_2O$, 250 mM KCl |
| Freezing mixture | 12.6 g/L $K_2HPO_4$, 0.9 g/L $Na_3C_6H_5O_7.2H_2O$, 0.18 g/L $MgSO_4$, 1.8 g/L $(NH_4)_2SO_4$, 3.6 g/L $KH_2PO_4$, 96.0 g/L glycerol |

### 2.2.4 Growth of bacteria for medium scale expression of proteins

From frozen stocks, bacteria were streaked on to LB agar plates (1.5%) containing carbenicillin (100 µg/ml) and incubated at 37 $^oC$ overnight. A single colony was inoculated into LB medium (100 ml) containing carbenicillin (100 µg/ml) and grown overnight at 37 $^oC$ for 16 hours, with 220rpm shaking. A 5 ml (1% inoculum) addition of the overnight culture was made to 10 flasks containing each 500 ml of LB media supplemented with 20 mM glycerol and carbenicillin (100 µg/ml). The cells were grown to an $OD_{680}$ of 0.6 and induced with appropriate IPTG concentration. After induction growth was continued for 3 hours before being harvested at 12000 x g for 20 minutes at 4 $^oC$. The cell pellets were resuspended in double volume of buffer (20 mM Tris-HCl pH 7.6, 10% v/v glycerol and 0.5 mM EDTA) and stored at - 80 $^oC$. Same procedure for fermentations (30-100 litre) were completed by Mr David Sharples.

### 2.2.5 Growth and preparation of cells for efflux assay

Frozen stocks of bacterial cells expressing the desired proteins were streaked on an LB-agar plate supplemented with carbenicillin (100 μg/ml). A single colony was transferred to LB medium (10 ml) containing carbenicillin (100 μg/ml) and grown for 16 hours at 37 ºC with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with IPTG (0.2 mM) and then grown for a further 1 hour. Harvested cells were washed three times using buffer (Tris 20mM, pH 7.0) and resuspended in the same buffer to an accurate $A_{680}$ of 1.0. Cells were loaded with 10 μM acriflavine in the presence of 10 μM CCCP and incubated for 30 minutes at 37 ºC. The loaded cells were again washed three times and resuspended in the same buffer.

## 2.3 Techniques for preparation and manipulation of recombinant DNA

### 2.3.1 Cloning strategy

In order to amplify protein expression, target genes were inserted into the multiple cloning site (MCS) of plasmid pTTQ18-$His_6$ immediately downstream of the *tac* promoter (Ward et al., 2001; Saidijam et al., 2003). Cutting of the DNA with two different restriction enzymes, *Eco*RI and *Pst*I, were performed to ensure correct orientation of the gene when ligated into plasmid pTTQ18-$His_6$, as well as preventing re-ligation of the plasmid. Membrane protein genes were amplified by PCR using bacterial genomic DNA as the template, followed by digestion with *Eco*RI and *Pst*I and ligation with *Eco*RI-*Pst*I digested pTTQ18-$His_6$. Plasmid constructs containing the relevant gene inserts were then transformed into the *E. coli* Omnimax strain. Antibiotic selections of cell colonies were followed by restriction digestion analysis and automated DNA sequencing of the isolated plasmids. Positive clones were transformed into *E. coli* BL21 (DE3) for optimisation of protein expression.

### 2.3.2 PCR primer design

Sequences of the genes of interest were obtained from the Transporter Protein Analysis Databse (http://www.membranetransport.org/) (Paulsen et al., 1998: Ren et al., 2006) or the UniProt KnowledgeBase (http://www.uniprot.org/). Restriction sites in the genes of interest were mapped using Webcutter 2, which allowed

checking for the presence of any *Eco*RI or *Pst*I restriction sites that are used for cloning. Only genes that did not contain internal *Eco*RI or *Pst*I sites were kept for cloning. Primers were designed to introduce in-frame *Eco*RI sites (GAATTC) at the 5` ends and *Pst*I sites (CTGCAG) at the 3` ends of the genes to allow subsequent ligation with *Eco*RI-*Pst*I digested pTTQ18-$His_6$ (Table 2.3).

Primers were designed with the following ideal properties where possible; a length of between 25 and 45 bases, a melting temperature ≥70 ºC, a minimum GC content of 40%, termination with a G or C base, absence of primer dimers or other secondary structures. Primer properties and quality were predicted using OligoAnalyzer 3.1 software.

**Table 2.3 Summary of DNA primers used for cloning in this project.** Forward and reverse primers are represented by (For) and (Rev), respectively. Restriction sites are coloured for *EcoR*I (red) and *Pst*I (green).

| Oligoname | Sequence |
|---|---|
| CUB18073_For | CCGGAATTCGCATATGTTGAAAGTAGAAAGGCGAACGATTG |
| CUB18073_Rev | AAAACTGCAGCTAGAGAAGACAGCTTTTCTTGATAACTCCG |
| EIQ13585_For | CCGGAATTCGCATATGGAACATCAGAGAAAACTATTCCAGC |
| EIQ13585_Rev | AAAACTGCAGCACCTATGGTTTTTTGCTCTCCTGTTTTTTCTG |
| CJK90608_For | CCGGAATTCGCATATGAATCACGACGGCTTTCAGATCAAGCAGATCGAC |
| CJK90608_Rev | AAAACTGCAGCCCCCAGCGGCGCGGGTTGTACCG |
| COI77568_For | CCGGAATTCGCATATGGAAAAGCAGTTTGGTCGC |
| COI77568_Rev | AAAACTGCAGCCTTCTTAATTAGCTTCGTTGCAACGTAG |
| CAC11736_For | CCGGAATTCGCATATGACGCATTCTACCGATATGAGC |
| CAC11736_Rev | AAAACTGCAGCGCTGATCTCTTTGCGAGTCTTTTCTCC |
| VPA1242_For | CCGGAATTCGCATATGGCTGGAGACAATAACTACAGTCTTGGACCAG |
| VPA1242_Rev | AAAACTGCAGCCGCTGGCTGTGTAGCCAGTACTTTTTTGTTGAG |
| NMB2067_For | CCGGAATTCGCATATGTCGGGCAATGCCTCCTCTCC |
| NMB2067_Rev | AAAACTGCAGCTGACGGGTTCCTTTGTAAAGATTGG |
| BAA80379_For | CCGGAATTCGCATATGGGTCTGGGCAATACCCCCGAG |
| BAA80379_Rev | AAAACTGCAGCCGTCTCGTTCTTGAACTTAACTTTAGCTGGGAC |
| AAN69889_For | CCGGAATTCGCATATGAGTAGCAGCCTCGACCTTGCCCCTG |
| AAN69889_Rev | AAAACTGCAGCATGGCCGCTGTGGTCGACGGCGATGCAC |
| EFQ62020_For | CCGGAATTCGCATATGCGTACAAGCCTGTCGAATGACC |
| EFQ62020_Rev | AAAACTGCAGCATGACTGACGTTATCGACTGCAATGCTTTCGC |
| ELQ14219_For | CCGGAATTCGCATATGAGCCATTCGACGCGTATCGAAACC |
| ELQ14219_Rev | AAAACTGCAGCCGAATGGGCGGTGGCTGAC |

### 2.3.3 Polymerase chain reaction (PCR)

PCR is an *in vitro* technique for amplifying specific chosen regions of DNA using a thermostable DNA polymerase enzyme by repeated rounds of DNA synthesis (Figure 2.1). PCR is simple and relatively cheap and involves a process of heating and cooling called thermal cycling. The reaction uses a DNA template containing the target sequence to be amplified. The DNA polymerase is an enzyme that creates new strands of DNA complementary to the target sequence. The first and most commonly used of these enzymes is Taq DNA polymerase (from *Thermis aquaticus*), whilst Pfu DNA polymerase (from *Pyrococcus furiosus*) is used widely because of its higher fidelity when copying DNA. Although these enzymes are slightly different, they both have two important capabilities making them suitable for PCR; they can make new strands of DNA using primers and a DNA template and they are thermostable. Other components in the PCR reaction are the primers, which are short stretches of single-stranded DNA that are complementary to the target sequence. The polymerase begins synthesizing new DNA from the end of the primer. The reaction also contains the four nucleotide bases dATP, dTTP, dGTP, and dCTP as "building blocks" for new DNA strands and a suitable buffer.

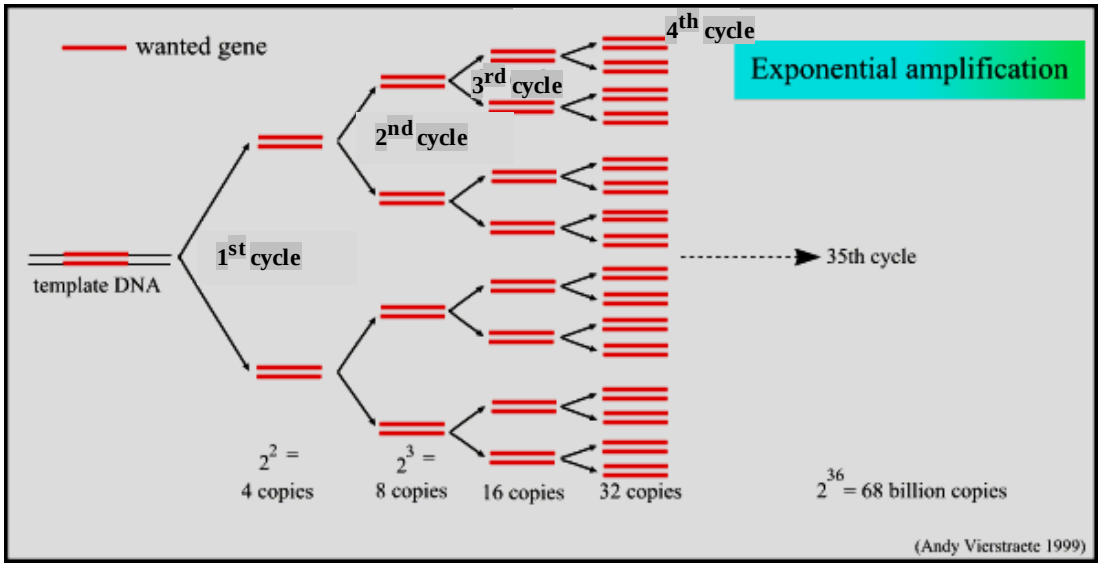



**Figure 2.1. Polymerase chain reaction principle.** This picture was reproduced from Vierstraete, 1999 (http://users.ugent.be/~avierstr/principles/pcr.html).

There are three main stages during the PCR procedure referred to as denaturation, annealing and extension (Figure 2.2). The denaturation stage involves application of high temperature (94-96 °C) to the original double-stranded DNA molecule to separate the strands from each other. This breaks the hydrogen bonds between the bases resulting in two single strands of DNA, which act as templates for the production of the new strands of DNA. The annealing stage involves binding of primers to the template strands of DNA. Cooling to a temperature of 50-65 °C promotes the formation of hydrogen bonds between the primers and specific sites on the template DNA dictated by their complementary sequences. The extension stage involves making a new strand of DNA by the Taq polymerase enzyme. During this stage the temperature is increased to 72 °C and the DNA polymerase extends the strands suing the nucleobases. These stages are repeated multiple times (30-40 cycles) to obtain multiple copies of the desired DNA sequence.

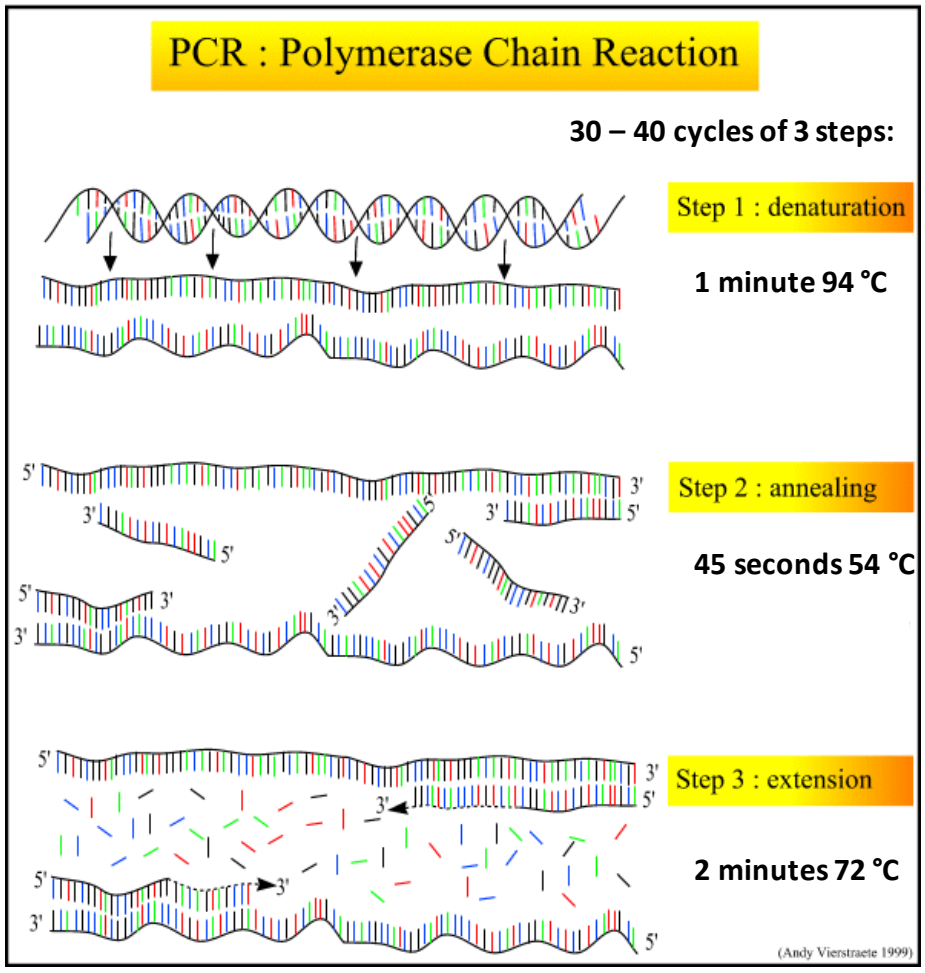



**Figure 2.2 Polymerase chain reaction stages.** This picture was reproduced from Vierstraete, 1999 (http://users.ugent.be/~avierstr/principles/pcr.html).

### 2.3.4 Isolation of genes from bacterial genomic DNA using PCR

PCR was performed using a Bio-rad thermocycler (Waltham, USA) with total reaction volumes of 50 µl by adding the reagents shown in Table 2.4 and using the conditions shown in Table 2.5. Primers were prepared in sterilised water (10 pmol/µl) and stored at -20 ºC.

**Table 2.4 PCR reaction components**

| PCR reaction components | Final concentration | Final volume |
|---|---|---|
| Sterile water | - | 39 µl |
| pfu Turbo buffer (10 x) | 1 x | 5 µl |
| dNTPs (10 mM each) | 0.2 mM | 1 µl |
| Genomic DNA (50 ng/µl) | 50 ng | 1 µl |
| Forward primer (10 µM) | 0.3 µM | 1.5 µl |
| Reverse primer (10 µM) | 0.3 µM | 1.5 µl |
| pfu Turbo polymerase (100 units) | 2.5 units | 1 µl |
| **Total volume** | | **50 µl** |

**Table 2.5 PCR reaction conditions**

| Steps | Temperature | Time | Number of cycles |
|---|---|---|---|
| Warm up | 95 | 3 minutes | x 1 |
| Denaturation | 95 | 30 seconds | x 30 |
| Annealing | 58 | 30 seconds | |
| Extension | 72 | 120 seconds | |
| Final extension | 72 | 10 minutes | x 1 |
| Hold | 4 | ∞ | x 1 |

### 2.3.5 Plasmid DNA preparation

Plasmid DNA was isolated from bacterial overnight cultures and purified using a QIAprep Spin Miniprep Kit and protocol from QIAGEN Ltd., UK. Isolated plasmid DNA was stored at a temperature of -20 ºC.

### 2.3.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used to separate, analyse, identify and purify DNA fragments and to check the progression of restriction enzyme digestion. The rate of DNA migration through the agarose gel is dependent upon the agarose concentration, the molecular size of the DNA, the conformation of the DNA and the applied current. Prior to gel casting, dried agarose (0.5 g) was dissolved in 1 x TAE buffer (50 ml) (Table 2.6) by heating in a microwave (30-45 seconds) and the solution was cooled to 60 ºC before pouring into the mould. The percentage of agarose used in all gels was 1%. Cyber safe dye (1 µl) was added to the agarose gel matrix before pouring into the mould for staining the DNA to visualize under ultraviolet light. DNA samples were prepared by mixing with the appropriate volume of 6x DNA loading dye and loaded into wells of the gel. The gel was run in 1x TAE buffer at 100 V for 30 minutes to achieve separation of the DNA fragments.

**Table 2.6 Composition of 50X Tris Acetate EDTA (TAE) buffer**

| Component | Amount |
|---|---|
| Tris base pH 8.0 | 242 g |
| Glacial acetic acid | 57.1 ml |
| 0.5 M EDTA pH 8.0 | 100 ml |
| Made up to 1 litre with MilliQ water | |

### 2.3.7 Size and concentration of DNA

The sizes of DNA bands on agarose gels were determined approximately by running samples alongside a 1 Kb DNA ladder (Table 2.7). DNA concentrations were determined by measuring the absorbance at 260 nm and 280 nm using a Nano-drop spectrometer. The DNA concentration was calculated using the rule that 1 mg/ml of DNA has an $A_{260nm}$ of 20. For a pure DNA sample the $A_{260nm}$ / $A_{280nm}$ = 1.8 whilst values above 1.8 suggest RNA contamination and below 1.8 suggest protein contamination. A ratio of $A_{260nm}$ / $A_{280nm}$ therefore predicts the purity of the DNA sample.

**Table 2.7 DNA molecular weight markers used for the determination of DNA size.**

| Fragments | 1 kb DNA ladder (Invitrogen™) (1 µl = 50 ng) | |
|---|---|---|
| | Base pairs | Amount (ng) |
| 1 | 12216 | 10 |
| 2 | 11198 | 7 |
| 3 | 10180 | 7 |
| 4 | 9162 | 7 |
| 5 | 8144 | 7 |
| 6 | 7126 | 7 |
| 7 | 6108 | 6 |
| 8 | 5090 | 6 |
| 9 | 4072 | 6 |
| 10 | 3054 | 6 |
| 11 | 2036 | 6 |
| 12 | 1636 | 10 |
| 13 | 1018 | 5 |
| 14 | 506 | 3 |

### 2.3.8 Isolation of DNA fragments from agarose gels

DNA fragments were isolated from agarose gels using the QIAgen gel extraction kit and protocol. Following DNA agarose gel electrophoresis, the region of gel containing the DNA of interest was excised with a clean sharp scalpel. The gel slice was weighed in a colourless tube and three volumes of the binding and solubilisation buffer QG was added to 1 volume of gel (i.e. ~300 µl added to 100 mg). The mixture was incubated at 50 ºC for 10 minutes until the gel slice had completely dissolved and for DNA fragments (<500 bp and >4 kb) one gel volume of isopropanol was added to the sample and mixed. To bind DNA, the sample was applied to a spin column and centrifuged for 1 minute at 13000 rpm in a micro centrifuge. Buffer QG (0.5 ml) was added to wash the sample, which was then centrifuged for 1 minute. The column was washed with buffer PE (low salt wash buffer and ethanol) (0.75 ml) and then centrifuged for 1 minute. The column was placed in a clean Eppendorf tube, and the DNA was eluted by addition of elution buffer EB (10 mM Tris.HCl, pH 8.5)

(50 µl) to the centre of the QIAquick membrane and centrifugation of the column for 1 min at 13000 rpm.

### 2.3.9 Restriction digestion of DNA

In order to obtain DNA fragments (genes and plasmid DNA) with the desired restriction sites at their ends for successful ligations it was necessary to cut the DNA with the relevant restriction enzymes. The enzymes used were *Eco*RI and *Pst*I and the reaction conditions for digestion are shown in Table 2.8. Reactions were incubated at 37 °C for 1 hour and then enzymes were deactivated by incubating at 80 °C for 10 minutes.

**Table 2.8 Conditions used for restriction digestion**

| **Reaction component** | **Volume added** |
|---|---|
| DNA | 10.5 µl (3 µg) |
| *Eco*RI | 1.5 µl (3 units) |
| *Pst*I | 1.5 µl (3 units) |
| Restriction enzyme buffer (Buffer 3) | 2.5 µl |
| MilliQ water | 9 µl |
| Incubation temperature and time: 37 ºC for 1 hour | |

### 2.3.10 Ligation of gene insert and plasmid

Ligation of the genes of interest into plasmid DNA was catalysed by T4 DNA ligase. The gene insert to plasmid vector molar ratio has significant effects on the outcome of ligation and subsequent transformation. Ligation reactions were therefore performed using various plasmid to gene molar ratios of 1:1, 1:3 and 1:5 (Table 2.9). Two controls were also conducted, the first contained plasmid DNA and gene insert but did not contain T4 DNA ligase (reaction 4), whilst the second contained plasmid DNA and T4 DNA ligase but no gene insert (reaction 5). The purpose of these controls was to evaluate whether there was any self-ligated plasmid or undigested plasmid DNA which would confer carbenicillin resistance on the transformed bacteria, but would not contain the gene of interest. The reaction mixture was

incubated at 16 °C overnight and the reaction was terminated by incubating at 65 °C for 10 minutes before performing transformation.

**Table 2.9 Ligation reactions using various ratios of plasmid and gene insert**

| | **Ligation reaction mixture** | | | | |
|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** |
| **pTTQ18** | 5 ng/µl | 5 ng/µl | 5 ng/µl | 5 ng/µl | 5 ng/µl |
| **Gene** | 5 ng/µl | 15 ng/3µl | 25 ng/5µl | 15 ng/3µl | no gene |
| **DNA T4 ligase** | 1 µl | 1 µl | 1 µl | no ligase | 1 µl |
| **10 x T4 ligase buffer** | 1.5 µl | 1.5 µl | 1.5 µl | 1.5 µl | 1.5 µl |
| **Sterile water** | 10.5 µl | 8.5 µl | 6.5 µl | 9.5 µl | 11.5 µl |
| **Total volume** | 15 µl | 15 µl | 15 µl | 15 µl | 15 µl |

### 2.3.11 Preparation of competent cells

*E. coli* competent cells were prepared using the method of Inoue et al., (1990). LB agar plates containing carbenicillin (100 µg/ml) were incubated overnight and 10 colonies were used to inoculate SOB medium (250 ml) in a 1 litre baffled flask. This was incubated at 18 $^{o}$C with shaking at 220 rpm until the $OD_{600}$ reached ~0.6. Cells were placed on ice for 10 minutes before sedimentation (2500 x g, 10 minutes, 4 $^{o}$C). Cells were resuspended in ice-cold transformation buffer (80 ml) (Table 2.2) using pre-cooled pipette tips and left on ice for 10 minutes. The cells were sedimented again (2500 x g, 10 minutes, 4 $^{o}$C) and then gently resuspended in ice-cold transformation buffer (20 ml). DMSO was added to a final concentration of 7% (v/v) and the cells were left on ice for 10 minutes before dispensing in 50 µl aliquots, snap freezing in liquid nitrogen and storage at -80 $^{o}$C.

### 2.3.12 Transformation of competent cells

Omnimax competent cells (50 µl) were thawed on ice prior to the addition of 5 µl (10-15 ng) of ligation mixture and incubated on ice for 30 minutes. The cells were

heat shocked at 42 °C for 45 seconds, followed by cooling on ice for 2 minutes. LB medium (450 µl) was added and the cells were incubated at 37 °C with shaking at 220 rpm for 1 hour to allow expression of the antibiotic resistance gene. The cells were sedimented by centrifugation (3000 x g) for 120 seconds and resuspended in LB medium (200 µl). An aliquot of the resuspended cells (100 µl) was streaked onto an LB agar plate containing carbenicillin (100 µg/ml) and incubated at 37 °C overnight.

### 2.3.13 Plasmid DNA sequencing and alignment

Plasmid DNA was sequenced by Beckman Coulter Genomics (BCG, UK) using minimum volumes of 15 µl at 100 ng/µl concentration. The clone or mutant of interest was identified and confirmed by determination of the DNA sequence alignment. For mutant constructs, the DNA sequence of the gene insert was aligned with the DNA sequence of the wild-type gene insert using the multiple alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Also for reconfirmation the wild-type and mutant gene insert DNA sequence were translated using the Expasy Translate tool (http://web.expasy.org/translate/). The amino acid sequence of the mutant was again aligned against the amino acid sequence of the wild-type gene to confirm presence of the single point mutation. Alignments were important for validation of the site directed specific mutation and avoiding any undesirable mutation. After confirmation of each mutant identity by sequencing, each plasmid was transformed into *E. coli* strain BL21 (DE3) cells for further studies.

### 2.3.14 Preparation of transformed *E. coli* BL21(DE3) cell deeps

Frozen stocks of BL21(DE3) cells containing plasmid DNA were prepared by transferring a single colony of cells into LB medium (10 ml) containing carbenicillin (100 µg/ml). Cells were cultured for 6-8 hours at 37 ºC with shaking at 220 rpm until an $OD_{680}$ of 1-1.5 was reached. Aliquots (0.75 ml) of the culture were added to 0.75 ml of freezing mixture (Table 2.2), snap frozen in liquid nitrogen and stored at -80 ºC.

### 2.3.15 Site directed mutagenesis

Site-specific mutants of the Mhp1 protein (D229N, D229A and D229E) were generated using a Q5 site-directed mutagenesis kit and quick protocol from New England Biolab Company. Mutant primers with the desired base change (Table 2.10) were designed using the New England Biolabs online software (NEBaseChanger.neb.com.). The pTTQ18 plasmid containing the wild type Mhp1 gene insert, was used as the template for the PCR reaction. The mutated Mhp1 gene insert was created and amplified using PCR with the reaction conditions given in Tables 2.11 and 2.12.

**Table 2.10 Primers used for site directed mutagenesis**

| **Mutant** | **Primer** | **Base change** |
|---|---|---|
| **D229N** | 5'GAGCATCCACaacATCGTGAAGGAG3'<br>5'ACCACGACCGCGATCCAG3' | GAC $\rightarrow$AAC |
| **D229A** | 5'GAGCATCCACgcaATCGTGAAGGAGG3'<br>5'ACCACGACCGCGATCCAG3' | GAC $\rightarrow$GCA |
| **D229E** | 5'GAGCATCCACgagATCGTGAAGG3'<br>5'ACCACGACCGCGATCCAG3' | GAC $\rightarrow$GAG |

**Table 2.11 Reaction components for site-directed mutagenesis PCR**

| **PCR reaction component** | **25 µl RXN** | **Final** |
|---|---|---|
| Template DNA (1–25 ng/µl) | 1 µl | 1-25 ng |
| 10 µM Forward primer | 1.25 µl | 0.5 µM |
| 10 µM Reverse primer | 1.25 µl | 0.5 µM |
| Nuclease-free water | 9.0 µl | - |
| Q5 Hot start high-fidelity 2X master mix | 12.5 µl | 1x |

**Table 2.12 PCR conditions used to create the desired mutant plasmids**

| **Step** | **Temperature (° C)** | **Time** |
|---|---|---|
| Initial denaturation | 98 | 30 seconds |
| 25 cycles | 98 | 10 seconds |
| | 67 | 10-30 seconds |
| | 72 | 20-30 seconds |
| Final extension | 72 | 120 seconds |
| Hold | 4-10 | ∞ |

### 2.3.16 Kinase, ligase and *DpnI* (KLD) treatment

KLD treatment of PCR products allows efficient phosphorylation, intramolecular ligation/circularization and digestion of the parental methylated wild-type dsDNA template plasmid. The reagents given in Table 2.13 were gently mixed using a pipette and incubated at room temperature for 5 minutes.

**Table 2.13 Reaction components for Kinase, ligase and *DpnI* treatment**

| Reaction component | Volume | Final |
|---|---|---|
| PCR product | 1 µl | |
| 2 x KLD reaction buffer | 5 µl | 1 x |
| 10 x KLD enzyme mixture | 1 µl | 1 x |
| Nuclease free water | 3 µl | |

### 2.3.17 Transformation of competent cells with mutant constructs

NEB 5-alpha competent *E. coli* cells, supplied with the Q5 site-directed mutagenesis kit, were transformed in order to amplify the plasmid PCR products prior to sequencing. Aliquots of NEB 5-alpha competent cells (50 µl) were thawed on ice prior to the addition of the KLD treated plasmid DNA (5 µl). The cells were then incubated on ice for 30 minutes before heat shocking at 42°C for 45 seconds and then incubated for a further 5 minutes on ice. LB medium (450 µl) was added followed by incubation at 37ºC for 1 hour with shaking at 220 rpm. Transformed cells (50-100 µl) were plated onto LB/carbenicillin (100 µg/ml)/agar plates and incubated overnight at 37ºC. Plasmid DNA isolated was sequenced by Beckman Coulter Genomics (BCG, UK) using minimum volumes of 15 µl at 100 ng/µl concentration. The successfully sequenced gene inserts containing the desired point mutation were transformed into *E. coli* BL21 (DE3) cells. The desired mutation was confirmed by aligning the mutant gene DNA sequence with the wild-type gene DNA sequence (Appendix 1).

## 2.4 Techniques for preparation and manipulation of membrane Protein

### 2.4.1 Preparation of *E. coli* mixed membranes by the water lysis method

Small scale membranes were prepared by the water lysis method (Witholt et al., 1976; Ward et al., 2000) from 50 ml cultures of *E. coli* BL21(DE3) cells harbouring the appropriate pTTQ18-based construct. Two batches of cells were grown to $OD_{600}$ = 0.8 in LB medium supplemented with the antibiotic carbenicillin (100 µg/ml). One batch was induced with 0.2 mM IPTG while the other was left uninduced and incubation was continued for a further 2 hours. Cells were harvested at 4500 x g for 10 minutes at 4ºC. The cells were resuspended in 10 ml 0.2 M Tris-HCl (pH 8.0) and were shaken for 20 minutes. At time zero, 4.85 ml of 1 M sucrose; 0.2 M Tris-HCl pH 8.0; 1 mM EDTA was added followed by the addition of 65 µl of 10 mg/ml lysozyme solution at 1.5 minutes. At 2 minutes the reaction was terminated with the addition of 9.6 ml MilliQ water and left stirring for 20 minutes. The resultant spheroplasts were sedimented at 45000 x g for 20 minutes at 4ºC and were resuspended and disrupted in 15 ml MilliQ water using a homogeniser. A further sedimentation step (45000 x g for 20 minutes at 4ºC) followed for the fractionation of the cytoplasm (supernatant), and the generation of mixed cell membranes (pellet). Membrane preparations were washed three times with 30 ml 0.1 M sodium phosphate buffer pH 7.2, 1 mM β-mercaptoethanol. The final pellet was resuspended in 300 or 500 µl of the above sodium phosphate buffer and stored at -80ºC.

### 2.4.2 Preparation of inner and outer membrane fractions from large-scale cultures

Inner and outer membrane preparation process was accomplished in two stages. The first stage involved the disruption of cells using a cell disrupter and the second stage involved sucrose gradient ultracentrifugation to separate the inner and the outer membrane fractions.

**2.4.2.1 Cell disruption.** *E. coli* cells obtained from large-scale fermenter cultures (30-100 litres) were stored at -80 °C in buffer (20 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 10% v/v glycerol). Cells were thawed gradually at 4 °C with gentle mixing and resuspended in disruption buffer (20 mM Tris-HCl pH 7.5 containing 0.5 mM EDTA). Usually 3 or 4 ml of disruption buffer was used for 1 g (wet weight) of bacterial cells. The suspension was stirred for 1 minute using an ultra tarex

homogeniser and thawed cells, which typically had a yoghurt-like consistency, were passed through a needle and syringe to ensure that the cells were fully defrosted and completely homogenised. The cell disrupter was operated in accordance with the manufacturer's instructions. The cell disruptor was switched on and cooled to an operating temperature of 4 °C, washed thoroughly with $dH_2O$ (200 ml) and then with cell disruption buffer (100 ml). The cells were disrupted by explosive decompression by being passed through the cell disrupter twice at a pressure of 30 kpsi and were collected in a 500 ml conical flask pre-cooled on ice. Undisrupted cells and cell debris were sedimented and removed by centrifugation at 12000 x g for 45 minutes at 4 °C. The supernatant containing total/mixed membranes was collected and retained.

**2.4.2.2 Separation of inner and outer membrane fractions by sucrose density gradient ultracentrifugation.** Mixed membranes were collected from the supernatant of disrupted cells by ultracentrifugation at 131000 x g for 2 hours at 4 $^{o}$C. The membranes were resuspended in an appropriate volume, typically 5 ml, of 25% w/v sucrose dissolved in 20 mM Tris-HCl pH 7.6, 0.5 mM EDTA. Six sucrose step gradients were prepared by layering 10 ml quantities of 55, 50, 45, 40, 35 and 30% w/v sucrose solution starting with 55% w/v solution at the bottom and finishing with 30% at the top of 70 ml ultracentrifuge tubes. This dispensing was done very carefully with a needle and syringe slowly to avoid mixing of layers. The mixed membrane suspension (4 or 6 ml) was layered gently on top of the sucrose gradients held on ice. The inner and outer membranes were separated by ultracentrifugation at 131000 x g for 16 hours at 4 °C with minimal acceleration and no braking in a Beckman ultracentrifuge. After centrifugation a dense golden inner membrane band appeared at the 35-40% interface and a paler outer membrane band at the 45-50% interface. The inner and outer membrane fractions were carefully drawn off using a clean needle and syringe and resuspended in 20 mM tris-HCl pH 7.5. The membranes were then washed three times using 20 mM tris-HCl pH 7.5 by ultracentrifugation at 131000 x g for 1 hour at 4 °C to remove the sucrose. The resulting membrane pellet was resuspended in 2-5 ml of 20 mM tris-HCl pH 7.5 and dispensed into 0.5 ml Eppendorf tubes, which were frozen in liquid nitrogen and stored at -80 $^{o}$C. The expression level of the protein of interest in each of these fractions was analysed by SDS-PAGE and Western blotting.

### 2.4.3 Protein purification using immobilised metal affinity chromatography (IMAC)

Effective solubilisation using a detergent was an important first step in the purification of integral membrane proteins. Inner membrane preparations from 30 or 100 litre fermentations with the desired protein expressed were resuspended in solubilisation buffer (Table 2.14) and mixed for 2-3 hours at 4 °C. The membranes were then sedimented using Ti-45 centrifuge tubes at 131000 x g at 4 °C for 1 hour to remove the insoluble part. The supernatant containing the soluble fraction was incubated with Ni-NTA resin at 4 °C for 16 hours with mixing. The Ni-NTA supernatant mixture was run through a BioRad column to elute unbound components. The Ni-NTA resin was washed with 150-200 ml of wash buffer 1 (Table 2.15) to remove any remaining unbound constituents. The bound $His_6$-tagged protein was removed by the addition of elution buffer (Table 2.16). Volumes of eluted samples were reduced to 3 ml using Vivaspin 20 tube concentrators (4000 xg) with molecular weight cut off (MWCO) 100 kDa or 30 kDa for NCS1 and PACE family proteins, respectively. To remove the high concentration of imidazole, 3 ml sample was then applied to a BioRadEcono-pac 10 DG desalting column. Wash buffer-2 (5 ml; Table 2.17) was applied to the column and the eluted fraction was collected in a Vivaspin 6 tube MWCO 30 kDa or 100 kDa and spun at 4500 x g. The purified eluted fraction that had been concentrated to 4-20 mg/ml was flash frozen in liquid nitrogen and stored at -80 °C.

**Table 2.14 Solubilisation buffer for 60 ml**

| Stock solutions | Final concentration | Volume added |
|---|---|---|
| 1 M Tris-HCl pH 8 | 10 mM | 600 µl |
| 1 M Imidazole pH 8 | 20 mM | 1.2 ml |
| 50% Glycerol | 20 % | 24 ml |
| 4 M NaCl | 0.3 M | 4.5 ml |
| 10% DDM | 1 % | 6 ml |
| MilliQ water | - | 23.7 ml |

**Table 2.15 Wash buffer-1 for 250 ml.**

| Stock solutions | Final concentration | Volume added |
|---|---|---|
| 1 M Tris-HCl pH 8 | 10 mM | 2.5 ml |
| 1 M Imidazole pH 8 | 20 mM | 5 ml |
| 50% Glycerol | 10 % | 50 ml |
| 10% DDM | 0.05 % | 1.25 ml |
| MilliQ water | - | 191.25 ml |

**Table 2.16 Elution buffer for 50 ml**

| Stock solutions | Final concentration | Volume added |
|---|---|---|
| 1 M Tris-HCl pH 8 | 10 mM | 500 µl |
| 1 M Imidazole pH 8 | 200 mM | 10 ml |
| 50% Glycerol | 2.5 % | 2.5 ml |
| 10% DDM | 0.05 % | 0.25 ml |
| MilliQ water | | 36.75 ml |

**Table 2.17 Wash buffer-2 for 250 ml**

| Stock solutions | Final concentration | Volume added |
|---|---|---|
| 1 M Tris-HCl pH 8 | 10 mM | 2.5 ml |
| 50% Glycerol | 2.5 % | 12.5 ml |
| 10% DDM | 0.05 % | 1.25 ml |
| MilliQ water | - | 233.75 |

### 2.4.4 Determination of protein concentration

The Schaffner and Weissmann protein assay (1973) was used to determine the concentration of purified protein. It is based on the principle that trichloroacetic acid (TCA) quantitatively precipitated protein and filtered on to membranes and washed to remove chromogenic contaminants. A calibration graph was created using a stock of bovine serum albumin (0.1 mg/ml) in the range of 0-20 µg (Figure 2.3). Each protein sample was made up to 270 µl with MilliQ water followed by 30 µl of 1M

Tris-HCl pH 7.6, 60 µl w/v TCA (60%) and w/v SDS (10%). The precipitated protein samples were transferred drop wise onto 0.45 µm HAWP Millipore membrane filter. Filters were then washed twice with 6% w/v TCA (2 ml). Filters were stained in a petri dish containing 30 ml volumes of 0.1% naphthalene black in 45:10:45 v/v methanol/acetic acid/water for 3 minutes followed by washes in MilliQ water for 30 seconds, destain (90:2:8 v/v methanol/acetic acid/water) for 1 minute, and finally MilliQ water for 1 minute before drying. Each coloured spot was cut out and transferred to a 1.5 ml tube. By addition of 1 ml elution buffer (0.05 mM EDTA, 25 mM NaOH, 50% aqueous ethanol), eluted the stained protein and incubated for 30 minutes at 37 ºC. A WPA Biowave II UV/Visible spectrophotometer (Biochrom) was used to measure $OD_{630}$ of samples and concentrations were calculated with reference to a calibration curve from BSA (Figure 2.3).

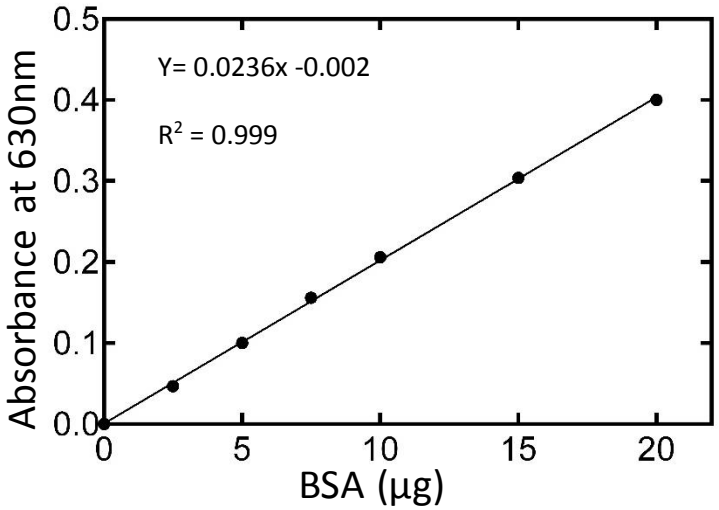



**Figure 2.3 A typical calibration curve using bovine serum albumin (BSA).**

### 2.4.5 Separation and analysis of proteins by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

The method was based on the procedure of Laemmli (1970), using the gel composition defined by Henderson and Macpherson (1986). The purity of the protein samples was assessed using SDS-PAGE gel. The running gel (15% v/v) was poured first and left to set for about 40 minutes and then the stacking gel (4% v/v) (Table 2.18). For standard SDS-PAGE analysis, protein (16 µg) samples were solubilised in 4x sample loading buffer and incubated at 37 °C for 30 minutes. Samples

electrophoresis was performed at 120V for 80-90 minutes. Coomassie brilliant blue was used for staining the gel for 2-3 hours, followed by destaining until the migrated protein bands were appropriately visualized. The molecular weight markers are shown in Tables 2.19 and 2.20.

**Table 2.18 Compositions of 15% resolving and 4% stacking gels for SDS-PAGE**

| **Stock solutions** | **Resolving gel (15%)** | **Stacking gel (4%)** |
|---|---|---|
| 40% acrylamide | 4.22ml | 0.77ml |
| 2% bis-acrylamide | 0.49ml | 0.39ml |
| 1.5M TrisHCl (pH8.8) | 2.81ml | - |
| 1.5M TrisHCl (pH6.8) | - | 0.75ml |
| 10% w/v SDS | 100µl | 50µl |
| Deionised water | 3.48ml | 3.20ml |
| 10% (w/v) ammonium persulphate | 37µl | 30µl |
| TEMED | 12µl | 9µl |

**Table 2.19 SDS-PAGE SDS 7 molecular weight markers.**

| **Protein** | **Band number** | **Molecular weight (kDa)** |
|---|---|---|
| Bovine serum albumin | 1 | 66.0 |
| Ovalbumin | 2 | 45.0 |
| G-3-P dehydrogenase | 3 | 36.0 |
| Carbonic anhydrase | 4 | 29.0 |
| Trypsinogen | 5 | 24.0 |
| Trypsin inhibitor | 6 | 20.1 |
| Alpha-lactalbumin | 7 | 14.2 |

**Table 2.20 High-range rainbow molecular weight markers.**

| **Band number** | **Molecular weight (kDa)** |
|---|---|
| 1 | 225.0 |
| 2 | 76.0 |
| 3 | 52.0 |
| 4 | 38.0 |
| 5 | 31.0 |
| 6 | 24.0 |
| 7 | 17.0 |
| 8 | 12.0 |

A linear relationship was observed between the distance migrated and the log of molecular weight standards for the SDS7 molecular weight markers (Figure 2.4).

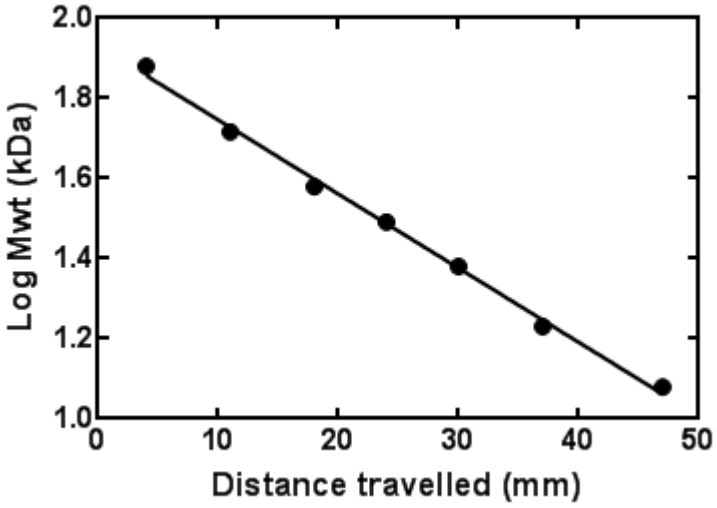



**Figure 2.4 A typical log molecular weight-distance plot for determining the size of proteins resolved by SDS-PAGE.** The plot represents distances migrated by SDS7 molecular weight markers.

### 2.4.6 Detection and visualisation of the $His_6$-tagged proteins by Western blotting

To identify the target proteins Western blotting is one of the most widely used and sensitive analytical techniques that exploits antibodies to detect and identify the target protein (Sambrook et al., 1989). The samples containing protein (4 µg) were first separated on the gel and were transferred to a membrane using a Biorad Trans-Blot semi-dry transfer cell operating at 18 volts for 35 minutes. Four pieces of filter paper were pre-soaked in 0.5 x SDS-PAGE running buffer (Towbin et al., 1979). Two pieces of filter paper followed by the membrane, the polyacrylamide gel and two further pieces of filter paper were layered onto one another. The membrane was incubated with BSA (3%) in TBST (20 mM Tris-HCl pH 7.6, 0.05% v/v Tween-20, 0.5M NaCl) for 3 hour at 4 ºC to block non-specific binding sites. The membrane was washed twice with 20 ml of TBST at room temperature for 10 minutes. The membrane was then incubated for 1 hour with HisProbe-HRP (10 ml) diluted to 1:5000 with TBST followed by three washes with 20 ml TBST for 10 minutes each.

A 6 ml SuperSignal West Pico chemiluminescent solution was prepared by mixing 3 ml West Pico luminol/enhancer solution and 3ml West Pico stable peroxide solution and membrane was incubated for 3 minutes before wrapping in acetate paper for exposure using the Syngene G:Box. Rainbow marker was used to provide an estimate of molecular weight of the detected protein band.

## 2.5 Biochemical and biophysical techniques to characterise membrane proteins

### 2.5.1 Circular dichroism (CD) spectroscopy of the proteins

The secondary structure and thermal stability of purified proteins was investigated using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant nitrogen flushing. CD buffer (10 mM NaPi pH 7.5, 0.05% DDM) alone was measured and subsequently subtracted from all CD data although the signal produced by the buffer alone was almost negligible. The samples were analysed in a Hellma quartz cuvette with a 1.0 mm pathlength. The secondary structure in 0.15 mg/ml purified protein was analysed from 180-260 nm at 20 ºC. The changes in secondary structure were monitored at 222 nm or 209 nm. The thermal stability of purified proteins was also analysed by ramping the temperature from 5-90 ºC and finally back to 5 ºC. The desired temperature was held for 60 seconds at each increment before a scan was completed. Data in which the high tension (HT) voltage was below 600 V were only analysed as data became defective when the HT voltage raises much above this value due to the high light absorbance of the sample (Kelly et al., 2005).

### 2.5.2 Steady-state spectrophotofluorimetry and ligand titrations

Steady-state spectrophotofluorimetry was undertaken on purified proteins in order to analyse ligand binding using a method based on that described by Ward et al., (2000). This exploited the intrinsic fluorescence properties of tryptophan residues in the proteins and a quench in its intensity caused by the binding of ligands. Purified NCS-1 protein (2.5 µM) or PACE family protein (4.4 µM) in fluorescence buffer (1 ml) (Table 2.21) were analyzed using a Photon Technology International spectrophotofluorimeter in the presence and absence of 15 mM NaCl and stirred at 18 °C. For ligands that were not soluble in 100% fluorescence buffer, a ligand stock

solution in 100% DMSO was used, giving a final DMSO concentration of no more than 2% in the assay sample. Fluorescence was excited at the wavelength that is absorbed by tryptophan residues (295 nm) and emission monitored over the wavelength range that is emitted during fluorescence (310-360 nm). In order to allow direct comparisons between measurements, slit widths of 0.6 nm and a step time of 0.5 sec were used for all measurements. Titrations of ligands at a range of concentrations (0-2 mM for NCS-1 proteins, 0-40 µM for PACE family proteins) were performed by additions of appropriate stock solutions into the sample. Following each ligand addition, samples were mixed for 1.5 minutes before the fluorescence emission was monitored. Data were analyzed to obtain apparent binding affinities ($K_d$ values) using the Michaelis-Menten analysis or nonlinear regression tool in GraphPad Prism 7 software.

**Table 2.21 Buffer for steady-state spectrophotofluorimetry**

| | **A- No added $Na^+$** | | **B- With added $Na^+$** | |
|---|---|---|---|---|
| **Stock solution** | **Final concentration** | **Volume** | **Final Concentration** | **Volume** |
| 1M Choline chloride | 140 mM | 700 µl | 125 mM | 625 µl |
| 1 M tris pH7.5 | 10 mM | 50 µl | 10 mM | 50 µl |
| 10% DDM | 0.05% | 25 µl | 0.05% | 25 µl |
| DMSO | 2% | 100 µl | 2% | 100 µl |
| 4 M NaCl | - | - | 15 mM | 18.75 µl |
| MilliQ water | - | 4.125 ml | - | 4.181 ml |
| Total | | 5 ml | 5 ml | |

### 2.5.3 Stopped-flow spectrophotofluorimetry to measure time dependence of ligand binding to purified proteins

The time-dependence of ligand binding to purified proteins following rapid mixing was measured by stopped-flow spectrophotofluorimetry using a SX.20MV-R stopped-flow reaction analyzer from Applied PhotoPhysics. Time courses for the change in tryptophan fluorescence from 0-2 seconds following mixing of protein and ligand were measured using an excitation wavelength of 280 nm, a cut off filter of

330 nm and a maximum PMT value of 330 V. The fluorescence quench caused by mixing of ligand (0-2 mM for NCS1 proteins) dissolved in 100% DMSO with purified NCS1 protein (2.5 µM) or PACE family protein (4.4 µM) in fluorescence buffer (Table 2.22) was monitored in the presence and absence of 15 mM NaCl at 18 $^{o}$C. The mixing dead time was less than 5 ms and each binding curve represented an average of at least four experimental runs. Data were analyzed using OriginPro and GraphPad Prism 7 software.

**Table 2.22 Buffer for stopped-flow spectrophotofluorimetry**

| **Stock solution** | **Final concentration** | **Volume of stock solution required for 100 mL** |
|---|---|---|
| 1M Tris pH 7.5 | 10 mM | 20 mL |
| 1 M choline chloride | 120 mM | 12 mL |
| 10% DDM | 0.05% | 0.5 mL |
| 50% glycerol | 2.5% | 5 mL |
| 100% DMSO | 2% | 2 mL |
| Adjust pH to 7.5 and make to 100 mL with milliQ water | | |

## 2.6 Assay for uptake of radioisotope-labelled substrate into energised whole cells

### 2.6.1 Growth and preparation of cells

Frozen stocks of *E. coli* BL21(DE3) cells expressing the desired proteins were streaked on an LB-agar plate supplemented with carbenicillin (100 µg/ml). A single colony was transferred to LB medium (10 ml) containing carbenicillin (100 µg/ml) and grown for 6 hours at 37 $^{o}$C with shaking at 220 rpm. A 1% inoculum was transferred to two cultures of LB medium (50 ml) containing glycerol (20 mM) and carbenicillin (100 µg/ml) and grown at 37 $^{o}$C with shaking at 220 rpm until the absorbance at $A_{600}$ reached 0.8. One culture was induced with IPTG (0.2 mM) while the other was left uninduced. Growth was continued for a further 1 hour under the same conditions. The cells were harvested at 4500 x g for 10 minutes at 20 $^{o}$C. Harvested cells were washed three times by resuspension and sedimentation (4500 x

g, 10 minutes, 20 $^{o}$C) using buffer (Table 2.23). The washed cells were then resuspended in the same buffer to a final concentration of $A_{680}$ = 2.

### 2.6.2 Incubation and sampling procedure

Once the concentration of cells was adjusted, an aliquot of cells (935 µl) was transferred to a bijou bottle, mixed with 20 mM glycerol and incubated at 25 ˚C for 3 minutes with constant bubbled aeration throughout the assay. At 3 minutes, radioactive substrate (50 µM) was added and samples of 100 µl were taken at time points of 0.25, 1, 2, 5, 7.5 and 10 minutes. Samples were immediately applied to cellulose nitrate filters (0.45 µm pore size) on a vacuum manifold already pre-soaked in same buffer. Filters were immediately washed with the same buffer (2 x 2 ml) and then transferred to a 20 ml scintillation vial followed by scintillation fluid (10 ml). Liquid scintillation counting was used to measure the retained radioactivity.

### 2.6.3 Calculation of transport activity

Standard counts were determined using 8µmoles of the radiolabelled substrate and filter followed by 10ml scintillation fluid. Background counts were determined using filters that had been washed on the vacuum manifold and then transferred into vials with 10 ml scintillation fluid. The uptakes were calculated in nmol/mg cells (1 ml of cells at $A_{680}$ = 1.0 contains 0.68 mg dry mass, Ashworth and Kornberg, 1966). Rates of radioisotope-labelled substrate uptake were determined using the following equation:

$$\textbf{Uptake (nmol/mg)} = \textbf{(dpm-blank)} \times \frac{\textbf{[total volume]}}{\textbf{[sample volume]}} \times \frac{\textbf{1}}{\textbf{[No. of. mg]}} \times \frac{\textbf{1}}{\textbf{[dpm std/nmol std]}}$$

No. of mg: cell volume x $A_{680}$ x 0.68

dpm= disintegration per minute

**Table 2.23 Buffer for radiolabelled transport assay**

| Stock solution | Final concentration | Volume of stock solution |
|---|---|---|
| 1 M MOPS (3-[N-Morpholino] propane sulphonic acid) | 20 mM | 20 ml |
| 4 M NaCl | 140 mM | 35 ml |
| 0.1 M KCl | 10 mM | 1 ml |
| Adjust pH to 6.6 and make to 1 L with MilliQ water | | |

## 2.7 Computational approaches to membrane proteins

### 2.7.1 BLAST search

Basic Local Alignment Search Tool (BLAST) searches were used to compare nucleotide or protein sequences to sequence databases. This is one of the most widely used computational tools for sequence searching (Casey, 2005). Known DNA or protein sequences were submitted to the BLAST search tool (https://blast.ncbi.nlm.nih.gov/) to identify homologous sequences.

### 2.7.2 Sequence alignment

Alignment is an approach for arranging the sequences of DNA, RNA or protein to identify regions of similarity that may be a consequence of functional, structural or evolutionary relationships between the sequences. DNA or protein sequences were aligned using the multiple alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/).

### 2.7.3 Phylogenetic analysis

Amino acid sequences of proteins were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and/or from NCBI (http://www.ncbi.nlm.nih.gov/) and aligned using the multiple alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The resultant nearest-neighbour phylogenetic tree was exported in Newick format and re-drawn using iTol: Interactive Tree Of Life (http://itol.embl.de/) (Letunic and Bork, 2016).

### 2.7.4 Membrane topology predictions

Amino acid sequences of proteins were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and/or from NCBI (http://www.ncbi.nlm.nih.gov/) and analysed by the membrane topology prediction tools TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001) and TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009).

# Chapter-3

# Characterisation of the $Na^+$-hydantoin transport protein, Mhp1, and its mutants D229E, D229N and D229A

## 3.1 Introduction

Mhp1 is a hydantoin transport protein from the Gram-positive bacterium *Microbacterium liquefaciens* (Suzuki and Henderson, 2006; Jackson et al., 2013) and belongs to the family of nucleobase-cation-symport-1 (NCS1) transporters (Saier et al., 2006). Based on 3D structural similarity the NCS-1, Neurotransmitter-Sodium-Symporter (NSS), Sodium-Solute-Symporter (SSS), Betaine / Carnitine / Choline Transporters (BCCT) and Amino Acid-Polyamine-Organocation (APC) families have been grouped into the LeuT superfamily despite the low levels of protein sequence similarity (Wong et al., 2012). The LeuT superfamily also known as the APC or 5-helix inverted repeat superfamily (5HIRT) (Shi, 2013; Diallinas, 2014; Kazmier et al., 2014). According to the Transporter Classification Database (TCDB), the NCS-1 family are designated as secondary active transporters 2.A.39.5 (Busch & Saier, 2002; Saier et al., 2006; Ren and Paulsen, 2006). Proteins of the NCS1 family function as transporters for nucleobases or vitamins including nucleosides, allantoin, 5-substituted hydantoins, thiamine, uric acid and other related compounds (Ren et al., 2007; Pantazopoulou and Diallinas, 2007; Ma et al., 2016). The common transport mechanism catalysed by NCS1 family proteins is simplified as:

Nucleobase or vitamin (out) + cation (out) → Nucleobase or vitamin (in) + cation(in)

Mhp1 is a sodium dependent symporter (Weyand et al., 2008) and transports hydantoins substituted at the 5-position with aromatic groups across membranes (Suzuki and Henderson, 2006). Hydantoins are heterocyclic organic compounds interesting commercially for the production of amino acids. They are important compounds in salvage pathways for nitrogen balance in yeasts and plants (Bommarius et al., 1998; Altenbuchner et al., 2001; Suzuki et al., 2005). From a biotechnology point of view the Mhp1 gene was associated with an important cluster of genes that produced optically pure amino acids e.g. L-phenylalanine and L-tryptophan (Suzuki et al., 2005; Suzuki and Henderson, 2006). Mhp1 consists of 489 amino acids with a 54.6 kDa molecular weight (Suzuki and Henderson, 2006; Shimamura et al., 2010). The N- and C-terminus are moderately modified in a genetic construct, where a $His_6$-tag is attached to the C-terminus to facilitate over-expression and purification (Suzuki and Henderson, 2006; Shimamura et al., 2008).

Different conformational states of the 3D structure of Mhp1 are established in outward-facing open (PDB 2JLN), occluded with substrate (PDB 2JLO) and inward-facing open conformations (PDB 2X79) (Weyand et al., 2008; Shimamura et al., 2010; Simmons et al., 2012). The Mhp1 protein structure is comprised of twelve transmembrane helices (TMHs), ten of which are organized in two inverted repeats of five helices (Figure 3.1). The remaining two helices, 11 and 12, have no defined function in the protein and are poorly conserved among the transporters of NCS1 family.

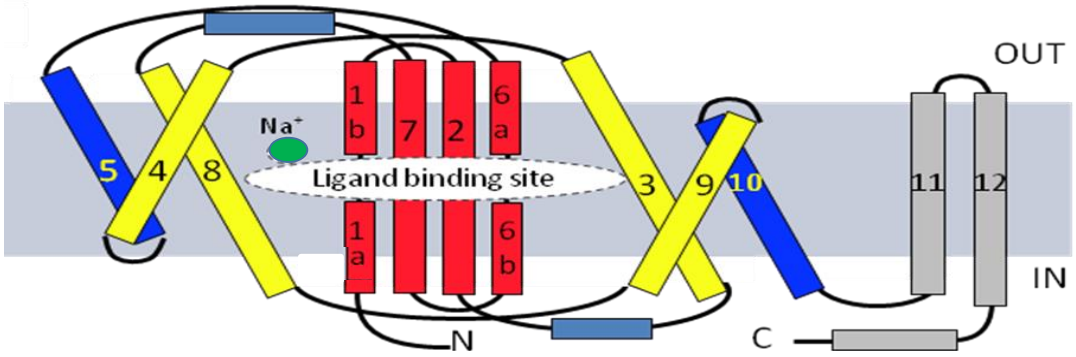



**Figure 3.1 Mhp1 topology diagram viewed in the plane of membrane.** The positions of the cation and substrate binding sites are shown as a green circle and white ellipsoid, respectively. The 'bundle' helices (TMHs 1, 2, 6 and 7) are in red, the 'hash motif' helices (TMHs 3, 4, 8 and 9) are in yellow, the 'thin gate' helices (TMHs 5 and 10) are in blue, two additional helices (TMHs 11, 12) are in grey and the membrane is in light blue. This diagram was adapted from Shimamura et al., (2010).

The Mhp1 protein has a single sodium and substrate binding site. Like other secondary transporters, Mhp1 utilises a mechanism regarded as "alternating access" membrane transport (Jardetzky, 1966) by switching from the outside to the inside facing state, releasing sodium and 5-substituted hydantoin. As stated before, three different conformational states of the Mhp1structure are already established whereas during ligand transport many other conformational states are likely to occur (Weyand et al., 2011; Jackson et al., 2013). Combining crystal structures of Mhp1 with molecular dynamics simulations and stopped flow fluorimetry, ten-steps were proposed to be involved during the transport process (Figure 3.2) (Jackson, 2012). In the absence of sodium, Mhp1 exists in an inward-facing state that converts to the outward-facing open and binds to the sodium ion first, followed by a ligand binding

(step 1-4); outward-facing occluded conformation formed by closing the external 'thin' gate (step 5); in relation to the 'bundle' movement of the 'hash motif' converts Mhp1 into an inward-facing occluded conformation from an outward-facing occluded conformation (step 6); intracellular thin gate opening release the sodium ion first followed by the ligand (steps 7-8). The protein facing inward and the exterior gate stays shut (steps 9-10).

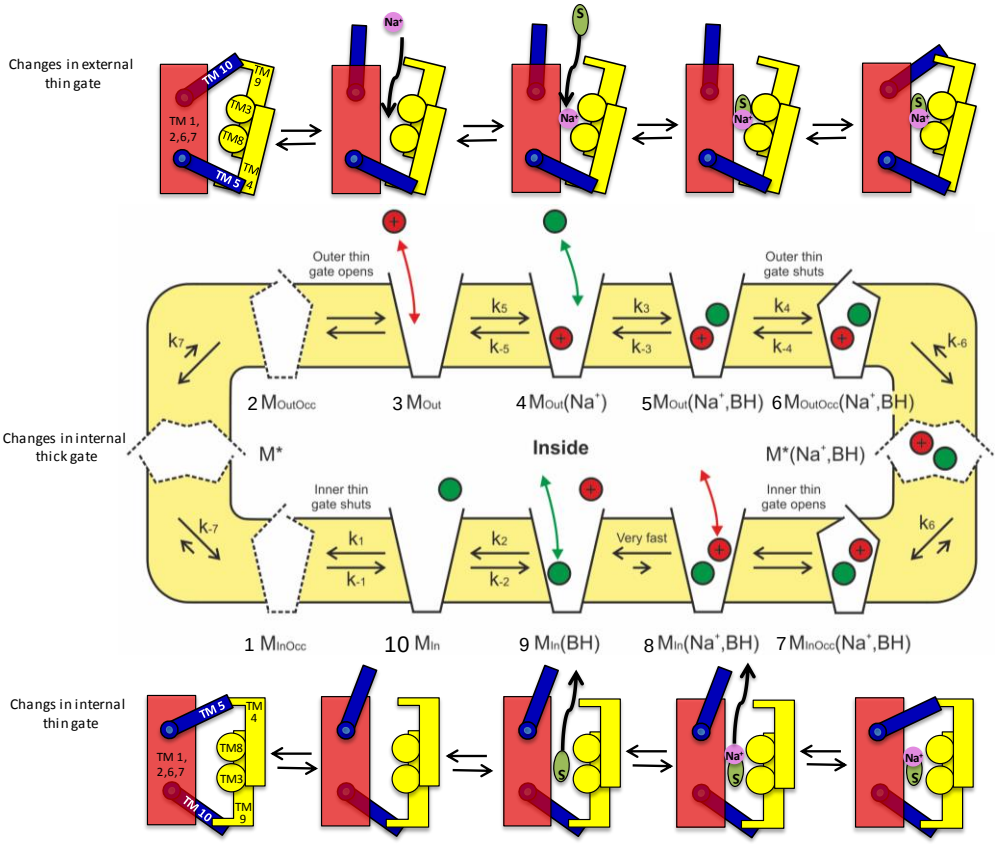



**Figure 3.2 Schematic diagrams of Mhp1 possible conformations during transport mechanism.** Helices 1, 2, 6 and 7 are the 'bundle' (red), helices 3, 4, 8 and 9 are the 'hash' motif (yellow), helices 5 and 10 are the 'thin' gates (blue), the ligand (green), the sodium ion (pink). The transport process follows 1-10 steps as described above. This diagram was provided by Professor Henderson, University of Leeds, UK.

To investigate the structure and biological activity of protein molecules, a single residue point mutation is an efficient method to make specific and intentional changes to the amino acid sequence of a gene. Previously a number of important residues in the Mhp1 protein have been identified with ligand and/or sodium binding

and transport (Jackson, 2012). Generally it is believed that conserved charged residues in many transport proteins play critical structural or functional roles in transmembrane domains (Guan and Nakae, 2001). The aspartate 229 is conserved among the bacterial NCS1 family proteins (Appendix 2). Based on the resolved crystal structures of Mhp1 it is located on the unwound segment of helix 6 in Mhp1 and might therefore have an important function. Therefore, this residue was mutated to a variety of amino acids (glutamate, asparagine and alanine) with different side chains to explore the possible effects on ligand binding.

## 3.2 Topology of Mhp1

As described earlier Mhp1 has 12 transmembrane spanning alpha-helices with both its N- and C-terminal ends at the cytoplasmic side of the membrane. This topology diagram was made based on the crystal structure of Mhp1 to show conserved residues among transporters of the NCS1 family and residues involved in cation or substrate binding (Figure 3.3).

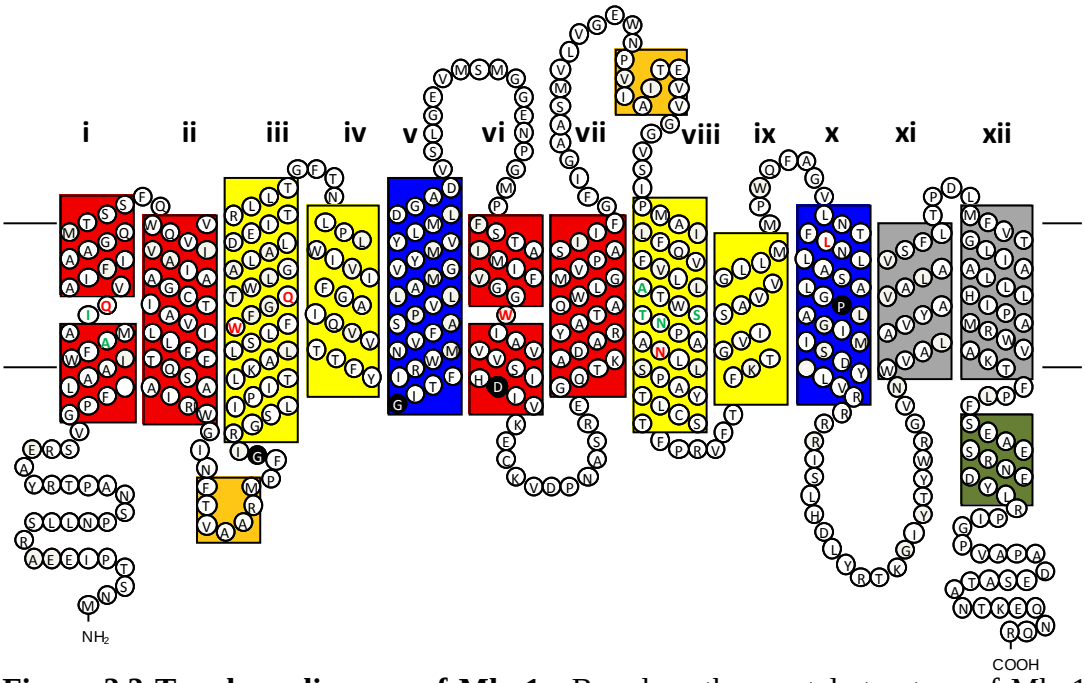



**Figure 3.3 Topology diagram of Mhp1.** Based on the crystal structure of Mhp1, the 12 transmembrane helices are indicated by roman numerals from the N- to C-terminus (Weyand et al., 2008). NCS1 family conserved residues are highlighted in black according to the alignment in Appendix 2. Residues involved in cation or substrate binding sites are highlighted in green and red respectively. This diagram was modified from Ma, (2010).

### 3.3 Purification of wild-type Mhp1

Sufficient amounts of purified wild-type Mhp1 protein were essential for biochemical and biophysical analysis. Inner membranes of *E. coli* BL21(DE3) expressing the Mhp1 protein were solubilised using 1% detergent DDM. Mhp1 was purified by immobilised metal affinity chromatography (IMAC) by exploiting the engineered C-terminal $His_6$-tag (Section 2.4.3). Various fractions from the purification were analysed by SDS-PAGE and Western blotting (Figure 3.4).

Under conditions of the SDS-PAGE separation the purified protein in lane-6 is migrating at a position of around 37 kDa. Comparing the lane-6 with lanes 2, 3, 4 and 5 reveals a number of contaminating bands as compared to lane-6, indicating success of the IMAC protein purification procedure. The migrated size of Mhp1 is less than the predicted molecular weight of 54.6 kDa; it has been reported previously that membrane proteins migrate anomalously on SDS-PAGE gel, which could be due to more SDS binding or only partial unfolding (Robinson, 2011; Ratha et al., 2009; Ma et al., 2016).

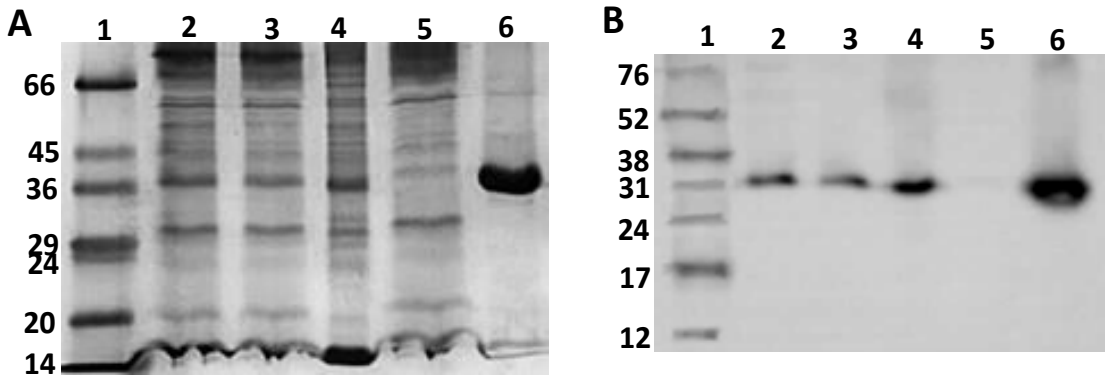



**Figure 3.4 SDS-PAGE and Western blot analysis of purification of the Mhp1 protein.** Mhp1 protein was purified from the inner membrane of *E. coli* BL21(DE3) pTTQ18/Mhp1-$His_6$. The inner membranes were solubilised in 1% DDM. Coomassie blue stained 15% SDS-PAGE (A) and Western blot analysis (B). Samples were loaded as follows: (1) molecular weight markers (kDa); (2) inner membranes; (3) detergent extract; (4) insoluble material pellet; (5) unbound flow; (6) purified protein. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

## 3.4 Secondary structure integrity and thermal stability of Mhp1 protein

Far UV (180 – 260 nm) CD spectroscopy was used to examine the secondary structure content and thermal stability of the purified wild-type Mhp1 protein (section 2.5.1). It is well established that CD scans in the range of 180-260 nm produce a typical spectrum for predominantly alpha helical proteins with a peak at ~192 nm and troughs at 209 nm and 222 nm (Wallace et al., 2003; Kelly et al., 2005; Bulheller et al., 2007). The results of CD experiments demonstrate that the wild-type Mhp1 protein was largely alpha-helical (Figure 3.5A) and confirm it has retained its alpha helical secondary structure after passing through the various steps of purification. Also the results in Figure 3.5A suggest that wild-type protein was correctly folded. This was an important control for further biophysical analyses which also allowed the structural integrity of each of the mutated proteins to be compared to wild-type protein.

To determine the thermal stability of wild-type Mhp1 protein, the loss of alpha-helical structure at increasing temperature was measured (Figure 3.5B) by ramping the temperature from 5-90 ºC and finally back to 5 ºC. With increasing temperature the signal at 209 nm began to change significantly indicating the instability of the protein and a melting temperature of 48.9 ºC was estimated using Global Analysis CD software 3. The wild-type Mhp1 protein is therefore reasonably stable for performing biophysical assays up to a temperature of 25 ºC. On returning the temperature to 5 ºC, there was no evidence that suggests refolding of the protein, consistent with reported results for other transport proteins (Bettaney, 2008; Ma, 2010; Sukumar, 2012; Jackson, 2012).

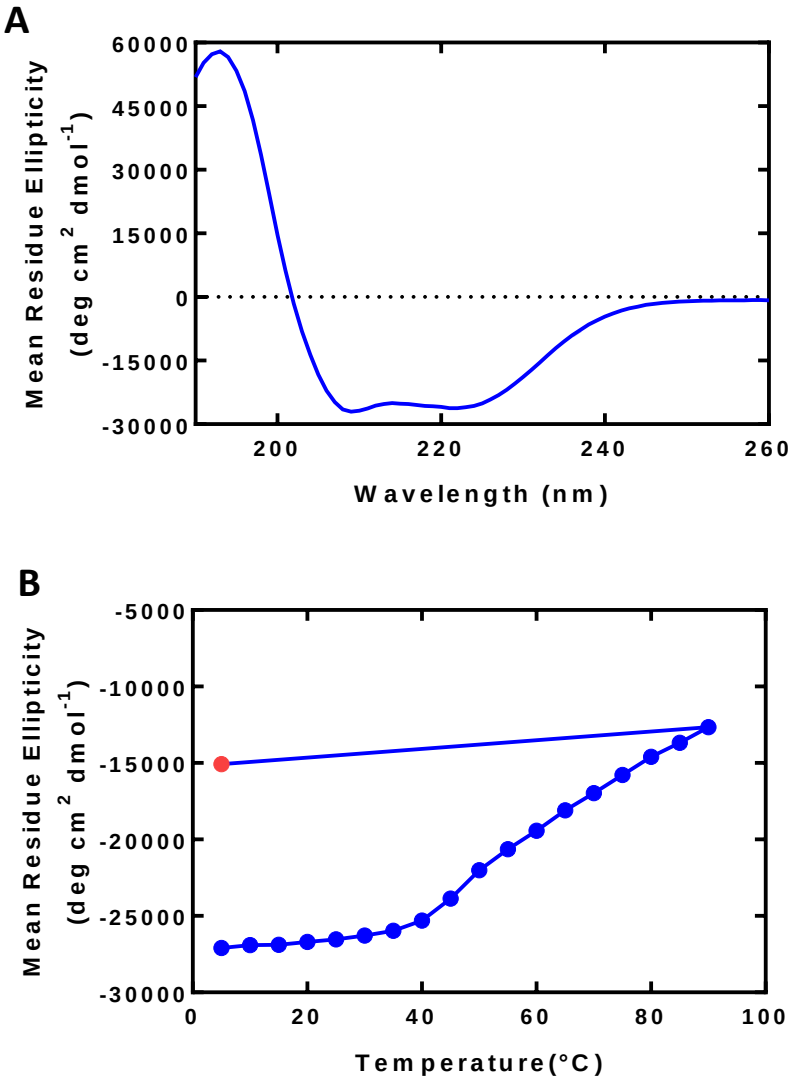



**Figure 3.5 Far-UV CD spectra of the secondary structure and thermal stability of purified Mhp1 protein.** (A) Far-UV (180-260 nm) CD spectrum for purified Mhp1 protein. Measurements were performed using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant liquid nitrogen flushing. Samples were prepared in a Hellma quartz cuvette of 1.0 mm path length in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM). The changes in secondary structure were monitored at 209 nm or 222 nm, representative of alpha helices. (B) The thermal stability of proteins was analysed by ramping the temperature from 5-90 ºC and finally back to 5 ºC (red). Spectra were recorded at 5 ºC intervals in the range of 180 to 260 nm.

## 3.5 Fluorimetric studies of Mhp1

Spectrophotofluorimetry is a fast and effective technique of measuring the fluorescence change in a protein upon the addition of ligand (Walmsley, 2000; Weyand et al., 2008; Simmons et al., 2012). Such fluorescence principally originates from tryptophan residues and is especially useful when tryptophan residues are proposed to be in, or close to, the ligand binding site. Cloned Mhp1 protein has 15 tryptophan residues (Figure 3.3), two (Trp117 and Trp220) of which are located in the ligand binding site, making spectrophotofluorimetry an interesting possibility (Weyand et al., 2008). This method was used as described in section (2.5.2) to measure the binding affinity of various ligands to wild-type Mhp1 and its mutants D229E, D229N and D229A.

### 3.5.1 Cation specificity of the Mhp1 protein

In secondary active transporters, sodium ions or protons often drive substrate translocation. Mhp1 is a sodium-dependent secondary active symporter of hydantoins (Weyand et al., 2008). To know the effect of other cations on the binding of L-benzylhydantoin (L-BH) to Mhp1, a range of different cations (ChCl, LiCl, KCl, RbCl and CsCl) were tested at a concentration of 140 mM in the binding assay with 2mM L-BH (Figure 3.6). With sodium chloride a significantly higher quench in fluorescence intensity was measured than with all of the other cations (ChCl, KCl, LiCl, CsCl and RbCl) used. So these results recognise that Mhp1 is specific for sodium ions whereas other cations did not stimulate the L-BH binding to purified Mhp1 protein, except perhaps a minor effect with $Li^+$.

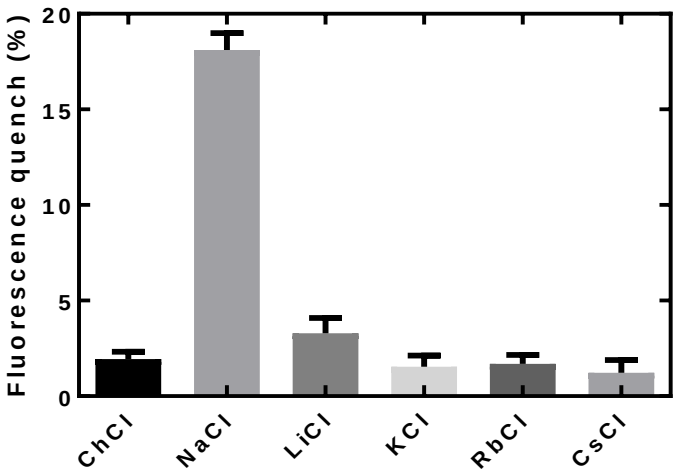



**Figure 3.6 Cation specificity of L-BH binding to Mhp1.** Fluorescence quench percentage was measured by using 140 µg/ml Mhp1 protein in fluorescence buffer (Table 2.21) and 2 mM L-BH at a temperature of 18 $^{o}$C. Fluorescence was excited at a wavelength of 295 nm and quenching was measured at a wavelength of 332 nm. To make the mixing more efficient a magnetic stirring bar was used in a quartz cuvette with a path-length of 1 cm. The fluorescence quench values are the mean of four repeats and the error bars represent the standard error of the means on two batches of protein.

### 3.5.2 Binding affinity of D and L-benzylhydantoin to purified Mhp1

Previously it has been shown that uptake of L-enantiomers of BH and IMH by whole cells expressing Mhp1 were higher than for the D-enantiomers (Suzuki and Henderson, 2006). This implies that purified Mhp1 will bind to L-enantiomers with a higher affinity than the analogous D-enantiomers. Therefore, spectrophotofluorimetry was used to measure the concentration dependent binding of L-BH or D-BH to the Mhp1 protein in the absence and presence of 15 mM NaCl over the concentration range 0-2 mM. Figure 3.7 indicates how L-BH bound with a higher affinity to purified Mhp1 than D-BH in both the presence and absence of sodium. With the addition of ligand the percentage of fluorescence quench was increased. For L-BH it was observed that in the presence of 15 mM NaCl the value of an apparent $K_d$ was 0.043±0.003mM compared to 0.109±0.006mM for D-BH whereas in the absence of sodium the $K_d$ for L-BH binding was 0.67±0.04mM compared to 4.30±1.35mM for D-BH. The maximal fluorescence quench for D-BH

and L-BH in the presence and absence of sodium were comparable (Table 3.1). These results demonstrate that there was stereo selectivity between the L and D-enantiomers and also the sodium dependence of binding of either enantiomer of BH. Such stereo selectivity has also been reported for selectivity of substrate in the LeuT superfamily members ApcT and CaiT (Shaffer et al., 2009; Jung et al., 2002).

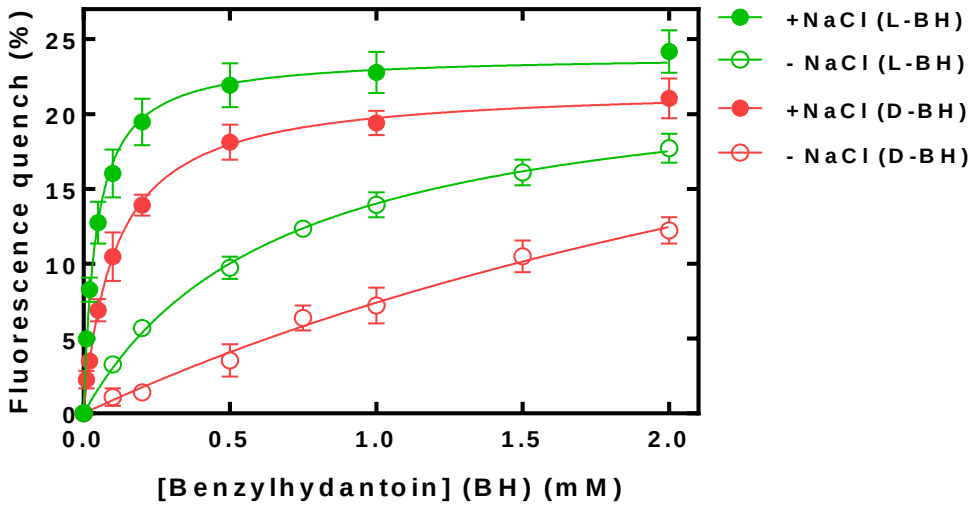



**Figure 3.7 Stimulation by sodium of L-BH or D-BH binding to the wild-type Mhp1 protein.** Using steady-state fluorimetry measurements on purified proteins were performed using a Photon Technology International spectrophotofluorimeter (section 2.5.2). Samples containing purified Mhp1 (140 µg/ml) in fluorescence buffer (Table 2.21) were titrated with benzylhydantoin (0-2 mM) for L-BH (green) and D-BH (red) in the presence of 15 mM NaCl (closed circles) and absence (open circles) at 18 ºC. Following ligand additions, samples were stirred for 1.5 minutes to equilibrate before making the measurements. Data were analysed using the Michaelis-Menten analysis tool on Graph Pad Prism 7. The data represents a minimum of four repeats with error bars showing the standard error of the mean.

**Table 3.1 Comparison of $K_d$ and $\Delta F_{max}$ values for L-BH or D-BH interacting with wild-type Mhp1 using spectrophotofluorimetry.**

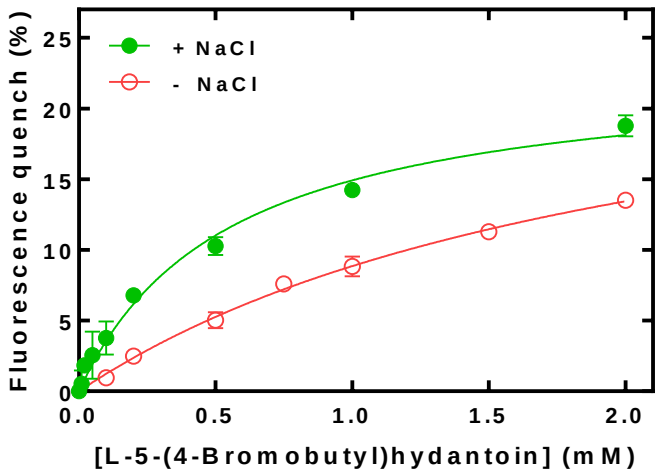
| Ligand | Apparent $K_d$ (mM) | | $F_{max}$ (%) | |
|---|---|---|---|---|
| | +NaCl | -NaCl | +NaCl | -NaCl |
| **L-BH** | 0.043±0.003 | 0.67±0.04 | 23.9±0.3 | 23.3±0.6 |
| **D-BH** | 0.109±0.006 | 4.30±1.35 | 21.8±0.3 | 39.1±9.2 |

### 3.5.3 Aliphatic group substitution at the 5-position of the hydantoin moiety reduces affinity

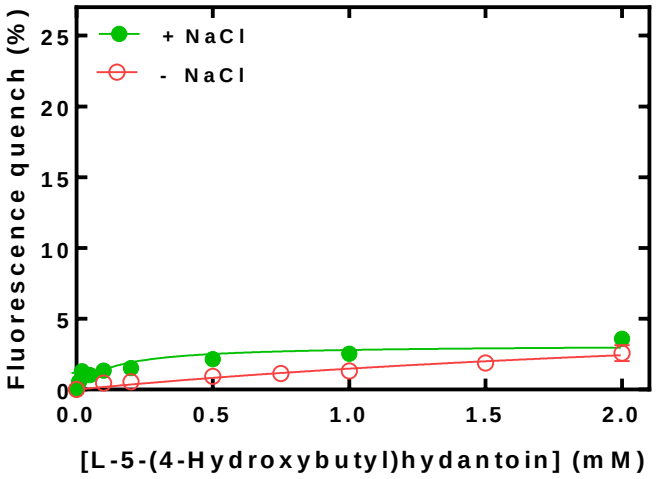
It has been proposed that both the hydantoin head group and an aromatic hydrophobic R group attached at the 5-position are required to generate a ligand that binds to Mhp1 protein with a reasonable affinity. If either substituent is removed then ligand binding affinity is reduced (Weyand et al., 2008; Simmons et al., 2012). It has also been recognized that the ligand binding pocket is hydrophobic in nature and that the benzene ring on the ligand can form a face-to-face and edge-to-face π stacking interaction with Mhp1 protein residues Trp117 and Trp220, respectively. In the light of above statement it has been proposed that compounds with aliphatic groups attached to the hydantoin are likely to have a reduced potency (Weyand et al., 2008; Shimamura et al., 2010; Weyand et al., 2011; Simmons et al., 2012). Therefore three ligands substituted with an aliphatic group at position 5 of hydantoin i.e. L-5-(4-bromobutyl) hydantoin, L-5-(4-hydroxybutyl) hydantoin and L-5-(neo-pentyl) hydantoin (Figure 3.9) were titrated to the wild type Mhp1 protein from 0-2mM in the absence and presence of 15 mM NaCl at 18ºC. This will investigate how modification of aromatic R groups attached at the 5-position of the hydantoin moiety affected ligand potency. Spectrophotofluorometery was used to measure the fluorescence change upon the addition of aliphatic group hydantoins substituted at 5-position (Figure 3.8).

It was observed that the replacement of the benzyl group of hydantoin with a bromobutyl resulted in a weak affinity in the presence of NaCl with a $K_d$ value of 0.54±0.07 mM compared to the $K_d$ value of 0.043±0.003 mM for L-BH (Figure 3.8A, Table 3.2). Similarly replacement of the benzyl group with a hydroxybutyl hydantoin and pentyl hydantoin resulted in a drastic loss of potency (Figure 3.8B, Figure 3.8C, Table 3.2). The ineffectiveness could be due to the unfavourable stacking interactions in the ligand binding site. These results recognise that the hydrophobic ligand binding site of Mhp1 prefers the aromatic ring at the 5-position of hydantoin to be bound at the ligand binding site. Removal of aromatic interactions impairs ligand binding affinity.

A


+ NaCl
- NaCl
Fluorescence quench (%)
0
5
10
15
20
25
0.0
0.5
1.0
1.5
2.0
[L-5-(4-Bromobutyl)hydantoin] (mM)

B


+ NaCl
- NaCl
Fluorescence quench (%)
0
5
10
15
20
25
0.0
0.5
1.0
1.5
2.0
[L-5-(4-Hydroxybutyl)hydantoin] (mM)

C

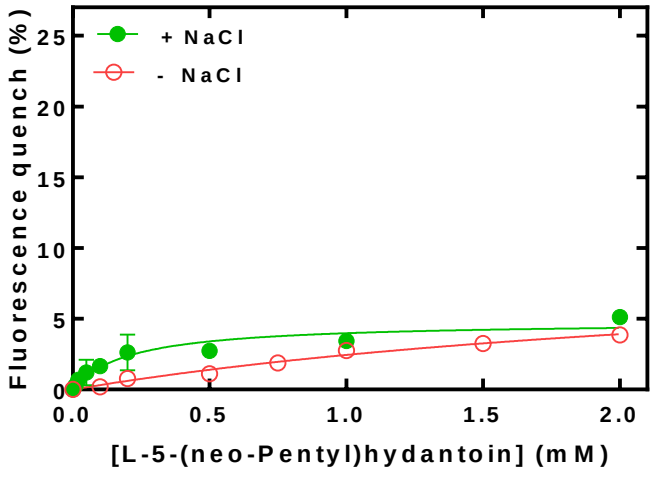

+ NaCl
- NaCl
Fluorescence quench (%)
0
5
10
15
20
25
0.0
0.5
1.0
1.5
2.0
[L-5-(neo-Pentyl)hydantoin] (mM)

**Figure 3.8 Fluorescence change upon the addition of different hydantoins compounds to the wild-type Mhp1 protein.** L-5-(4-brombutyl) hydantoins (A), L-5-(4-hydroxybutyl) hydantoins (B) and L-5-(neo-pentyl) hydantoin (C) were titrated from 0-2mM in the presence of 15 mM NaCl (closed circles, green) and absence (open circles, red). The ligand was added to the protein sample and mixed for 1.5 minutes to equilibrate before the fluorescence spectra were obtained. To make the mixing more efficient a magnetic stirring bar was used in a quartz cuvette with a path-length of 1 cm. Binding curves were generated using the Michaelis-Menten analysis tool on Graph Pad Prism 7.

**Table 3.2 Comparison of $K_d$ and $\Delta F_{max}$ values for different hydantoins compounds interacting with the wild-type Mhp1 using spectrophotofluorimetry.**

| Ligand | Apparent $K_d$ (mM) | | $\Delta F_{max}$ (%) | |
|---|---|---|---|---|
| | +NaCl | -NaCl | +NaCl | -NaCl |
| **L-benzylhydantoin** | 0.043±0.003 | 0.67±0.04 | 23.9±0.3 | 23.3±0.6 |
| **L-5-(4-bromobutyl) hydantoins** | 0.54±0.07 | 2.15±0.29 | 23.08±1.28 | 27.88±2.37 |
| **L-5-(4-hydroxybutyl) hydantoins** | 0.12±0.05 | 3.62±2.51 | 3.15±0.33 | 6.84±3.41 |
| **L-5-(neo-pentyl) hydantoin** | 0.20±0.06 | 2.99±1.12 | 4.80±0.45 | 9.77±2.49 |

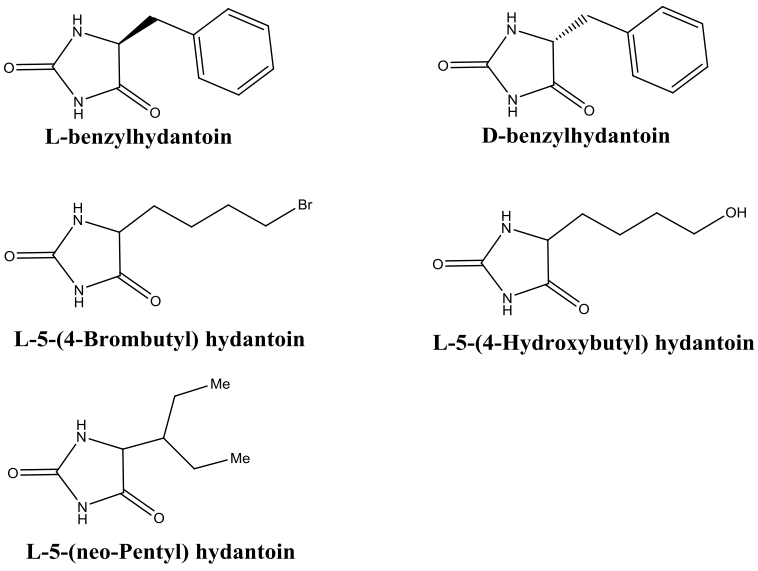



**Figure 3.9 Chemical structures of hydantoins compounds**

### 3.5.4 Kinetic characterisation of Mhp1 binding to L-benzylhydantoin by stopped-flow fluorimetry

Experiments were performed to investigate the binding of L-BH to Mhp1 using stopped-flow fluorimetry, which measures the initial rate of the fluorescence response following rapid mixing of protein and ligand (section 2.5.3). The rapid mixing of the ligand with protein allows the fluorescence changes in the protein to be measured on the millisecond time scale. Deconvolution of exponential contributions to the observed rate changes allows direct ligand interaction with protein to be differentiated from conformational fluorescence changes (Walmsley, 2000). The fluorescence response over 10 seconds was monitored in the absence and presence of added NaCl (15 mM) following additions of L-BH up to a concentration of 2 mM (Figure 3.10). The observed rate constant and amplitude of fluorescence change were taken from these data.

There was a high rate of fluorescence change and amplitude in the presence of sodium chloride whereas a smaller rate of fluorescence change and amplitude was observed in the absence of sodium chloride (Figure 3.11). In the presence of NaCl (15 mM), an apparent $K_d$ value of 0.051±0.006 mM for L-BH binding was generated, whereas an apparent $K_d$ value of 0.152±0.135 mM was produced in the absence of sodium. These results showed a significant effect of sodium on the rate constant and amplitude, indicating that L-BH binds to Mhp1 in a sodium dependent manner.

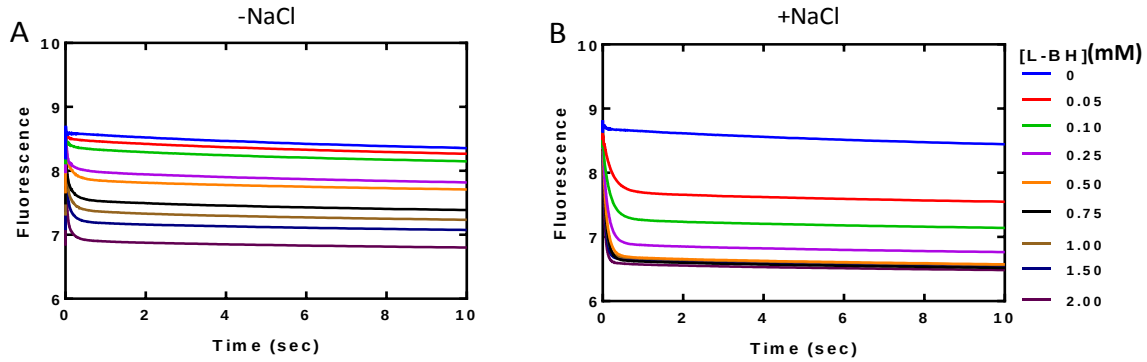



**Figure 3.10 Fluorescence change for binding of L-benzyl hydantoin to Mhp1.** Using stopped-flow fluorimetry at 18 ºC purified Mhp1 (140 µg/ml) in fluorescence buffer (Table 2.22) in absence (A) and presence (B) of added NaCl (15 mM) was mixed with increasing concentrations of L-BH (0-2 mM) and the fluorescence response was monitored over 10 seconds following each addition.

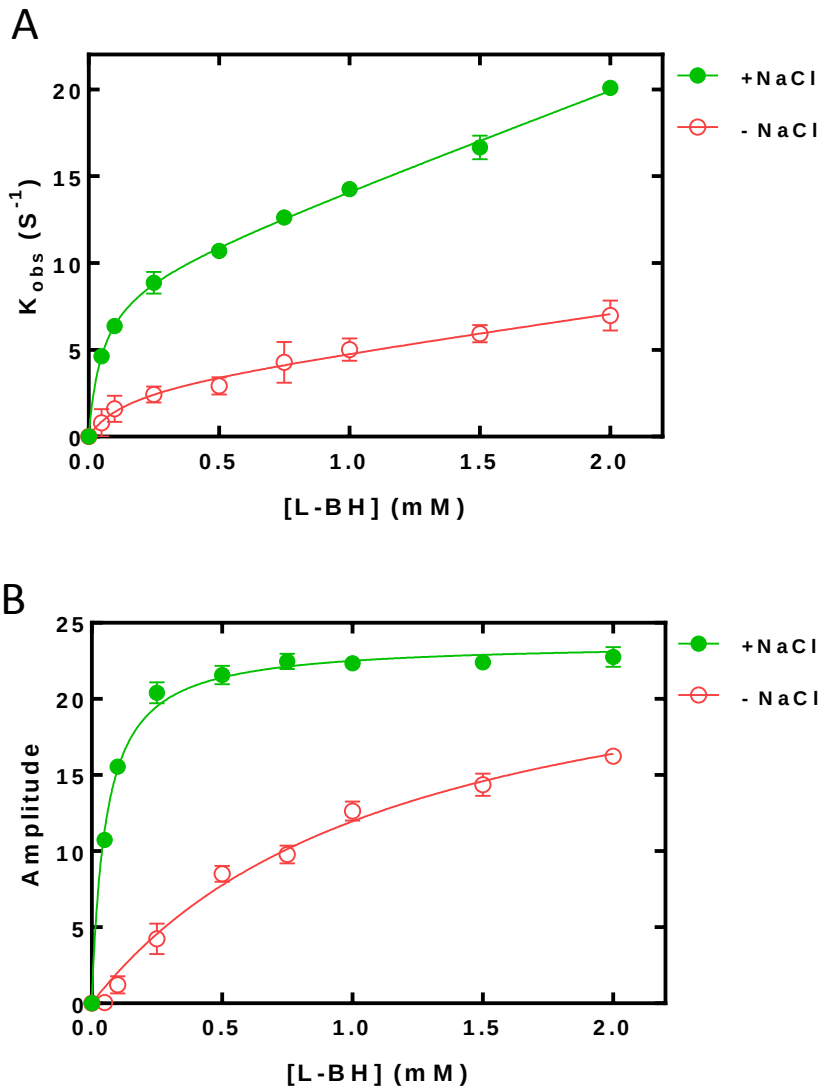



**Figure 3.11 Kinetic parameters for sodium-dependent binding of L-BH to Mhp1.** Observed rate constant (A) and amplitude (B) were determined using stopped-flow fluorimetry (SX-20 of Applied Photophysics, UK) at 18 Cº. Purified Mhp1 protein (140 µg/ml) was mixed with increasing concentrations of L-BH (0-2 mM) in fluorescence buffer (Table 2.22) in the presence of 15 mM NaCl (closed circles, green) or its absence (open circles, red). The fluorescence response was monitored over 10 seconds following each addition (Figure3.10). Values of $k_{obs}$ were calculated using a double exponential equation ($Y= Y_0 + A_1e^{-x/t1} + A_2e^{-x/t2}$) of OriginPro. The $K_d$ value of $k_{obs}$ in the presence and absence of sodium were determined by non-linear regression equation (Y=ΔFmax*X/(Kd+X) + NS*X + Background) on GraphPad prism 7.

## 3.6 Site directed mutagenesis to probe the role of Mhp1 residue Asp229

### 3.6.1 Selecting residue Asp229 for mutation

Previously it has been reported that there are interacting sodium and ligand binding sites in Mhp1, and binding of sodium stabilizes the Mhp1 outward-facing open conformation and stimulates ligand binding (Weyand et al., 2011). Therefore, sodium binding may cause some conformational changes that create a higher affinity site for ligand binding. The aspartic acid residue at the position corresponding to Asp229 in Mhp1 is conserved amongst the NCS-1 family transporters (Appendix 2). This residue is not directly in the ligand or sodium binding site but might be important due to its location in the discontinuous helix 6 (Figure 3.3). Often, interactions between neighbouring residues are as important as the interactions with the ligand. The disruption of interactions between residues could cause conformational changes that will affect the affinity of ligand binding if this interaction is essential. Therefore, this residue was mutated to a variety of amino acids (glutamate, asparagine and alanine) with different side chains to investigate their possible effect on L-BH binding.

The residue was mutated from Asp to Glu, keeping the negative charge on the R-group with a modest increase in size of the side-chain. Another mutation was constructed from Asp to Asn to abolish the charge but keep a similar size of side chain. A third mutant of the same residue was created from Asp to Ala to diminish its acidic charge and reduce its size.

### 3.6.2 Predicting the impact of Asp229 mutations based on the resolved crystal structures of Mhp1

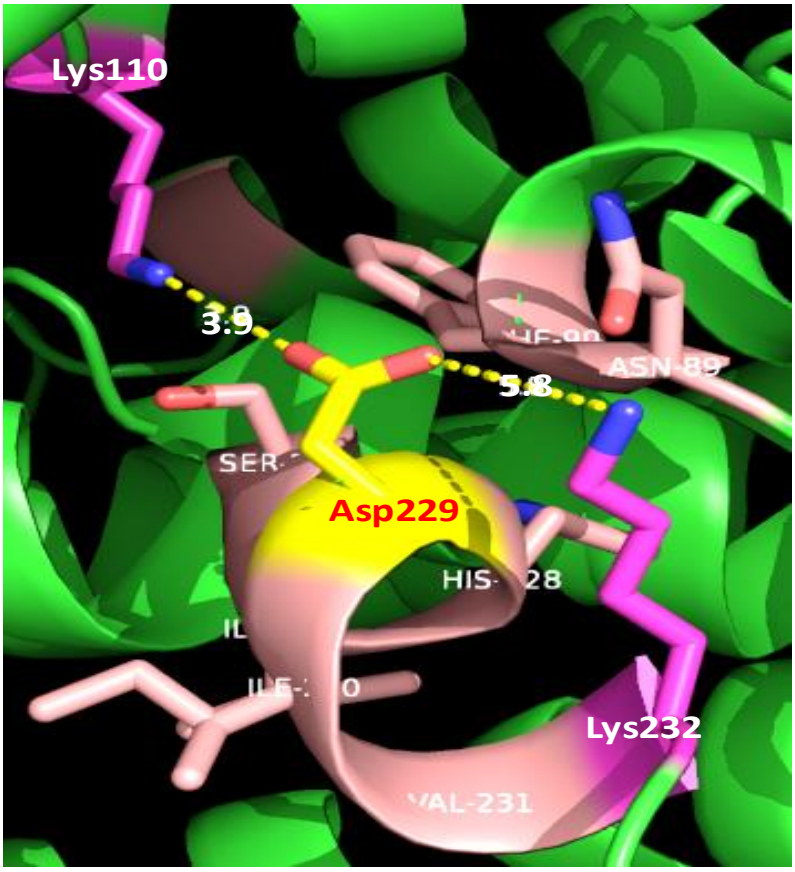
Based on the crystal structures of Mhp1 in outward-facing open, ligand bound occluded and inward-facing open conformations, residue Asp229 is located in Mhp1 on discontinuous helix 6 (Figure 3.3). Asp229 has a side chain of negative charge that may interact with the positive charge of Lys110 (helix 3) and Lys232 (loop 6-7) forming salt bridges in the inward-facing structure (PDB 2X79) (Figure 3.12). In membrane proteins, salt bridges are not only structurally important but also critical for functional and conformational changes (Kumar and Nussinov, 1999; Law et al.,

2008; Bosshard et al., 2004; Walther and Ulrich, 2014). Normally Mhp1 exists in the inward-facing state in the absence of sodium and benzylhydantoin. The sodium and benzylhydantoin together convert the protein from the inward-facing to the outward-facing state (Weyand et al., 2010). In the outward-facing open conformation, sodium is likely to bind first followed by hydantoin as shown in Figure 3.2.

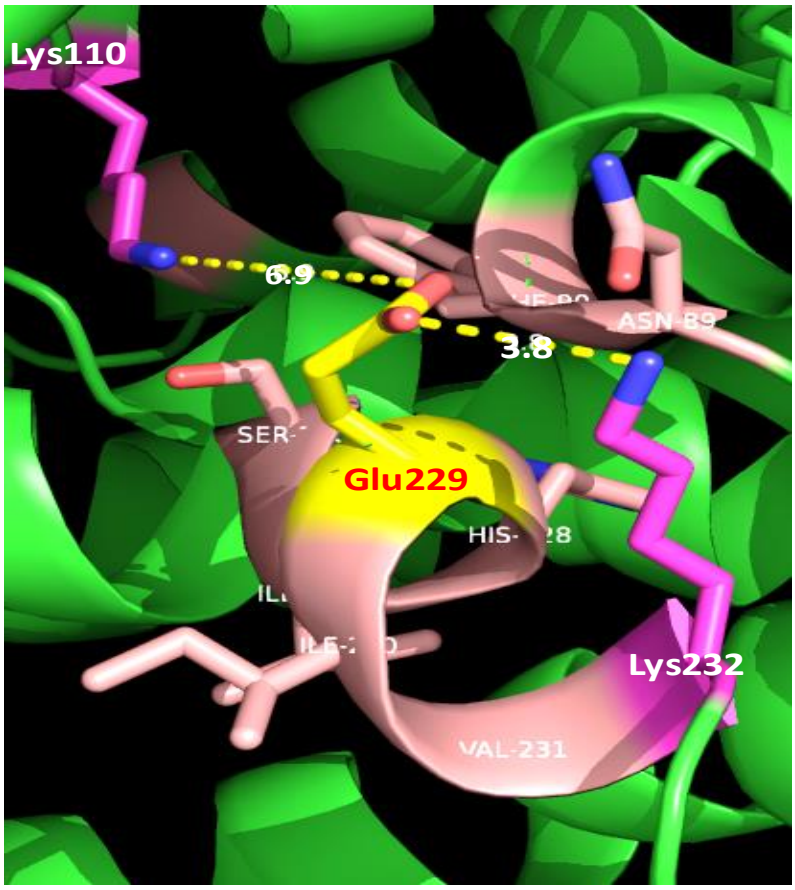
According to the pymol software the distance between the oxygen atoms of Asp229 and nitrogen atoms of Lys110 during salt bridge formation was typically 3.9 Å, whereas the distance between Asp229 and Lys232 was 5.8 Å (Figure 3.12A). Mutating residue from Asp229 to Glu was predicted to retain biological activity due to its negative charge that will retain a similar structure and functional stability. Figure 3.12B shows that mutating Asp229 to Glu has now formed salt bridges with Lys110 and Lys232 with a distance of 6.9 Å and 3.8 Å, respectively. Mutation of Asp229 to Asn and Ala prevents formation of the salt-bridges and would probably destabilise the structure and lose function for ligand binding.

**Figure 3.12 Salt bridge formation in the inward-facing open (PDB 2X79) structure of Mhp1 and mutant D229E.** Salt bridge formation between Asp229 or Glu229 in wild-type Mhp1 (A) and mutant D229E (B), respectively, with lysine residues 110 and 232. The yellow dashed lines represent hydrogen bonds. This diagram was generated in Pymol (http://www.pymol.org) using the inward-facing open crystal structure of Mhp1 (PDB 2X79).

A


Lys110
3.9
5.8
Asp229
HIS-
28
Lys232
VAL-231

B


Lys110
6.9
3.8
Glu229
HIS-
28
Lys232
VAL-231

## 3.7 Purification of Mhp1 mutants D229E, D229N and D229A

Inner membranes were prepared by sucrose density gradient ultracentrifugation (section 2.4.2) from the *E. coli* BL21(DE3) cells expressing Mhp1 mutant D229E, D229N and D229A and were solubilised using 1% DDM. The same methodology was applied for purification of all the proteins using immobilised metal affinity chromatography by exploiting the engineered C-terminal $His_6$-tag (Section 2.4.3). Various fractions from the purification were analysed by SDS-PAGE and Western blotting (Figure 3.13). For each protein lane 2 contains the inner membranes solubilised with 1% DDM and lane 3 contains the supernatant obtained by ultracentrifugation of the solubilised inner membranes. Lane 4 shows the pellet obtained following solubilisation and ultracentrifugation (insoluble fraction). Lane 5 shows protein that did not bind to the Ni-NTA column (unbound fraction) and the blots suggest that most of the required proteins have been bound to the resin. Lane 6 shows the purified protein, migrating at ~37 KDa in the eluted fraction is less than the predicted molecular weight of 54 kDa but it is widely recognised that membrane proteins migrate anomalously on SDS-PAGE gels at lower molecular weight positions than their actual molecular weights due to their hydrophobic nature, partial unfolding, high binding of SDS or the retention of secondary structure facilitating the migration through the gel (Ward et al., 2000; Robinson, 2011; Rath et al., 2009; Rath and Deber, 2013).

The comparison of yield and purity of wild-type Mhp1 and its mutant proteins are detailed in Table-3.3. Which show that the mutant D229E yield was similar to that of the wild-type Mhp1 whereas the mutant D229N and D229A purification yield was less than that for the wild-type Mhp1 as shown in the Table 3.3. Overall, the yield and purification quality of wild-type Mhp1 was better than all of the mutants (D229E, D229N and D229A).

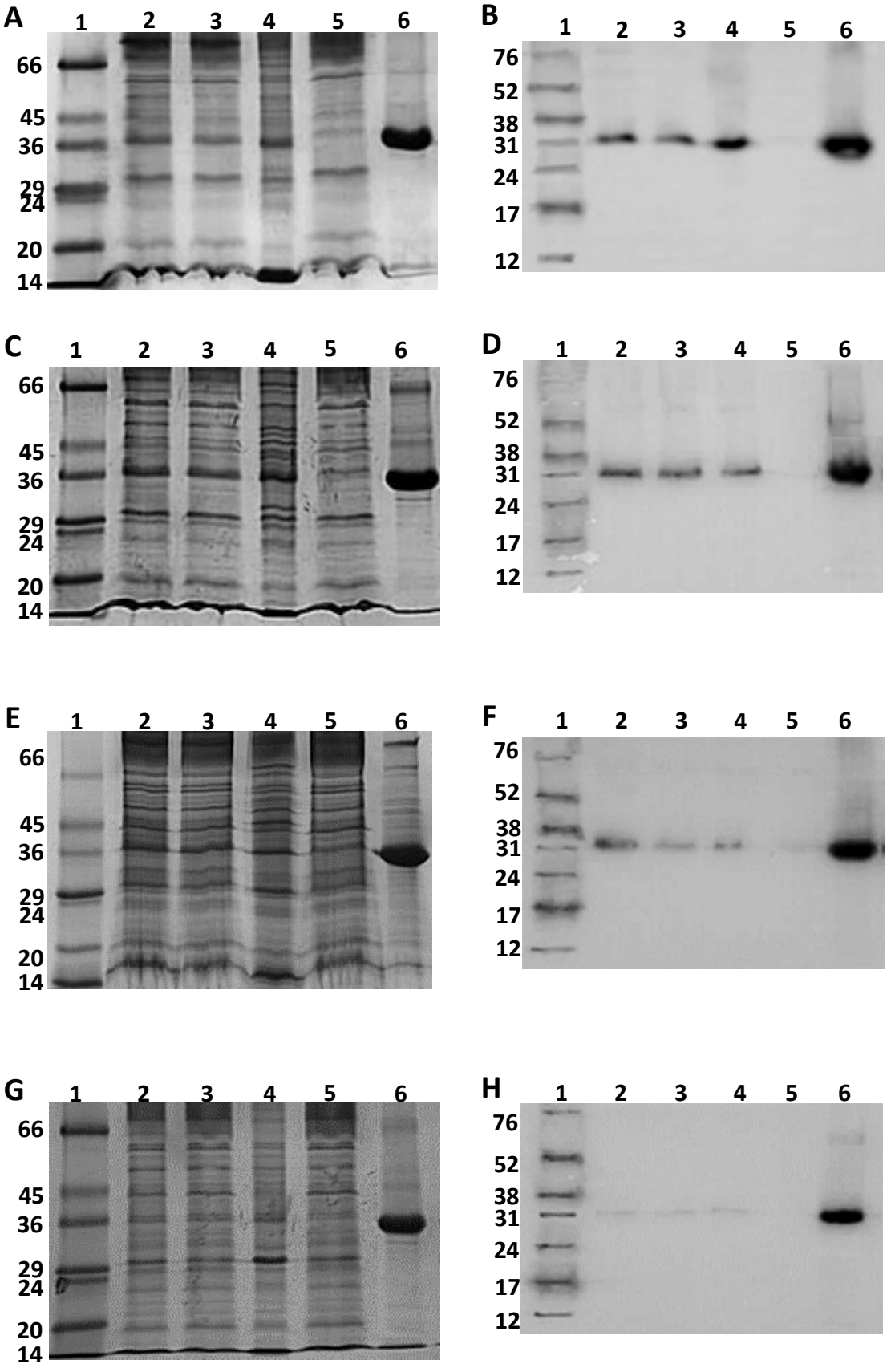

A
1 2 3 4 5 6
66
45
36
29
24
20
14
B
1 2 3 4 5 6
76
52
38
31
24
17
12
C
1 2 3 4 5 6
66
45
36
29
24
20
14
D
1 2 3 4 5 6
76
52
38
31
24
17
12
E
1 2 3 4 5 6
66
45
36
29
24
20
14
F
1 2 3 4 5 6
76
52
38
31
24
17
12
G
1 2 3 4 5 6
66
45
36
29
24
20
14
H
1 2 3 4 5 6
76
52
38
31
24
17
12

**Figure 3.13 Purification of proteins from inner membranes**. Inner membranes of *E. coli* BL21(DE3) expressing the Mhp1 (A, B), mutant D229E (C, D), D229N (E, F) and D229A (G, H) proteins were prepared using the sucrose density gradient procedure (section 2.4.2). Membrane preparations were soblubilised using solubilisation buffer containing 1% DDM (section 2.4.3). Different fractions from the purification were subjected to analysis on a 15% SDS-PAGE gel (left) and Western blot (right). The samples were loaded as follows: (1) molecular weight marker (kDa); (2) inner membranes; (3) detergent extract; (4) insoluble material pellet; (5) unbound flow through; and (6) purified protein. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

**Table 3.3 Purity and yield of proteins from 30 litre fermenter cultures.**

| Protein name | Wet cell pellet from 30L (g) | Inner membrane obtained (ml) | Purification yield (mg per litre) | Purification purity (%) |
|---|---|---|---|---|
| **Mhp1** | 102.1 | 24.9 | 1.6 | 89 |
| **D229E** | 106.5 | 24.7 | 1.5 | 81 |
| **D229N** | 100.2 | 21.4 | 1.3 | 77 |
| **D229A** | 93.9 | 19.6 | 1.1 | 83 |

## 3.8 Analysing the secondary structure and thermal stability of Mhp1 mutants D229E, D229N and D229A using circular dichroism

Wild-type Mhp1 and its mutants D229E, D229N and D229A proteins were purified (section 2.4.3). The secondary structure integrity was analysed by far UV (180-260 nm) CD spectroscopy (Section 2.5.1). It was important to confirm as far as possible the structural integrity of Mhp1 and its mutants prior to ligand binding analysis. CD spectra of mutants were compared to that of the wild-type Mhp1 protein (Figure 3.14). The shape of the resultant CD spectra of all the tested mutants confirm that they are predominantly alpha-helical and matched very closely with that of the wild-type Mhp1 protein with a positive peak at ~192 nm and negative peaks at 209 nm and 222 nm (Wallace et al., 2003; Kelly et al., 2005; Miles and Wallace, 2016). The CD spectra similarities of all the mutants with that of the wild-type Mhp1 suggest that the gross conformations of the mutant proteins have retained their alpha helical secondary structure. There was some small variability in the CD signal intensities probably due to differences in the protein concentration of the samples. These measurements also confirm that all the purified proteins were correctly folded and had retained their secondary structure architecture after passing through various steps of the purification.

Far UV CD spectroscopy was also used to determine the thermal stability of all mutants and wild-type protein (section 2.5.1). The purified proteins were heated at 5ºC intervals by ramping the temperature from 5-90ºC and finally back to 5ºC (Figure 3.15). With increasing temperature, the signal at 209 nm began to change significantly indicating thermal stability of the proteins. The thermal stabilities of all the mutants were comparable to the wild-type Mhp1 protein, using Global Analysis CD software-3 (Table 3.4). The thermal denaturation in most cases is irreversible aggregation (Miles and Wallace, 2016), and the final scans at 5 ºC show that the purified proteins had all lost their ability to refold into their native secondary structure. These results are similar to reports on a wider range of membrane transport proteins (Bettaney, 2008; Jackson, 2012).

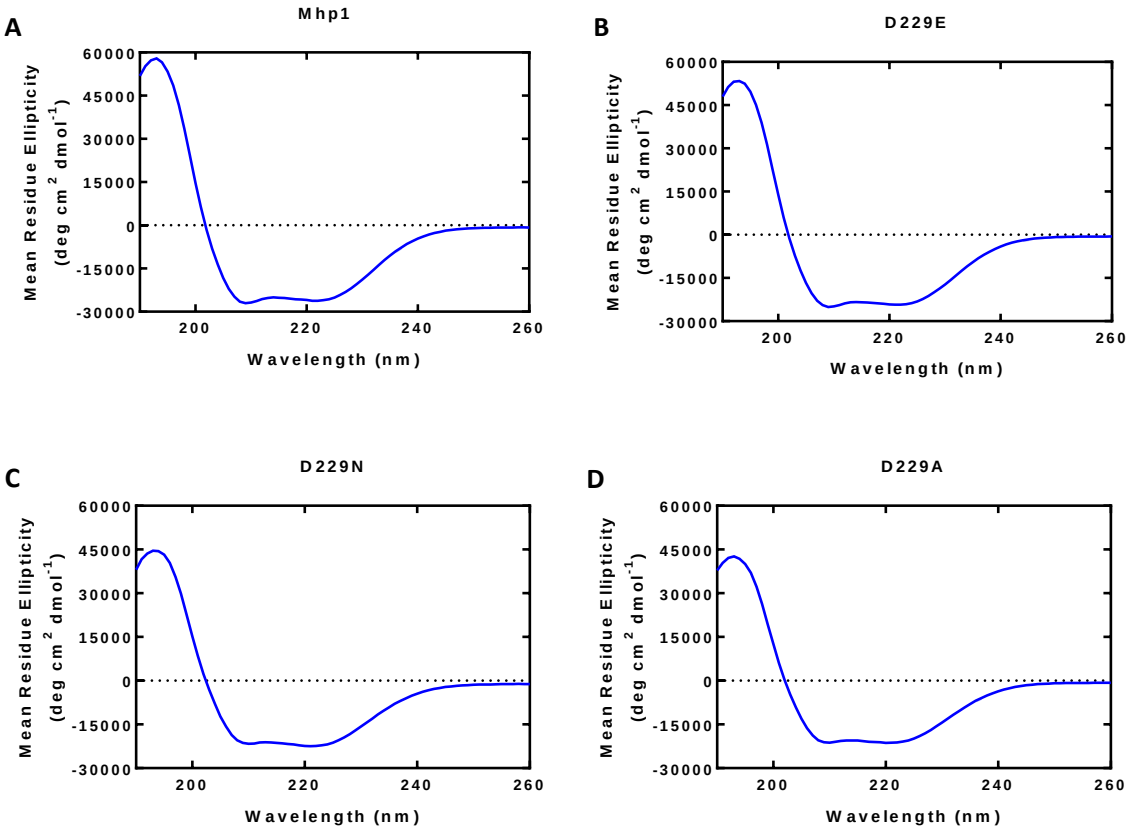



**Figure 3.14 Secondary structure spectra of purified proteins.** Far-UV (180-260 nm) CD spectroscopy was performed for purified proteins (A) Mhp1; (B) D229E; (C) D229N and (D) D229A. Measurements were taken using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant liquid nitrogen flushing. Samples were prepared as described in methods section 2.5.1 and analysed in a Hellma quartz cuvette of 1.0 mm pathlength in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM).

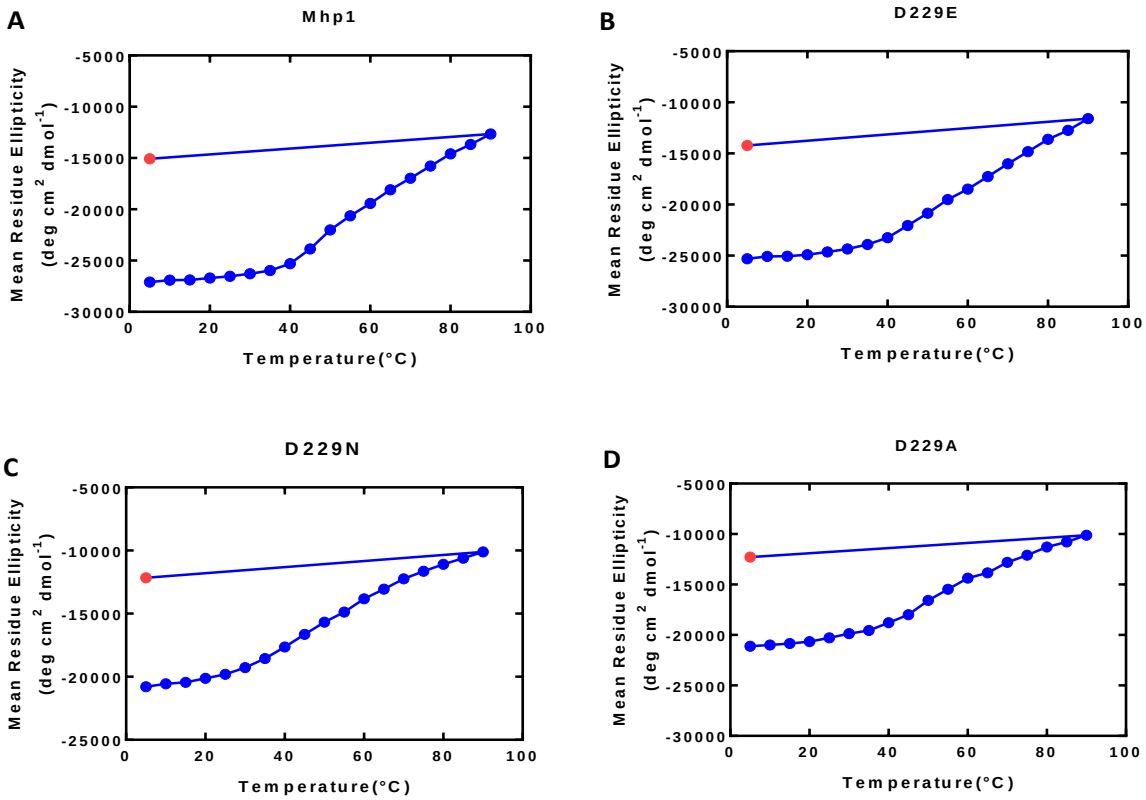



**Figure 3.15 Thermal stability analysis spectra of purified proteins using far UV CD.** Far-UV (180-260 nm) CD spectroscopy was performed for purified proteins (A) Mhp1; (B) D229E; (C) D229N and (D) D229A. Measurements were taken using a CHIRASCAN instrument (Applied Photophysics, UK) with constant liquid nitrogen flushing. Samples were prepared as described in methods section 2.5.1 and analysed in a Hellma quartz cuvette of 1.0 mm pathlength in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM). The change in the CD signal at 209 nm shows the thermal stability of the proteins. The thermal unfolding of proteins was analysed by ramping the temperature from 5-90ºC and the final spectrum was taken upon cooling of the sample back to 5ºC (red).

**Table 3.4 Estimated temperatures of thermal denaturation for the wild-type Mhp1 and its mutants D229E, D229N and D229A**

| Protein name | Estimated melting temperature (ºC) |
|---|---|
| Mhp1 | 48.9 |
| D229E | 48.9 |
| D229N | 48.8 |
| D229A | 48.9 |

## 3.9 Binding affinity of L-BH to purified Mhp1 mutants D229E, D229N and D229A

After successful protein purification of correctly folded proteins confirmed by secondary structure analyses, the binding affinities of these mutants for L-BH were probed using fluorimetry. Steady state spectrophotofluorimetry was used to establish the apparent $K_d$ values for binding of L-BH to mutants D229E, D229N and D229A, compared with wild-type Mhp1 by measuring the intrinsic tryptophan fluorescence changes (Table 3.5). Binding of L-BH produces the highest quenching in the intrinsic fluorescence of wild-type Mhp1 and comparable quenching was produced by the mutant D229E, whereas much lower quenching was exhibited by the mutants D229N and D229A. In the presence of sodium, apparent $K_d$ values were 0.085±0.006 mM and 0.043±0.003 mM for mutant D229E and the wild-type Mhp1, respectively. In the absence of sodium apparent $K_d$ values were 1.04±0.15 mM and 0.67±0.04 mM for mutant D229E and the wild-type Mhp1, respectively (Figure 3.16A). This suggests that mutating Asp229 to Glu has caused only a slight impairment of L-BH binding, which was expected due to their similar sizes and charges.

In mutants D229N and D229A, L-BH binding was substantially reduced compared to that for wild-type Mhp1 in the presence and absence of sodium chloride (Figure 3.16 B, Figure 3.16 C and Table 3.5). The apparent $K_d$ values in the presence and absence of sodium for the D229N and D229A mutants, are detailed in Table 3.5, which are much higher compared to those for wild type Mhp1. This suggests that mutation of Asp229 to Asn and Ala has caused a major perturbation in L-BH binding.

Wild-type Mhp1 binds L-BH with a higher affinity in the presence of sodium, whereas in the absence of sodium the protein still retains some activity for binding of L-BH. Mutants D229N and D229A showed very low binding affinity in the absence and presence of sodium demonstrating that these mutants could destabilise the inward-facing state of the protein. The importance of a carboxyl group at this position is evident from the results of mutating Asp229 to Glu (Figure 3.16A), which showed good binding activity when compared to the other two mutants. The mutation of Asp229 to Glu had retained some affinity for L-BH binding and it would

be interesting to mutate this residue to one with a positively charged side chain e.g. lysine, histidine or arginine that will likely abolish L-BH binding.

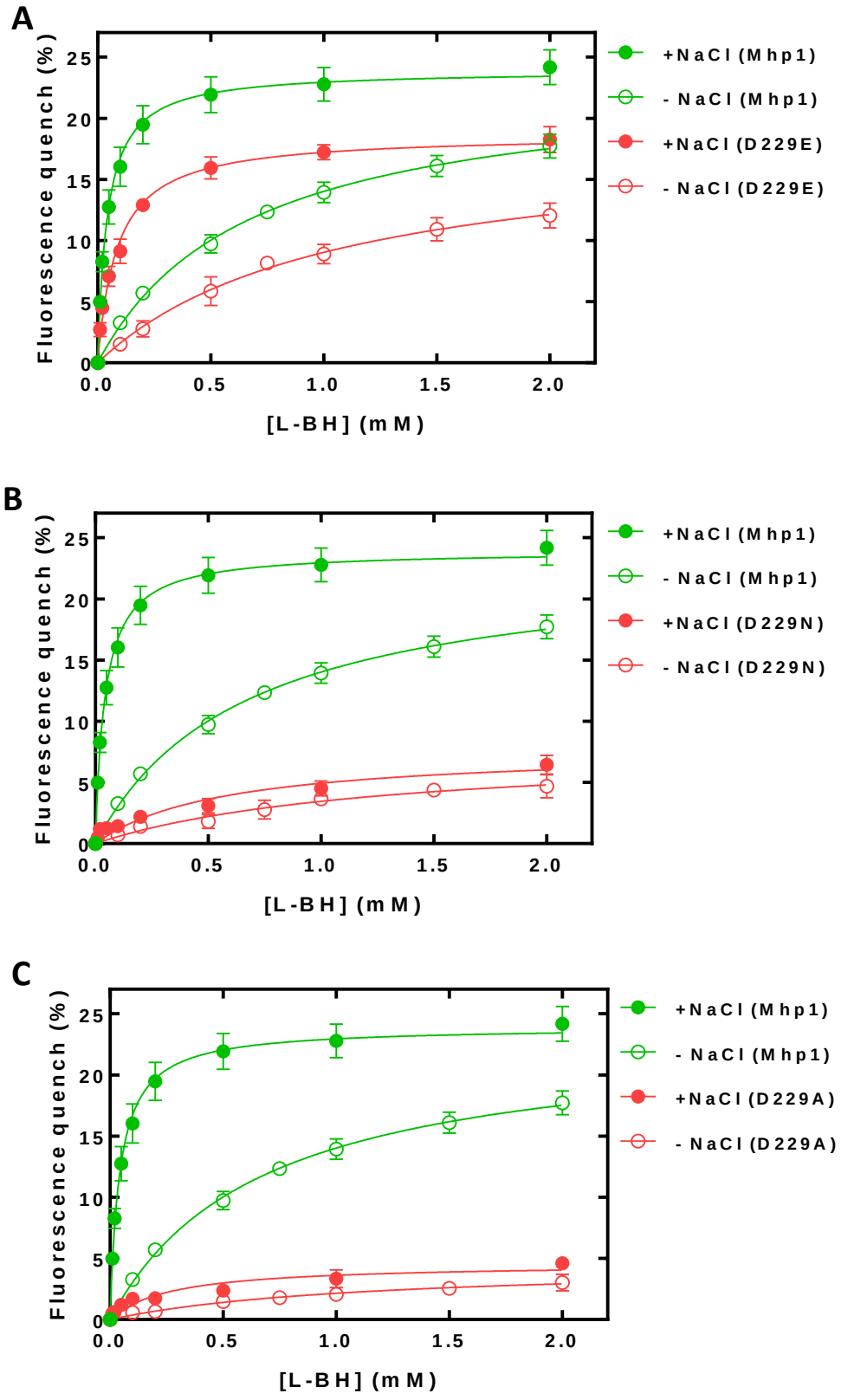


**Figure 3.16 Effect of mutating Asp229 on the binding affinity of L-BH.** Steady-state fluorimetry measurements on purified proteins (A) D229E, (B) D229N and (C) D229A were performed using a Photon Technology International spectrophotofluorimeter (section 2.5.2). Samples containing purified protein (140 µg/ml) in fluorescence buffer (Table 2.21) were titrated with increasing concentrations of L-BH (0-2 mM) for wild type (green) and mutant (red) in the presence of 15 mM NaCl (closed circles) and absence of sodium (open circles) at 18 $^{o}$C. Following ligand additions, samples were stirred for 1.5 minutes to equilibrate before making the measurements. Fluorescence was excited at a wavelength of 295 nm and quenching was measured at a wavelength of 332 nm. Data were analysed using the Michaelis-Menten analysis tool on Graph Pad Prism 7. The data represents a minimum of four repeats with error bars showing the standard error of the mean.

**Table 3.5 Comparison of $K_d$ and $\Delta F_{max}$ values for L-BH interacting with wild-type Mhp1 and mutants D229E, D229N and D229A using spectrophotofluorimetry.**

| Ligand | Apparent $K_d$ (mM) | | $\Delta F_{max}$ (%) | |
|---|---|---|---|---|
| | +NaCl | -NaCl | +NaCl | -NaCl |
| **Mhp1** | 0.043±0.003 | 0.67±0.04 | 23.93±0.38 | 23.35±0.66 |
| **D229E** | 0.085±0.006 | 1.04±0.15 | 18.69±0.36 | 18.45±1.32 |
| **D229N** | 0.565±0.169 | 1.26±0.42 | 7.72±0.90 | 7.84±1.33 |
| **D229A** | 0.660±0.064 | 1.25±0.36 | 4.58±0.35 | 4.56±0.72 |

## 3.10 Kinetic parameters for binding of L-BH to Mhp1 and mutants D229E, D229N and D229A

Stopped-flow fluorimetric measurements were performed (section 2.5.3) to monitor the rapid changes in fluorescence for L-BH binding to wild-type Mhp1 and mutants D229E, D229N and D229A in the absence and presence of 15 mM sodium chloride. A comparison of the rate constants and amplitudes of L-BH binding with those obtained previously for wild-type Mhp1 are shown in Figure 3.18. The fluorescence changes over time were plotted up to 10 seconds and the observed rate constant and amplitude of fluorescence change were calculated from these data (Figure 3.17). Conservative mutation of Asp229 to Glu has some effect on lowering the binding affinity for L-BH. This mutant retained almost similar sodium dependent activity as observed in wild-type Mhp1 (Figure 3.18A). In the presence of 15mM NaCl the effect on the apparent $K_d$ was only small giving values of 0.051±0.006 mM and 0.062±0.023 mM for wild-type and mutant D229E, respectively. In the absence of NaCl, the apparent $K_d$ values were 0.152±0.135 mM and 0.582±1.772 mM for wild-type and mutant D229E, respectively (Table 3.6). The results show a high rate of fluorescence change and amplitude in the presence of sodium chloride and a smaller rate of fluorescence change and amplitude in the absence of sodium chloride for both wild-type and D229E (Figure 3.18A, B). These results suggest that this mutation has slightly reduced the binding affinity for L-BH and retained a similar pattern of sodium dependent L-BH binding activity.

Mutation of Asp229 to Asn, which is a similar sized residue without a carboxyl group, resulted in drastic loss of binding activity for L-BH. Compared to wild-type a significantly increased $K_d$ value of 0.331±0.253 mM was produced in the presence of 15mM NaCl whereas in the absence of sodium the $K_d$ value was not measurable. These results suggest that a carboxyl group at this position is essential for L-BH binding. This was confirmed by mutation of Asp229 to Ala which also resulted in drastic loss of binding activity for L-BH, producing an even higher $K_d$ value of 1.17±1.4 mM in the presence of 15 mM NaCl (Table 3.6). Similarly in the absence of sodium its $K_d$ value was not measurable. The results of this mutation further validate the importance of a carboxyl group at this position. The effect of mutations was reflected in both the observed rate constant and amplitude of fluorescence changes for L-BH binding (Figure 3.18C, D; Figure 3.18 E, F).

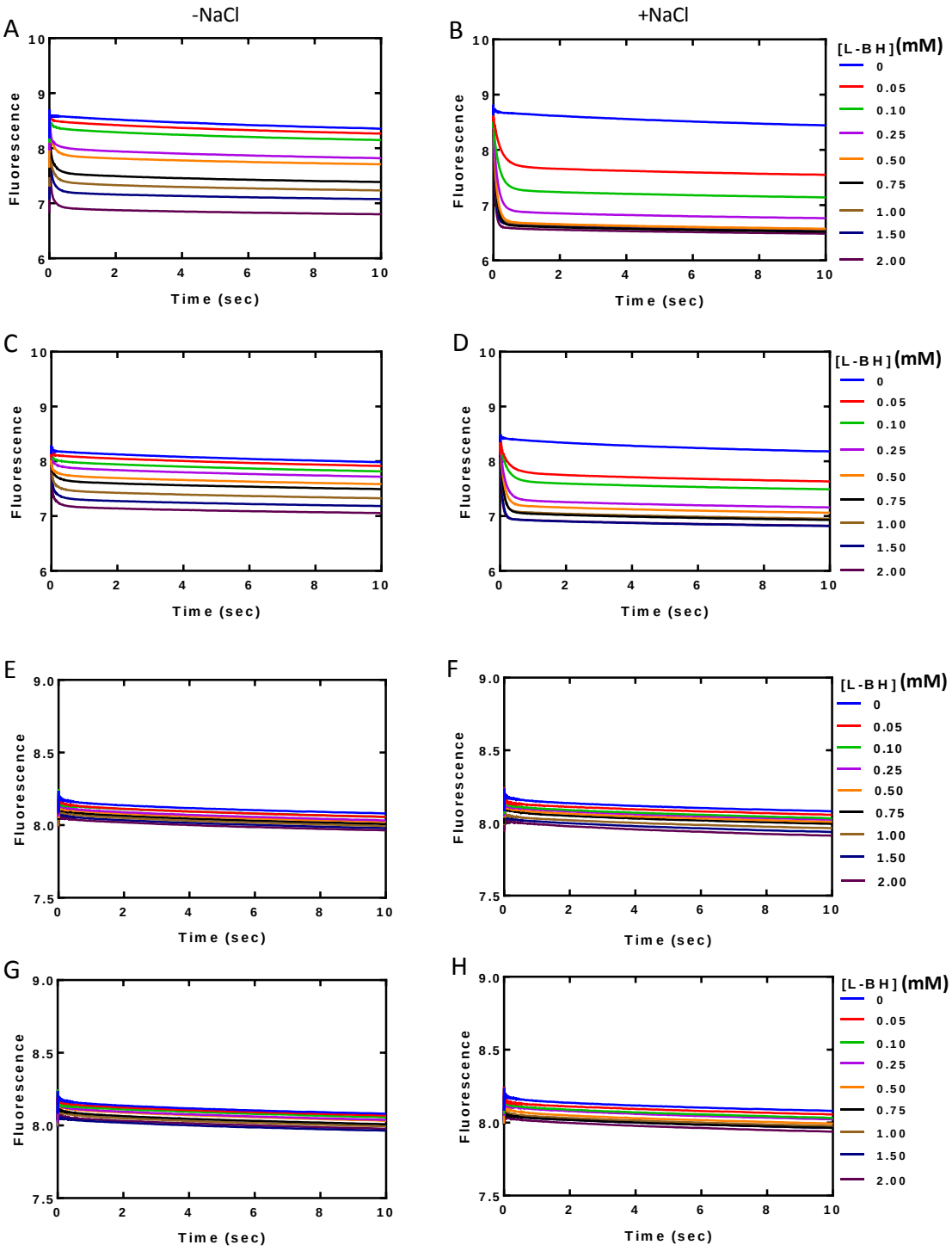



**Figure 3.17 Fluorescence change for binding of L-BH to Mhp1 and its mutants.** Using stopped-flow fluorimetry at 18 ºC purified Mhp1 (A, B); D229E (C, D) D229N (E, F) and D229A (G, H) in fluorescence buffer (Table2.22) in absence (left) and presence (right) of added NaCl (15 mM) was mixed with increasing concentrations of L-BH (0-2 mM) and the fluorescence response was monitored over 10 seconds following each addition.

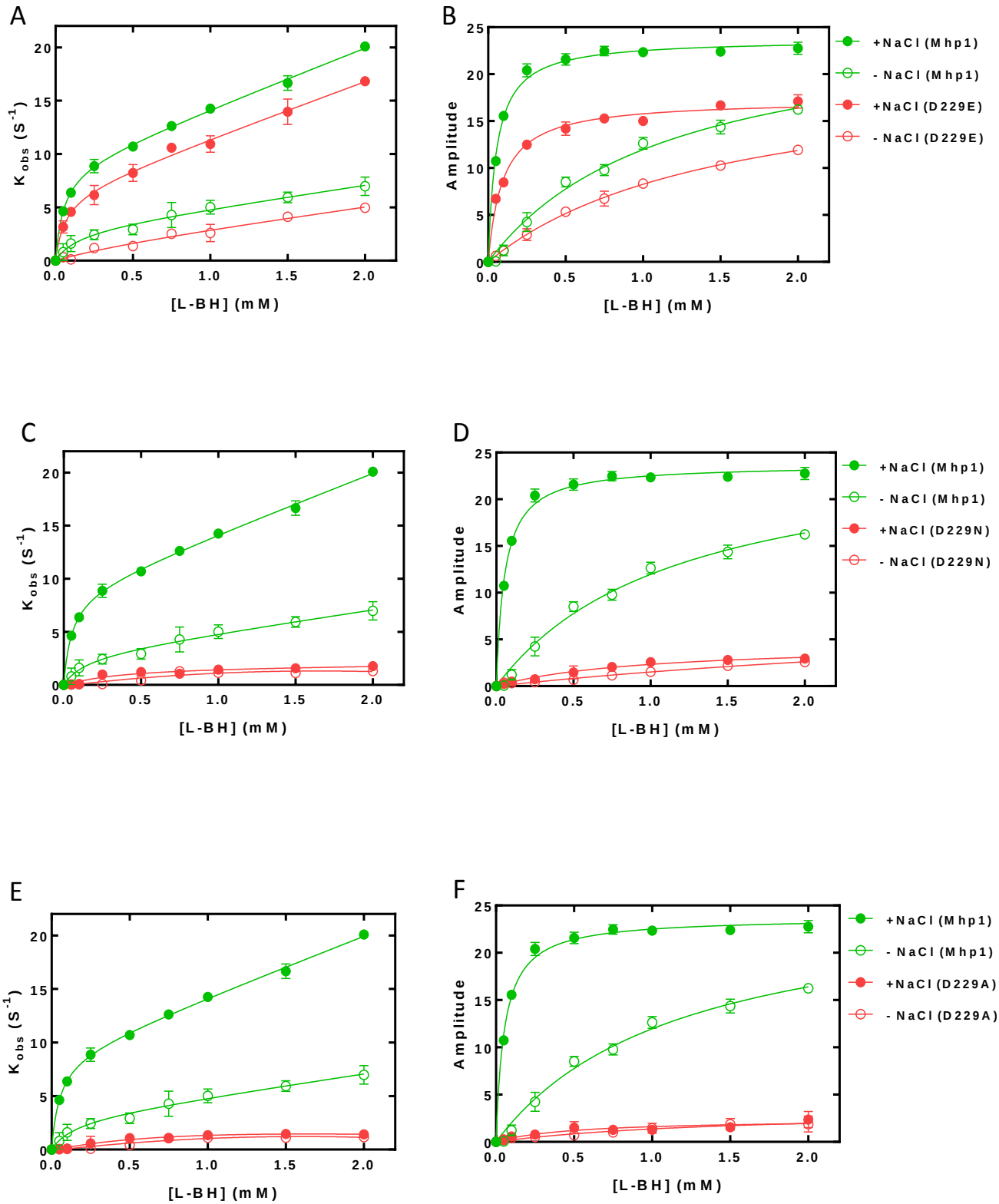

A
B
C
D
E
F
$K_{obs}$ ($S^{-1}$)
Amplitude
[L-BH] (mM)
+NaCl (Mhp1)
- NaCl (Mhp1)
+NaCl (D229E)
- NaCl (D229E)
+NaCl (D229N)
- NaCl (D229N)
+NaCl (D229A)
- NaCl (D229A)

**Figure 3.18 Kinetic parameters for L-BH binding to Mhp1 and mutants using stopped-flow fluorimetry.** Observed rate constant and amplitude of wild-type Mhp1 (green) were compared to those of its mutants (red). Mutant D229E rate constant (A), amplitude (B); D229N rate constant (C), amplitude (D); and D229A rate constant (E), amplitude (F) were observed using stopped-flow fluorimetry (SX-20 of Applied Photophysics, UK) at 18 ºC. Purified proteins of mutants (140µg/ml) were mixed with increasing concentrations of L-BH (0-2 mM) in fluorescence buffer (Table 2.22) in the presence of 15 mM NaCl (closed circles) or its absence (open circles). The fluorescence response was monitored over 10 seconds following each addition (Figure3.17). Values of $k_{obs}$ were calculated using a double exponential equation ($Y= Y_0 + A_1e^{-x/t1} + A_2e^{-x/t2}$) of OriginPro. The $K_d$ value of $k_{obs}$ in the presence and absence of sodium were determined by non-linear regression equation (Y=ΔFmax*X/(Kd+X) + NS*X + Background) on GraphPad prism 7.

**Table 3.6 Comparison of $K_d$ values for L-BH interacting with the wild-type Mhp1 and its mutants D229E, D229N and D229A using stopped-flow spectrophotofluorimetry.** ND indicates not determined.

| Ligand | Apparent $K_d$ (mM) | |
|---|---|---|
| | +NaCl | -NaCl |
| **Mhp1** | 0.051±0.006 | 0.152±0.135 |
| **D229E** | 0.062±0.023 | 0.582±1.772 |
| **D229N** | 0.331±0.253 | ND |
| **D229A** | 1.170±1.437 | ND |

## 3.11 Conclusions

This chapter focused on characterisation of wild-type Mhp1 protein and its mutants D229E, D229N and D229A. Wild-type Mhp1 protein was purified by immobilised metal affinity chromatography using a Ni-NTA resin. Far-UV circular dichroism measurements accomplished with Mhp1 protein demonstrate that the detergent-solubilised proteins had retained a typical alpha helical secondary structure after purification and was correctly folded. A melting temperature of 48.9 ºC was estimated using Global Analysis CD software-3 and therefore Mhp1 protein was reasonably stable for performing biophysical and biochemical assays in a temperature range of 18-25 ºC. The wild-type protein was an important control to be compared to each of the mutant proteins structural integrity and stability.

Ligand binding experiments have shown that DDM-solubilised Mhp1 protein was functionally active outside its native membrane environment. The cation specificity experiments showed high affinity of L-BH to Mhp1 in the presence of sodium confirming a sodium dependency as previously suggested (Weyand et al., 2008; Ma, 2010). All other cations (ChCl, LiCl, KCl, RbCl, CsCl) were much less of a stimulant for L-BH binding. Sodium dependent ligand binding of L-BH and D-BH to purified Mhp1 protein was confirmed using spectrophotofluorimetry. Stereo selectivity of Mhp1 was shown by binding to the L-enantiomers of BH in preference to the D-enantiomers. In the absence of sodium, the Mhp1 protein still has the binding affinity for L-BH but with less affinity whereas L-BH binding was dramatically increased in the presence of sodium. The binding of sodium may cause some conformational changes that create a higher affinity site for ligand binding. Weak binding affinity of various hydantoins compounds substituted at position 5' with an aliphatic group confirm that Mhp1 lacks broad ligand specificity. The higher binding affinity of L-BH to Mhp1 shows that hydantoin moieties with hydrophobic R groups bind with stronger affinity than those with other entities. These results consolidate the earlier reports that Mhp1 possesses a hydropobic ligand binding site and prefers ligands having hydrophobic R groups at the 5-position attached to the hydantoin moiety (Jackson, 2012; Simmon et al., 2014).

Site directed mutagenesis is an effective approach to investigate the role of specific residues in the structure and function of membrane proteins. The aspartic acid

residue at the position corresponding to Asp229 in Mhp1 is conserved amongst the NCS-1 family transporters. Its location in discontinuous helix 6 might also be important. Therefore, this residue was mutated to a variety of amino acids (glutamate, asparagine and alanine) with different side chains to investigate their possible effect on L-BH binding. To compare the impact of mutated proteins with wild type protein, various assays were performed and quantified including purification and yield of proteins, secondary structure integrity, thermal stability and ligand binding assays.

For purification, the same strategy was used for all mutants that was applied to the wild-type Mhp1. The highest yield and purity was obtained for the wild-type protein as compared to the mutants. Far-UV circular dichroism measurements were accomplished with the purified Mhp1 and its mutants D229E, D229N and D229A. The results demonstrated that all the detergent-solubilised proteins had retained alpha helical secondary structure after purification and mutations are not detrimental to the integrity of proteins. Thermostability trials exhibited that wild-type Mhp1 protein and D229 mutants had effectively identical thermal stability with melting temperatures of 48.9 ºC or 48.8 ºC as determined by Global Analysis CD software-3. In view of the CD measurements it is reasonable to conclude that all proteins are stable for performing biophysical assays using a temperature range of 18-25 ºC. Steady-state and stopped-flow spectrophotofluorimetry measurements were performed to observe the effect of Asp229 residue mutations on the L-BH binding of purified Mhp1 and mutants proteins. Drastic loss of binding activity for L-BH was observed for mutants Asp229 to Asn and Asp229 to Ala whilst the conservative mutation Asp229 to Glu has slightly impaired L-BH binding affinity. Based on the fluorimetry results, it can be concluded that a carboxyl group side chain of Asp229 residue is important for binding of L-BH.

# Chapter-4

# Cloning and expression of bacterial NCS1 family proteins

## 4.1 Introduction

This chapter describes a strategy for cloning and expression of Nucleobase Cation Symporter-1 (NCS1) family proteins. Currently the NCS1 family comprises over 2000 sequenced proteins derived from Gram-negative and Gram-positive bacteria, archaea, fungi and plants (de Koning and Diallinas, 2000; Pantazopoulou and Diallinas, 2007; Saier et al., 2009; Weyand et al., 2010; Witz et al., 2014; Krypotou et al., 2015; Ma et al., 2016; Sioupouli et al., 2017). These protein function as transporters for purine and pyrimidine nucleobases and nucleosides, hydantoins and other compounds including pyridoxine, thiamine and uric acid. Members of this family posses twelve transmembrane spanning alpha-helices (Saier et al., 2009; Witz et al., 2014) and function using a symport mechanism driven by a proton or sodium gradient (Krypotou et al., 2015). The structural model for the NCS-1 family is the sodium-coupled hydantoin transport protein, Mhp1 from *Microbacterium liquefaciens* (Suzuki and Henderson, 2006), for which crystal structures have been determined in three different conformations, i.e outward-facing open, occluded with substrate and inward-facing open (Weyand et al., 2008; Shimamura et al., 2010; Simmons et al., 2012). These structures of Mhp1 have provided the principal model for understanding the alternating access mechanism of membrane transport and for the mechanism of ion-coupling (Shimamura et al., 2010; Weyand et al., 2011; Adelman et al., 2011; Shi, 2013; Kazmier et al., 2014). The other well characterised bacterial NCS1 family proteins are the allantoin transporter PucI from *Bacillus subtilis* (Ma et al., 2016) and the cytosine transporter CodB from *E. coli* (Danielsen et al., 1995). The functionalities of many NCS1 proteins from fungi and plants have also been characterised experimentally (Schwacke et al., 2003; Krypotou et al., 2015). The fungal proteins are Fur-type transporters FurA (allantoin), FurD (uracil/uric acid), FurE (uracil/uric acid/allantoin), Fur4 (uracil), Dal4 (allantoin), Fui1 (uridine) and the Fcy-type transporters FcyB (purines/cytosine), Fcy2 (purines/cytosine), Thi7 (thiamine), Tpn1 (pyridoxine), Nrt1 (nicotinamide riboside) from *Aspergillus nidulans* and *Saccharomyces cerevisiae* (Hamari et al., 2009; Krypotou et al., 2015). The plant transporters are AtNCS1 (PLUTO) from *Arabidopsis thaliana* (adenine/guanine/uracil) (Mourad et al., 2012; Witz et al., 2014), CtNCS1 from *Chlamydomonas reinhardtii* (adenine/guanine/uracil/allantoin) (Schein et al., 2013), ZmNCS1 from *Zea mays* (adenine/guanine/cytosine) and

SvNCS1 from *Setaria viridis* (adenine/guanine/hypoxanthine/cytosine/allantoin) (Rapp et al., 2016). In order to provide further information about the functions and relationships of NCS1 proteins, the aim of the work in this chapter was to clone and express further bacterial members of the family for functional and structural studies. This used a well-established cloning strategy for bacterial transport proteins using plasmid pTTQ18.

## 4.2 Cloning strategy for bacterial membrane transport proteins

The cloning strategy used in this project was based on a traditional restriction enzyme-based method involving digestion of both vector and amplified DNA fragments with the relevant restriction enzymes to enable DNA ligation. The vector selected for production of the constructs was pTTQ18 (Stark, 1987; Ward et al., 2000; Ward et al., 2001; Saidijam et al., 2003) (Figure 4.1), which has been used successfully for amplifying expression of numerous bacterial and archaeal membrane transport proteins (Hoyle, 2000; Ward et al., 2000; Saidijam et al., 2003; Saidijam et al., 2005; Suzuki and Henderson, 2006; Szakonyi et al., 2007; Ma et al., 2013; Bettaney et al., 2013; Ma et al., 2016). The pTTQ18 plasmid (Figure 4.1) contains polylinker/*lac*Zα region flanked by a hybrid *trp-lac* (*tac*) promoter. The tac promoter consists of the -35 region of the trp RNA polymerase binding site fused with the lacUV5 -10 repressor binding region of the *lac* promoter. Downstream of the *tac* promoter is a multicloning site, which permits the use of *Eco*RI restriction enzyme site at the 5' end of the amplified gene and *Pst*I enzyme site at the 3' end of the gene, for successful ligation. pTTQ18 also contains the *bla* gene for expression of ß-lactamase, conferring resistance to carbenicillin, and the *lac*$I^q$ gene for amplified expression of the LacI repressor to prevent derepression of the plasmid in the absence of inducer.

The chosen affinity tag was RGS$(His)_6$, which has been added to a modified pTTQ18 and is placed at the C-terminus of the protein (Hoyle, 2000). The location of the C-terminus of the protein is important since previous experience has shown that a periplasmic C-terminus generally renders expression unsuccessful. This is likely caused by an inability of the hydrophilic histidine tag to cross the hydrophobic membrane. In cases where the C terminus of the target membrane protein is

predicted to be periplasmic, introduction of a C-terminal Strep-tag can be successful if the RGS(His)$_6$-tag is not (Szakonyi et al., 2007). Before cloning, it was therefore necessary to assess the predicted topology of the protein and investigate whether the location of the C-terminus is expected to be cytoplasmic or periplasmic using topology prediction tools TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001) and TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009). The location of the RGS(His)$_6$ tag on the plasmid also dictates use of the *Pst*I restriction site to enable correct fusion of the protein with the tag. If *Pst*I cannot be used because of an internal *Pst*I site within the gene of interest, then the RGS(His)$_6$ tag can instead be added at the primer level, and the *Hin*dIII restriction site used for restriction/ligation cloning.

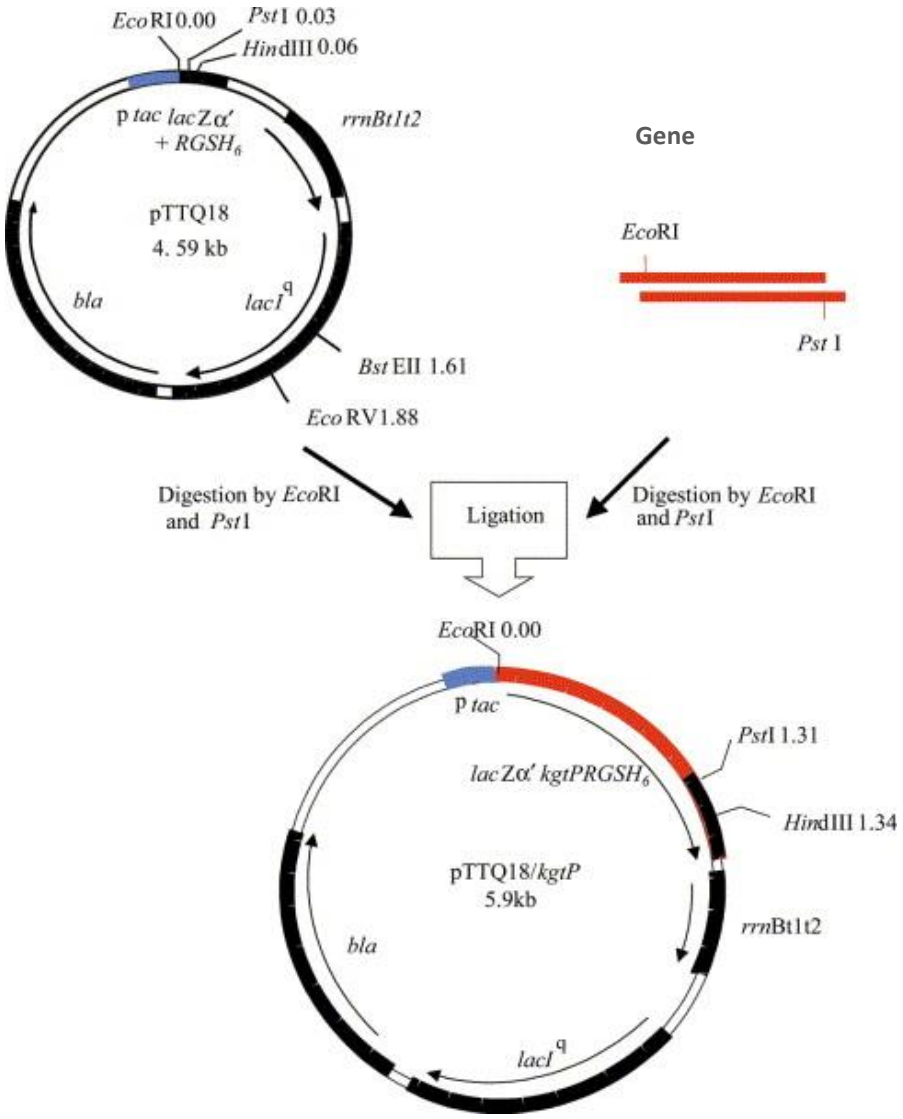



**Figure 4.1 Cloning strategy for membrane proteins using plasmid pTTQ18.**
This figure was reproduced from Saidijam et al., (2003).

## 4.3 Selection of bacterial NCS1 proteins for cloning and characterisation

Selection of bacterial NCS1 family proteins for cloning and characterisation was made by searches of the databases UniProt KnowledgeBase (http://www.uniprot.org/), NCBI (http://www.ncbi.nlm.nih.gov/) and the Transporter Protein Analysis Database (http://www.membranetransport.org/) (Ren et al., 2006). NCS1 proteins that had not previously been cloned or characterised were chosen based on a number of criteria. The availability of genomic DNA as the template for gene cloning dictated the bacterial species/strain of origin that could be chosen. Proteins with putative substrates that we had available in radiolabelled form to allow testing of transport activity was the next important criterion for choosing the desired proteins. In order to allow the use of the restriction enzymes *Eco*RI and *Pst*I for cloning with plasmid pTTQ18, gene sequences were checked for presence of internal restriction sites that would be cut by these enzymes using the tool Webcutter 2 (http://rna.lundberg.gu.se/cutter2/). The putative number of transmembrane spanning alpha-helices in the proteins and the predicted locations of N- and C-termini was determined using topology prediction tools to determine if a condidate protein has a predicted membrane topology and structural organisation consistent with characterised members of bacterial NCS1 family proteins. Based on previous literatures on cloning of similar proteins, it was important that the C-terminus was predicted to be at the cytoplasmic side of the membrane to allow the cloning strategy with the C-terminal $His_6$-tag to be successful. Finally, PCR primers of suitable properties were designed for amplifying expression of the genes of interest. Based on these criteria, which are listed below, eleven proteins were chosen for cloning. Details of these proteins are given in Tables 4.1 and 4.2.

- Check for availability of genomic DNA
- Check for availability of radiolabelled substrates
- Check gene sequences for undesirable internal restriction sites
- Perform membrane topology predictions to determine the putative number of transmembrane spanning α-helices and locations of N- and C-termin
- Design PCR primers based on a cloning strategy using plasmid pTTQ18
- Calculate the protein sequence identities and similarity with the characterised bacterial NCS1 proteins, Mhp1, PucI and CodB.

**Table 4.1 Bacterial NCS-1 family proteins**. Characterised bacterial NCS-1 family proteins and proteins chosen for cloning, expression, purification and characterisation. The name refers to the gene name taken from the NCBI database. The sizes of proteins and putative substrates are also taken from the NCBI database. Sequence identities with Mhp1, PucI and CodB were calculated using the alignment tool Clustal Omega (Sievers and Higgins, 2014).

| No. | Protein name (bacterium) | Size of protein (amino acids) | Known/putative substrate(s) | Identity (%)<br>Mhp1 (top)<br>PucI (middle)<br>CodB (bottom) |
|---|---|---|---|---|
| 0 | Mhp1<br>*Microbacterium liquefaciens*<br>PucI<br>*Bacillus subtilis*<br>CodB<br>*E. coli* | 501<br><br>490<br><br>419 | Hydantoins<br><br>Allantoin<br><br>Cytosine | 100.0<br><br>100.0<br><br>100.0 |
| 1 | CUB18073<br>*Bacillus cereus* | 456 | Allantoin | 20.1<br>23.8<br>20.3 |
| 2 | EIQ13585<br>*Shigella flexneri* | 484 | Allantoin | 25.5<br>37.0<br>20.6 |
| 3 | CJK90608<br>*Streptococcus pneumoniae* | 490 | Allantoin | 25.2<br>34.0<br>23.7 |
| 4 | COI77568<br>*Streptococcus pneumoniae* | 449 | Purine-cytosine | 18.3<br>23.2<br>22.9 |
| 5 | CAC11736<br>*Thermoplasma acidophilum* | 460 | Purine-cytosine | 23.6<br>23.8<br>22.3 |
| 6 | VPA1242<br>*Vibrio parahaemolyticus* | 412 | Cytosine | 22.8<br>24.1<br>75.2 |
| 7 | NMB2067<br>*Neisseria meningitides* | 407 | Hydroxymethyl-pyrimidine | 22.6<br>22.5<br>24.7 |
| 8 | BAA80379<br>*Aeropyrum pernix* | 450 | Nucleosides | 24.1<br>23.4<br>24.0 |
| 9 | AAN69889<br>*Pseudomonas putida* | 510 | Nucleosides | 27.0<br>33.8<br>22.6 |
| 10 | EFQ62020<br>*Pseudomonas fluorescens* | 509 | Nucleosides | 27.3<br>35.1<br>24.1 |
| 11 | ELQ14219<br>*Pseudomonas syringae* | 500 | Nucleosides | 22.1<br>22.3<br>23.1 |

**Table 4.2 Details of topology predictions for the chosen NCS1 family proteins.** Amino acid sequences of the NCS1 family proteins given in Table 1 were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and/or from NCBI (http://www.ncbi.nlm.nih.gov/) and analysed by the membrane topology prediction tools TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001) and TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009). The predicted positions of the N- and C-termini and the number of transmembrane spanning α-helices are shown (Appendix 3).

| Protein name | Predicted position of | | | | Number of predicted transmembrane spanning α-helices | |
|---|---|---|---|---|---|---|
| | N-terminus | | C-terminus | | | |
| | TMHMM | TOPCONS | TMHMM | TOPCONS | TMHMM | TOPCONS |
| CUB18073 | Inside | Inside | Inside | Inside | 12 | 12 |
| EIQ13585 | Inside | Inside | Inside | Inside | 12 | 12 |
| CJK90608 | Inside | Inside | Inside | Inside | 12 | 12 |
| COI77568 | Inside | Inside | Inside | Inside | 12 | 12 |
| CAC11736 | Inside | Inside | Inside | Inside | 12 | 12 |
| VPA1242 | Outside | Inside | Inside | Inside | 11 | 12 |
| NMB2067 | Inside | Inside | Outside | Inside | 11 | 12 |
| BAA80379 | Inside | Inside | Inside | Inside | 12 | 12 |
| AAN69889 | Outside | Inside | Inside | Inside | 11 | 12 |
| EFQ62020 | Outside | Inside | Inside | Inside | 11 | 12 |
| ELQ14219 | Inside | Inside | Inside | Inside | 12 | 12 |

## 4.4 Evolutionary relationships of chosen NCS1 family proteins

In order to demonstrate the evolutionary relationships of NCS1 family proteins chosen for cloning, the phylogenetic analysis was performed for the eleven proteins (Table 4.1) alongside the characterised bacterial NCS1 family proteins Mhp1, PucI and CodB (Figure 4.2). Based on this analysis, the protein most closely related to Mhp1 is the putative nucleoside transporter (BAA80379) from *Aeropyrum pernix* with a combined sequence homology of 43.9% (24.1% identical, 19.7% highly similar). The protein most closely related to PucI is putative allantoin transporter (EIQ13585) from *Shigella flexneri* with the combined sequence homology of 65.5% (37.0% identical, 28.5% highly similar). The protein most closely related to CodB is the putative cytosine transporter (VPA1242) from *Vibrio parahaemolyticus* with a combined sequence homology of 89.0% (75.2% identical, 13.8% highly similar). Interestingly, the three putative allantoin transporters (CUB18073, EIQ13585, CJK90608) are not clustered together and also are not the nearest neighbours of PucI. Similarly, the two putative purine-cytosine transporters (COI77568, CAC11736) and the four putative nucleoside transporters (BAA80379, AAN69889, EFQ62020, ELQ14219) are not clustered with each other. In the whole analysis, the two most closely related proteins are the putative nucleoside transporters AAN69889 and EFQ62020 with a shared sequence identity of 80.3%.

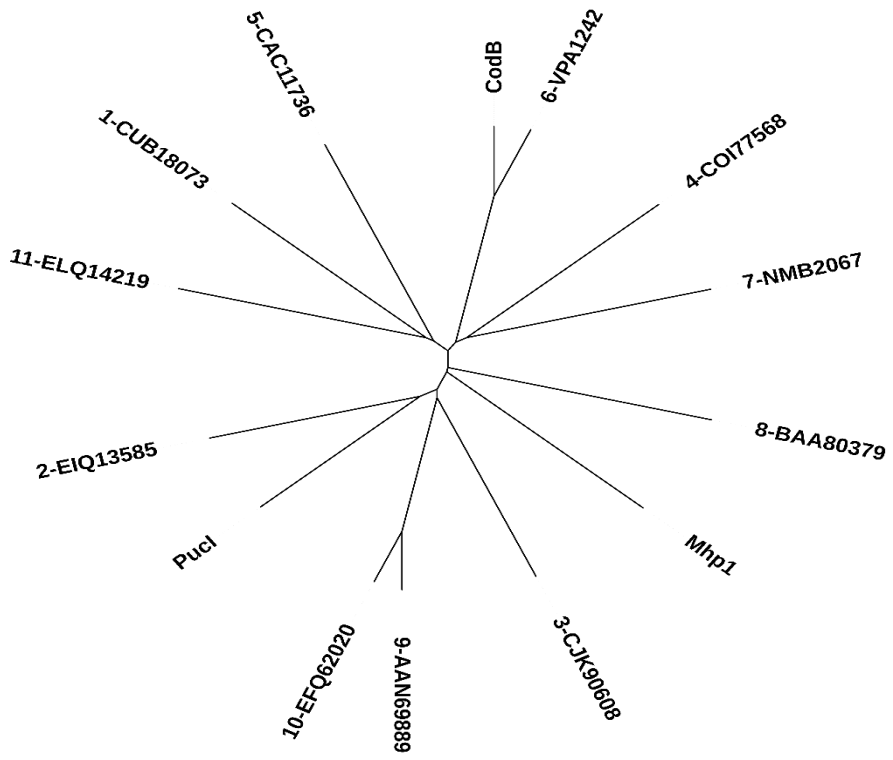



**Figure 4.2 Phylogenetic tree of chosen and characterised bacterial NCS1 family proteins.** Amino acid sequences of the NCS1 family proteins given in Table 4.1 were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and/or from NCBI (http://www.ncbi.nlm.nih.gov/) and aligned using the multiple alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). The resultant nearest-neighbour phylogenetic tree was exported in Newick format and re-drawn using iTol: Interactive Tree Of Life (http://itol.embl.de/) (Letunic and Bork, 2016).

## 4.5 Isolation of NCS1 family genes from bacterial genomic DNA using PCR

Out of the eleven bacterial genes selected from the NCS1 family, six were successfully amplified by PCR (Figure 4.3). All six amplified genes were successfully ligated into plasmid pTTQ18 as confirmed by restriction digestion analysis (Figure 4.4) and DNA sequencing (Appendix 4). Recombinant clones were transformed into *E. coli* BL21(DE3) cells for expression trials. Amplification of the two genes from *Streptococcus pneumoniae* was not successful probably due to the high GC content in the primers. Various conditions might be helpful such as addition of DMSO and a lowered annealing temperature.

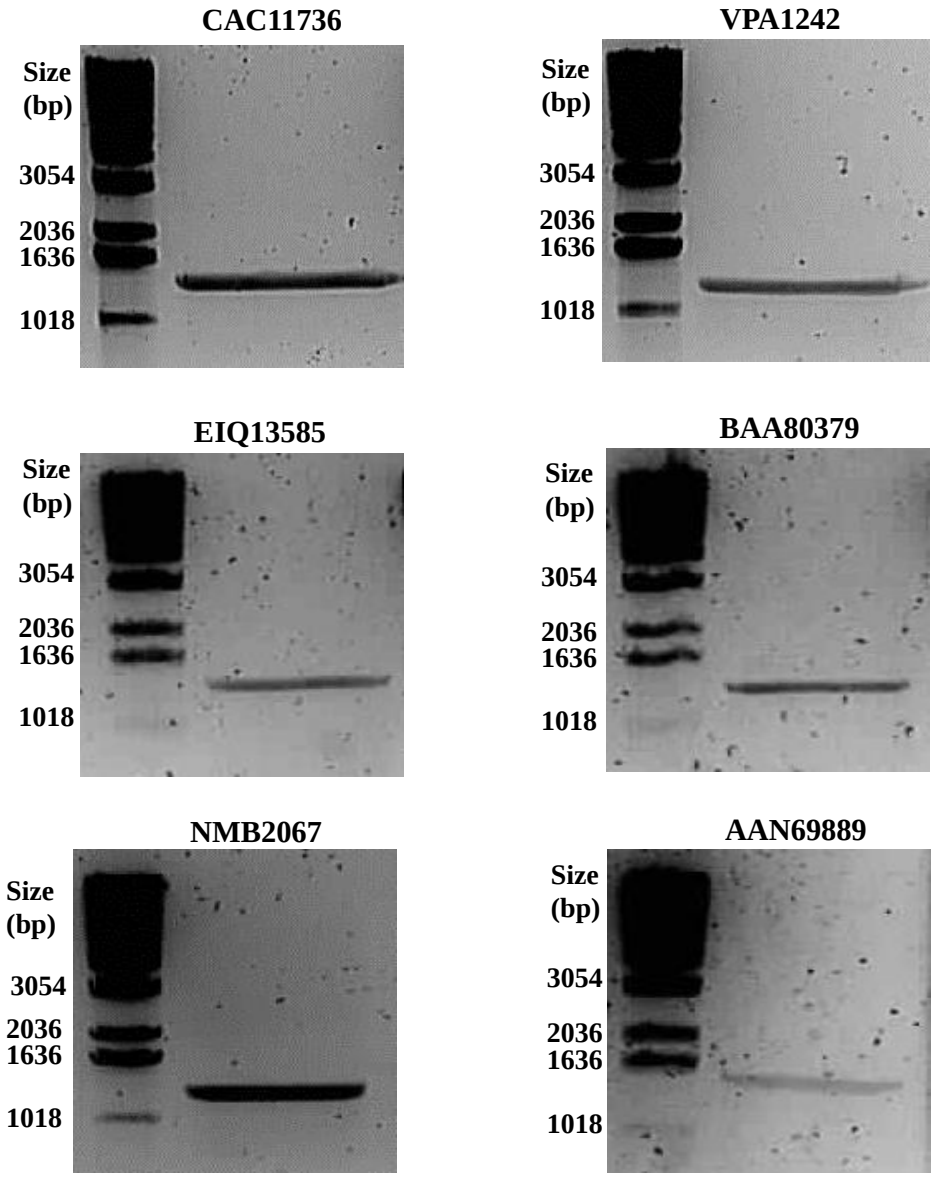



**Figure 4.3 PCR amplification of NCS1 family genes.** PCR products were analysed on a 1% agarose gel. Lanes: 1 kb DNA ladder (left) and PCR products (right).

### 4.5.1 Digestion, ligation and transformation of pTTQ18

Plasmid pTTQ18 containing gene *kgtP* was isolated from expression strain BL21(DE3) (Section 2.3.5)and transformed (Section 2.3.12) into the cloning strain Omnimax. Plasmid DNA isolated from the Omnimax cells was digested (Section 2.3.9) with *Eco*RI and *Pst*I enzymes and then electropheresed on a 1% agarose gel in order to remove the reaction enzymes and extract the digested pTTQ18 (Figure 4.4).

Amplified genes and plasmid pTTQ18 were ligated with T4 DNA ligase (Section 2.3.10). Omnimax cells were transformed with the ligation products and grown on carbenicillin-containing plates.

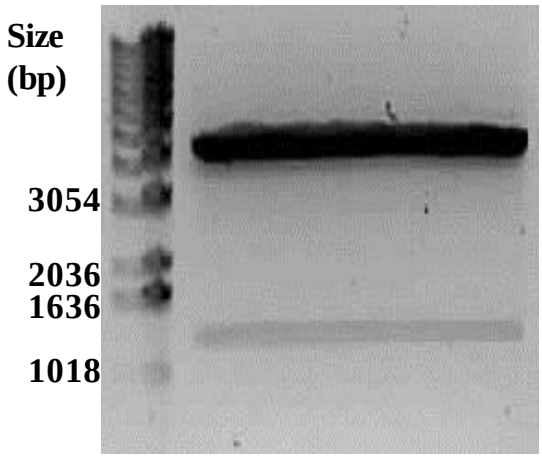



**Figure 4.4. Digestion and purification of pTTQ18 from an agarose gel.** Plasmid pTTQ18 was digested with *Eco*RI and *Pst*I for 1 hour at 37 °C. Samples were loaded as 1kb DNA ladder (1) and digested plasmid pTTQ18 (2). The top band is the digested pTTQ18 used later in ligation reactions and the bottom band is the excised gene *kgtP*.

### 4.5.2 Restriction digestion analysis for identification of correct plasmids

Up to four colonies from each plate were picked and grown in carbenicillin-containing LB liquid medium for plasmid purification. To identify the plasmids containing the correct gene, restriction digestion analysis was performed using *Eco*RI and *Pst*I enzymes and samples were analysed on a 1% agarose gel (Figure-4.5) followed by automated DNA sequencing. The gene EIQ13585 was successfully amplified by PCR as confirmed by its band on the agarose gel (Figure 4.3) but sequencing data failed. Cloning of this gene could be repeated. The other five plasmids were transformed into *E. coli* BL21(DE3) cells for expression level testing.

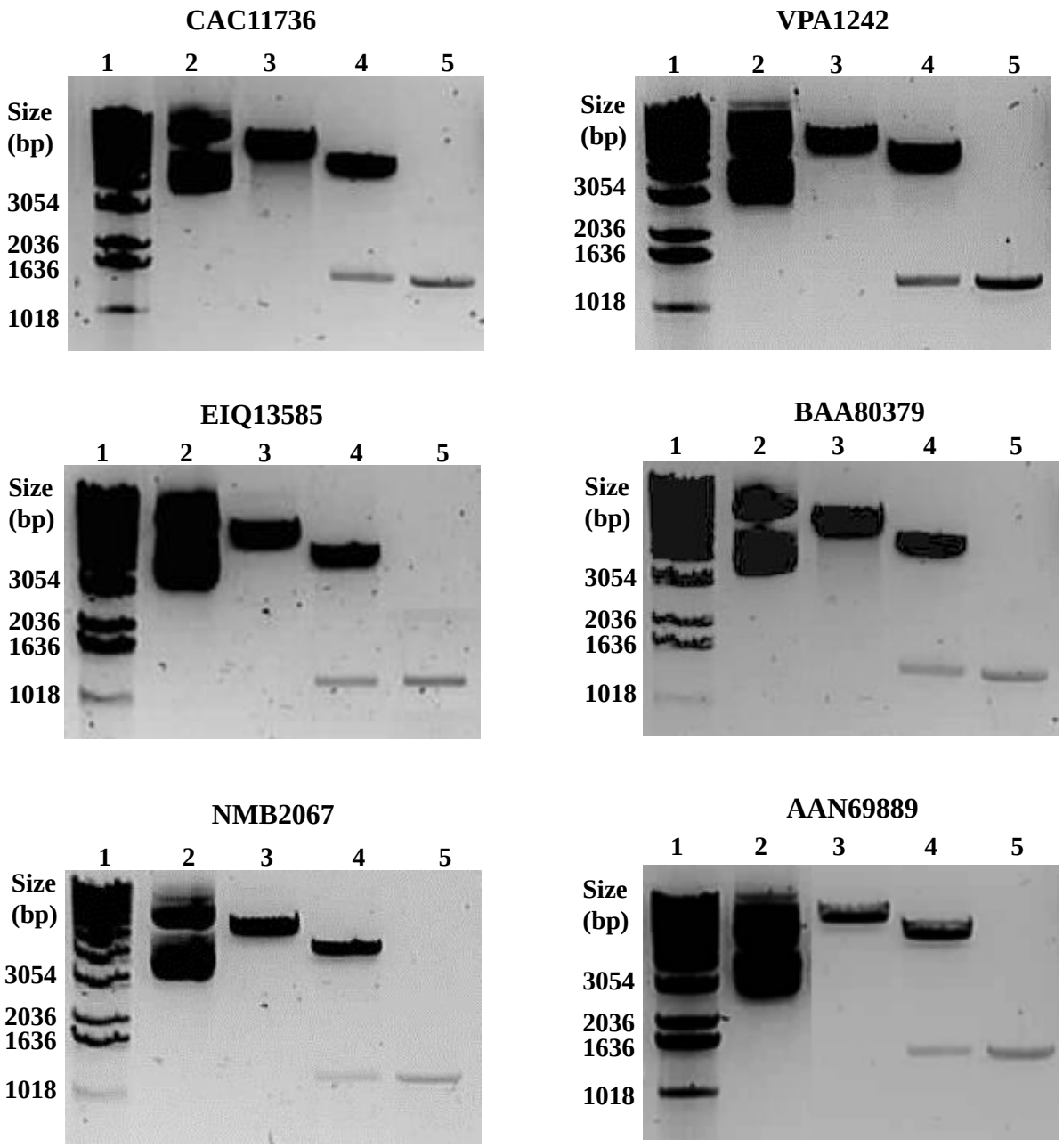



**Figure 4.5 Restriction digestion analysis of isolated plasmids to confirm the approximate size of gene insert.** Isolated pTTQ18 plasmids containing the putative gene were digested following electrophoresis on an 1% agarose gel. The lanes were loaded as follow: (1) 1 kb DNA ladder; (2) undigested pTTQ18; (3) *Eco*RI digested pTTQ18; (4) *Eco*RI and *Pst*I digested pTTQ18; (5) *Eco*RI and *Pst*I digested PCR product.

### 4.6 Detection of putative proteins and determination of expression level

Five proteins were tested for expression from small scale cultures. BL21(DE3) cells transformed with the isolated plasmids were grown in LB medium (50 ml) supplemented with carbenicillin (100 µg/ml) up to $OD_{680}$ = 0.6. Cells were initially induced with IPTG (0.5 mM) and then grown for a further 2 hours. Harvested cells were subjected to the water lysis method to obtain mixed membranes. Only two proteins with amplified expression (AAN69889 and VPA1242) were detected on both the SDS-PAGE gel and Western blot (Figure 4.6 and 4.7).

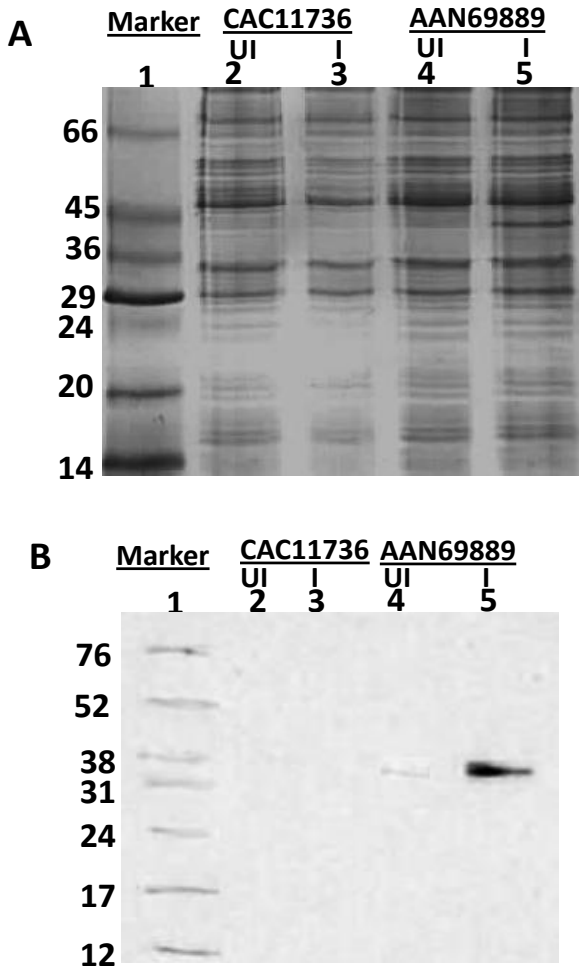



**Figure 4.6 Test for amplified expression of CAC11736 and AAN69889.** SDS-PAGE (A) and Western blot (B) analysis of total membrane fractions of uninduced (UI) and induced (I) cells. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) CAC11736 (*Thermoplasma acidophilum*) uninduced; (3) CAC11736 (*Thermoplasma acidophilum*); induced; (4) AAN69889 (*Pseudomonas putida*) uninduced; (5) AAN69889 (*Pseudomonas putida*) induced.

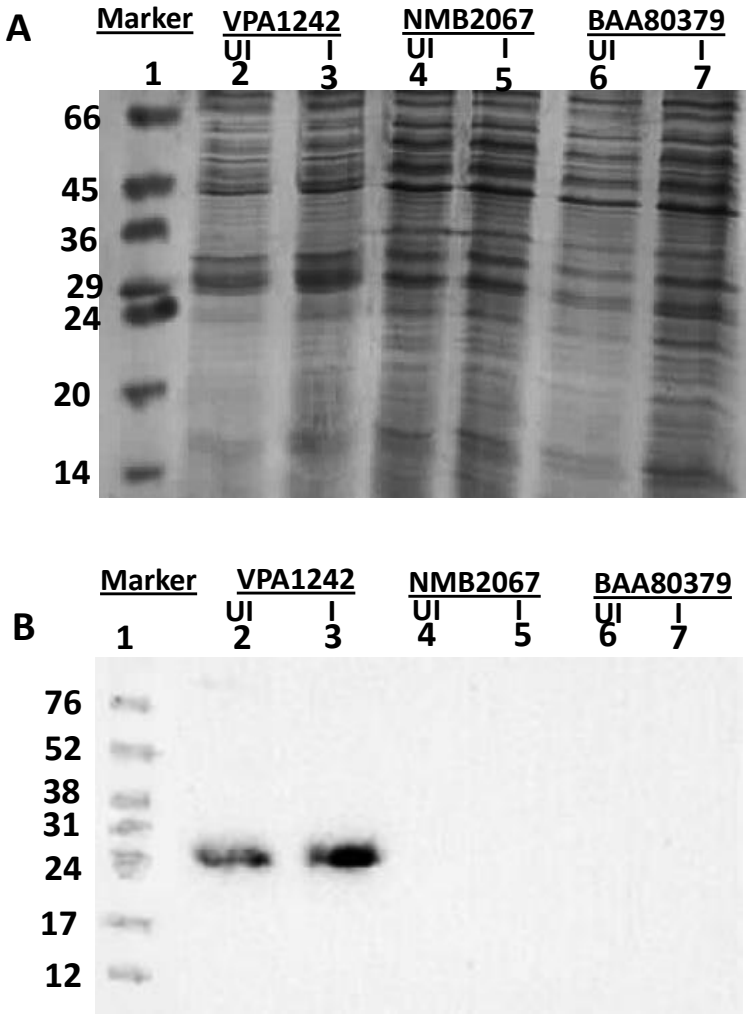



**Figure 4.7 Test for amplified expression of VPA1242, NMB2067 and BAA80379.** SDS- PAGE (A) and Western blot (B) analysis of total membrane fractions of uninduced (UI) and induced (I) cells. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) VPA1242 (*Vibrio parahaemolyticus*) uninduced; (3) VPA1242 (*Vibrio parahaemolyticus*) induced; (4) NMB2067 (*Neisseria meningitides*) uninduced; (5) NMB2067 (*Neisseria meningitides*) induced; (6) BAA80379 (*Aeropyrum pernix*) uninduced; (7) BAA80379 (*Aeropyrum pernix*) induced.

### 4.6.1 Test of proteins expression in *E. coli* strain BL21 star (DE3)

The results in section 4.6 demonstrate that BL21 (DE3) strain was not ideal for expression of CAC11736, NMB2067 and BAA80379 proteins. There could be number of problems of not expressing proteins but some of the most common typical hurdles are instability of mRNA, presence of high number of rare codons, degradation of protein, premature termination, formation of inclusion bodies, different vectors, various host and their growth conditions etc (Streatfield, 2007; Angov, 2011; Mauro et al., 2014; Rosano and Ceccarelli, 2014; Jia and Jeon, 2016). To find a solution for expression problems at the level of strains, BL21-star (DE3) cells were considered. The BL21-star (DE3) is a derivative of BL21 (DE3) strain that improves the mRNA stability due to a mutation in the RNaseE gene (rne131), which decreases the production of endogenous RNases and reduce the mRNA degradation, therefore enhanced stability of mRNA transcripts results in increased protein yield. This strain has been sugessted for higher levels of heterologous expression due to improvement of mRNAs stability and the higher yield of recombinant protein (Kido et al., 1996; Lopez et al., 1999). Thus, the three proteins that failed to express in BL21 (DE3) cells were tested for expression in another *E. coli* strain i.e. BL21-star (DE3). The cells were grown as described in section 2.2.5. Harvested cells were subjected to the water lysis method to obtain mixed membranes (section 2.4.1). Two proteins (NMB2067 and BAA80379) were now detected by the SDS-PAGE gel and Western blot analysis whereas the protein (CAC11736) was still not spotted (Figure 4.8). Protein expression level of BAA80379 was low as compared to NMB2067 and needs further optimisation. The expression of NMB2067 and BAA80379 in BL21-star (DE3) suggests that mRNAs stability is one of the mechanisms involved in their expression and the RNAse deficient strain (BL21-star) has the capability of solving this problem.

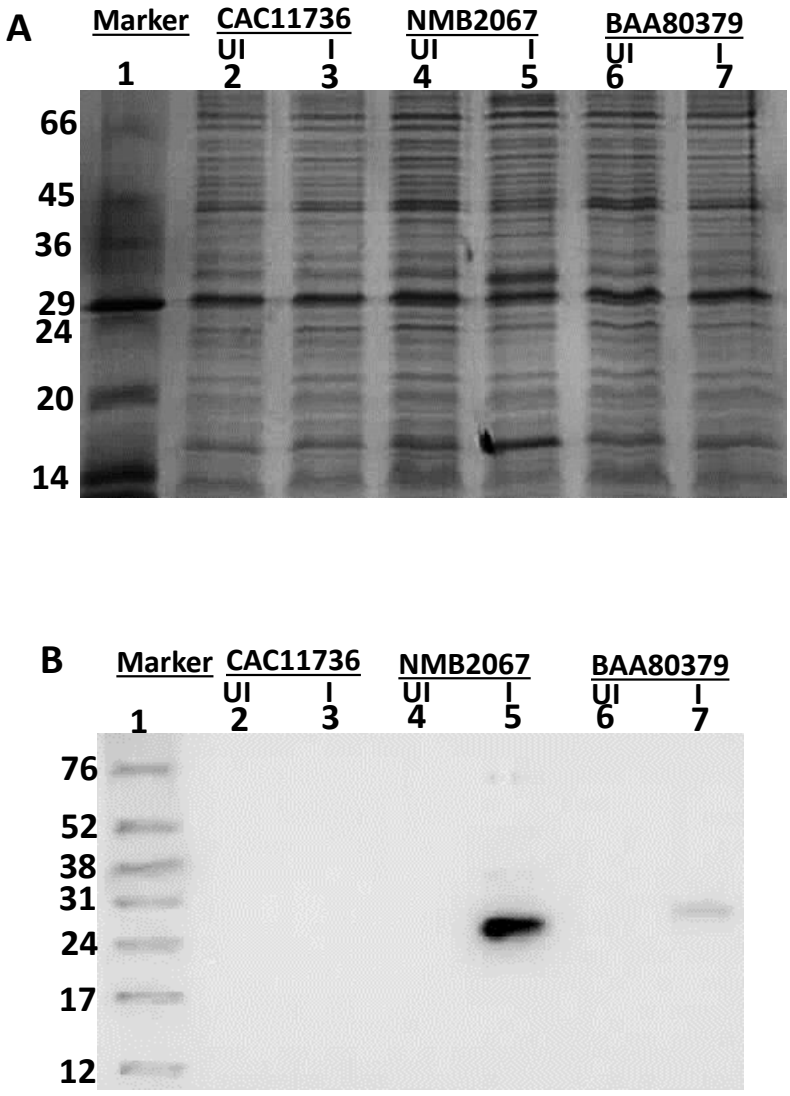



**Figure 4.8 Test for amplified expression of CAC11736, NMB2067 and BAA80379.** SDS- PAGE (A) and Western blot (B) analysis of total membrane fractions of uninduced (UI) and induced (I) cells. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) CAC11736 (*Thermoplasma acidophilum*) uninduced; (3) CAC11736 (*Thermoplasma acidophilum*) induced; (4) NMB2067 (*Neisseria meningitides*) uninduced; (5) NMB2067 (*Neisseria meningitides*) induced; (6) BAA80379 (*Aeropyrum pernix*) uninduced; (7) BAA80379 (*Aeropyrum pernix*) induced.

### 4.6.2 Expression test of CAC11736 (*Thermoplasma acidophilum*) in different strains of *E. coli*

The CAC11736 protein was intractable to expression in the former strains i.e. (BL21(DE3) and BL21-star (DE3). Use of different strains of *E. coli* with desirable characteristics is one of the strategies for optimal overexpression of membrane proteins. Various organisms use codons at different frequencies so to check rare codon usage problem, four different strains of *E. coli* were selected (BL21-star (DE3) pRARE2; BL21-gold (DE3) pRARE2; C41(DE3) pRARE2; and C43(DE3) pRARE2). Indeed, these *E. coli* strains have strong proven history of successful prokaryotic integral membrane proteins overexpression. Each strain harbouring pRARE2 plasmids was activated for rare codon optimisation (Burgess-Brown et al., 2008). The BL21-star (DE3) is the *E. coli* strain that increases the stability of mRNAs while BL21 gold (DE3) is an improved versions of BL21 competent cells lacking the Lon and OmpT proteases, which degrade recombinant proteins (Jerpseth et al., 1998). The C41(DE3) and C43(DE3) are mutant strains of BL21(DE3) and far superior than the parental strain for the overexpression of many membrane proteins (Miroux and Walker, 1996; Dumon-Seignovert et al., 2004). The strain C41(DE3) has a mutation, which enable to prevent cell death related with overexpression of various recombinant toxic proteins. The strain C43(DE3) is derived from C41(DE3) and can express a different set of toxic proteins to C41(DE3). Regardless of these efforts, none of the strains was able to express the CAC11736 protein. Further use of the most appropriate vector (different promoters and tags) and optimising expression conditions (temperature range, induction level, induction concentration, various media's) can be the keys for successful membrane proteins expression in *E. coli*, and need to be tested.

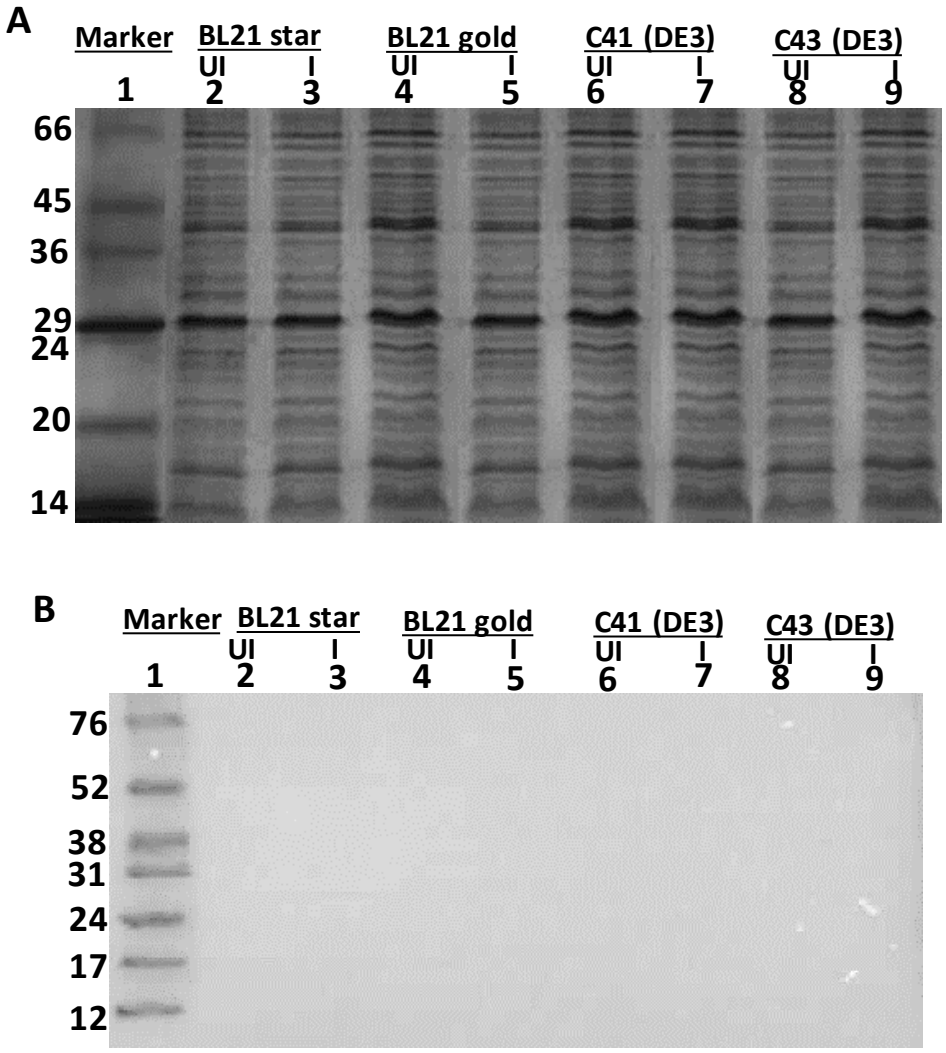



**Figure 4.9 Expression test of CAC11736 (*Thermoplasma acidophilum*) in various strains of *E. coli*.** SDS- PAGE (A) and Western blot (B) analysis of total membrane fractions of uninduced (UI) and induced (I) cells. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) CAC11736 uninduced in BL21 star (DE3), pRARE2; (3) CAC11736 induced in BL21 star (DE3), pRARE2; (4) CAC11736 uninduced in BL21 gold (DE3), pRARE2; (5) CAC11736 induced in BL21 gold (DE3), pRARE2; (6) CAC11736 uninduced in BL21 C41 (DE3), pRARE2; (7) CAC11736 induced in BL21 C41 (DE3), pRARE2; (8) CAC11736 uninduced in BL21 C43 (DE3), pRARE2; (9) CAC11736 induced in BL21 C43 (DE3), pRARE2.

**Table 4.3 Summary of progress for cloning, amplified expression, purification and characterisation of bacterial NCS1 family proteins.** Out of eleven proteins, six were successfully cloned, five had correct sequencing and were tested for expression, four proteins NMB2067, BAA80379, VPA1242 and AAN69889 were successfully expressed. Two proteins VPA1242 and AAN69889 were further characterised (Chapter 5). *EIQ13585 was cloned but sequencing was incorrect.

| Protein name | Cloned | Expression |
|---|---|---|
| CUB18073 | No | - |
| EIQ13585 | Yes* | - |
| CJK90608 | No | - |
| COI77568 | No | - |
| CAC11736 | Yes | No |
| VPA1242 | Yes | Yes |
| NMB2067 | Yes | Yes |
| BAA80379 | Yes | Yes |
| AAN69889 | Yes | Yes |
| EFQ62020 | No | - |
| ELQ14219 | No | - |

## 4.7 Conclusions

The only NCS1 family protein with high-resolution crystal structures available is Mhp1 (Weyand et al., 2008; Shimamura et al., 2010; Simmons et al., 2014). A genomic approach was therefore used here for studying further bacterial NCS1 family transporters. Eleven protein homologous of Mhp1 were selected for cloning, amplified expression, purification and characterisation (Table 4.1). *In silico* analysis of these proteins suggested that most of them have twelve transmembrane spanning α-helices and possess cytoplasmic N- and C- terminal ends (Table 4.2 and Appendix 3). Six of the selected proteins were successfully cloned from genomic DNA using PCR with introduction of the $His_6$-tag at the C-terminus. The PCR primers designed for use in this study introduced *Eco*RI and *Pst*I restriction sites at the 5` and 3` ends of the gene, respectively. Following successful amplification, the resulting DNA fragment was purified and the integrities of the cloned genes were established by automated DNA sequencing to confirm that each gene had been cloned without mutation and was inserted into pTTQ18 with a correct orientation. Of the six cloned genes, five were successfully sequenced while one sequence was incorrect. The pTTQ18 plasmids containing the sequenced genes were used to transform *E. coli* BL21(DE3) cells for expression studies. Out of the five proteins initially only two (AAN69889 and VPA1242) showed successful amplified expression according to the SDS-PAGE gel and Western blot analysis of small-scale membrane preparations (Figure 4.6 and 4.7). Three proteins (CAC11736, NMB2067 and BAA80379) that failed to express in BL21 (DE3) cells were tested for expression in *E. coli* strain BL21-star (DE3). Two of these proteins (NMB2067 and BAA80379) were detected by the SDS-PAGE gel and Western blot analysis after expression in this strain, whereas the third protein (CAC11736) was not spotted (Figure 4.8). To check rare codon usage problem and toxicity, four different strains of *E. coli* for CAC11736 expression were selected (BL21 star (DE3) pRARE2; BL21 gold (DE3) pRARE2; C41(DE3) pRARE2; C43(DE3) pRARE2). Regardless of these efforts, none of the strains was able to express the CAC11736 protein (Figure 4.9). There could be many possible reasons for unsuccessful expression of the successfully cloned gene. These might be an appropriate expression vector, appropriate host or overexpression might cause aggregation or lack of correct chaperones, thus limiting the level of protein being expressed. The two proteins (AAN69889 and VPA1242) were further

subjected to expression optimisation, large-scale membrane preparation, purification and functional characterisation (Chapter 5).

# Chapter-5

# Characterisation of NCS1 family proteins AAN69889 and VPA1242

## 5.1 Introduction to *Pseudomonas putida*

The opportunistic Gram negative bacterium *Pseudomonas putida* is the paradigm for a subclass of proteobacteria found widespread in terrestrial and aquatic environments (Palleroni, 1984). The important metabolic activities of these bacteria include element cycling and the degradation of biogenic and xenobiotic pollutants (Timmis, 2002). Pseudomonas species have great potential for biotechnological applications, especially in areas of bioremediation (Dejonghe et al., 2001), biocatalysis (Schmid et al., 2001), as biocontrol agents in plant protection (Walsh et al., 2001) and for production of novel bioplastics (Kahlon, 2016). *P. putida* strain KT2440 is the best characterised pseudomonas species they all have and was the first Gram negative soil bacterium to be certified as a safety strain by the Recombinant DNA Advisory Committee (Register, 1982).

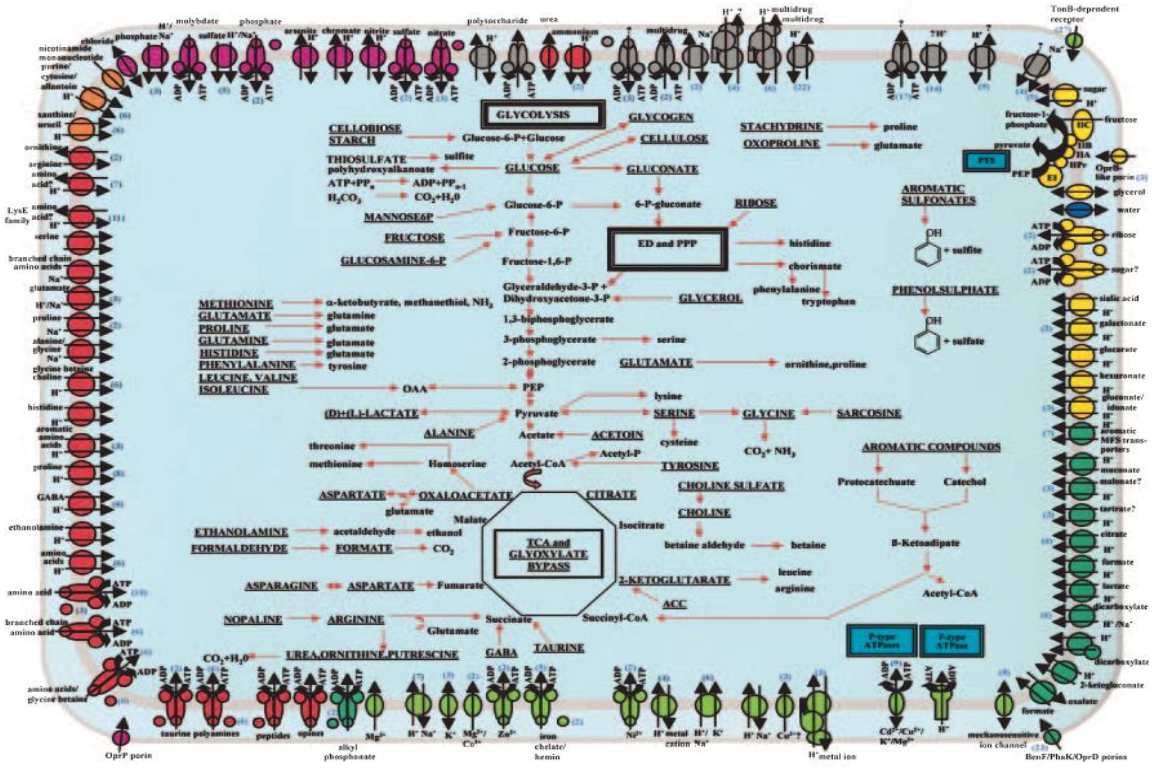



**Figure 5.1 Overview of transport and metabolism in *Pseudomonas putida* strain KT2440.** Predicted transporters are grouped by substrate specificity: inorganic anions (pink), inorganic cations (light green), amino acids/peptides/amines/ purines/pyrimidines and other nitrogenous compounds (red), carbohydrates (yellow), drug efflux and other (dark grey), water (blue), carboxylates, aromatic compounds and other carbon sources (dark green). Uncertainty about the substrate transported is indicated by question marks.This figure was reproduced from Nelson et al., (2002).

It has been estimated that *P. putida* strain KT2440 possesses 350 cytoplasmic membrane transport systems, five of these are classed as belonging to the NCS1 family (Figure 5.1) (Nelson et al., 2002). According to the Pseudomonas Genome Database (http://beta.pseudomonas.com/) (Winsor et al., 2015), the gene that codes for the NCS1 protein AAN69889 appears at the locus position PP_4309, which predicts a cytoplasmic membrane protein that is putatively a nucleoside transporter. Interestingly, this gene is directly upstream from gene locus PP_4310 that codes for a hydantoin racemase (Figure 5.2). The UniProt KnowledgeBase (http://www.uniprot.org/) entry for this protein (Q88EZ1) also suggests that it is a putative nucleoside transporter of the NCS1 family. AAN69889 contains 510 amino acids that putatively form 11 or 12 transmembrane spanning alpha-helices according to membrane topology predictions (Figure 5.3). Based on structurally characterised NCS1 family proteins and the demonstrated reliability for the TOPCONS prediction tool (Hennerdal and Elofsson, 2011), twelve transmembrane helices are most likely to be correct. In this case, both the N- and C-terminal ends of the protein are predicted to occur on the cytoplasmic side of the membrane.

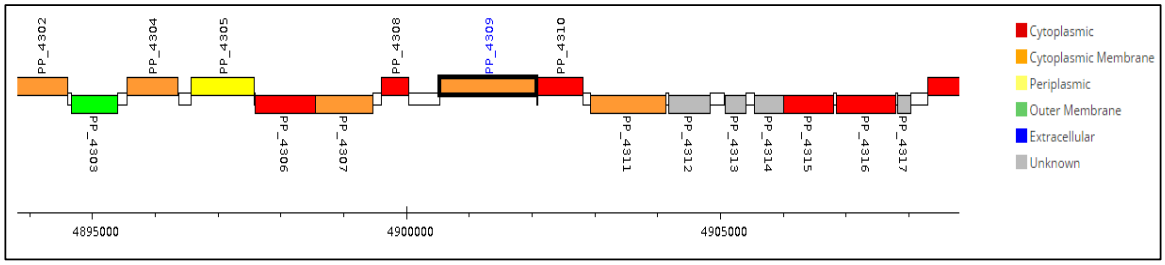



**Figure 5.2 Map of the *Pseudomonas putida* AAN69889 region.** This diagram was taken from the Pseudomonas Genome Database (http://beta.pseudomonas.com/) (Winsor et al., 2015).

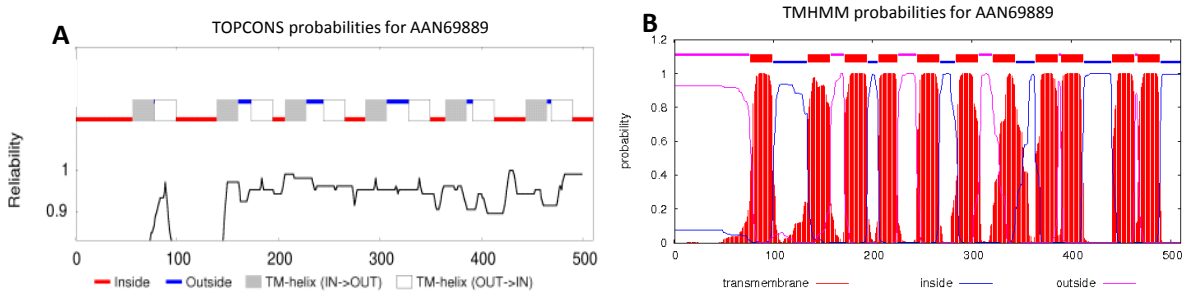



**Figure 5.3 Predictions of transmembrane helices in AAN69889.** The amino acid sequence of AAN69889 was analysed by the membrane topology prediction tools TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009) (A) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001) (B).

## 5.2 Conservation of residues between AAN69889 and characterised bacterial NCS1 family proteins

In the absence of a high-resolution crystal structure, analysis of conserved residues between the protein of interest and closely homologous characterised proteins can give clues about its function and structural organisation. Based on separate sequence alignments with characterised bacterial NCS1 family proteins, AAN69889 from *P. putida* shares overall homologies of 50.8% (27.0% identical, 23.8% highly similar) with Mhp1 from *Microbacterium liquefaciens*, 58.5% (33.8% identical, 24.7% highly similar) with PucI from *Bacillus subtilis* and 46.7% (22.6% identical, 24.1% highly similar) with CodB from *E. coli* (Table 4.1). This confirms that AAN69889 is most closely related with allantoin transporter PucI, as demonstrated by the phylogenetic analysis shown in Figure 4.2. An alignment between AAN69889, Mhp1 and PucI identifies the residues that are conserved between these three proteins (Figure 5.4), which are also shown in a topology diagram of AAN69889 based on the putative positions of transmembrane helices given by the TOPCONS tool (Figure 5.5). This includes twelve residues in the sodium and substrate binding site of Mhp1 based on crystal structures. Of these, the positions corresponding to Mhp1 residues Trp117, Gln121, Asn314, Asn318 and Leu363 are identically conserved in AAN69889 and in PucI. The position corresponding to Mhp1 residue Gln42 is occupied by asparagine in both AAN69889 and PucI. The position corresponding to Mhp1 residue Ile41 is not conserved and is occupied by histidine in both AAN69889 and PucI. The position corresponding to Mhp1 residue Trp220 is conserved in PucI and replaced by tyrosine in AAN69889. Overall, the conservation of these residues in AAN69889 more closely match to those in PucI than Mhp1, therefore suggesting that AAN69889 could be a transporter of allantoin.

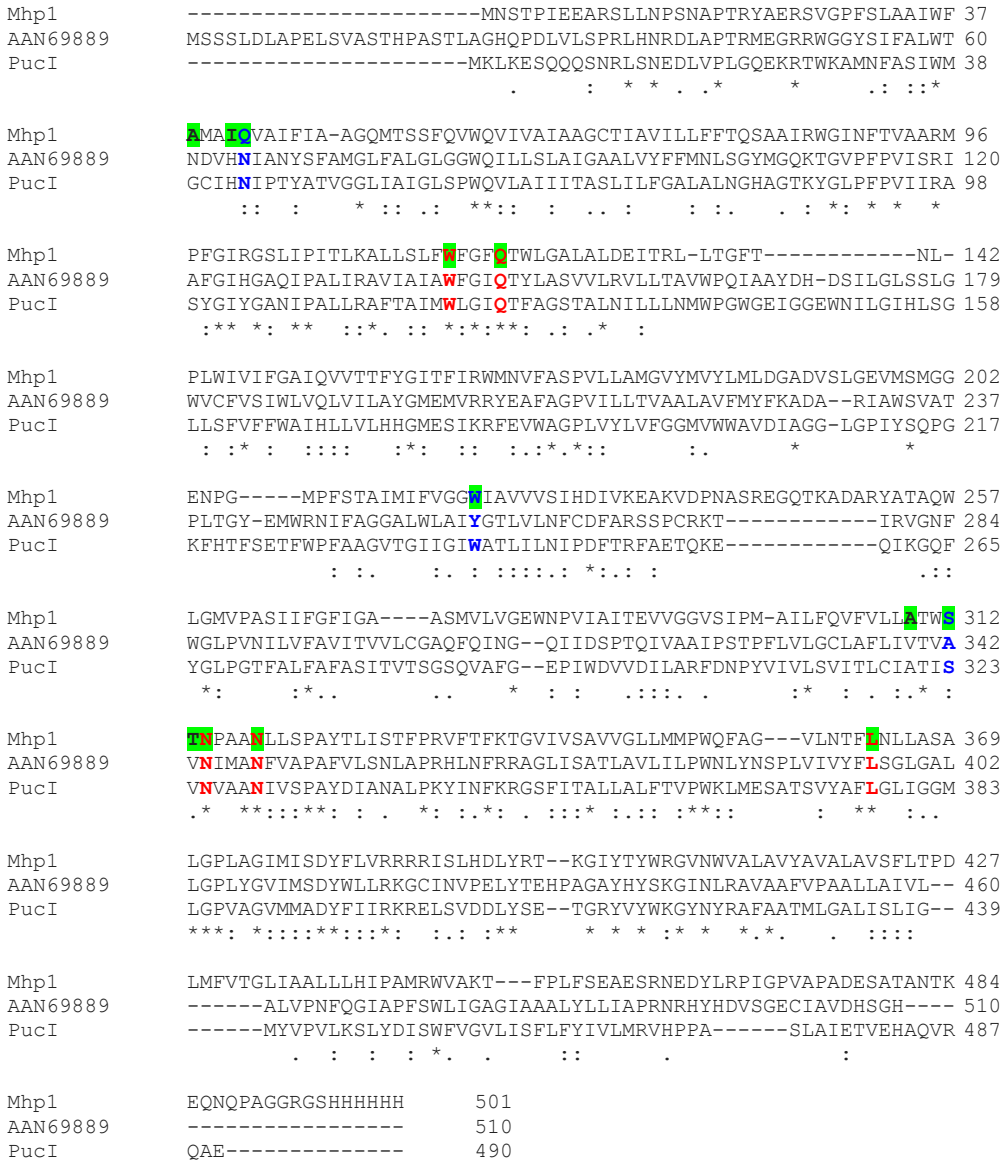



**Figure 5.4 Conservation of residues between the AAN69889 protein of *Pseudomonas putida*, Mhp1 and PucI.** Amino acid sequences of AAN69889 from *Pseudomonas putida*, Mhp1 from *Microbacterium liquefaciuens* and PucI from *Bacillus subtilis* were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and aligned using the multiple sequence alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Conserved residues are indicated below the sequences as identical (*), highly similar (:) and similar (.). Residues in the ligand and sodium binding site of Mhp1 are highlighted in green and those in AAN69889 and PucI that are identical or highly similar in this region are coloured red and blue, respectively.

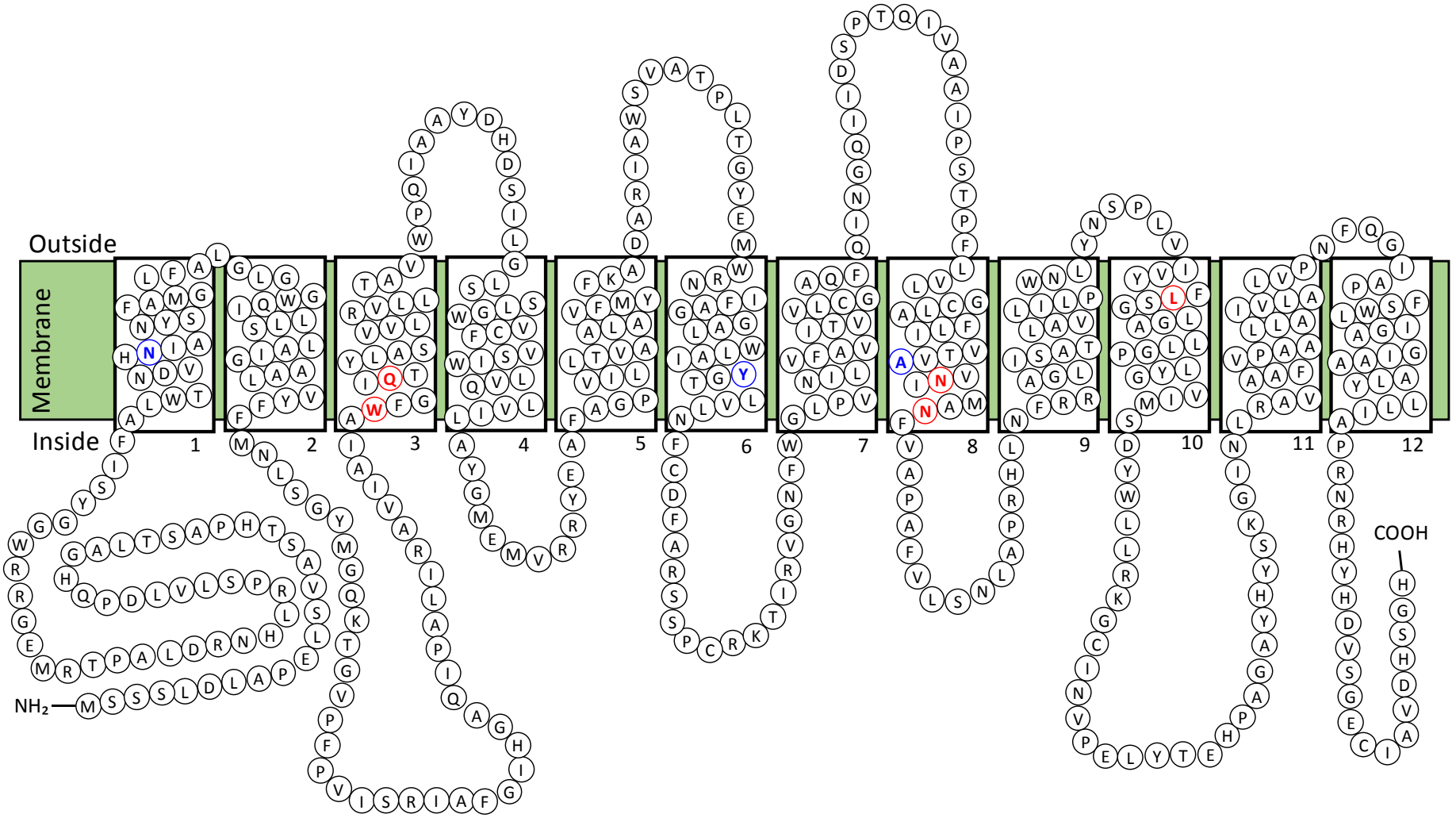



**Figure 5.5 Predicted topology of the *Pseudomonas putida* membrane transport protein AAN69889.** Diagram for the putative topology of the AAN69889 protein of *Pseudomonas putida* based on the TOPCONS prediction for transmembrane helices. Residues are coloured to show those that are identical (red) or highly similar (blue) compared with corresponding positions in the sodium and substrate binding site of Mhp1.

## 5.3 Optimisation of AAN69889 expression in *E. coli* strain BL21(DE3)

For performing functional and structural studies on membrane proteins it is necessary to obtain sufficient quantities of protein. Because the natural expression levels of membrane proteins are usually too low, amplified expression must be achieved (Bernaudat et al., 2011). *E. coli* is the organism most widely used to overexpress bacterial membrane transport proteins owing to the fact that it is easily accessible and easy to handle (Henderson et al., 2000; Ward et al., 2000; Saidijam et al., 2003; Drew et al., 2003; Drew et al., 2005). Despite the common use of *E. coli* for amplifying expression of membrane proteins, their overexpression can have toxic effects on the cells. This is most likely caused by exceeding the folding capacity of cytoplasmic and periplasmic chaperones (Wagner et al., 2006) or by exceeding the capacity of the membrane protein biogenesis/protein secretion machinery (Wagner et al., 2007; Wagner et al., 2008). The result could be expression of the protein in cytoplasmic inclusion bodies (Drew et al., 2003; Wagner et al., 2006; Wagner et al., 2007), from which isolation of correctly folded, functional membrane proteins is very difficult, though there are a few successful examples (Drew et al., 2003; Wagner et al., 2006; Wagner et al., 2007). Whilst proteases can be a problem for expressing membrane proteins in *E. coli*, as they can cause rapid protein degradation (Kihara et al., 1995), protease inhibitors can be added during membrane protein purification to minimise possible degradation (Weyand et al., 2008). In this research project, the plasmid pTTQ18/AAN69889 was transformed into *E. coli* BL21(DE3) host strain for protein expression (Chapter 4). Strain BL21(DE3) has been widely used for achieving high levels of membrane protein expression for a range of bacterial and archaeal transport proteins (Ward et al., 2000; Saidijam et al., 2003; Saidijam et al., 2005; Szakonyi et al., 2007; Ma et al., 2013; Bettaney et al., 2013), including the NCS1 proteins Mhp1 and PucI (Suzuki and Henderson, 2006; Weyand et al., 2008; Ma et al., 2016).

### 5.3.1 Optimisation of IPTG concentration for AAN69889 protein overexpression

Experiments were performed to evaluate the optimal IPTG concentration for amplified expression of protein AAN69889 in *E. coli* BL21(DE3) cells. All cultures induced with IPTG in the range 0.1 - 1.0 mM showed marginal decrease in growth rate after induction compared with cells that were left uninduced, indicating that overexpression was happening (Figure 5.6). Following an induction period of 2 hours, all induced cultures reached an $A_{680}$ of around 2.4. SDS-PAGE and Western blot analyses of mixed membranes prepared from these cultures indicated that an IPTG concentration of 0.25 mM was best for producing the highest level of AAN69889 protein expression (Figure 5.7). All further cell growth experiments were therefore conducted using 0.25 mM IPTG for induction. Note that in this and subsequent experiments, the cloned protein would still express in uninduced cells. This leakiness is a characteristic of expression from the operator-promoter in plasmid pTTQ18 (Stark, 1987).

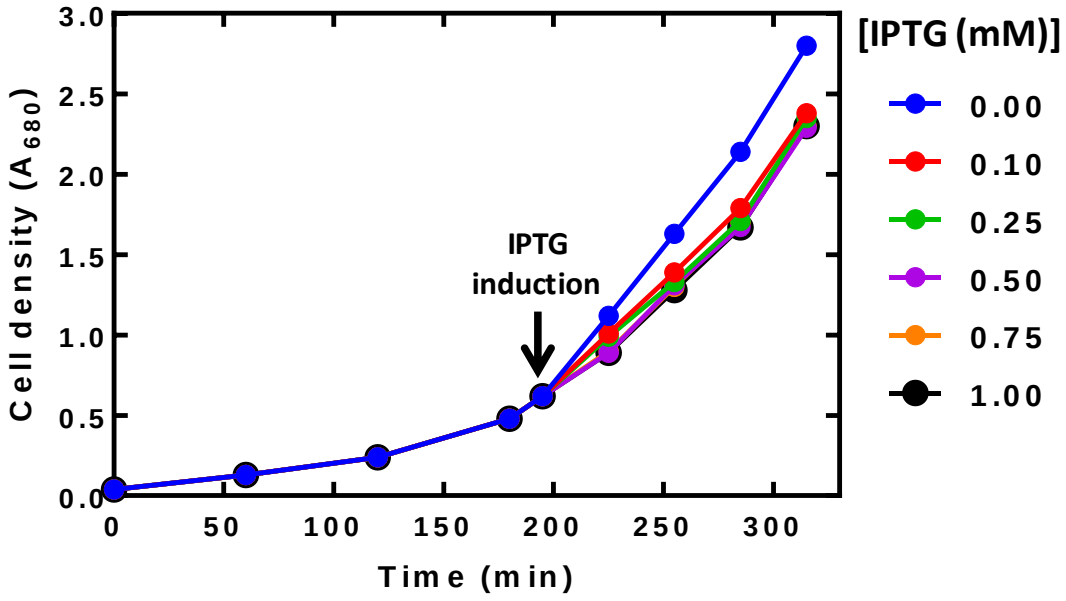



**Figure 5.6 Effect of various IPTG concentrations on the growth of *E. coli* BL21(DE3) cells harbouring plasmid pTTQ18/AAN69889.** Growth curves for *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 grown in 50 ml LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37℃ with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with the given concentration of IPTG and then grown for a further 2 hours.

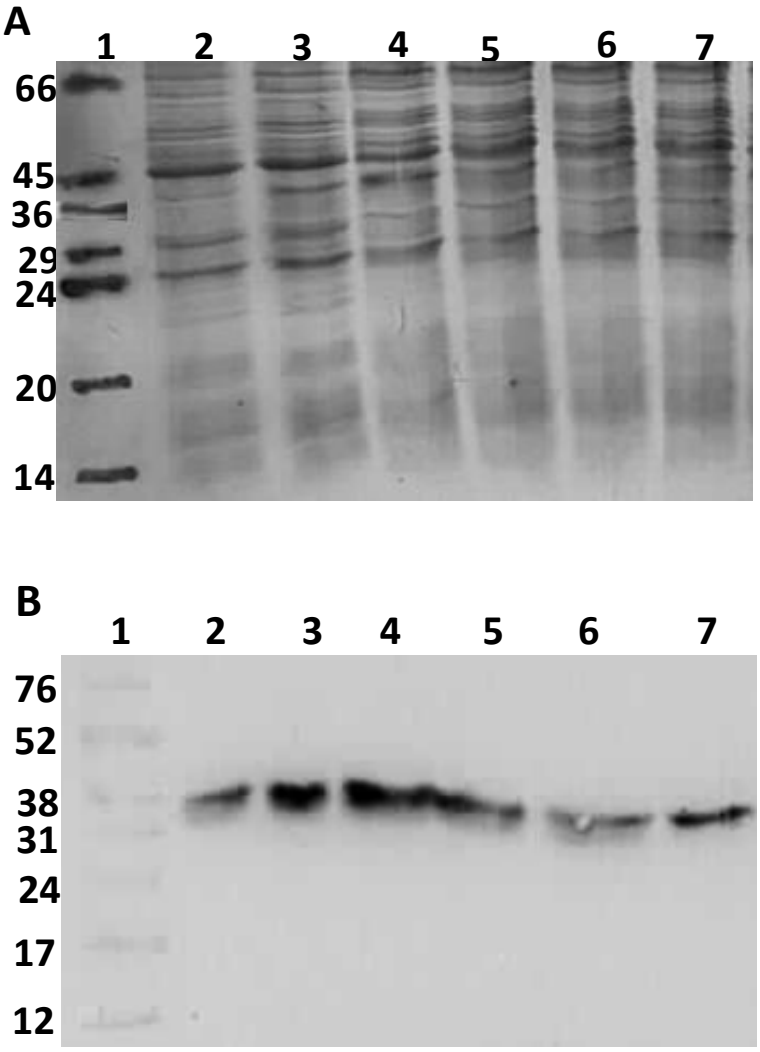



**Figure 5.7 Analysis of the effect of various IPTG concentrations on the overexpression level of the AAN69889 protein.** (A) SDS-PAGE and (B) Western blot. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) uninduced cells; (3) 0.1 mM IPTG; (4) 0.25 mM IPTG; (5) 0.5 mM IPTG; (6) 0.75 mM IPTG; (7) 1 mM IPTG. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

### 5.3.2 Optimisation of the period of induction for AAN69889 protein overexpression

Experiments were performed to evaluate the optimum length of post-induction time for amplified expression of protein AAN69889 in *E. coli* BL21(DE3) cells. All the cultures induced with 0.25 mM IPTG were grown for further periods in the range 2 to 5 hours showed gradual increase in the final $A_{680}$ from around 2.4 to 2.9 (Figure 5.8). SDS-PAGE and Western blot analyses of mixed membranes prepared from these cultures indicated that a post-induction period of 3 hours was optimum for producing the highest level of AAN69889 protein expression (Figure 5.9). All further large-scale cell growth experiments for producing membranes were therefore conducted using a post induction period of 3 hours.

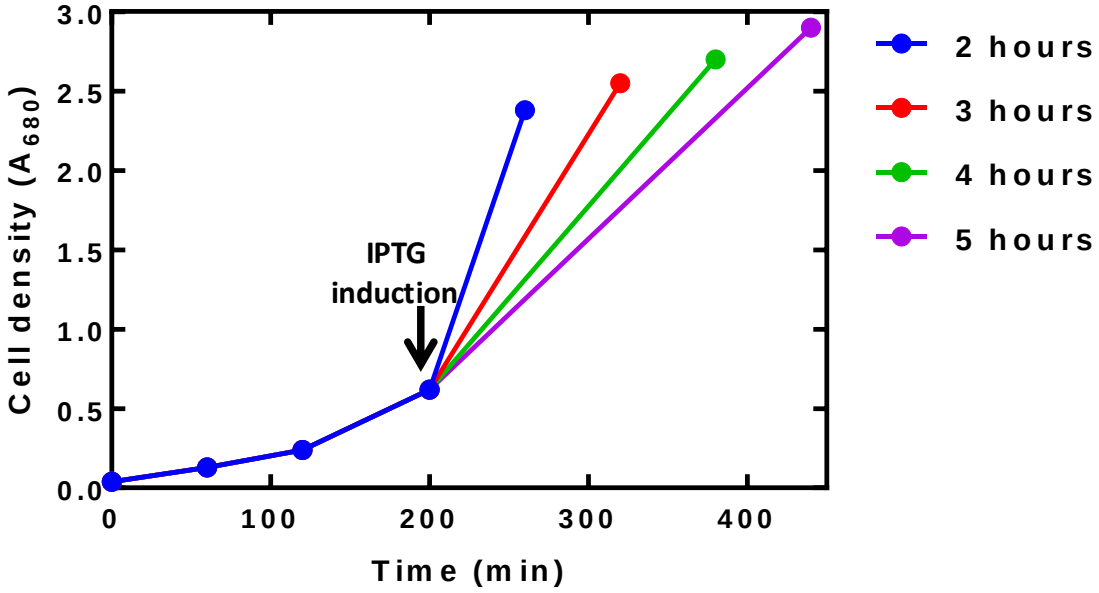



**Figure 5.8 Effect of post induction period on the growth of *E. coli* BL21 DE3) cells harbouring pTTQ18/AAN69889.** Growth curves for *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 grown in 50 ml LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for the given periods of time.

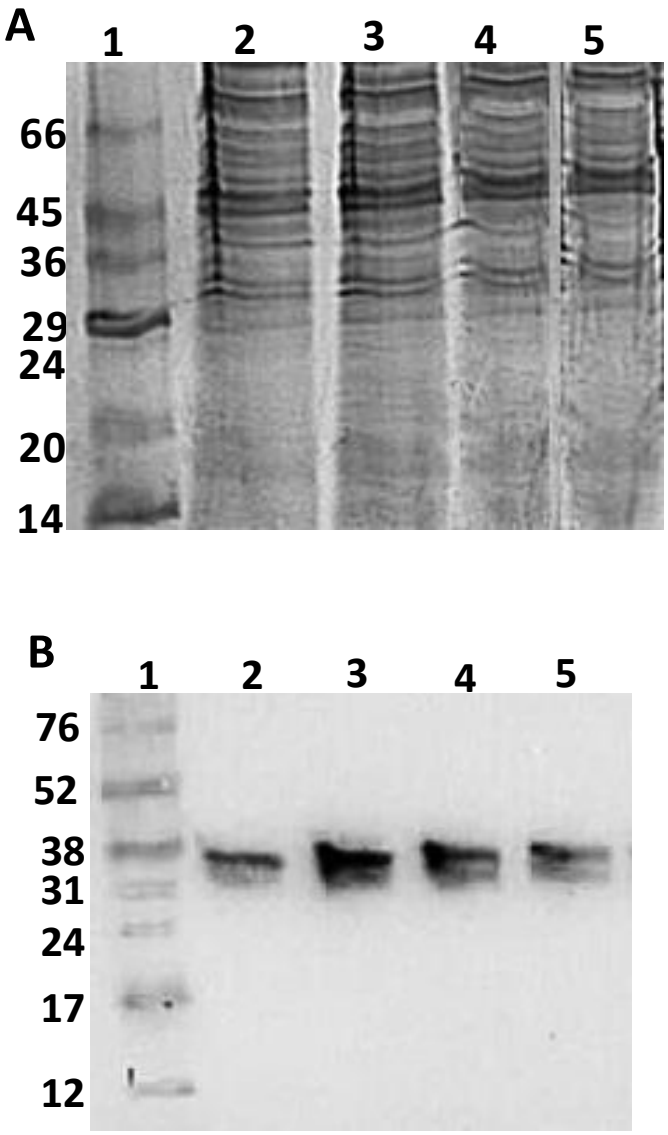



**Figure 5.9 Investigation of post induction period on the over expression level of AAN69889 protein.** (A) SDS-PAGE and (B) Western blot. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) 2 hours post induction; (3) 3 hours post induction; (4) 4 hours post induction; (5) 5 hours post induction. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

### 5.3.3 Effect of medium composition on the cell growth and on overexpression of the AAN69889 protein

Experiments were performed to evaluate the effect of medium composition on the cell growth and amplified expression of the protein AAN69889 in *E. coli* BL21(DE3) cells. The cultures using LB medium achieved the highest cell density of $A_{680}$ = ~3.2 (Figure 5.10). The cultures using 2TY medium reached a slightly lower cell density with an $A_{680}$ = ~2.9 and those using M9 minimal medium reached to a significantly lower $A_{680}$ = ~1.8 (Figure 5.11). In each respective medium, cell growth in induced cells was marginally lower than in uninduced cells, which is indicative of protein overexpression. SDS-PAGE and Western blot analysis of mixed membranes prepared from these cultures indicated that LB medium produced a significantly higher level of amplified expression than the 2TY and M9 minimal media (Figure 5.10), which is an important observation. All further large-scale cell growth experiments for producing membranes were therefore conducted using LB medium.

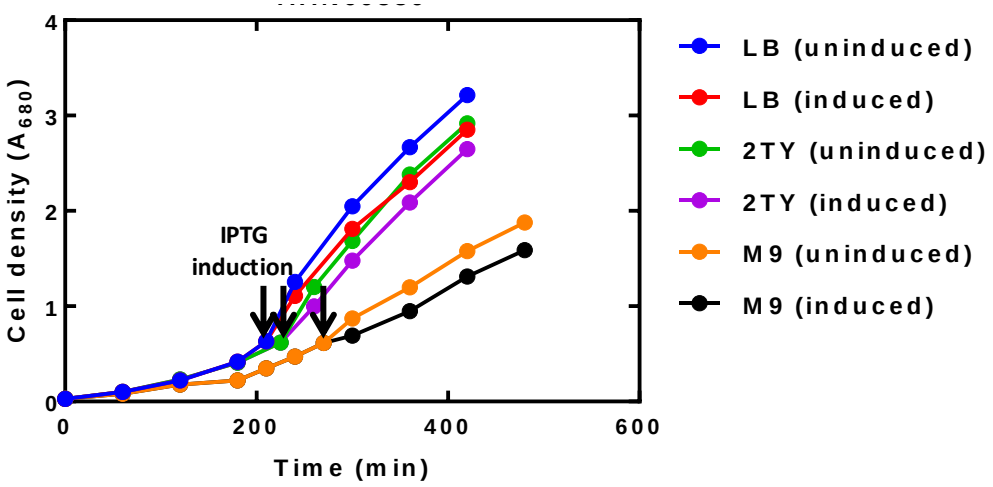



**Figure 5.10 Effect of medium composition on the growth of *E. coli* BL21(DE3) cells harbouring pTTQ18/AAN69889.** Growth curves for *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 grown in 50 ml volumes of LB medium supplemented with 20 mM glycerol, 2TY medium supplemented with 20 mM glycerol and M9 minimal medium supplemented with 20 mM glycerol. All cultures contained 100 µg/ml of carbenicillin and were grown at 37 °C with shaking at 220 rpm. Cells were left uninduced or were induced at $A_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for a further 3 hours.

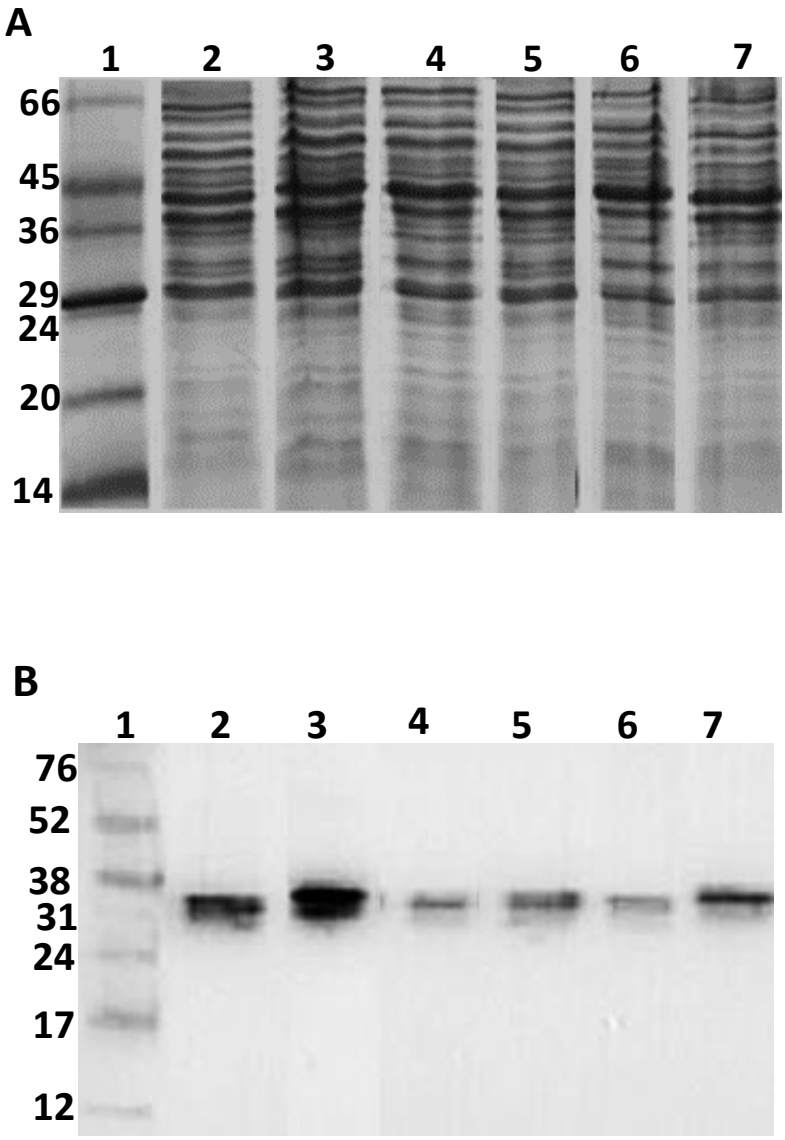



**Figure 5.11 Analysis of medium composition on the overexpression level of AAN69889 protein.** (A) SDS-PAGE and (B) Western blot. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) uninduced LB; (3) induced LB; (4) uninduced 2TY; (5) induced 2TY; (6) uninduced minimal; (7) induced minimal. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

## 5.4 Substrate specificity of AAN69889

Based on the substrates of characterised NCS1 family proteins and those putatively identified as substrates in databases, substrates of protein AAN69889 are likely to be nucleobases, nucleosides, hydantoins or related compounds (Figure 5.12).

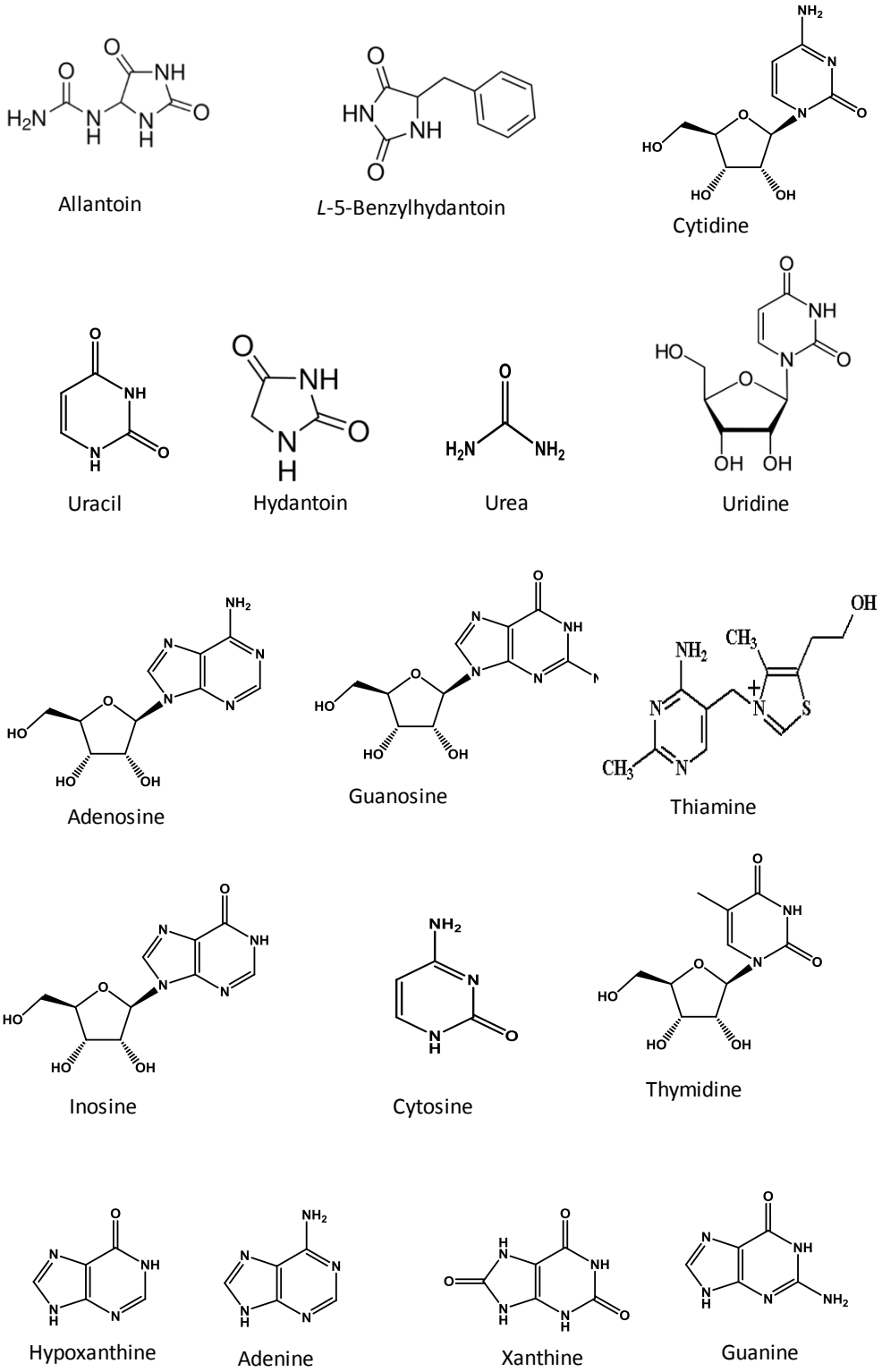



**Figure 5.12 Chemical structures of possible substrates/inhibitors for the AAN69889 protein.** Structures were drawn using ChemDraw software.

### 5.4.1 Evaluation of potential substrates for protein AAN69889

Transport experiments in energised whole *E. coli* cells were performed to test a number of radiolabelled compounds at 50µM final concentration as potential substrates of the AAN69889 protein. The compounds tested were; $^{14}$C-allantoin, $^{3}$H-benzylhydantoin, $^{14}$C-indolylmethylhydantoin, $^{14}$C-uracil, $^{3}$H-cytosine and $^{3}$H-thiamine (Figure 5.13). Despite the *in silico* prediction for AAN69889 to be a nucleoside transporter, only allantoin out of the tested compounds showed significant uptake into the cells. Uptake of $^{14}$C-allantoin was significantly greater in induced cells compared with uninduced cells with values after 2 minutes of around 19.5 and 3.6 nmol mg$^{-1}$ cells, respectively. These initial transport measurements suggested that the main substrate of AAN69889 is allantoin.

### 5.4.2 Uptake of allantoin but not cytosine, uridine, thiamine, or hydantoins by AAN69889 into energised whole cells

Further transport experiments were then performed to test if allantoin is the main substrate of AAN69889. *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 and the empty plasmid pTTQ18 were tested for transport of $^{14}$C-allantoin. Uptake of $^{14}$C-allantoin was significantly greater in induced cells compared with uninduced cells containing pTTQ18/AAN69889 (Figure 5.14). As a control, induced and uninduced cells containing plasmid pTTQ18 with no gene insert were tested for the ability to transport $^{14}$C-allantoin (Figure 5.14). The results indicate negligible uptake in these cells, which also confirms that uptake into uninduced cells containing pTTQ18/AAN69889 is due to leaky expression. These results confirmed that allantoin is the main substrate of AAN69889, which could have been predicted based from its closest evolutionary relationship with PucI. The results also demonstrate how important it is to perform experimental assessments of substrate specificity when characterising transport protein function and not to rely on database predictions.

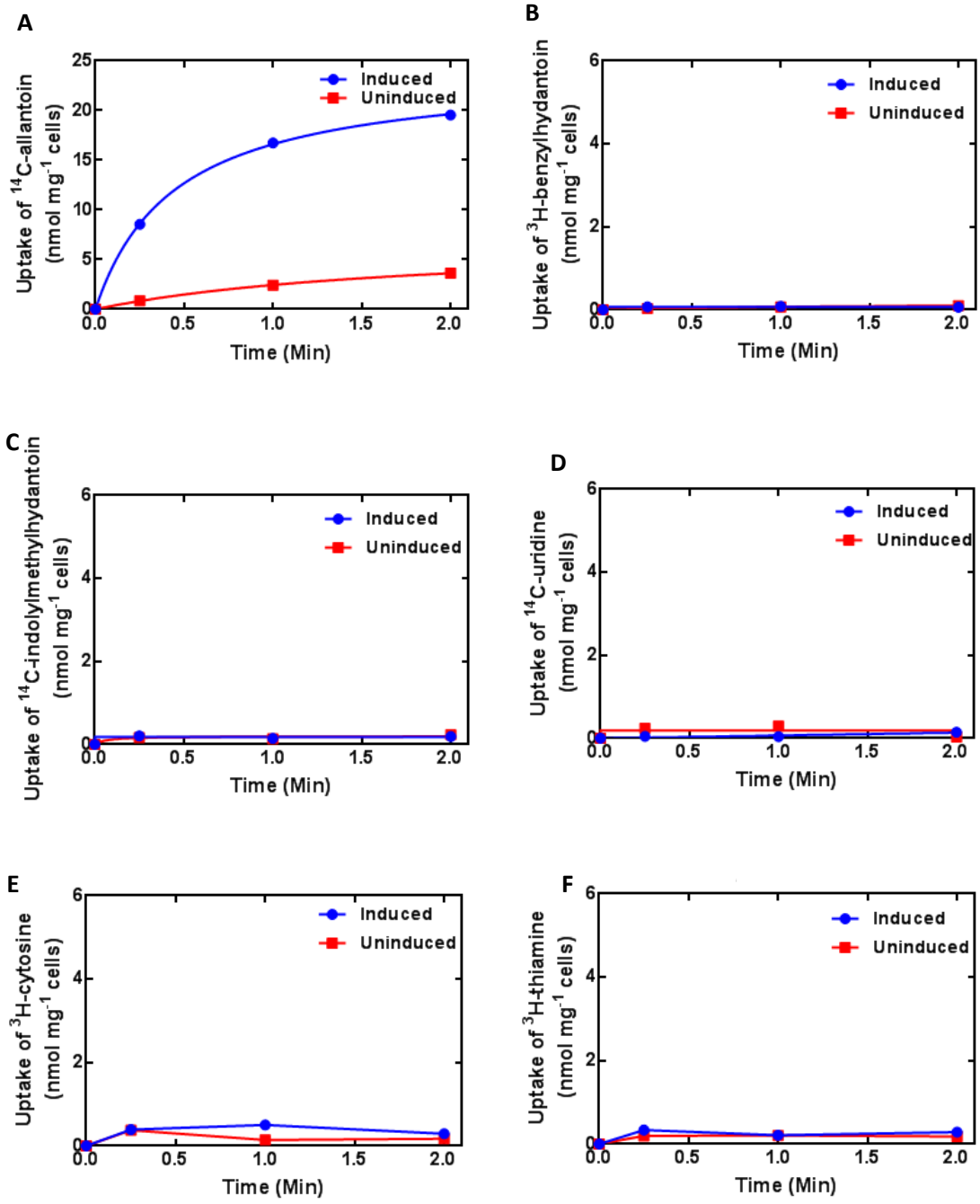



**Figure 5.13 Possible transport of radiolabelled substrates by AAN69889.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were left uninduced or were induced at $A_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer (150 mM KCl, 5 mM MES pH 6.6) and resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Cells were energised with 20mM glycerol and tested for uptake of 50 µM of different potential radiolabelled substrates.

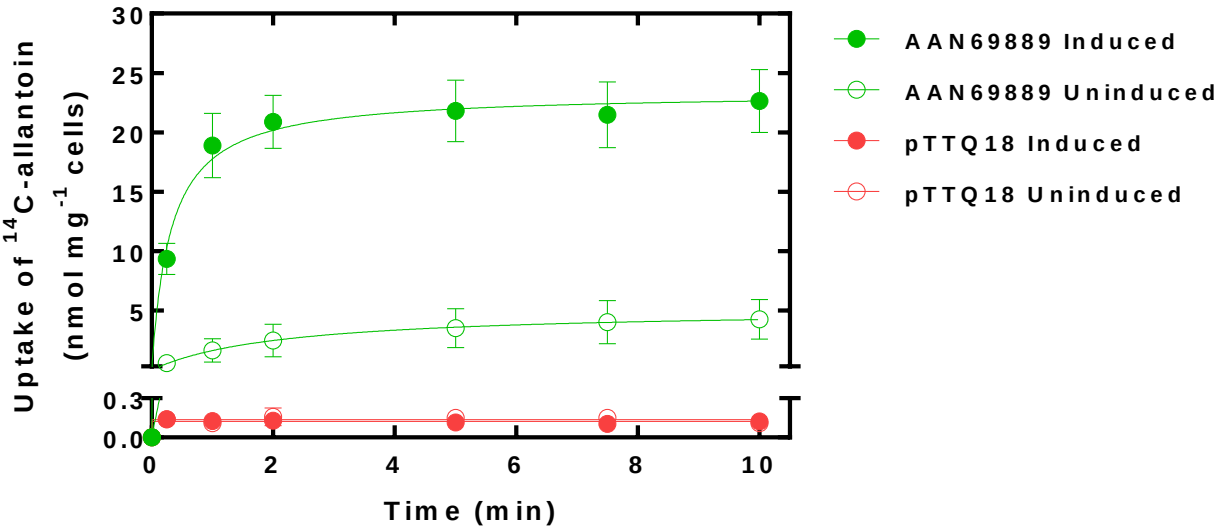



**Figure 5.14 Confirmation of $^{14}$C-allantoin transport byAAN69889.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 (green) or the empty plasmid pTTQ18/no gene (red) were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37°C with shaking at 220 rpm. Cells were left uninduced (open circles) or were induced (closed circles) at $A_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer (150 mM KCl, 5 mM MES pH 6.6) and resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Cells were energised with 20mM glycerol and tested for uptake of 50 µM $^{14}$C-allantoin at the given time points. The data represent mean of duplicate measurements.

### 5.4.3 Assessing the effect of sodium ions on the uptake of $^{14}$C-allantoin by AAN69889

In order to assess the effect of sodium ions on the uptake of $^{14}$C-allantoin by AAN69889, transport of $^{14}$C-allantoin was measured in the presence of various concentrations of NaCl in the range 0-150 mM. In each case, the overall salt concentration was kept constant at 150 mM by using the appropriate concentration of KCl in the transport buffer. Under conditions of zero NaCl, crown ether at a concentration of 10 mM was included to mop up any traces of NaCl. At time points of both 15 seconds and 2 minutes, no significant difference in uptake of $^{14}$C-allantoin was observed over the entire range of NaCl concentrations (Figure 5.15). These

results demonstrate that $^{14}$C-allantoin transport by AAN69889 is unlikely to be dependent on sodium.

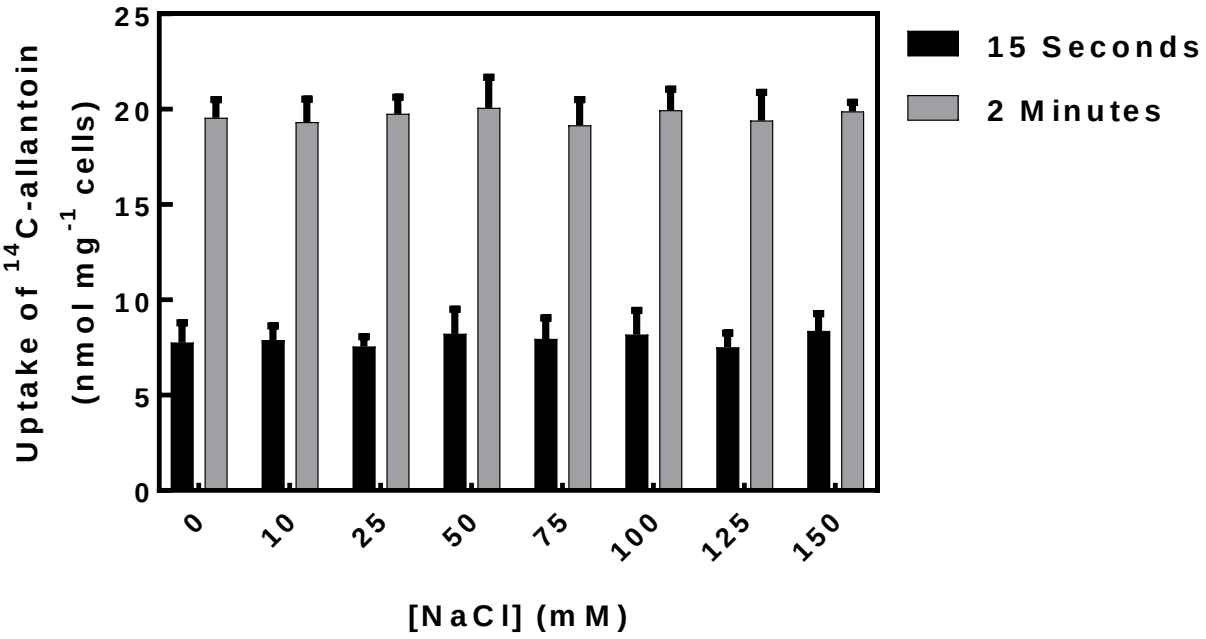



**Figure 5.15 Effect of sodium ion concentration on uptake of $^{14}$C-allantoin by cells expressing AAN69889.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer containing 5 mM MES (pH 6.6) and a range of NaCl concentrations from 0-150 mM, balanced by a range of KCl concentrations to maintain an overall salt concentration of 150 mM. Cells were resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Aliquots of cells were energised with 20 mM glycerol and tested for uptake of 50 µM $^{14}$C-allantoin at the given time points. The data represent mean of duplicate measurements.

### 5.4.4 Ligand specificity of AAN69889 determined by competition assay

To investigate ligand specificity of AAN69889, $^{14}$C-allantoin uptake in *E. coli* cells expressing AAN69889 was measured in the presence of a ten-fold molar excess of potential unlabelled competing compounds (Figure 5.16). The greatest competitive effect was produced by allantoin, which reduced uptake to 11.2% and 8.4% at time points of 15 seconds and 2 minutes, respectively. The next most effective competitors were uracil and hydantoin, which reduced uptake to 66.6/51.9% and 77.0/64.6% after 15 seconds and 2 minutes, respectively. It should be pointed out that addition of DMSO alone reduced the uptake to 76.3% and 85.1% at 15 seconds and 2 minutes, respectively. All other compounds had uptake values greater than 80% compared with non-competed cells. Overall the results demonstrate high specificity for recognition of allantoin by AAN69889, which is more specific than ligand recognition by PucI (Ma, 2010; Ma et al., 2016).

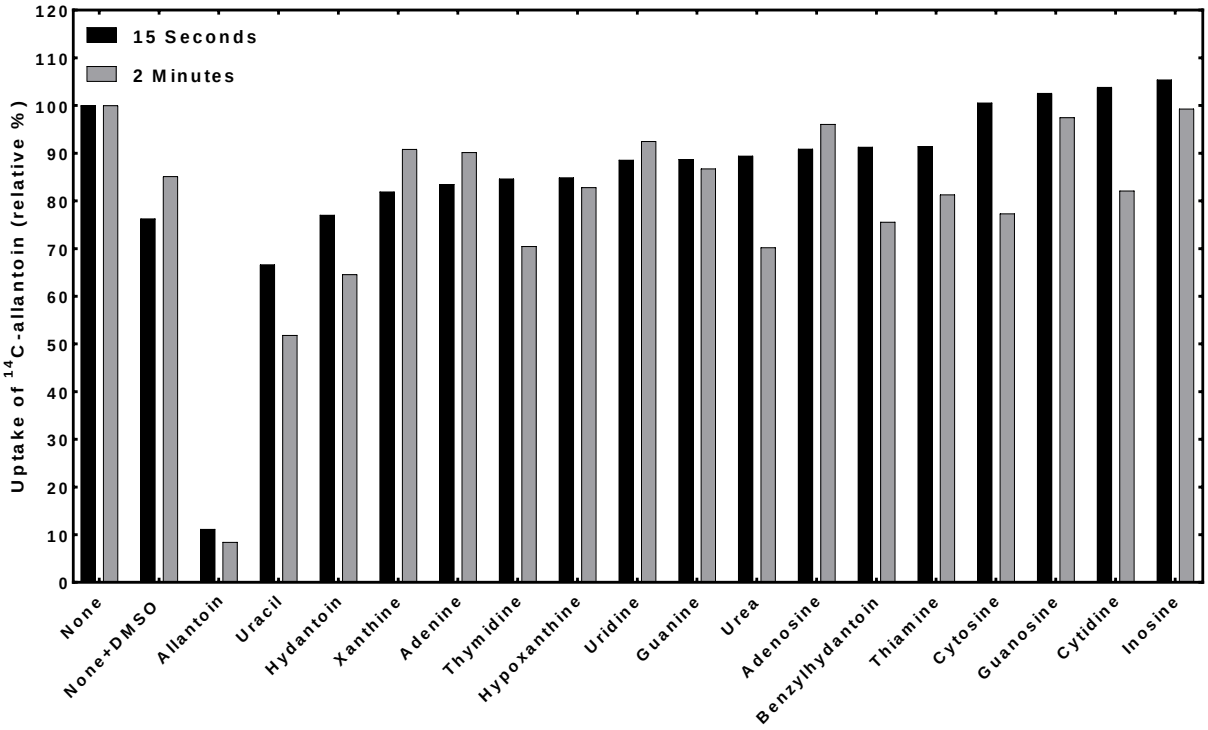



**Figure 5.16 Ligand specificity of AAN69889.** Competition of $^{14}$C-allantoin (50 µM) uptake into *E. coli* BL21(DE3) cells expressing AAN69889 in the presence of a ten-fold molar excess of potential competitors. The non-competed uptake rate was taken as 100% corresponding to 15 seconds and 2 minutes post-addition of $^{14}$C-allantoin, respectively. All data represent the average of duplicate measurements. None = no competitors.

### 5.4.5 Concentration-dependence of $^{14}$C-allantoin uptake by AAN69889

An experiment was performed to obtain kinetic parameters for uptake of $^{14}$C-allantoin by AAN69889 into energised whole cells. It was observed that uptake of $^{14}$C-allantoin into energised *E. coli* BL21(DE3) cells expressing AAN69889 was concentration-dependent and conformed to a rectangular hyperbola over the concentration range 0-500 µM (Figure 5.17). The data were fitted to the Michaelis-Menten equation to produce values for the apparent affinity of initial rate uptake ($K_m$) and maximum velocity ($V_{max}$) of 39.17 ± 2.66 µM and 14.77 ± 0.27 nmol/mg cells/15 sec, respectively. Uptake into uninduced cells also showed concentration-dependence, but this was negligible, producing $K_m$ and $V_{max}$ values of 20.56 ± 4.32 µM and 0.51 ± 0.02 nmol/mg cells/15 sec, respectively. Kinetic parameters for $^{14}$C-allantoin uptake by AAN69889 are similar to those obtained for PucI, which produced $K_m$ and $V_{max}$ values of 24.4 ± 3 µM and 14.8 nmol/mg cells/15 sec, respectively (Ma et al., 2016).

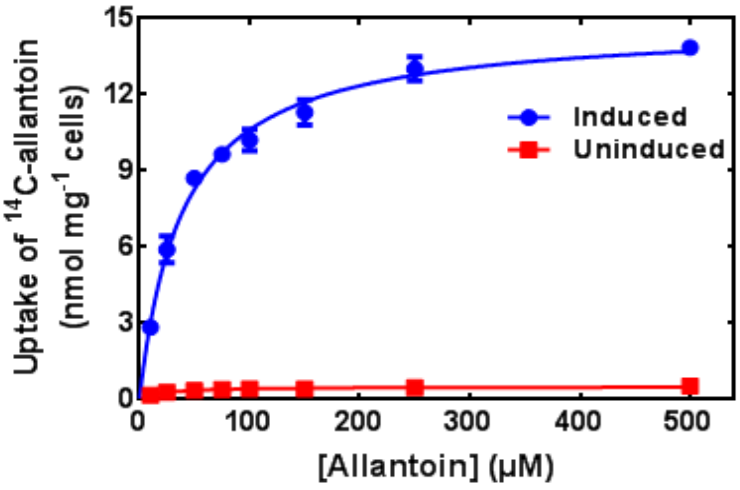



**Figure 5.17 Concentration-dependence of $^{14}$C-allantoin uptake by AAN69889.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/AAN69889 were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were left uninduced or were induced at an $OD_{680}$ = 0.6 with IPTG (0.25 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer (150 mM KCl, 5 mM MES pH 6.6) and resuspended in the same buffer to an accurate $OD_{680}$ of around 2.0. Cells were energised with 20 mM glycerol and tested for uptake of a range of concentrations of $^{14}$C-allantoin (0-500 µM) 15 seconds post addition.

## 5.5 Purification and yield of AAN69889

Inner membranes containing amplified expression of AAN69889 were solubilised with 1% DDM and the $His_6$-tagged protein was purified by immobilised metal affinity chromatography using a Ni-NTA resin (Section 2.4.3). The NTA ligand contains four chelating sites that are interacting with metal ions. NTA usually occupies four of the ligand binding sites in the co-ordination sphere of the $Ni^{+2}$ ion, leaving two sites to interact with the hexahistidine tag (Qiagen, 2003). Fractions from the purification of AAN69889 were analysed by SDS-PAGE and Western blotting (Figure 5.18). Lane 2 contains the inner membranes solubilised with 1% DDM and lane 3 contains the supernatant obtained by ultracentrifugation of the solubilised inner membranes. The relative amount of AAN69889 in the supernatant is similar to the level in inner membranes, indicating that the large majority of membrane proteins have been solubilised by the DDM. Lane 4 shows proteins in the pellet obtained following solubilisation and ultracentrifugation (insoluble fraction), which does contain some AAN69889 protein. Due to the small volume of this pellet (resuspended in 500-700 µl) compared with the volume of the supernatant (60 ml), the total amount of AAN69889 in the pellet is negligible. Lane 5 shows protein that did not bind to the Ni-NTA column (unbound fraction) and the blot suggests that most of the AAN69889 protein has been bound to the resin. Following removal of further unbound material by washing with buffer containing a low concentration of imidazole (20 mM), purified protein was eluted by washing with buffer containing a high concentration of imidazole (200 mM); this dissociates the $His_6$-tagged proteins because they can no longer compete for binding sites on the Ni-NTA resin (Lane 6). Two associated bands in the purified protein are visible which might be due to unfolding property of protein. Application of Mass spectrometry would be best to analyse this. The purity of AAN69889 protein was 80% determined by densitometry. The yield of purified protein was estimated from a typical 30-litre fermenter culture of *E. coli* BL21(DE3) cells harbouring plasmid pTTQ18/AAN69889 in LB medium supplemented with 20 mM glycerol, which produced 252.9 g of cell pellet. Membranes prepared from these cells equated to a total volume of 43.6 ml of inner membranes. Based on the purification shown in Figure 5.18, the yield of purified protein was 1.2 mg per litre of cell culture.

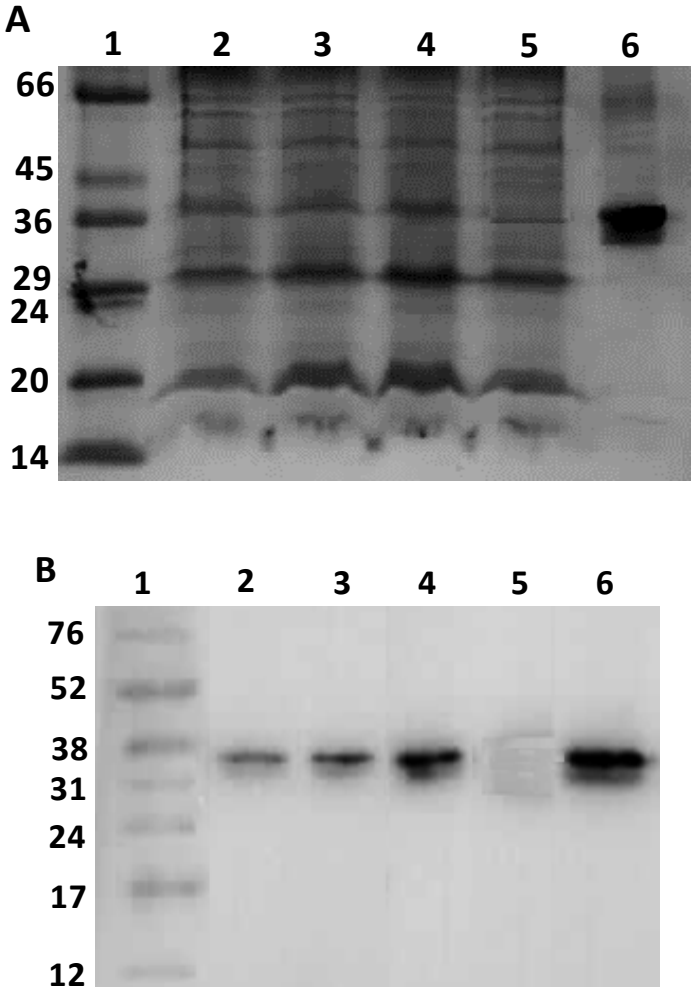



**Figure 5.18 Purification of AAN69889 from inner membranes**. SDS-PAGE (A) and Western blot (B) analysis for purification of AAN69889. Inner membranes (section 2.4.2) were solubilised in 1% DDM. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) inner membranes; (3) supernatant (soluble fraction); (4) membrane pellet (insoluble fraction); (5) column flow-through (unbound fraction); (6) purified protein. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

### 5.5.1 Estimation of molecular weight for AAN69889

Using the polypeptide sequence of AAN69889($His_6$) and the Expasy $M_w$ calculator (http://web.expasy.org/compute_pi/), the theoretical mass of AAN69889($His_6$) was 56693.72Da. The size of the AAN69889 protein according to SDS-PAGE was determined by comparison with distances migrated by molecular weight marker from

a log molecular weight-distance curve (Section 2.4.5). By SDS-PAGE AAN69889($His_6$) representing two bands that travel to a distance of 21mm and 19mm, which corresponds to a $Log_{10}M_w$ of 1.57 and 1.58, respectively. Therefore a molecular weight of 37 kDa and 38 kDa determined (Table 5.1; Figure 5.19). The predicted size of AAN69889($His_6$) protein is 56.6kDa and it migrates on SDS-PAGE at a position of 37 kDa and 38.0 kDa, which is a size difference of ~68%. This is consistent with SDS-PAGE analysis of other NCS1 family proteins, which migrate at approximately 62-74% of their predicted sizes (Bettaney, 2008; Ma, 2010; Ma et al., 2016). It is widely recognised that membrane proteins migrate anomalously on SDS-PAGE gels at lower molecular weight positions than their actual molecular weights predicted from amino acid composition due to their hydrophobic nature, high binding of SDS or the retention of secondary structure facilitating the migration through the gel (Ward et al., 2000; Rath and Deber, 2013).

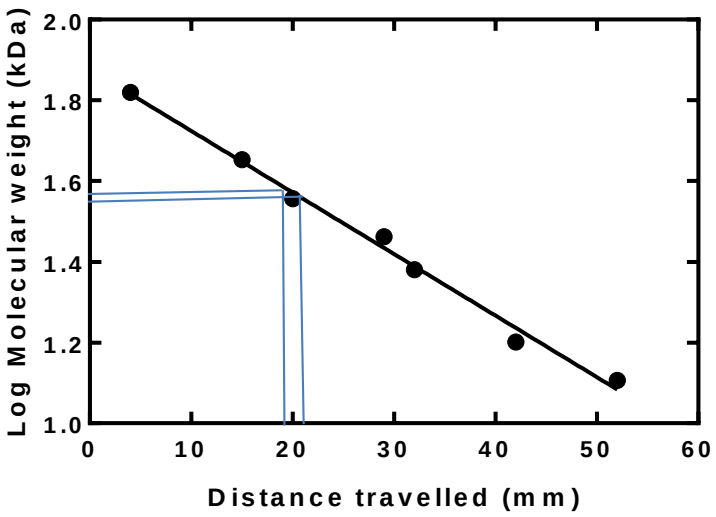



**Figure 5.19 Estimation of AAN69889 molecular weight on SDS-PAGE gel.** Calibration graphs for the distance travelled by proteins on SDS-PAGE gel (Figure 5.18) based on migration of marker proteins of known molecular weight. The blue lines represent the distance travelled by AAN69889, which equates to a molecular weight of 37 and 38 kDa.

**Table 5.1 Distances travelled by molecular weight marker and AAN69889 protein on SDS-PAGE gel.**

| Lane Contents | Distance travelled (mm) | MW (kDa) | Log10 (kDa) |
|---|---|---|---|
| **(1) Mw Marker** | 4 | 66 | 1.81 |
| | 15 | 45 | 1.65 |
| | 20 | 36 | 1.55 |
| | 29 | 29 | 1.46 |
| | 32 | 24 | 1.38 |
| | 42 | 20 | 1.30 |
| | 52 | 14 | 1.14 |
| **(2) Purified protein (first band)** | 21.0 | 37 | 1.57 |
| **(3) Purified protein (second band)** | 19.0 | 38 | 1.58 |

## 5.6 Secondary structure integrity and thermal stability of AAN69889

Circular dichroism (CD) spectroscopy (Section 2.5.1) is a valuable biophysical technique for examining the folding, conformational changes and determination of protein secondary structure. This can also be used to monitor the membrane proteins thermal unfolding property (Wallace et al., 2003; Miles and Wallace. 2016). Far-UV CD spectroscopy (180-260 nm) was used to assess the secondary structure content and thermal stability of the purified AAN69889 protein solubilised in DDM detergent (Figure 5.20A). The resultant spectra were characteristic of predominantly alpha helical proteins with a peak at ~192 nm and troughs at 209 nm and 222 nm (Wallace et al., 2003; Kelly et al., 2005; Bulheller et al., 2007; Bettaney et al., 2013). These measurements also confirmed that the purified AAN69889 protein was correctly folded and had retained its alpha helical secondary structure after passing

through various steps of the purification. This was an important control prior to further biophysical analyses using the purified protein.

The thermal stability of AAN69889 was analysed by ramping the temperature from 5-90 ºC and finally back to 5 ºC (Figure 5.20 B and C). The final scan at 5 ºC shows whether the purified protein had refolded into its native secondary structure. AAN69889 loses it secondary structure with rising temperature and comparisons of spectra at 5 ºC prior to thermal denaturation and at 90 ºC following denaturation showed three main differences (Figure 5.20C). Firstly, a decrease in the negative signals at both 209 nm and 222 nm and secondly significant reductions in the amplitude of the signal at 192 nm. These changes were indicative of reductions in the alpha helical content and therefore of thermal denaturation. At 90 ºC there was still a positive signal at ~192 nm and negative signals at ~209 nm and ~222 nm suggesting that the protein was not fully unfolded. With increasing temperature the signal at 209 nm began to change significantly at ~40 ºC and a melting temperature of 46.2 ºC was estimated using Global Analysis CD software 3. The AAN69889 protein is therefore reasonably stable for performing biophysical assays using a temperature range of 18-25 ºC. On returning the temperature to 5 ºC, there was no evidence that suggests refolding of the protein, similar to reported results for other transport proteins (Bettaney, 2008; Ma, 2010; Sukumar, 2012; Jackson, 2012).

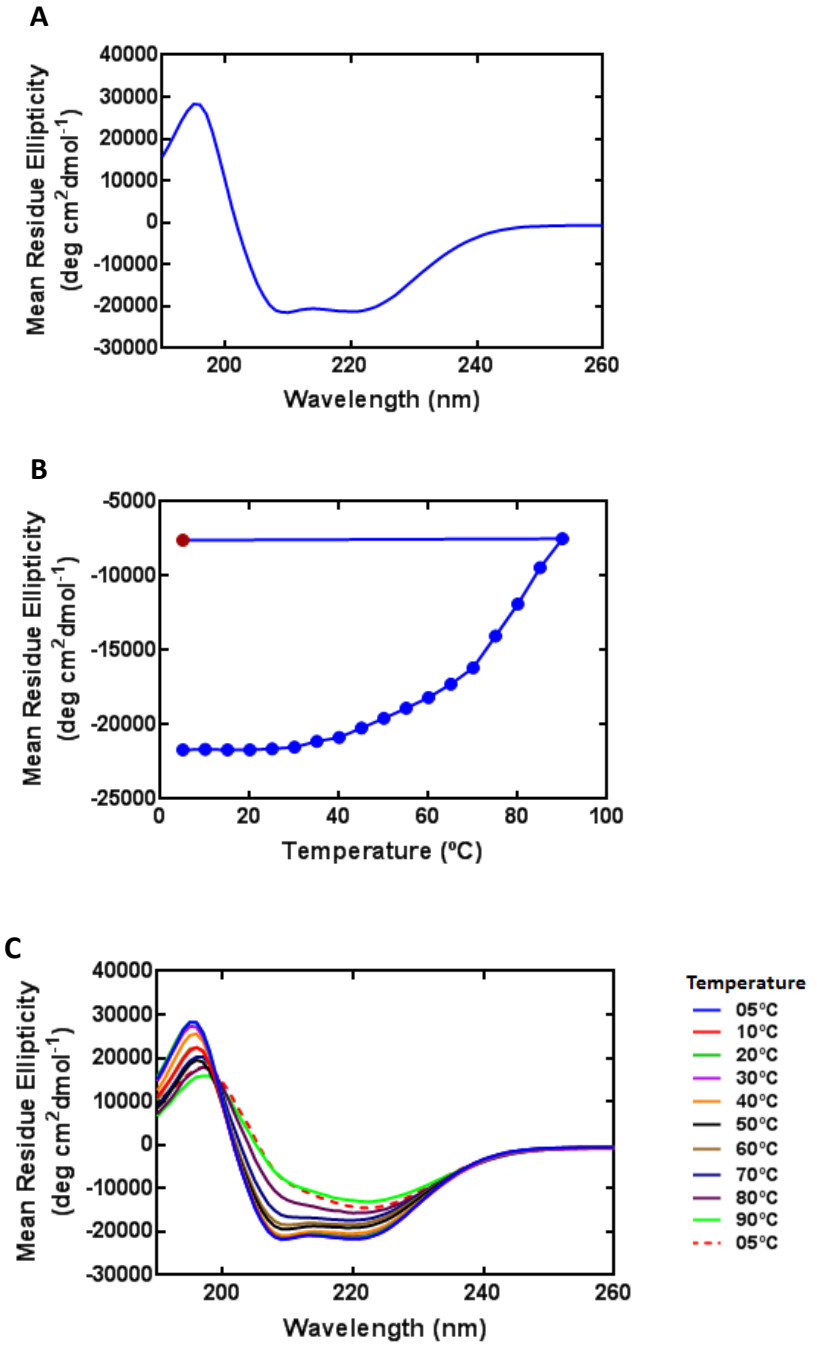



**Figure 5.20 Far-UV CD analysis of purified AAN69889 protein.** (A) Far-UV (180-260 nm) CD spectrum for purified AAN69889 protein. Measurements were performed using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant nitrogen flushing. Samples were prepared in a Hellma quartz cuvette of 1.0 mm pathlength at a final protein concentration of 0.15 mg/ml in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM). (B) Thermal unfolding of AAN69889 protein over the concentration range 5-90 ºC and finally back to 5 ºC monitored at a wavelength of 209 nm. (C) Full spectra showing thermal unfolding of AAN69889 protein over the given temperature range.

## 5.7 Possible fluorimetric analysis of ligand binding to AAN69889

Fluorescence spectroscopy is a convenient method of assessing ligand binding by exploiting changes in the intrinsic protein fluorescence. Such fluorescence principally originates from tryptophan residues and is especially useful when tryptophan residues are proposed to be in or close to the ligand binding site. Fluorimetric detection of allantoin binding to AAN69889 and its dependency on NaCl was tested by titrating allantoin in the range of 0-2 mM against the purified protein (2.5 µM) (Figure 5.21). The magnitude of fluorescence quench at 332 nm was only up to around 3.2% at an allantoin concentration of 2 mM, which is relatively small compared with the fluorescence quench of around 20% obtained for L-BH binding to Mhp1 at the same concentration (Weyand et al., 2008; Jackson, 2012; Simmons et al., 2014; Figure 3.7). About half of this quenching effect was due to addition of the DMSO that was used to solubilise the allantoin. Nevertheless, there was no significant difference in the fluorescence quench obtained with allantoin in the absence and presence of 15 mM NaCl. Fluorescence of tryptophan residues is highly sensitive to the environment (Chen and Barkley, 1998; Ghisaidoobe and Chung, 2014) and it is possible that the binding site in the AAN69889 does not involve tryptophan. Mhp1 has two tryptophan residues (Trp117 and Trp220) in its substrate binding site whereas AAN69889 has tyrosine residue at this position make it less sensitive for this type of assay. The fluorescence results with AAN69889 do not mean that there is no binding of allantoin to the protein; indeed the uptake of radiolabelled allantoin by AAN69889 was already established. It seems that binding does not have a significant effect on the fluorescence of tryptophan residues at least at concentration up to 2mM, considered to be a physiological range of concentration 50x Km value (Figure 5.17).

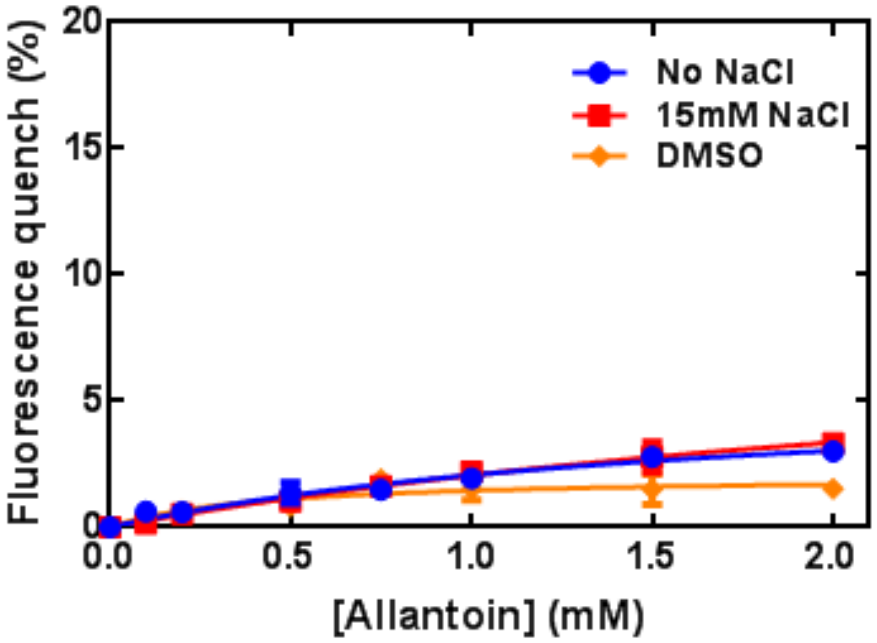



**Figure 5.21 Possible detection of binding of allantoin to AAN69889 by quenching of intrinsic fluorescence.** Steady-state spectrophotofluorimetry measurements were performed for detecting allantoin binding to purified AAN69889 (Section 2.5.2). Samples containing protein (2.5 µM) in fluorescence buffer (Table 2.21) at 18 ºC were excited at 295 nm and fluorescence emission was measured at 332 nm. Protein was pre-equilibrated with stirring for 1.5 minutes and then following additions of allantoin or DMSO, samples were stirred for 0.5 minutes before making the measurements. Data represent triplicate measurements and were fitted to the Michaelis-Menten equation in GraphPad Prism 7 software.

## 5.8 Introduction to *Vibrio parahaemolyticus*

*Vibrio parahaemolyticus* is a gram-negative marine bacterium and is a worldwide leading cause of food-borne gastroenteritis in humans, especially in the areas with high consumption of raw, or undercooked seafoods (Letchumanan et al., 2014; Letchumanan et al., 2015). The organism naturally found in estuarine, marine and coastal environments and possess enormous adaptive capabilities: a planktonic cell or attached to submerged inert surfaces, like bottoms of boats or to other ocean surfaces like fish, shellfish and zooplankton (McCarter, 1999; Iwamoto et al., 2010; Nelapati et al., 2012). *V. parahaemolyticus* either exist as a swimmer cell or a swarmer cell, adopted for locomotion in various environments. (Sar et al., 1990; Makino et al., 2003). Under appropriate circumstances the organism has an exceptionally short generation time of 8-12 minutes (Ulitzur, 1974; Makino et al., 2003). A haemolysin is thought to be an important virulence factor produced by the bacterium but the mechanisms of pathogenesis are still unclear (Hiyoshi et al., 2010; West et al., 2013; Lee et al., 2015; Kumaran and Citarasu, 2016).

VP1242 is the only predicted NCS1 family transporter in *V. parahaemolyticus*. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/) (Kanehisa et al., 2017), the gene that codes for *V. parahaemolyticus* protein VPA1242 is predicted to be a cytosine permease and is immediately upstream from the gene that codes for protein VPA1243, which is predicted to be a cytosine deaminase (Figure 5.22). The UniProt KnowledgeBase (http://www.uniprot.org/) entry for this protein (Q87GS3) also suggests that it is a putative cytosine permease, also referred to as CodB. Indeed, VPA1242 shares 75.2% sequence identity with *E. coli* cytosine permease CodB (Table 4.1) and the *E. coli* cytosine-inducible operon *codBA* encodes CodB followed by a cytosine deaminase (Danielsen et al., 1992; Qi and Turnbough, 1995). *E. coli* can utilise cytosine as its sole source of nitrogen by using CodB to mediate the uptake of exogenous cytosine and CodA to catalyse hydrolytic deamination of cytosine to uracil (a source of pyrimidines) and ammonia (a source of nitrogen) (Danielsen et al., 1992). VPA1242 contains 412 amino acids that putatively form eleven or twelve transmembrane spanning α-helices according to membrane topology predictions (Figure 5.23). Based on structurally characterised NCS1 family proteins and a demonstrated reliability for the TOPCONS prediction tool (Hennerdal and Elofsson,

2011), twelve transmembrane helices is most likely to be correct. In this case, both the N- and C-terminal ends of the protein are predicted to be at the cytoplasmic side of the membrane. This is the same as the membrane topology predicted for *E. coli* CodB (Danielsen et al., 1995).

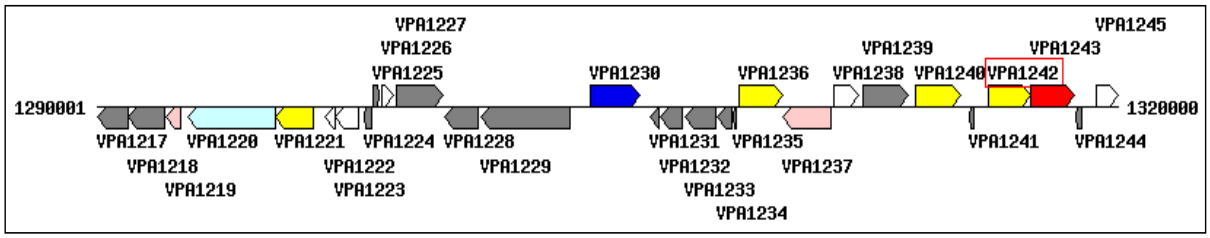



**Figure 5.22 Map of the *Vibrio parahaemolyticus* VPA1242 region.** This diagram was taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg/) (Kanehisa et al., 2017).

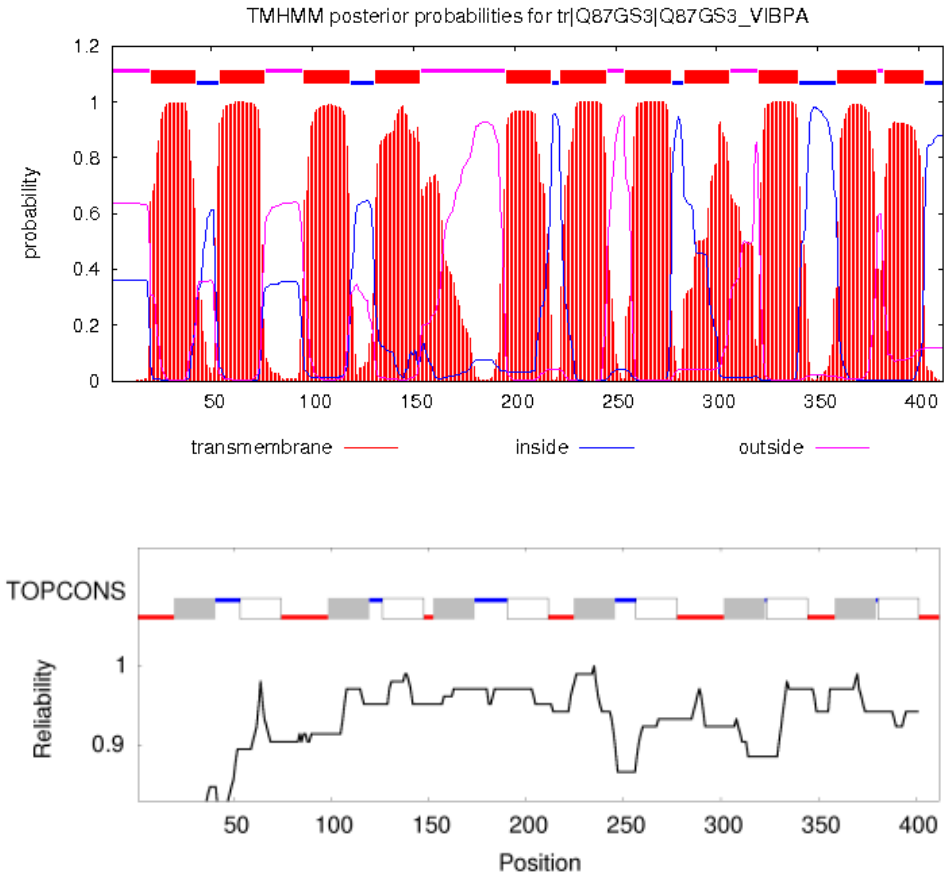



**Figure 5.23 Predictions of transmembrane helices in VPA1242.** The amino acid sequence of VPA1242 was analysed by the membrane topology prediction tools TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001) (top) and TOPCONS (http://topcons.cbr.su.se/) (Bernsel et al., 2009) (bottom).

## 5.9 Conservation of residues between VPA1242 and characterised bacterial NCS1 family proteins

Based on separate sequence alignments with characterised bacterial NCS1 family proteins, VPA1242 from *V. parahaemolyticus* shares overall homologies of 47.3% (22.8% identical, 24.5% highly similar) with Mhp1 from *Microbacterium liquefaciens*, 45.1% (24.8% identical, 20.3% highly similar) with PucI from *Bacillus subtilis* and 89.5% (75.2% identical, 14.3% highly similar) with CodB from *E. coli*. This confirms that VPA1242 is most closely related with the *E. coli* cytosine transporter CodB, as demonstrated by the phylogenetic analysis shown in Chapter 4 Figure 4.2. An alignment between VPA1242, Mhp1 and CodB identifies the residues that are conserved between these three proteins (Figure 5.24). Twelve residues in the sodium and substrate binding site of Mhp1 based on crystal structures are also highlighted (Figure 5.25). Only four of these positions, corresponding to Mhp1 residues Trp117, Thr313, Asn314 and Leu363, are identically conserved in VPA1242 and in CodB.

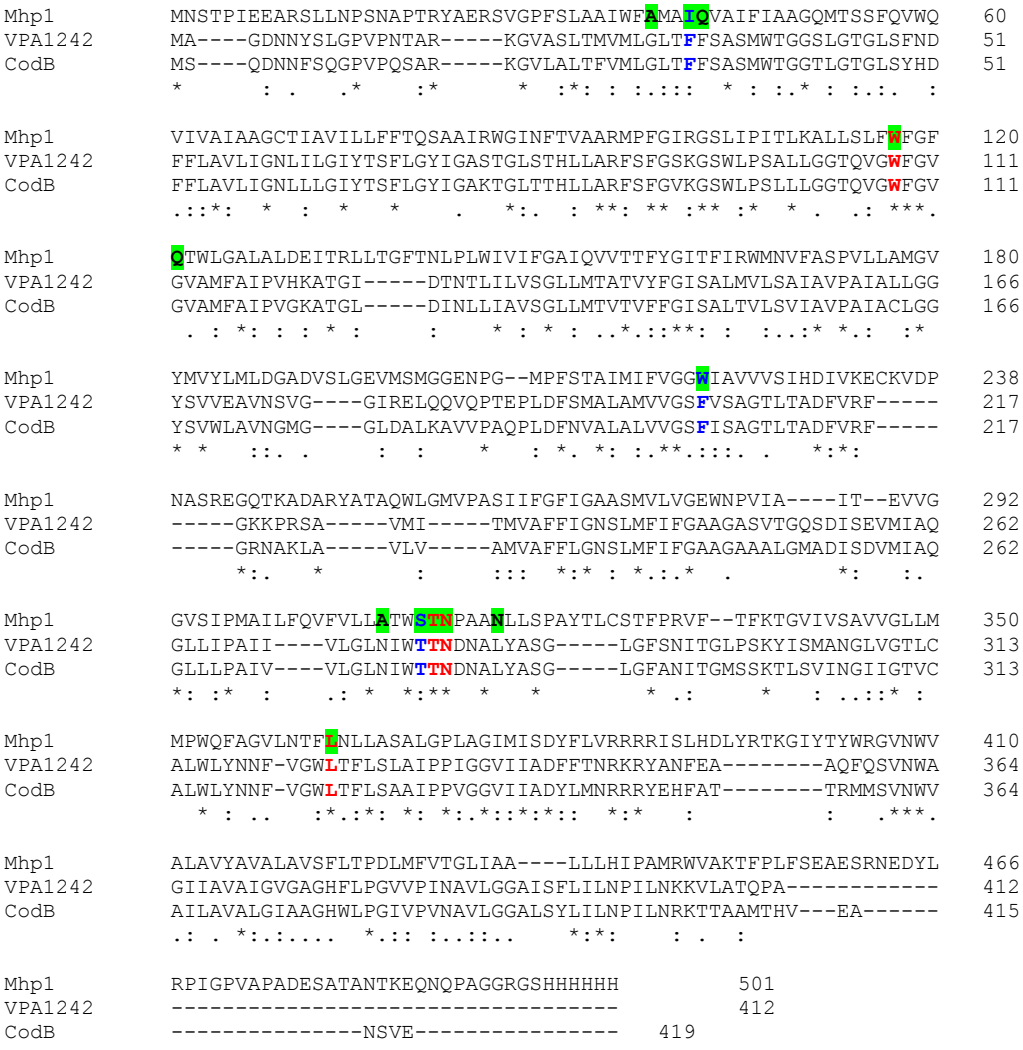



**Figure 5.24 Conservation of residues between the VPA1242 protein of *Vibrio parahaemolyticus*, Mhp1 and CodB.** Amino acid sequences of VPA1242 from *Vibrio parahaemolyticus* (Q87GS3), Mhp1 from *Microbacterium liquefaciuens* (D6R8X8) and CodB from *E. coli* (P0AA82) were taken from the UniProt KnowledgeBase (http://www.uniprot.org/) and aligned using the multiple sequence alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Conserved residues are indicated below the sequences as identical (*), highly similar (:) and similar (.). Residues in the ligand and sodium binding site of Mhp1 are highlighted in green and those in VPA1242 and CodB that are identical or highly similar in are coloured red and blue, respectively.

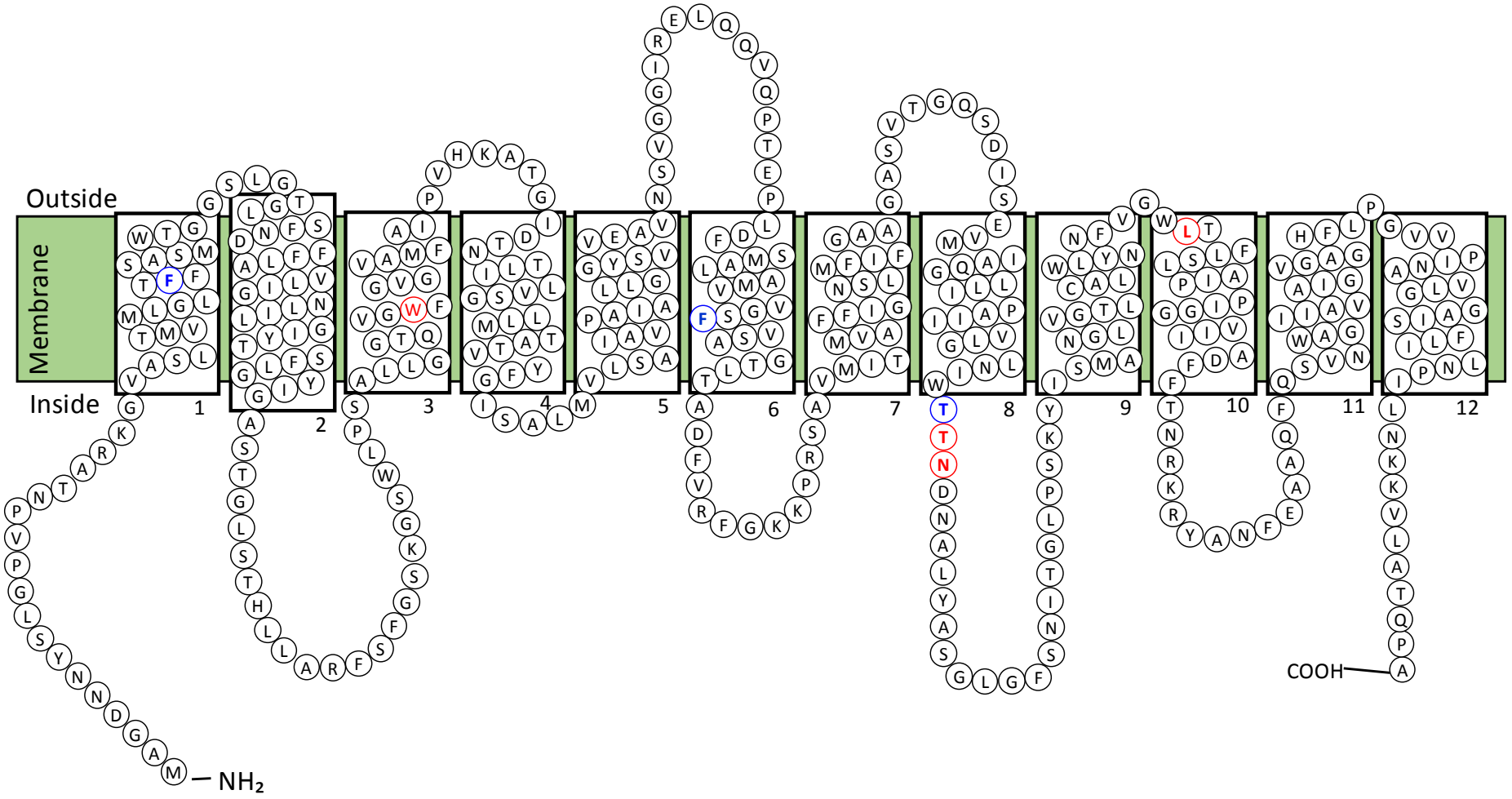



**Figure 5.25 Predicted topology of the *Vibrio parahaemolyticus* membrane transport protein VPA1242.** Diagram for the putative topology of the VPA1242 protein of *Vibrio parahaemolyticus* based on the TOPCONS prediction for transmembrane helices. Residues are coloured to show those that are identical (red) or highly similar (blue) compared with corresponding positions in the sodium and substrate binding site of Mhp1.

## 5.10 Substrate specificity for protein VPA1242

### 5.10.1 Potential substrates for protein VPA1242 in energised whole cells

Transport experiments in energised whole *E. coli* cells were performed to test a number of radiolabelled compounds as potential substrates of the VPA1242 protein. The compounds tested were: $^{3}$H-benzylhydantoin, $^{14}$C-uracil, $^{3}$H-cytosine, $^{3}$H-thiamine. The results are shown in Figure 5.26. Uptake of $^{3}$H-cytosine was significantly greater in induced cells compared with uninduced cells. These initial transport measurements suggest that the main substrate of VPA1242 is cytosine, which is consistent with a high homology of VPA1242 with CodB from *E. coli*.

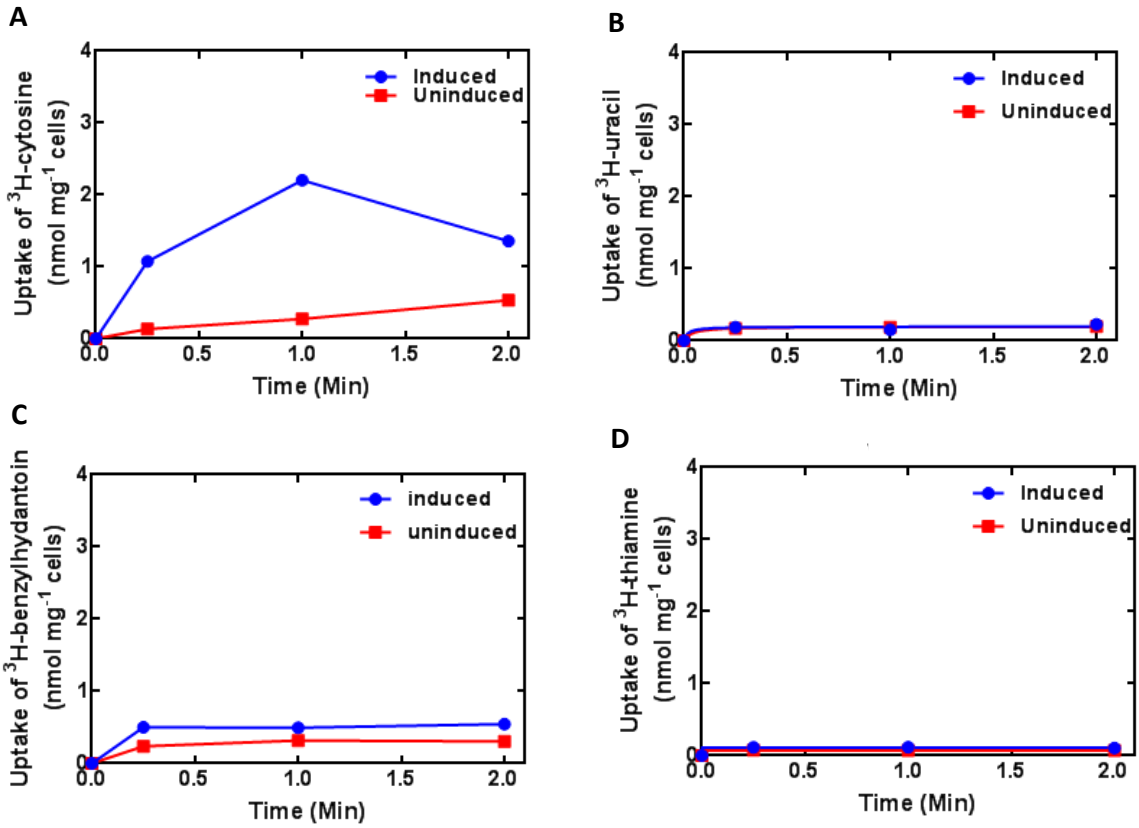



**Figure 5.26 Substrate specificity of the VPA1242 protein.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/VPA1242 were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were left uninduced or were induced at $A_{680}$ = 0.6 with IPTG (0.5 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer (150 mM KCl, 5 mM MES pH 6.6) and resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Cells were energised and tested for uptake of 50 µM different potential radiolabelled substrates.

### 5.10.2 Uptake of $^{3}$H-cytosine by VPA1242

To test that cytosine is the main substrate of VPA1242, *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/VPA1242 and the empty plasmid pTTQ18 were tested for transport of $^{3}$H-cytosine. Uptake of $^{3}$H-cytosine was significantly greater in induced cells compared with uninduced cells containing pTTQ18/VPA1242 (Figure 5.27). As a control, induced and uninduced cells containing plasmid pTTQ18 with no gene insert were tested for the ability to transport $^{3}$H-cytosine (Figure 5.27). The results indicate negligible uptake in these cells, which also confirms that uptake into uninduced cells containing pTTQ18/VPA1242 is due to leaky expression. The rapid decrease in cytosine uptake after the peak might indicate cytosine metabolism inside the cells. The radiolabelled cytosine used in this study is $^{3}$H-cytosine and it is most likely that cytosine metabolism leads to loss of $^{3}$H as $^{3}H_2O$. These results confirmed that cytosine is a main substrate of VPA1242.

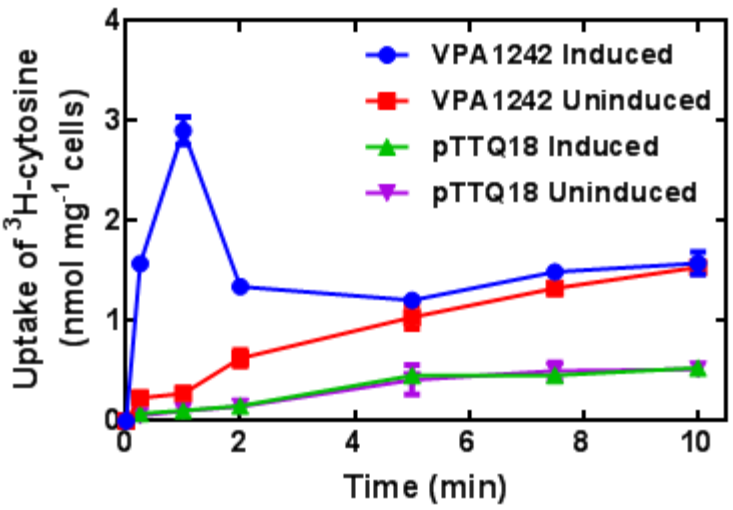



**Figure 5.27 Confirmation of $^{3}$H-cytosine transport by VPA1242.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/VPA1242 or the empty plasmid pTTQ18/no gene were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were left uninduced or were induced at $A_{680}$ = 0.6 with IPTG (0.5 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer and resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Cells were energised and tested for uptake of 50 µM $^{3}$H-cytosine at the given time points.

### 5.10.3 Assessing the effect of sodium ions on the uptake of cytosine by VPA1242

In order to assess the effect of sodium ions on the uptake of $^{3}$H-cytosine by VPA1242, transport of $^{3}$H-cytosine was measured in the presence of various concentrations of NaCl in the range 0-150 mM. In each case, the overall salt concentration was kept constant at 150 mM by using the appropriate concentration of KCl in the transport buffer. Under conditions of zero NaCl, crown ether at a concentration of 10 mM was included to remove any traces of NaCl. At time points of both 15 seconds and 2 minutes, no significant difference in uptake of $^{3}$H-cytosine was observed over the entire range of NaCl concentrations (Figure 5.28). These results demonstrate that $^{3}$H-cytosine transport by VPA1242 is not dependent on sodium.

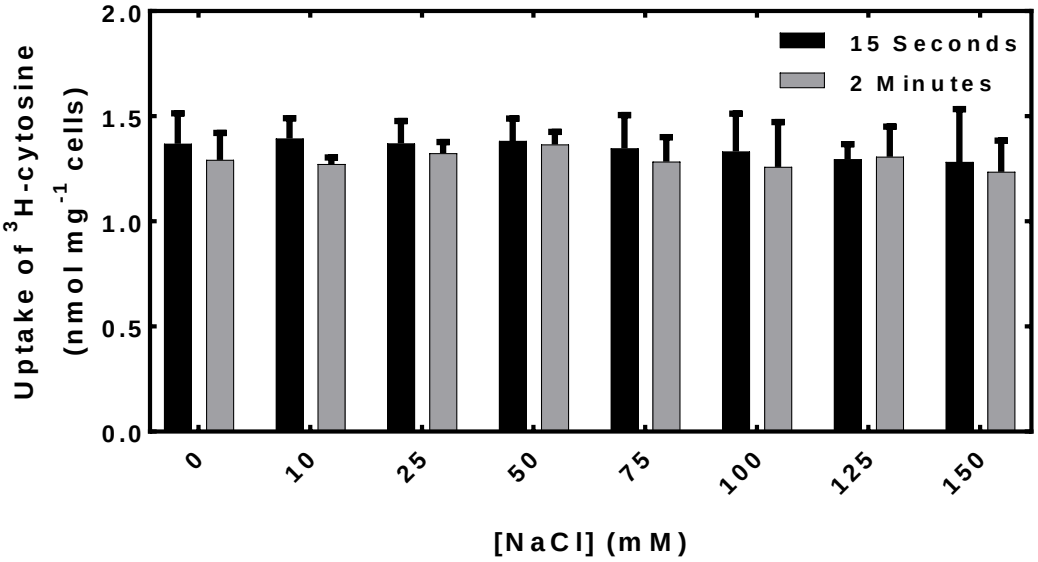



**Figure 5.28 Effect of sodium ion concentration on uptake of $^{3}$H-cytosine by cells expressing VPA1242.** *E. coli* BL21(DE3) cells harbouring the plasmid pTTQ18/ VPA1242 were grown in LB medium supplemented with 20 mM glycerol and 100 µg/ml of carbenicillin at 37 °C with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with IPTG (0.5 mM) and then grown for a further 1 hour. Harvested cells were washed three times using transport assay buffer containing 5 mM MES (pH 6.6) and a range of NaCl concentrations from 0-150 mM balanced by a range of KCl concentrations to maintain an overall salt concentration of 150 mM. Cells were resuspended in the same buffer to an accurate $A_{680}$ of around 2.0. Aliquots of cells were energised with 20 mM glycerol and tested for uptake of 50 µM $^{3}$H-cytosine at the given time points. The data represent the mean of duplicate measurements.

### 5.10.4 Substrate specificity of VPA1242 using competition assay

The substrate specificity of VPA1242 was measured using radiolabelled $^{3}$H-cytosine uptake in *E. coli* cells expressing VPA1242 in the presence of a ten-fold molar excess of potential unlabelled competing substrates (Figure 5.29). The best competing effect was produced by cytosine, which reduced uptake to 10.7% and 11.5% at time points of 15 seconds and 2 minutes, respectively, confirming that VPA1242 mediates the transport of cytosine. The next most effective competitors were hydantoin, benzylhydantoin and uracil which reduced uptake to 52.7/70.7%, 58.1/86.1% and 58.4/87.8% after 15 seconds and 2 minutes, respectively. The results suggest that these are also possible substrates transported by VPA1242. The structure of cytosine contains a six-membered ring, but still share some common structural features with hydantoin, benzylhydantoin and uracil. Overall the results demonstrate high specificity for recognition of cytosine by VPA1242 and no other nucleobases that were tested here apart from an effect of uracil at 15 seconds.

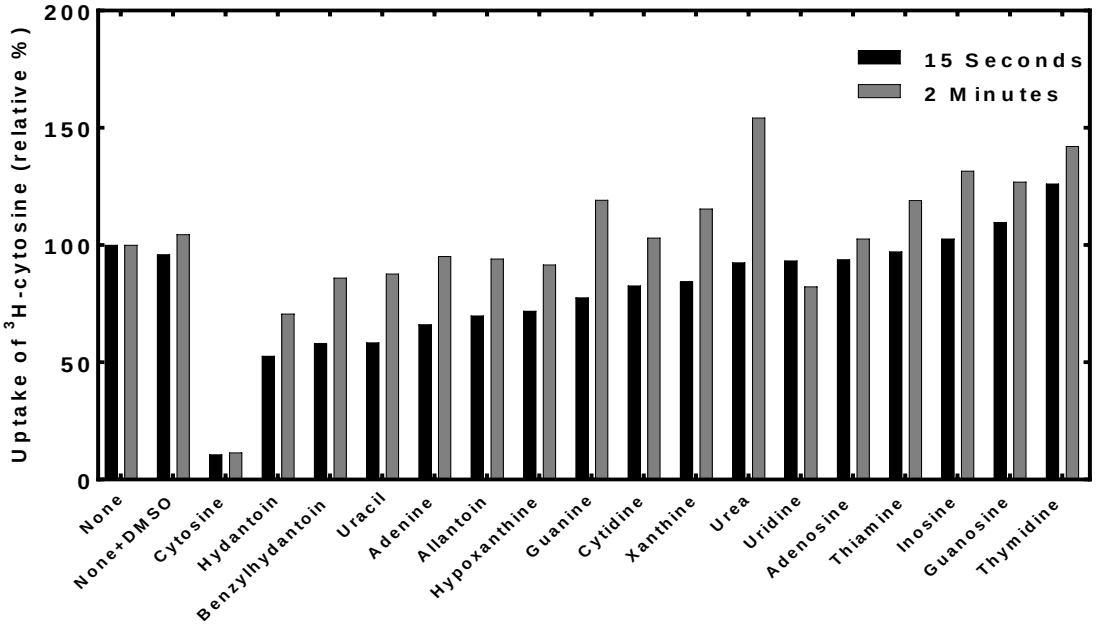



**Figure 5.29 Ligand specificity of VPA1242.** Competition of $^{3}$H-cytosine (50 µM) uptake into *E. coli* BL21(DE3) cells expressing VPA1242 in the presence of a ten-fold molar excess of potential competitors. The non-competed uptake rate was taken as 100% corresponding to 15 seconds and 2 minutes post-additon of $^{3}$H-cytosine, respectively. Some of the compounds were added from a stock solution in 100% DMSO and the final concentration of DMSO in the assay mixture was 1%. All data represent the average of duplicate measurements. None = no competitors.

## 5.11 Purification and yield of VPA1242

From preparations of mixed membranes, inner and outer membranes were isolated by sucrose density gradient ultracentrifugation (Section 2.4.2) (Ward et al., 2000). VPA1242 was purified from inner membranes by solubilisation in 1% DDM and the recombinant protein was purified by Ni-NTA affinity column chromatography to exploit the C-terminal $His_6$-tag. Purification was achieved by washing the column with 20 mM imidazole followed by elution with 200 mM imidazole (Section 2.4.3). Various fractions from the purification of VPA1242 were analysed by SDS-PAGE and Western blotting (Figure 5.30). Lane 6 shows the purified VPA1242 protein, which has a purity of 85% determined by densitometry; this includes the upper band at around 50 kDa, which is likely to be the dimer. The monomer migrates at a molecular weight position of around 30 kDa, which is lower than the calculated molecular weight of 42,821 Da. This is consistent with SDS-PAGE analysis of other NCS1 family proteins, which migrate at approximately 62-74% of their predicted sizes (Bettaney, 2008; Ma, 2010; Ma et al., 2016). On the gel and the Western blot, the intensity of the bands in lane 6 relative to the other lanes demonstrates an efficient purification of VPA1242. The yield of purified protein was estimated from a typical 30-litre fermenter culture of *E. coli* BL21(DE3) cells harbouring plasmid pTTQ18/VPA1242 in LB medium supplemented with 20 mM glycerol, which produced 282.5 g of cell pellet. Membranes prepared from these cells equated to a total volume of 38.8 ml of inner membranes. Based on the purification shown in Figure 5.28, the yield of purified protein is 0.9 mg per litre of cell culture.

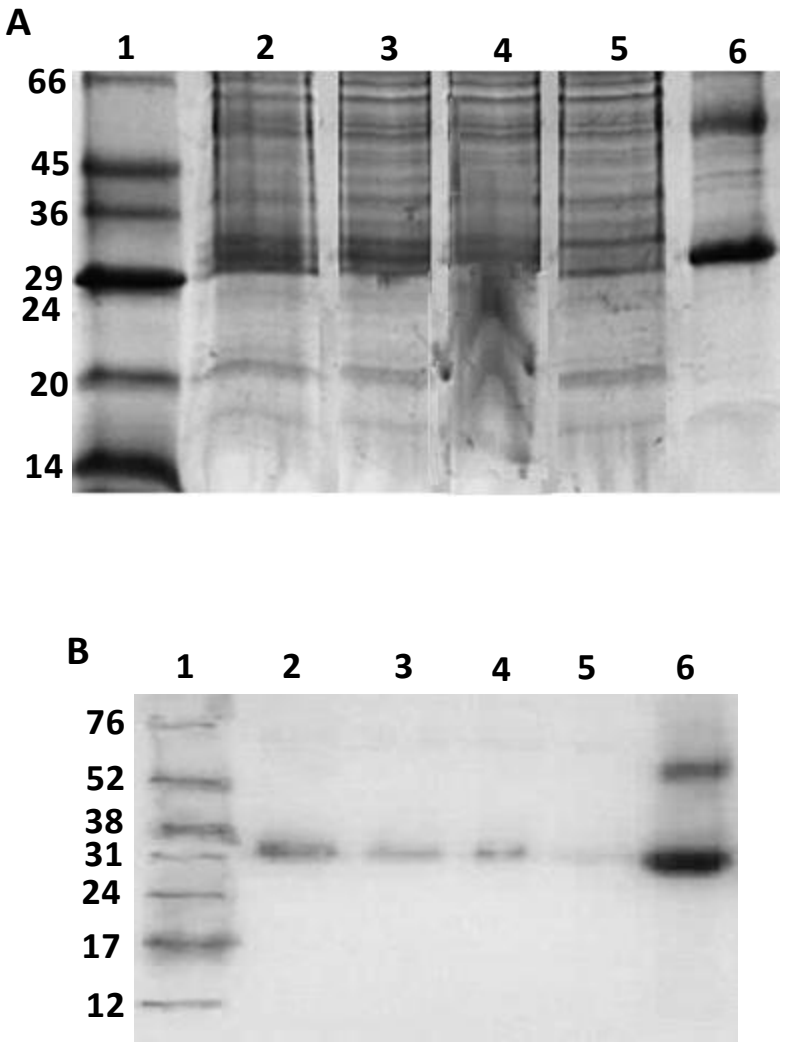



**Figure 5.30 Purification of VPA1242 from inner membranes**. SDS-PAGE (A) and Western blot (B) analysis for purification of VPA1242. Inner membranes were solubilised in 1% DDM. Samples were loaded as follows: (1) molecular weight markers (kDa); (2) inner membranes; (3) supernatant (soluble fraction); (4) membrane pellet (insoluble fraction); (5) column flow-through (unbound fraction); (6) purified protein. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

## 5.12 Secondary structure integrity and thermal stability of VPA1242 protein

Far UV CD (180-260 nm) is a convenient method to analyse the secondary structural integrity of membrane proteins (Wallace et al., 2002; Kelly et al., 2005; Whitmore and Wallace, 2008). VPA1242 protein was purified and the secondary structure was analysed by far UV CD spectroscopy (section 2.5.1). The CD spectrum obtained for purified $His_6$-tagged VPA1242 protein (Figure 5.31A) showed two characteristic negative peaks at 209 nm and 222 nm and a positive peak at ~192 nm consistent with high alpha helix content (Wallace et al., 2003; Kelly et al., 2005; Bulheller et al., 2007). This also confirms that VPA1242 has retained its secondary structure during the purification process.

A thermal stability experiment of VPA1242 was carried out by varying the temperature from 5 to 90 ºC and then back to 5 ºC (Figure 5.31 B and C). The CD spectra of the purified VPA1242 protein demonstrated three stages in the unfolding process. From 5 ºC to 30 ºC there was a slow increase in the ellipticity value indicating little change of secondary structure. From 30 ºC to 60 ºC there was a greater increase in the ellipticity value indicating a significant change of the secondary structure. From 60 ºC to 90 ºC the increase in ellipticity was slow again, indicating that the change in secondary structure of the purified VPA1242 was complete. Refolding of many alpha-helical membrane proteins have been reported and bacteriorhodopsin was the first alpha-helical membrane protein that was successfully refolded from SDS denatured state (Huang et al., 1981; London and Khorana, 1982). Since then, many other alpha-helical membrane proteins have been testified including lactose permease LacY116; diacylglycerol kinase114 and a potassium channel KcsA115 (Nagy et al., 2001; Valiyaveetil et al., 2002; Harris et al., 2014). In our results the CD unfolding experiment indicates that purified VPA1242 had retained its structure prior to performing the thermal unfolding experiment. On returning the temperature from 90 ºC to 5 ºC, there was no evidence to suggest refolding of the protein and the same is reported by other researchers (Bettaney, 2008; Ma, 2010; Sukumar, 2012; Jackson, 2012). Finally, a melting temperature of 42.8 ºC was estimated using Global Analysis CD software 3. The VPA1242 protein is therefore reasonably stable for performing biophysical assays using a temperature range of 18-25 ºC.

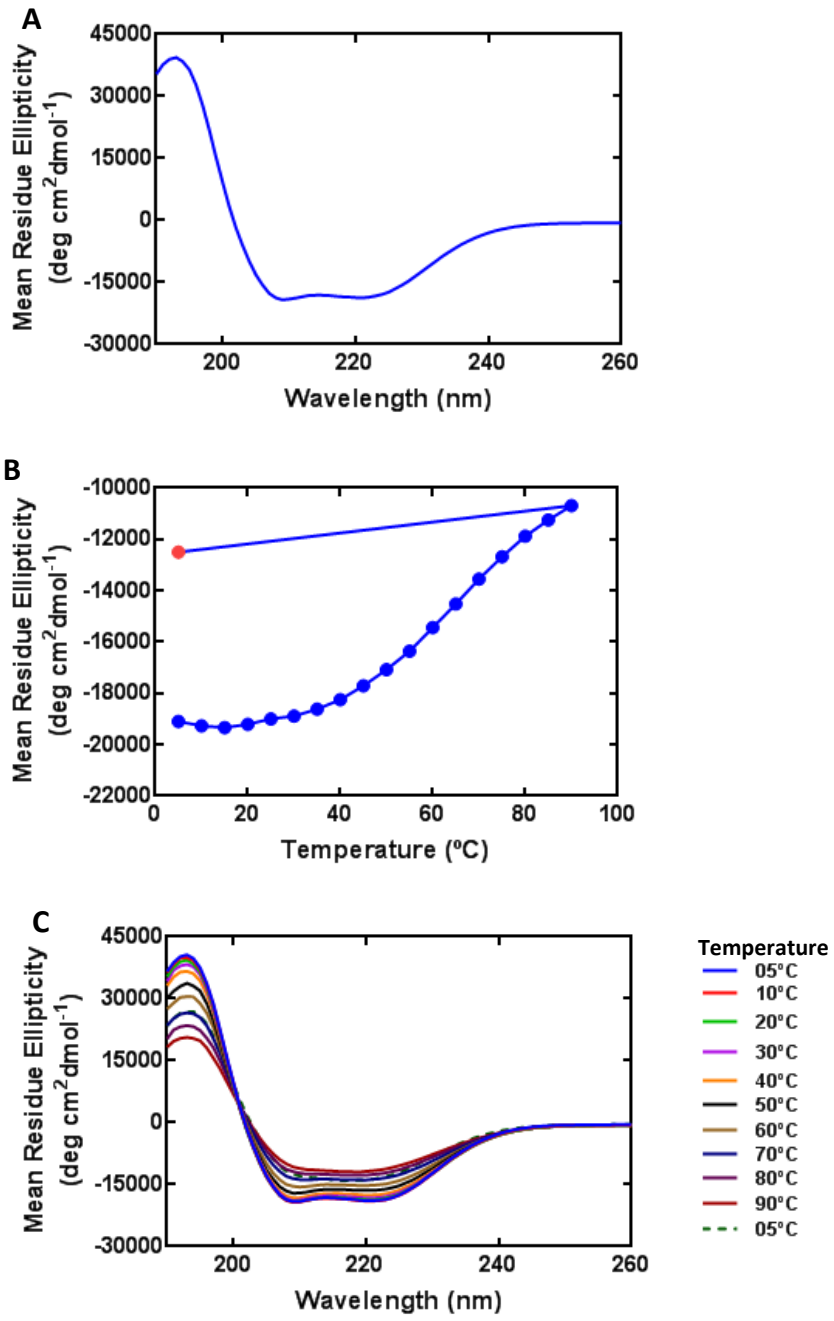



**Figure 5.31 Far-UV CD analysis of purified VPA1242 protein.** (A) Far-UV (180-260 nm) CD spectrum for purified VPA1242 protein. Measurements were performed using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant liquid nitrogen flushing. Samples were prepared in a Hellma quartz cuvette of 1.0 mm pathlength at a final protein concentration of 0.15 mg/ml in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM). (B) Thermal unfolding of VPA1242 protein over the concentration range 5-90 ºC and finally back to 5 ºC monitored at a wavelength of 209 nm. (C) Full spectra showing thermal unfolding of VPA1242 protein over the given temperature range.

### 5.13 Fluorimetric analysis of ligand binding to VPA1242

Steady-state fluorimetric measurements were performed for detecting binding of cytosine to purified VPA1242 by exploiting the fluorescence properties of intrinsic protein tryptophan residues. The quench in fluorescence intensity observed for titration of cytosine with VPA1242 in absence and presence of added NaCl was similar in both cases and have an approximately linear relationship (Figure 5.32). This suggested that the observed quenching effect may be due to non-specific interactions or absorbance of the excitation or emission wavelengths of light by cytosine itself. Therefore, a control experiment was performed by titrating cytosine without protein, indicated a negative fluorescence quenching percentage (Figure 5.32) suggesting that cytosine itself absorb more light in this range of wavelength.

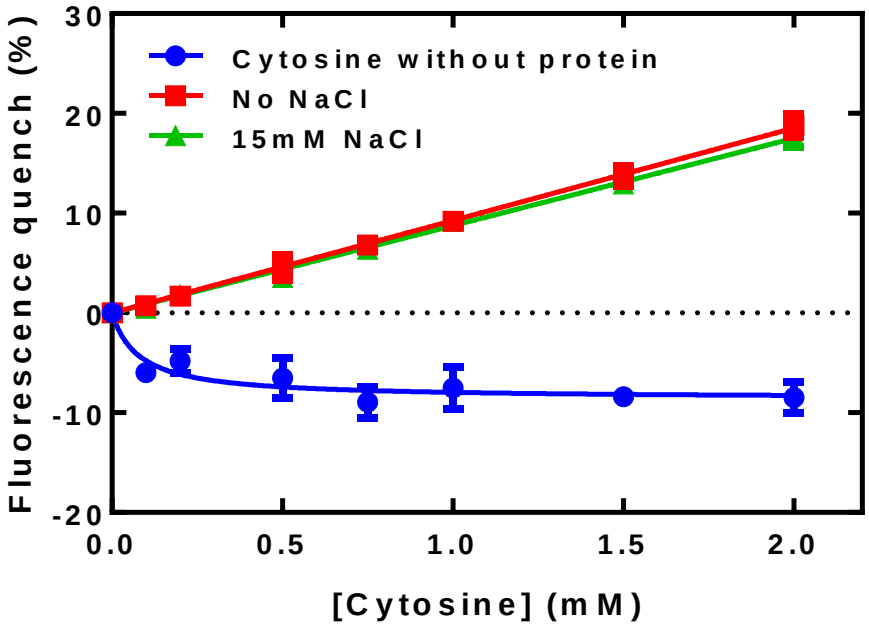



**Figure 5.32 Binding of cytosine to VPA1242 by quenching of intrinsic fluorescence.** Steady-state spectrophotofluorimetry measurements were performed for detecting cytosine binding to purified VPA1242. Samples containing protein (2.5 µM) in fluorescence buffer (10 mM Tris-HCl pH 7.6; 0.05% DDM; 2% DMSO; 140 mM choline chloride) at 18 ºC were excited at 295 nm and fluorescence emission was measured at 332 nm. Following cytosine additions, samples were stirred for 1.5 minutes before making the measurements. Various samples were titrated with cytosine that contained buffer alone (blue), VPA1242 with 15 mM NaCl (green) and VPA1242 without NaCl (red). The data represent triplicates measurements and were analysed using the Michaelis-Menten equation in GraphPad Prism 7 software.

## 5.14 Conclusions

This chapter described characterisation of two bacterial transport proteins of the NCS1 family, AAN69889 from *Pseudomonas putida* and VPA1242 from *Vibrio parahaemolyticus*. In the genome of *P. putida* the gene for AAN69889 (510 residues) has a locus position directly upstream from a gene that encodes a putative hydantoin racemase. Consistent with characterised members of the NCS1 family, AAN69889 is predicted to contain twelve transmembrane spanning apha-helices with both the N- and C-terminal ends in the cytoplasm. Out of the other characterised bacterial NCS1 family proteins, AAN69889 is most closely related to the allantoin transporter PucI from *Bacillus subtilis* with an overall sequence homology of 58.5% (33.8% identical plus 24.7% highly similar residues). Transport activities of various compounds in energised whole cells were measured using a range of potential radiolabelled substrates that initially demonstrated highest uptake for $^{14}$C-allantoin and no significant uptake of any other compounds including $^{3}$H-*L*-5-benzylhydantoin, $^{14}$C-*L*-5-indolylmethylhydantoin, $^{14}$C-uridine, $^{3}$H-cytosine and $^{3}$H-thiamine. Similar to PucI, $^{14}$C-allantoin uptake by AAN69889 was not dependent on sodium, suggesting that transport is driven by a proton gradient. The concentration-dependence of $^{14}$C-allantoin uptake by AAN69889 conformed to Michaelis-Menten kinetics producing values for the apparent affinity for substrate ($K_m$) and maximum velocity (Vmax) of 39.2 ± 2.7 μM and 14.8 ± 0.3 nmol/mg cells/15 sec, respectively. The same parameters for $^{14}$C-allantoin uptake by PucI were 24.4 ± 3 μM and 14.8 nmol/mg cells/15 sec, respectively (Ma et al., 2016). Based on competition of $^{14}$C-allantoin uptake by a ten-fold excess of unlabelled compounds, AAN69889 was highly specific for allantoin (Figure 5.12). The only other compounds that had small but significant competitive effects were uracil and hydantoin, which are structurally similar to allantoin. Protein databases (e.g. UniProt, NCBI) describe AAN69889 as a nucleoside transporter but our results have demonstrated that AAN69889 is a highly specific transporter of allantoin with no significant recognition of nucleosides or nucleobases. This emphasises the importance of laboratory experiments for confirming substrate specificity and other characterisations of membrane transport proteins. It is fair to say that it is not always appropriate to rely on databases and/or *in silico* predictions for such information. IMAC purification of AAN69889 using a Ni-NTA resin produced protein with a

purity of around 80% and a yield of around 1.2 mg/litre from fermenter cultures using LB medium supplemented with glycerol. Circular dichroism measurements confirmed the integrity of the secondary structure of AAN69889 and indicated that the detergent-solubilised purified protein has an alpha-helix content. A melting temperature of 46.2 ºC was estimated using Global Analysis CD software-3 and therefore AAN69889 protein is reasonably stable for performing biophysical assays using a temperature range of 18-25 ºC. Fluorimetry measurements found no significant difference in the fluorescence quench obtained with allantoin suggesting that the binding site in AAN69889 may not involve tryptophan residues.

Protein databases describe VPA1242 from *V. parahaemolyticus* as a putative cytosine permease and is most closely related to cytosine transporter CodB from *E. coli* (Table 4.1). Transport measurements with VPA1242 in energised whole cells using a range of potential radiolabelled substrates demonstrated highest uptake for $^{3}$H-cytosine and no significant uptake of any other compounds. Based on competition of $^{3}$H-cytosine uptake by a ten-fold excess of unlabelled compounds, VPA1242 was highly specific for cytosine. The only other compounds that had small but significant competitive effects were hydantoin, benzyhydantoin, and uracil, which share some structural similarity to cytosine. $^{3}$H-cytosine uptake by VPA1242 was not dependent on sodium, suggesting that transport is driven by a proton gradient. IMAC purification of VPA1242 using a Ni-NTA resin produced protein with a purity of around 85% and a yield of around 0.9 mg/litre from fermenter cultures using LB medium supplemented with glycerol. The circular dichroism confirmed the integrity of the secondary structure of VPA1242 and indicated that the detergent-solubilised purified protein has an alpha-helix content. Thermal unfolding and refolding was analysed by ramping the temperature from 5-90ºC and finally back to 5ºC. On returning the temperature to 5ºC, there was no evidence to suggest refolding of the protein. A melting temperature of 42.8ºC was estimated using Global Analysis CD software-3. The quench in fluorescence intensity observed for titration of cytosine with VPA1242 in the absence and presence of added NaCl was similar in both cases and had an approximately linear relationship. This suggested that the observed quenching effect may be due to non-specific interactions or absorbance of the excitation or emission wavelengths of light by cytosine itself. Therefore, a control experiment was performed by titrating cytosine without protein,

indicated negative fluorescence quenching percentage suggesting that cytosine itself absorbs more light in this range of wavelength.

In this project, AAN69889 and VPA1242 were successfully cloned into plasmid pTTQ18 and expressed in *E. coli* strain BL21(DE3) followed by large scale production and purification of proteins. Biochemical and biophysical studies have confirmed that protein AAN69889 and VPA1242 were correctly folded and functional. This study will help to progress further research, especially mutagenesis studies to identify residues in AAN69889 and VPA1242 that could be important for substrate binding and recognition and crystallisation trials for structural studies.

# Chapter-6

# Expression screening, purification and characterisation of AceI and its homologues

## 6.1 Introduction

Currently, all over the world, the multidrug resistance rate is dramatically increasing and the mechanisms of resistance are complicated. Active multi-drug efflux systems are one of the major mechanisms of bacterial resistance to drugs and are an alarming threat to antibiotic therapy. There are diverse mechanisms which contribute to inherent and acquired resistance to antimicrobial compounds. Gram-negative bacteria are commonly intrinsically more resistant to many biocides as a result of their cell structure and the activity of multidrug efflux pumps (Alibert-Franco et al., 2010; Nikaido and Pagès, 2012; Kumar and Varela, 2012; Blair et al., 2014; Li et al., 2015). In prokaryotes efflux pumps have various important physiological functions including self-defence systems (Webber and Piddock, 2003). Owing to the widespread presence of these pumps in pathogens, five families of bacterial multidrug efflux systems were identified until 2014 (Kourtesi et al., 2013). These families are: the adenosine triphosphate (ATP)-binding cassette (ABC) superfamily (Lubelski et al., 2007); the major facilitator superfamily (MFS) (Marger and Saier, 1993; Pao et al., 1998; Law et al., 2008); the small multidrug resistance (SMR) family (Chung et al., 2001); the resistance-nodulation-division (RND) family (Tseng et al., 1999; Seeger et al., 2008; Nikaido and Takatsuka, 2009); and the multidrug and toxic compound extrusion (MATE) family (Kuroda and Tsuchiya, 2009).

Recently 23 homologues of the AceI protein have been reported to represent a new sixth family of bacterial multidrug efflux pumps, designated the Proteobacterial Antimicrobial Compound Efflux (PACE) family of transport proteins (Hassan et al., 2015). The ABC, MFS, MATE and SMR families are widely distributed in Gram-negative and Gram-positive bacteria, whereas the RND and PACE family are specific to Gram-negative microorganisms. These families are classified on the basis of their amino acid sequence similarity, substrate specificity and the energy source used to export their substrates (Blanco et al., 2016). Only the ABC superfamily is a primary transporter utilizing ATP hydrolysis to expel substrates, while the other families are secondary transporters using the proton or sodium gradient as the energy source. The PACE family prototypical protein AceI (*Acinetobacter* chlorhexidine efflux) from the Gram negative bacterium *Acinetobacter baumanii* has been shown to confer resistance to chlorhexidine, *via* an active efflux mechanism (Hassan et al., 2013; Hassan et al., 2015). Chlorhexidine is a synthetic compound (Figure 6.1)

widely used as an antiseptic and antimicrobial agent, effective against Gram-positive and Gram-negative organisms, facultative anaerobes, aerobes and yeasts (Leikin and Paloucek, 2008). It has been commercially available for more than 60 years and plays an important role in controlling the spread of diseases, notably hospital-acquired infections.

It has been reported that multidrug-resistant *A. baumannii* exposed to 0.2M NaCl up-regulated expression of 150 genes, 20% of which were categorised as transport proteins and specifically 14 belonged to diverse classes of efflux pumps (Hood et al., 2010). In bacterial pathogens, multi-drug efflux mechanisms are a major area of research, so that ultimately measures may be discovered/developed to inhibit these active multi-drug efflux pumps. It is also of prime importance to investigate, whether other identified members of the PACE family show affinity for chlorhexidine. The current lack of a high-resolution structure for PACE family proteins means that any information about their substrate binding and transport mechanism must be indirectly obtained from other types of experiments. In this work I have employed an expression screening and purification strategy to identify PACE family proteins suitable for progression into a structural biology pipeline and for functional characterisation in a similar manner to that used with other bacterial and archaeal transport proteins, including drug efflux pumps (Gordon et al., 2008; Bettaney et al., 2013 ; Ma et al., 2013). This was accompanied by circular dichroism, fluorimetry and transport assays for confirming structural and functional integrity of the proteins and for elucidating structure-function relationships.

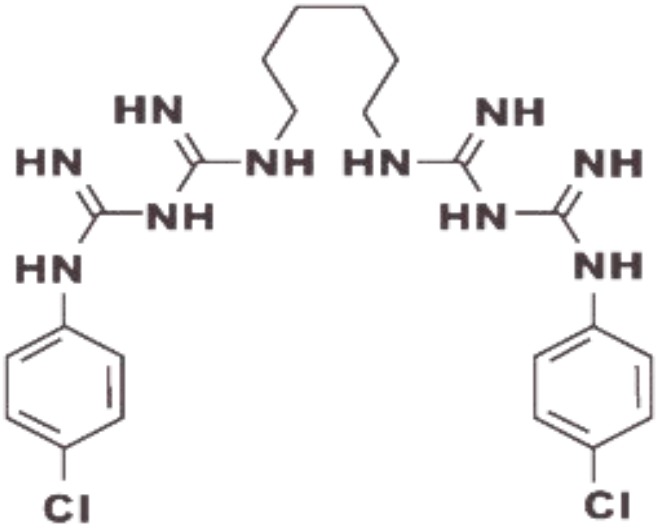



**Figure 6.1 Structure of chlorhexidine**

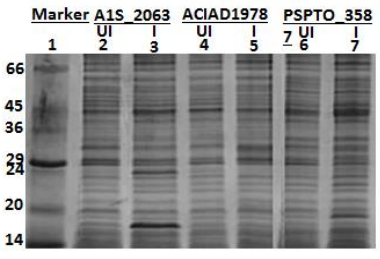
## 6.2 Screening for expression level

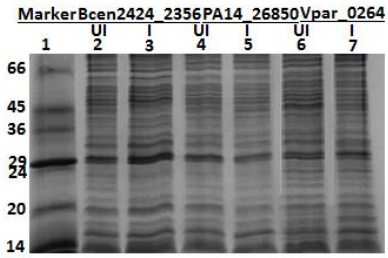
Expression screening was the initial step carried out to test for a sufficient amount of appropriate quality protein, which is one of the main hindrances in the development of structural biology research. To accelerate the purification of proteins, small-scale expression screening was carried out to identify quickly those that would allow sufficient quantities of protein to be produced from larger-scale cultures for structural and functional analysis. All twenty four genes of the PACE family transport proteins were cloned by Karl Hassan (Macquarie University, Australia). The genes and their organism of origin that were cloned into the vector pTTQ18 for recombinant protein expression in *E. coli* BL21 (DE3) are listed in Table 6.1. *E. coli* BL21 (DE3) is the most common expression host, used for recombinant protein production because of its fast growth, reachable high cell densities and culture on inexpensive media (Shiloach and Fass, 2005; Terpe, 2006; Sezonov et al., 2007; Zhang et al., 2015). The expression levels of the twenty four recombinant proteins from small-scale *E. coli* cultures are shown in Figure 6.2.

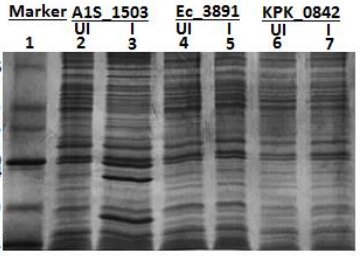
Usually the proteins that are identified by Western blot analysis but fail to visualise in SDS-PAGE, are not considered the best candidates to pursue for purification by Ni-NTA chromatography due to low yields containing high contaminants. The proteins (AceI, Fbal_3166, STY_3166, Tmarg_opt, PFL_4558, A1S_1503 and PSPTO_3587) were expressed at a level sufficient for producing larger-scale cultures, purification, and for structural and functional characterisation. All the above proteins showed clear and intense bands on both gels and Western blots analysis (Figure 6.2). The detection of a signal on the Western blot confirmed that the recombinant expressed protein had retained its $His_6$-tag. Focusing on a smaller number of proteins that show the most potential at this stage was important to make the best use of time and materials. This type of strategy is commonly used in pipelines for structural biology of membrane proteins (Ward et al., 2000; Gordon et al., 2008; Bettaney et al., 2013; Ma et al., 2013; Moraes et al., 2014; Alegre and Law, 2015).

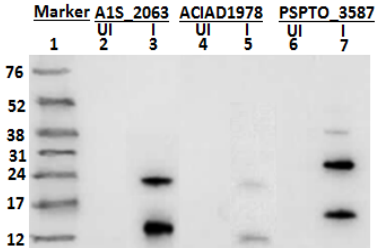
**Table 6.1 PACE family proteins**. Proteins chosen for expression, purification and characterisation. The name refers to the gene name and sizes of proteins taken from the NCBI database. Sequence identy with AceI were calculated using the multiple sequence alignment tool Clustal Omega (Sievers et al., 2011).

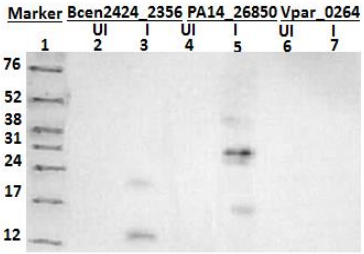
| **Protein name (bacterium)** | **Number of amino acids** | **Identity with AceI (%)** | **Similarity with AceI (%)** |
|---|---|---|---|
| AceI<br>*Acinetobacter baumannii* | 144 | 100.0 | 100.0 |
| ACIAD1978<br>*Acinetobacter baumannii* ATCC 17978 | 152 | 62.6 | 78.3 |
| PSPTO_3587<br>*Pseudomonas syringae pv. tomato str.* DC3000 | 169 | 29.6 | 55.0 |
| Bcen2424_2356<br>*Burkholderia cenocepacia* HI2424 | 148 | 30.4 | 59.5 |
| PA14_26850<br>*Pseudomonas aeruginosa* PA14 | 171 | 31.0 | 55.0 |
| Vpar_0264<br>*Veillonella parvula* DSM 2008 | 134 | 32.4 | 55.6 |
| A1S_1503<br>*Acinetobacter baumannii* ATCC 17978 | 145 | 34.5 | 59.3 |
| Ec_3891 (ECTW07793_0407)<br>*E. coli* TW07793 | 156 | 28.9 | 54.6 |
| KPK_0842<br>*Klebsiella pneumoniae* 342 | 153 | 29.4 | 58.2 |
| PFL_4585<br>*Pseudomonas protegens* Pf-5 | 143 | 31.0 | 54.9 |
| Arad_01702 (HMPREF0018_01702)<br>*Acinetobacter radioresistens* SH164 | 144 | 62.6 | 80.6 |
| PFL_4558<br>*Pseudomonas protegens* Pf-5 | 144 | 34.0 | 62.5 |
| ROS217_23162<br>*Roseovarius* sp. 217 | 145 | 30.2 | 58.6 |
| Entcl_2273<br>*Enterobacter cloacae* SCF1 | 152 | 31.6 | 58.6 |
| Mlut_15630<br>*Micrococcus luteus* NCTC 2665 | 179 | 24.9 | 46.4 |
| PP_3512<br>*Pseudomonas putida* KT2440 | 146 | 28.3 | 62.3 |
| Yreg_01962 (HMPREF0880_01962)<br>*Yokenella regensburgei* ATCC 43003 | 160 | 27.5 | 55.0 |
| Fbal_3166<br>*Ferrimonas balearica* DSM 9799 | 146 | 26.5 | 54.8 |
| Tmarg_opt<br>*Tepidiphilus margaritifer* DSM 15129 | 159 | 29.5 | 56.0 |
| VP1155<br>(*Vibrio parahaemolyticus* RIMD 2210633) | 140 | 30.3 | 55.6 |
| STY3166<br>*Salmonella enterica subsp. enterica serovar Typhi str.* CT18 | 160 | 27.5 | 52.5 |
| RP_mp1531 (RPS107_mp1531)<br>*Ralstonia solanacearum* PSI07 | 145 | 35.1 | 64.1 |
| MHA_0890<br>*Mannheimia haemolytica* PHL213 | 137 | 33.3 | 59.0 |
| P20429_2969<br>*Pseudoalteromonas sp.* BSi20429 | 145 | 32.9 | 61.4 |

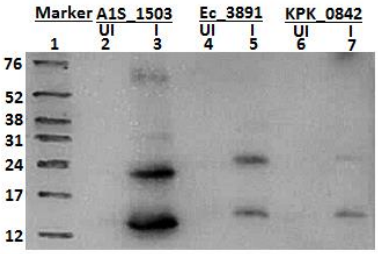

Marker A1S_2063 ACIAD1978 PSPTO_358
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

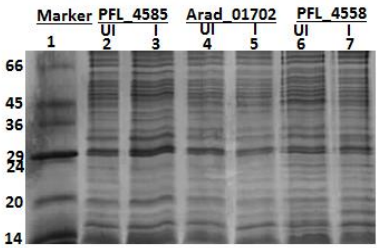

Marker Bcen2424_2356 PA14_26850 Vpar_0264
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

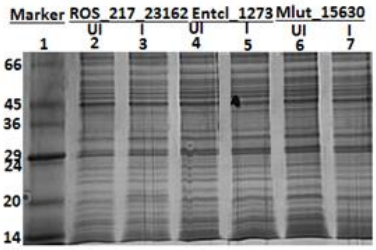

Marker A1S_1503 Ec_3891 KPK_0842
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

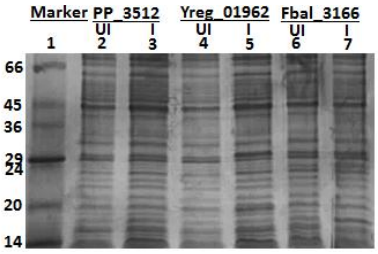

Marker A1S_2063 ACIAD1978 PSPTO_3587
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12

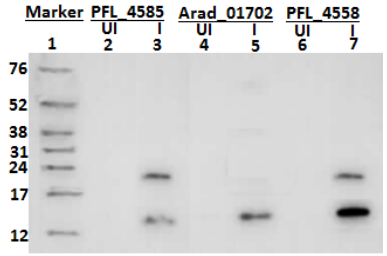

Marker Bcen2424_2356 PA14_26850 Vpar_0264
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12

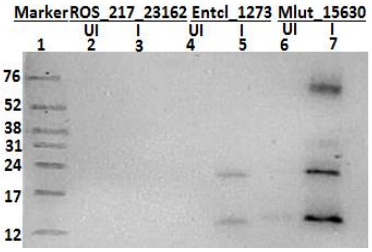

Marker A1S_1503 Ec_3891 KPK_0842
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12

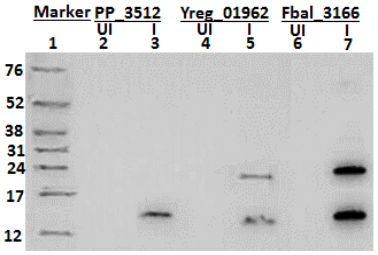

Marker PFL_4585 Arad_01702 PFL_4558
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

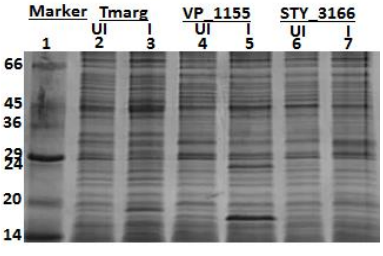

Marker ROS_217_23162 Entcl_1273 Mlut_15630
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

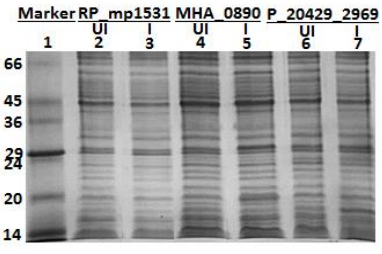

Marker PP_3512 Yreg_01962 Fbal_3166
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14

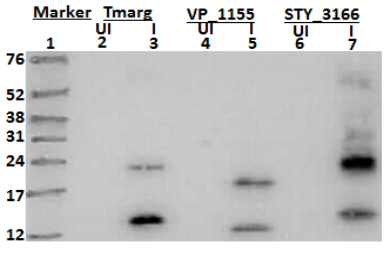

Marker PFL_4585 Arad_01702 PFL_4558
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12


Marker ROS_217_23162 Entcl_1273 Mlut_15630
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12


Marker PP_3512 Yreg_01962 Fbal_3166
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12


Marker Tmarg VP_1155 STY_3166
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14


Marker RP_mp1531 MHA_0890 P_20429_2969
UI I UI I UI I
1 2 3 4 5 6 7
66
45
36
29
24
20
14


Marker Tmarg VP_1155 STY_3166
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12

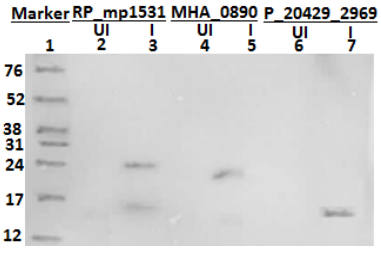

Marker RP_mp1531 MHA_0890 P_20429_2969
UI I UI I UI I
1 2 3 4 5 6 7
76
52
38
31
24
17
12

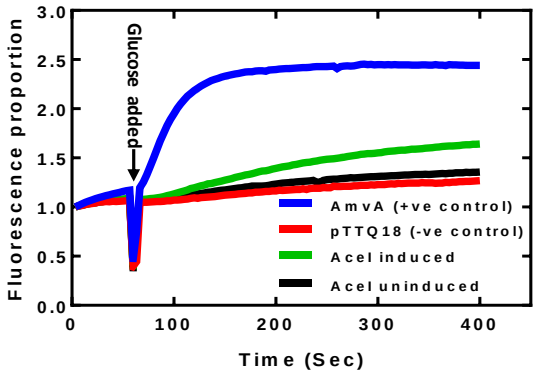
**Figure 6.2 Expression screening of PACE family proteins**. SDS-PAGE (top panel) and Western blot (lower panel) analysis of membrane preparations from small-scale cultures of *E. coli* BL21 (DE3) expressing the recombinant proteins listed in Table 6.1. *E. coli* mixed membranes were prepared using the water lysis method as described in section 2.4.1. Cells were cultured in LB medium containing 20 mM glycerol and 100 μg/ml carbenicillin at 37 °C with shaking at 220 rpm. Cells were left uninduced (UI) or were induced (I) at $OD_{600}$ = 0.8 with 0.2 mM IPTG. Cells were harvested 2 hours after IPTG induction. In each gel and blot, lane 1 shows molecular weight markers.

## 6.3 Efflux mediated fluorimetric transport assay by AceI and its homologs

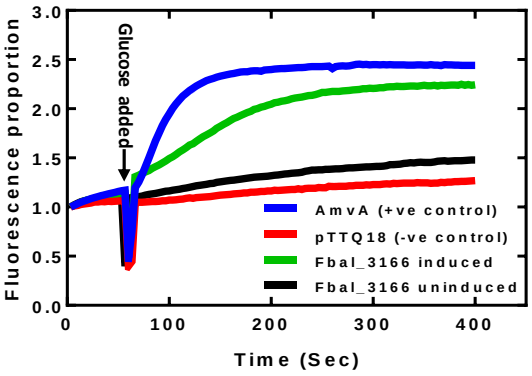
The PACE family prototypical protein AceI was identified to confer resistance to chlorhexidine using an active efflux mechanism (Hassan et al., 2013). Following from this discovery, the resistance function of several AceI homologs were characterised and some were found to confer resistance to other biocides including acriflavine and benzalkonium (Hassan et al., 2015). Acriflavine transport was demonstrated for protein, VP1155 from *Vibrio parahaemolyticus*, using a simple whole cell transport assay (Hassan et al., 2015). The high level of amino acid sequence homology of the proteins in Table 6.1 with AceI (Appendix 5) suggested that these proteins share related biocide transport functions. In order to test this idea, the best expressed proteins, AceI, STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587, were assayed for acriflavine efflux after expression in *E. coli* cells. Cells containing the plasmid pTTQ18 only and cells expressing MFS multidrug efflux pump AmvA were used as a negative and positive control for acriflavine transport, respectively. Acriflavine loaded and energy-starved cells were prepared as described in section 2.2.5. Spectrophotofluorimetry was used to measure acriflavine fluorescence intensity over time to reflect efflux from the loaded cells. Acriflavine is a fluorescent antimicrobial dye and is fluorescent upon excitation at 450 nm (Ma et al., 1995; Dreier and Ruggerone, 2015; Rajamohan et al., 2010). Its fluorescence intensity is lower, when intercalated into nucleic acids, therefore, it is less fluorescent inside cells compared to outside cells and efflux can be observed as an increase in acriflavine fluorescence over time (Chen et al., 2002; Huda et al., 2001; Huda et al., 2003; Dreier and Ruggerone, 2015). The cells were re-energised

with 1% glucose at the point marked with an arrow in Figure 6.3. The change in fluorescence intensity was maximum for the positive control harbouring AmvA and was minimum for the empty plasmid pTTQ18 containing cells. However, the empty plasmid pTTQ18 containing cells have shown some low-level efflux activity, which could be due to the presence of other efflux transporters encoded by *E. coli*. The rate of change in fluorescence intensity of PSPTO_3587, Fbal-3166 and AceI expressing cells upon energisation suggest that these proteins are also able to mediate acriflavine efflux. However, the rate of change in the AceI expressing cells was lower compared to PSPTO_3587 and Fbal_3166. These measurements suggest that acriflavine is a weak substrate for AceI but has some transport and binding affinity. The uninduced samples of these three strains showed some acriflavine efflux actively above pTTQ18 which might be due to leaky expression of PSPTO_3587, Fbal_3166 and AceI. The level of fluorescence in the case of PFL_4558, A1S_1503, STY_3166 and Tmarg_opt from induced and uninduced cells was almost the same as that of the negative control pTTQ18 cells suggesting that acriflavine is not a substrate for these proteins. These results agree with the earlier MIC reported data where PSPTO_3587 and Fbal-3166 have been reported to confer resistance acriflavine (Hassan et al., 2015). However, acriflavine resistance was not reported for AceI. This may be because the fluorescence transport assay used here is likely to be a more sensitive assay for substrates than broth dilution MIC assays.

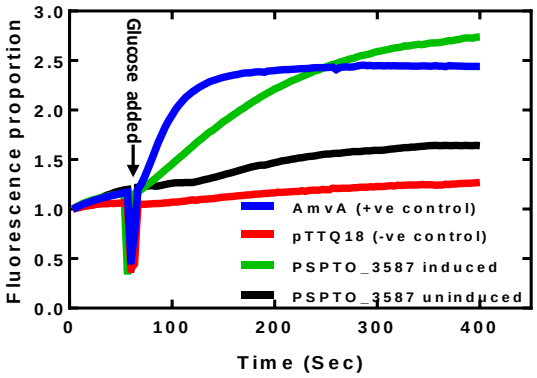
A


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
AceI induced
AceI uninduced
Time (Sec)

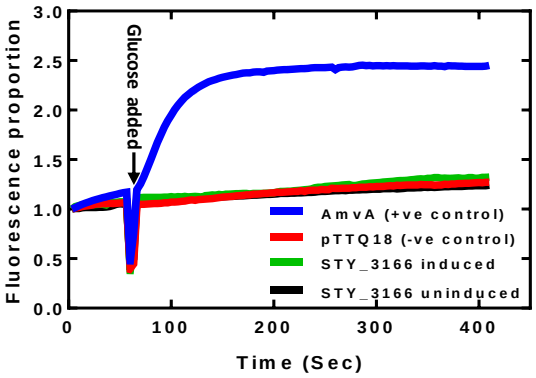
B


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
FbaI_3166 induced
FbaI_3166 uninduced
Time (Sec)

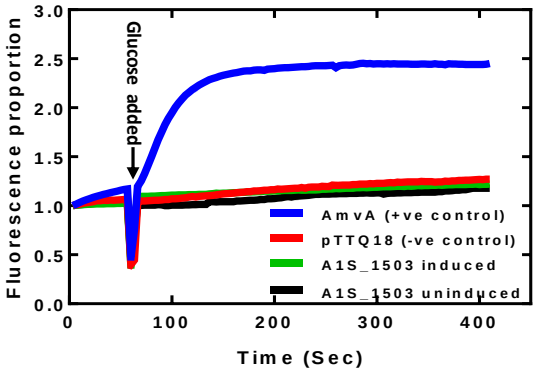
C


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
PSPTO_3587 induced
PSPTO_3587 uninduced
Time (Sec)

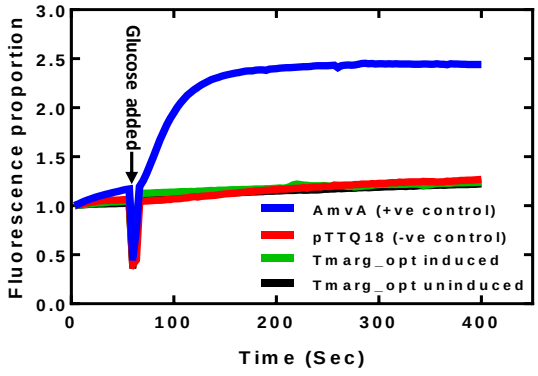
D


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
STY_3166 induced
STY_3166 uninduced
Time (Sec)

E


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
A1S_1503 induced
A1S_1503 uninduced
Time (Sec)

F


Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
Tmarg_opt induced
Tmarg_opt uninduced
Time (Sec)

G

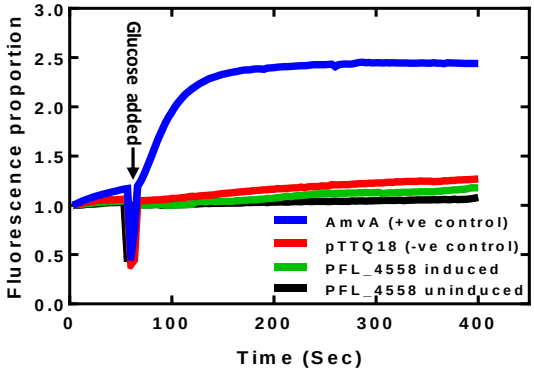

Fluorescence proportion
Glucose added
AmvA (+ve control)
pTTQ18 (-ve control)
PFL_4558 induced
PFL_4558 uninduced
Time (Sec)

**Figure 6.3 Efflux of acriflavine from bacterial cells.** *E. coli* cells carrying empty plasmid pTTQ18 (red); AmvA (blue); AceI (A), Fbal_3166 (B), PSPTO_3587 (C), STY_3166 (D), A1S_1503 (E), Tmarg_opt (F), and PFL_4558 (G), (green [induced] and black [uninduced]) were grown in LB medium containing 100 µg/ml carbenicillin at 37 °C with shaking at 220 rpm. Cells were induced at $A_{680}$ = 0.6 with IPTG (0.2 mM) and then grown for a further 1 hour. Harvested cells were washed three times using buffer (HEPES, pH 7.0) and resuspended in the same buffer to an accurate $A_{680}$ of 1.0 and loaded with 10 µM acriflavine in the presence of 10 µM CCCP. The loaded cells were again washed three times and resuspended in the same buffer. Cells were energised with 1% glucose addition at the point marked with an arrow and transport was initiated. Upward changes in the curve represent efflux of acriflavine from cells. The assay was performed with 1ml of cells at 20 °C using excitation wavelength 450 nm and emission was measured at 510 nm with a bandwidth 0.6 nm. The data represent the means of four replicates measurements.

Efflux mechanisms involve transport proteins that require energy to export substrates. To demonstrate that the efflux mediated by PSPTO_3587, Fbal_3166 and AceI is energy dependent, the levels of acriflavine fluorescence were measured by pre-incubating the cells in 10uM carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), an uncoupling agent of the membrane proton gradient. Glucose was added to the cells as above but no fluorescence change was observed in all types of cells (Figure 6.4). These results suggest that the efflux of acriflavine from *E. coli* cells is energy dependent but the precise mode of energisation is still unknown. Further experiments were performed to determine ion dependency of transport by these proteins.

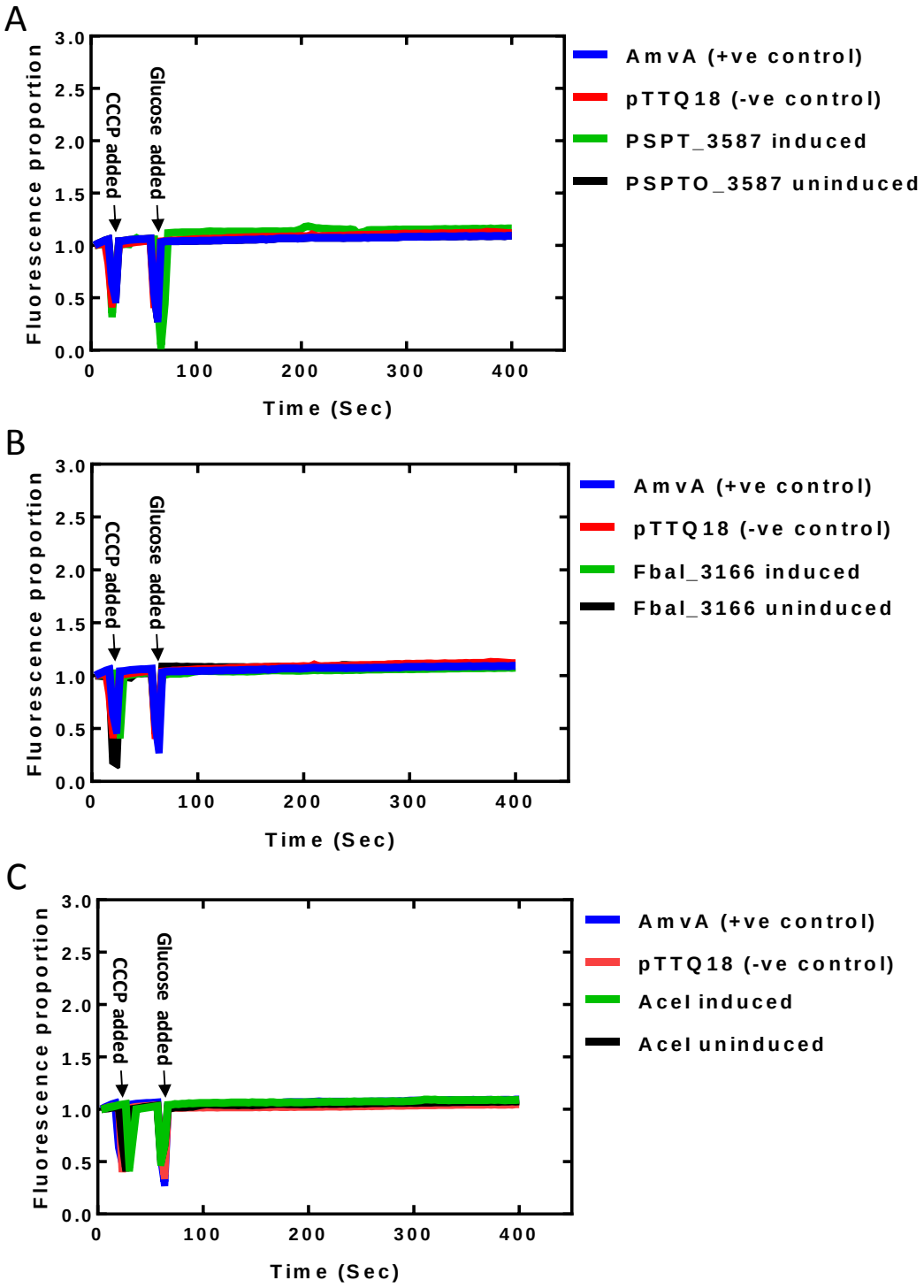
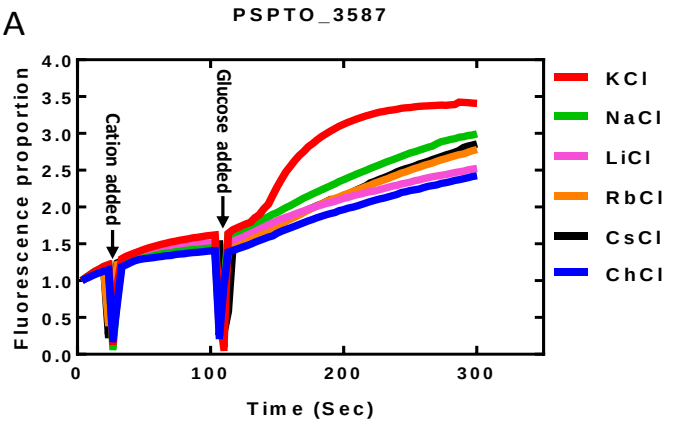


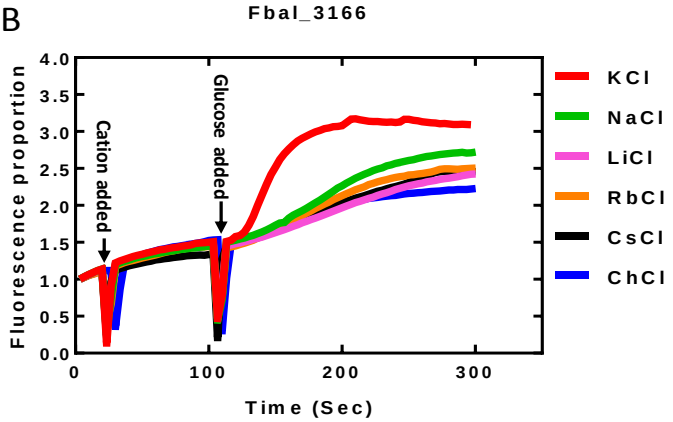

**Figure 6.4 Dependence on energy of acriflavine efflux from *E. coli* cells.** Acriflavine loaded cells were prepared as described in section 2.2.5. *E. coli* cells carrying empty plasmid pTTQ18 (red) as negative control; AmvA (blue) as positive control; PSPTO_3587 (A); Fbal_3166 (B) and AceI (C) (green [induced] and black [uninduced]). To deenergise the cells, initially 10µM CCCP was added indicated by the first arrow followed by 1% addition of glucose indicated by the second arrow. The data represent the means of four replicates measurements.

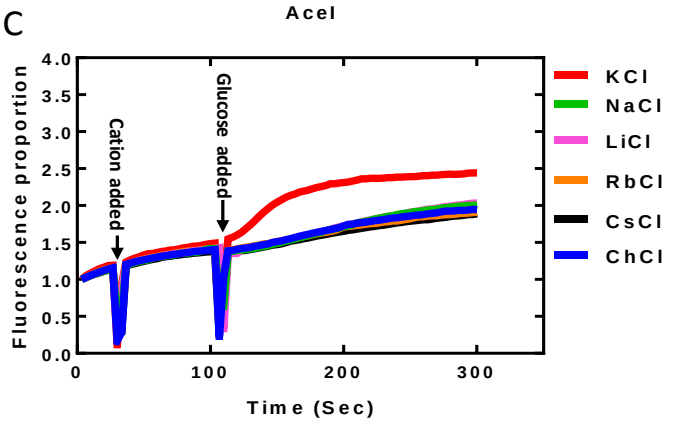
### 6.3.1 Effect of various cations on the acriflavine efflux from bacterial cells

As stated above, different multidrug efflux pumps from different families can have different energy coupling mechanisms. Only the ABC superfamily primary transporters utilize ATP hydrolysis to expel substrates, while the other families are secondary transporters using the proton or sodium gradient as the energy source (Figure 1.18). Several multidrug efflux pumps of the MATE family are reported to utilize $Na^+$ to transport their substrates (Morita et al., 2000; Huda et al., 2001; Chen et al., 2002; Huda et al., 2003; Long et al., 2010).

Furthermore, Mhp1, a major focus of this thesis is a sodium-dependant hydantoin transporter structurally related to other sodium dependent transporter, such as the *Aquifex aeolicus* leucine transporter (LeuT), *E. coli* betaine transporter (BetP) and *Vibrio parahaemolyticus* glucose transporter (vSGLT) (Weyand et al., 2008; Abramson and Wright, 2009; Yamashita et al., 2005; Faham et al., 2008; Krishnamurthy et al., 2009; Ressl et al., 2009). Therefore, it was of interest whether PACE family transporters display any cation dependence for efflux of acriflavine from the cells. The cation specificities of the PSPTO_3587, Fbal_3166 and AceI proteins were investigated by using different cations at a concentration of 15 mM in the fluorescence assay with acriflavine (Figure 6.5). A significantly higher change in fluorescence intensity was obtained with KCl than with all of the other cations (ChCl, LiCl, NaCl, RbCl & CsCl) used, suggesting that rate of efflux may be higher with potassium in the assay buffer. However, the rate of fluorescence change observed during acriflavine transport assays in the cells containing empty plasmid pTTQ18 was also higher in the presence of potassium. Therefore, the addition of potassium may have a general effect on the rate of acriflavine efflux from whole cells, possibly due to some changes in the membrane permeability or a decreased affinity of acriflavine for nucleic acids. The results from this experiment confirm that as compared to the other cations, potassium plays some important role in the acriflavine efflux, which may be specific to this type of assay.


A
PSPTO_3587
Fluorescence proportion
Cation added
Glucose added
Time (Sec)
KCl
NaCl
LiCl
RbCl
CsCl
ChCl


B
Fbal_3166
Fluorescence proportion
Cation added
Glucose added
Time (Sec)
KCl
NaCl
LiCl
RbCl
CsCl
ChCl


C
Acel
Fluorescence proportion
Cation added
Glucose added
Time (Sec)
KCl
NaCl
LiCl
RbCl
CsCl
ChCl

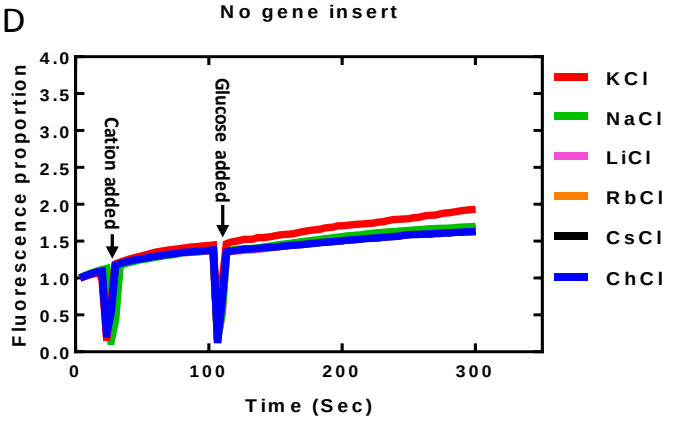
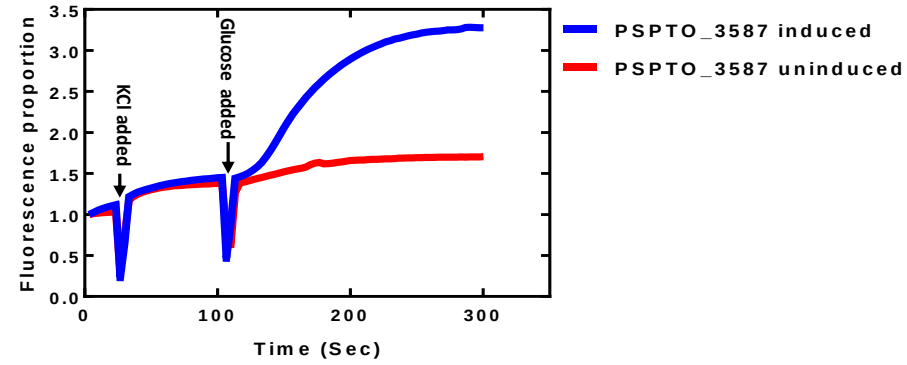

D
No gene insert
Fluorescence proportion
Cation added
Glucose added
Time (Sec)
KCl
NaCl
LiCl
RbCl
CsCl
ChCl

**Figure 6.5 Effect of cations on acriflavine efflux.** *E. coli* cells carrying PSPTO_3587 (A); Fbal_3166 (B); AceI (C) and empty plasmid pTTQ18 (D) were grown and induced as described in section 2.2.5. Spectrophotofluorimeter was used to measure the change in fluorescence upon addition of KCl (red), NaCl (green), LiCl (pink), RbCl (orange), CsCl (black) and ChCl (blue). Initially 15 mM cation was added indicated by the first arrow followed by 1% addition of glucose indicated by the second arrow. The assay was performed with 1ml of cells at 20 °C using excitation wavelength 450 nm and emission was measured at 510 nm with a bandwidth 0.6 nm. The data represent the means of four replicates measurements.

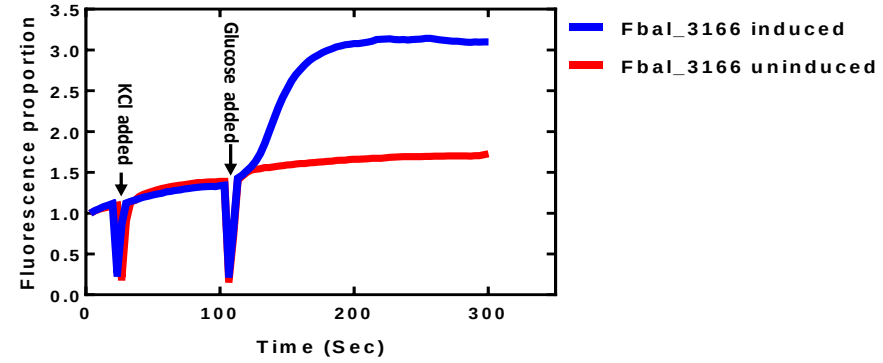
### 6.3.2 Confirmation of acriflavine efflux mediated by PACE pumps in the presence of potassium

In the previous section 6.3.1, it was observed that KCl was the most specific in releasing acriflavine from the cells. In order to validate further the efflux property of PSPTO_3587, Fbal_3166 and AceI (induced and uninduced), transport experiments were performed. As a control, induced and uninduced cells containing plasmid pTTQ18 with no gene insert were also tested for the ability to transport acriflavine (Figure 6.6). Using Tris buffer (20mM pH 7), these experiments were accomplished in the presence of 15mM KCl, followed by addition of 1% glucose as shown with an arrow indicated in Figure 6.6. The results obtained indicate that acriflavine efflux was significantly greater in all induced cells compared with their uninduced cells (Figure 6.6). These finding confirm that PSPTO_3587, Fbal_3166 and AceI are active efflux pumps and acriflavine is their substrate. There is some low level of efflux has been observed in all uninduced cells which might be due to leaky expression of these proteins. The empty plasmid pTTQ18 containing cells have also revealed a little efflux activity, which could be due to the presence of other efflux transporters encoded by *E. coli*.

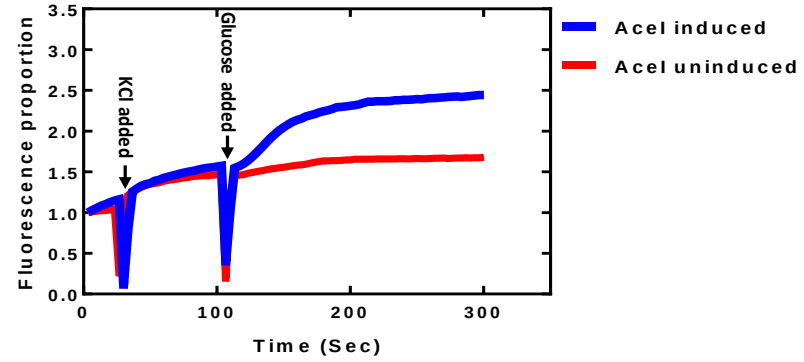
A


Fluorescence proportion
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
KCl added
Glucose added
0
100
200
300
Time (Sec)
PSPTO_3587 induced
PSPTO_3587 uninduced

B


Fluorescence proportion
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
KCl added
Glucose added
0
100
200
300
Time (Sec)
Fbal_3166 induced
Fbal_3166 uninduced

C


Fluorescence proportion
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
KCl added
Glucose added
0
100
200
300
Time (Sec)
AceI induced
AceI uninduced

D

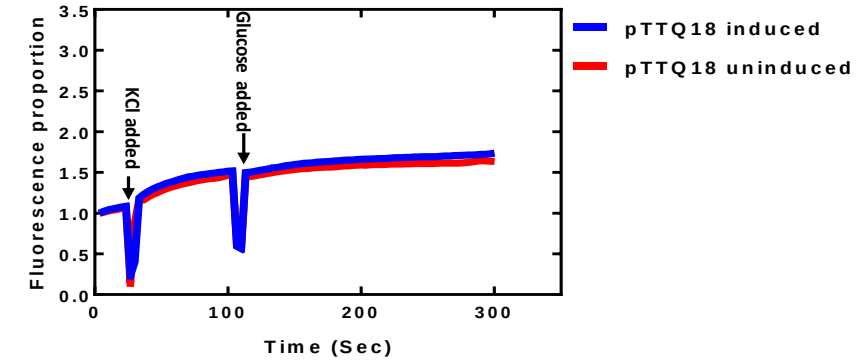

Fluorescence proportion
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
KCl added
Glucose added
0
100
200
300
Time (Sec)
pTTQ18 induced
pTTQ18 uninduced

**Figure 6.6 Measurement of acriflavine efflux from induced and uninduced cells.** Acriflavine loaded cells were prepared as described in section 2.2.5. *E. coli* cells carrying PSPTO_3587 (A); Fbal_3166 (B); AceI (C) and empty plasmid pTTQ18 (D) were induced (blue) and uninduced (red). Initially 15 mM KCl was added indicated by the first arrow followed by 1% addition of glucose indicated by the second arrow. The assay was performed with 1ml of cells at 20 °C using excitation wavelength 450 nm and emission was measured at 510 nm with a bandwidth 0.6 nm. The data represent the means of four replicates measurements.

### 6.3.3 Inhibition by chlorhexidine of acriflavine efflux

An active efflux of acriflavine by PSPTO_3587, Fbal_3166 and AceI was established as described above (Section 6.3.1 and 6.3.2). Further fluorimetric experiments were designed to investigate ligand specificity of these proteins. Acriflavine-chlorhexidine loaded and energy-starved cells were prepared (section 2.2.5) and efflux was recognised by measuring the fluorescence intensity. Various concentrations of chlorhexidine were used to determine their inhibitor role. Dose dependant inhibition of acriflavine efflux by chlorhexidine was observed (Figure 6.7). The results indicate that chlorhexidine is another possible substrate of these efflux transporters. The greatest competitive effect was produced by 20µM chlorhexidine with almost complete loss of fluorescence. These findings agree with the previous MIC reports of chlorhexidine transport by AceI (Hassan et al., 2013) and of chlorhexidine resistance by all three proteins (Hassan et al., 2015).

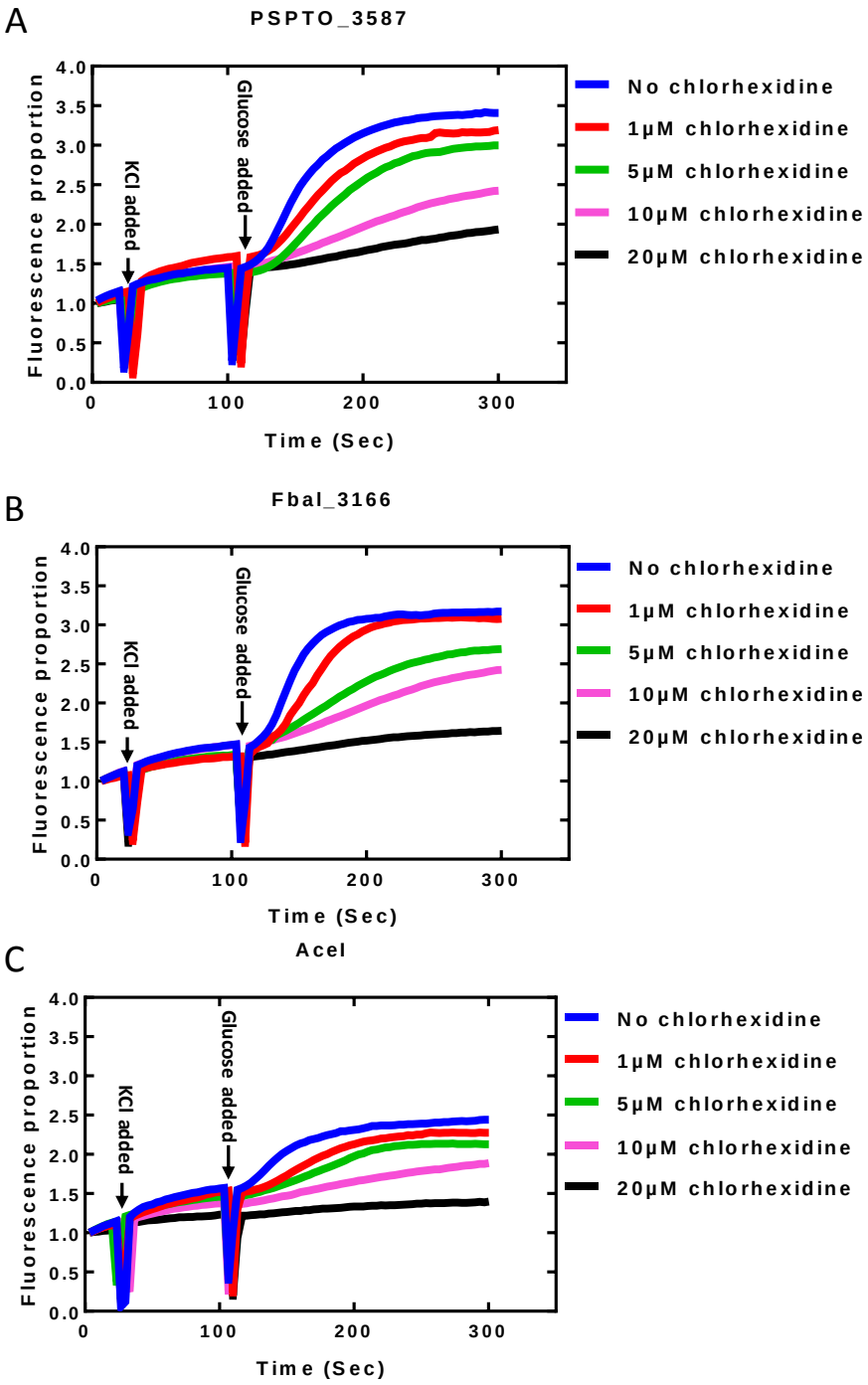



**Figure 6.7 Inhibition effect of chlorhexidine on acriflavine efflux by PSPTO_3587, Fbal_3166 and AceI.** Acriflavine-chlorhexidine loaded and energy-starved cells were prepared (section 2.2.5). *E. coli* cells carrying PSPTO_3587 (A); Fbal_3166 (B) and AceI (C). A dose-dependent effect on acriflavine efflux was observed in the presence of chlorhexidine. Cells loaded with no chlorhexidine (blue); 1µM chlorhexidine (red); 5µM chlorhexidine (green); 10µM chlorhexidine (pink); 20µM chlorhexidine (black). The data represent the means of four replicates measurements.

### 6.3.4 Inhibition of acriflavine efflux by spermidine

The PACE family proteins are multidrug efflux transporters (Hassan et al., 2015). As described above, acriflavine and chlorhexidine have been identified as substrates of some PACE proteins. Spermidine is a putative natural substrate of AceI since externally added spermidine leads to strong induction of AceI expression in *A. baumannii* (Hassan et al., unpublished data). It has been reported that a high concentration of spermidine is toxic to the cells (Gaboriau et al., 2004) and efflux pumps may be required to actively efflux from cells. Therefore, spermidine was also tested to determine its potential to inhibit acriflavine transport. *E. coli* cells expressing PSPTO_3587, Fbal_3166 and AceI were preloaded with various concentrations of spermidine in the presence of 10µM acriflavine and 10µM CCCP to de-energise the cells, as described in section 2.2.5. Dose dependent inhibition of acriflavine efflux by spermidine was investigated by measuring the fluorescence change (Figure 6.8). There was no significant difference in the rate of acriflavine efflux at all concentrations of spermidine tested. This could mean that spermidine is a very weak substrate. Alternatively, due to the hydrophilic nature of spermidine, it may not be able to move effectively into the cell in the absence of energy and therefore, may not be able to inhibit acriflavine binding and transport.

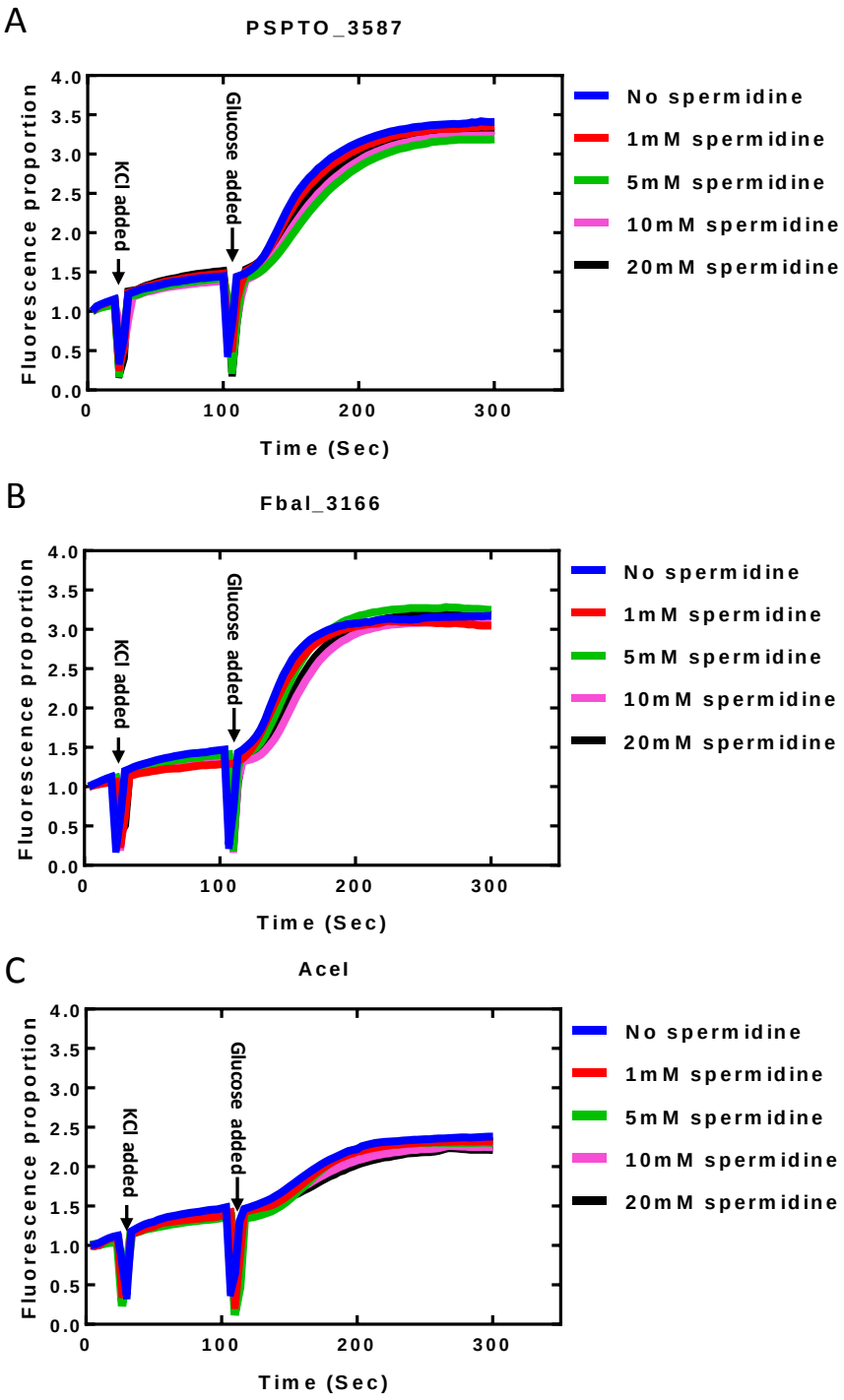



**Figure 6.8 Inhibition effect of spermidine on acriflavine efflux by PSPTO_3587, Fbal_3166 and AceI.** Acriflavine-spermidine loaded and energy-starved cells were prepared (section 2.2.5). *E. coli* cells carrying PSPTO_3587 (A); Fbal_3166 (B) and AceI (C). Dose-dependent effect on acriflavine efflux was tested in the presence of spermidine. Cells loaded with no spermidine (blue); 1mM spermidine (red); 5mM spermidine (green); 10mM spermidine (pink); 20mM spermidine (black). The data represent the means of four replicates measurements.

## 6.4 $^{3}$H-spermidine accumulation in energised *E. coli* cells expressing AceI, PSPTO_3587, STY_3166, Fbal_3166, PFL_4558 and Tmarg_opt

It has been recognised that PACE family proteins are active efflux transporters (Hassan et al., 2013; Hassan et al., 2015). In the previous experiments, no inhibition of acriflavine transport was observed by spermidine, possibly because spermidine is highly hydrophilic and unable to cross the lipid bilayer membrane effectively in the presence of CCCP, which de-energised the cells. Therefore to test whether these proteins could mediate spermidine efflux in *E. coli* a different approach was employed using radio-labelled $^{3}$H-spermidine accumolation measurements. In the literature, various methodologies are described to measure the efflux pumps activity in bacterial cells (Blair and Piddock, 2016). In this assay, a substrate ($^{3}$H-spermidine) accumulated inside the cells was measured and a lower level of substrate was used to infer more efflux activity. Cells were cultured and prepared as described in section 2.6. Experiments were performed to measure the uptake of $^{3}$H-labelled spermidine compound into energised whole cells. *E. coli* cells containing the empty plasmid pTTQ18, AceI and six AceI homologues (STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587) were tested for transport of $^{3}$H-spermidine at a final concentration of 20 mM are shown in Figure 6.9. The uptake of $^{3}$H-spermidine was similar in all uninduced cells and cells containing only plasmid pTTQ18 (both induced and uninduced), showing the baseline accumulation level under the conditions used (Figure 6.9). Compared to this baseline, cells expressing AceI, Fbal_3166, PFL_4558, Tmarg_opt and PSPTO_3587 showed lower accumulation of $^{3}$H-spermidine (Figure 6.9), which is interpreted as being related to their active efflux activity. In contrast to cells expressing these proteins, cells expressing STY_3166 and particularly A1S_1503 accumulate more spermidine in the induced cells than the uninduced controls (Figure 6.9 G, H). The higher level of uptake in the induced cells expressing A1S_1503 and STY_3166 proteins is intriguing. Efflux function by these transporters has not been demonstrated so far and their role in the cells is unknown. These results indicate that spermidine is another substrate of some PACE family proteins.

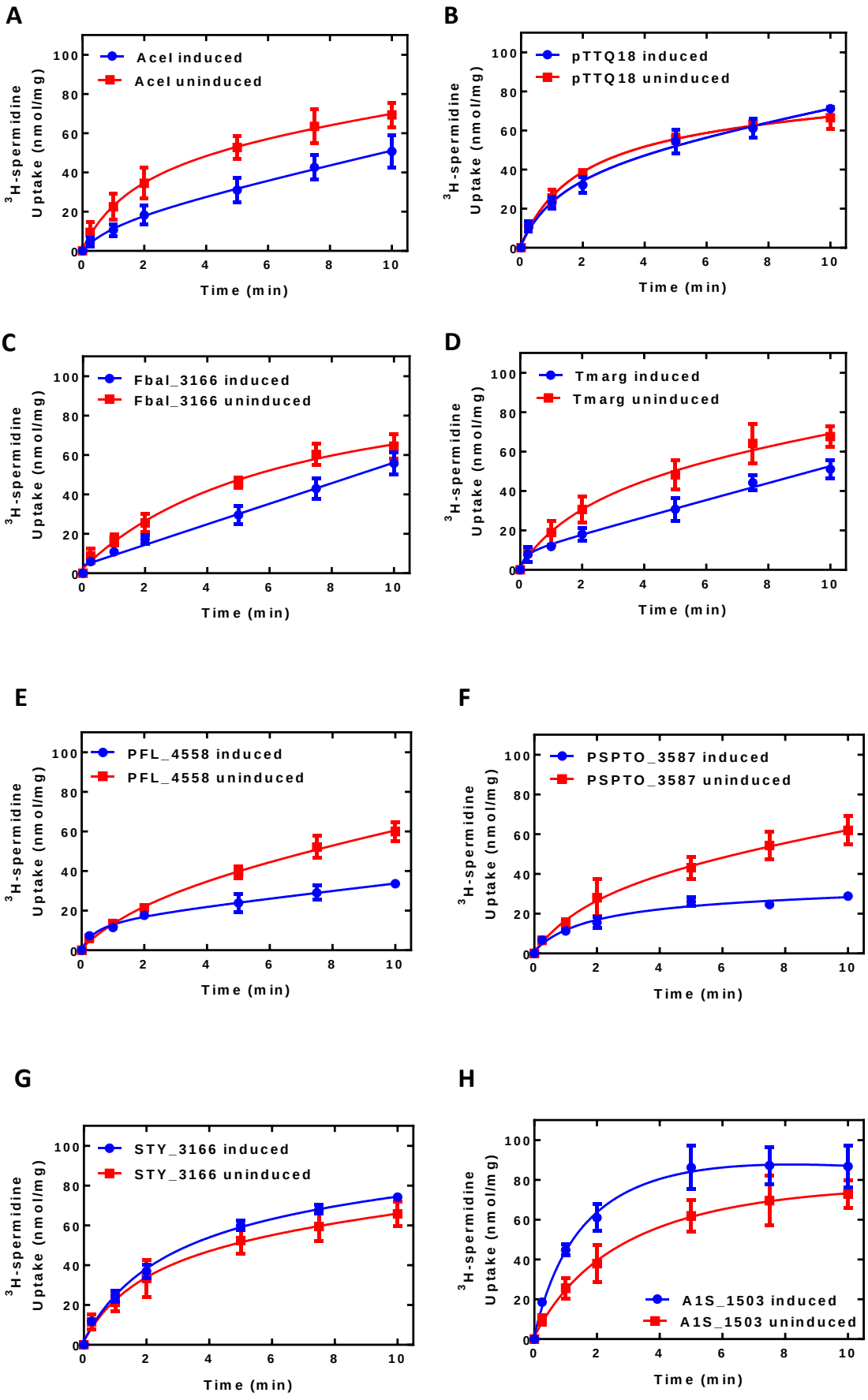

A
AceI induced
AceI uninduced
B
pTTQ18 induced
pTTQ18 uninduced
C
Fbal_3166 induced
Fbal_3166 uninduced
D
Tmarg induced
Tmarg uninduced
E
PFL_4558 induced
PFL_4558 uninduced
F
PSPTO_3587 induced
PSPTO_3587 uninduced
G
STY_3166 induced
STY_3166 uninduced
H
A1S_1503 induced
A1S_1503 uninduced
$^3$H-spermidine Uptake (nmol/mg)
Time (min)

**Figure 6.9 Measurements of $^{3}$H-spermidine accumulation in energised whole cells.** Uptake of $^{3}$H-spermidine into energised whole cells containing the constructs for expressing AceI (A) or containing the empty plasmid pTTQ18 (B), Fbal_3166 (C), Tmarg_opt (D), PFL_4558 (E), PSPTO_3587 (F), STY_3166 (G), and A1S_1503 (H). *E. coli* cells carrying the pTTQ18-based clone were grown, induced and prepared as described in section 2.6. Two batches of *E. coli* cells were grown to $OD_{600}$=0.8 in LB medium containing carbenicillin (100 ug/ml). One batch was induced with 0.2 mM IPTG (blue) while the other was left uninduced (red) and incubation was continued for one hour further. Cells were harvested at 4500 x g for 10 minutes at room temperature and were washed three times in buffer (140 mM NaCl; 10 mM KCl; 20 mM MOPS (3-[N-Morpholino] propane sulphonic acid) pH 6.6. The washed cells were resuspended in the same buffer and cell density was adjusted to $OD_{680}$ = 2. Glucose was added as energy source and the assay was initiated by the addition of 20 mM $^{3}$H-spermidine under constant bubbled aeration. Samples were taken at 0.25, 1, 2, 5, 7.5 and 10 minutes. Liquid scintillation counting was used to measure the retained radioactivity in nmol/mg cells.

## 6.5 Larger-scale cultures

On the basis of expression levels shown in small-scale cultures (Figure 6.2), seven proteins (AceI, Fbal_3166, PFL_4558, PSPTO_3587, A1S_1503, STY_3166 and Tmarg_opt) were recommended for larger-scale cultures. Out of seven, four proteins AceI, Fbal_3166, STY_3166 and Tmarg_opt, which represent four phylogenetic groups of PACE proteins (Figure 6.10; Hassan et al., 2015), were chosen for purification and further characterisation.

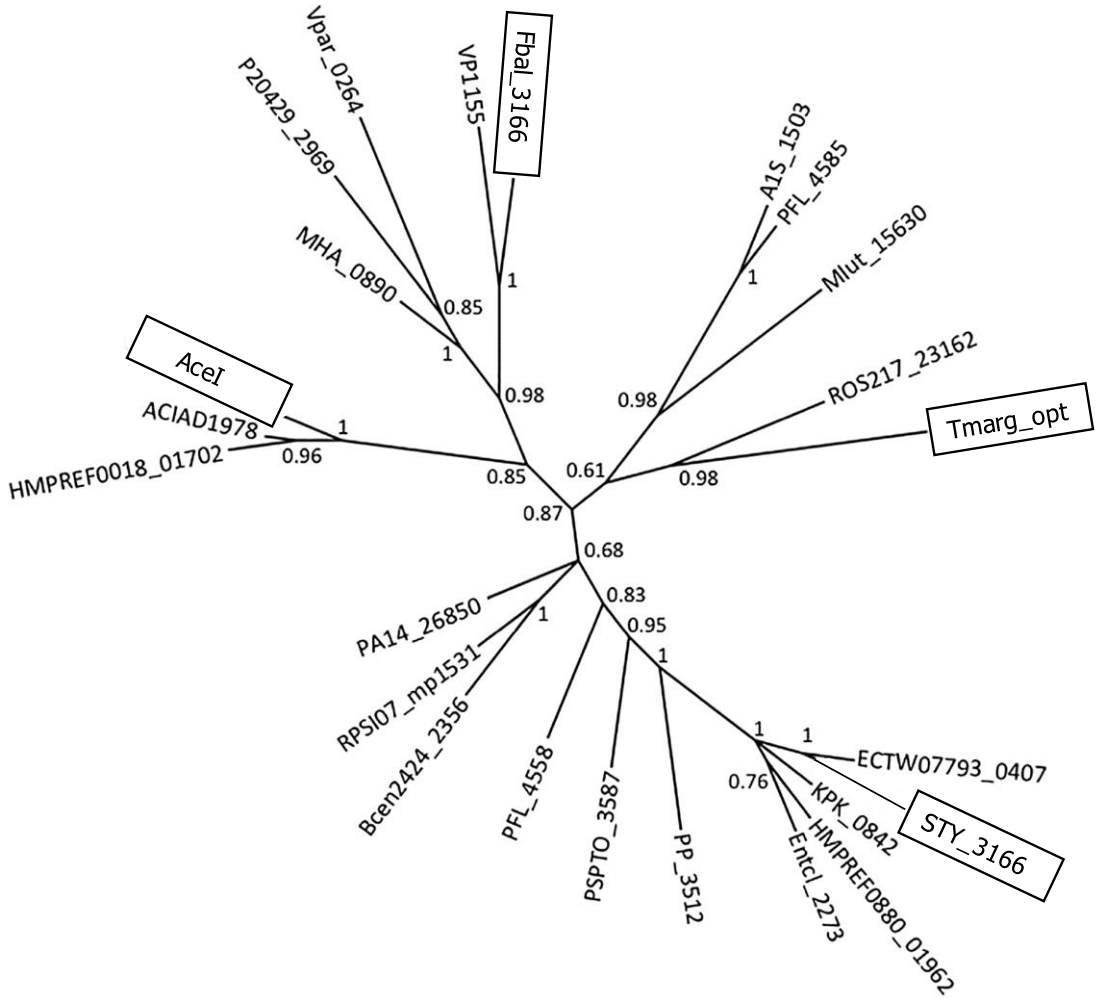



**Figure 6.10 Phylogenetic tree showing the relationships of PACE family proteins.** The encircled proteins in the tree were chosen for large scale cultures, purification and characterisation. This diagram was adapted from Hassan et al., (2015).

### 6.5.1 Predicted topology of AceI, Fbal_3166, STY_3166 and Tmarg_opt proteins

Membrane proteins represent about 20-30% of all the genes in most sequenced genome (Liszewski, 2015). However, due to the experimental difficulties, the number of high-resolution structures is relatively small compared with soluble proteins (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). To study a membrane protein in the absence of high-resolution structure, various computational tools are used to predict their structural and functional features based on amino acid sequence and composition. Using the topology prediction tool TMHMM (Krogh et al., 2001) showed that the proteins (AceI, Fbal_3166, STY_3166 and Tmarg_opt) consist of four putative transmembrane spanning α-helices with both N- and C-terminal ends at the cytoplasmic side of the membrane. Highly conserved residues play an important role in the structural and functional features of proteins (Sitborn and Pietrokovski, 2007). In order to determine the highly conserved residues among the twenty four PACE family proteins (Table 6.1), it was essential to compare the sequences of proteins. Multiple sequence alignment tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to identify conserved residues (Appendix 5), which are highlighted in the model of AceI, STY_3166, Fbal_3166 and Tmarg_opt (Figures 6.11 and 6.12) are positioned in all helices and the C-terminal cytoplasmic tail following TM4. A highly conserved motif WN in TM2 of these proteins may be more important, involved in substrate binding or transport. Creating mutants of this motif would be useful to investigate the roles of these residues.

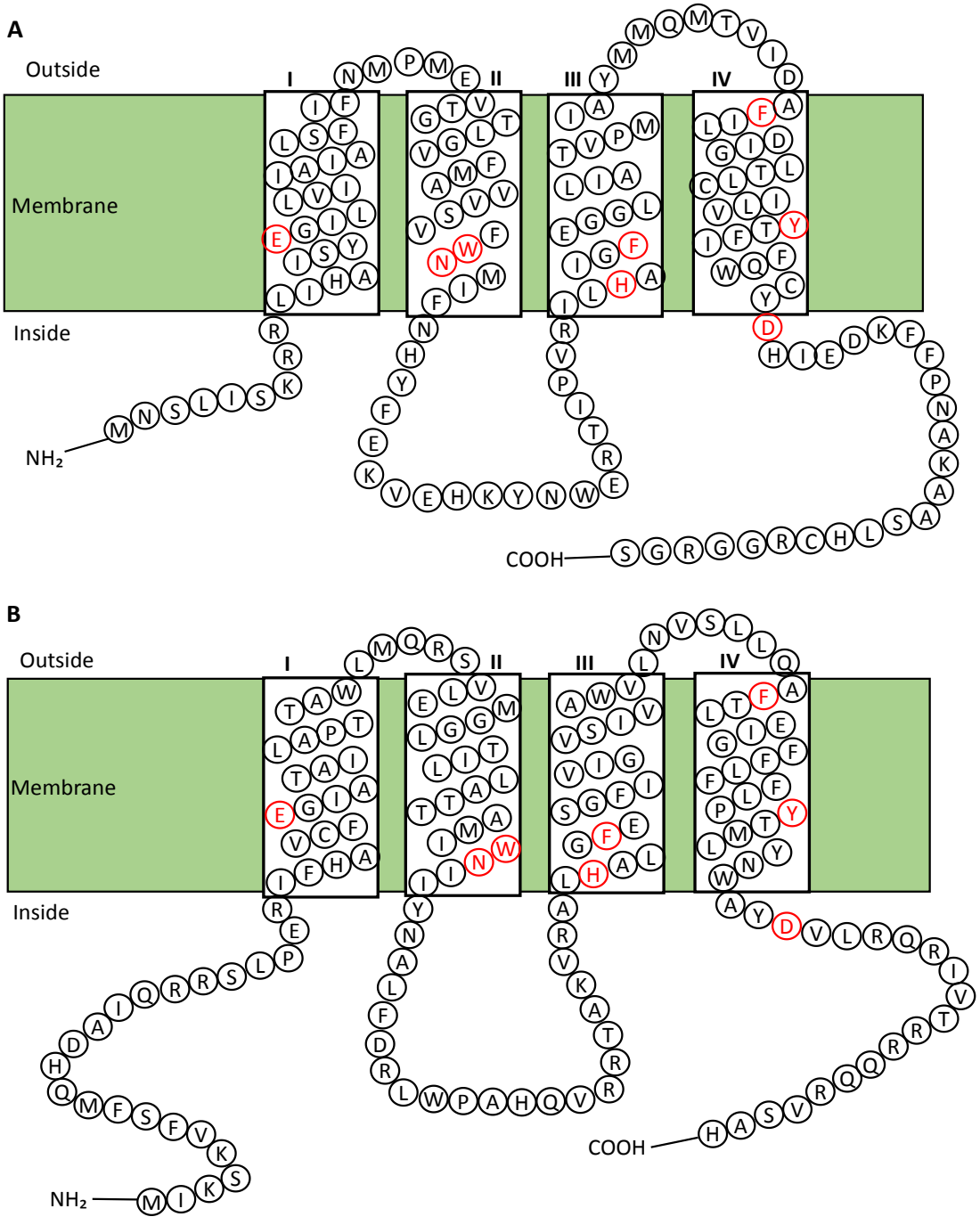



**Figure 6.11 Predicted membrane topology model of AceI and STY_3166.** Diagram for the putative membrane topology of the AceI (A) and STY_3166 (B) proteins with four transmembrane spanning α-helices. Both the N- and C-terminal ends are found on the cytosolic face of the membrane. The positions of transmembrane helices are based on the result of the TMHMM prediction tool and are shown in roman numerals from N- to C-terminus. Highly conserved residues are highlighted in red according to the alignment in Appendix 5.

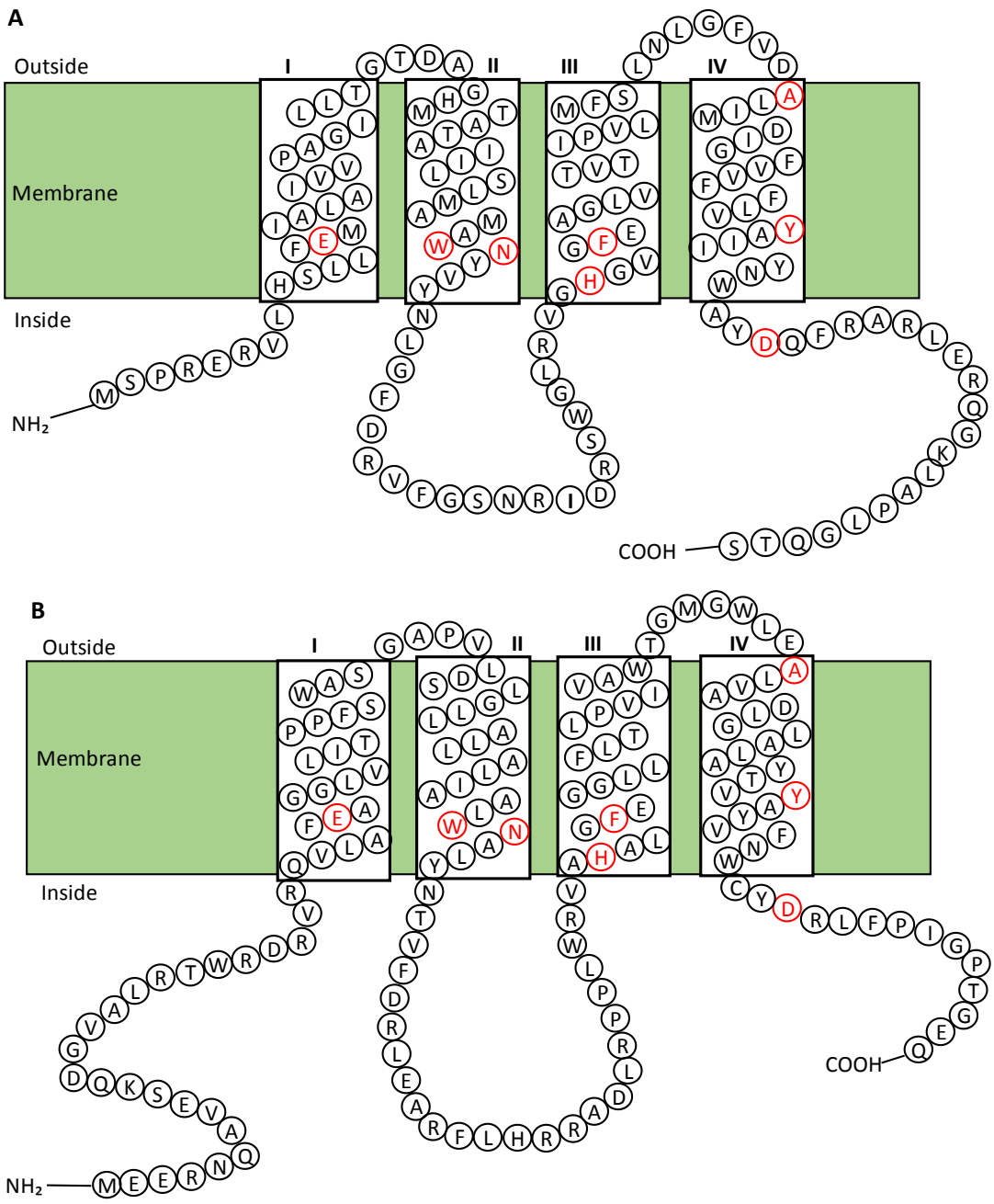



**Figure 6.12 Predicted membrane topology model of Fbal_3166 and Tmarg_opt.** Diagrams for the putative membrane topology of the Fbal_3166 (A) and Tmarge_opt (B) proteins with four transmembrane spanning α-helices. Both the N- and C-terminal ends are found on the cytosolic face of the membrane. The positions of transmembrane helices are based on the result of the TMHMM prediction tool and are shown in roman numerals from N- to C-terminus. Highly conserved residues are highlighted in red according to the alignment in Appendix 5.

### 6.5.2 Purification of AceI, Fbal_3166, STY_3166 and Tmarg_opt

Purification of the AceI and homologs proteins (Fbal_3166, STY_3166 and Tmarg_opt) was an important step to allow biochemical and biophysical measurements to be performed. Inner membranes were prepared by sucrose density gradient ultracentrifugation (section 2.4.2) from the *E. coli* BL21(DE3) cells expressing the AceI, Fbal_3166, STY_3166 and Tmarg_opt. Inner membranes were solubilised using 1% DDM containing solubilisation buffer. The same methodology (immobilised metal affinity chromatography) was applied for purification of all the proteins by exploiting the engineered C-terminal $His_6$-tag (Section 2.4.3). SDS-PAGE and Western blot analyses of the various fractions of purifications for AceI and three homologues are shown in Figure 6.13. For each protein lane 2 contains the inner membranes solubilised with 1% DDM and lane 3 contains the supernatant obtained by ultracentrifugation of the solubilised inner membranes. Lane 4 shows the pellet obtained following solubilisation and ultracentrifugation (insoluble fraction). Lane 5 shows protein that did not bind to the Ni-NTA column (unbound fraction) and the blots suggest that most of the required proteins have been bound to the resin. Following removal of further unbound material by washing with buffer containing 40 mM imidazole, purified protein was eluted by washing with buffer containing a high concentration of imidazole (200 mM) that dissociates the $His_6$-tagged proteins because they can no longer compete for binding sites on the Ni-NTA resin (Lane 6).

Under conditions of the SDS-PAGE separation, the AceI and Fbal_3166 purified proteins are largely present in their monomeric forms at a position of around 16 kDa whereas the STY_3166 and Tmarg_opt proteins are present at a position of around 18 kDa. The higher band(s) most likely representing a dimer or trimer in each protein. Comparisons of AceI and homologous proteins from 30 litre cultures are detailed in Table-6.2. Based on densitometric analysis, AceI was obtained with highest purity of 86% whereas Fbal_3166, STY_3166 and Tmarg_opt were 84%, 80% and 78%, respectively (Figure 6.13). The highest yield was obtained for STY_3166 was 1.3 mg/litre whereas AceI, Fbal_3166 and Tmarg_opt was 1.1, 1.1 and 1.0 mg/litre from the cell cultures, respectively.

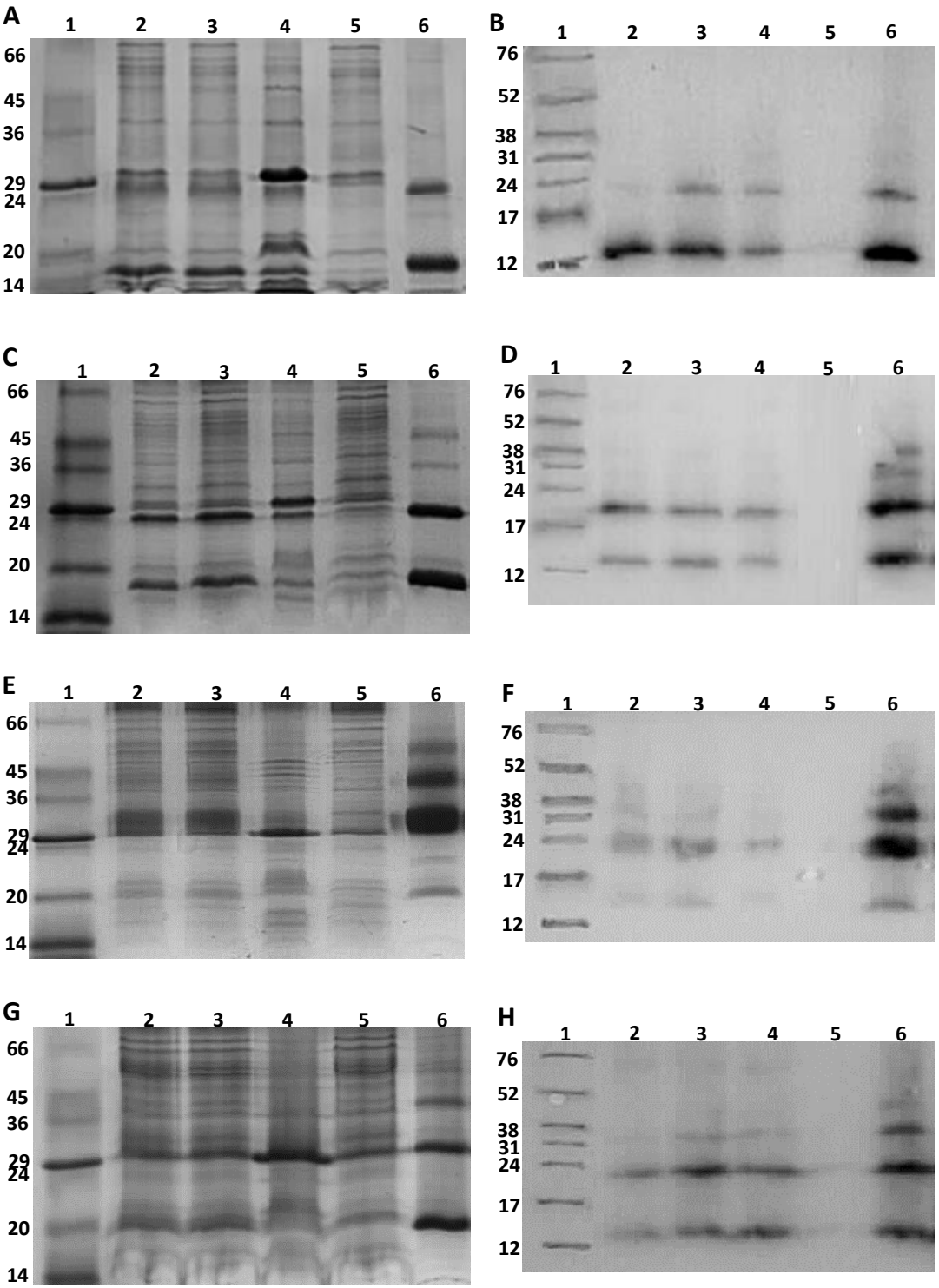

A
1 2 3 4 5 6
66
45
36
29
24
20
14
B
1 2 3 4 5 6
76
52
38
31
24
17
12
C
1 2 3 4 5 6
66
45
36
29
24
20
14
D
1 2 3 4 5 6
76
52
38
31
24
17
12
E
1 2 3 4 5 6
66
45
36
29
24
20
14
F
1 2 3 4 5 6
76
52
38
31
24
17
12
G
1 2 3 4 5 6
66
45
36
29
24
20
14
H
1 2 3 4 5 6
76
52
38
31
24
17
12

**Figure 6.13 Purification of proteins from inner membranes**. Inner membranes of *E. coli* BL21(DE3) expressing the AceI (A, B); Fbal_3166 (C, D); STY_3166 (E, F); and Tmarg_opt (G, H); were prepared using sucrose density gradient procedure (section 2.4.2). Membrane preparations were solubilised using solubilisation buffer containing 1% DDM (section 2.4.3). Different fractions from the purification were subjected to analysis on a 15% SDS-PAGE gel (left) and Western blot (right) as described in section 2.4.5. The samples were loaded on the gel as follows: (1) molecular weight marker (kDa); (2) inner membranes; (3) supernatant; (4) membrane pellet; (5) unbound flow through; (6) purified protein. All samples for the gels contained 16 µg protein and all samples for the Western blots contained 4 µg protein.

**Table 6.2 Purity and yield of proteins from 30 litre fermenter cultures.**

| Protein name | Wet cell pellets from 30L (g) | Inner membrane obtained (ml) | Purification yield (mg per litre) | Purification purity (%) |
|---|---|---|---|---|
| **AceI** | 129.37 | 14.3 | 1.1 | 86 |
| **Fbal_3166** | 112.69 | 12.9 | 1.1 | 84 |
| **STY_3166** | 116.9 | 16.5 | 1.3 | 80 |
| **Tmarg_opt** | 120.2 | 14.1 | 1.0 | 78 |

### 6.5.3 Analysing the secondary structure and thermal stability of AceI and its homologs Fbal_3166, STY_3166 and Tmarg_opt using CD

AceI and homologs (Fbal_3166, STY_3166 and Tmarg_opt) were purified (section 2.4.3) and the secondary structure was analysed by far UV CD spectroscopy according to section 2.5.1. The resultant CD spectra (Figure 6.14) of all the tested proteins (AceI, Fbal_3166, STY_3166 and Tmarg_opt) confirm that they are predominantly alpha-helical since their CD spectra showed the characteristic positive peak at ~192 nm and negative peaks at 209 nm and 222 nm (Wallace et al., 2003; Kelly et al., 2005; Miles and Wallace, 2016). There was some small variability in the CD signal intensities probably due to unintentional differences in protein concentration between the samples. These measurements also suggest that all the purified proteins were correctly folded and had retained their secondary structure architecture after passing through various steps of the purification.

To determine the thermal stability of AceI, Fbal_3166, STY_3166 and Tmarg_opt proteins, far UV CD spectroscopy was used to measure the loss of alpha-helical structure at increasing temperature (section 2.5.1). The purified proteins were heated at 5ºC increments by ramping the temperature from 5-90ºC and finally back to 5ºC (Figure 6.15). With increasing temperature, the signal at 209 nm began to change significantly when the proteins begin to melt, indicating their thermal stability. The thermal stabilities of all the proteins were measured by using Global Analysis CD software-3. For Fbal_3166 and STY_3166 from 5 ºC to 30 ºC there was a slow increase in the ellipticity value indicating little change of secondary structure. From 30 ºC to 60 ºC there was a greater increase in the ellipticity value indicating a significant change of the secondary structure. From 60 ºC to 90 ºC the increase in ellipticity was slow again, indicating that the change in secondary structure of the purified proteins was complete. In contrast, the AceI and Tmarg_opt were slowly denaturing as compared to Fbal_3166 and STY_3166. The melting temperatures of AceI, Tmarg_opt, Fbal_3166 and STY_3166 were 46.7 ºC, 37.6 ºC, 34.2 ºC and 32.6 ºC, respectively. These results suggest that AceI is the most stable protein as compared to the others and might be best for structure determination. The thermal denaturation in most cases is irreversible aggregation (Miles and Wallace, 2016), and the final scans at 5 ºC show that the purified proteins had loss their ability of refolding into their native secondary structure. These results are similar with reports on a wider range of

transport proteins (Bettaney, 2008; Ma, 2010; Sukumar, 2012; Jackson, 2012). The melting temperatures suggest that with these proteins performing experiments up to a temperature of 25 ºC should have no significant effects on the secondary structure stability of the protein.

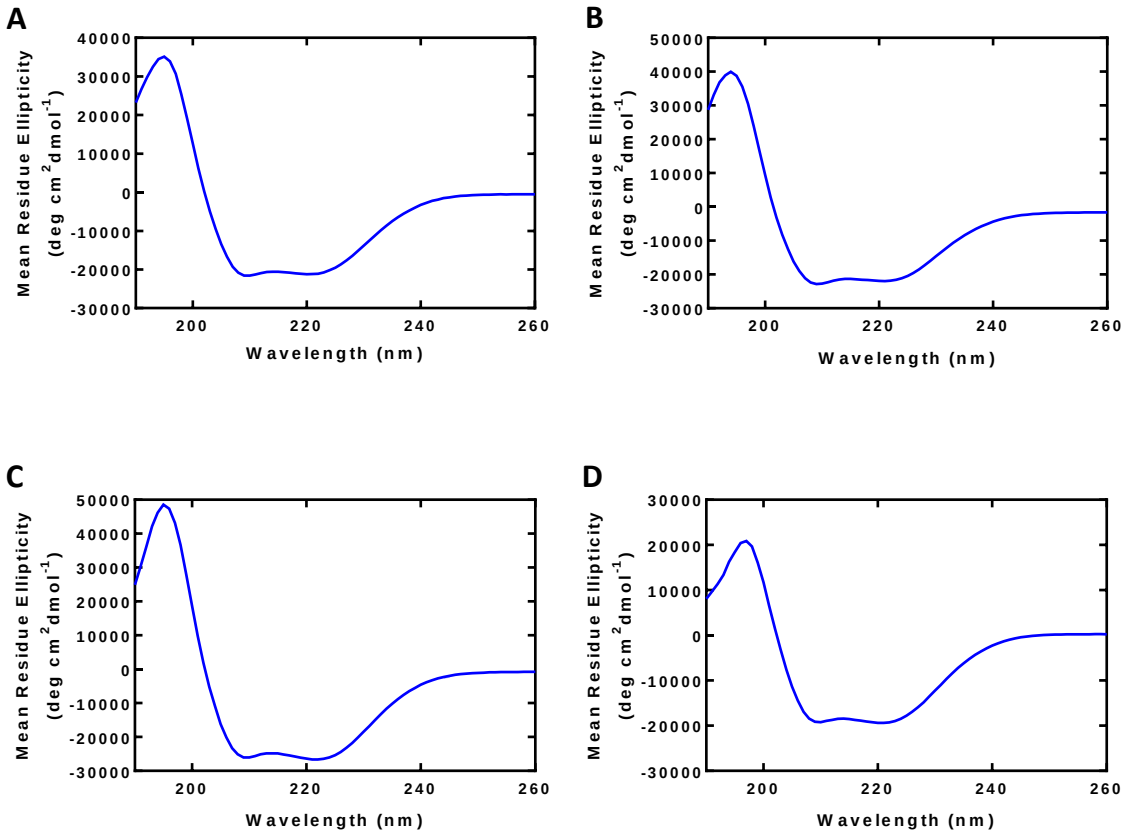



**Figure 6.14 Secondary structure integrity spectra of purified proteins.** Far-UV (180-260 nm) CD spectroscopy was performed for purified proteins (A) AceI; (B) Fbal_3166; (C) STY_3166 and (D) Tmarg_opt. Measurements were tkaen using a CHIRASCAN instrument (Applied Photophysics, UK) at 20 ºC with constant liquid nitrogen flushing. Samples were prepared as described in methods section 2.5.1 and analysed in a Hellma quartz cuvette of 1.0 mm pathlength at a final protein concentration of 0.15 mg/ml in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM).

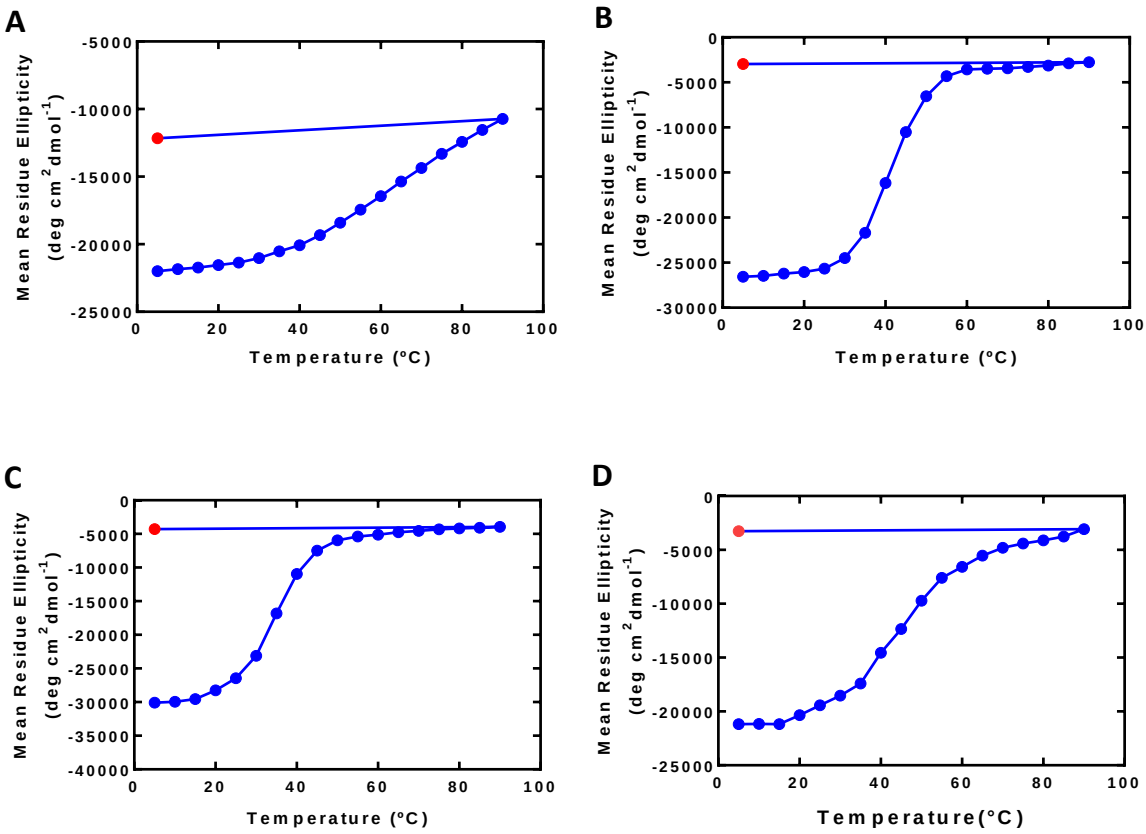



**Figure 6.15 Thermal stability analysis of purified proteins using far UV CD.** Far-UV (180-260 nm) CD spectroscopy was performed for purified proteins (A) AceI; (B) Fbal_3166; (C) STY_3166 and (D) Tmarg_opt. Measurements were taken using a CHIRASCAN instrument (Applied Photophysics, UK) with constant liquid nitrogen flushing. Samples were prepared as described in methods section 2.5.1 and analysed in a Hellma quartz cuvette of 1.0 mm pathlength at a final protein concentration of 0.15 mg/ml in CD buffer (10 mM NaPi pH 7.5; 0.05% DDM). The change in the CD signal at 209 nm show thermal stability of the proteins. The samples were heated at 5ºC increments by ramping the temperature from 5-90ºC and the final spectrum was taken upon cooling of the sample back to 5ºC (red).

### 6.5.4 Ligand binding activity of AceI, Fbal_3166, STY_3166 and Tmarg_opt purified proteins

Steady-state spectrophotofluorimetry experiments were performed (section 2.5.2), to test binding of chlorhexidine and of the putative natural substrates spermidine, cadaverine and putrescine to purified AceI protein and its homologs Fbal_3166, STY_3166, and Tmarg_opt. (Figure 6.16). The binding affinity of each protein for these compounds was interacted by monitoring the intrinsic fluorescence properties of aromatic tryptophan residues as a function of ligand concentration. The resultant binding curves were used to obtain values for the apparent dissociation constant ($K_d$) and the maximum quenching of fluorescence intensity ($\Delta F_{max}$) (Table 6.3). Of the ligands tested, chlorhexidine induced the highest fluorescence quench for each protein (Figure 6.16). The apparent $K_d$ values obtained for binding of chlorhexidine to AceI and its homologs Fbal_3166, STY_3166, and Tmarg_opt were 2.8±1.7 µM, 2.4±0.7 µM, 11.8±7.1 µM and 4.5±3.8 µM, respectively, and maximal fluorescence quench values of 10.1±2.4%, 21.7±2.4%, 17.7±7.6% and 10.5±4.1%, respectively. The $K_d$ values suggest that binding affinity of the Fbal_3166 and AceI proteins for chlorhexidine is higher than STY_3166 and Tmarg_opt proteins. These results agree with earlier MIC reports where Fbal_3166 and AceI showed chlorhexidine resistance, but STY_3166 and Tmarg_opt did not facilitate detectable resistance levels (Hassan et al., 2013; Hassan et al., 2015). AceI and its homologs Fbal_3166, STY_3166, and Tmarg_opt showed no significant difference in fluorescence quench upon the addition of spermidine, cadaverine and putrescine over the concentration range of 0-40µM (Figure 6.16). A low percentage quenching for spermidine, cadaverine and putrescine suggest very weak or no binding affinity of these compounds.

As a control the background fluorescence quenching of ligands was determined by titrating the compounds at the same concentration against free tryptophan (Figure 6.16F). The cloned AceI protein has three tryptophan residues and the protein concentration of AceI was 4.4 µM in this experiment. Therefore, free tryptophan of 13.2 µM was used to generate equal values of fluorescence in the fluorimetry buffer. It was revealed that each of the polyamines has a small quenching effect on free tryptophan and chlorhexidine has comparatively more quenching than these compounds. However, the fluorescence quench with chlorhexidine of AceI and its

homologs was significantly higher allowing valid measurements of binding affinity above the background. Further to confirm chlorhexidine nonspecific binding, an experiment was performed with the NCS1 family protein Mhp1, which showed an almost linear relationship (Figure 6.16E), but with a higher level of non-specific quenching than seen for free tryptophan. In contrast to Mhp1, the AceI, Fbal_3166, STY_3166 and Tmarg_opt proteins show non-linear quenching by chlorhexidine, indicative of a binding interaction with chlorhexidine (Figure 6.16).

The higher level of background fluorescence quenching from chlorhexidine in the Mhp1 and Tryptophan titrations may be due to its absorbing light at the excitation wavelength 295 nm or emission wavelength 330 nm. Therefore, an additional experiment was performed to reveal the chlorhexidine absorbance spectrum in the range of excitation (295nm) and emission wavelength (310-360nm) using an absorbance spectrometer. It was revealed that chlorhexidine did absorb light across these wavelengths, and was higher at the excitation wavelength (295nm) compared to the emission region (330nm) (Figure 6.17). Therefore, the background fluorescence quenching from chlorhexidine is likely to be due to it absorbing some of the excitation light.

Overall, these steady-state fluorimetry measurements show that chlorhexidine is a preferred substrate for PACE family proteins, in contrast to spermidine, cadaverine and putrescine, which appear to bind very weakly or not at all. Chlorhexidine is a synthetic compound commercially available from the last few decades and this suggests that its binding and transport by AceI is fortuitous, rather than by the action of a selection pressure. However, this direct interaction between AceI and chlorhexidine recognises the earlier reports that AceI is an active chlorhexidine efflux protein (Hassan et al., 2013; Hassan et al., 2015).

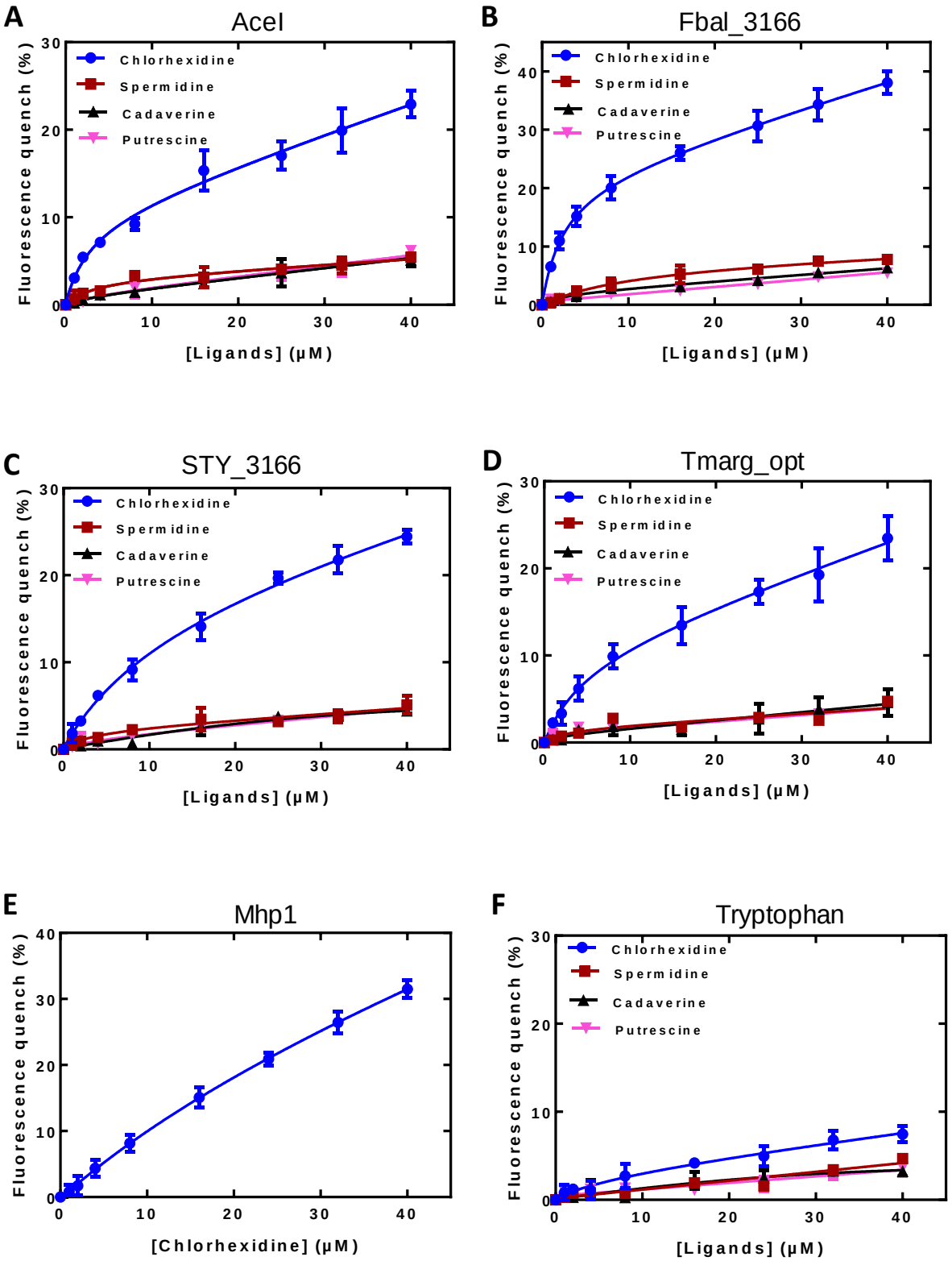

A
AceI
B
Fbal_3166
C
STY_3166
D
Tmarg_opt
E
Mhp1
F
Tryptophan
Chlorhexidine
Spermidine
Cadaverine
Putrescine
Fluorescence quench (%)
[Ligands] (µM)
[Chlorhexidine] (µM)

**Figure 6.16 Ligand-binding affinity curves for chlorhexidine, spermidine, cadaverine and putrescine binding to purified proteins.** Steady-state fluorimetry measurements on purified proteins were performed using a Photon Technology International spectrofluorimeter (section 2.5.2). Samples containing protein (4.4 µM) in fluorescence buffer (Table 2.21) at 18 ºC were excited at 295 nm and fluorescence emission was measured at 330 nm. Following ligand additions, samples were stirred for 1.5 minutes before making the measurements. Ligands were represented as chlorhexidine (blue), spermidine (red), cadaverine (black) and putrescine (pink). Data were analysed using a nonlinear regression equation (Y=ΔFmax*X/(Kd+X) + NS*X + Background) on GraphPad Prism 7 software.

**Table 6.3 Comparison of $K_d$ and $\Delta F_{max}$ values for chlorhexidine, spermidine, cadaverine and putrescine interacting with the AceI, Fbal_3166, STY_3166 and Tmarg_opt using spectrophotofluorimetry.** ND indicates not determined.

| | **Chlorhexidine** | **Spermidine** | **Cadaverine** | **Putrescine** |
|---|---|---|---|---|
| | **$K_d$ (µM) $\Delta F_{max}$ (%)** | **$K_d$ (µM) $\Delta F_{max}$ (%)** | **$K_d$ (µM) $\Delta F_{max}$ (%)** | **$K_d$ (µM) $\Delta F_{max}$ (%)** |
| **AceI** | 2.8±1.7 10.1±2.4 | 3.2±3.3 3.1±1.3 | 4.0±10.7 1.1±1.2 | 3.5±1.2 1.1±9.0 |
| **Fbal_3166** | 2.4±0.7 21.7±2.4 | 9.4±7.8 7.1±3.7 | 2.0±2.3 2.0±0.8 | ND |
| **STY_3166** | 11.8±7.1 17.7±7.6 | 3.1±4.4 2.3±1.2 | ND | 4.2±7.8 1.1±0.9 |
| **Tmarg_opt** | 4.5±3.8 10.5±4.1 | 2.1±3.9 1.6±1.1 | 1.1±5.3 0.7±1.1 | 0.5±1.2 1.0±0.5 |

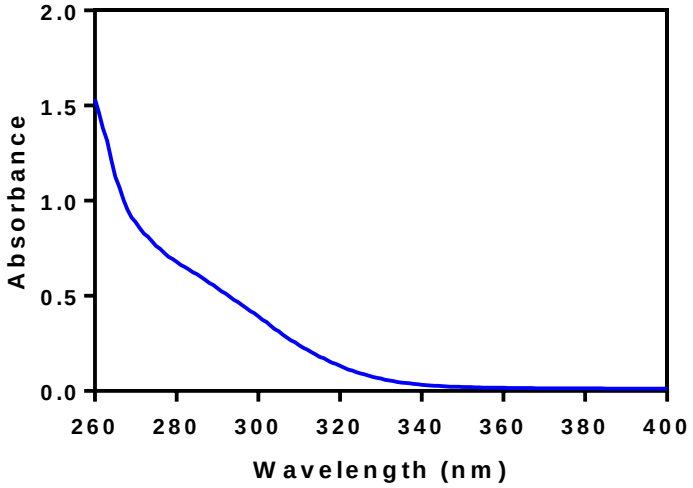



**Figure 6.17 Absorbance spectrum of chlorhexidine.** Samples containing chlorhexidine (20µM) were analysed using a UV-Vis DS-11 spectrophotometer.

The binding curves do not approach saturation level at 40 µM in the case of spermidine, cadaverine and putrescine titrations, which might be due to their weak affinity for AceI and homologs. To achieve the saturation curve and determine the concentration dependent binding of these compounds to the AceI protein, experiments were repeated over a higher concentration range 0-5 mM (Figure 6.18). Again the titration curves did not approach saturation even using 5 mM compound. Control experiments were also performed to evaluate the effect of higher concentration of these ligands on free tryptophan. Almost similar levels of quenching on free tryptophan were observed suggesting that spermidine, cadaverine and putrescine have a very weak or no binding affinity to the AceI protein. Alternatively, these compounds may bind in a different region of the PACE proteins and not induce a change in the environment of tryptophan residues and hence no significant fluorescence change.

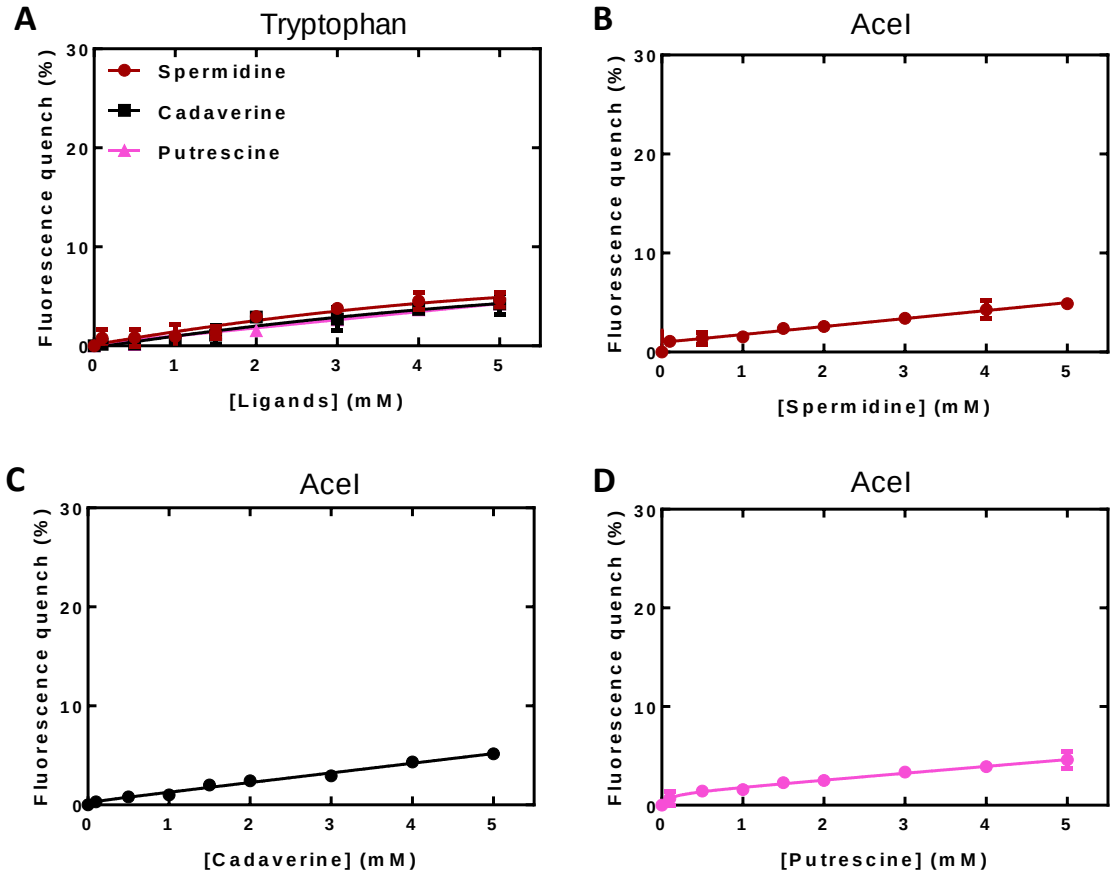



**Figure 6.18 Binding affinity of AceI protein for various polyamine compounds.** Steady-state spectrophotofluorimetry measurements on purified AceI protein were performed using a Photon Technology International spectrofluorimeter. Samples containing free tryptophan (13.2 µM) (A); or AceI protein (4.4 µM) (B, C, and D) in fluorescence buffer (Table 2.21) at 18 ºC were excited at 295 nm and fluorescence emission was measured at 330 nm. Following ligand additions, samples were stirred for 1.5 minutes before making the measurements. Ligands were represented as spermidine (red), cadaverine (black) and putrescine (pink). Data were analysed using a nonlinear regression equation (Y=ΔFmax*X/(Kd+X) + NS*X + Background) on GraphPad Prism 7 software.

## 6.6 Conclusions

A strategy has been demonstrated for the expression screening, purification and characterization of PACE family transport proteins preliminary to understanding structural biology. Starting with twenty-four proteins, small-scale expression screening identified seven (AceI, STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587) that were suitable for scaling up to larger cultures, inner membrane preparations and purifications.

The best expressed proteins, AceI, STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587, were tested for acriflavine efflux after expression in *E. coli* cells. The rate of acriflavine efflux was observed in induced cells expressing PSPTO_3587, Fbal-3166 and AceI by measuring the change in fluorescence intensity over time after energisation. The rate of change in the AceI expressing cells was lower as compared to PSPTO_3587 and Fbal_3166, suggested that acriflavine is a weak substrate for AceI. The uninduced samples of these three strains showed some acriflavine efflux actively above negative control pTTQ18 which might be due to leaky expression of PSPTO_3587, Fbal_3166 and AceI. The level of fluorescence in the case of PFL_4558, A1S_1503, STY_3166 and Tmarg_opt from induced and uninduced cells was almost the same as that of the negative control pTTQ18 cells suggesting that acriflavine is not a substrate for these proteins. It was demonstrated that the efflux mediated by PSPTO_3587, Fbal_3166 and AceI is energy dependant, but the precise mode of energisation is still unknown. Further experiments were performed to determine ion dependency of these proteins. A significantly higher change in fluorescence intensity was obtained with KCl than with all of the other cations (ChCl, LiCl, NaCl, RbCl & CsCl) used, suggesting that rate of efflux may be higher with potassium in the assay buffer. However, the rate of fluorescence change observed during acriflavine transport assays in the cells containing empty plasmid pTTQ18 was also higher in the presence of potassium. Therefore, the addition of potassium may have a general effect on the rate of acriflavine efflux from whole cells, possibly decreased affinity of acriflavine for nucleic acids.

As described above the PACE family proteins are most likely to be multidrug efflux transporters. In order to investigate ligand specificity of PSPTO_3587, Fbal_3166 and AceI, which can transport acriflavine, chlorhexidine and spermidine were used at various concentrations to determine their possible competitive inhibition of

acriflavine transport. Dose dependant inhibition of acriflavine efflux by chlorhexidine was observed for all three proteins, whereas no significant difference was found for all tested concentrations of spermidine in any protein. This could mean that spermidine is either a very weak substrate or due to its hydrophilic nature, it may not be able to move effectively into the cell in the absence of energy, and thus be unable to compete effectively for binding into the binding site. Therefore, to test these proteins for spermidine recognition a different approach was employed using energised whole cells to measure the uptake of radio-labelled $^{3}$H-spermidine accumulation. *E. coli* cells containing the empty plasmid pTTQ18, AceI and six AceI homologues (STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587) were tested for transport of $^{3}$H-spermidine at a concentration of 20 mM (Figure 6.9). The uptake of $^{3}$H-spermidine was significantly greater in all uninduced cells compared with cells expressing, AceI, Fbal_3166, PFL_4558, PSPTO_3587, and Tmarg_opt, suggesting that the overexpression of these proteins reduces the level of $^{3}$H-spermidine uptake, due to an active efflux mechanism. The higher level of uptake in the induced cells expressing A1S_1503 and STY_3166 proteins is intriguing. Efflux function by these transporters has not been demonstrated so their function in the cells is unknown. These results confirm that spermidine is another substrate of some PACE family transporters. Assays performed in this project so far measure the efflux or accumulation of a substrate by efflux proteins using whole cells of *E. coli*, where several other factors could affect the assay results. Therefore, purified proteins reconstituted into liposomes will be a good model for such type of assays.

On the basis of expression levels shown in small-scale cultures (Figure 6.2), seven proteins (AceI, Fbal_3166, PFL_4558, PSPTO_3587, A1S_1503, STY_3166 and Tmarg_opt) were recommended for larger-scale cultures. Focusing on a smaller number of proteins that show the most potential at this stage was important to make the best use of time and materials. Out of seven, four proteins (AceI, Fbal_3166, STY_3166 and Tmarg_opt) that represent four phylogenetic groups of PACE family proteins were chosen for purification and further characterisation. The proteins were purified with variable success and their secondary structural integrity was analysed using circular dichroism spectroscopy. All the tested proteins produced spectra with shapes characteristic of alpha-helix content indicating that the detergent-solubilised

purified proteins have retained their secondary structure after purification. Thermostability trials showed that the AceI protein was the most stable as compared to Fbal_3166, STY_3166 and Tmarg_opt. In view of the CD measurements, AceI is the best protein to pursue for studies of high resolution structure determination. From CD results, it was also concluded that all tested proteins are reasonably stable for performing biophysical assays using a temperature range of 18-25 ºC. In steady-state spectrophotofluorimetry measurements, chlorhexidine was the preferred substrate for AceI, Fbal_3166, STY_3166 and Tmarg_opt whereas a very weak or no binding affinity of the putative natural substrates spermidine, cadaverine and putrescine was observed. These results indicate that PACE family proteins are multidrug efflux transporters and that the substrate specificity is different among the proteins.

# Chapter 7

# Conclusions and future perspectives

## 7.1 General conclusions

Membrane proteins are biologically important and technically challenging to study. Nearly 30% of all the genes in most sequenced genomes encode membrane proteins and are targeted by more than 50% of commercially available drugs (Kathy, 2015; Rawson et al., 2016). About one-third of membrane proteins are transport proteins by which desirable nutrients are imported and toxic materials are exported across biological membranes (Ren and Paulsen, 2005). This demonstrates the significance of membrane transporters but despite this importance significantly fewer crystal structures have been determined compared to soluble proteins (www.pdb.org). The reason that there are fewer structures of membrane proteins is that they are notoriously resistant to the determination of high resolution structures because of their hydrophobicity and dynamic flexibility. This study describes applications of molecular, biochemical, biophysical and computational techniques to elucidate the structure-activity relationships of NCS1 and PACE family proteins in order to improve understanding of the molecular mechanisms of these transporters.

Overexpression of recombinant membrane proteins is a bottleneck in the field of membrane protein structural biology. Expression screens on a small-scale developed in this project show that it is possible to find best candidates amongst a number of proteins for further structural and functional studies in a short time frame. The results in chapter-5 show that optimisation of membrane protein expression levels requires screening of various parameters including medium composition, IPTG concentration and period of induction etc. In all cases significant progress in membrane protein overexpression and purification was made and sufficient quantity and quality of proteins were achieved for downstream analyses.

Mhp1 is a model protein for secondary transporters in general and for the NCS1 proteins in particular, and for which crystal structures have been determined (Weyand et al., 2008; Shimamura et al., 2010; Simmons et al., 2012). More structural and functional information was obtained by creating mutants of a conserved residue which highlighted importance of the carboxyl group at the specific position. The information obtained on these mutants has increased our knowledge and depth of understanding about structural changes that affect the function. This information will be applicable to investigate the structure-function relationships of other NCS1 family proteins.

The current lack of an X-ray crystal structure for PACE family members means that any information about the protein's substrate binding and transport mechanism must be indirectly obtained from other experiments to predict their structural and functional features. Here, circular dichroism, fluorimetry, radiolabelled transport assay, and fluorimetric efflux assays were performed to elucidate the transport mechanism and structure-function relationship for these transporters. Such experiments have provided some useful information regarding the novel PACE family but the X-ray crystal or NMR structures of a PACE family member overexpressed and purified in this work would greatly advance our understanding of the data obtained, as well as the structure-function relationship of the transporter group.

## 7.2 Summary of results

A general literature review on membrane proteins and the objectives of this research were undertaken in chapter-1. All the materials and methods that were used to carry out this research are concisely described in chapter-2 including recombinant DNA technologies, proteins production and purification, biochemical, biophysical and computational techniques.

Chapter-3 of this thesis highlights the roles of a single amino acid (Asp229) in ligand binding to the Mhp1 protein. Three mutants of the aspartic acid residue (Asp229) were created to a variety of amino acids (glutamate, asparagine and alanine) with different side chain to investigate their possible effects on ligand binding. All three mutants (D229E, D229N and D229A) were overexpressed and purified using the IMAC procedure. Far-UV circular dichroism measurements of Mhp1 and mutants D229E, D229N and D229A demonstrated that these mutations were not detrimental to the integrity of proteins and had retained alpha helical secondary structure after purification. Thermostability trials showed that the wild-type Mhp1 and the Asp229 mutants had effectively identical thermal stabilities as determined by Global Analysis CD software-3. Stopped-flow and steady-state spectrophotofluorimetry measurements were performed to study the effect of Asp229 residue mutations on the L-BH binding of the purified Mhp1 and mutants proteins. Drastic loss of binding activity for L-BH was observed for the mutants Asp229 to Asn and Asp229 to Ala whereas the conservative mutation Asp229 to Glu has slightly impaired the L-BH

binding affinity. Based on the fluorimetry results, it can be concluded that the carboxyl group side chain of the Asp229 residue is important for binding of L-BH. This residue negative charge might be interacting with the positive charge of Lys110 (helix 3) and Lys232 (loop 6-7) forming salt bridges and have an important role in stabilising the inward-facing structure (PDB 2X79). The above statement were recognised by the mutant D229E that possess similar carboxyl group on side chain that could form likewise salt bridges with Lys110 and Lys232 and had retained ligand binding activity. These results suggesting that a single point mutation can disrupt interactions between residues that cause severely impairment of ligand binding.

The cloning and expression of bacterial NCS1 family proteins are described in chapter-4. Eleven proteins homologous with Mhp1 were selected for cloning and expression. Six proteins (AAN69889, VPA1242, NMB2067, BAA80379, CAC11736 and EIQ13585) were successfully cloned from bacterial genomic DNA. Out of the six cloned genes, two proteins (AAN69889 and VPA1242) were expressed in *E. coli* BL21 (DE3) whereas two proteins (NMB2067 and BAA80379) were expressed in *E. coli* BL21-star (DE3). The gene CAC11736 was tested in various strains of *E. coli* but was not expressed at all while the gene EIQ13585 sequence was incorrect and would need to be repeated. There could be many possible reasons for unsuccessful expression of the successfully cloned gene. These might be inappropriate codon usage, mRNA instabilities, choice of a vector, overexpression might cause aggregation or lack of correct chaperones, thus limiting the level of protein being expressed. The two proteins (AAN69889 and VPA1242) were further subjected to expression optimisation, large-scale membrane preparation, purification and functional characterisation.

Chapter-5 describes characterisation of two bacterial transport proteins, AAN69889 from *Pseudomonas putida* and VPA1242 from *Vibrio parahaemolyticus*. For optimal expression of AAN69889 various conditions (IPTG concentration, period of induction and different medium composition) were optimised. Transport of various compounds in energised whole cells were measured using a range of potential radiolabelled substrates that demonstrated highest uptake for $^{14}$C-allantoin and no significant uptake of any other compounds. No sodium effect was observed on $^{14}$C-allantoin uptake by AAN69889 suggesting that transport is possibly driven by a

proton gradient. Based on lack of competition for $^{14}$C-allantoin uptake by a ten-fold excess of unlabelled compounds, AAN69889 was highly specific for allantoin. The only other compounds that had small but significant competitive effects were uracil and hydantoin, which are structurally similar to allantoin. Protein databases (e.g. UniProt, NCBI) describe AAN69889 as a nucleoside transporter but our results have demonstrated that AAN69889 is a highly specific transporter of allantoin with no significant recognition of nucleosides or nucleobases. This emphasises the importance of laboratory experiments for confirming substrate specificity and other characterisations of membrane transport proteins. It is fair to say that it is not always appropriate to rely on databases and/or *in silico* predictions for such information. IMAC purification of AAN69889 using a Ni-NTA resin produced protein with a purity of around 80% and a yield of around 1.2 mg/litre from fermenter cultures using LB medium supplemented with glycerol. The circular dichroism confirmed the integrity of the secondary structure of AAN69889 and indicated that the detergent-solubilised purified protein has an alpha-helix content. A melting temperature of 46.2 ºC was estimated using Global Analysis CD software-3 and therefore AAN69889 protein is reasonably stable for performing biophysical assays using a temperature range of 18-25 ºC. Fluorimetry measurements found no significant difference in the fluorescence quench obtained with allantoin suggesting that the binding site in AAN69889 may not involve tryptophan residues.

Protein databases describe VPA1242 as a putative cytosine permease and the protein is most closely related to cytosine transporter CodB from *E. coli* (Table 4.1). Transport measurements with VPA1242 in energised whole cells using a range of potential radiolabelled substrates demonstrated highest uptake for $^{3}$H-cytosine and no significant uptake of any other compounds. Based on competition of $^{3}$H-cytosine uptake by a ten-fold excess of unlabelled compounds, VPA1242 was highly specific for cytosine. The only other compounds that had small but significant competitive effects were hydantoin, benzyhydantoin, and uracil, which share some structural similarity to cytosine. $^{3}$H-cytosine uptake by VPA1242 was not dependent on sodium, suggesting that transport is driven by a proton gradient. IMAC purification of VPA1242 using a Ni-NTA resin produced protein with a purity of around 85% and a yield of around 0.9 mg/litre from fermentor cultures using LB medium. Far UV circular dichroism confirmed alpha helical secondary structure of VPA1242 and

a melting temperature of 42.8 ºC was estimated using Global Analysis CD software-3. The quench in fluorescence intensity observed for titration of cytosine with VPA1242 in the absence and presence of added NaCl was similar in both cases and had an approximately linear relationship. This suggested that the observed quenching effect may be due to non-specific interactions or absorbance of the excitation or emission wavelengths of light by cytosine itself. A control experiment by titrating cytosine without protein indicated negative fluorescence quenching suggesting that cytosine itself absorbs more light in this range of wavelength. Overall, these biochemical and biophysical studies have confirmed substrate specificity, structure integrity and functional activities of AAN69889 and VPA1242 proteins. This study will help to proceed further research, especially mutagenesis studies to identify residues in AAN69889 and VPA1242 that could be important for substrate binding and recognition and crystallisation trials for structural studies.

In Chapter-6 a strategy has been demonstrated for the expression screening, purification and characterization of PACE family transport proteins. Starting with twenty-four proteins, small-scale expression screening identified seven proteins (AceI, STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587) that were suitable for scaling up to larger cultures, inner membrane preparations and purifications. The best expressed proteins were tested for acriflavine efflux after expression in *E. coli* cells. The rate of acriflavine efflux was measured in induced cells expressing PSPTO_3587, Fbal-3166 and AceI. It was demonstrated that the efflux mediated by PSPTO_3587, Fbal_3166 and AceI is energy dependent, but the precise mode of energisation is still unknown. No effect of different cations (ChCl, LiCl, NaCl, KCl, RbCl and CsCl) on acriflavine efflux was observed except KCl, which may be an artefact in this assay possibly a general effect on the rate of acriflavine efflux from whole cells, through a decreased affinity of acriflavine for nucleic acids. Dose dependent inhibition of acriflavine efflux by chlorhexidine was observed for all three proteins, whereas no significant difference was found for all tested concentrations of spermidine in any protein. This could mean that spermidine is either a very weak substrate or due to its hydrophilic nature, it may not be able to effectively move into the cell in the absence of energy, and thus be unable to effectively compete for binding site.

For spermidine recognition a different approach was employed using energised whole cells to measure the uptake of radio-labelled $^{3}$H-spermidine accumulation. *E. coli* cells containing the empty plasmid pTTQ18, AceI and six AceI homologues (STY_3166, Fbal_3166, PFL_4558, A1S_1503, Tmarg_opt and PSPTO_3587) were tested for transport of $^{3}$H-spermidine (Figure 6.9). The uptake of $^{3}$H-spermidine was significantly greater in all uninduced cells compared with induced cells expressing, AceI, Fbal_3166, PFL_4558, PSPTO_3587, and Tmarg_opt, suggesting that the overexpression of these proteins reduces the level of $^{3}$H-spermidine uptake, due to an active efflux mechanism. The higher level of uptake in the induced cells expressing A1S_1503 and STY_3166 proteins is intriguing. Efflux function by these transporters has not been demonstrated so their function in the cells is unknown. These results indicate that spermidine is a substrate of some PACE family transporters.

On the basis of expression levels shown in small-scale cultures (Figure 6.2), seven proteins (AceI, Fbal_3166, PFL_4558, PSPTO_3587, A1S_1503, STY_3166 and Tmarg_opt) were recommended for larger-scale cultures. Focusing on a smaller number of proteins that show the most potential was important to make the best use of time and materials. Out of seven, four proteins (AceI, Fbal_3166, STY_3166 and Tmarg_opt) were chosen for purification and further characterisation. The proteins were purified with variable success and their secondary structural integrity was analysed using circular dichroism spectroscopy. All the tested proteins had retained alpha helical secondary structure integrity after purification. Thermostability trials showed that the AceI protein was the most stable as compared to Fbal_3166, STY_3166 and Tmarg_opt. In view of the CD measurements, AceI is the best protein to pursue for studies of high resolution structure determination. From CD results, it was also concluded that all tested proteins are reasonably stable for performing biophysical assays using a temperature range of 18-25 $^{o}$C. In steady-state spectrophotofluorimetry measurements, chlorhexidine was the preferred substrate for AceI, Fbal_3166, STY_3166 and Tmarg_opt whereas a very weak or no binding affinity of the putative natural substrates spermidine, cadaverine and putrescine was observed. These results indicate that PACE family proteins are multidrug efflux transporters and that the substrate specificity is different among the proteins.

## 7.3 Future directions

This study will help to progress further research in a number of directions. This project was based on the expression screening, large scale production, purification and characterisation of a number of different proteins. However, there is additional research work that could be conducted to further characterise these proteins. For the continuation of structural research a good starting point could be to improve purification of proteins to achieve closer to 100 % purity using, for example size exclusion chromatography and/or ion exchange chromatography. Membrane proteins purification is challenging and the proteins often aggregate and lose their functional integrity after removal from the lipid bilayer. As every protein is different, different strategies and protocols are used to purify each individual protein according to its intended use. For structural determination purposes high throughput purification is required. This would likely be followed by initiation of crystallisation trials for all suitable proteins, with the long term aim of protein(s) 3-D structure determination using X-ray crystallography, which is a well-established technique. Detailed 3D structural information unveils real facts regarding the mechanism of these proteins. Such information is important in the designing and developing of effective drugs with improved selectivity. Drug discovery based on structure is reported to be more efficient and can shorten the drug development process (Nicola and Abagyan, 2009; Paul et al., 2010).

Studies conducted with bacterial NCS-1 family proteins would be advantageous compared to eukaryotic homologues due to their higher stability and ease of production. Based on sequence alignments AAN69889 from *P. putida* shares overall homologies of 50.8% (27.0% identical, 23.8% highly similar) with Mhp1 and there is some residue conservation at the substrate and cation binding sites (Figure 5.4). AAN69889 may be a good target for crystallisation trials and it would be interesting to determine whether AAN69889 has a similar structural fold to Mhp1 and other LeuT superfamily members. Proteins with similar three-dimensional (3D) structure are likely to have similar function. An understanding of the structure-activity relationships in AAN69889 may not only give new insights into the mechanisms of NCS-1 family transporters but also into the whole LeuT superfamily.

Studies conducted with PACE family proteins provide further evidence that efflux is the mechanism of resistance in this family of proteins (Chapter 6). An active efflux of chlorhexidine, spermidine and acriflavine have been identified in several members of PACE family using fluorimetric and radiolabelled substrate efflux assays. In future studies such assays will be highly valuable to find physiological substrates for these transporters. Chlorhexidine and acriflavine are synthetic compounds commercially available from the last few decades, their binding and transport by PACE proteins is fortuitous to the host cell, but these compounds are unlikely to be physiological substrates. The basic information obtained in this study about PACE family transporters will entail atomic level study using NMR and electron microscopy for deeper understanding of transport mechanism operating in these proteins. Which will help in designing and developing of inhibitors that could increase the susceptibility of organisms for antimicrobials.

Additionally, the following recommendations are more specific targets that were pointed out during the progress of this research project. Proteins that were poorly expressed need to be optimised for their expression by testing various hosts, different expression vectors as well as growth conditions. Three proteins (PFL_4558, PSPTO_3587 and A1S_1503) of the PACE family and two proteins (NMB2067 and BAA80379) of the NCS1 family that were well expressed but not pursued for large scale production and further characterisation due to time limit of this project are suggested to be tested for the correct substrate, structural integrity and thermal stability.

Three mutants of aspartate residue (Asp229) in Mhp1 protein were generated and the importance of negatively charged side chains has been demonstrated (Chapter 3). This aspartate residue is highly conserved among the NCS1 family proteins (Appendix 2). It would be interesting to mutate this residue in AAN69889 and VPA1242 to provide much information regarding the structure-activity relationships within these transporters. Site directed mutagenesis may prove beneficial in determining ligand binding sites within AAN69889 and VPA1242.

## 7.4 Concluding remarks

Significant progress has been made in the area of membrane protein structural biology. The relative scarcity of high resolution structures limits the information available for elucidating the molecular mechanisms and ligand interactions of membrane proteins and for using them in structure-based drug design. There is therefore a huge amount of information yet to be learned about the structure and function of membrane proteins. The much needed progress in this area will open their potential to improving treatment of infectious diseases and assist in the design and development of novel therapeutics. It is an exciting time for membrane transporter field to exploit the available diverse sophisticated techniques of molecular, biochemical, biophysical and computational methodologies for structural and functional characterisation of these transporters.

# Appendices

**Appendix 1 An alignment of the DNA sequences of the gene and amino acids of the wild-type Mhp1 and its mutants.**

**Asp229Asn DNA sequence alignment**

```
Mhp1        ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
D229N       ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
            ************************************************************

Mhp1        CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
D229N       CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
            ************************************************************

Mhp1        ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
D229N       ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
            ************************************************************

Mhp1        GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
D229N       GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
            ************************************************************

Mhp1        AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
D229N       AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
            ************************************************************

Mhp1        CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
D229N       CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
            ************************************************************

Mhp1        CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
D229N       CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
            ************************************************************

Mhp1        AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
D229N       AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
            ************************************************************

Mhp1        ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
D229N       ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
            ************************************************************

Mhp1        TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
D229N       TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
            ************************************************************

Mhp1        GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
D229N       GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
            ************************************************************

Mhp1        ATCGCGGTCGTGGTGAGCATCCACGACATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
D229N       ATCGCGGTCGTGGTGAGCATCCACAACATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
            ********************** ***********************************

Mhp1        TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
D229N       TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
            ************************************************************

Mhp1        GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
D229N       GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
            ************************************************************

Mhp1        TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
D229N       TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
            ************************************************************

Mhp1        CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
D229N       CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
            ************************************************************


Mhp1        TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
D229N       TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
            ************************************************************
```

```
Mhp1        ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
D229N       ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
            ************************************************************

Mhp1        ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC     1140
D229N       ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC     1140
            ************************************************************

Mhp1        TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC     1200
D229N       TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC     1200
            ************************************************************

Mhp1        TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG     1260
D229N       TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG     1260
            ************************************************************

Mhp1        GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG     1320
D229N       GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG     1320
            ************************************************************

Mhp1        CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC     1380
D229N       CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC     1380
            ************************************************************

Mhp1        CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT     1440
D229N       CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT     1440
            ************************************************************

Mhp1        GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC     1500
D229N       GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC     1500
            ************************************************************

Mhp1        CATA    1504
D229N       CATA    1504
            ****
```

## Asp229Asn Amino acid sequence alignment

```
Mhp1        MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
D229N       MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
            ************************************************************

Mhp1        VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
D229N       VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
            ************************************************************

Mhp1        QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
D229N       QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
            ************************************************************

Mhp1        YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHDIVKEAKVDPNA     240
D229N       YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHNIVKEAKVDPNA     240
            ********************************************** ***********

Mhp1        SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
D229N       SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
            ************************************************************

Mhp1        LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
D229N       LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
            ************************************************************

Mhp1        TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
D229N       TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
            ************************************************************

Mhp1        VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
D229N       VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
            ************************************************************

Mhp1        ANTKEQNQPAGGRGSHHHHHH      501
D229N       ANTKEQNQPAGGRGSHHHHHH      501
            **********************
```

## Asp229Ala DNA sequence alignment

```
Mhp1        ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
D229A       ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
            ************************************************************

Mhp1        CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
D229A       CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
            ************************************************************

Mhp1        ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
D229A       ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
            ************************************************************

Mhp1        GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
D229A       GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
            ************************************************************

Mhp1        AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
D229A       AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
            ************************************************************

Mhp1        CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
D229A       CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
            ************************************************************

Mhp1        CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
D229A       CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
            ************************************************************

Mhp1        AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
D229A       AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
            ************************************************************

Mhp1        ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
D229A       ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
            ************************************************************

Mhp1        TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
D229A       TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
            ************************************************************

Mhp1        GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
D229A       GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
            ************************************************************

Mhp1        ATCGCGGTCGTGGTGAGCATCCACGACATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
D229A       ATCGCGGTCGTGGTGAGCATCCACGCAATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
            ************************  ********************************

Mhp1        TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
D229A       TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
            ************************************************************

Mhp1        GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
D229A       GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
            ************************************************************

Mhp1        TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
D229A       TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
            ************************************************************

Mhp1        CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
D229A       CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
            ************************************************************



Mhp1        TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
D229A       TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
            ************************************************************

Mhp1        ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
D229A       ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
            ************************************************************
```

```
Mhp1        ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC    1140
D229A       ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC    1140
            ************************************************************

Mhp1        TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC    1200
D229A       TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC    1200
            ************************************************************

Mhp1        TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG    1260
D229A       TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG    1260
            ************************************************************

Mhp1        GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG    1320
D229A       GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG    1320
            ************************************************************

Mhp1        CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC    1380
D229A       CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC    1380
            ************************************************************

Mhp1        CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT    1440
D229A       CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT    1440
            ************************************************************

Mhp1        GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC    1500
D229A       GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC    1500
            ************************************************************

Mhp1        CATA    1504
D229A       CATA    1504
            ****
```

## Asp229Ala Amino acid sequence alignment

```
Mhp1        MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
D229A       MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
            ************************************************************

Mhp1        VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
D229A       VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
            ************************************************************

Mhp1        QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
D229A       QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
            ************************************************************

Mhp1        YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHDIVKEAKVDPNA     240
D229A       YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHAIVKEAKVDPNA     240
            ********************************************** ***********

Mhp1        SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
D229A       SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
            ************************************************************

Mhp1        LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
D229A       LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
            ************************************************************

Mhp1        TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
D229A       TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
            ************************************************************

Mhp1        VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
D229A       VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
            ************************************************************

Mhp1        ANTKEQNQPAGGRGSHHHHHH      501
D229A       ANTKEQNQPAGGRGSHHHHHH      501
            **********************
```

## Asp229Glu DNA sequence alignment

```
Mhp1        ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
D229E       ATGAACTCGACACCCATCGAAGAGGCTCGCAGCCTCCTGAACCCATCCAATGCACCCACT     60
            ************************************************************

Mhp1        CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
D229E       CGATACGCCGAGCGCTCCGTCGGCCCGTTCTCCCTCGCGGCCATCTGGTTCGCCATGGCG     120
            ************************************************************

Mhp1        ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
D229E       ATCCAGGTGGCGATCTTCATCGCCGCGGGACAGATGACGAGCAGCTTCCAGGTCTGGCAG     180
            ************************************************************

Mhp1        GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
D229E       GTGATCGTCGCCATCGCCGCAGGCTGCACGATCGCAGTGATCCTGCTCTTCTTCACCCAG     240
            ************************************************************

Mhp1        AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
D229E       AGCGCGGCGATCCGCTGGGGCATCAACTTCACGGTCGCCGCGCGGATGCCTTTCGGCATC     300
            ************************************************************

Mhp1        CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
D229E       CGCGGATCGCTGATCCCGATCACCCTCAAGGCCCTGCTCTCGCTGTTCTGGTTCGGCTTC     360
            ************************************************************

Mhp1        CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
D229E       CAGACGTGGCTGGGCGCGCTGGCGCTCGATGAGATCACGCGTCTCCTCACCGGATTCACG     420
            ************************************************************


Mhp1        AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
D229E       AACCTGCCGCTGTGGATCGTCATCTTCGGCGCGATCCAGGTCGTGACGACCTTCTACGGG     480
            ************************************************************

Mhp1        ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
D229E       ATCACGTTCATCCGCTGGATGAACGTCTTCGCCTCGCCGGTGCTCCTCGCGATGGGCGTG     540
            ************************************************************

Mhp1        TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
D229E       TACATGGTGTACCTGATGCTCGACGGCGCCGACGTGAGCCTCGGCGAGGTCATGTCGATG     600
            ************************************************************

Mhp1        GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
D229E       GGTGGCGAGAACCCTGGCATGCCGTTCTCGACCGCGATCATGATCTTCGTCGGCGGCTGG     660
            ************************************************************

Mhp1        ATCGCGGTCGTGGTGAGCATCCACGACATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
D229E       ATCGCGGTCGTGGTGAGCATCCACGAGATCGTGAAGGAGGCCAAGGTCGACCCGAACGCG     720
            ************************** ***********************************

Mhp1        TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
D229E       TCGCGAGAAGGTCAGACGAAGGCCGACGCGCGATACGCCACGGCGCAGTGGCTCGGCATG     780
            ************************************************************

Mhp1        GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
D229E       GTGCCGGCATCCATCATCTTCGGATTCATCGGCGCCGCCTCGATGGTGCTGGTGGGGGAG     840
            ************************************************************

Mhp1        TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
D229E       TGGAACCCGGTCATCGCCATCACCGAGGTGGTCGGCGGCGTGTCGATCCCGATGGCGATC     900
            ************************************************************

Mhp1        CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
D229E       CTCTTCCAGGTCTTCGTGCTGCTCGCCACCTGGTCGACCAACCCCGCAGCGAATCTCCTC     960
            ************************************************************

Mhp1        TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
D229E       TCGCCGGCGTACACGCTGATCAGCACGTTCCCGCGGGTGTTCACGTTCAAGACCGGTGTG     1020
            ************************************************************

Mhp1        ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
D229E       ATCGTCTCGGCGGTCGTCGGCCTGCTGATGATGCCGTGGCAGTTCGCCGGCGTGCTCAAC     1080
            ************************************************************
```

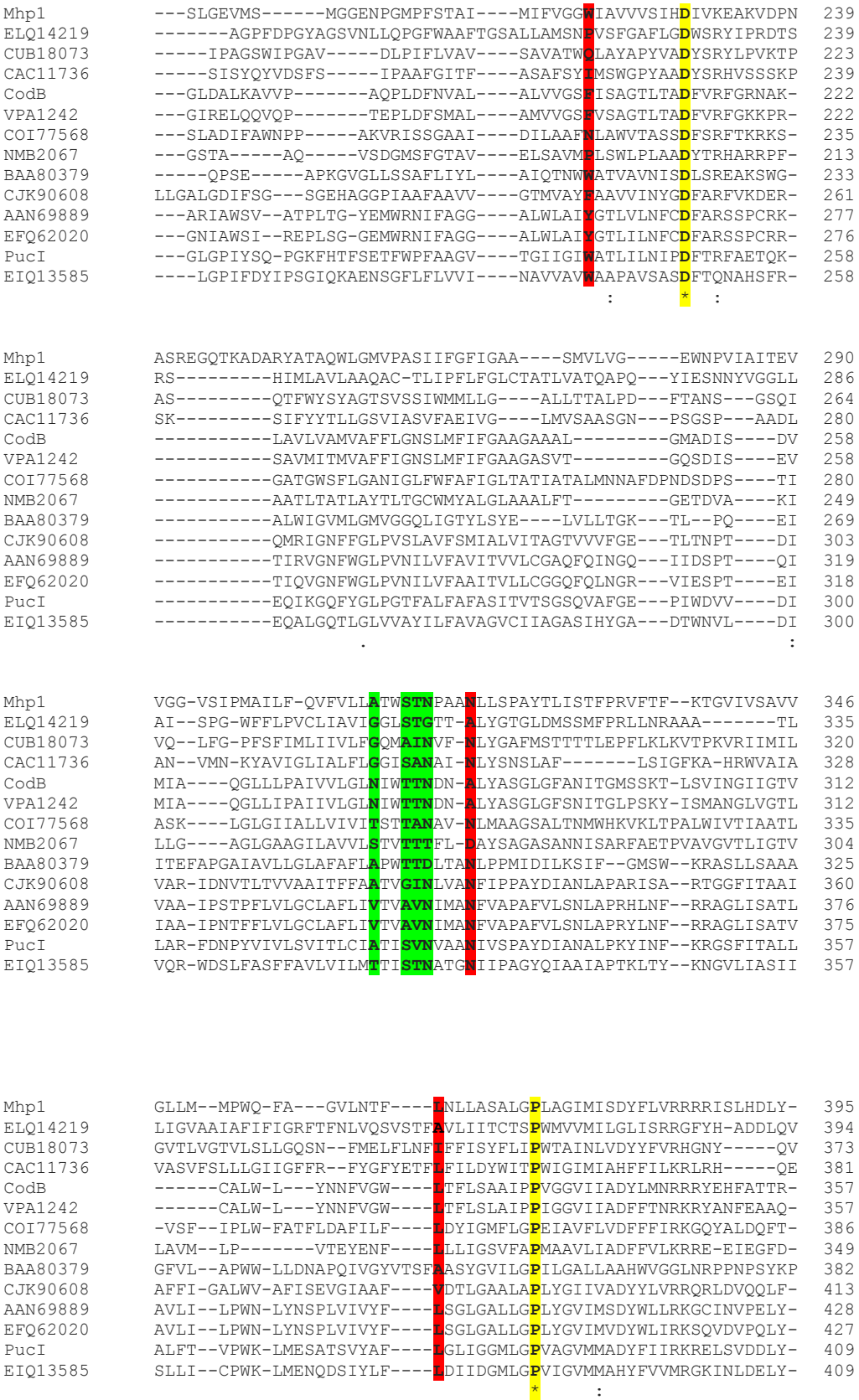
```
Mhp1        ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC    1140
D229E       ACCTTCCTGAACCTGCTTGCGAGTGCTCTCGGCCCGCTCGCGGGGATCATGATCAGCGAC    1140
            ************************************************************

Mhp1        TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC    1200
D229E       TACTTCCTCGTGCGCCGTCGCCGCATCAGCCTGCATGACCTGTATCGGACCAAGGGCATC    1200
            ************************************************************

Mhp1        TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG    1260
D229E       TACACGTACTGGCGAGGGGTCAACTGGGTCGCACTCGCGGTCTACGCGGTCGCGCTGGCG    1260
            ************************************************************

Mhp1        GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG    1320
D229E       GTGTCGTTCCTCACTCCGGACCTGATGTTCGTGACCGGCCTGATCGCCGCCCTTCTGCTG    1320
            ************************************************************

Mhp1        CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC    1380
D229E       CACATCCCGGCGATGCGATGGGTGGCGAAGACCTTCCCGCTGTTCTCCGAAGCCGAGAGC    1380
            ************************************************************

Mhp1        CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT    1440
D229E       CGGAACGAGGACTACCTGCGACCGATCGGCCCTGTGGCGCCGGCGGACGAATCAGCGACT    1440
            ************************************************************


Mhp1        GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC    1500
D229E       GCGAACACGAAGGAGCAGAACCAGCCTGCAGGCGGTCGTGGCAGCCACCATCACCATCAC    1500
            ************************************************************

Mhp1        CATA    1504
D229E       CATA    1504
            ****
```

## Asp229Glu Amino acid sequence alignment

```
Mhp1        MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
D229E       MNSTPIEEARSLLNPSNAPTRYAERSVGPFSLAAIWFAMAIQVAIFIAAGQMTSSFQVWQ     60
            ************************************************************

Mhp1        VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
D229E       VIVAIAAGCTIAVILLFFTQSAAIRWGINFTVAARMPFGIRGSLIPITLKALLSLFWFGF     120
            ************************************************************

Mhp1        QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
D229E       QTWLGALALDEITRLLTGFTNLPLWIVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGV     180
            ************************************************************

Mhp1        YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHDIVKEAKVDPNA     240
D229E       YMVYLMLDGADVSLGEVMSMGGENPGMPFSTAIMIFVGGWIAVVVSIHEIVKEAKVDPNA     240
            ************************************************:***********

Mhp1        SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
D229E       SREGQTKADARYATAQWLGMVPASIIFGFIGAASMVLVGEWNPVIAITEVVGGVSIPMAI     300
            ************************************************************

Mhp1        LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
D229E       LFQVFVLLATWSTNPAANLLSPAYTLISTFPRVFTFKTGVIVSAVVGLLMMPWQFAGVLN     360
            ************************************************************

Mhp1        TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
D229E       TFLNLLASALGPLAGIMISDYFLVRRRRISLHDLYRTKGIYTYWRGVNWVALAVYAVALA     420
            ************************************************************

Mhp1        VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
D229E       VSFLTPDLMFVTGLIAALLLHIPAMRWVAKTFPLFSEAESRNEDYLRPIGPVAPADESAT     480
            ************************************************************

Mhp1        ANTKEQNQPAGGRGSHHHHHH      501
D229E       ANTKEQNQPAGGRGSHHHHHH      501
            **********************
```

**Appendix 2 Amino acids sequence alignment of Mhp1 and its homologues.** Using multiple sequence alignment tool "Clustal W" Mhp1 and its homologues amino acids sequence were aligned. The conserved residues are shown in yellow. The residues in red are involved in substrate recognition and green in cation coupling in Mhp1.

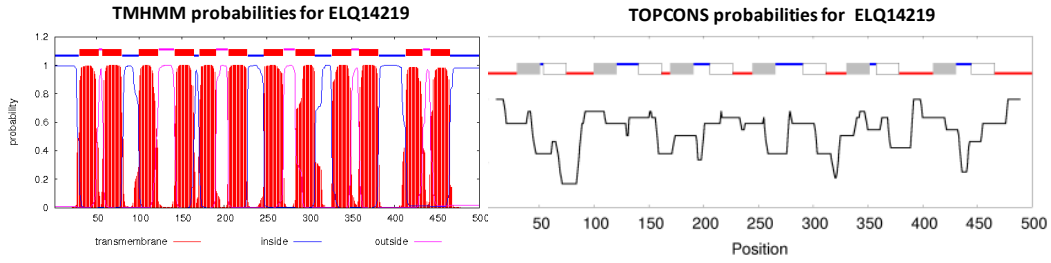
```
Mhp1          ------------------------MN-STPIEEARSLLNPSNAPTRY----AERSVGPFS   31
ELQ14219      ----------------MSHSTR--------------IETF-GVEQIP---DTQRDASPID   26
CUB18073      ---------------------MK------------VERRT-IEYIP---NEERHGKAKD   22
CAC11736      ----------------MTHSTDMSTE----HEFGYDIDKYSKANKEI---DIENDKHPSS   37
CodB          ----------------------MSQD-----------NNFSQGPVPQSARKGVLALTFVM   27
VPA1242       ----------------------MAGD-----------NNYSLGPVPNTARKGVASLTMVM   27
COI77568      ------------------------ME-----------KQFGRVEVISA--D-KRTMSNWD   22
NMB2067       ---------------------------------------------MSG--NASSPSSSSA   13
BAA80379      ------------------------------------MGLGNTPE----KGGGLGALE   17
CJK90608      ---------------------MNHDGFQIKQIDPSLYNADLAPLPP----AERKWGWFE   34
AAN69889      MSS--SLDLAPELSVASTHPASTLAGHQPDLVLSPRLHNRDLAPTRM----EGRRWGGYS   54
EFQ62020      MRTSLSNDLALDLPSSTLTP---EAASPGPLVLSPRLHNKDLAPTKA----EGRRWGRYS   53
PucI          ------------------------MKLKESQQQSNRLSNEDLVPLGQ----EKRTWKAMN   32
EIQ13585      ------------------------ME-HQRKLFQQRGYSEDLLPKTQ----SQQTWKTFN   31

Mhp1          LAAIWFAMAIQVAIFIAAGQM--TSSFQVWQVIVAIAAGCTIAVILLFFTQSAAIRWGIN   89
ELQ14219      LFRLIFGGANTFATAVLGSFPV-LFGLSFQAGVWAIVLGVTMGALILAPMGLFGAINGTN   85
CUB18073      LFPVWFGANMHITTLVTGTIPV-AMGLNLFWSVAAIICGTLIGAIFMASHSAQGPQLGIP   81
CAC11736      MFYIWFASNLITVGDFAVGFIPV-YLGLPILYSIVAIAVGTIAGGIMLAYMSQLGAVYRVP   96
CodB          LGLTFFSA----SMWTGGTL---GTGLSYHDFFLAVLIGNLLLGIYTSFLGYIGAKTGLT   80
VPA1242       LGLTFFSA----SMWTGGSL---GTGLSFNDFFLAVLIGNLILGIYTSFLGYIGASTGLS   80
COI77568      MFATWVGANANNGTWYIGGV---IAAAGMYTASTTLIISGLVSYALLALASFMGYKTGLP   79
NMB2067       IGLIWFGAAVSIAEISTGTL---LAPLGWQRGLAALLLGHAVGGALFFAAAYIGALTGRS   70
BAA80379      YTFIMFSMASCLPLFFLGPIAF-NLGLSLQEALLAALVGNLVVAVAMALNGHAGIKHKID   76
CJK90608      IFNVWSNDIQSLFGYTLAATLFISYGLNGWAVLTGIVLAGFIVMGLVHLTGKPSVKYGIP   94
AAN69889      IFALWTNDVHNIANYSFAMGLF-ALGLGGWQILLSLAIGAALVYFFMNLSGYMGQKTGVP   113
EFQ62020      IFALWTNDVHNIANYSFAIGLY-ALGLGGWQIVLSLGIGAALVYFFMNLSGYMGQKTGVP   112
PucI          FASIWMGCIHNIPTYATVGGLI-AIGLSPWQVLAIIITASLILFGALALNGHAGTKYGLP   91
EIQ13585      YFTLWMGSVHNVPNYVMVGGFF-ILGLSTFSIMLAIILSAFFIAAVMVLNGAAGSKYGVP   90

                                                          .         .

Mhp1          FTVAARMPFGIRGSLIPITLKA-LLSLFWFGFQTWLGALALDEITR-LLTGFTNLPL---   144
ELQ14219      NAVSSGAHFGVHGRIIGSFLSL-LTAVAFFSLSVWSSGDALVGGAQ----RLVGLPENNT   140
CUB18073      QMIQSRAQFGVIGAILPLFLVM-FIYLGFFASSTILAAGT----LS----SFVPIPGSWS   132
CAC11736      QMYMGRAPFGTVGGSLLSILQW-GNTAGWLTVNVILASLA----IY----QMVKIPYYAI   147
CodB          THLLARFSFGVKGSWLPSLLLG-GTQVGWFGVGVAMFAIPV--------------------   120
VPA1242       THLLARFSFGSKGSWLPSALLG-GTQVGWFGVGVAMFAIPV--------------------   120
COI77568      MMALTRASFGLRGSFIPSIINI-VQFIGWAAANTFIAAISISVIFK-----DLFGWQAYT   133
NMB2067       SMESVRLSFGKRGSVLFSVANM-LQLAGWTAVMIYAGATVSSALGKVLWDGESFVWWALA   129
BAA80379      FPEQAVRSLGELTGKAAVVMRG-LVGAMWFGVEAYNGALALNLILLFAL-GLTGAA---L   131
CJK90608      FPVMARASMGVRGANFPAVVRG-IVAIFWYGVQTYFASTAVALLLRAFM---GTDPQAAT   150
AAN69889      FPVISRIAFGIHGAQIPALIRA-VIAIAWFGIQTYLASVVLRVLLTAVWPQIAAYDHD-S   171
EFQ62020      FPVISRISFGIHGAQIPALIRA-VIAIAWFGIQTYLASVVLRVLLTAIHPGFADYDHN-A   170
PucI          FPVIIRASYGIYGANIPALLRA-FTAIMWLGIQTFAGSTALNILLLNMWPGWGEIGGEWN   150
EIQ13585      FAMILCASYGVRGALFPGLLRGGIAAIMWFGLQCYAGSLACLILIGKIWPGFLTLGGDFT   150
                        *                    :         .

Mhp1          --------W---IVIFGAIQVVTTFYGITFIRWMNVFASPVLLAMGVYMVYLMLDGADV-   192
ELQ14219      SLGL-------AYGLFAVLVLIVCIFGFRFMLWINKVAVWASSLLFLSGILAF-------   186
CUB18073      IIGL------------SAVCFLLTIFGHDLIHKMQKILSWTSFAVFFAATI-LIFQLP--   177
CAC11736      VLLI------------VAIVALTALVGYRAIRILERSMSYVLGLLFVFLVLFLIHSHA--   193
CodB          -GKATGLDINLLIAVSGLLMTVTVFFGISALTVLSVIAVPAIACLGGYSVWLAVNGM--G   177
VPA1242       -HKATGIDTNTLILVSGLLMTATVYFGISALMVLSAIAVPAIALLGGYSVVEAVNSV--G   177
COI77568      AGGKAGLVIGIII--MSLLHLASISLGERSVRMIERIG---IVLVFIFVLWESIVVFQHV   188
NMB2067       NGAL--------I-------VLWLVFGARKTGGLKTVS---MLLMLLAVLWLSAEVFSTA   171
BAA80379      LEKATVLIP-AALVLYLGSMYLVLKLGVKGIGKAATLAGPLLLLYFAWLWIWMKNSGF--   188
CJK90608      LLGLTAIDW-AAYVMVCVFQVALFVRGVDWVTRFLNWAGPLVYLVMIVLMIAIWYQAGPS   209
AAN69889      ILGLSSLGW-VCFVSIWLVQLVILAYGMEMVRRYEAFAGPVILLTVAALAVFMYFKAD--   228
EFQ62020      ILGLSTLGW-ACFVVIWFVQLVILAYGMEMVRRYEGFAGPVILATMAALAGWMYFQAG--   227
PucI          ILGIHLSGL-LSFVFFWAIHLLVLHHGMESIKRFEVWAGPLVYLVFGGMVWWAVDIAG--   207
EIQ13585      LLGLSLPGL-ITFLLFWLVNVGIGFGGGKVLNKFTAILNPCIYIVFGGMAIWAISLVG--   207
                                        *
```


Mhp1 ---SLGEVMS------MGGENPGMPFSTAI----MIFVGGWIAVVVSIHDIVKEAKVDPN 239
ELQ14219 -------AGPFDPGYAGSVNLLQPGFWAAFTGSALLAMSNPVSFGAFLGDWSRYIPRDTS 239
CUB18073 -----IPAGSWIPGAV-----DLPIFLVAV----SAVATWQLAYAPYVADYSRYLPVKTP 223
CAC11736 -----SISYQYVDSFS-----IPAAFGITF----ASAFSYIMSWGPYAADYSRHVSSSKP 239
CodB ---GLDALKAVVP-------AQPLDFNVAL----ALVVGSFISAGTLTADFVRFGRNAK- 222
VPA1242 ---GIRELQQVQP-------TEPLDFSMAL----AMVVGSFVSAGTLTADFVRFGKKPR- 222
COI77568 ---SLADIFAWNPP-----AKVRISSGAAI----DILAAFNLAWVTASSDFSRFTKRKS- 235
NMB2067 ---GSTA-----AQ-----VSDGMSFGTAV----ELSAVMPLSWLPLAADYTRHARRPF- 213
BAA80379 -----QPSE-----APKGVGLLSSAFLIYL----AIQTNWWATVAVNISDLSREAKSWG- 233
CJK90608 LLGALGDIFSG---SGEHAGGPIAAFAAVV----GTMVAYFAAVVINYGDFARFVKDER- 261
AAN69889 ---ARIAWSV--ATPLTG-YEMWRNIFAGG----ALWLAIYGTLVLNFCDFARSSPCRK- 277
EFQ62020 ---GNIAWSI--REPLSG-GEMWRNIFAGG----ALWLAIYGTLILNFCDFARSSPCRR- 276
PucI ---GLGPIYSQ-PGKFHTFSETFWPFAAGV----TGIIGIWATLILNIPDFTRFAETQK- 258
EIQ13585 ----LGPIFDYIPSGIQKAENSGFLFLVVI----NAVVAVWAAPAVSASDFTQNAHSFR- 258
ASREGQTKADARYATAQWLGMVPASIIFGFIGAA----SMVLVG-----EWNPVIAITEV 290
ELQ14219 RS---------HIMLAVLAAQAC-TLIPFLFGLCTATLVATQAPQ---YIESNNYVGGLL 286
CUB18073 AS---------QTFWYSYAGTSVSSIWMMLLG----ALLTTALPD---FTANS---GSQI 264
CAC11736 SK---------SIFYYTLLGSVIASVFAEIVG----LMVSAASGN---PSGSP---AADL 280
CodB ----------LAVLVAMVAFFLGNSLMFIFGAAGAAAL---------GMADIS----DV 258
VPA1242 ----------SAVMITMVAFFIGNSLMFIFGAAGASVT---------GQSDIS----EV 258
COI77568 ----------GATGWSFLGANIGLFWFAFIGLTATIATALMNNAFDPNDSDPS----TI 280
NMB2067 ----------AATLTATLAYTLTGCWMYALGLAAALFT---------GETDVA----KI 249
BAA80379 ----------ALWIGVMLGMVGGQLIGTYLSYE----LVLLTGK---TL--PQ----EI 269
CJK90608 ----------QMRIGNFFGLPVSLAVFSMIALVITAGTVVVFGE---TLTNPT----DI 303
AAN69889 ----------TIRVGNFWGLPVNILVFAVITVVLCGAQFQINGQ---IIDSPT----QI 319
EFQ62020 ----------TIQVGNFWGLPVNILVFAAITVLLCGGQFQLNGR---VIESPT----EI 318
PucI ----------EQIKGQFYGLPGTFALFAFASITVTSGSQVAFGE---PIWDVV----DI 300
EIQ13585 ----------EQALGQTLGLVVAYILFAVAGVCIIAGASIHYGA---DTWNVL----DI 300
VGG-VSIPMAILF-QVFVLLATWSTNPAANLLSPAYTLISTFPRVFTF--KTGVIVSAVV 346
ELQ14219 AI--SPG-WFFLPVCLIAVIGGLSTGTT-ALYGTGLDMSSMFPRLLNRAAA-------TL 335
CUB18073 VQ--LFG-PFSFIMLIIVLFGQMAINVF-NLYGAFMSTTTTLEPFLKLKVTPKVRIIMIL 320
CAC11736 AN--VMN-KYAVIGLIALFLGGISANAI-NLYSNSLAF-------LSIGFKA-HRWVAIA 328
CodB MIA----QGLLLPAIVVLGLNIWTTNDN-ALYASGLGFANITGMSSKT-LSVINGIIGTV 312
VPA1242 MIA----QGLLIPAIIVLGLNIWTTNDN-ALYASGLGFSNITGLPSKY-ISMANGLVGTL 312
COI77568 ASK----LGLGIIALLVIVITSTTANAV-NLMAAGSALTNMWHKVKLTPALWIVTIAATL 335
NMB2067 LLG----AGLGAAGILAVVLSTVTTTFL-DAYSAGASANNISARFAETPVAVGVTLIGTV 304
BAA80379 ITEFAPGAIAVLLGLAFAFLAPWTTDLTANLPPMIDILKSIF--GMSW--KRASLLSAAA 325
CJK90608 VAR-IDNVTLTVVAAITFFAATVGINLVANFIPPAYDIANLAPARISA--RTGGFITAAI 360
AAN69889 VAA-IPSTPFLVLGCLAFLIVTVAVNIMANFVAPAFVLSNLAPRHLNF--RRAGLISATL 376
EFQ62020 IAA-IPNTFFLVLGCLAFLIVTVAVNIMANFVAPAFVLSNLAPRYLNF--RRAGLISATV 375
PucI LAR-FDNPYVIVLSVITLCIATISVNVAANIVSPAYDIANALPKYINF--KRGSFITALL 357
EIQ13585 VQR-WDSLFASFFAVLVILMTTISTNATGNIIPAGYQIAAIAPTKLTY--KNGVLIASII 357
GLLM--MPWQ-FA---GVLNTF----LNLLASALGPLAGIMISDYFLVRRRRISLHDLY- 395
ELQ14219 LIGVAAIAFIFIGRFTFNLVQSVSTFAVLIITCTSPWMVVMILGLISRRGFYH-ADDLQV 394
CUB18073 GVTLVGTVLSLLGQSN--FMELFLNFIFFISYFLIPWTAINLVDYYFVRHGNY-----QV 373
CAC11736 VASVFSLLLGIIGFFR--FYGFYETFLFILDYWITPWIGIMIAHFFILKRLRH-----QE 381
CodB ------CALW-L---YNNFVGW----LTFLSAAIPPVGGVIIADYLMNRRRYEHFATTR- 357
VPA1242 ------CALW-L---YNNFVGW----LTFLSLAIPPIGGVIIADFFTNRKRYANFEAAQ- 357
COI77568 -VSF--IPLW-FATFLDAFILF----LDYIGMFLGPEIAVFLVDFFFIRKGQYALDQFT- 386
NMB2067 LAVM--LP-------VTEYENF----LLLIGSVFAPMAAVLIADFFVLKRRE-EIEGFD- 349
BAA80379 GFVL--APWW-LLDNAPQIVGYVTSFAASYGVILGPILGALLAAHWVGGLNRPPNPSYKP 382
CJK90608 AFFI-GALWV-AFISEVGIAAF----VDTLGAALAPLYGIIVADYYLVRRQRLDVQQLF- 413
AAN69889 AVLI--LPWN-LYNSPLVIVYF----LSGLGALLGPLYGVIMSDYWLLRKGCINVPELY- 428
EFQ62020 AVLI--LPWN-LYNSPLVIVYF----LSGLGALLGPLYGVIMVDYWLIRKSQVDVPQLY- 427
PucI ALFT--VPWK-LMESATSVYAF----LGLIGGMLGPVAGVMMADYFIIRKRELSVDDLY- 409
EIQ13585 SLLI--CPWK-LMENQDSIYLF----LDIIDGMLGPVIGVMMAHYFVVMRGKINLDELY- 409

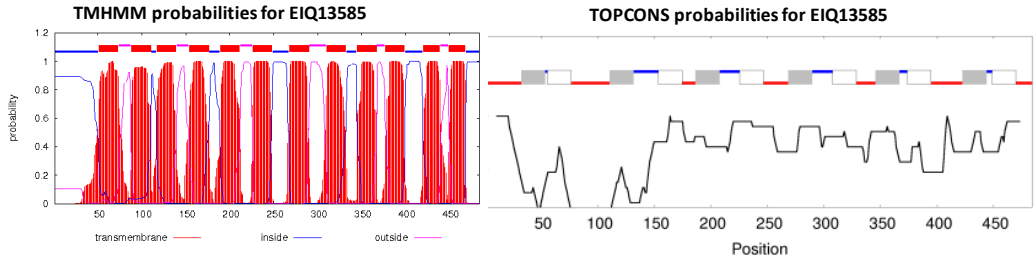
```
Mhp1          --RTKG--IYTYWRGVNWVALAVYAVALAVSFLTPDLMFVTGLIAALLLHIPAMRWVAK-  450
ELQ14219      FTRGQRGGRYWFHHGWNWRGLGAWIPSAAVGLCFVNLPGQFVGPLGELAAGIDISLPIT-  453
CUB18073      KAMFDVNGPY---GKVNWITTIAFVLSILLEIPFINTS-FYIGPLAKMFGGGDIAWIVG-  428
CAC11736      FEDL---------PKIVKPGIYAYIISIAVSIPFMSPAGIINMPLASMLHGVDISYFVS-  431
CodB          ----M-MSV---------NWVAILAVALGIA-AGHWLPGI-----------VPVNAVLG-  390
VPA1242       ----F-QSV---------NWAGIIAVAIGVG-AGHFLPGV-----------VPINAVLG-  390
COI77568      ----KVDGKYWYNGGLN--WIAIASWAVGIG-LYFGLKSV-----------NIISQTIGV  428
NMB2067       ----FAGLVLWLAGFILYRFLLSSGWESSIG-LTAPVMSA-----------VAIAT----  389
BAA80379      VILPTTAG---LLAGLIVSYAIAYPLGMVTSVLGVPF-------------PQGPIWYVG-  425
CJK90608      --CAERTGIYHFNAGWNRKAMIAFGVSAVFSVASVWTPGL--------ESLSGFAWLLG-  462
AAN69889      --TEHPAGAYHYSKGINLRAVAAFVPAALLAIVLALVPNF--------QGIAPFSWLIG-  477
EFQ62020      --SEDPKGVYYYSRGVNLRAVAAFVPAAVIAILLALLPGF--------AGVAPFSWLFG-  476
PucI          --SETGR-YVYW-KGYNYRAFAATMLGALISLIGMYVPVL--------KSLYDISWFVG-  456
EIQ13585      --TAPGD-YKYYDNGFNLTAFSVTLVAVILSLGGKFIPFM--------EPLSRVSWFVG-  457


Mhp1          ------TFPLFSE----AESRNEDYLRPIGPVAPADESATANTKEQNQPAGGRGSHHHHHH 501
ELQ14219      ----LGLAALLYLT----LLRSFPEPVSVYGPKGAHPMLGRKAISSAT-PSATA--HS--  500
CUB18073      ----LAVPSVLYYV----LMKPRLKKRS-YQEKL------SSL-----------------  456
CAC11736      ----FFSAMIIYLY-----LSRKLDDRN-HGEKT------RKEIS---------------  460
CodB          --------GALSYLILNPILNRKTTAAMTHVEANSVE----------------------  419
VPA1242       --------GAISFLILNPILNKKVLATQPA----------------------------  412
COI77568      TFVAMALTGLIYYVA-TKLIKK-------------------------------------  449
NMB2067       ------VSVRLFFKK-TQSLQRNPS----------------------------------  407
BAA80379      ----VAVSVIAAAL----LLKLVPAKVKFKNET--------------------------  450
CJK90608      ----ALFGGTLHYV----LMRKQPLLVTP---------VQPA-------PLG--------  490
AAN69889      ----AGIAAALYLL----IAPRNRHYHDVSGECIAVDHSGH--------------------  510
EFQ62020      ----AGIAGLLYLL----IAKRQAFYADVSGESIAVDNVSH--------------------  509
PucI          ----VLISFLFYIV----LMRVHPPA------SLAIETVEHA-------QVRQAE-----  490
EIQ13585      ----VIVAFAAYAL----LKKRTTAE------KTGEQ----K-------TIG--------  484
```

## Appendix 3 Topology prediction by TOPCONS and TMHMM of the selected NCS1 proteins

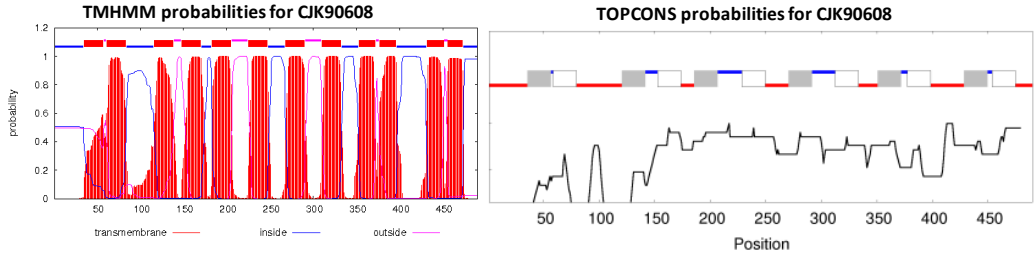
### 1. CUB18073 (*Bacillus cereus*)



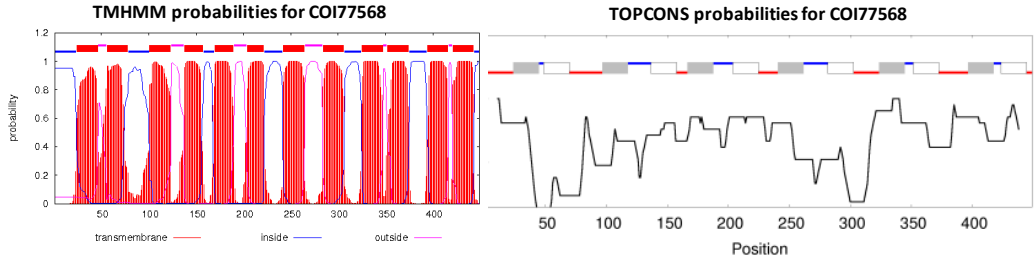

### 2. EIQ13585 (*Shigella flexneri)*



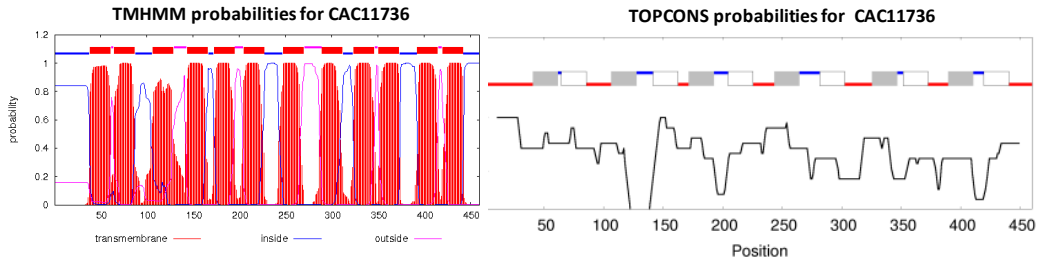

### 3. CJK90608 (*Streptococcus pneumoniae)*



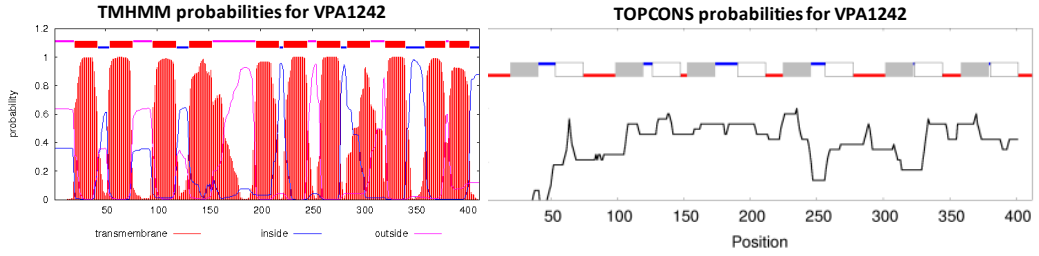

### 4. COI77568 (*Streptococcus pneumoniae)*



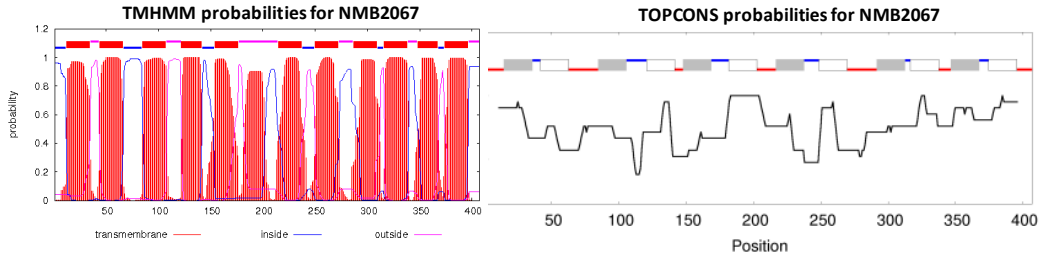
## 5. CAC11736 (*Thermoplasma acidophilum)*



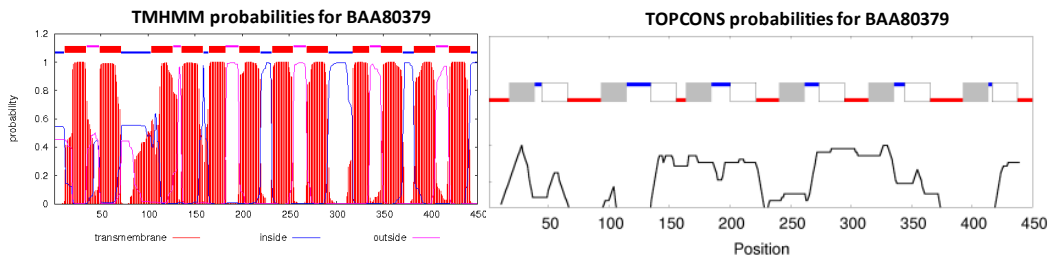

## 6. VPA1242 (*Vibrio parahaemolyticus)*



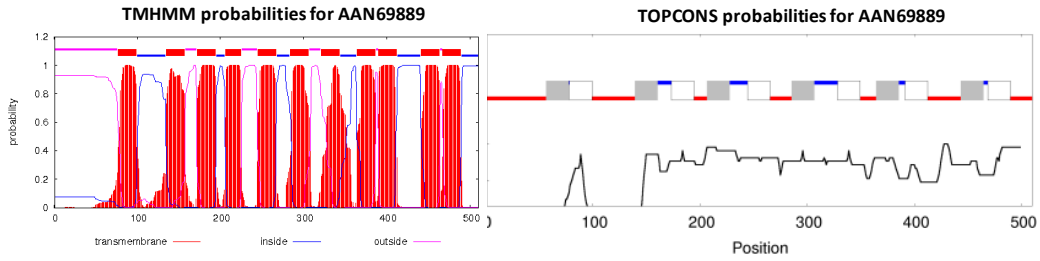

## 7. NMB2067 (*Neisseria meningitides)*



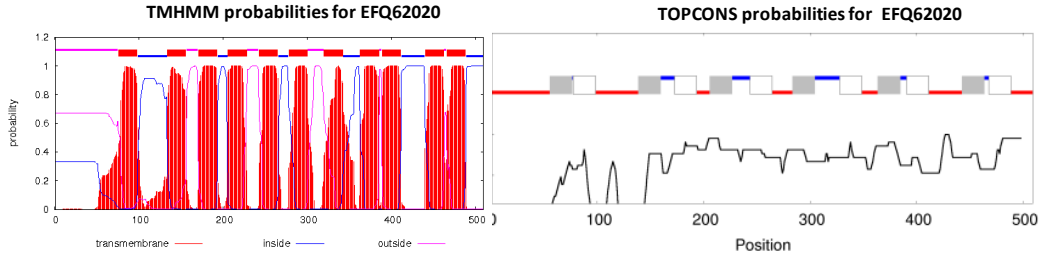

## 8. BAA80379 (*Aeropyrum pernix)*



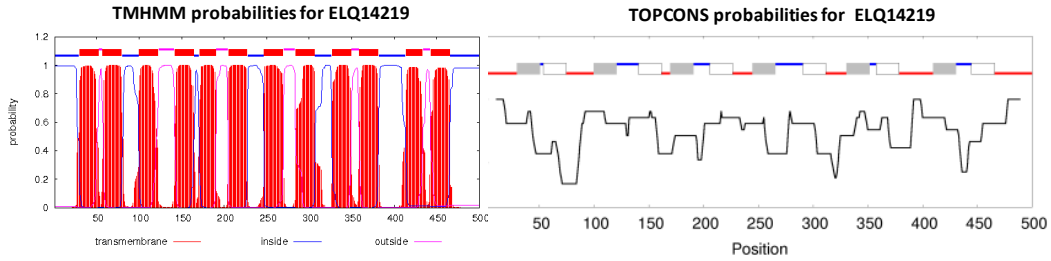
## 9. AAN69889 (*Pseudomonas putida)*




## 10. EFQ62020 (*Pseudomonas fluorescens)*




## 11. ELQ14219 (*Pseudomonas syringae)*



**Appendix 4 DNA sequences of the recombinant NCS1 clones.** The N-terminal MNSH (blue) and C-terminal $His_6$-tag (red).

**CAC11736**

atgaattcgcatatgacgcattctaccgatatgagcaccgaacatgagttcggatatgat
M N S H M T H S T D M S T E H E F G Y D
atcgataaatacagcaaagcaaacaaggaaatagacatcgaaaacgacaagcacccatca
I D K Y S K A N K E I D I E N D K H P S
tccatgttctacatatggtttgcttctaacctgaccgttggagactttgcggttggtttt
S M F Y I W F A S N L T V G D F A V G F
attcctgtgtatcttgggttgcctatactttactccatcgttgcaatagcggttgggacc
I P V Y L G L P I L Y S I V A I A V G T
attgcaggtggcataatgcttgcttatatgagccagctgggcgcagtctacagggtacca
I A G G I M L A Y M S Q L G A V Y R V P
cagatgtacatggggcgtgccccatttggtacagtagggggtctctattatccatcctg
Q M Y M G R A P F G T V G G S L L S I L
caatggggaaataccgcaggctggttgaccgtgaacgtgattctggcatcactggccatt
Q W G N T A G W L T V N V I L A S L A I
tatcaaatggtgaaaataccgtattatgccatagtgcttctcattgttgctatagtcgcg
Y Q M V K I P Y Y A I V L L I V A I V A
ctaactgctcttgttgggtacagagctataaggatccttgagagatccatgtcatatgtc
L T A L V G Y R A I R I L E R S M S Y V
cttggattgctatttgtttttcttgtcctgtttcttatccattcccatgcatctatatcg
L G L L F V F L V L F L I H S H A S I S
taccagtatgtggattcgtttagcattcccgcagccttcggtataacattcgcttcagca
Y Q Y V D S F S I P A A F G I T F A S A
ttctcctatataatgtcttggggaccttatgcagcggactattccaggcatgtatcatca
F S Y I M S W G P Y A A D Y S R H V S S
agcaagccttcaaaatcgatcttctattacacactcctgggttcagtgattgcctctgta
S K P S K S I F Y Y T L L G S V I A S V
tttgcggagattgtgggactcatggtttcggcagcctcaggaaatccttccgggtctccg
F A E I V G L M V S A A S G N P S G S P
gcagcagatctggcaaatgtcatgaacaaatatgcggttatagggctaatagccctgttt
A A D L A N V M N K Y A V I G L I A L F
ctgggtgggatatcggcaaatgcaataaacctctattctaattctcttgctttcctctca
L G G I S A N A I N L Y S N S L A F L S
ataggctttaaggcccatagatgggttgcaatagcggtggcttcagtttttttcgctttta
I G F K A H R W V A I A V A S V F S L L
cttggcataataggcttcttccggttttatggcttctatgagacattcctattcatactt
L G I I G F F R F Y G F Y E T F L F I L
gattattggattacgccatggataggcataatgatagcacacttcttcatattgaagaga
D Y W I T P W I G I M I A H F F I L K R
ttgagacatcaggagttcgaagatctgccaaagatagtgaaaccgggaatatatgcatac
L R H Q E F E D L P K I V K P G I Y A Y
atcatttccattgccgtatcgataccgtttatgtcccctgcgggaataataaatatgcca
I I S I A V S I P F M S P A G I I N M P
ctagcttcaatgctgcatggagtcgatataagctatttcgtttcgtttttctcagcgatg
L A S M L H G V D I S Y F V S F F S A M
ataatttacctctatctgtcccgaaaattggatgaccgcaatcatggagaaaagactcgc
I I Y L Y L S R K L D D R N H G E K T R
aaagagatcagcgctgcaggcggtcgtggcagccaccatcaccatcaccat
K E I S A A G G R G S H H H H H H

**VPA1242**

atgaattcgcatatggctggagacaataactacagtcttggaccagttcccaacacggcc
M N S H M A G D N N Y S L G P V P N T A
agaaaaggcgtggcgtcactaaccatggtgatgttaggactcacttttttctccgcaagt
R K G V A S L T M V M L G L T F F S A S
atgtggacaggtggttctctcggtactgggctctcctttaatgatttttttcctcgctgtt
M W T G G S L G T G L S F N D F F L A V
ctcatcggtaacctaatcctcggtatttacacttctttcctcggttacatcggcgcttct
L I G N L I L G I Y T S F L G Y I G A S
actggtctctctactcaccttcttgctcgtttctcttttggttctaaaggctcttggctt
T G L S T H L L A R F S F G S K G S W L
ccttctgctcttcttggcggtactcaagttgggtggtttggagttggtgtcgcgatgttt
P S A L L G G T Q V G W F G V G V A M F
gcgattccggtgcataaagcaacgggcatcgataccaatactttgatccttgtctccggt
A I P V H K A T G I D T N T L I L V S G
ttgctcatgaccgcgaccgtttactttggaatctcagcactgatggtgctgtcagcaatc
L L M T A T V Y F G I S A L M V L S A I
gccgttcctgccattgcgctacttggcggttactctgtcgtagaagcagtaaatagcgtt
A V P A I A L L G G Y S V V E A V N S V
ggtggtattcgcgaactacaacaagttcaaccaaccgaaccactcgacttttcaatggct
G G I R E L Q Q V Q P T E P L D F S M A
ctcgcgatggttgtaggctcttttgtcagtgcaggcacattaacggcagatttcgtacgc
L A M V V G S F V S A G T L T A D F V R
tttggtaaaaagccgcgcagcgccgtgatgatcaccatggtggcatttttcatcggtaac
F G K K P R S A V M I T M V A F F I G N
tcgttaatgtttatctttggcgcggcaggtgcgtcggtgactggtcaatcagacatctct
S L M F I F G A A G A S V T G Q S D I S
gaagtgatgattgcacaaggtttattgattcctgccatcatcgtacttgggttaaacatt
E V M I A Q G L L I P A I I V L G L N I
tggacgaccaacgacaacgcgctctacgcatcaggtcttggattttccaacatcactggc
W T T N D N A L Y A S G L G F S N I T G
ttaccaagtaaatacatctcgatggcgaacggtcttgttggcacactttgcgctctctgg
L P S K Y I S M A N G L V G T L C A L W
ctctacaataactttgtaggctggctgaccttcttatcgttggcaattccaccaattggc
L Y N N F V G W L T F L S L A I P P I G
ggcgtgatcatcgctgacttcttcacgaaccgtaaacgctacgcaaactttgaagccgcc
G V I I A D F F T N R K R Y A N F E A A
cagttccaaagcgtaaactgggctggcattatcgctgtagcgattggcgttggtgcggga
Q F Q S V N W A G I I A V A I G V G A G
cacttccttcctggcgtagtaccgattaacgctgtacttggcggcgcaattagcttccta
H F L P G V V P I N A V L G G A I S F L
attcttaatcctattctcaacaaaaaagtactggctacacagccagcggctgcaggcggt
I L N P I L N K K V L A T Q P A A A G G
cgtggcagccaccatcaccatcaccat
R G S H H H H H H

**NMB2067**

atgaattcgcatatgtcgggcaatgcctcctctccttcatcttcctccgccatcgggctg
M N S H M S G N A S S P S S S S A I G L
atttggttcggcgcggcggtatcgattgccgaaatcagcacgggtacgctgcttgcgcct
I W F G A A V S I A E I S T G T L L A P
ttgggctggcagcgcggtctggcggctctacttttgggtcatgccgtcggcggcgcgctg
L G W Q R G L A A L L L G H A V G G A L
ttttttgcggcggcgtatatcggcgcactgaccggacgcagctcgatggaaagcgtgcgc
F F A A A Y I G A L T G R S S M E S V R
ctgtcgttcggcaaacgcggttcagtgctgttttccgtggcgaatatgctgcaactggcc
L S F G K R G S V L F S V A N M L Q L A
ggctggacggcggtgatgatttacgccggcgcaacggtcagctccgctttgggcaaagtg
G W T A V M I Y A G A T V S S A L G K V

ttgtgggacggcgaatcttttgtctggtgggcattggcaaacggcgcgctgattgtgctg
 L  W  D  G  E  S  F  V  W  W  A  L  A  N  G  A  L  I  V  L
tggctggttttcggcgcacgcaaaacaggcgggctgaaaaccgtttcgatgctgctgatg
 W  L  V  F  G  A  R  K  T  G  G  L  K  T  V  S  M  L  L  M
ctgttggcggttctgtggctgagtgccgaagtcttttccacggcaggcagcaccgccgca
 L  L  A  V  L  W  L  S  A  E  V  F  S  T  A  G  S  T  A  A
caggtttcagacggcatgagtttcggaacggcagtcgagctgtccgccgtgatgccgctt
 Q  V  S  D  G  M  S  F  G  T  A  V  E  L  S  A  V  M  P  L
tcctggctgccgcttgccgccgactacacgcgccacgcgcgccgcccgtttgcggcaacc
 S  W  L  P  L  A  A  D  Y  T  R  H  A  R  R  P  F  A  A  T
ctgacggcaacgctcgcctacacgctgaccggctgctggatgtatgccttgggtttggca
 L  T  A  T  L  A  Y  T  L  T  G  C  W  M  Y  A  L  G  L  A
gcggcgttgttcaccggagaaaccgacgtggcaaaaatcctgctgggcgcaggtttgggt
 A  A  L  F  T  G  E  T  D  V  A  K  I  L  L  G  A  G  L  G
gcggcaggcattttggcggtcgtcctctccaccgttaccacaacgtttctcgatgcctat
 A  A  G  I  L  A  V  V  L  S  T  V  T  T  T  F  L  D  A  Y
tccgccggcgcgagtgcgaacaacatttccgcgcgttttgcggaaacacccgtcgctgtc
 S  A  G  A  S  A  N  N  I  S  A  R  F  A  E  T  P  V  A  V
ggcgttaccctgatcggcacggtacttgccgtcatgctgcccgttaccgaatatgaaaac
 G  V  T  L  I  G  T  V  L  A  V  M  L  P  V  T  E  Y  E  N
ttcctgctgcttatcggctcggtatttgcgccgatggcggcggttttgattgccgacttt
 F  L  L  L  I  G  S  V  F  A  P  M  A  A  V  L  I  A  D  F
ttcgtcttgaaacggcgtgaggagattgaaggctttgactttgccggactggttctgtgg
 F  V  L  K  R  R  E  E  I  E  G  F  D  F  A  G  L  V  L  W
cttgcgggcttcatcctctaccgcttcctgctctcgtccggctgggaaagcagcatcggt
 L  A  G  F  I  L  Y  R  F  L  L  S  S  G  W  E  S  S  I  G
ctgaccgccccgtaatgtctgccgttgccattgccaccgtatcggtacgccttttcttt
 L  T  A  P  V  M  S  A  V  A  I  A  T  V  S  V  R  L  F  F
aaaaaaacccaatctttacaaaggaacccgtcagctgcaggcggtcgtggcagccaccat
 K  K  T  Q  S  L  Q  R  N  P  S  A  A  G  G  R  G  S  H  H
caccatcaccat
 H  H  H  H

**BAA80379**

atgaattcgcatatgggtctgggcaataccccgagaaaggaggggcctaggagccctt
 M  N  S  H  M  G  L  G  N  T  P  E  K  G  G  G  L  G  A  L
gagtatacttttattatgttcagcatggcatcctgcctgccacttttcttcctaggcccc
 E  Y  T  F  I  M  F  S  M  A  S  C  L  P  L  F  F  L  G  P
atagccttcaacctcggcctctcgcttcaagaggctctactggcagcactggtaggaaac
 I  A  F  N  L  G  L  S  L  Q  E  A  L  L  A  A  L  V  G  N
ctggtagtggcagtggcaatggcgctcaacgggcatgcagggattaagcacaagatagac
 L  V  V  A  V  A  M  A  L  N  G  H  A  G  I  K  H  K  I  D
tttccagagcaggccgtaagaagcctaggagagctcactggcaaggcggctgtggtgatg
 F  P  E  Q  A  V  R  S  L  G  E  L  T  G  K  A  A  V  V  M
agagggctagtgggtgctatgtggttcggggtggaagcctacaacggcgccctagcccta
 R  G  L  V  G  A  M  W  F  G  V  E  A  Y  N  G  A  L  A  L
aacctaatactcctcttcgccctagggcttactggagcagcgctgctggagaaggccacg
 N  L  I  L  L  F  A  L  G  L  T  G  A  A  L  L  E  K  A  T
gtcctcataccagcagctctagtactctaccttggctccatgtatctagtcctaaagcta
 V  L  I  P  A  A  L  V  L  Y  L  G  S  M  Y  L  V  L  K  L
ggggtaaagggtatagggaaggctgccaccctcgcgggacccctttctgctcctctacttc
 G  V  K  G  I  G  K  A  A  T  L  A  G  P  L  L  L  L  Y  F
gcatggctctggatctggatgaagaactctgggttccagccgtccgaggcacccaagggt

A W L W I W M K N S G F Q P S E A P K G
gtaggcttgctaagctcagccttcctaatctacctggcgatacagacaaactggtgggcg
V G L L S S A F L I Y L A I Q T N W W A
accgttgcggtgaacataagcgacctttcccgtgaggccaagagctggggcgccctctgg
T V A V N I S D L S R E A K S W G A L W
atcggagtcatgcttggcatggtgggaggccagcttataggtacctacctcagctacgag
I G V M L G M V G G Q L I G T Y L S Y E
ctagtcttgctaaccgggaaaacgctaccgcaggagatcataacagagttcgctccagga
L V L L T G K T L P Q E I I T E F A P G
gctatagcagtccttctagggctggccttcgcctttctcgcaccctggacaacagatcta
A I A V L L G L A F A F L A P W T T D L
acggccaatctacccccgatgatagatatactgaaaagcattttcgggatgagctggaag
T A N L P P M I D I L K S I F G M S W K
agggccagcctcctctcggcagcggcaggcttcgtgctggctccatggtggctgctggac
R A S L L S A A A G F V L A P W W L L D
aacgcaccccagatagtaggctatgtaacctccttcgcagcaagctacggtgtaatacta
N A P Q I V G Y V T S F A A S Y G V I L
ggcccgatactcggtgcactcctcgcagcacactgggtaggaggcctaaacaggcccccc
G P I L G A L L A A H W V G G L N R P P
aacccaagttacaagccagttatactacccactacagccggcctactagcaggcctcata
N P S Y K P V I L P T T A G L L A G L I
gtctcatacgcaatagcctacccgctgggtatggtaaccagcgtcctcggagtaccattc
V S Y A I A Y P L G M V T S V L G V P F
ccgcaggggccaatatggtacgtgggagtcgccgtgagtgtgattgcagccgccctgctt
P Q G P I W Y V G V A V S V I A A A L L
ctcaaacttgtcccagctaaagttaagttcaagaacgagacggctgcaggcggtcgtggc
L K L V P A K V K F K N E T A A G G R G
agccaccatcaccatcaccat
S H H H H H H

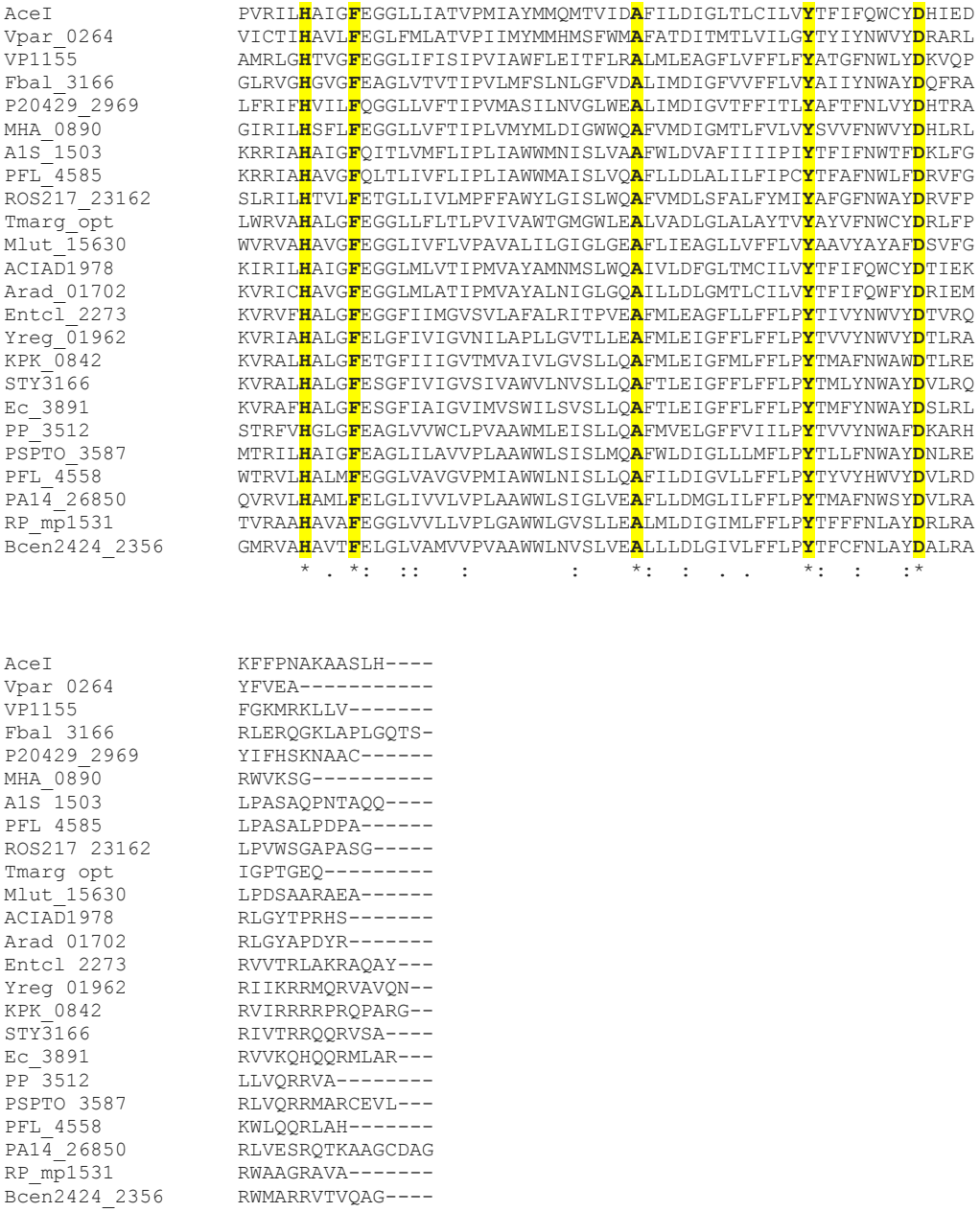
**AAN69889**

atgaattcgcatatgagtagcagcctcgaccttgcccctgaactatccgtcgccagcaca
M N S H M S S S L D L A P E L S V A S T
caccccgcgtccacgcttgccggccaccagccggaccttgttctcagcccgcgcctgcat
H P A S T L A G H Q P D L V L S P R L H
aaccgcgacctcgcgccaacgcgtatggaaggccgtcgctggggcggctacagcatcttc
N R D L A P T R M E G R R W G G Y S I F
gcgctgtggacaaacgatgtgcacaatatcgccaactattcgttcgccatgggcttgttc
A L W T N D V H N I A N Y S F A M G L F
gccctcggcctgggtggctggcaaattctgctgtcgctggccatcggcgcggcactggtg
A L G L G G W Q I L L S L A I G A A L V
tacttcttcatgaacctgtcgggttacatggggcagaaaaccggggtaccgttcccggtc
Y F F M N L S G Y M G Q K T G V P F P V
atcagccgcattgccttcggcatccacggcgcgcagatcccggcgctgatccgtgccgtc
I S R I A F G I H G A Q I P A L I R A V
atcgccatcgcctggttcggcatccagacctacctggcatcggtggtgctgcgggtgctg
I A I A W F G I Q T Y L A S V V L R V L
ctcaccgccgtttggccgcaaatagcagcctacgaccacgacagcatcctcggcctgtcg
L T A V W P Q I A A Y D H D S I L G L S
agcctgggctgggtgtgtttcgtgtcgatctggctggtgcagctggtgatcctggcctac
S L G W V C F V S I W L V Q L V I L A Y
ggcatggagatggtgcgccgttacgaggcctttgccgggcctgtgatcctgctgaccgtc
G M E M V R R Y E A F A G P V I L L T V

```
gccgcactggcggtgttcatgtacttcaaggccgacgcgcgcattgcctggtcggtggct
 A  A  L  A  V  F  M  Y  F  K  A  D  A  R  I  A  W  S  V  A
acgccgctgaccggctacgagatgtggcgcaacatctttgccggcggcgcactgtggctg
 T  P  L  T  G  Y  E  M  W  R  N  I  F  A  G  G  A  L  W  L
gcgatctacggcaccctggtgctgaacttctgtgacttcgcccgctcctcgccttgccgc
 A  I  Y  G  T  L  V  L  N  F  C  D  F  A  R  S  S  P  C  R
aagaccatccgcgtgggcaacttctggggcctgccggtaaacatcctggtattcgccgtg
 K  T  I  R  V  G  N  F  W  G  L  P  V  N  I  L  V  F  A  V
atcaccgtggtgctgtgcggcgcgcaattccagatcaacggccagatcatcgacagcccg
 I  T  V  V  L  C  G  A  Q  F  Q  I  N  G  Q  I  I  D  S  P
acgcagatcgttgccgccatacccagcacgccattcctggtgctcggctgcctggccttc
 T  Q  I  V  A  A  I  P  S  T  P  F  L  V  L  G  C  L  A  F
ctgatcgtcaccgtagcggtgaacatcatggccaacttcgtcgccccggctttcgtactc
 L  I  V  T  V  A  V  N  I  M  A  N  F  V  A  P  A  F  V  L
agcaacctggcgccgcgccacctgaacttccgccgtgccgggctgatcagcgccaccctg
 S  N  L  A  P  R  H  L  N  F  R  R  A  G  L  I  S  A  T  L
gcggtgctgatcctgccctggaacctgtacaacagcccgctggtgatcgtgtacttcctg
 A  V  L  I  L  P  W  N  L  Y  N  S  P  L  V  I  V  Y  F  L
tcgggcctgggcgccctgctaggcccgctgtacggcgtgatcatgtccgactactggttg
 S  G  L  G  A  L  L  G  P  L  Y  G  V  I  M  S  D  Y  W  L
ctgcgcaaaggctgcatcaacgtgccggagctgtataccgagcacccggccggcgcctat
 L  R  K  G  C  I  N  V  P  E  L  Y  T  E  H  P  A  G  A  Y
cactacagcaagggcatcaacctgcgcgccgtggccgccttcgtgccggccgcactgctg
 H  Y  S  K  G  I  N  L  R  A  V  A  A  F  V  P  A  A  L  L
gccatcgtgctggccctggtgcccaacttccagggcatcgcgccgttctcctggctgatt
 A  I  V  L  A  L  V  P  N  F  Q  G  I  A  P  F  S  W  L  I
ggtgccggcatcgccgccgccctatacctgcttatcgcgccacgtaaccgccattaccac
 G  A  G  I  A  A  A  L  Y  L  L  I  A  P  R  N  R  H  Y  H
gacgtcagcggcgagtgcatcgccgtcgaccacagcggccatgctgcaggcggtcgtggc
 D  V  S  G  E  C  I  A  V  D  H  S  G  H  A  A  G  G  R  G
agccaccatcaccatcaccat
 S  H  H  H  H  H  H
```

**Appendix 5 Amino acids sequence alignment of AceI and its homologues.** Using multiple sequence alignment tool "Clustal W" AceI and its homologues amino acids sequence were aligned. The conserved residues are shown in yellow.

```
AceI                ---------------------------------------------MLISKRRLIHAISYEG
Vpar_0264           ---------------------------------------------MSATERVVQSILYEL
VP1155              ---------------------------------------------MTRNERIFHAVLFEL
Fbal_3166           ---------------------------------------------MSPRERVLHSLLFEM
P20429_2969         ------------------------------------------MDAKMGSLERMFQAVLFEV
MHA_0890            ---------------------------------------------MTAWERVFHALLFEC
A1S_1503            ---------------------------------------------MQGLKRRIVYVSSYEI
PFL_4585            ---------------------------------------------MQGVKRKLVYVSLFEV
ROS217_23162        ---------------------------------------------MRSPLDRLRHALSFEI
Tmarg_opt           --------------------MEERNQA--------VESKQDGVALRTWRDRVRQVLAFEA
Mlut_15630          MSETHRARPGVDPASAPPSDVPRGPGGRDAL-------VRRRVFRTPLLRRVVYAVVFEL
ACIAD1978           ------------------------------MHLIQI---KECFKMLISKRRMIHALSYEV
Arad_01702          ------------------------------------------MEVMLLSKRRLIHALSYEI
Entcl_2273          ---------------------------------------MQQKSMQHRTLAERIFHAVSFEG
Yreg_01962          ------------------------------MKLRFIFMQQNNLHRRSILERIFHAIGFEG
KPK_0842            ---------------------------------------MQQIPHQRKTLTERVIHAITFEG
STY3166             ----------------------------MIKSKVFSFMQHDAIQRRSLPERIFHAVCFEG
Ec_3891             ---------------------------------------MQHNAIQRRSLLERIFHAICFEG
PP_3512             ------------------------------------MECAMKNVSFTERLVHAVGYEV
PSPTO_3587          --------------------MNEQKDSLNMTHREQPTRQQPPVVHKTIRERALHATLFEV
PFL_4558            ------------------------------------------MSLQKSLNERIFQAVGFEL
PA14_26850          --------------MTIQPDFNDEPGAFAMT-------HTHTALDKTLKERVFHALAFEG
RP_mp1531           ---------------------------------------MSVPQKTPLERVCHALAFEL
Bcen2424_2356       ---------------------------------------MKQINKTVTERLLHALTFEL
                                                                  :     :*

AceI                ILLVIIAIALSFIFNMPMEVTGTLGVFMAVVSVFWNMIFNHYFEKVEHKYNWE----RTI
Vpar_0264           GCVLIGCLVMQFVPH--EGQPLVLMIIFSLLAMVWNFVFNWIFDKLVPGDRLA----RGP
VP1155              MALAIIVPAAALITGKGSSDLALVGIGLSLYTVVWNYIYNLYFDKWFGSNRAD----RSL
Fbal_3166           IALAIVVPAGILLTGTDAGHMTATAIILSLMAMAWNYVYNLGFDRVFGSNRID----RSW
P20429_2969         LAVTLSIIGLAVFTEHAISALSGTMIIIATIAMCWNFVFNWFFDKVATGAKEQ----RSV
MHA_0890            FAIIFTVILTSWLTSHRMVDLTAVIVMISVIAMVWNVIFNWGFDKVFTGERVQ----RGL
A1S_1503            IGMVISSVGLALLAGDSVEHTGPLSVMITTIAVTWNFIYNILYEKWEARQESK---SRTV
PFL_4585            FGMTFSALGLALLSGTSPSSTGPLAVIITSIAVTWNFIYTTLFERWESRQPSR---TRTV
ROS217_23162        IALLLVVPLGAVAFHVPIHDIGVVGIVSATLATLWNMIYNYVFDVALQRLSGT--TKKSA
Tmarg_opt           GGLVLITPPFSWASGAPVLDSLGLLALLALIAALWNALYNTVFDRLEARFLHRRADLRPP
Mlut_15630          LAILFTTGILAALGNGG-GPSLAVAVVSSTVALLWNIVFNSVFETLERRLGIT---GRPW
ACIAD1978           ILLVIIAIALSFIFDVPLEVTGTLGIVMAVTSVFWNMIFNHFFEKFERKHQLE----RTV
Arad_01702          ILLIIIAIALSFLFEIPMEVSGTLGVVMAITSVIWNMIFNHFFEKFEHRHQLK----RTI
Entcl_2273          LATLILAPTAAWIMQRSVLEMGGVTVLLATIAMVWNIIYNAIFDRFWPVSRVA----RTA
Yreg_01962          LATLVLAPTAAWLMQRSVVEMGGVAILLATLAMVWNIIYNAGFDKLWPVTRVT----RTA
KPK_0842            LATLILAPTAAWLMQRSVVEMGGLSILLATLAMVWNIIYNAMFDRLWPVSRFP----RQL
STY3166             IATAILAPTTAWLMQRSVLEMGGLTILLATTAMIWNIIYNALFDRLWPAHQVR----RTA
Ec_3891             IATAILAPTTAWLMQRTILEMGGMTILLASTAMIWNIIYNMLFDRYWPAHQVK----RTA
PP_3512             FAVLLCAPLLSWIMGKSLATAGTLAVTLSVIAMLWNMVYNVLVDRWVQTERIN----WKA
PSPTO_3587          GGVILVAPLLAWIMNQSLAMIGIMTVMISTVAMLWNMVYNALFDRLRSRFGFA----MTA
PFL_4558            LAVLICTPLLSWIMDKPMADMGLVTIGIGLLALAWNVVFNGLFDRLLKRLQWV----HNG
PA14_26850          LAVLLTAPVLSLVMNKPLAHMGALTLMFSTVAMLWNMLFNSLFDRAQRRMGFQ----RTL
RP_mp1531           IATLICAPLLSWLMALPLMQMGALTILFALVAMAWNMVFNAGFERIERRCGWA----RTL
Bcen2424_2356       VAIALCAPIGAWLLDMPVSHVGVLTVMVSLIAMAWNMTFNTLFDRFERRAGLS----RTL
                        .                          :  **  :.   :
```


AceI PVRILHAIGFEGGLLIATVPMIAYMMQMTVIDAFILDIGLTLCILVYTFIFQWCYDHIED
Vpar_0264 VICTIHAVLFEGLFMLATVPIIMYMMHMSFWMAFATDITMTLVILGYTYIYNWVYDRARL
VP1155 AMRLGHTVGFEGGLIFISIPVIAWFLEITFLRALMLEAGFLVFFLFYATGFNWLYDKVQP
Fbal_3166 GLRVGHGVGFEAGLVTVTIPVLMFSLNLGFVDALIMDIGFVVFFLVYAIIYNWAYDQFRA
P20429_2969 LFRIFHVILFQGGLLVFTIPVMASILNVGLWEALIMDIGVTFFITLYAFTFNLVYDHTRA
MHA_0890 GIRILHSFLFEGGLLVFTIPLVMYMLDIGWWQAFVMDIGMTLFVLVYSVVFNWVYDHLRL
A1S_1503 KRRIAHAIGFQITLVMFLIPLIAWWMNISLVAAFWLDVAFIIIIPIYTFIFNWTFDKLFG
PFL_4585 KRRIAHAVGFQLTLIVFLIPLIAWWMAISLVQAFLLDLALILFIPCYTFAFNWLFDRVFG
ROS217_23162 SLRILHTVLFETGLLIVLMPFFAWYLGISLWQAFVMDLSFALFYMIYAFGFNWAYDRVFP
Tmarg_opt LWRVAHALGFEGGLLFLTLPVIVAWTGMGWLEALVADLGLALAYTVYAYVFNWCYDRLFP
Mlut_15630 WVRVAHAVGFEGGLIVFLVPAVALILGIGLGEAFLIEAGLLVFFLVYAAVYAYAFDSVFG
ACIAD1978 KIRILHAIGFEGGLMLVTIPMVAYAMNMSLWQAIVLDFGLTMCILVYTFIFQWCYDTIEK
Arad_01702 KVRICHAVGFEGGLMLATIPMVAYALNIGLGQAILLDLGMTLCILVYTFIFQWFYDRIEM
Entcl_2273 KVRVFHALGFEGGFIIMGVSVLAFALRITPVEAFMLEAGFLLFFLPYTIVYNWVYDTVRQ
Yreg_01962 KVRIAHALGFELGFIVIGVNILAPLLGVTLLEAFMLEIGFFLFFLPYTVVYNWVYDTLRA
KPK_0842 KVRALHALGFETGFIIIGVTMVAIVLGVSLLQAFMLEIGFMLFFLPYTMAFNWAWDTLRE
STY3166 KVRALHALGFESGFIVIGVSIVAWVLNVSLLQAFTLEIGFFLFFLPYTMLYNWAYDVLRQ
Ec_3891 KVRAFHALGFESGFIAIGVIMVSWILSVSLLQAFTLEIGFFLFFLPYTMFYNWAYDSLRL
PP_3512 STRFVHGLGFEAGLVVWCLPVAAWMLEISLLQAFMVELGFFVIILPYTVVYNWAFDKARH
PSPTO_3587 MTRILHAIGFEAGLILAVVPLAAWWLSISLMQAFWLDIGLLLMFLPYTLLFNWAYDNLRE
PFL_4558 WTRVLHALMFEGGLVAVGVPMIAWWLNISLLQAFILDIGVLLFFLPYTYVYHWVYDVLRD
PA14_26850 QVRVLHAMLFELGLIVVLVPLAAWWLSIGLVEAFLLDMGLILFFLPYTMAFNWSYDVLRA
RP_mp1531 TVRAAHAVAFEGGLVVLLVPLGAWWLGVSLLEALMLDIGIMLFFLPYTFFFNLAYDRLRA
Bcen2424_2356 GMRVAHAVTFELGLVAMVVPVAAWWLNVSLVEALLLDLGIVLFFLPYTFCFNLAYDALRA
AceI KFFPNAKAASLH----
Vpar_0264 YFVEA----------
VP1155 FGKMRKLLV-------
Fbal_3166 RLERQGKLAPLGQTS-
P20429_2969 YIFHSKNAAC------
MHA_0890 RWVKSG----------
A1S_1503 LPASAQPNTAQQ----
PFL_4585 LPASALPDPA------
ROS217_23162 LPVWSGAPASG-----
Tmarg_opt IGPTGEQ---------
Mlut_15630 LPDSAARAEA------
ACIAD1978 RLGYTPRHS-------
Arad_01702 RLGYAPDYR-------
Entcl_2273 RVVTRLAKRAQAY---
Yreg_01962 RIIKRRMQRVAVQN--
KPK_0842 RVIRRRRPRQPARG--
STY3166 RIVTRRQQRVSA----
Ec_3891 RVVKQHQQRMLAR---
PP_3512 LLVQRRVA--------
PSPTO_3587 RLVQRRMARCEVL---
PFL_4558 KWLQQRLAH-------
PA14_26850 RLVESRQTKAAGCDAG
RP_mp1531 RWAAGRAVA-------
Bcen2424_2356 RWMARRVTVQAG----

# References

Aboulwafa, M. and Saier Jr, M.H. 2013. Lipid dependencies, biogenesis and cytoplasmic micellar forms of integral membrane sugar transport proteins of the bacterial phosphotransferase system. *Microbiology.* **159**(11), pp.2213-2224.

Abramson, J. and Wright, E.M. 2009. Structure and function of $Na^+$-symporters with inverted repeats. *Current opinion in structural biology.* **19**(4), pp.425-432.

Adelman, J.L., Dale, A.L., Zwier, M.C., Bhatt, D., Chong, L.T., Zuckerman, D.M. and Grabe, M. 2011. Simulations of the alternating access mechanism of the sodium symporter Mhp1. *Biophysical journal.* **101**(10), pp.2399-2407.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. 2002. Molecular biology of the cell. new york: Garland science; 2002. *Classic textbook now in its 5th Edition.*

Alberts, B., Johnson, A., Walter, P., Lewis, J., Raff, M. and Roberts, K. 2008. Molecular cell biology. *New York and London: Garland Science.* **4**.

Alegre, K.O. and Law, C.J. 2015. Purification of a multidrug resistance transporter for crystallization studies. *Antibiotics.* **4**(1), pp.113-135.

Alibert-Franco, S., Mahamoud, A., Bolla, J.-M., Davin-Regli, A., Chevalier, J. and Garnotel, E. 2010. Efflux pumps of gram-negative bacteria, a new target for new molecules. *Current topics in medicinal chemistry.* **10**(18), pp.1848-1857.

Altenbuchner, J., Siemann-Herzberg, M. and Syldatk, C. 2001. Hydantoinases and related enzymes as biocatalysts for the synthesis of unnatural chiral amino acids. *Current Opinion in Biotechnology.* **12**(6), pp.559-563.

Alvarez-Ortega, C., Olivares, J. and Martínez, J.L. 2013. RND multidrug efflux pumps: what are they good for? *Frontiers in microbiology.* **4**.

Andersen, K.R., Leksa, N.C. and Schwartz, T.U. 2013. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins: Structure, Function, and Bioinformatics.* **81**(11), pp.1857-1861.

Anes, J., McCusker, M.P., Fanning, S. and Martins, M. 2015. The ins and outs of RND efflux pumps in *Escherichia coli*. *Frontiers in microbiology.* **6**.

Angov, E. 2011. Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology journal.* **6**(6), pp.650-659.

Arinaminpathy, Y., Khurana, E., Engelman, D.M. and Gerstein, M.B. 2009. Computational analysis of membrane proteins: the largest class of drug targets. *Drug discovery today.* **14**(23), pp.1130-1135.

Arutyunov, D. and Frost, L.S. 2013. F conjugation: back to the beginning. *Plasmid.* **70**(1), pp.18-32.

Ashworth, J. and Kornberg, H.L. 1966. The anaplerotic fixation of carbon dioxide by *Escherichia coli*. *Proceedings of the Royal Society of London B: Biological Sciences.* **165**(999), pp.179-188.

Auer, M., Kim, M.J., Lemieux, M.J., Villa, A., Song, J., Li, X.-D. and Wang, D.-N. 2001. High-yield expression and functional analysis of *Escherichia coli* glycerol-3-phosphate transporter. *Biochemistry.* **40**(22), pp.6628-6635.

Bartlett, J.G., Gilbert, D.N. and Spellberg, B. 2013. Seven ways to preserve the miracle of antibiotics. *Clinical Infectious Diseases.* **56**(10), pp.1445-1450.

Bay, D.C., Rommens, K.L. and Turner, R.J. 2008. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1778**(9), pp.1814-1838.

Bay, D.C. and Turner, R.J. 2009. Diversity and evolution of the small multidrug resistance protein family. *BMC evolutionary biology.* **9**(1), p140.

Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.-B., Salvi, D., Seigneurin-Berny, D. and Richaud, P. 2011. Heterologous expression of membrane proteins: choosing the appropriate host. *PloS one.* **6**(12), pe29191.

Bernsel, A., Viklund, H., Hennerdal, A. and Elofsson, A. 2009. TOPCONS: consensus prediction of membrane protein topology. *Nucleic acids research.* **37**(suppl_2), pp.W465-W468.

Berridge, M.J. and Irvine, R.F. 1989. Inositol phosphates and cell signalling. *Nature.* **341**(6239), pp.197-205.

Bettaney K (2008) Characterisation of membrane transport proteins of bacteria – from gene to crystal. PhD Thesis, University of Leeds, Leeds, UK.

Bettaney, K.E., Sukumar, P., Hussain, R., Siligardi, G., Henderson, P.J. and Patching, S.G. 2013. A systematic approach to the amplified expression, functional characterization and purification of inositol transporters from *Bacillus subtilis*. *Molecular membrane biology.* **30**(1), pp.3-14.

Bill, R.M., Henderson, P.J., Iwata, S., Kunji, E.R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G. and Vogel, H. 2011. Overcoming barriers to membrane protein structure determination. *Nature biotechnology.* **29**(4), pp.335-340.

Birnbaum, S. and Bailey, J. 1991. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnology and bioengineering.* **37**(8), pp.736-745.

Blair, J.M. and Piddock, L.J. 2016. How to measure export via bacterial multidrug resistance efflux pumps. *MBio.* **7**(4), pp.e00840-00816.

Blair, J.M., Richmond, G.E. and Piddock, L.J. 2014. Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future microbiology.* **9**(10), pp.1165-1177.

Blair, J.M., Webber, M.A., Baylay, A.J., Ogbolu, D.O. and Piddock, L.J. 2015. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology.* **13**(1), pp.42-51.

Blanco, P., Hernando-Amado, S., Reales-Calderon, J.A., Corona, F., Lira, F., Alcalde-Rico, M., Bernardini, A., Sanchez, M.B. and Martinez, J.L. 2016. Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. *Microorganisms.* **4**(1), p14.

Bolla, J.R., Su, C.-C., Do, S.V., Radhakrishnan, A., Kumar, N., Long, F., Chou, T.-H., Delmar, J.A., Lei, H.-T. and Rajashankar, K.R. 2014. Crystal structure of the *Neisseria gonorrhoeae* MtrD inner membrane multidrug efflux pump. *PloS one.* **9**(6), pe97903.

Bommarius, A.S., Schwarm, M. and Drauz, K. 1998. Biocatalysis to amino acid-based chiral pharmaceuticals—examples and perspectives. *Journal of Molecular Catalysis B: Enzymatic.* **5**(1), pp.1-11.

Booth, P.J. 2003. The trials and tribulations of membrane protein folding in vitro. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1610**(1), pp.51-56.

Borges-Walmsley, M.I., McKEEGAN, K.S. and Walmsley, A.R. 2003. Structure and function of efflux pumps that confer resistance to drugs. *Biochemical Journal.* **376**(2), pp.313-338.

Bornhorst, J.A. and Falke, J.J. 2000. [16] Purification of proteins using polyhistidine affinity tags. *Methods in enzymology.* **326**, pp.245-254.

Bosshard, H.R., Marti, D.N. and Jelesarov, I. 2004. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition.* **17**(1), pp.1-16.

Brochet, M., Couvé, E., Zouine, M., Poyart, C. and Glaser, P. 2008. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *Journal of bacteriology.* **190**(2), pp.672-680.

Brown, D. and London, E. 1998. Functions of lipid rafts in biological membranes. *Annual review of cell and developmental biology.* **14**(1), pp.111-136.

Brown, M.H., Paulsen, I.T. and Skurray, R.A. 1999. The multidrug efflux protein NorM is a prototype of a new family of transporters. *Molecular microbiology.* **31**(1), pp.394-395.

Bulheller, B.M., Rodger, A. and Hirst, J.D. 2007. Circular and linear dichroism of proteins. *Physical Chemistry Chemical Physics.* **9**(17), pp.2020-2035.

Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U. and Gileadi, O. 2008. Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein expression and purification.* **59**(1), pp.94-102.

Busch, W. and Saier, M.H. 2002. The transporter classification (TC) system, 2002. *Critical reviews in biochemistry and molecular biology.* **37**(5), pp.287-337.

Cabrita, M.A., Baldwin, S.A., Young, J.D. and Cass, C.E. 2002. Molecular biology and regulation of nucleoside and nucleobase transporter proteins in eukaryotes and prokaryotes. *Biochemistry and cell biology.* **80**(5), pp.623-638.

Cafiso, D.S. 2005. Structure and interactions of C2 domains at membrane surfaces. *Protein-lipid interactions: from membrane domains to cellular networks.* pp.403-422.

Calos, M.P. 1978. DNA sequence for a low-level promoter of the lac repressor gene and an'up'promoter mutation. *Nature.* **274**(5673), pp.762-765.

Carruthers, A. 1990. Facilitated diffusion of glucose. *Physiological Reviews.* **70**(4), pp.1135-1176.

Casey, R. 2005. BLAST sequences aid in genomics and proteomics. *Business Intelligence Network.*

Cecchetto, G., Amillis, S., Diallinas, G., Scazzocchio, C. and Drevet, C. 2003. The AzgA purine transporter of *Aspergillus nidulans*: characterisation of a protein belonging to a new phylogenetic cluster. *Journal of Biological Chemistry.*

Chaga, G., Hopp, J. and Nelson, P. 1999. Immobilized metal ion affinity chromatography on $Co2^{+}$-carboxymethylaspartate–agarose Superflow, as demonstrated by one-step purification of lactate dehydrogenase from chicken breast muscle. *Biotechnology and applied biochemistry.* **29**(1), pp.19-24.

Chaudhary, A.S. 2016. A review of global initiatives to fight antibiotic resistance and recent antibiotics′ discovery. *Acta Pharmaceutica Sinica B.* **6**(6), pp.552-556.

Chen, J., Morita, Y., Huda, M.N., Kuroda, T., Mizushima, T. and Tsuchiya, T. 2002. VmrA, a member of a novel class of $Na^{+}$-coupled multidrug efflux pumps from *Vibrio parahaemolyticus*. *Journal of bacteriology.* **184**(2), pp.572-576.

Chen, Y. and Barkley, M.D. 1998. Toward understanding tryptophan fluorescence in proteins. *Biochemistry.* **37**(28), pp.9976-9982.

Chitsaz, M. and Brown, M.H. 2017. The role played by drug efflux pumps in bacterial multidrug resistance. *Essays In Biochemistry.* **61**(1), pp.127-139.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Gerstein, M., Arrowsmith, C.H. and Edwards, A.M. 2000. Structural proteomics: prospects for high throughput

sample preparation. *Progress in biophysics and molecular biology.* **73**(5), pp.339-345.

Chung, Y. and Saier Jr, M. 2001. SMR-type multidrug resistance pumps. *Current opinion in drug discovery & development.* **4**(2), pp.237-245.

Cournia, Z., Allen, T.W., Andricioaei, I., Antonny, B., Baum, D., Brannigan, G., Buchete, N.-V., Deckman, J.T., Delemotte, L. and Del Val, C. 2015. Membrane protein structure, function and dynamics: a perspective from experiments and theory. *The Journal of membrane biology.* **248**(4), p611.

Courvalin, P. 1996. The Garrod Lecture Evasion of antibiotic action by bacteria. *Journal of Antimicrobial Chemotherapy.* **37**(5), pp.855-869.

Cowan, S., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R., Jansonius, J. and Rosenbusch, J. 1992. Crystal structures explain functional properties of two *E. coli* porins. *Nature.* **358**(6389), pp.727-733.

Crowe, J., Dobeli, H., Gentz, R., Hochuli, E., Stiiber, D. and Henco, K. 1994. 6xffis-ni-nta chromatography as a superior technique in recombinant protein expression/purification. *Protocols for gene analysis.* pp.371-387.

Crumplin, G. and Odell, M. 1986. Development of resistance to ofloxacin. *Drugs.* **34**, pp.1-8.

Daley, D.O., Rapp, M., Granseth, E., Melén, K., Drew, D. and Von Heijne, G. 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science.* **308**(5726), pp.1321-1323.

Danielsen, S., Kilstrup, M., Barilla, K., Jochimsen, B. and Neuhard, J. 1992. Characterization of the *Escherichia coli* codBA operon encoding cytosine permease and cytosine deaminase. *Molecular microbiology.* **6**(10), pp.1335-1344.

Dassa, E. and Bouige, P. 2001. The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Research in microbiology.* **152**(3), pp.211-229.

Davies, J. 1994. Inactivation of antibiotics and the dissemination of resistance genes. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information.* **264**(5157), pp.375-381.

Davies, J. and Davies, D. 2010. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews.* **74**(3), pp.417-433.

Davis, H.L., Whalen, R.G. and Demeneix, B.A. 1993. Direct gene transfer into skeletal muscle in vivo: factors affecting efficiency of transfer and stability of expression. *Human gene therapy.* **4**(2), pp.151-159.

De Boer, H.A., Comstock, L.J. and Vasser, M. 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences.* **80**(1), pp.21-25.

de Koning, H. and Diallinas, G. 2000. Nucleobase transporters. *Molecular membrane biology.* **17**(2), pp.75-94.

Dejonghe, W., Boon, N., Seghers, D., Top, E.M. and Verstraete, W. 2001. Bioaugmentation of soils by increasing microbial richness: missing links. *Environmental Microbiology.* **3**(10), pp.649-657.

Delcour, A.H. 2009. Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1794**(5), pp.808-816.

Demain, A.L. and Vaishnav, P. 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances.* **27**(3), pp.297-306.

Desimone, M., Catoni, E., Ludewig, U., Hilpert, M., Schneider, A., Kunze, R., Tegeder, M., Frommer, W.B. and Schumacher, K. 2002. A novel superfamily of transporters for allantoin and other oxo derivatives of nitrogen heterocyclic compounds in Arabidopsis. *The Plant Cell.* **14**(4), pp.847-856.

Deuschle, U., Gentz, R. and Bujard, H. 1986. lac repressor blocks transcribing RNA polymerase and terminates transcription. *Proceedings of the National Academy of Sciences.* **83**(12), pp.4134-4137.

Diallinas, D., Gorfinkiel, G., Arst, H.N., Cecchetto, C. and Scazzocchio, S. 1995. Genetic and molecular characterization of a gene encoding a wide specificity purine permease of *Aspergillus nidulans* reveals a novel family of transporters conserved in prokaryotes and eukaryotes. *Journal of Biological Chemistry.* **270**(15), pp.8610-8622.

Diallinas, G. 2014. Understanding transporter specificity and the discrete appearance of channel-like gating domains in transporters. *Frontiers in pharmacology.* **5**.

Diallinas, G. and Gournas, C. 2008. The ubiquitous Nucleobase-Ascorbate Transporter (NAT) family: lessons from model microbial genetic systems. *Channels.* **2**(5), pp.363-372.

Dobson, L., Reményi, I. and Tusnády, G.E. 2015. The human transmembrane proteome. *Biology direct.* **10**(1), p31.

Donovan, R.S., Robinson, C.W. and Glick, B. 1996. Optimizing inducer and culture conditions for expression of foreign proteins under the control of thelac promoter. *Journal of industrial microbiology & biotechnology.* **16**(3), pp.145-154.

Dowhan, W. 1997. Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annual review of biochemistry.* **66**(1), pp.199-232.

Drapeau, C., Grilli, E. and Petrosillo, N. 2010. Rifampicin combined regimens for Gram-negative infections: data from the literature. *International journal of antimicrobial agents.* **35**(1), pp.39-44.

Dreier, J. and Ruggerone, P. 2015. Interaction of antibacterial compounds with RND efflux pumps in *Pseudomonas aeruginosa*. *Frontiers in microbiology.* **6**.

Drew, D., Fröderberg, L., Baars, L. and de Gier, J.-W.L. 2003. Assembly and overexpression of membrane proteins in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1610**(1), pp.3-10.

Drew, D., Slotboom, D.J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N.M., Lambert, W., Lerch, M. and Daley, D.O. 2005. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein science.* **14**(8), pp.2011-2017.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21 (DE3), C41 (DE3), and C43 (DE3). *Protein expression and purification.* **37**(1), pp.203-206.

Elkins, C.A. and Nikaido, H. 2003. Chimeric analysis of AcrA function reveals the importance of its C-terminal domain in its interaction with the AcrB multidrug efflux pump. *Journal of bacteriology.* **185**(18), pp.5349-5356.

Engel, A. and Gaub, H.E. 2008. Structure and mechanics of membrane proteins. *Annu. Rev. Biochem.* **77**, pp.127-148.

Erni, B. 1992. Group translocation of glucose and other carbohydrates by the bacterial phosphotransferase system. *International review of cytology.* **137**, pp.127-148.

Fàbrega, A., Madurga, S., Giralt, E. and Vila, J. 2009. Mechanism of action of and resistance to quinolones. *Microbial biotechnology.* **2**(1), pp.40-61.

Faham, S., Watanabe, A., Besserer, G.M., Cascio, D., Specht, A., Hirayama, B.A., Wright, E.M. and Abramson, J. 2008. The crystal structure of a sodium galactose transporter reveals mechanistic insights into $Na^+$/sugar symport. *Science.* **321**(5890), pp.810-814.

Franklin, F., Bagdasarian, M., Bagdasarian, M. and Timmis, K. 1981. Molecular and functional analysis of the TOL plasmid pWWO from *Pseudomonas putida* and cloning of genes for the entire regulated aromatic ring meta cleavage pathway. *Proceedings of the National Academy of Sciences.* **78**(12), pp.7458-7462.

Frillingos, S. 2012. Insights to the evolution of Nucleobase-Ascorbate Transporters (NAT/NCS2 family) from the Cys-scanning analysis of xanthine permease XanQ. *International journal of biochemistry and molecular biology.* **3**(3), p250.

Gaboriau, F., Kreder, A., Clavreul, N., Moulinoux, J.-P., Delcros, J.-G. and Lescoat, G. 2004. Polyamine modulation of iron uptake in CHO cells. *Biochemical pharmacology.* **67**(9), pp.1629-1637.

Gao, F.P. and Cross, T.A. 2006. Recent developments in membrane-protein structural genomics. *Genome biology.* **6**(13), p244.

Garavito, R.M. and Ferguson-Miller, S. 2001. Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry.* **276**(35), pp.32403-32406.

Garrett, R.H., Grisham, C.M. (1999) Biochemistry, 2nd Edition. Harcourt Brace, Custom Publishers, San Diego.

Gautier, A. 2014. Structure determination of α-helical membrane proteins by solution-state NMR: Emphasis on retinal proteins. *Biochimica et Biophysica Acta (BBA)-Bioenergetics.* **1837**(5), pp.578-588.

Ghisaidoobe, A.B. and Chung, S.J. 2014. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences.* **15**(12), pp.22518-22538.

Gillissen, B., Bürkle, L., André, B., Kühn, C., Rentsch, D., Brandl, B. and Frommer, W.B. 2000. A new family of high-affinity transporters for adenine, cytosine, and purine derivatives in Arabidopsis. *The Plant Cell.* **12**(2), pp.291-300.

Girke, C., Daumann, M., Niopek-Witz, S. and Möhlmann, T. 2014. Nucleobase and nucleoside transport and integration into plant metabolism. *Frontiers in plant science.* **5**.

Goldberg, M., Pribyl, T., Juhnke, S. and Nies, D.H. 1999. Energetics and topology of CzcA, a cation/proton antiporter of the resistance-nodulation-cell division protein family. *Journal of Biological Chemistry.* **274**(37), pp.26065-26070.

Golkar, Z., Bagasra, O. and Pace, D.G. 2014. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *The Journal of Infection in Developing Countries.* **8**(02), pp.129-136.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U. and Putnam, N. 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research.* **40**(D1), pp.D1178-D1186.

Gordon, E., Horsefield, R., Swarts, H.G., de Pont, J.J.H., Neutze, R. and Snijder, A. 2008. Effective high-throughput overproduction of membrane proteins in Escherichia coli. *Protein expression and purification.* **62**(1), pp.1-8.

Gournas, C., Papageorgiou, I. and Diallinas, G. 2008. The nucleobase–ascorbate transporter (NAT) family: genomics, evolution, structure–function relationships and physiological role. *Molecular Biosystems.* **4**(5), pp.404-416.

Gräslund, S., Nordlund, P., Weigelt, J., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R. and Ming, J. 2008. Protein production and purification. *Nature methods*. **5**(2), pp.135-146.

Griffith, J.K., Baker, M.E., Rouch, D.A., Page, M.G., Skurray, R.A., Paulsen, I.T., Chater, K.F., Baldwin, S.A. and Henderson, P.J. 1992. Membrane transport proteins: implications of sequence comparisons. *Current opinion in cell biology*. **4**(4), pp.684-695.

Grinius, L.L. and Goldberg, E.B. 1994. Bacterial multidrug resistance is due to a single membrane protein which functions as a drug pump. *Journal of Biological Chemistry*. **269**(47), pp.29998-30004.

Guan, L. and Nakae, T. 2001. Identification of Essential Charged Residues in Transmembrane Segments of the Multidrug Transporter MexB of *Pseudomonas aeruginosa*. *Journal of bacteriology*. **183**(5), pp.1734-1739.

Gutman, N., Steiner-Mordoch, S. and Schuldiner, S. 2003. An amino acid cluster around the essential Glu-14 is part of the substrate-and proton-binding domain of EmrE, a multidrug transporter from *Escherichia coli*. *Journal of Biological Chemistry*. **278**(18), pp.16082-16087.

Hamari, Z., Amillis, S., Drevet, C., Apostolaki, A., Vágvölgyi, C., Diallinas, G. and Scazzocchio, C. 2009. Convergent evolution and orphan genes in the Fur4p-like family and characterization of a general nucleoside transporter in *Aspergillus nidulans*. *Molecular microbiology*. **73**(1), pp.43-57.

Harada, N. and Inagaki, N. 2012. Role of sodium-glucose transporters in glucose uptake of the intestine and kidney. *Journal of diabetes investigation*. **3**(4), pp.352-353.

Harris, N.J., Findlay, H.E., Simms, J., Liu, X. and Booth, P.J. 2014. Relative domain folding and stability of a membrane transport protein. *Journal of molecular biology*. **426**(8), pp.1812-1825.

Hassan, K.A., Elbourne, L.D., Li, L., Gamage, H.K., Liu, Q., Jackson, S.M., Sharples, D., Kolstø, A.-B., Henderson, P.J. and Paulsen, I.T. 2015a. An ace up their sleeve: a transcriptomic approach exposes the AceI efflux protein of *Acinetobacter baumannii* and reveals the drug efflux potential hidden in many microbial pathogens. *Frontiers in microbiology*. **6**.

Hassan, K.A., Jackson, S.M., Penesyan, A., Patching, S.G., Tetu, S.G., Eijkelkamp, B.A., Brown, M.H., Henderson, P.J. and Paulsen, I.T. 2013. Transcriptomic and biochemical analyses identify a family of chlorhexidine efflux proteins. *Proceedings of the National Academy of Sciences*. **110**(50), pp.20254-20259.

Hassan, K.A., Liu, Q., Henderson, P.J. and Paulsen, I.T. 2015b. Homologs of the *Acinetobacter baumannii* AceI transporter represent a new family of bacterial multidrug efflux systems. *MBio*. **6**(1), pp.e01982-01914.

Hegstad, K., Langsrud, S., Lunestad, B.T., Scheie, A.A., Sunde, M. and Yazdankhah, S.P. 2010. Does the wide use of quaternary ammonium compounds enhance the selection and spread of antimicrobial resistance and thus threaten our health? *Microbial drug resistance.* **16**(2), pp.91-104.

Heir, E., Sundheim, G. and Holck, A. 1999. The qacG gene on plasmid pST94 confers resistance to quaternary ammonium compounds in *staphylococci* isolated from the food industry. *Journal of applied microbiology.* **86**(3), pp.378-388.

Henderson, P., Hoyle, C. and Ward, A. 2000. *Expression, purification and properties of multidrug efflux proteins*. Portland Press Limited.

Henderson, P. and Maiden, M. 1990. Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences.* **326**(1236), pp.391-410.

Henderson, P.J. 1998. Function and Structure of Membrane Transport Proteins-1.

Henderson, P.J. 2013. Membrane Transport: Energetics and Overview. *Encyclopedia of Biophysics.* pp.1496-1504.

Hennerdal, A. and Elofsson, A. 2011. Rapid membrane protein topology prediction. *Bioinformatics.* **27**(9), pp.1322-1323.

Higgins, C.F. 1992. ABC transporters: from microorganisms to man. *Annual review of cell biology.* **8**(1), pp.67-113.

Higgins, C.F. 2001. ABC transporters: physiology, structure and mechanism–an overview. *Research in microbiology.* **152**(3), pp.205-210.

Hildreth, S.B., Gehman, E.A., Yang, H., Lu, R.-H., Ritesh, K., Harich, K.C., Yu, S., Lin, J., Sandoe, J.L. and Okumoto, S. 2011. Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. *Proceedings of the National Academy of Sciences.* **108**(44), pp.18179-18184.

Hiyoshi, H., Kodama, T., Iida, T. and Honda, T. 2010. Contribution of *Vibrio parahaemolyticus* virulence factors to cytotoxicity, enterotoxicity, and lethality in mice. *Infection and immunity.* **78**(4), pp.1772-1780.

Hochuli, E. 1989. Genetically designed affinity chromatography using a novel metal chelate absorbent. *Biologically active molecules.* Springer, pp.217-239.

Hochuli, E., Döbeli, H. and Schacher, A. 1987. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography A.* **411**, pp.177-184.

Holmes, R. and Jobling, M. 1996. Genetics: conjugation. *Barons medical microbiology*. **4**.

Hood, M.I., Jacobs, A.C., Sayood, K., Dunman, P.M. and Skaar, E.P. 2010. *Acinetobacter baumannii* increases tolerance to antibiotics in response to monovalent cations. *Antimicrobial agents and chemotherapy.* **54**(3), pp.1029-1041.

Hoyle, C. J. (2000). Recombinant expression, purification and characterisation of bacterial multidrug efflux proteins. Ph.D. Thesis, University of Leeds, Leeds, UK.

Huang, K.-S., Bayley, H., Liao, M.-J., London, E. and Khorana, H. 1981. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *Journal of Biological Chemistry.* **256**(8), pp.3802-3809.

Huda, M., Chen, J., Morita, Y., Kuroda, T., Mizushima, T. and Tsuchiya, T. 2003. Gene Cloning and Characterization of VcrM, a $Na^+$-Coupled Multidrug Efflux Pump, from *Vibrio cholerae* Non-O1. *Microbiology and immunology.* **47**(6), pp.419-427.

Huda, M.N., Morita, Y., Kuroda, T., Mizushima, T. and Tsuchiya, T. 2001. Na+-driven multidrug efflux pump VcmA from *Vibrio cholerae* non-O1, a non-halophilic bacterium. *FEMS microbiology letters.* **203**(2), pp.235-239.

Hunte, C., Von Jagow, G. and Schagger, H. 2003. *Membrane protein purification and crystallization: a practical guide.* Academic press.

Hvorup, R.N., Winnen, B., Chang, A.B., Jiang, Y., Zhou, X.F. and Saier, M.H. 2003. The multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) exporter superfamily. *The FEBS Journal.* **270**(5), pp.799-813.

Hyde, R.J., Cass, C.E., Young, J.D. and Stephen A. Baldwin, J.D. 2001. The ENT family of eukaryote nucleoside and nucleobase transporters: recent advances in the investigation of structure/function relationships and the identification of novel isoforms. *Molecular membrane biology.* **18**(1), pp.53-63.

Inoue, H., Nojima, H. and Okayama, H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene.* **96**(1), pp.23-28.

Iwamoto, M., Ayers, T., Mahon, B.E. and Swerdlow, D.L. 2010. Epidemiology of seafood-associated infections in the United States. *Clinical microbiology reviews.* **23**(2), pp.399-411.

Iwata, S. 2003. Crystallization informatics of membrane proteins. *Methods and results in crystallization of membrane proteins (ed. S. Iwata).* pp.283-297.

Jackson, S.M (2012) Elucidating the molecular mechanisms of ligand binding and transport by the $Na^+$-Hydantoin transport protein, Mhp1. PhD thesis, University of Leeds, Leeds, UK.

Jackson, S.M., Patching, S.G., Ivanova, E., Simmons, K., Weyand, S., Shimamura, T., Brueckner, F., Suzuki, S.i., Iwata, S. and Sharples, D.J. 2013. Mhp1, the $Na^+$-hydantoin membrane transport protein. *Encyclopedia of Biophysics.* pp.1514-1521.

Jardetzky, O. 1966. Simple allosteric model for membrane pumps. *Nature.* **211**(5052), pp.969-970.

Jarvik, J.W. and Telmer, C.A. 1998. Epitope tagging. *Annual review of genetics.* **32**(1), pp.601-618.

Jelesko, J.G. 2012. An expanding role for purine uptake permease-like transporters in plant secondary metabolism. *Frontiers in plant science.* **3**.

Jerpseth, M., Jerpseth, B., Breister, L. and Greener, A. (1998) High-efficiency Derivatives of the BL21 Series for Protein Expression. Strategies, **11**, 3-4.

Jia, B. and Jeon, C.O. 2016. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. *Open biology.* **6**(8), p160196.

Johansson, L.C., Wöhri, A.B., Katona, G., Engström, S. and Neutze, R. 2009. Membrane protein crystallization from lipidic phases. *Current opinion in structural biology.* **19**(4), pp.372-378.

Johnson, J.E. and Cornell, R.B. 1999. Amphitropic proteins: regulation by reversible membrane interactions. *Molecular membrane biology.* **16**(3), pp.217-235.

Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J.-P. 2014. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology.* **12**(3), pp.181-196.

Jones, P. and George, A. 2004. The ABC transporter structure and mechanism: perspectives on recent research. *Cellular and Molecular Life Sciences.* **61**(6), pp.682-699.

Jung, B., Flörchinger, M., Kunz, H.-H., Traub, M., Wartenberg, R., Jeblick, W., Neuhaus, H.E. and Möhlmann, T. 2009. Uridine-ribohydrolase is a key regulator in the uridine degradation pathway of Arabidopsis. *The Plant Cell.* **21**(3), pp.876-891.

K Liszewski, M. and Atkinson, J.P. 2015. Complement regulators in human disease: lessons from modern genetics. *Journal of internal medicine.* **277**(3), pp.294-305.

Kadner, R. 1996. Cytoplasmic membrane. *Escherichia coli.* pp.58-87.

Kahlon, R.S. 2016. *Pseudomonas: Molecular and Applied Biology.* Springer.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research.* **45**(D1), pp.D353-D361.

Karatza, P., Panos, P., Georgopoulou, E. and Frillingos, S. 2006. Cysteine-scanning Analysis of the Nucleobase-Ascorbate Transporter Signature Motif in YgfO Permease of *Escherichia coli* Gln-324 AND Asn-325 ARE ESSENTIAL, AND Ile-

329–Val-339 FORM AN α-HELIX. *Journal of Biological Chemistry*. **281**(52), pp.39881-39890.

Kashmiri, S. and Hotchkiss, R.D. 1975. Evidence of tandem duplication of genes in a merodiploid region of pneumococcal mutants resistant to sulfonamide. *Genetics.* **81**(1), pp.21-31.

Kaslow, D.C. and Shiloach, J. 1994. Production, purification and immunogenicity of a malaria transmission-blocking vaccine candidate: TBV25H expressed in yeast and purified using nickel-NTA agarose. *Nature Biotechnology.* **12**(5), pp.494-499.

Kazmier, K., Sharma, S., Islam, S.M., Roux, B. and Mchaourab, H.S. 2014. Conformational cycle and ion-coupling mechanism of the $Na^+$/hydantoin transporter Mhp1. *Proceedings of the National Academy of Sciences.* **111**(41), pp.14752-14757.

Kelly, S.M., Jess, T.J. and Price, N.C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1751**(2), pp.119-139.

Kido, M., Yamanaka, K., Mitani, T., Niki, H., Ogura, T. and Hiraga, S. 1996. RNase E polypeptides lacking a carboxyl-terminal half suppress a mukB mutation in *Escherichia coli*. *Journal of bacteriology.* **178**(13), pp.3917-3925.

Kihara, A., Akiyama, Y. and Ito, K. 1995. FtsH is required for proteolytic elimination of uncomplexed forms of SecY, an essential protein translocase subunit. *Proceedings of the National Academy of Sciences.* **92**(10), pp.4532-4536.

King, A.E., Ackley, M.A., Cass, C.E., Young, J.D. and Baldwin, S.A. 2006. Nucleoside transporters: from scavengers to novel therapeutic targets. *Trends in pharmacological sciences.* **27**(8), pp.416-425.

Koteiche, H.A., Reeves, M.D. and Mchaourab, H.S. 2003. Structure of the substrate binding pocket of the multidrug transporter EmrE: site-directed spin labeling of transmembrane segment 1. *Biochemistry.* **42**(20), pp.6099-6105.

Kourtesi, C., Ball, A.R., Huang, Y.-Y., Jachak, S.M., Vera, D.M.A., Khondkar, P., Gibbons, S., Hamblin, M.R. and Tegos, G.P. 2013. Suppl 1: Microbial efflux systems and inhibitors: approaches to drug discovery and the challenge of clinical implementation. *The open microbiology journal.* **7**, p34.

Krishnamurthy, H., Piscitelli, C.L. and Gouaux, E. 2009. Unlocking the molecular secrets of sodium-coupled transporters. *Nature.* **459**(7245), pp.347-355.

Krogh, A., Larsson, B., Von Heijne, G. and Sonnhammer, E.L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology.* **305**(3), pp.567-580.

Krypotou, E., Evangelidis, T., Bobonis, J., Pittis, A.A., Gabaldón, T., Scazzocchio, C., Mikros, E. and Diallinas, G. 2015. Origin, diversification and substrate

specificity in the family of NCS1/FUR transporters. *Molecular microbiology.* **96**(5), pp.927-950.

Kumar, S. and Nussinov, R. 1999. Salt bridge stability in monomeric proteins. *Journal of molecular biology.* **293**(5), pp.1241-1255.

Kumar, S. and Varela, M.F. 2012. Biochemistry of bacterial multidrug efflux pumps. *International Journal of Molecular Sciences.* **13**(4), pp.4484-4495.

Kumaran, T. and Citarasu, T. 2016. Isolation and characterization of virulence-related properties of pathogenic *Vibrio Parahaemolyticus* isolated from aquatic environments. *Environmental Journal.* **2**, pp.57-65.

Kundig, W., Ghosh, S. and Roseman, S. 1964. Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proceedings of the National Academy of Sciences.* **52**(4), pp.1067-1074.

Kuroda, T. and Tsuchiya, T. 2009. Multidrug efflux transporters in the MATE family. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1794**(5), pp.763-768.

Kuusinen, A., Arvola, M., Oker-Blom, C. and Keinänen, K. 1995. Purification of Recombinant GluR-D Glutamate Receptor Produced in Sf21 Insect Cells. *The FEBS Journal.* **233**(3), pp.720-726.

Laage, R. and Langosch, D. 2001. Strategies for prokaryotic expression of eukaryotic membrane proteins. *Traffic.* **2**(2), pp.99-104.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature.* **227**(5259), pp.680-685.

Lambert, P.A. 2005. Bacterial resistance to antibiotics: modified target sites. *Advanced drug delivery reviews.* **57**(10), pp.1471-1485.

Law, C.J., Almqvist, J., Bernstein, A., Goetz, R.M., Huang, Y., Soudant, C., Laaksonen, A., Hovmöller, S. and Wang, D.-N. 2008. Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *Journal of molecular biology.* **378**(4), pp.828-839.

Lee, C.-T., Chen, I.-T., Yang, Y.-T., Ko, T.-P., Huang, Y.-T., Huang, J.-Y., Huang, M.-F., Lin, S.-J., Chen, C.-Y. and Lin, S.-S. 2015. The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by acquiring a plasmid that expresses a deadly toxin. *Proceedings of the National Academy of Sciences.* **112**(34), pp.10798-10803.

Leikin, J.B. and Paloucek, F.P. 2008. Chlorhexidine gluconate. *Poisoning and Toxicology Handbook (4th ed.), Informa.* pp.183-184.

Lekshmi, M., Ammini, P., Kumar, S. and Varela, M.F. 2017. The Food Production Environment and the Development of Antimicrobial Resistance in Human Pathogens of Animal Origin. Microorganisms. 5(1), p11.

Letchumanan, V., Chan, K.-G. and Lee, L.-H. 2014. *Vibrio parahaemolyticus*: a review on the pathogenesis, prevalence, and advance molecular identification techniques. *Frontiers in microbiology.* **5**.

Letchumanan, V., Pusparajah, P., Tan, L.T.-H., Yin, W.-F., Lee, L.-H. and Chan, K.-G. 2015. Occurrence and antibiotic resistance of *Vibrio parahaemolyticus* from shellfish in Selangor, Malaysia. *Frontiers in microbiology.* **6**.

Letunic, I. and Bork, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research.* **44**(W1), pp.W242-W245.

Leung, J., Karachaliou, M., Alves, C., Diallinas, G. and Byrne, B. 2010. Expression and purification of a functional uric acid–xanthine transporter (UapA). *Protein expression and purification.* **72**(1), pp.139-146.

LeVine, M.V., Cuendet, M.A., Khelashvili, G. and Weinstein, H. 2016. Allosteric mechanisms of molecular machines at the membrane: transport by sodium-coupled symporters. *Chemical reviews.* **116**(11), pp.6552-6587.

Li, G., Liu, K., Baldwin, S.A. and Wang, D. 2003. Equilibrative nucleoside transporters of *Arabidopsis thaliana* cDNA cloning, expression pattern, and analysis of transport activities. *Journal of Biological Chemistry.* **278**(37), pp.35732-35742.

Li, X.-Z. and Nikaido, H. 2009. Efflux-mediated drug resistance in bacteria. *Drugs.* **69**(12), pp.1555-1623.

Li, X.-Z., Nikaido, H. and Poole, K. 1995. Role of mexA-mexB-oprM in antibiotic efflux in *Pseudomonas aeruginosa*. *Antimicrobial agents and chemotherapy.* **39**(9), pp.1948-1953.

Li, X.-Z., Plésiat, P. and Nikaido, H. 2015. The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clinical microbiology reviews.* **28**(2), pp.337-418.

Lilie, H., Schwarz, E. and Rudolph, R. 1998. Advances in refolding of proteins produced in *E. coli*. *Current opinion in biotechnology.* **9**(5), pp.497-501.

Liszewski, K. 2015. Dissecting the structure of membrane proteins.

London, E. and Khorana, H.G. 1982. Denaturation and renaturation of bacteriorhodopsin in detergents and lipid-detergent mixtures. *Journal of Biological Chemistry.* **257**(12), pp.7003-7011.

Long, F., Su, C.-C., Zimmermann, M.T., Boyken, S.E., Rajashankar, K.R., Jernigan, R.L. and Edward, W.Y. 2010. Crystal structures of the CusA efflux pump suggest methionine-mediated metal transport. *Nature.* **467**(7314), pp.484-488.

Lopez, P.J., Marchand, I., Joyce, S.A. and Dreyfus, M. 1999. The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Molecular microbiology.* **33**(1), pp.188-199.

Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X. and Wang, J. 2011. Structure and mechanism of the uracil transporter UraA. *Nature.* **472**(7342), pp.243-246.

Lubelski, J., Konings, W.N. and Driessen, A.J. 2007. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiology and Molecular Biology Reviews.* **71**(3), pp.463-476.

Luckey, M. 2014. *Membrane structural biology: with biochemical and biophysical foundations.* Cambridge University Press.

Lundstrom, K. 2007. Structural genomics and drug discovery. *Journal of cellular and molecular medicine.* **11**(2), pp.224-238.

Ma, D., Cook, D.N., Alberti, M., Pon, N.G., Nikaido, H. and Hearst, J.E. 1995. Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli. Molecular microbiology.* **16**(1), pp.45-55.

Ma P (2010) Structure-activity relationships of membrane proteins: The NCS-1 family of transports and sensor kinases of two-component systems. PhD Thesis, University of Leeds, Leeds, UK.

Ma, P., Patching, S.G., Ivanova, E., Baldwin, J.M., Sharples, D., Baldwin, S.A. and Henderson, P.J. 2016. Allantoin transport protein, PucI, from *Bacillus subtilis*: evolutionary relationships, amplified expression, activity and specificity. *Microbiology.* **162**(5), pp.823-836.

Ma, P., Varela, F., Magoch, M., Silva, A.R., Rosário, A.L., Brito, J., Oliveira, T.F., Nogly, P., Pessanha, M. and Stelter, M. 2013. An efficient strategy for small-scale screening and production of archaeal membrane transport proteins in *Escherichia coli*. *PloS one.* **8**(10), pe76913.

Magee, J., Pritchard, E.L., Fitzgerald, K.A., Dunstan, F. and Howard, A. 1999. Antibiotic prescribing and antibiotic resistance in community practice: retrospective study, 1996-8. *Bmj.* **319**(7219), pp.1239-1240.

Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M. and Yamashita, A. 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae. The Lancet.* **361**(9359), pp.743-749.

Makoff, A. and Oxer, M. 1991. High level heterologous expression in *E. coli* using mutant forms of the lac promoter. *Nucleic acids research.* **19**(9), pp.2417-2421.

Makrides, S.C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological reviews.* **60**(3), pp.512-538.

Mao, W., Warren, M.S., Black, D.S., Satou, T., Murata, T., Nishino, T., Gotoh, N. and Lomovskaya, O. 2002. On the mechanism of substrate specificity by resistance nodulation division (RND)-type multidrug resistance pumps: the large periplasmic loops of MexD from *Pseudomonas aeruginosa* are involved in substrate recognition. *Molecular microbiology.* **46**(3), pp.889-901.

Marger, M.D. and Saier, M.H. 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends in biochemical sciences.* **18**(1), pp.13-20.

Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M. and Striedner, G. 2013. Evaluation of three industrial *Escherichia coli* strains in fed-batch cultivations during high-level SOD protein production. *Microbial cell factories.* **12**(1), p58.

Marston, H.D., Dixon, D.M., Knisely, J.M., Palmore, T.N. and Fauci, A.S. 2016. Antimicrobial resistance. *Jama.* **316**(11), pp.1193-1204.

Martinez, J. and Baquero, F. 2000. Mutation frequencies and antibiotic resistance. *Antimicrobial agents and chemotherapy.* **44**(7), pp.1771-1777.

Martins, M., McCusker, M.P., Viveiros, M., Couto, I., Fanning, S., Pagès, J.-M. and Amaral, L. 2013. A simple method for assessment of MDR bacteria for over-expressed efflux pumps. *The open microbiology journal.* **7**, p72.

Mauro, V.P., Chappell, S.A., Zhou, W. and Edelman, G.M. 2014. *Reengineering mRNA primary structure for enhanced protein production*. Google Patents.

McCarter, L. 1999. The multiple identities of *Vibrio parahaemolyticus*. *Journal of molecular microbiology and biotechnology.* **1**(1), pp.51-57.

McManus, M.C. 1997. Mechanisms of bacterial resistance to antimicrobial agents. *American Journal of Health-System Pharmacy.* **54**(12), pp.1420-1433.

Menger, F., Zana, R. and Lindman, B. 1998. Portraying the structure of micelles. *J. Chem. Educ.* **75**(1), p115.

Mercier, A., Kay, E., Vogel, T.M. and Simonet, P. 2007. 14 Gene Flow in the Rhizosphere. *The Rhizosphere: Biochemistry and Organic Substances at the Soil-Plant Interface.* p401.

Midgett, C.R. and Madden, D.R. 2007. Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies. *Journal of structural biology.* **160**(3), pp.265-274.

Miles, A.J. and Wallace, B. 2016. Circular dichroism spectroscopy of membrane proteins. *Chemical Society Reviews.* **45**(18), pp.4859-4872.

Miroux, B. and Walker, J.E. 1996. Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of molecular biology.* **260**(3), pp.289-298.

Misquitta, Y. and Caffrey, M. 2003. Detergents destabilize the cubic phase of monoolein: implications for membrane protein crystallization. *Biophysical journal.* **85**(5), pp.3084-3096.

Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. and Stewart, P.D.S. 2014. Membrane protein structure determination—the next generation. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1838**(1), pp.78-87.

Morais, V.A., Haddad, D., Craessaerts, K., De Bock, P.-J., Swerts, J., Vilain, S., Aerts, L., Overbergh, L., Grünewald, A. and Seibler, P. 2014. PINK1 loss-of-function mutations affect mitochondrial complex I activity via NdufA10 ubiquinone uncoupling. *Science.* **344**(6180), pp.203-207.

Morita, Y., Kataoka, A., Shiota, S., Mizushima, T. and Tsuchiya, T. 2000. NorM of *Vibrio parahaemolyticus* is an $Na^+$-driven multidrug efflux pump. *Journal of bacteriology.* **182**(23), pp.6694-6697.

Morita, Y., Kodama, K., Shiota, S., Mine, T., Kataoka, A., Mizushima, T. and Tsuchiya, T. 1998. NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in Escherichia coli. *Antimicrobial agents and chemotherapy.* **42**(7), pp.1778-1782.

Moriyama, Y., Hiasa, M., Matsumoto, T. and Omote, H. 2008. Multidrug and toxic compound extrusion (MATE)-type proteins as anchor transporters for the excretion of metabolic waste products and xenobiotics. *Xenobiotica.* **38**(7-8), pp.1107-1118.

Mourad, G.S., Tippmann-Crosby, J., Hunt, K.A., Gicheru, Y., Bade, K., Mansfield, T.A. and Schultes, N.P. 2012. Genetic and molecular characterization reveals a unique nucleobase cation symporter 1 in Arabidopsis. *FEBS letters.* **586**(9), pp.1370-1378.

Müller, T., Meyer, H.E., Egensperger, R. and Marcus, K. 2008. The amyloid precursor protein intracellular domain (AICD) as modulator of gene expression, apoptosis, and cytoskeletal dynamics—relevance for Alzheimer's disease. *Progress in neurobiology.* **85**(4), pp.393-406.

Muniesa, M., Colomer-Lluch, M. and Jofre, J. 2013. Potential impact of environmental bacteriophages in spreading antibiotic resistance genes. *Future microbiology.* **8**(6), pp.739-751.

Murakami, S., Nakashima, R., Yamashita, E. and Yamaguchi, A. 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature.* **419**(6907), pp.587-593.

Murakami, S. and Yamaguchi, A. 2003. Multidrug-exporting secondary transporters. *Current opinion in structural biology.* **13**(4), pp.443-452.

Nagy, J.K., Lonzer, W.L. and Sanders, C.R. 2001. Kinetic study of folding and misfolding of diacylglycerol kinase in model membranes. *Biochemistry.* **40**(30), pp.8971-8980.

Narhi, L.O. 2013. *Biophysics for therapeutic protein development.* Springer Science & Business Media.

Natale, P., Brüser, T. and Driessen, A.J. 2008. Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1778**(9), pp.1735-1756.

Nelapati, S., Nelapati, K. and Chinnam, B. 2012. *Vibrio parahaemolyticus*-An emerging foodborne pathogen-A Review. *Vet. World.* **5**(1), pp.48-62.

Nelson, D.L., Lehninger, A.L. and Cox, M.M. 2008. *Lehninger principles of biochemistry.* Macmillan.

Nelson, K., Weinel, C., Paulsen, I., Dodson, R., Hilbert, H., Martins dos Santos, V., Fouts, D., Gill, S., Pop, M. and Holmes, M. 2002. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environmental microbiology.* **4**(12), pp.799-808.

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

Nikaido, H. 2003. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and molecular biology reviews.* **67**(4), pp.593-656.

Nikaido, H. and Pagès, J.-M. 2012. Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria. *FEMS microbiology reviews.* **36**(2), pp.340-363.

Nikaido, H. and Takatsuka, Y. 2009. Mechanisms of RND multidrug efflux pumps. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1794**(5), pp.769-781.

Olsen, I. and Jantzen, E. 2001. Sphingolipids in bacteria and fungi. *Anaerobe.* **7**(2), pp.103-112.

Omote, H., Hiasa, M., Matsumoto, T., Otsuka, M. and Moriyama, Y. 2006. The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations. *Trends in Pharmacological Sciences.* **27**(11), pp.587-593.

Overington, J.P., Al-Lazikani, B. and Hopkins, A.L. 2006. How many drug targets are there? *Nature reviews Drug discovery.* **5**(12), pp.993-996.

Padan, E. 2002. Bacterial membrane transport: organization of membrane activities. *eLS.*

Padan, E., Kozachkov, L., Herz, K. and Rimon, A. 2009. NhaA crystal structure: functional–structural insights. *Journal of Experimental Biology.* **212**(11), pp.1593-1603.

Palleroni, N.J. 1984. *Pseudomonas. Bergey's Manual of Systematics of Archaea and Bacteria.*

Palty, R., Hershfinkel, M. and Sekler, I. 2012. Molecular identity and functional properties of the mitochondrial $Na^{+}/Ca2^{+}$ exchanger. *Journal of Biological Chemistry.* **287**(38), pp.31650-31657.

Pantazopoulou, A. and Diallinas, G. 2007. Fungal nucleobase transporters. *FEMS microbiology reviews.* **31**(6), pp.657-675.

Pao, S.S., Paulsen, I.T. and Saier, M.H. 1998. Major facilitator superfamily. *Microbiology and molecular biology reviews.* **62**(1), pp.1-34.

Paulsen, I.T., Brown, M.H. and Skurray, R.A. 1996a. Proton-dependent multidrug efflux systems. *Microbiological reviews.* **60**(4), pp.575-608.

Paulsen, I.T. and Skurray, R.A. 1993. Topology, structure and evolution of two families of proteins involved in antibiotic and antiseptic resistance in eukaryotes and prokaryotes—an analysis. *Gene.* **124**(1), pp.1-11.

Paulsen, I.T., Skurray, R.A., Tam, R., Saier, M.H., Turner, R.J., Weiner, J.H., Goldberg, E.B. and Grinius, L.L. 1996b. The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. *Molecular microbiology.* **19**(6), pp.1167-1175.

Paulsen, I.T., Sliwinski, M.K. and Saier, M.H. 1998. Microbial genome analyses: global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. *Journal of molecular biology.* **277**(3), pp.573-592.

Piddock, L.J. 2006. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. *Clinical microbiology reviews.* **19**(2), pp.382-402.

Pielak, G.J. and Tian, F. 2012. Membrane proteins, magic-angle spinning, and in-cell NMR. *Proceedings of the National Academy of Sciences.* **109**(13), pp.4715-4716.

Pope, B. and Kent, H.M. 1996. High efficiency 5 min transformation of *Escherichia coli*. *Nucleic acids research.* **24**(3), pp.536-537.

Porath, J., Carlsson, J., Olsson, I. and Belfrage, G. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature.* **258**, pp.598-599.

Pos, K.M. 2009. Drug transport mechanism of the AcrB efflux pump. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics.* **1794**(5), pp.782-793.

Pos, K.M. and Diederichs, K. 2002. Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Crystallographica Section D: Biological Crystallography.* **58**(10), pp.1865-1867.

Postis, V.L., Deacon, S.E., Roach, P.C., Wright, G.S., Xia, X., Wright, G.S., Xia, X., Ingram, J.C., Hadden, J.M. and Henderson, P.J. 2008. A high-throughput assay of membrane protein stability. *Molecular membrane biology.* **25**(8), pp.617-624.

Qiagen. (2003) The QIAexpressionist- A handbook for high level expression and purification of 6 x His-tagged proteins, QIAGEN Gmbh, QIAGEN Inc. and QIAGEN LTD.

Qi, F. and Turnbough Jr, C.L. 1995. Regulation ofcodBAOperon Expression in *Escherichia coli* by UTP-dependent Reiterative Transcription and UTP-sensitive Transcriptional Start Site Switching. *Journal of molecular biology.* **254**(4), pp.552-565.

Raetz, C.R. 1996. Bacterial lipopolysaccharides: a remarkable family of bioactive macroamphiphiles. *Escherichia coli and Salmonella: cellular and molecular biology.* **1**, pp.1035-1063.

Rajamohan, G., Srinivasan, V.B. and Gebreyes, W.A. 2010. Molecular and functional characterization of a novel efflux pump, AmvA, mediating antimicrobial and disinfectant resistance in *Acinetobacter baumannii. Journal of antimicrobial chemotherapy.* **65**(9), pp.1919-1925.

Rapp, M., Schein, J., Hunt, K.A., Nalam, V., Mourad, G.S. and Schultes, N.P. 2016. The solute specificity profiles of nucleobase cation symporter 1 (NCS1) from *Zea mays* and *Setaria viridis* illustrate functional flexibility. *Protoplasma.* **253**(2), p611.

Rath, A. and Deber, C.M. 2013. Correction factors for membrane protein molecular weight readouts on sodium dodecyl sulfate–polyacrylamide gel electrophoresis. *Analytical biochemistry.* **434**(1), pp.67-72.

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G. and Deber, C.M. 2009. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences.* **106**(6), pp.1760-1765.

Rawson, S., Davies, S., Lippiat, J.D. and Muench, S.P. 2016. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Molecular Membrane Biology.* **33**(1-2), pp.12-22.

Read, A.F. and Woods, R.J. 2014. Antibiotic resistance management. *Evolution, medicine, and public health.* **2014**(1), p147.

Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I. and Saier, M.H. 2012. The major facilitator superfamily (MFS) revisited. *The FEBS journal.* **279**(11), pp.2022-2035.

Rees, D.C., Johnson, E. and Lewinson, O. 2009. ABC transporters: the power to change. *Nature reviews Molecular cell biology.* **10**(3), pp.218-227.

Regenhardt, D., Heuer, H., Heim, S., Fernandez, D., Strömpl, C., Moore, E. and Timmis, K. 2002. Pedigree and taxonomic credentials of *Pseudomonas putida* strain KT2440. *Environmental microbiology.* **4**(12), pp.912-915.

Register, F. 1982. Certified host-vector systems. *Washington, DC.* **47**, p17197.

Ren, Q., Chen, K. and Paulsen, I. 2007. TransportDB: a comprehensive database 4 resource for cytoplasmic membrane transport systems and outer membrane channels. *5 Nucleic Acids Res* **35**. D274-9.

Ren, Q., Chen, K. and Paulsen, I.T. 2006. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic acids research.* **35**(suppl_1), pp.D274-D279.

Ren, Q. and Paulsen, I.T. 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Computational Biology.* **1**(3), pe27.

Renaud, J.-P., Chung, C.-w., Danielson, U.H., Egner, U., Hennig, M., Hubbard, R.E. and Nar, H. 2016. Biophysics in drug discovery: impact, challenges and opportunities. *Nature Reviews Drug Discovery.* **15**(10), pp.679-698.

Ressl, S., van Scheltinga, A.C.T., Vonrhein, C., Ott, V. and Ziegler, C. 2009. Molecular basis of transport and regulation in the $Na^+$/betaine symporter BetP. *Nature.* **458**(7234), pp.47-52.

Rettner, R.E. and Saier Jr, M.H. 2010. The autoinducer-2 exporter superfamily. *Journal of molecular microbiology and biotechnology.* **18**(4), pp.195-205.

Roberts, M. 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS microbiology reviews.* **19**(1), pp.1-24.

Robinson, A.S. 2011. *Production of membrane proteins: strategies for expression and isolation.* John Wiley & Sons.

Rosano, G.L. and Ceccarelli, E.A. 2014a. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in microbiology.* **5**.

Rosenbusch, J.P. 2001. Stability of membrane proteins: relevance for the selection of appropriate methods for high-resolution structure determinations. *Journal of structural biology.* **136**(2), pp.144-157.

Saidijam, M., Bettaney, K.E., Szakonyi, G., Psakis, G., Shibayama, K., Suzuki, S., Clough, J.L., Blessie, V., Abu-Bakr, A. and Baumberg, S. 2005. *Active membrane transport and receptor proteins from bacteria*. Portland Press Limited.

Saidijam, M., Psakis, G., Clough, J.L., Meuller, J., Suzuki, S.i., Hoyle, C.J., Palmer, S.L., Morrison, S.M., Pos, M.K. and Essenberg, R.C. 2003. Collection and characterisation of bacterial membrane proteins. *FEBS letters.* **555**(1), pp.170-175.

Saier Jr, M.H., Beatty, J.T., Goffeau, A., Harley, K.T., Heijne, W., Huang, S.-C., Jack, D.L., Jahn, P., Lew, K. and Liu, J. 1999. The major facilitator superfamily. *J Mol Microbiol Biotechnol.* **1**(2), pp.257-279.

Saier Jr, M.H., Reddy, V.S., Tamang, D.G. and Västermark, Å. 2013. The transporter classification database. *Nucleic acids research.* **42**(D1), pp.D251-D258.

Saier Jr, M.H., Tran, C.V. and Barabote, R.D. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic acids research.* **34**(suppl_1), pp.D181-D186.

Saier, M., Tam, R., Reizer, A. and Reizer, J. 1994. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Molecular microbiology.* **11**(5), pp.841-847.

Sambrook J, Fritsch ER and Maniatis T. 1989. Molecular cloning - A laboratory manual, 2nd edition. Cold Spring Harbour Laboratory Press, New York.

Sar, N., McCARTER, L., Simon, M. and Silverman, M. 1990. Chemotactic control of the two flagellar systems of *Vibrio parahaemolyticus*. *Journal of bacteriology.* **172**(1), pp.334-341.

Schaffner, W. and Weissmann, C. 1973. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Analytical biochemistry.* **56**(2), pp.502-514.

Schmid, A., Dordick, J., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. 2001. Industrial biocatalysis today and tomorrow. *nature.* **409**(6817), pp.258-268.

Schmieger, H. and Schicklmaier, P. 1999. Transduction of multiple drug resistance of *Salmonella enterica* serovar Typhimurium DT104. *FEMS Microbiology Letters.* **170**(1), pp.251-256.

Schuldiner, S., Lebendiker, M. and Yerushalmi, H. 1997. EmrE, the smallest ion-coupled transporter, provides a unique paradigm for structure-function studies. *Journal of experimental biology.* **200**(2), pp.335-341.

Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W.B., Flügge, U.-I. and Kunze, R. 2003. ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiology.* **131**(1), pp.16-26.

Seddon, A.M., Curnow, P. and Booth, P.J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA)-Biomembranes.* **1666**(1), pp.105-117.

Seeger, M.A., Diederichs, K., Eicher, T., Brandstatter, L., Schiefner, A., Verrey, F. and Pos, K.M. 2008. The AcrB efflux pump: conformational cycling and peristalsis lead to multidrug resistance. *Current drug targets.* **9**(9), pp.729-749.

Sengupta, S., Chattopadhyay, M.K. and Grossart, H.-P. 2013. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in microbiology.* **4**.

Sezonov, G., Joseleau-Petit, D. and D'Ari, R. 2007. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of bacteriology.* **189**(23), pp.8746-8749.

Shaffer, P.L., Goehring, A., Shankaranarayanan, A. and Gouaux, E. 2009. Structure and mechanism of a $Na^+$-independent amino acid transporter. *Science.* **325**(5943), pp.1010-1014.

Shah, N.R., Hancock, R.E. and Fernandez, R.C. 2014. Bordetella pertussis lipid A glucosamine modification confers resistance to cationic antimicrobial peptides and increases resistance to outer membrane perturbation. *Antimicrobial agents and chemotherapy.* **58**(8), pp.4931-4934.

Shi, Y. 2013. Common folds and transport mechanisms of secondary active transporters. *Annual review of biophysics.* **42**, pp.51-72.

Shiloach, J. and Fass, R. 2005. Growing *E. coli* to high cell density—a historical perspective on method development. *Biotechnology advances.* **23**(5), pp.345-357.

Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N.G., Hadden, J.M., Sharples, D., Sansom, M.S., Iwata, S., Henderson, P.J. and Cameron, A.D. 2010. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science.* **328**(5977), pp.470-473.

Shimamura, T., Yajima, S., Suzuki, S.i., Rutherford, N.G., O'Reilly, J., Henderson, P.J. and Iwata, S. 2008. Crystallization of the hydantoin transporter Mhp1 from Microbacterium liquefaciens. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications.* **64**(12), pp.1172-1174.

Sievers, F. and Higgins, D.G. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. *Multiple sequence alignment methods.* pp.105-116.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. and Söding, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology.* **7**(1), p539.

Simmons, K.J., Jackson, S.M., Brueckner, F., Patching, S.G., Beckstein, O., Ivanova, E., Geng, T., Weyand, S., Drew, D. and Lanigan, J. 2014. Molecular mechanism of

ligand recognition by membrane transport protein, Mhp1. *The EMBO journal.* pe201387557.

Singer, S.J. and Nicolson, G.L. 1972. The fluid mosaic model of the structure of cell membranes. *Science.* **175**(4023), pp.720-731.

Sioupouli, G., Lambrinidis, G., Mikros, E., Amillis, S. and Diallinas, G. 2017. Cryptic purine transporters in *Aspergillus nidulans* reveal the role of specific residues in the evolution of specificity in the NCS1 family. *Molecular microbiology.* **103**(2), pp.319-332.

Sitbon, E. and Pietrokovski, S. 2007. Occurrence of protein structure elements in conserved sequence regions. *BMC structural biology.* **7**(1), p3.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. and Li, Q. 2009. Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli. Protein Science.* **18**(5), pp.936-948.

Sota, M. and Top, E.M. 2008. Horizontal gene transfer mediated by plasmids. *Plasmids: Current Research and Future Trends.* p263.

Spratt, B.G. 1994. Resistance to antibiotics mediated by target alterations. *Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information.* **264**(5157), pp.388-396.

Srigunapalan, S., Lam, C., Wheeler, A.R. and Simmons, C.A. 2011. A microfluidic membrane device to mimic critical components of the vascular microenvironment. *Biomicrofluidics.* **5**(1), p013409.

Stark, M.J. 1987. Multicopy expression vectors carrying the lac represser gene for regulated high-level expression of genes in *Escherichia coli*. *Gene.* **51**(2), pp.255-267.

Starling, A.P., Hughes, G., Sharma, R.P., East, J.M. and Lee, A.G. 1995. The Hydrophilic Domain of Phospholamban Inhibits the $Ca2^{+}$-ATPase-The Importance of the Method of Assay. *Biochemical and biophysical research communications.* **215**(3), pp.1067-1070.

Streatfield, S.J. 2007. Approaches to achieve high-level heterologous protein production in plants. *Plant biotechnology journal.* **5**(1), pp.2-15.

Sukumar P (2012) Molecular dissection of substrate and inhibitor binding to the D-Galactose-$H^{+}$ symport protein (GalP) from *Escherichia coli* – the bacterial homologue of GLUTT1. Ph.D. Thesis, University of Leeds, Leeds, UK.

Sun, J., Deng, Z. and Yan, A. 2014. Bacterial multidrug efflux pumps: mechanisms, physiology and pharmacological exploitations. *Biochemical and biophysical research communications.* **453**(2), pp.254-267.

Suzuki, S.i. and Henderson, P.J. 2006. The hydantoin transport protein from *Microbacterium liquefaciens*. *Journal of bacteriology.* **188**(9), pp.3329-3336.

Suzuki, S.i., Takenaka, Y., Onishi, N. and Yokozeki, K. 2005. Molecular Cloning and Expression of the hyu Genes from *Microbacterium liquefaciens* AJ 3912, Responsible for the Conversion of 5-Substituted Hydantoins to α-Amino Acids, in *Escherichia coli*. *Bioscience, biotechnology, and biochemistry.* **69**(8), pp.1473-1482.

Szakonyi, G., Leng, D., Ma, P., Bettaney, K.E., Saidijam, M., Ward, A., Zibaei, S., Gardiner, A.T., Cogdell, R.J. and Butaye, P. 2007. A genomic strategy for cloning, expressing and purifying efflux proteins of the major facilitator superfamily. *Journal of antimicrobial chemotherapy.* **59**(6), pp.1265-1270.

Takida, S. and Wedegaertner, P.B. 2004. Exocytic pathway-independent plasma membrane targeting of heterotrimeric G proteins. *FEBS letters.* **567**(2-3), pp.209-213.

Terpe, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology.* **72**(2), p211.

Thorens, B. 1993. Facilitated glucose transporters in epithelial cells. *Annual review of physiology.* **55**(1), pp.591-608.

Timmis, K.N. 2002. *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environmental Microbiology.* **4**(12), pp.779-781.

Towbin, H., Staehelin, T. and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences.* **76**(9), pp.4350-4354.

Tsai, C.S. 2003. *An introduction to computational biochemistry.* John Wiley & Sons.

Tseng, T.-T., Gratwick, K.S., Kollman, J., Park, D., Nies, D.H., Goffeau, A. and Saier Jr, M.H. 1999. The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. *Journal of molecular microbiology and biotechnology.* **1**(1), pp.107-125.

Ulitzur, S. 1974. *Vibrio parahaemolyticus* and *Vibrio alginolyticus*: Short generation-time marine bacteria. *Microbial ecology.* **1**(1), pp.127-135.

Valiyaveetil, F.I., MacKinnon, R. and Muir, T.W. 2002. Semisynthesis and folding of the potassium channel KcsA. *Journal of the American Chemical Society.* **124**(31), pp.9113-9120.

Van Bambeke, F., Balzi, E. and Tulkens, P.M. 2000. Antibiotic efflux pumps. *Biochemical pharmacology.* **60**(4), pp.457-470.

Van Dyke, M.W., Sirito, M. and Sawadogo, M. 1992. Single-step purification of bacterially expressed polypeptides containing an oligo-histidine domain. *Gene.* **111**(1), pp.99-104.

Ventola, C.L. 2015. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics.* **40**(4), p277.

Vidana, R. 2015. *Origin of intraradicular infection with Enterococcus faecalis in endodontically treated teeth.* Inst för odontologi/Dept of Dental Medicine.

Vierstraete, A. 1999. Principle of the PCR. *Εκδόσεις University of Ghent. http://users. ugent. be/~ avierstr/principles/pcr. html.*

Viswanathan, V. 2014. *Off-label abuse of antibiotics by bacteria*. Taylor & Francis.

Voet, D., Voet, J.G. and Pratt, C.W. 2008. Principles of biochemistry. *Hoboken: Wiley.*

Wagner, S., Baars, L., Ytterberg, A.J., Klussmeier, A., Wagner, C.S., Nord, O., Nygren, P.-Å., van Wijk, K.J. and de Gier, J.-W. 2007. Consequences of membrane protein overexpression in *Escherichia coli. Molecular & Cellular Proteomics.* **6**(9), pp.1527-1550.

Wagner, S., Bader, M.L., Drew, D. and de Gier, J.-W. 2006. Rationalizing membrane protein overexpression. *Trends in biotechnology.* **24**(8), pp.364-371.

Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., Van Wijk, K.J., Slotboom, D.J. and Persson, J.O. 2008. Tuning *Escherichia coli* for membrane proteinoverexpression. *Proceedings of the National Academy of Sciences.* **105**(38), pp.14371-14376.

Wallace, B., Lees, J., Orry, A., Lobley, A. and Janes, R.W. 2003. Analyses of circular dichroism spectra of membrane proteins. *Protein Science.* **12**(4), pp.875-884.

Walsh, U.F., Morrissey, J.P. and O'Gara, F. 2001. *Pseudomonas* for biocontrol of phytopathogens: from functional genomics to commercial exploitation. *Current Opinion in Biotechnology.* **12**(3), pp.289-295.

Walther, T.H. and Ulrich, A.S. 2014. Transmembrane helix assembly and the role of salt bridges. *Current opinion in structural biology.* **27**, pp.63-68.

Walmsley, 2000. Spectroscopic and kinetic approaches for probing the mechanisms of solute transporters, In Membrane Transport – Practical Approach, SA Baldwin, ed. (Oxford: Oxford University Press), pp. 167-189.

Wang, M., Tran, J.H., Jacoby, G.A., Zhang, Y., Wang, F. and Hooper, D.C. 2003a. Plasmid-mediated quinolone resistance in clinical isolates of *Escherichia coli* from Shanghai, China. *Antimicrobial agents and chemotherapy.* **47**(7), pp.2242-2248.

Wang, M.Y., Wang, X. and Guo, D. 2003b. A level set method for structural topology optimization. *Computer methods in applied mechanics and engineering.* **192**(1), pp.227-246.

Ward, A., Hoyle, C., Palmer, S., O'Reilly, J., Griffith, J., Pos, M., Morrison, S., Poolman, B., Gwynne, M. and Henderson, P. 2001. Prokaryote multidrug efflux proteins of the major facilitator superfamily: amplified expression, purification and characterisation. *Journal of molecular microbiology and biotechnology.* **3**(2), pp.193-200.

Ward, A., Sanderson, N., O'Reilly, J., Rutherford, N., Poolman, B. and Henderson, P. 2000. *The Amplified Expression, Indentificaiton, Purification, Assay and Properties of Histidine-Tagged Bacterial Membrane Transport Proteins*. Blackwell Press: Oxford.

Waugh, D.S. 2011. An overview of enzymatic reagents for the removal of affinity tags. *Protein expression and purification.* **80**(2), pp.283-293.

Webber, M. and Piddock, L. 2003. The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy.* **51**(1), pp.9-11.

Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. 1996. Optimization of heterologous protein production in *Escherichia coli. Current opinion in biotechnology.* **7**(5), pp.494-499.

West, C.K.G., Klein, S.L. and Lovell, C.R. 2013. High frequency of virulence factor genes tdh, trh, and tlh in *Vibrio parahaemolyticus* strains isolated from a pristine estuary. *Applied and environmental microbiology.* **79**(7), pp.2247-2252.

Weyand, S., Shimamura, T., Beckstein, O., Sansom, M.S., Iwata, S., Henderson, P.J. and Cameron, A.D. 2011. The alternating access mechanism of transport as observed in the sodium-hydantoin transporter Mhp1. *Journal of synchrotron radiation.* **18**(1), pp.20-23.

Weyand, S., Shimamura, T., Yajima, S., Suzuki, S.i., Mirza, O., Krusong, K., Carpenter, E.P., Rutherford, N.G., Hadden, J.M. and O'reilly, J. 2008. Structure and molecular mechanism of a nucleobase–cation–symport-1 family transporter. *Science.* **322**(5902), pp.709-713.

Whitmore, L. and Wallace, B.A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers.* **89**(5), pp.392-400.

Widdowson, C.A. and Klugman, K.P. 1998. The molecular mechanisms of tetracycline resistance in the pneumococcus. *Microbial drug resistance.* **4**(1), pp.79-84.

Wikström, M., Kelly, A.A., Georgiev, A., Eriksson, H.M., Klement, M.R., Bogdanov, M., Dowhan, W. and Wieslander, Å. 2009. Lipid-engineered *Escherichia coli* membranes reveal critical lipid headgroup size for protein function. *Journal of Biological Chemistry.* **284**(2), pp.954-965.

Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A. and Brinkman, F.S. 2015. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic acids research.* **44**(D1), pp.D646-D653.

Wirmer-Bartoschek, J. and Bartoschek, S. 2012. NMR in drug discovery on membrane proteins. *Future medicinal chemistry.* **4**(7), pp.869-875.

Witholt, B., Boekhout, M., Brock, M., Kingma, J., van Heerikhuizen, H. and de Leij, L. 1976. An efficient and reproducible procedure for the formation of spheroplasts from variously grown *Escherichia coli*. *Analytical biochemistry.* **74**(1), pp.160-170.

Wong, F.H., Chen, J.S., Reddy, V., Day, J.L., Shlykov, M.A., Wakabayashi, S.T. and Saier Jr, M.H. 2012. The amino acid-polyamine-organocation superfamily. *Journal of molecular microbiology and biotechnology.* **22**(2), pp.105-113.

Wright, G.D. 2005. Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced drug delivery reviews.* **57**(10), pp.1451-1470.

Wright, P.M., Seiple, I.B. and Myers, A.G. 2014. The evolving role of chemical synthesis in antibacterial drug discovery. *Angewandte Chemie International Edition.* **53**(34), pp.8840-8869.

Xiong, J. 2006. *Essential bioinformatics.* Cambridge University Press.

Yamashita, A., Singh, S.K., Kawate, T., Jin, Y. and Gouaux, E. 2005. Crystal structure of a bacterial homologue of $Na^+$/Cl--dependent neurotransmitter transporters. *Nature.* **437**(7056), pp.215-223.

Yin, H. and Flynn, A.D. 2016. Drugging membrane protein interactions. *Annual review of biomedical engineering.* **18**.

Yin, Y., He, X., Szewczyk, P. and Chang, G. 2006. Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science.* **312**(5774), pp.741-744.

Young, J.D., Yao, S.Y., Baldwin, J.M., Cass, C.E. and Baldwin, S.A. 2013. The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Molecular aspects of medicine.* **34**(2), pp.529-547.

Zgurskaya, H.I., López, C.A. and Gnanakaran, S. 2015. Permeability barrier of Gram-negative cell envelopes and approaches to bypass it. *ACS infectious diseases.* **1**(11), pp.512-522.

Zhang, Z., Kuipers, G., Niemiec, Ł., Baumgarten, T., Slotboom, D.J., de Gier, J.-W. and Hjelm, A. 2015. High-level production of membrane proteins in *E. coli* BL21 (DE3) by omitting the inducer IPTG. *Microbial cell factories.* **14**(1), p142.

Zhou, Y. and Bowie, J.U. 2000. Building a thermostable membrane protein. *Journal of Biological Chemistry.* **275**(10), pp.6975-6979.