UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL & ENVIRONMENTAL SCIENCES

Biological Sciences

Volume 1 of 1

Using GFP to investigate protein localisation, function and global cellular response

by

Benjamin Yarnall

Thesis for the degree of Doctor of Philosophy

September_2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL & ENVIRONMENTAL SCIENCES

Biological Sciences

Thesis for the degree of Doctor of Philosophy

USING GFP TO INVESTIGATE PROTEIN LOCALISATION, FUNCTION AND GLOBAL CELLULAR RESPONSE

By Benjamin Paul Yarnall

Integral membrane proteins (IMPs) make up 20-30 % of genes in all walks of life. They are major determinants of disease pathology, making them prime therapeutic targets for cancer, bacterial infection and genetic disorders. Despite this, they are underrepresented in the literature; this is often attributed to complications with IMP expression, purification and characterisation. This thesis aimed to tackle the difficulties with IMP characterisation. Exploitation of tagged versions of the IMPs has opened up new research fields, initially based on simple observations. The tag used is the well characterised GFP which has allowed for the simple optimisation of conditions for protein over-expression, determination of intracellular localisation of *H. sapiens* SWEET sugar transporter, development of a ligand binding method that does not rely on properties of the ligand and demonstrated that protein over-expression using the *Escherichia coli* pET system only occurs in mutant forms of this organism. The implication being that protein over-expression only occurs via genetically modified organisms.

Ubiquitously expressed, the MFS is one of the largest protein superfamilies, with roles including metabolite and xenobiotic transport. They have been implicated in the development of antimicrobial resistance in *E. coli*, making them a clinically relevant target. An expression library of 63 GFP-tagged proteins was produced, before screening for optimal expression conditions. The larger amounts of transporter obtained via this approach enabled the implementation of a ligand-binding assay using the technique of thermophoresis. This approach produced novel binding substrates as well as identifying a new binding event for the well characterised drug efflux transporter mdfA. Significantly, the approach has shown that cyclic AMP also binds to mdfA and another MFS transporter kgtP, potentially identifying a novel role for these transporters in the catabolite repression process.

The work presented in this thesis has shown the versatility of a reporter system like GFP to uncover fundamental properties at the cellular level (protein localisation experiments), the biochemical level (optimisation of protein over-expression and ligand binding studies) but also at the cellular level (*E. coli*'s use of mutants during protein over-expression). This research is a building block to identifying new drug targets to tackle the global problem of antimicrobial resistance.

Table of Contents

Table of Tables

Table of Tables

Table of Tables

Table of Figures

Table of Figures

Table of Figures

Academic Thesis: Declaration of Authorship

I, BEN YARNALL

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Using GFP to investigate protein localisation, function and global cellular response

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: .....

Date: .....

Acknowledgements

This thesis is dedicated to the following people, without whom it would not have been possible.

The BBSRC funding body for providing the means to undergo this research programme; and along with the MRC, NERC, SPITFIRE and Kerkut for funding my friends and colleague that made the experience so enjoyable. The Southampton Biological department for providing a platform to present work and discuss science with so many distinguished minds. My supervisors Declan Doyle, Lorraine Williams and Matthew Terry for believing I could undertake the project. Particular thanks to Declan, for his patience, expertise and training in all matters throughout my project. Thank you to Phil Williamson for his support in NMR experiments and his sense of enthusiasm about MATLAB coding. Ivo Tews for his advice on all matters crystallography. Thank you to David Tumbarello for supplying eukaryotic vectors and helping with confocal microscopy. Thank you to Neville Wright for understanding the manual for every piece of equipment in building 85.

The rest of my lab group for providing such an enjoyable working environment, especially Monika for helping to maintain my sanity when results did not go my way. I would like to thank Chris and Matt for undertaking the task of thermoprofiling with me, Sayeed for the endless purifications and Dan Noel for his work with mutant cell lines. Thank you to Matthew Bellamy for his support in undergraduate demonstrating.

With thanks to all the friends I have made along the way. Simon, Nancy, Katie and Flora for providing an enjoyable and friendly environment to start a new job. To Nico, Alex giving me the opportunity to try duck from a can. To Garrick, Oli, Oli and Connor for taking me climbing and making me go to hair metal concerts. To Liam, Liam, Laura, Harry, Scott, Bevill and Sophie for introducing me to the joys of karaoke paired with table football. To the Southampton Taekwondo club for teaching me to touch my toes, with thanks to masters Dave, Cliff, Kike, Vince, Darren and Booker.

To the people facing similar struggles to my own, who helped me face my problems with a grin. To Andy Hutchin, Matt Rodriguez, Moritz Machelett and Chris Holes for help with protein and crystallography work; and help destroying

all challengers at table football. To Luke Evans, Mary Beton for doing NMR to remind me that there are many difficult methodologies. To Megha Rai and John Butler for understanding the difficulties attributed to membrane protein purification.

To my housemates and co-conspirators. Mike Alright, Emma Joslin, Seb Shepherd and Miguel Ramirez for living in a tropical hell, with light stair surfing and a BBQ area. Adham Ashton-Butt and Connie Tremlett for subsidising my final year and providing table football nemeses, may your skills wane in the absence of an appropriate practise arena. To Sarmi Sri and Chrysia Pegasiou for putting up with me when my focus wavered and refraining from strangling me when I was distracting.

To Polly, Sam, Deej and Smee for letting me talk at length about things they had zero interest in. To Catherine and Nick for providing a safe space with all the dog walking I could muster. To Mark and Tom for developing a liking for Cricket and quickly outpacing me in earning potential during this process. To Rosie and Mike for always being interested in what I am doing, no matter how dull it must be to them and no matter how much else they have going on. To princess bubble-Em who supported me throughout, helpfully undergoing a similar process to prepare me for what to expect. Teaching me how to use capital letters and semi-colons and putting up with me during the writing stages.

Definitions and Abbreviations

Abbreviation	Definition	Abbreviation	Definition
AA	Amino Acid	MFS	Major Facilitator Superfamily
ABC transporter	ATP-binding cassette transporter	MIC	Minimum Inhibitory Concentration
AP1	activator protein 1	MT	Microtubule
BLAST	Basic Local Alignment Search Tool	MWCO	Molecular Weight Cut Off
BP	Base Pair	MxD	Cysteamine-Cysteine Mixed Disulfide
CBM	Cellulose-binding Module	NCBI	National Centre for Biotechnology Information
CCMB	competent Cell Preparation Buffer	NM	Non-mutants
CDR	Complementarity Determining Region	OD600	Optical density at 600nm
chocl	Choline chloride	ORF	Open Reading Frame
CP	Cross Polarisation	PCR	Polymerase Chain Reaction
CRP	cAMP-receptor protein	PDB	Protein Data Bank
CTNS	Cystinosin Transporter	POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
DARPin	Designed Ankyrin Repeat Protein	PTS	Phosphotransferase System
DDM	n-dodecyl-β-D-maltopyranoside	PVDF	Polyvinylidene Fluoride
DHA	Dihydroartemisinin	RCR	root mean square correlation ration
DM	n-decyl-β-D-maltopyranoside	RFU	Relative Fluorescence Unit
DxC	Sodium Deoxycholate	RFUmax	Maximum RFU

ESRF	European Synchrotron Radiation Facility	ROS	Reactive Oxygen Species
EtBr	Ethidium Bromide	RPE	Retinal Pigmented Epithelial
FT	Flow Through	SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
GBVR	*Gloeobacter* Biliverdin Reductase	SDW	Sterile Distilled Water
GFP	Green Fluorescent Protein	sfGFP	Superfolder Green Fluorescent Protein
GLUT	Glucose Tranporter	SGLT	Sodium-Glucose Linked Transporter
GPCR	G-protein coupled receptor	SLC2	Solute Carrier Family 2
GST	Glutathione-S-transferase	SLC5	Solute Carrier 5
hBVR	*H. sapiens* biliverdin reductase	SSF	Solute:Sodium Symporter Superfamily
HEK	Human Epithelial Kidney	SWEET	Sugars Will Eventually Be Exported Transporters
His-tag	Histidine Tag	TB	Terrific Broth
HsSWEET	*H. sapiens* SWEET	tDBDF	“two dinucleotide binding domains” flavoproteins
IMAC	Immobilised Metal Affinity Chromatography	TLH	Thermolabile Haemolysin
IMP	Integral Membrane Protein	Tm	Melting Temperature
LacY	Lactose Permease	TMH	Transmembrane Helices
LAMP1	Lysosome Associated Membrane Glycoprotein 1	TPP+	Tetraphenylphosphonium
LB	Luria Broth	UTI	Urinary Tract Infection

LDAO	N,N-dimethyldodecylamine N-oxide	UVL	Unilamellar Vesicles
LIC	Ligation-Independent Cloning	vSGLT	V. *parahaemolyticus* SGLT
LMNG	Lauryl Maltose Neopentyl Glycol	Z-3-12	N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate
MBP	Maltose Binding Protein	Z-3-14	3-(N,N-dimethylmyristylammonio)propanesulfonate
		β-mer	mercaptoethanol

Chapter 1 Using GFP to investigate protein localisation, function and global cellular response

1.1 Characterised membrane proteins are underrepresented in the literature

Membrane proteins are ubiquitously expressed throughout the biological kingdoms and are indispensable for a number of cellular processes. Different types of membrane-bound protein are visualised in Figure 1.1 A. Transporter proteins are required to facilitate passive diffusion, such as the aquaporin transport of water (for review, see Groszmann et al 2017) or to actively transport substrates, such as xylose transported by the MFS transporter xylE (for review, see Shi 2013) across the plasma membrane. Others act as receptors to translate an extracellular signal to an intracellular response, such as intracellular nucleotide exchange activity of GPCR proteins (for review, see Gurevich and Gurevich 2008). Enzymes, O-acyltransferase, may also be membrane bound to catalyse localised reactions (for review, see Masumoto et al 2015).

Correct cellular physiology and function relies on operational membrane proteins. Incorrectly folded integral membrane proteins (IMPs) cause diseases including cystic fibrosis caused by mutation of the cystic fibrosis transmembrane conductance regulator (Zeng et al 1997), deafness caused by mutation of the connexin 26 (Patel et al 2015), developmental deformities caused by mutation of the SLC26 anion exchangers (Kere et al 2006), cerebellum degradation caused by mutation of the ATP7A protein (Lenartowicz et al 2015) and predisposition to cancer caused by mutation of the FANCO gene (Bogliolo and Surrallés 2015). In spite of this, they are underrepresented in literature, compared to soluble, cytoplasmic proteins. While membrane proteins make up around 26 % of protein coding genes in *H. sapien*s (Sundararaj et al 2004) they make up <2 % of the characterised structures (Kozma et al 2013).

This underrepresentation is, in part, due to the additional difficulties associated with producing membrane proteins at almost every stage of the protein production process. This process is summarised in Figure 1.1 B. There is less space within the plasma membrane than within the cytoplasm, contributing to lower yields

compared to soluble proteins (Seddon et al 2004). Expression of recombinant proteins can have toxic effects on the host system (Gubellini et al 2011). Protein purification requires additional steps when compared to soluble protein extraction (Smith 2011), such as membrane solubilisation, which each require optimisation to maintain correct protein structure (Anandan and Vrielink 2016). These difficulties in protein production and purification cause a bottle neck in the characterisation of membrane proteins (Moraes et al 2014). This is reflected in their underrepresentation in the published literature.

This thesis aims to tackle this underrepresentation using green fluorescent protein (GFP) as a reporter gene, to investigate some of these bottleneck stages. In addition to answering questions on protein localisation and function, GFP will be used to optimise protein expression and purification and to investigate the global cellular response to recombinant protein expression. Improving the techniques available for the characterisation of membrane proteins could have vast implications for biology. A key application would be to tackle the growing prevalence of antibiotic resistance in bacteria, an increasingly important healthcare problem in which membrane transporters are proposed to play an important role (Chitsaz and Brown 2017). A better understanding of how membrane transporters facilitate this trend could begin to address the problem.

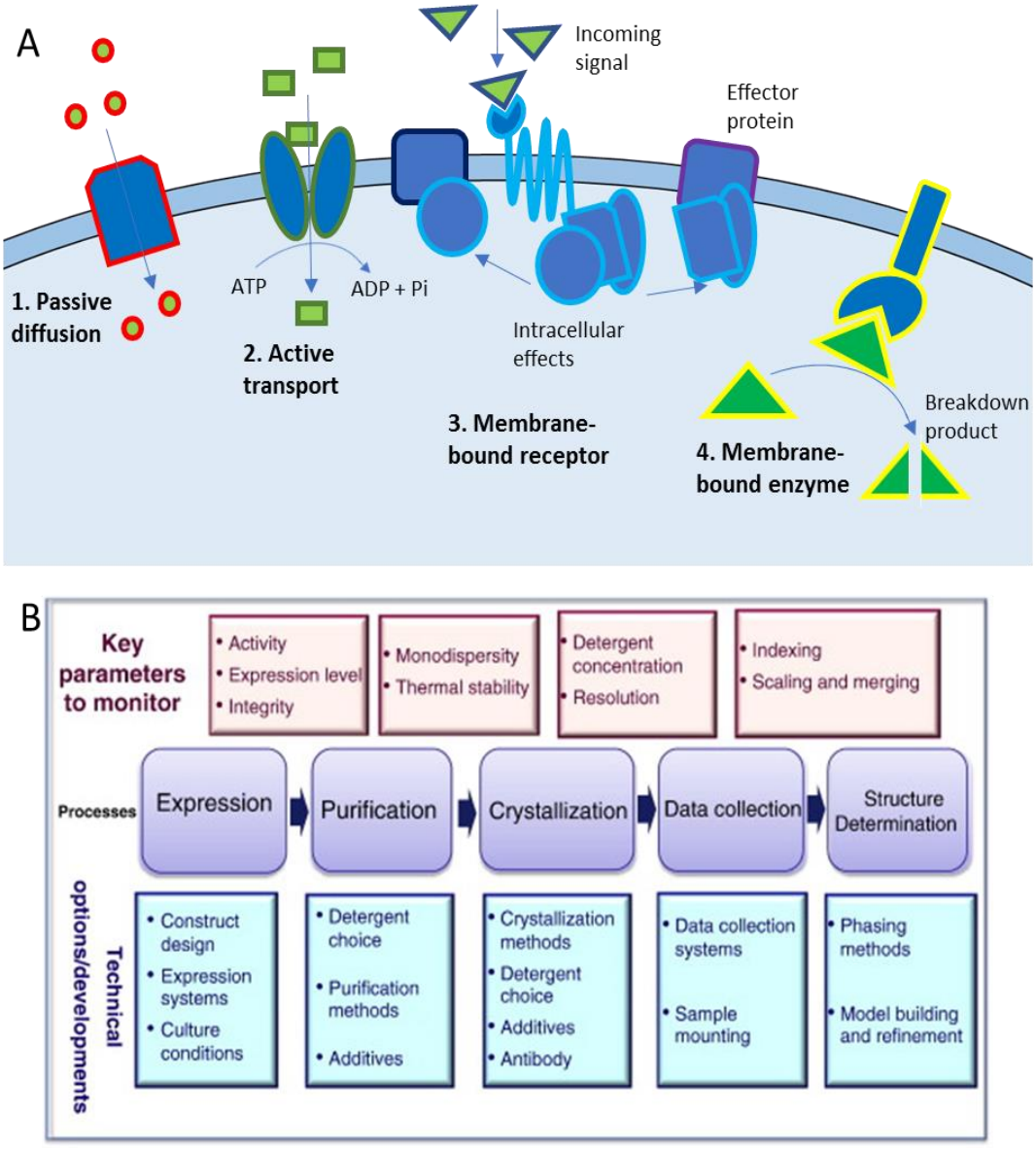


Figure 1.1 Characterisation of membrane proteins

A) Diagrammatic representation of some key membrane protein functions: 1. Passive diffusion and 2. Active transport across the plasma membrane, in exchange for ATP hydrolysis. 3. Membrane-bound receptors to detect incoming extracellular signals, which initiate downstream intracellular effects. 4. Membrane-bound enzymes to catalyse intracellular or extracellular reactions. Membrane proteins are displayed in blue; substrates and signals are displayed in green. B) Key steps in the expression and structural determination of membrane proteins. Includes key parameters to be monitored (top line) and key considerations (lower line) at each step. Adapted from Moraes et al (2014).

1.2 Green Fluorescent Protein discovery and physiology

An interest in bioluminescence was recorded in writing as early as 384 BCE (Harvey 1957). However, with little knowledge of the intricacies of cellular and molecular biology, the mysteries of this phenomenon would not be uncovered for thousands of years. Green fluorescent protein (GFP) was partially purified from *Aequorea*, a genus of bioluminescent jellyfish, by Shimomura et al (1962). Their work identified a novel fluorophore activated by the presence of calcium in aerobic conditions. The new protein was deemed to differ from the previously discovered luciferase, which required ATP and magnesium to fluoresce (Green and McElroy 1956). Further characterisation defined peak excitation emission wavelengths of the purified GFP from *Aequorea*, with values matching those of the original organism tissue (Johnson et al 1962).

Fluorescence is absorption of electromagnetic radiation followed by emission of electromagnetic radiation in the visible spectrum. When incoming radiation hits a substance, it changes the outer shell electron's orbital from a ground state to a higher energy level or excited state (Llères et al 2007). This higher energy orbital can dissipate by relaxation, vibrationally releasing the energy as heat. Alternatively, the electron energy level can jump down to the grounded state releasing the energy as a photon, the visible output of fluorescence. These changes in the electron energy level are visually represented in Figure 1.2 A. In molecules such as GFP the absorbed energy is released in a combination of dissipation to the surrounding solvent and photon expulsion. The emitted radiation is therefore at a lower energy, or longer wavelength, than the absorbed radiation (Phillips 2016).

The native function of fluorescent proteins is still uncertain, although many potential roles have been proposed. A phylogenetic analysis estimated 30 independent examples of fluorescent protein evolution, for example in *Aequorea* and in *Renilla*, a type of coral, suggesting they can provide an adaptive advantage (Hastings 1983). In coral, fluorescent proteins absorb light in wavelengths of low photosynthetic value, reflecting the potentially damaging energy as light and providing photoprotection (Salih et al 2000). Alternatively, GFP evolution may have been driven by the endosymbiosis between coral and algae, providing photoprotection for the algae (Field et al 2006) and protecting the coral from reactive oxygen species (ROS) produced by the algae photosynthesising (Bou-Abdallah et al 2006). Evidence exists for both theories: inactivation of GFP leads to the endosymbiotic algae leaving the coral (Dove, 2004), while changing light

exposure results in a change in GFP concentration produced by the coral (Roth et al, 2010). GFP is also able to reduce key cellular components, such cytochrome C and NAD^+, indicating it may have evolved to facilitate the light-induced exchange of electrons, or possibly acting as a proton pump (Agmon 2005; Bogdanov et al 2009).

While its role was unclear, there was a lot of interest in the light emitting capabilities of fluorescent proteins, particularly in how this was achieved. A common chromophore was identified in the *Aequorea* and *Renilla* GFP. Spectrophotometric analysis was carried out on isolated protein in the native state and after chemical denaturation. The denatured proteins had identical spectrophotometric patterns, as a result of matching chromophores. However, the native proteins had vastly different absorbance peaks, indicating the tertiary structure impacted the chromophores' properties (Ward et al 1980). Successful cloning and sequencing of GFP from *A. victoria* cDNA revealed a 238-residue protein (Prasher et al 1992), allowing creation of expression vectors and paving the way to uncover the mechanism behind GFP fluorescence and its potential for biology.

The smallest components containing active chromophore were identified by high pressure liquid chromatography as having the ability to absorb light at 380 nm. The chromophore also matched the extinction characteristics of those shown by Ward et al (1980; Cody et al 1983). Using the PICO-Tag system (White et al 1986), the amino acids making up the smallest component of the chromophore were identified as Ser, Tyr and Gly (Cody et al 1993). Formation of this chromophore was shown to be hindered when GFP was expressed anaerobically, but fluorescence was restored upon exposed to oxygen, indicating an oxidation event is required for the formation of the chromophore (Heim et al 1994). This reoxidation process did not occur on the same time scale as that of protein folding post-denaturation and eventually it was concluded that GFP was folded in three independent steps: formation of the tertiary structure, circularisation to form the chromophore, and oxidation of the chromophore (Reid and Flynn 1997). This is summarised in Figure 1.2 B. Comparison of fluorescent proteins shows strong conservation of the residue Glycine 67, indicating it may be important in the nucleophilic movement required for cyclization (Tsien 1998).

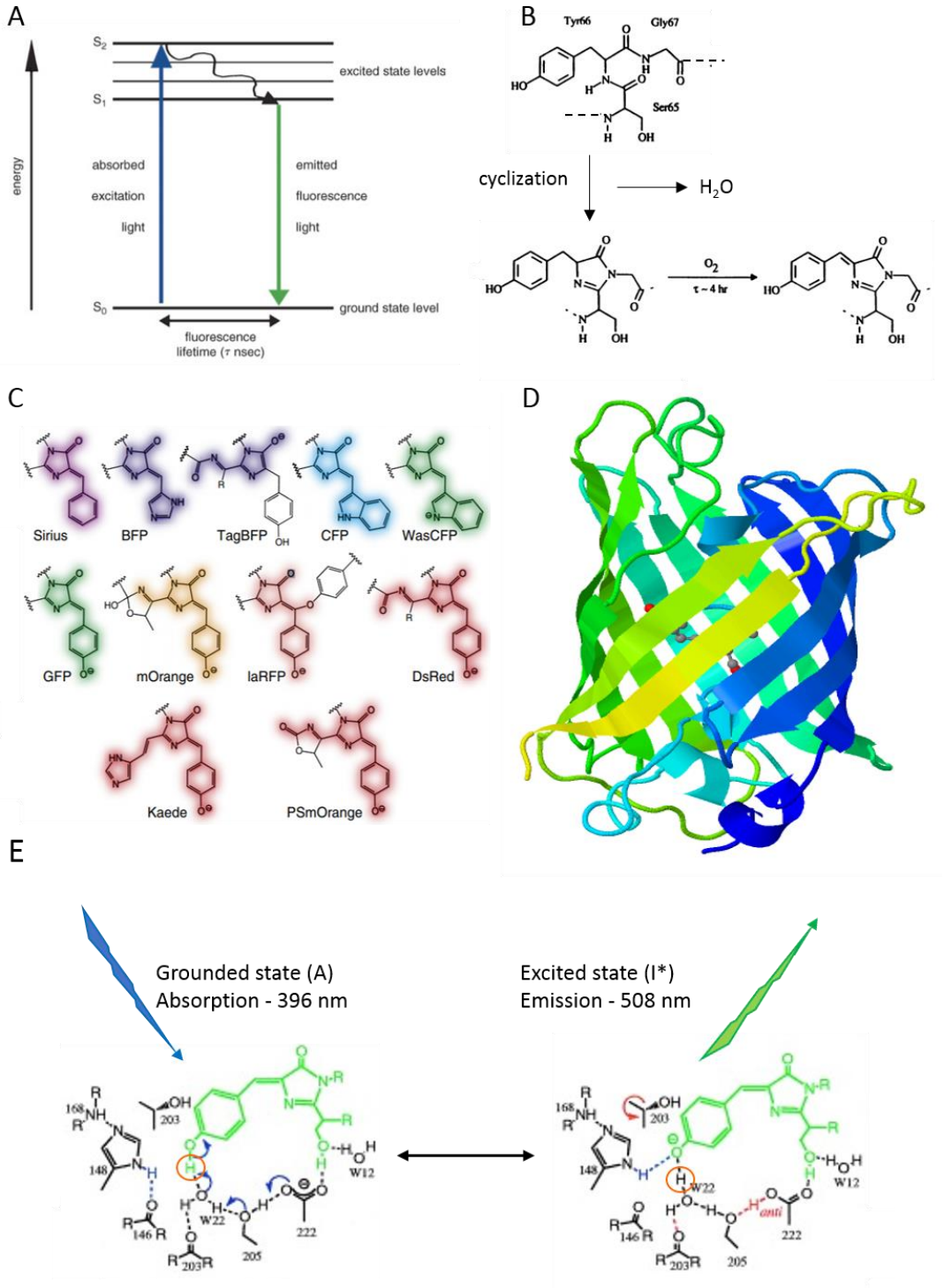


Figure 1.2 Visual representation of fluorescent proteins chromophore and interactions with the surrounding amino acids during fluorescence

A) Visual representation of electron energy excitation levels during absorption and fluorescent emission (Llères et al 2007). B) Proposed formation mechanism of the GFP chromophore region, following enzyme folding assays. During formation of the tertiary structure, the chromophore position enables cyclization of the glycine and serine residues. Cyclization is followed by oxidation of the tyrosine R-group completing the active chromophore (Heim et al 1994). C) Primary structural representation of a range of

chromophores including GFP and other fluorescent proteins (Mishin et al 2015). D) The tertiary structure of *A. victoria* GFP. Protein rainbow colouring from N-terminal blue to C-terminal yellow. Chromophore regions shown as ball and stick (Ormo et al 1996). E) Structural representation of GFP chromophore and surrounding amino acids, in grounded (A) and excited (I*) states. The energy from photon absorption enables electron movement (blue arrows) culminating in proton transfer (circled in orange). The emission of a photon from the excited state reverses these changes, allowing return to the grounded state. Dotted lines represent hydrogen bonds. Adapted from Brejc et al (1997)

An understanding of the chromophoric region of GFP enabled further study of this area through specific mutations, showing even single amino acid changes can alter the absorbance and emission spectra. Heim et al (1994), for instance, found replacing tyrosine with histidine at residue 66 (Y66H) altered the spectra to produce a blue variant. A review by Mishin et al (2015) summarises some of the available chromophore alternatives, including purples, reds and oranges (Fig. 1.2 C).

Expression of GFP in systems such as *E. coli* and *C. elegans* demonstrated correct fluorophore production and detected no toxic effects (Chalfie et al 1994), fuelling speculation of the potential uses of recombinant GFP in molecular biological investigations. The crystal structure of GFP is shown in Figure 1.2 D, as identified in parallel by two groups (Ormo et al, 1996; Yang et al, 1996). GFP from *A. victoria* was shown to be formed of an 11 stranded β-barrel cylinder with an alpha helix running through the centre, containing the chromophore region. The centrally-located chromophore is protected from extramolecular solvent, although some water molecules are present in the cavity. The cylinder is maintained by inter β-sheet hydrogen bonds, possibly explaining the stability of the protein. Analysis of 266 fluorescent proteins in the Protein Data Bank (Berman et al 2000) showed the "lid" residues at the top and bottom of the barrel are highly conserved, indicating functional importance (Ong et al 2011).

The fluorescence mechanism was modelled by comparing wildtype and mutant structures of GFP (Brejc et al 1997), shown in Figure 1.2 E. In the chromophore-grounded state (A) there are two networks of amino acids surrounding the serine-tyrosine-glycine chromophore. The first is comprised of the amino acids Glu222, Ser205 and Thr203 interacting with the chromophore through two water molecules. Glu222 interacts with the chromophore serine, while Ser205, Thr203 and the chromophore tyrosine all interact with the same water molecule.

Absorption of UV light promotes the electrons in the double bond between tyrosine and glycine to a higher energy, non-bonding orbital. This is facilitated by

deprotonation of the tyrosine hydroxyl group enabling rearrangement of the phenyl side group electrons (Weber et al 1999). The anionisation of tyrosine is facilitated by electron movement of surrounding amino acids Glu222, Ser205 and Thr203, and proton transfer to Ser205. The newly formed anion is stabilised by interacting with the side chain of His148. This newly formed tertiary structure reduces vibrational dissipation of the higher energy orbital (Remington 2011). This enables fluorescence, as the excited orbital of the I* state lowers to the orbital of the grounded state, it releases surplus energy as a photon. This reverses the movement of protons and electrons, returning the system to the grounded state.

The generation of expression plasmids enabled optimisation of GFP. Codon optimisation followed by DNA shuffling was used to produce more efficient mutants (Crameri et al 1996). The fluorescent nature of the protein enables simple identification of beneficial mutations. Crameri et al (1996) demonstrated a GFP variant that was more soluble, locating more frequently to the cytoplasm than wild type GFP. Tests with CHO cells also showed an improvement in fluorescence when expressed in eukaryotic cell lines. The technique of DNA shuffling was later used to generate super-folder GFP, a variant resistant to aggregation that is triggered by recombinant expression with a poorly folding polypeptide tag (Pedelacq et al 2005).

1.3 Alternative fluorescent proteins

GFP is not the only fluorescent protein of scientific significance. As previously mentioned, characterisation of luciferase ran in parallel to that of GFP. Luciferase was first isolated in high yields from fire flies by Green and McElroy (1956), noting that luminescence requires magnesium and energy in the form of ATP. As well as the firefly, *P. pyralis*, luciferases can be found in bacteria (Cui et al 2014) and sea life such as the coral *Renilla*. Interestingly, in *Renilla*, GFP can be excited by either light or by luciferase: luciferase is activated by ATP and emits light at a lower wavelength, which is transferred to excite GFP (Ward and Cormier 1979).

The *P. pyralis* luciferase was further characterised by cloning and expressing in *E. coli* (De Wet et al 1985). Its use as a reporter gene when characterising the constitutive CaMV 35S plant promoter highlighted its' potential for biotechnology (Ow et al 1987). Figure 1.3 A shows the structure of the *P. pyralis* luciferase: the 62 kDa protein comprises of two compact domains, formed of both α-helices and β-sheets (Conti et al 1996). The structure is dissimilar to that of GFP, demonstrating

the GFP β-barrel is not the only structure enabling fluorescence. However, the environmental protection the β-barrel provides to the chromophore makes it a common structure: fluorescent proteins isolated from *A. japonica, A. victoria*, soft corals and sea anemones have all been structurally characterised and shown to rely on a β-barrel structure (Wiedenmann et al 2004; Gurskaya et al 2006; Andresen et al 2007; Kumagai et al 2013). One possible disadvantage of barrel structure of reporter genes in molecular biology is the size. For this reason, there has been an interest in smaller fluorescent proteins formed with an exposed chromophore. Several examples of alternative fluorescent proteins are summarised in Table 1.1 and illustrated in Figure 1.3 C – G, alongside the structure of GFP (Fig. 1.3 B).

Table 1.1 Examples of fluorescent proteins utilised as reporters in molecular biology

Structural description concerns symmetry and secondary structure make up, these are shown in Figure 1.3, as listed. Internal chromophores are protected from the environment by the proteins tertiary structure. External chromophores are exposed to their solvent environment. Sources dictate the structural characterisation of each protein.

Protein	**Organism**	**Structural details**	**Chromophore**	**Source and PDB ref**
GFP (Fig. 1.3B)	*A. victoria*	β-barrel structure surrounding central chromophore. Regular appearance.	Internal	Ormo et al 1996 - 1EMA
IFP2.0 (Fig. 1.3C)	*D. radiodurans*	Globular structure formed of α-helices and short β-sheets. Chromophore is located at one edge of the protein. Requires biliverdin for fluorescence.	External	Yu et al 2014 - 4CQH
LOV (Fig. 1.3D)	*A. thaliana*	Compact protein formed of α-helices	External	Christie et al

		and β-sheets. Regular appearance. Flavin mononucleotide based chromophore		2012 - 4EES
UnaG (Fig. 1.3E)	*A. japonica*	β-barrel structure, smaller than that of GFP. Chromophore formed of bilirubin.	External	Kumagi et al 2013 - 4I3B
SNAP (Fig. 1.3F)	*H. sapiens*	Compact protein formed of α-helices and β-sheets. O6-alkylguanine-DNA alkyltransferase exposed to fluorescent O6-benzylguanine (BG) derivatives.	External	Mollwitz et al 2012 - 3KZY
Haloalk-ane dehaloge-nase (Fig. 1.3G)	*R. rhodocrous*	Protein larger than GFP (294 amino acids). Mostly formed of α-helices. Enzymatic activity used to covalently attach fluorophore.	External	Newman et al 1999 - 1BN7

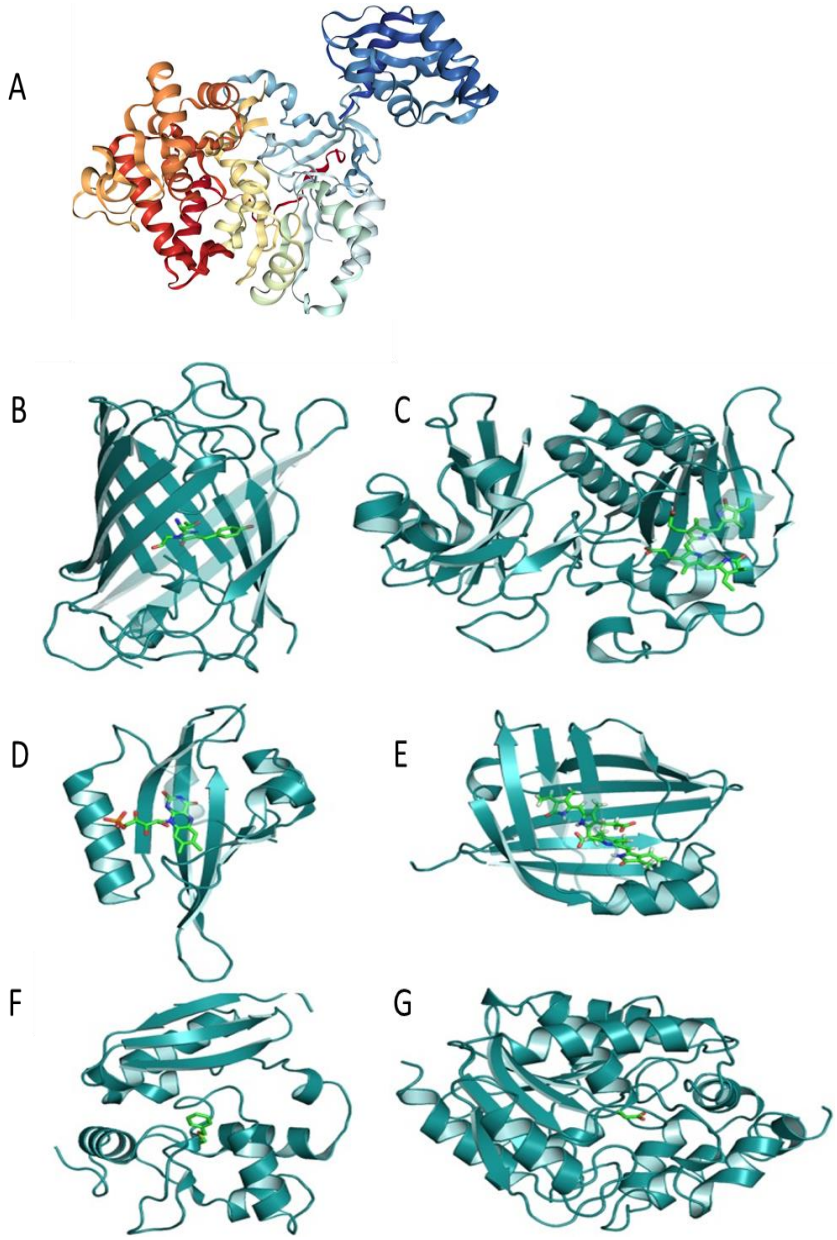


Figure 1.3 Structural comparison of fluorescent proteins

A) Tertiary structure of firefly luciferase. Rainbow colouring from N-terminal red to C-terminal blue (Conti et al 1996, PDB - 1LCI). B – G) Comparison of GFP (B) tertiary structure of with a range of other fluorescent proteins in cartoon format; chromophore is represented by a ball and stick model. B) GFP from *A. victoria* PBD - 1EMA; C) IFP2.0 from *D. radiodurans* PDB - 4CQH; D) LOV from *A. thaliana* PDB - 4EES; E) UnaG from *A. japonica* PDB - 4I3B; F) SNAP tag PDB - 3KZY; G) Halo tag PDB - 1BN7. Adapted from Thorn (2017). Proteins summarised in Table 1.1.

1.4 Using GFP fluorescence as a reporter

Although there are many alternative reporting signals for use in cellular and molecular biology, the relative inertness and ability to function without cofactors may account for the continued popularity of GFP. This section will focus on the different uses of GFP as a reporter gene.

- GFP as a reporter of protein yield and expression

Reporter genes attach to proteins, promoters or characteristics of a cellular system and can be used to investigate a variety of problems (Rosochacki et al 2002). Although it is not perfect, as described in more detail in Section 1.5, GFP is a useful reporter as it is relatively small and inert, enabling its incorporation into a system with minimal disruption (Jensen 2012). This has led to the use of GFP as a method of screening for protein expression, as its fluorescence is directly proportional to gene expression (Soboleski et al 2005) and protein concentration (Lo et al 2015), allowing direct quantification of expression without any extra sample preparation; whole cells, suspended pellets or purified samples can be tested neat if required. Plate readers can be used to analyse hundreds of samples in rapid succession enabling large scale screening experiments.

Protein expression and purification is a time-consuming and labour-intensive process. As described at the beginning of this introduction, there are many reasons why it is not always possible to purify the precise protein of interest, and often it is necessary to investigate a homologous protein instead. For instance, Brunner et al (2014) screened 80 homologues of the novel TMEM16 family to find one suitable for crystallography. As well as identifying appropriate targets, the purification conditions need to be considered, requiring further screening (Ericsson et al 2006), as summarised in Figure 1.4 A. This process is simplified by GFP-tagging of proteins, generating numerical values for fluorescence and quantifying protein concentration, enabling easy comparison of different expression and purification conditions. The use of GFP in this process has been widely utilised in a variety of expression systems, including *E. coli* (Bird et al 2015) *P. pastoris* (Brooks et al 2013), HEK293 (Chaudhary et al 2011) and SF9 (Chen et al 2012).

Other methods exist to quantify protein expression and compare expression conditions, but they may be less suitable than GFP. Western blotting with primary antibodies for a recombinant protein's tag can be used to screen expression conditions (Psakis et al 2007). Using Western blots also ensures the purified protein

is of the correct size, minimising noise generated by degraded protein. The protocol requires cell lysis, SDS-PAGE and Western blotting, however, making it less appropriate for larger screens. It is possible to use fluorophores and stains, such as SYPRO Orange, to quantify and compare the level of soluble protein denaturation in different purification conditions, but this is not possible with membrane proteins (Ericsson et al 2006). Gel filtration may be used to optimise purification conditions for membrane proteins, (Hammon et al 2009) but this is much more laborious than spectrofluorometric analysis, making it more difficult for larger screens.

GFP may, therefore, be more appropriate as a key part of the purification protocol, when screening large numbers of samples or purification conditions. The solubility of Superfolder GFP (sfGFP) has been shown to improve the correct folding of some more unstable peptide sequences (Pedelacq et al 2005). GFP can also be used as a tag for affinity purification, using immobilised GFP antibodies or small synthetic ligands (Pina et al 2015). Part of this thesis will focus on the use of GFP as a reporter of protein yield and expression, when screening for different conditions for purification and expression.

- GFP as a reporter of cellular processes

As described at the start of this introduction, one possible reason that it is difficult to purify membrane proteins is the toxic effect of expressing recombinant protein on cellular physiology. GFP can be used as a reporter to monitor cell viability or to track cellular stress after exposure to different environments (Elliott et al 2000; Landete et al 2014).

Cell viability is concerned with the number of healthy cells in a sample. It is important for analysis of frozen tissue samples and toxicity assays (Riss and Moravec 2004; Weinber et al 2009). EtBr is an alternative option, staining for dead cells, but is toxic to the experimenter. The relatively inert nature of GFP makes it a suitable cellular sensor and is a good indicator of cell viability, for example in rat adenocarcinoma cells transfected with GFP (Elliot et al 2000). Other GFP variants have been constructed, allowing tracking of various cellular conditions including pH (Chen et al 2013; Fig. 1.4 B), and intracellular calcium concentrations (Tantama et al 2011; 2013). There are very few alternative methods for detecting intracellular pH, making these adapted proteins useful in research (Chen et al 2013). The stability of GFP enables exposure to a variety of alterations without interfering with fluorescent function.

It is also possible to use GFP as a reporter for the presence of certain compounds within a cell. Thermolabile haemolysin (TLH) is a toxin from the lethal food-borne bacterium, *V. parahaemolyticu*s. By inserting the antigen recognition region for TLH, known as the Complementarity Determining Region (CDR), into sfGFP, it is possible to create hybrid GFP-antigens capable of recognising TLH without the need for additional tags or secondary binding reagents (Wang et al 2014). This highlights the potential for GFP as a reporter to quickly identify therapeutically relevant compounds in samples, which may improve existing antibody therapies.

- GFP as a reporter of protein function

Resolving protein function improves understanding of how a protein contributes to cellular development, homeostasis and interaction with the environment. To determine this function, it is important to investigate a protein's subcellular location and how it interacts with the rest of the proteome and elements in its environment. GFP has been well utilised to answer these questions.

GFP can be used to investigate interaction between proteins. This is summarised in Figure 1.5 A: co-expressed proteins tagged with GFP fragments will fluoresce if the fragments are brought into contact with each other, indicating interaction (Kanno et al 2011). Many techniques are available to show binding or transport of proteins, including NMR, coprecipitation, competition assays and radiolabelled substrates (Gilman 1970; Elion 2006; Fluman et al 2014; Teilum et al 2017). While NMR, for instance, can provide information regarding binding sites (Teilum et al 2017), some of these alternative techniques are impractical for screening purposes, as they require labelling of substrates or acquisition of specific antibodies.

It is also possible to assess function using less direct methods, such as in plate-based studies. Observing the effect of overexpressing a protein on a cell's ability to grow on different environmental stresses, such as antimicrobials, can help predict function (Nishino 2001). However, changing the expression profile of one protein can alter expression levels of hundreds of other native proteins (Zhou et al 2013). It is therefore possible that the resultant phenotype results as an effect of expression interference, rather than directly caused by the protein of interest. For example, deletion of the xenobiotic transporter mdfA (Heng et al 2015) causes increased concentrations of cellular arabinose (Koita 2012) through a currently uncharacterised method. It may therefore be necessary to characterise a variety of phenotypic effects on knockout mutants or overexpressing cell lines to fully understand a genes role.

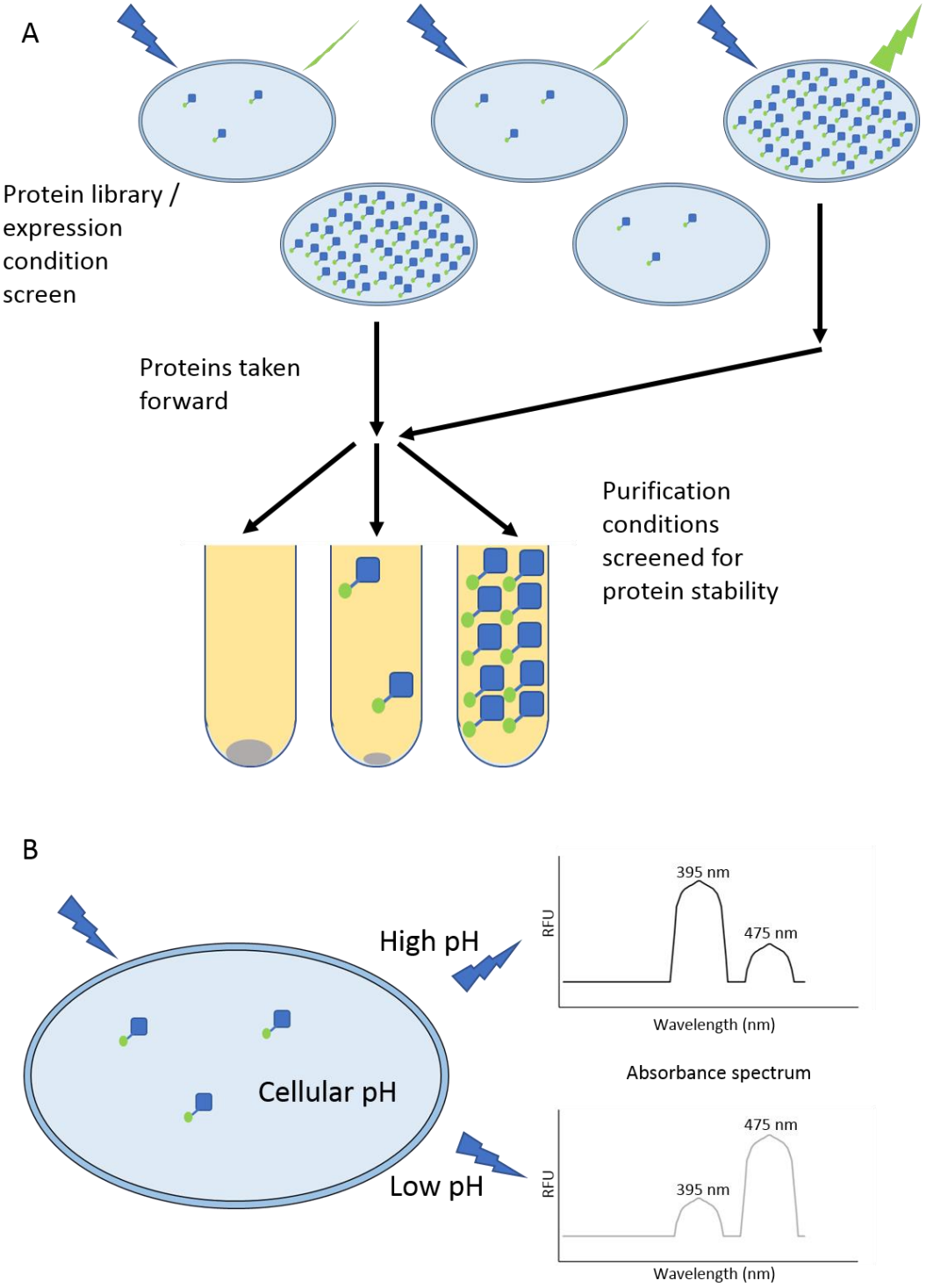


Figure 1.4 GFP can be used as a reporter for various assays, including A) purification screening and B) intracellular pH measurement

A) GFP used for protein expression and purification screening. A protein library screen tests multiple homologues and expression conditions simultaneously. Well-expressed proteins are taken forward to purification screening. Different purification conditions are screened and tested for protein stability with spectrofluorimetry used to optimise yields. B) Cells transformed with pH-sensitive GFP, to measure intracellular or compartmental pH.

The ratio of Absorbance spectrum peaks can be used to measure cellular pH in response to stimuli.Cells shown in blue and purified sample shown in yellow.

It is also necessary to determine protein localisation within the cell when investigating protein function (Rudner and Losick 2010, Scott et al 2005). Although not perfect, as described further in Section 1.5 GFP-tagging provides information on protein localisation in bacterial and eukaryotic systems (Figure 1.5 B; Britton et al 1998, Lippincott-Schwartz and Smith 2002). Cell fractionation is an alternative method for determining subcellular localisation, but this can lead to issues with artefacts (Suzuki et al 2010). Western blots with antibodies targeting the protein of interest can also be used to investigate subcellular localisation (Liu 2002). Alternatively, immunocytochemical or histochemical staining can provide different expression information. Proteins are not uniformly expressed and are often regulated by cell type or environmental stimulus. Placing GFP under control of the same promoter as the gene of interest can provide information on tissue-specific or situational gene expression, shown in Figure 1.5 C (Cormack 1998). Visible proteins, such as luciferase, or proteins enabling histochemical staining, such as β-glucuronidase, can also be used for this process (Smirnova and Ugarova 2017). Both GFP and staining techniques are used in current literature and the protein under investigation may dictate which technique is most suitable.

It is also possible to use GFP to take real time measurements of the cellular environment. This can be done using GFP sensors or attaching GFP to promoters controlled by the characteristic of interest. The effect of protein-upregulation or disruption on specific cellular components or processes that can be quantified with GFP, such as pH or sugar concentration, can be used to investigate function (Tantama et al 2011; Koita et al 2013). Using CRISPR/Cas9 genome editing, GFP can be precisely placed into a genome in the place of the native gene (Ota et al 2016). This could theoretically be used to investigate gene upregulation in response to a stimulus, by quantifying changes in promoter activity using GFP fluorescence.

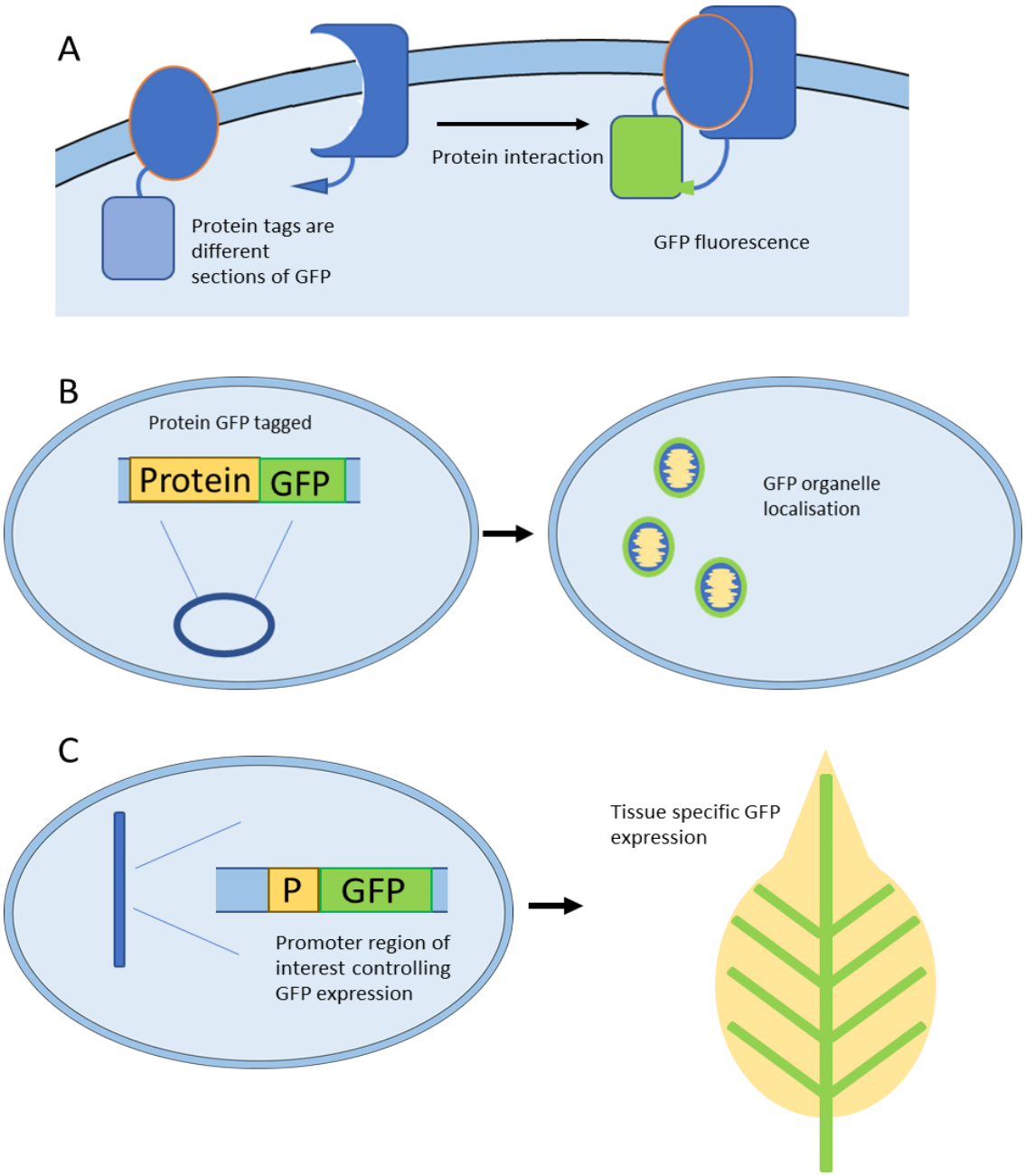


Figure 1.5 GFP can be used to investigate elements of protein function

A) Protein interaction assay using GFP. The proteins under investigation are each tagged with a fragment of the GFP protein. When protein interaction occurs, GFP is formed, producing a measurable signal. B) Protein subcellular localisation assay using GFP. Cells are transfected with an expression plasmid containing a GFP-tagged version of the query protein. Fluorescence or confocal microscopy can determine the subcellular localisation of the protein. Co-expressing fluorescent markers or organelles stains such as DAPI, can be used to observe colocalization. C) Protein tissue specificity assay using GFP. GFP is stably integrated into a multicellular organism under the control of the query promoter (P). GFP expression shows the tissues in which the promoter is active; this example shows a hypothetical plant vascular promoter.

1.5 Complications associated with the use of GFP as a reporter

This chapter has discussed the wide-ranging applications of GFP. There have, however, been reports of GFP interfering with protein function or creating artefacts, potentially leading to altered protein yield, function or localisation.

Hammon et al (2009) found expression yields were hindered by GFP tags, suggesting that new expression vectors without GFP were required for purification (Hammon et al 2009). Subcellular localisation may also be affected by GFP tagging. Palmer and Freeman (2004) screened the localisation of 16 proteins with N- or C-terminal GFP tags. When N-terminally tagged, expression was completely abolished in 4 cases, while the subcellular localisation was disrupted in a further 4 (Palmer and Freeman 2004). Skube et al (2011) tagged EB1, a regulator of microtubule (MT) dynamics, with GFP. When N-terminally tagged, the protein localised to the cytosol and was non-functional. When C-terminally tagged, EB1 correctly localised to MT tips, but it no longer interacted with CLIP-170, another important MT regulatory protein (Skube et al 2011).

Cellular disruption of protein function is likely to cause knock-on effects in multicellular organisms. NMR-based metabolomics found the quantity of 16 out of 57 metabolites differed between GFP tagged transgenic and control mice (Li et al 2013). Additionally, work with recombinant mice found GFP expression under the control of the cardiac α-myosin promoter caused cardiomyopathy (Huang et al 2000). It is therefore worth considering the potential implications of GFP tagging when interpreting results. However, while the tag may have implications for any expression system, it also vastly improves the screening potential when used appropriately.

1.6 Project aims and applications

As described at the beginning of this introduction, membrane proteins make up a significant proportion of the proteome but are comparatively underrepresented in the literature. Difficulties associated with the production and isolation include protein yield, stability during purification and toxicity caused by overexpression.

Working to reduce the impact of these problems could have huge implications for future investigations, improving structural and functional characterisation of vital membrane proteins. Mutations in membrane proteins cause debilitating diseases

such as cystic fibrosis (Zeng et al 1997) while their incorrect regulation is associated with the development of diseases such as cancer (Yamamoto et al 1990). Other membrane proteins contribute to drug resistance in pathogens (Nishino et al 2001) and cancerous cells (Sun et al 2012), both increasingly important problems for healthcare.

Improving the ability to produce and characterise these proteins improves our ability to develop new medicines and drugs (Callaghan et al 2014). Sugar transporters have also been proposed as therapeutic targets in cancer medicine (Yamamoto et al 1990), reducing energy to cancer cells, interrupting their metabolic process and hindering their proliferation (Fantin et al 2006). A recent review commissioned by the UK government estimated the death toll of antimicrobial resistance bacteria would rise to 10 million a year by 2050 if current trends are not countered (O'Neill et al 2014). Isolating key membrane proteins in the process would help in the development of additional antimicrobials to reduce their impact (Tegos et al 2011).

This project aims to utilise GFP as a reporter to investigate the function of novel proteins. *E. coli* based expression systems will be utilised for the majority of this project, with the view to optimise purification systems and purify practical amounts of protein. RPE (Retinal Pigmented Epithelial cells) and HELA expression systems will be used to determine localisation of eukaryotic proteins. The aims are as follows:

- Characterisation of the *H. sapiens* SWEET1 protein

The SWEET1 gene will be cloned in *E. coli* expression vectors, with and without a GFP tag. Expression will be followed by purification-optimisation to produce practical protein yields. GFP-tagged protein will be used to investigate substrate binding. Untagged protein will be purified to homogeneity, with the view to set up crystallisation trials for structural characterisation. A second set of eukaryotic vectors will be cloned to investigate organelle localisation.

- Characterisation of the Major Facilitator Superfamily of membrane proteins

An expression library will be set up, by cloning MFS genes from *E. coli* MG1655 genomic DNA into a plasmid with a GFP tag. Proteins will be screened for the best expression and purification conditions in BL21 (DE3) cells, with a preference for MFS proteins thought to have implications for drug resistance. Once suitable

candidates are identified, GFP-tagged protein will be used to develop a ligand binding assay.

· Investigation into the toxic effects of T7 controlled protein expression in *E. coli*

Expression of proteins in *E. coli* can cause cellular toxicity, severely limiting protein yield. GFP can be used to investigate the toxicity of T7-induced protein expression in the BL21 (DE3) cell line. Strains have been developed for the expression of toxic heterologous integral membrane proteins, but little has been done to improve systems for native proteins. It will be investigated whether GFP can be used to determine whether protein expression is still active and investigate adaptations of cells to escape the toxic effects.

Chapter 2 Materials and methods

2.1 General molecular biology

2.1.1 Polymerase chain reaction (PCR) amplification for expression vector construction

PCR was used to construct and analyse protein expression vectors. GoTaq polymerase (Promega, UK) was used for analysis of constructed expression vectors. GoTaq reagents and run parameters are shown in Tables 2.1 and 2.2. Phusion High-Fidelity DNA Polymerase (ThermoFisher Scientific Ltd, UK) was used for cloning of expression vector inserts. Phusion polymerase reagents and run parameters are shown in Tables 2.3 and 2.4. All PCR was carried out on a DNA engine TETRAD 2 (Bio-Rad, UK). Primers are listed later in this chapter.

Table 2.1 Typical PCR conditions for analytical PCR using GoTaq polymerase

Reagent	**Stock concentration**	**volume µL**	**Final concentration**
GoTaq 10 X buffer	200 mM Tris HCl (pH 8.4), 500 mM KCl, 50 mM MgCl2	1	20 mM Tris HCl (pH 8.4), 50 mM KCl, 5 mM MgCl2
dNTP mix	10 mM (each dNTP)	0.2	200 µM
Forward primer	100 pmol/µl	0.5	0.5 µM
Reverse primer	100 pmol/µl	0.5	0.5 µM
DNA template	Various ng/µl	1	-
GoTaq	5 u/µl	0.05	0.25 u
Sterile distilled water (SDW)	n/a	6.65	n/a

Table 2.2 Typical PCR thermoprofile parameters for analytical PCR using GoTaq polymerase

Step	Temperature (˚C)	Time (s)	Cycle No
Initial Denature	98	300	1
Denature	98	30	35
Annealing	45	30	
Extension	72	60 s/kb	
Final extension	72	300	1
Hold	4	-	1

Table 2.3 Typical PCR conditions for cloning PCR using Phusion polymerase

Reagent	Concentration	Volume µL	Final concentration
HF 5X buffer	125 mM TAPS-HCl, 250 mM KCl, 10 mM $MgCl_2$, 5 mM β-mercaptoethanol	10	25 mM TAPS-HCl, 50 mM KCl, 2 mM $MgCl_2$, 1 mM β-mercaptoethanol
dNTP mix	10 mM (each dNTP)	1	200 µM
forward primer	100 pmol/µl	0.25	0.5 µM
reverse primer	100 pmol/µl	0.25	0.5 µM
DNA template	Various ng/µl	various	0.5 µg/50 µl
Phusion polymerase	3 u/µl	0.25	0.75 u
SDW	n/a	38.25 - template	n/a

Table 2.4 Typical PCR thermoprofile parameters for cloning PCR using Phusion polymerase

Step	Temperature (˚C)	Time (s)	Cycle No
Initial denature	98	300	1
Denature	98	10	35
Annealing	45-60 (45 unless stated)	10	
Extension	72	15 s/kb	
Final extension	72	300	1
Hold	4	-	1

PCR constructs and primers can be found throughout materials and methods. PCR products were analysed with DNA agarose gel electrophoresis. For cloning, PCR products were gel purified.

2.1.2 Gel electrophoresis to assess DNA and generate purified vector inserts

DNA was analysed by gel electrophoresis in an agarose gel. 1 % (w/v) agarose (ThermoFisher Scientific Ltd, UK) was heated in 1X TAE buffer (40mM Tris, 20 mM acetic acid, 1 mM EDTA) then allowed to cool to 50 ˚C. 2.5 µL of DNA stain NANCY-520 (Sigma-Aldrich, UK) per 100 mL was added and the gel was set in a tray. Samples were loaded with 1 X DNA loading buffer (4 % sucrose, 0.025 % bromophenol blue Sigma-Aldrich, UK) and alongside a molecular weight marker (Hyperladder 1kb, Bioline, UK or GeneRuler 1kb ladder, ThermoFisher Scientific Ltd, UK) run at 120 V. Gels were visualised with a UV Transilluminator.

2.1.3 DNA purification of expression vector inserts following PCR and gel electrophoresis

DNA was purified from both treated PCR products and from agarose gels. DNA was purified following PCR product treatment using Gene-JET PCR Purification Kit, and following gel electrophoresis using Gene-JET Gel Extraction Kit (both ThermoFisher Scientific Ltd, UK). Both protocols were undertaken according to manufacturer's instructions. All centrifuge steps were performed using a desktop centrifuge at 12

000 g. DNA was quantified using a Nanodrop 2000c UV-Vis spectrophotometer (ThermoFisher Scientific Ltd, UK)

2.1.4 Cloning

2.1.4.1 Ligation-dependent T4 cloning of PQ loop genes

Prior to T4 cloning, insert DNA was digested to generate sticky ends. 3 µg of insert DNA or 6 µg of vector DNA was mixed with 3 µL of restriction enzyme to a final concentration of 0.06 u/µL (Promega, UK, unless otherwise stated), 5 µL of appropriate restriction enzyme buffer and 0.1 µg of acetylated BSA. The remaining volume was made up to 50 µL with sterile distilled water (SDW). The digest reaction was incubated at 37 ˚C for 4 hours, before heating at 70 ˚C for 15 minutes to deactivate the restriction enzyme. Table 2.5 shows digestion detail for each insert and vector cloned in this way.

Ligation-dependent cloning reactions were ligated using T4 DNA Ligase (Promega, UK). Ligation reactions were set up as in Table 2.6 and incubated for 4 hours at 20 ˚C, followed by 10 minutes at 70 ˚C to deactivate the T4 ligase. Ligation reactions were transformed into *E. coli* on agar plates with appropriate selection pressure, as described in Section 2.2.3.

Table 2.5 Gene and vector components used in ligation dependent T4 cloning

HsSWEETCysless (Gene ID: 55974) and CTNSCysless (Gene ID: 1497) from *H. sapiens)* were *E. coli* codon optimised and synthesised by BioMatik (www.biomatik.com). Histidine codons in transmembrane regions are replaced with valine codons and those in soluble regions are replaced with serine codons. TOPCONS analysis was used to predict transmembrane regions. All codons were replaced with the most frequently used codons identified by Maloy et all (1996) PQLC1 (PlasmidID HsCD00332540), PQLC2 (MmCD00318405) and PQLC3 (HsCD00333475) were ordered from the Harvard Clone Collection. All restriction enzymes were purchased from Promega. Components were digested and purified prior to cloning.

Insert	**Enzyme to digest insert**	**Vector**	**Vector reference**	**Enzyme to digest vector**
GFP	SalI and HindIII	pET27b	Novagen, UK #69863-3	SalI and HindIII
HsSWEETcysless **PQLC1** **PQLC2** **PQLC3** **CTNSCysLess2** **CTNSCysLess3**	EcoRI and SalI	pET27b-GFP	See above	EcoRI and SalI
HsSWEETcysless	HindIII and BamHI	pEGFP-N1	Clontech, USA # 632469	HindIII and BamHI
HsSWEETcysless	HindIII and BamHI	pEGFP-C1	Clontech, USA # 632470	HindIII and BamHI
HsSWEETcysless	HindIII and BamHI	pDDGFP	Robert Campbell (Addgene plasmid # 40286)	HindIII and BamHI

Table 2.6 T4 ligation cloning conditions

DNA components were pre-treated with restriction enzymes prior to T4 cloning.

Reagent	Volume or concentration
Vector DNA	100 ng
Insert DNA	17 ng
Ligase 10 X buffer	40 mM Tris-HCl pH 7.8, 10 mM MgCl2, 10 mM DTT, 0.5 mM ATP
T4 DNA ligase	1 u
SDW	To final volume of 10 µL

2.1.4.2 Ligation-independent cloning (LIC) of HsSWEET and MFS genes

Table 2.7 lists the constructs that were cloned via ligation independent cloning (LIC). Ligation independent vectors pNIC28-Bsa4 (Savitsky et al 2010) and H6-msfGFP (Addgene plasmid # 29725) were used for cloning. H6-msfGFP and pNIC28-Bsa4 were linearised with SspI or SacI, respectively, and gel-purified as above. Gel-purified vectors and cloning PCR inserts were treated separately with T4 polymerase (Novagen, UK), as per Table 2.8 and Table 2.9. T4 polymerase was added as a final step. Vector reactions were incubated at 22 ˚C for 50 minutes. Insert reactions were incubated at 22 ˚C for 30 minutes. Following T4 treatment samples were incubated at 75 ˚C for 20 minutes before DNA extraction using the GeneJet DNA Purification kit (as above). 2 µL of T4 polymerase treated PCR product was mixed with 1 µL of T4 polymerase treated vector and incubated at room temperature for 30 minutes. This was then added to 30 µL of chemically competent Mach1 *E. coli* and transformation carried out, as in Section 2.2.3.

Table 2.7 Gene and vector components used in ligation independent T4 cloning

HsSWEET sequence information is shown in Chapter 3.2. All other genes were amplified from *E. coli* K21 MG1655 genomic DNA. All gene ids are shown in table 4.4.

Insert	Vector
HsSWEETcysless- codon optimised version of (Gene ID: 55974)	pNIC28-BSA4
ampG, amtB, araE, bcr, citT, cynX, dctA, dgoT, emrB, emrD, exuT, fsr, fucP, gltP, gntT, gudP, hcaT, hsrA, kgtP, lacY, lldP, lplT, mdfA, mdtD, mdtG, mdtH, mdtL, mdtM, mhtP, nepI, rcnA, rhtC, setA, setB, setC, shiA, sstT, tqsA, tsqA, ttdT, uhpC, xylE, xylH, yajR, ybaL, ybaT, ybbW, ybdA, yddG, ydeA, ydeE, ydfJ, ydhC, ydjE, ydjK, yebQ, yegT, yfaV, yfcJ, yhjE, yhjX, yihN, yjhB, ylcA, ynfM	H6-msfGFP

Table 2.8 Ligation independent cloning vector conditions

Component	Volume (µL)
SDW	19
Gel purified vector	50
10 x T4 polymerase buffer	10
dGTP (25 mM)	10
BSA (10 mg/mL)	1
DTT (100 mM)	5
T4 polymerase	5

Table 2.9 Ligation independent cloning gene insert treatment

Component	Volume (µL)
SDW	1.9
10 x T4 polymerase buffer	1
dCTP (25 mM)	1
BSA (10 mg/mL)	0.1
DTT (100 mM)	0.5
PCR product	5
T4 polymerase	0.5

2.1.4.3 Confirmation of gene incorporation into expression vectors

Plasmid DNA was purified from positive colonies using miniprep extraction as described in Section 2.2.4. Analytical PCR, see Section 2.1.1, with cloning primers or restriction digest were used to confirm gene incorporation. To analyse vectors by restriction digest 0.5 µg DNA was mixed with 0.5 µL restriction enzyme (Promega, UK), 1 X of appropriate restriction enzyme buffer, 0.1 µg of acetylated BSA, with the remaining volume made up to 10 µL with SDW. All digests were carried out alongside an empty control vector. The reaction was incubated at 37 ˚C for 4 hours before running on an agarose gel as above. Positive vectors were confirmed with Value Read sequencing (Eurofins, Germany) using the following promoters: H_6-msfGFP, pNIC28-BSA4 - T7 primer TAA TAC GAC TCA CTA TAG GG; pEGFPC1 - pEGFPC1for primer GAT CAC TCT CGG CAT GGA C; pEGFPN1, pDDGFP – CMVfor primer CGC AAA TGG GCG GTA GGC GTG.

2.2 General bacterial materials and methods

2.2.1 Recipes of different media used for bacterial growth

Different growth media was used for bacterial growth throughout this thesis. LB, TB and 2YT liquid media and LB agar were all purchased from Formedium, UK and made up using manufacturer's instructions. Media components are shown in Table 2.10.

Table 2.10 Composition of Formedium growth media used for cloning and protein expression with *E. coli*

Growth media	Tryptone (g/L)	Yeast extract (g/L)	Additional components
LB	10	5	NaCl 10 g/L
TB	12	24	KH_2PO_4
2YT	16	10	NaCl 5 g/L
LB agar	10	5	NaCl 10 g/L Agar 15 g/L

2.2.2 Creation of chemically competent *E. coli* cells

Two strains of *E coli* were used, BL21 (DE3) competent *E coli* NEB, UK # C25271 were used for protein expression and One Shot® Mach1 ™ T1 Phage-Resistant chemically competent *E coli* Invitrogen, UK # C8620-03 were used for cloning and vector propagation.

BL21 (DE3) or Mach1 cells were streaked on a non-selective LB agar plate and grown overnight at 37 ˚C. A single colony from the plate was picked using a sterile pipette tip and ejected into a 30 mL Universal container (ThermoFisher Scientific Ltd, UK) containing 5 mL of LB broth media, before overnight incubation at 37 ˚C and 220 rpm. 0.1 mL of this overnight culture was used to inoculate 100 mL of LB in a presterilised 500 mL conical flask, then incubated at 37 ˚C, 220 rpm until an OD660 of 0.5 for RecA- cells and 0.3-0.4 for RecA+ cells, before cells were cooled on ice for one hour. Cells were poured into two chilled 50 mL conical-bottomed centrifuge tube (ThermoFisher Scientific Ltd, UK) and harvested at 4000 g, 4 ˚C, for 15 minutes. Supernatants were discarded, and each pellet was resuspended in 20 mL of cool competent cell preparation buffer (CCMB: 80 mM CaCl2:2H2O Sigma-Aldrich, UK, 20 mM MnCl2:4H2O Sigma-Aldrich, UK, 10 mM MgCl2:6H2O Sigma-Aldrich, UK, 10 mM KOAc Sigma-Aldrich, UK, 10% (v/v) sterile glycerol). Both

samples were combined and incubated on ice for two hours. Cells were pelleted by centrifugation at 4000 g at 4 ˚C for 15 minutes. Supernatant was discarded, and the pellet resuspended in 6 mL of CCMB. Cells were aliquoted into 110 µL fractions, snap frozen and stored at -80 ˚C.

2.2.3 Transformation of expression vectors into competent *E. coli*

Chemically competent cells were thawed on ice and a 30 µL aliquoted into a 1.5 mL Eppendorf. 1 µL of DNA was pipetted into the Eppendorf and gently mixed with the cells. The mixture was then incubated on ice for 30 minutes. Cells were heat shocked at 42 ˚C for 45 seconds then placed on ice for two minutes. 250 µL of LB medium was added before incubating at 37 ˚C, 220 rpm for 45 minutes. 100 µL of the cell mixture was spread on an LB agar plate supplemented with the appropriate selection pressure and incubated at 37 ˚C for 16 hours. Table 2.11 shows the selection pressure used in these plates for each vector used, Antibiotics purchased from Sigma-Aldrich, UK.

Table 2.11 Vector selection pressures for cloning and protein expression

Antibiotics were used for expression and cloning. Additional selection pressures were only used for cloning purposes.

Vector	Concentration of antibiotic	Additional selection pressure
pNIC28-BSA4	50 µg/ml kanamycin	5 % (w/v) sucrose
H6-msfGFP	50 µg/ml kanamycin	
pET27b	50 µg/ml kanamycin	
pEGFP-N1	50 µg/ml kanamycin	
pEGFP-C1	50 µg/ml kanamycin	
pDD-GFP	100 µg/ml ampicillin	

2.2.4 Extraction of plasmid DNA from *E. coli*

Plasmids were purified from cell cultures with Gene-JET Plasmid Miniprep Kit (ThermoFisher Scientific Ltd, UK). Extraction was undertaken according to manufacturer's instructions. All centrifuge steps were performed using a desktop centrifuge at 12 000 g. DNA was quantified using a Nanodrop 2000c UV-Vis spectrophotometer.

2.3 General protein expression

2.3.1 Expression of cloned genes using T7 polymerase regulated vectors

Starter cultures were inoculated using a sterile pipette tip to harvest transformed colonies from an LB agar plate into 5 mL LB medium containing appropriate antibiotic (Table 2.10), before incubation at 37 ˚C, 220 rpm for 16 hours. The starter culture was then used to inoculate growth media (for details, see Section 2.3.7, 2.5.5 and 2.6.5/2.6.5) containing appropriate antibiotic to a starting OD_{600} of 0.005, incubated at 37 ˚C, 220 rpm and OD_{600} measurements monitored. All OD_{600} measurements were taken with a Nanodrop 2000c UV-Vis spectrophotometer. Upon reaching an OD_{600} of 0.4-5 the incubator temperature was set to expression temperature. Cells were grown until an expression OD_{600} of 0.8-1 was reached then induced with IPTG at the desired concentration (Sigma-Aldrich, UK). Exact expression conditions can be found in the appropriate methods section for each chapter: GFP in 2.3.7, HsSWEET in 2.5.5 and MFS in 2.6.4/2.6.5.

2.3.2 Affinity chromatography purification of bacterially expressed protein

Nickel resin volumes, Wash and Elution buffers for purification can be found in appropriate methods section; GFP in 2.3.7, HsSWEET in 2.5.5 and MFS in 2.6.4/2.6.5

Following expression, cells were harvested by centrifugation at 4000 g, 4 ˚C for 20 minutes. From this step onwards, cell contents were maintained at 4 ˚C. Cell pellets were then resuspended in lysis buffer (0.1 M Tris, pH 8; 0.15 M NaCl, Melford Laboratories Ltd, UK) using a tissue grinder. Cells were lysed by one of two methods. Sonication using the Misonix Sonicator Ultrasonic Processor XL2020 (ThermoFisher Scientific, UK). Sonication on setting 7 was performed for a total of 10 minutes, in 10 second pulses with 30 second rest. During sonication, samples were kept on ice with gentle mixing after the first 5 minutes. Alternatively, cells were lysed using a Cell Disruptor (Constant Systems, UK) with three rounds at 35 thousand pounds per square inch.

For both methods, unbroken cells and cell debris were removed by centrifugation at 8000 g for 10 minutes. The cell pellet was discarded and the supernatant ultracentrifuged at 100 000 g for 60 minutes, pelleting the insoluble cellular fraction. The supernatant was then discarded and the insoluble fraction suspended

in solubilisation buffer (unless otherwise stated: 0.1 M Tris, pH 8; 0.15 M NaCl; 40 mM Brij 35, ACROS Organics, UK) using a tissue homogeniser. The insoluble fraction was solubilised on a tube rotator at 10 rpm for four hours, unless otherwise stated. Insoluble material was separated by ultracentrifugation at 100 000 g for 60 minutes. Soluble material was decanted into a 50 mL conical-bottomed centrifuge tubes (one tube was used per 25 mL of supernatant) and mixed with His60 Ni Resin (Clontech, France) suspended in wash 1 buffer. The protein resin mixture was combined on a tube rotator set at 10 rpm. The mix resin was transferred to an open gravity chromatography column and the flow through (FT) collected. The resin was then washed and eluted with the following volumes unless otherwise stated, all steps were collected. 10 times the volume of nickel resin with Wash 1, 2 and 3. 2.5 times the volume of nickel resin with Elution buffer.

2.3.3 SDS-PAGE electrophoresis analysis of bacterially expressed protein

Collected samples were visualised using SDS-PAGE, stained with Coomassie Blue stain, made as below. SDS-PAGE was formed of a resolving gel (12 % Acrylamide/bisacrylamide ThermoFisher Scientific Ltd, UK, 1 M Tris pH 8.4, 0.06 % APS Sigma-Aldrich, UK, 0.13 % TEMED ThermoFisher Scientific Ltd, UK and 0.1 % SDS Sigma-Aldrich, UK) and a stacking gel (5 % Acrylamide/bisacrylamide, 0.72 M Tris pH 8.8, 0.025 % APS, 0.4 % TEMED and 0.1 % SDS). Samples were loaded into stacking gel wells with sample buffer (2 % SDS, 2 mM DTT Sigma-Aldrich, UK, 4 % glycerol, 0.04 M Tris pH 6.8, 0.01 % bromophenol blue) and run alongside molecular weight marker (EZ run protein marker ThermoFisher Scientific Ltd, UK # BP3600 or prestained protein molecular weight marker ThermoFisher Scientific Ltd, UK # 26612) at 120 V. Gels were soaked in fixing solution (40 % ethanol, ThermoFisher Scientific Ltd, UK and 10 % acetic acid, ThermoFisher Scientific Ltd, UK) for 10 minutes on an oscillating platform. Gels were then stained with Coomassie stain for 1 hour on an oscillating platform. Stained gels were imaged with a Gel Doc™ XR+ (Bio-Rad, UK) or Nikon D50 digital camera.

Coomassie Blue stain was made by dissolving 70 mg of Coomassive Blue G-250 (Sigma-Aldrich, UK) in 1.5 mL of HCl (ThermoFisher Scientific Ltd, UK) then diluted to 5 L with distilled water and filtered.

2.3.4 Concentration of expressed protein

Protein concentration was carried out using Vivaspin columns (Generon, UK) with appropriate molecular weight cut off. Concentration was undertaken according to

manufacturer's instructions. All centrifuge steps were performed using a Sorvall legend RT large bench centrifuge (ThermoFisher Scientific, UK) at 4000 g, 4 ˚C. Following concentration protein was quantified using a Nanodrop 2000c UV-Vis spectrophotometer.

2.3.5 Gel filtration analysis of bacterially expressed protein

Purified protein elution samples were analysed with gel filtration using a Superdex 5-150 column (GE Lifesciences, UK). Elution fractions from nickel affinity chromatography were pooled and concentrated to a final volume of 500 µL. Ethanol in the column was replaced by degassed column buffer; unless otherwise stated, column buffer was protein Elution buffer containing no imidazole.

The column was calibrated at 0.2 mL/min, with the void volume marker Blue dextran (Sigma-Aldrich, UK), 66.45 kDa albumin from bovine serum (Sigma-Aldrich, UK), 44.287 kDa albumin from chicken serum (Sigma-Aldrich, UK) and 20.1 kDa trypsin inhibitor from G. max (Sigma-Aldrich, UK). Pooled elution fractions were then run through the column and 500 µL fractions collected. The protein size (kDa) of gel filtered samples were calculated using a standard curve.

2.3.6 Western blot immunoanalysis of bacterially expressed protein

SDS-PAGE gels were run as above. Once complete, protein components were transferred to a Polyvinylidene fluoride (PVDF) membrane (Sigma-Aldrich,UK). The transfer was run at 30 V and 4 ˚C for 16 hours in transfer buffer (25 mM Tris pH 8, 190 mM glycine, Sigma-Aldrich, UK and 20 % methanol, ThermoFisher Scientific Ltd, UK). After transfer, the PVDF membrane was incubated for 1 hour in 20 mL of blocking buffer (1X PBS Sigma-Aldrich, UK (NaCl 138 mM; KCl - 27 mM; pH 7.4), 0.05 % TWEEN-20, Sigma-Aldrich, UK and 5 % BSA Sigma-Aldrich, UK). The blocking solution was then removed and the membrane incubated in 20 mL of WB-wash solution (PBS and 0.05 % TWEEN-20) for 10 minutes, 3 times. After washing the PVDF membrane was incubated for 1 hour in 10 mL of primary antibody solution (PBS, 0.05 % TWEEN-20 and 3.33 µL of 6x-His Tag Monoclonal Antibody ThermoFisher Scientific Ltd, UK # MA1-21315). The primary solution was then removed and the membrane was incubated in 20 mL of WB-wash solution for 10 minutes, 3 times. From this point onwards, the membranes exposure to UV light was minimised. After washing, the PVDF membrane was incubated for 1 hour in 10 mL of secondary antibody solution (PBS, 0.05 % TWEEN-20 and 1 µL of IRDye® 680LT Infrared Dye LI-COR, UK). The secondary antibody solution was then

removed and the membrane was incubated in 20 mL of WB-wash solution for 10 minutes 3 times. An Odyssey (LI-COR Biosciences, UK) was used for imaging of western blots.

2.3.7 GFP quantification using protein fluorescence

GFP measurements were taken with a RF-5301PC (Shimadzu, UK) spectrofluorometer. A wavelength of 488 nm was used for excitation and 500-530 nm was collected for emission output. A calibration curve was constructed using purified H6-sfGFP. H6-sfGFP was expressed in 500 mL of LB broth containing 50 µg/mL of kanamycin, using the empty H6-msfGFP vector. With no insertion, this vector has sfGFP in frame with the start codon and an N-terminal hexahistidine tag. Expression was induced with 0.2 mM IPTG at 18 ˚C for 16 hours. Protein purification was edited to reflect soluble protein. After the first 100 000 g membrane isolation, the soluble fraction was added to 1 mL of 0.5 g/mL nickel resin. The purification Wash and Elution buffers are shown in Table 2.12.

Table 2.12 Composition of Wash buffers and Elution buffers for affinity chromatography purification of *E. coli* expressed GFP

Buffer	**Tris pH8 (M)**	**NaCl (M)**	**Imidazole (M) (Sigma-Aldrich, UK)**	**Brij 35 (mM)**
Wash 1	0.1	0.15	0	10
Wash 2	0.1	1	0	5
Wash 3	0.1	0.15	0.02	5
Elution buffer	0.1	0.15	0.2	5

Purified GFP was diluted with gel filtration buffer to between 0.3975 and 159 ng/mL, to reflect the measurement capabilities of the spectrofluorometer. The generated calibration curve produced the equation: GFP concentration = spectrofluorometer RFU/403192. This equation was used to calculate protein concentrations during expression tests.

2.4 Expression, purification and characterisation of HsSWEET

2.4.1 Tertiary structural predictions of HsSWEET using amino acid sequence

The sequence of sugar transporter SWEET1 isoform a [Homo sapiens] (NCBI reference number NP_061333.2) was used for predictions. The TOPCONS web server was used to predict the topology of HsSWEET (Tsirigos et al 2015). Structural models were made using the SWISS-MODEL homology-modelling server, accessed via the ExPASy web server (Biasiniet al 2014). Clustal Omega (Sievers et al 2011) was used for alignment of HsSWEET with structurally characterised OsSWEET. Information from the alignment was used to manually annotate the OsSWEET PDB model # 5cwtg using the molecular visualisation software Chimera (Pettersen et al 2004).

2.4.2 Cloning of HsSWEET and PQ-loop proteins into vectors for subcellular localisation and protein expression

Table 2.13 lists the SWEET and PQ-loop constructs that were generated via cloning PCR throughout this chapter, including the names of the primers used for amplification; primer names match sequences listed in Appendix Table 7.3. LIC was used to construct the pNIC-HsSWEETCysless vector. All other vectors were constructed using ligation-dependent T4 cloning. Sequential restriction digests were carried out to prepare PCR inserts and vectors for ligation; restriction enzymes used are listed in Table 2.

Table 2.13 Vector constructs for expression and subcellular localisation of HsSWEET and PQ loop genes

Forward primers are labelled For. and reverse primers are labelled Rev. Primer sequence information can be found in Table 7.3. UniPROTKB HsSWEET1-Q9BRV3, PQLC1-Q8N2U9, PQLC2- Q6ZP29, PQLC3- Q8N755, CTNSCysLess2-O6093.

Target	Construct name	For. primer	Rev. primer	DNA template	Target vector	Construct + tags
HsSWEET cysless	pNIC-HsSWEETCysless	HsSWEETCysless_f001	HsSWEETCysless_r001	HsSWEETCysless	pNIC28-BSA4	START - Hexahis - TEV - target protein- STOP
HsSWEET cysless	pEGFP-N1-HsSWEET	HsSWEETCysLess-f002	HsSWEETCysLess-r003	HsSWEETCysless	pEGFP-N1	HindIII – START- HsSWEETCysLess – CG – BamHI – GFP - STOP
HsSWEET cysless	pDDGFP-HsSWEETCysless c004	HsSWEETCysLess-f003	HsSWEETCysLess-r002	HsSWEETCysless	pDDGFP	HindIII – START – HA – HsSWEETCysLess – STOP – BamHI
HsSWEET cysless	pDDGFP-HsSWEETCysless c005	HsSWEETCysLess-f002	HsSWEETCysLess-r004	HsSWEETCysless	pDDGFP	HindIII – START - HsSWEETCysLess – HA – STOP – BamHI
HsSWEET cysless	pEGFP-C1-HsSWEETCysless	HsSWEETCysLess-f004	HsSWEETCysLess-r002	HsSWEETCysless	pEGFP-C1	START – GFP – HindIII – HsSWEETCysLess – STOP – BamHI
GFP	pET27b-GFP	SalI-GFP-f001	GFP-HindIII-r001	H6-msfGFP	pET27b	START - pelB signal sequence – MCS – SalI – GFP – STOP – HindIII

Target	Construct name	For. primer	Rev. primer	DNA template	Target vector	Construct + tags
HsSWEET	SLC50A1-GFP	EcoRI-SLC50A1-f001	SLC50A1-SalI-r001	SLC50A1	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – SLC50A1 – SalI – GFP – STOP – HindIII
HsSWEET cysless	HsSWEETCysLess-GFP	EcoRI-hSWTCysLess-f001	hSWTCysLess-SalI-r001	HsSWEETCYSLESS	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – HsSWEETCysLess – SalI – GFP – STOP – HindIII
PQLC1	PQLC1-GFP	EcoRI-PQLC1-f001	PQLC1-SalI-r001	PQLC1	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – PQLC1– SalI – GFP – STOP – HindIII
PQLC2	PQLC2-GFP	EcoRI-mPQLC2-f001	mPQLC2-SalI-r001	PQLC2	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – mPQLC2– SalI – GFP – STOP – HindIII
PQLC3	PQLC3-GFP	EcoRI-PQLC3-f001	PQLC3-SalI-r001	PQLC3	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – PQLC32 – SalI – GFP – STOP – HindIII
CTNSCysLess1	CTNSCysLess1-GFP	EcoRI-CTNS-f001	CTNS-SalI-r001	CTNSCysless	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – deltaSS-CTNSCysless– SalI – GFP – STOP – HindIII

Target	Construct name	For. primer	Rev. primer	DNA template	Target vector	Construct + tags
CTNSCysLess2	CTNSCysLess2-GFP	EcoRI-CTNS-f002	CTNS-SalI-r001	CTNSCysless	pET27b-GFP	START - pelB signal sequence – linker - EcoRI – deltaSS-deltaNterm-CTNSCysless– SalI – GFP – STOP – HindIII
CTNSCysLess3	CTNSCysLess3-GFP	NcoI-CTNS-f001	CTNS-NheI-r001	CTNSCysless	pET27b-GFP	START - pelB signal sequence – linker - NcoI – Nterm-CTNSCysless– NheI – linker – Hexahis tag – STOP

2.4.3 Subcellular localisation of HsSWEET in human HEK and RPE cell lines

Transfection and confocal microscopy was kindly carried out by Dr David Tumbarello (University of Southampton). pDD-GFP, pEGFP-N1 and pEGFP-C1 vectors were transfected into propagated HEK293 and RPE cell lines for study.

Cell culture:

Hela cells – cultured in RPMI +Glutamax™ medium supplemented with 10% FBS (Fetal Bovine Serum) and 1% Pen/Strep

RPE cells – cultured in DMEM/F12 + Glutamine, 10% FBS and 1% Pen/Strep

Transfection and immunostaining:

Cells were seeded onto glass coverslips in 6-well dishes in complete media 24 hours prior to transfection. Cells were transfected with 1 μg of plasmid DNA using 3 μl Fugene6 (Promega) according to manufacturer's protocols. Cells were fixed 24 hours post-transfection in 4% Formaldehyde in PBS for 20 minutes, washed in PBS, and permeabilised in 0.02% Triton X-100 in PBS for 2 minutes. Fixed cells were blocked in 1% BSA in PBS for 15 minutes prior to incubation with appropriate primary antibodies for 2 hours at room temperature. Cells were then incubated for 45 minutes at room temperature with Alexa fluor 568 or 488 goat anti-rabbit or anti-mouse secondary antibodies. Following immunostaining cells were incubated in 0.005 mg/mL hoescht solution for 5 minutes, then rinsed with PBS solution. Immunostained coverslips were washed in PBS, dH20, and mounted with FluorSave reagent (Merck) on glass.

Microscopy:

Imaging was performed using a 100x oil immersion objective on a Zeiss Axioplan upright microscope and images taken using Metamorph software. Images were assembled with Adobe photoshop. Visualisation conditions are shown in Table 2.14.

Table 2.14 Confocal microscopy fluorescent antibodies and organelle stains for subcellular localisation of HsSWEET

Target	Visualised	Method
pDD-GFP	Anti-HA – clone HA.11, mouse monoclonal (BioLegend, UK)	Alexa Fluor™ goat anti-mouse 568 antibodies (Invitrogen, UK), excitation at 578 nm emission at 603 nm
pEGFP-N1 and pEGFP-C1	Anti-GFP – rabbit polyclonal (Invitrogen, UK)	Alexa Fluor™ goat anti-rabbit 488 (Invitrogen, UK), excitation at 495 nm emission at 518 nm
Lysosome	Anti-Lamp1 (CD107a) (BD Bioscience, UK)	Alexa Fluor™ goat anti-mouse 568 antibodies (Invitrogen, UK), excitation at 578 nm emission at 603 nm
Actin	Alexa Fluor™ 488-phalloidin (Invitrogen, UK)	Alexa Fluor™ 488, excitation at 578 nm emission at 603 nm
Nucleus	Hoechst Binding	Excitation at 350 nm and emission 461 nm

2.4.4 Predictions of HsSWEET subcellular localisation using primary sequence

Webservers for LocSigB (Negi et al 2015), PSORTII (Horton et al 1997) were used to predict HsSWEET localisation based on sequence.

2.4.5 Comparison of GFP-tagged HsSWEET expression at different temperatures

Expression temperature tests were carried out using 500 mL of LB broth with 50 µg/mL kanamycin. Expression was induced by 0.2 mM IPTG and incubated at a range of temperatures (15, 18, 21, 24, 27 ˚C), 200 rpm, for 16 hours. For samples at 15 ˚C, 24 ˚C and 27 ˚C, partial protein purification was carried out up to the initial collection of membrane fractions, after the first 100 000 g ultracentrifugation step of section 2.3.2. For samples at 18 ˚C and 21 ˚C, full protein purification was carried out (as per section 2.3.2) using solubilisation, Wash and Elution buffers, shown in Table 2.15. 1 mL of 0.5 g/mL nickel resin was used per 500 mL of LB broth.

Table 2.15 Buffer constituents for affinity chromatography purification of his-tagged HsSWEET using nickel resin

N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate (Z-3-12) purchased from (Sigma-Aldrich, UK).

Buffer	Tris pH8 (M)	NaCl (M)	Imidazole (M)	Z-3-12 (mM)
Solubilisation buffer	0.1	0.15	0	40
Wash 1	0.1	0.15	0	10
Wash 2	0.1	1	0	5
Wash 3	0.1	0.15	0.02	5
Elution buffer	0.1	0.15	0.2	5

2.4.6 Batch purification of HsSWEET with different detergents to identify solubilisation efficiency

Purification was set up to test the solubilisation efficiency of the following detergents: 3-(N,N-dimethylmyristylammonio) propanesulfonate (Z-3-14; Sigma-Aldrich,UK), Brij 010 (Sigma-Aldrich,UK), Brij 35 (Sigma-Aldrich,UK) and Brij 93 (Sigma-Aldrich,UK). Purification buffers are shown in Table 2.16.

Batch purification was carried out using 500 mL of LB broth with 50 µg/mL kanamycin. Buffers for this process are listed in Table 2.16. Expression was induced by 0.2 mM IPTG and incubated at 21˚C, 220 rpm for 16 hours. The membrane fraction was collected by ultracentrifugation at 100 000 g, 4 ºC for 60 minutes and divided into four. Each fraction was suspended in solubilisation buffer. 0.25 mL of 0.5 g/mL nickel resin was added to each membrane fraction. The combined protein-resin was mixed on a tube rotator at 10 rpm for 1 hour. The resin was then collected by centrifugation at 4000g, 4 ºC for 10 minutes and placed into a 1.5 mL microcentrifuge tube. 1 mL of Wash 1 was added to each microcentrifuge tube, followed by 10 minutes of mixing on a tube rotator. The resin was pooled by centrifugation and the Wash removed with a pipette. 0.1 mL of Elution buffer was added, the tubes were agitated and the elution pipetted off. The purification efficiency was compared by SDS-PAGE (as per section 2.3.3).

Table 2.16 Buffer constituents for batch affinity chromatography of his-tagged HsSWEET with nickel resin

Detergents used were: Z-3-14, Brij 010, Brij 35 and Brij93.

Buffer	Tris pH 8 (M)	NaCl (M)	Imidazole (M)	Detergent (mM)
Solubilisation buffer	0.1	0.15	0	40
Wash 1	0.1	0.15	0	10
Elution buffer	0.1	0.15	0.2	5

2.4.7 Using GFP-tagged HsSWEET to compare the yield of expression conditions growth comparison

Expression tests were carried out in 500 mL of broth containing 50 µg/mL kanamycin. Test expressions in LB and TB were compared at 18, 23 and 30 ˚C. Expression was induced with 0.2 mM IPTG for 16 hours at 220 rpm. Expressing cells were harvested by centrifugation at 4000 g, 4 ˚C for 20 minutes. The pellet was resuspended in 50 mL of lysis buffer (as per section 2.3.2) before calculating the GFP concentration (as per section 2.3.7). This concentration corresponded to the amount of expressed HsSWEET.

2.4.8 Gel filtration of HsSWEET

Following purification HsSWEET was gel filtered to purify to homogeneity. The proteins used to generate a standard curve are shown in Table 2.17. Gel filtration buffer was 100 mM Tris pH8, 150 mM NaCl and 40 mM Brij35. The log molecular weight of the protein standards was plotted against the volume of the peak position to create a standard curve. The standard curve equation was “Log molecular weight = -0.1494 * peak position + 2.6035”.

Table 2.17 Details of protein gel filtration to make a standard curve of HsSWEET purification

The Void volume was set using blue dextran. The void volume peak position was set at zero to formulate the standard curve.

Standard	Mwt kDa	log Mwt	Peak position	- Void volume (mL)
Blue dextran	Void volume	n/a	8.425	0
Bovine serum albumin	66.45	1.822495	13.415	4.99
Chicken serum albumin	44.287	1.646276	15.425	7
Trypsin inhibitor	20.1	1.303196	16.775	8.35

2.4.9 T_m thermostability test of HsSWEET titrated with glucose

Protein from section 2.4.6 was used for thermostability tests. Prior to the melting temperature (T_m) titration test, HsSWEET was partially purified as per section 2.5.6. Wash and Elution buffers were used as Table 2.15, but with all detergents replaced with Brij 35. The Solubilisation buffer used was 0.1 M Tris pH 8, 0.15 M NaCl and 40 mM Brij 35. After the insoluble material was pelleted via centrifugation, the purification was ended and solubilised material was used for the T_m titration. Glucose (Sigma-Aldrich, UK) dissolved in solubilisation buffer was added to HsSWEET at concentrations ranging from 0 to 100mM. The samples were divided in two and either maintained at 4 ˚C or incubated at 62.5 ˚C for 3 minutes, using a DNA engine TETRAD 2. Samples were then centrifuged at 100 000 g for 20 minutes to remove aggregated protein. GFP concentration (corresponding to SWEET concentration) in the samples were then measured using the spectrofluorimeter (as per section 2.3.7). GFP concentration was plotted on a bar chart to identify differences in protein between the 4 and 62.5 ˚C at different glucose concentrations.

2.5 Expression, purification and characterisation of MFS proteins

2.5.1 Cloning of MFS proteins into the *E. coli* expression vector H6-msfGFP

Constructs for the expression of *E. coli* MFS proteins were created via cloning PCR 2.1.1, using primers from Appendix Table 7.3. Forward primers were flanked with the sequence TACTTCCAATCCAATGCA and reverse primers were flanked with the sequence CTCCCACTACCAATGCC. These introduced ligation independent cloning sites to the PCR product.

E. coli K-12 MG1655 genomic DNA was extracted from cell lines using a Genomic DNA Purification Kit (ThermoFisher Scientific Ltd, UK). Extraction was undertaken according to manufacturer's instructions then diluted to 10 ng/µL in SDW, 10 ng was used per PCR reaction. PCR products were gel purified and used directly in ligation independent cloning with the vector H6-msfGFP. Vectors were sequenced using the T7 forward primer TAATACGACTCACTATAGGG.

2.5.2 Comparing expression yield of MFS proteins using different temperature and growth media

Expression tests were carried out in 50 mL of LB, TB or 2YT medium. Composition of these media are listed in section 2.2.1. Expression was induced with 0.2 mM IPTG or 0 mM IPTG (for control expressions) and incubated at 18 or 30 ºC for 16 hours. Following expression, OD_{600} was measured using a Nanodrop 2000c UV-Vis spectrophotometer and the cells were centrifuged at 4000 g for 20 minutes. Supernatant was discarded and the cells re-suspended in lysis buffer (see Section 2.3.2) to a total volume 5 mL. The GFP concentration was measured using spectrofluorimeter as per section 2.3.7. This concentration corresponded to the amount of expressed GFP-tagged MFS protein.

2.5.3 T_m (melting temperature) thermostability testing of MFS protein aggregation at different temperatures in the presence of substrates

The targets and corresponding substrates tested in the thermostability tests are listed in Table 2.18. All sugars, cAMP, kanamycin, EtBr and cysteine were dissolved in substrate buffer, as detailed in the following sections; Table 2.18 refers to specific sections. Ethanol was used to dissolve chloramphenicol to 250 mM and

TPP^+ to 500 mM; norfloxacin was dissolved in 100 mM HCl to a concentration of 100 mM. These stocks were further diluted with substrate buffer.

Table 2.18 Cloned target and substrates investigated in the T_m thermostability tests

Sections 2.5.3.1-3 gives further detail of the experimental method, including substrate buffers, as listed.

Target	Substrates tested [concentration range]	For details, see section
GFP		2.6.3.1
SetB-GFP	Lactose [0 - 5 mM]	2.6.3.1
LacY-GFP	Lactose [0 - 5 mM]	2.6.3.1
XylE-GFP	Sucrose [0 - 5 mM]	2.6.3.1
MdfA-GFP	Chloramphenicol [0 - 5 mM]	2.6.3.1
AraE-GFP	Arabinose [0 - 5 mM]	2.6.3.1
MdtM-GFP	Chloramphenicol [0 - 5 mM]	2.6.3.1
YdeA-GFP	cAMP [0 - 5 mM]	2.6.3.1
Detailed in Section 2.6.3.2		
GFP	Chloramphenicol [0 – 1 mM]	2.6.3.2
MdfA-GFP	Chloramphenicol [0 – 8.6 mM], Kanamycin [0 – 1 mM], TPP^+ [0 – 1 mM], EtBr [0 – 1 mM], norfloxacin [0 – 1 mM]	2.6.3.2
YdeA-GFP	Chloramphenicol [0 – 20 mM]	2.6.3.2
MdtM-GFP	Chloramphenicol [0 – 5 mM]	2.6.3.2
AraE-GFP	Arabinose [0 – 15 mM]	2.6.3.2
Detailed in Section 2.6.3.3		
MdfA-GFP	Chloramphenicol [0 – 1 mM]	2.6.3.3
HsrA-GFP	Cysteine [0 – 1 mM]	2.6.3.3
MdtM-GFP	Chloramphenicol [0 – 1 mM]	2.6.3.3
YdeA-GFP	cAMP [0 – 1 mM]	2.6.3.3

2.5.3.1 Thermostability tests of partially purified protein

For initial thermostability tests, partial purification of MFS proteins was carried out as per section 2.3.2. Protein was prepared in solubilisation buffer (0.1 M Tris pH 8, 0.15 M NaCl and 40 mM Brij 35). After insoluble material was centrifuged out of solution, the solubilised material was used for thermostability tests. Substrates, details shown in Table 2.18, were dissolved in substrate buffer (0.1 M Tris pH 8, 0.15 M NaCl and 40 mM Brij 35) unless otherwise stated and were added to solubilised MFS proteins, including a control with no substrate. Substrates and samples were incubated for 12 hours at 4 ˚C. The samples were pipetted into separate PCR tubes, one for each temperature in the range, and exposed to the desired temperature for 3 minutes, using a DNA engine Tetrad 2. The thermoprofile temperature range for these experiments was 4 ˚C, 20 ˚C, 38˚C, 45 ˚C, 50˚C, 55˚C, 60 ˚C, 66 ˚C, 72 ˚C, 78˚C, 84 ˚C and 90 ˚C. Samples were returned to 18 ºC for 3 minutes before storage at 4 ºC. Samples were then centrifuged at 100 000 g for 20 minutes to remove precipitated protein. GFP fluorescence (corresponding to MFS concentration) was measured using the spectrofluorimeter (as per section 2.3.7).

All data points were carried out in triplicate and an average RFU was plotted against temperature. The maximum RFU (RFU_{max}) was calculated as an average fluorescence of the lowest three temperatures, 4 ˚C, 20 ˚C and 38˚C. The RFU_{max} was divided by 2 to generate the fluorescence value of $^{1/2}RFU_{max.}$ The temperature at which the plotted thermoprofiles fluorescence reached the $^{1/2}RFU_{max}$ was calculated as the T_m. T_m values were plotted against substrate concentration to investigate the effect of substrate on protein precipitation.

2.5.3.2 Thermostability test using affinity chromatography purified MFS protein

The expression tests and determined temperature listed in Chapter 4.2 were used for purification. As previously, expression was carried out as in Section 2.3.1. Post expression, purification was carried out as per section 2.3.2; all purification buffers are shown in Table 2.19. 1 mL of 0.5 g/mL nickel resin was used per 500 mL of LB broth. Purified protein was dialysed in 0.1 M Tris pH 8, 0.15 M NaCl and 5 mM Brij 35, before use in thermostability tests, as described in section 2.5.3.1. Substrate buffers are shown in Table 2.20.

Table 2.19 Composition of Wash buffers and Elution buffers for affinity chromatography purification of MFS-GFP proteins

Buffer	Tris pH8 (M)	NaCl (M)	Imidazole (M)	Brij 35 (mM)
Solubilisation buffer	0.1	0.15	0	40
Wash 1	0.1	0.15	0	10
Wash 2	0.1	1	0	5
Wash 3	0.1	0.15	0.02	5
Elution buffer	0.1	0.15	0.2	5

Table 2.20 Composition of substrate solubilisation buffers for T_m thermostability test of affinity purified MFS-GFP proteins

Substrate Buffer	Tris pH8 (M)	NaCl (M)	Brij 35 (mM)	Solvent
kanamycin, EtBr and arabinose	0.1	0.15	5	n/a
Chloramphenicol	0.1	0.15	5	3.3 % EtOH
TPP^+	0.1	0.15	5	0.138 % EtOH
Norfloxacin	0.1	0.15	5	1 mM HCl

2.5.3.3 Thermostability T_m hold test using affinity chromatography purified MFS protein

The expression tests and determined temperature listed in Chapter 4.2 were used for purification. As previously, expression was carried out as in Section 2.3.1. Post expression, purification was carried out as per section 2.3.2; all purification buffers are shown in Table 2.21. 1 mL of 0.5 g/mL nickel resin was used per 500 mL of LB broth. Purified protein was dialysed in 0.1 M Tris pH 8, 0.15 M NaCl and 5 mM Brij 35, before use in thermostability hold tests. The substrates, listed in Table 2.18, were dissolved in substrate buffer shown in Table 2.20 and Table 2.21 and added

to the purified MFS-sfGFP transporter. The samples were divided into two, with half maintained at 4 ˚C and half incubated at the protein's T_m for 15 minutes, using a DNA engine Tetrad 2. Samples were then centrifuged at 100 000 g for 20 minutes to remove aggregated protein. Fluorescence (corresponding to MFS concentration) in the samples was then measured using the spectrofluorimeter (as per section 2.3.7) or the POLARstar Omega spectrofluorometer plate reader (BMG LABTECH, UK). POLARstar omega spectrofluorimeter settings: temperature 20 ºC; excitation filter = 485-12; emission filter = 520-530. Fluorescence was plotted on a line graph to identify differences in protein aggregation and the T_m temperature at different substrate concentrations.

Table 2.21 Composition of substrate solubilisation buffers for T_m thermostability test of affinity purified MFS-GFP proteins

Substrate buffer	Tris pH8 (M)	NaCl (M)	Brij 35 (mM)	Solvent
Cysteine and cAMP	0.1	0.15	5	n/a
Chloramphenicol	0.1	0.15	5	0.384 % EtOH

2.5.4 Using thermophoresis to test substrate binding to GFP-tagged MFS protein

Initial thermophoresis experiments were carried out using the Monolith NT.115 (NanoTemper, UK). For monolith testing, protein purification as described in section 2.3.2 was modified for three MFS protein targets: hsrA-sfGFP, mdfA-sfGFP 1 and mdtM-sfGFP. Second round of testing was carried out using a Monolith NT.115 in the laboratory of David Staunton (Oxford University); MFS proteins kgtP-sfGFP and mdfA-sfGFP 2. Buffers for solubilisation, purification and dialysis are shown in Table 2.22. Each protein was expressed in 500 mL of LB broth and induced with 0.2 mM IPTG at 18 ºC, 220 rpm for 16 hours. 2 mL of 0.5 g/mL nickel resin was bound to solubilised protein for 5 hours. Following purification elution, the purified protein was dialysed for 4 hours; the dialysis buffer was then replaced and a further dialysis carried out for 8 hours. Additionally, a second mdfA-sfGFP sample was purified as per Section 2.5.3.2. Following dialysis proteins binding was analysed using the Monolith NT.115, as per the manufacturer's instructions. The

following settings were used: Nano BLUE detector selected for GFP analysis; substrate concentration series 1 in 2. The substrate concentration ranges used are shown in Table 2.23. Monolith NT.115 software was used to plot protein kinetics against substrate concentration and calculate K_d.

Table 2.22 Composition of Wash buffers and Elution buffers for affinity chromatography purification of MFS-GFP proteins for monolith thermophoresis

Additionally, all mdfA and kgtP buffers contain 10 % glycerol and 5 mM β-mercaptoethanol (β-mer; Sigma-Aldrich, UK). After purification 0.002 % (w/v) cholesteryl hemisuccinate (Sigma-Aldrich, UK) was added to elutant. Additional reagents for these purifications include; n-decyl-β-D-maltopyranoside (DM; Anatrace, UK), n-dodecyl-β-D-maltopyranoside (DDM; Anatrace, UK), lauryl maltose neopentyl glycol (LMNG; Anatrace, UK), choline chloride (chocl; Sigma-Aldrich, UK), Na_2SO_4 (Sigma-Aldrich, UK).

HsrA						
Buffer	**HEPES pH 7.2 (M)**	**NaCl (M)**	**Imidazole (M)**	**DDM (mM)**	**Volume (mL)**	
Solubilisation buffer	0.1 (Tris pH 8)	0.15	0	5	25	
Wash 1	0.02	0.15	0	2	30	
Elution buffer	0.02	0.15	0.25	2	10	
Dialysis Buffer	0.02	0.15	0	2	n/a	
MdfA 1						
Buffer	**Tris pH8 (M)**	**NaCl (M)**	**Imidazole (M)**	**DM (%)**	**LMNG (%)**	**Volume (mL)**
Solubilisation buffer	0.02	0.3	0	0.5		25
Wash 1	0.02	0.3	0.02	0.2	0.01	30
Elution buffer	0.02	0.3	0.35	0.2	0.01	10
Dialysis buffer	0.04 (HEPES pH 7.2)	0.1 (chocl)	0	0.2	0.01	n/a
MdtM						
Buffer	**Tris pH8 (M)**	**NaCl (M)**	**Imidazole (M)**	**DDM (%)**	**Glycerol (%)**	**volume**

Solubilisati-on buffer	0.02	0.5	0.01	1.2	10	25
Wash 1	0.05	0.1	0.025	0.1	20	40
Elution buffer	0.05	0.1	0.5	0.1	20	5
Dialysis buffer	0.05	0.2 (Na_2SO_4)		0.05		n/a
MdfA 2						
Buffer	**Tris pH8 (M)**	**NaCl (M)**	**BRIJ-35 (M)**			**Volume**
Solubilisati on buffer	0.02	0.3	0.04			25
	HEPE pH7.2 (M)	**chocl (M)**	**Imidazole (M)**	**DDM (mM)**	**LMN-G (%)**	
Wash 1	0.04	0.1	0.02	2	0.01	30
Elution buffer	0.04	0.1	0.35	2	0.01	5
Dialysis buffer	0.04	0.1		2	0.01	n/a
KgtP						
Buffer	**Tris pH8 (M)**	**NaCl (M)**	**BRIJ-35 (M)**			**Volume**
Solubilisati-on buffer	0.02	0.3	0.04			25
	HEPE pH7.2 (M)	**NaCl (M)**	**Imidazole (M)**	**DDM (mM)**	**LMN G (%)**	
Wash 1	0.04	0.1	0.02	2	0.01	30
Elution buffer	0.04	0.1	0.35	2	0.01	5
Dialysis buffer	0.04	0.1		2	0.01	n/a

Table 2.23 MFS-sfGFP substrate binding test conditions with thermophoresis

Affinity purified sfGFP-tagged MFS proteins were tested using the Monolith NT.115.

GFP-tagged protein	Substrate	Concentration range
MdfA and mdfA 1	Chloramphenicol	0.0305 – 1000 µM
MdfA 1	TPP^+	0.0305 – 1000 nM
MdtM	Chloramphenicol	0.0305 – 1000 nM
HsrA	Cysteine	0.0305 – 1000 µM
MdfA	Chloramphenicol	0.09 – 3000 µM
MdfA	cAMP	0.0305 – 1000 µM
KgtP	cAMP	0.0305 – 1000 µM

2.6 Investigating adaptions to IPTG/expression vector toxicity in *E. coli*

Throughout this section, the Nanodrop 2000c UV-Vis spectrophotometer was used to measure cellular OD_{600}. LB broth and LB agar constituents are listed in Section 2.2.1.

2.6.1 Testing the impact of IPTG and lactose on cell lines with genomic copies of T7 polymerase

Cell lines KRX (Promega, UK # L3002), Lemo21 (DE3) (NEB, UK # C2528J), BL21 (DE3) (NEB, UK # C25271) and MACH1 (Invitrogen, UK # C8620-03) were plated on LB agar with no selection pressure and grown at 37 ºC for 16 hours. Single colonies were selected and grown in 10 mL of LB broth at 37 ºC, 220 rpm, until an OD_{600} of 0.1 was reached. Cells were then streaked on LB agar containing one of the following substrates: unaltered control; 0.01 mM, 0.1 mM or 1 mM IPTG; 2 g/L or 15 g/L lactose (Sigma-Aldrich, UK). Plates were incubated for 16 hours at 37 ºC. Cell growth was visually analysed to identify the effect of IPTG on cell viability.

2.6.2 Testing the impact of IPTG on BL21 (DE3) cell containing expression vectors

BL21 (DE3) cells were transformed with a control H6-msfGFP vector or H6-msfGFP vectors containing MFS proteins (For vector formation see chapter 4.2). Transformed cells were grown on LB agar containing 50 µg/mL kanamycin for 16 hours at 37 ºC. Single colonies were selected and grown in 10 mL of LB broth containing 50 µg/mL kanamycin at 37 ºC and 220 rpm until an OD_{600} of 0.1 was reached. Cells were then streaked on LB agar containing 50 µg/mL kanamycin and one of the following concentrations of IPTG; 0 mM, 0.01 mM, 0.05 mM, 0.1 mM, 0.5 mM, 1 mM. Plates were incubated for 72 hours at 37 ºC. Cell growth was visually analysed to identify effect of IPTG on cell viability.

To test the persistence of any changes, cells were selected for growth from plates containing IPTG and the above steps repeated.

2.6.3 Testing the persistence of IPTG adaptions in various growth media

This section was done in collaboration with Dan Noel (University of Southampton). BL21 (DE3) cells were transformed with H6-sfGFP or H6-mdfA-sfGFP. Cells were

incubated at 37 ºC and 220 rpm to an OD_{600} of 0.1 and plated on agar plates, containing either 0.01 mM or 1 mM IPTG. Colonies were then numbered and used for further studies. For each construct, one colony was selected that demonstrated; no GFP expression (colonies 1 and 4), GFP expression (colonies 2 and 5), irregular colony growth (colonies 3 and 6) and control growth (colonies nm 1 and nm 2).

Colonies were used to inoculate non-selective LB media, incubated overnight at 37 ºC, 220 rpm and then spread on agar plates supplemented with either 0, 0.01 or 1 mM IPTG. Cells from these plates were used to inoculate LB broth and incubated overnight as previously, before inoculating liquid LB broth containing either 0, 0.01 or 1 mM IPTG. Readings for OD_{600} and GFP RFU were recorded at 2, 4, 6, 8 and 24-hour time points. Readings where plotted against time to investigate the effect of initial physiology on cell growth and GFP expression.

2.6.4 Isolation of DNA from BL21 (DE3) cultures adapted to high concentrations of IPTG

Plasmids were obtained from adapted cell lines using Gene-JET Plasmid Miniprep Kit, as described in Section 2.2.4. Extracted plasmids were then transformed into non-adapted cell lines (Section 2.2.3). Newly transformed cell lines were tested using procedures in section 2.7.2. Visual inspection was used to analyse the effect of each plasmid on cell viability at different IPTG concentrations.

Genomic DNA was extracted from cell lines using a Genomic DNA Purification Kit (ThermoFisher Scientific Ltd, UK). Extraction was undertaken according to manufacturer's instructions. All centrifuge steps were performed using a desktop centrifuge at 12 000 g. DNA was quantified using a Nanodrop 2000c UV-Vis spectrophotometer. The genomic prep quality was analysed using gel electrophoresis (Section 2.1.2).

2.6.5 Testing the expression potential of BL21 (DE3) cells adapted to high concentrations of IPTG

Cells were transformed with H6-msfGFP isolated from colony nm 1, from section 2.7.3, and grown on LB agar containing 50 µg/mL kanamycin for 16 hours at 37 ºC. Single colonies were selected and grown in 10 mL of LB broth containing 50 µg/mL kanamycin at 37 ºC and 220 rpm until an OD_{600} of 0.1 was reached. Cells were then streaked on LB agar containing 50 µg/mL kanamycin and either 0 mM or 1 mM IPTG. Two cell lines were selected from 1 mM IPTG and a lawn swab was used

from 0 mM IPTG. The colonies or swab were used to inoculate LB broth. This was grown to an OD_{600} of 0.8 before protein expression was induced with either 0, 0.2 or 1 mM IPTG. After overnight expression, the cells were harvested and visually analysed.

Chapter 3 Expression and purification of the eukaryotic *H. sapiens* SWEET1 protein

3.1 Introduction to the human sugar transport proteins

Sugars such as glucose and sucrose are vital for many forms of life. They are a main source of energy when consumed in the form of sugars, or carbohydrate supplied by vegetables, fruits, legumes and grain (Hu 2002). Sugars provide a source of carbon, a key component of many of the molecules that make up the cellular machinery of life, including DNA and the skeleton of organic compounds. Sugars are broken down during respiration to provide chemical energy that powers cells. The ability to obtain from the environment and utilise sugar is found throughout the majority of life on the planet.

While there are alternative forms of energy, many bacterial species use glucose preferentially, suppressing alternate metabolic pathways when it is available, through a process called catabolite repression. When glucose is unavailable cAMP and associated proteins are upregulated; this activates operons for alternative carbon sources, such as lactose. The addition of glucose to minimal media containing *E. coli* causes downregulation of the cAMP-receptor protein (CRP) suppressing the transcription of operons for other energy sources (Borirak et al 2014). This preference is not only seen in single-cell organisms but exists in us as the human brain uses 20 % of all expended energy in the form of glucose when in the resting state (Clark and Sokoloff 1999).

Clearly, living organisms are not exposed to a continuous stream of a fixed sugar concentration. Metabolic pathways must therefore be in place to deal with a variety of energy sources at a range of concentrations, enabling cellular survival under times of deficiency and excess. For bacteria, they control a series of metabolic pathways that respond to different energy sources by altering the expression of specific metabolic genes (Green et al 2014). There are many different potential carbon sources such as glucose, lactose, xylose, arabinose, sucrose and fructose, therefore, a number of different uptake proteins are required as the first step in the catabolic process. For prokaryotes, these include the phosphotransferase system (PTS), Sodium-Glucose Linked Transporter (SGLT), ATP binding cassette transporters (ABC transporters) and the Major Facilitator Superfamily (MFS) of

transporters. These are represented in Figure 3.1 A. The three types of transport featured are facilitated diffusion, active transport and secondary transport.

Facilitated diffusion enables the movement of a substrate down a concentration gradient with no required energy input, this is utilised by MFS transporters such as the *H. sapiens* GLUT1 (Jung and Rampal 1977). Both active transport and secondary transport enable the movement of a substrate against a concentration gradient.

Active transport proteins move a substrate up a concentration gradient by utilising the hydrolysis of an energy source. Information on these transporter families can be found in reviews for ABC transporters (Beis 2015), P-ATPases (Palmgren 2011), F-ATPases and V-ATPases (Futai et al 2012). Briefly; the PTS transporter imports sugar from the extracellular space to the cytoplasm. Sugar transport by the IMP dimer is driven by two energy coupling proteins, these transfer a phosphoryl group from phosphoenolpyruvate PEP to the incoming sugar forming a sugar phosphate (Milton and Saier 2015). An ABC transporter is a homodimer, each chain consists of a nucleotide binding domain and a transmembrane domain. The transmembrane domains come together to form a transporter, alternating access through the transporter is powered by the ATPase function of the nucleotide binding domain (Tanaka et al 2017).

Transport against a substrate gradient can also be achieved by coupling it to a favourable electrochemical gradient. The movement down a favourable gradient controls access, ensuring the transporter is only open to one side of the membrane at a time. The conformational changes will occur in a sequence ensuring substrates only travel in the desired direction. These are known as secondary transporters. The mechanism of a symport secondary transporter is illustrated in Figure 3.1 B. Energy is required to maintain an electrochemical gradient of substrate 1, which moves down its gradient via the secondary transporter. This energetically-favourable movement of substrate 1 is coupled to the less energetically-favourable movement of substrate 2, up its concentration gradient (Shi 2013). Coupled transport also occurs with the substrates moving in opposite directions these are known as antiporters.

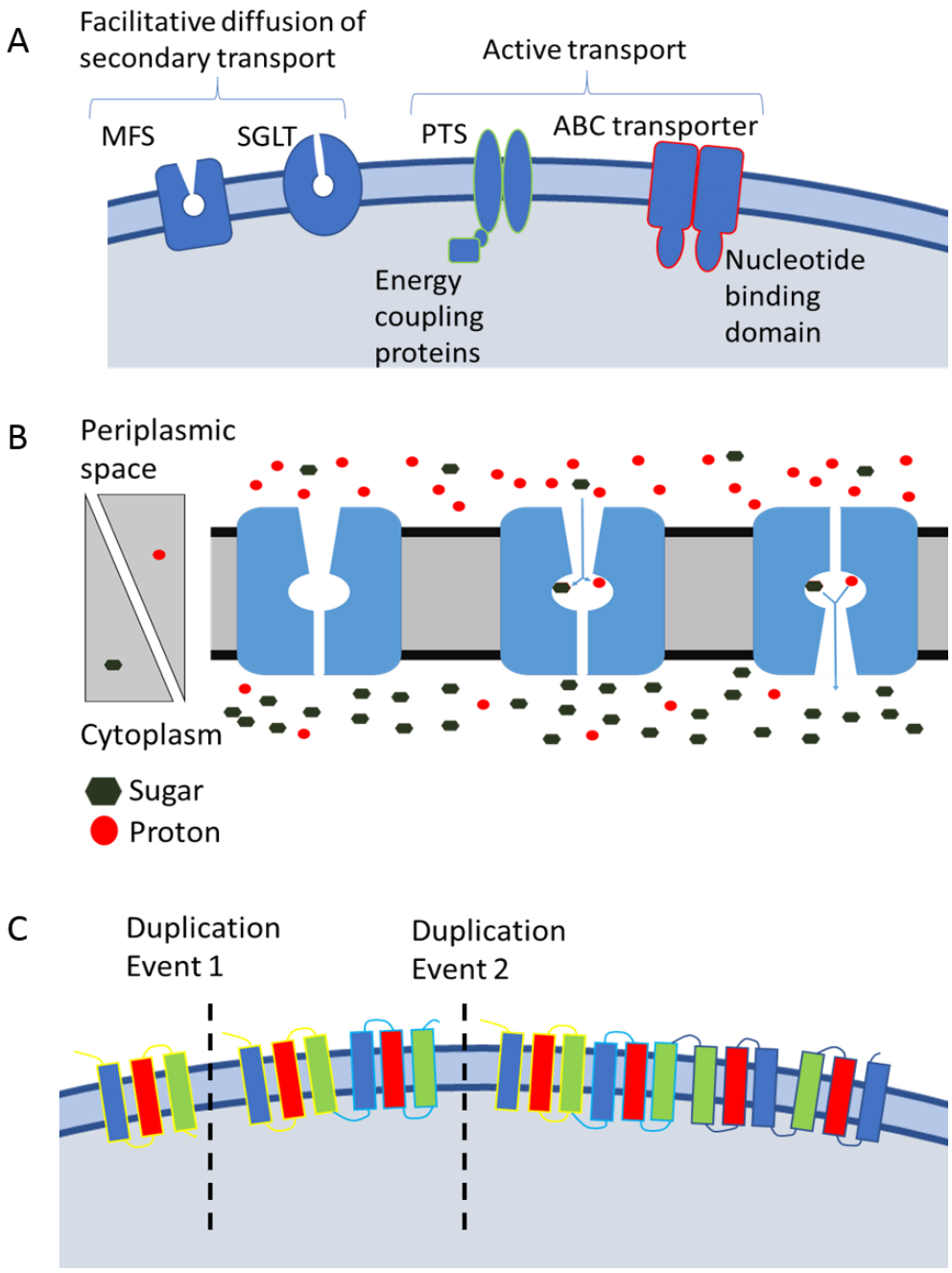


Figure 3.1 Schematics of key transport systems, including the mechanism of secondary transport and the evolution of the Multi Facilitator Superfamily (MFS)

A) Representation of membrane transport systems. The MFS and SGLTs (sodium-glucose transporters) are commonly secondary transporters (although some MFS transporters facilitate passive diffusion). The PTS (phosphotransferase system) and ABC (ATP-binding cassette) transporters use active transport powered by either their associated energy coupling proteins or nucleotide binding domains. B) Symport secondary transporter; the access alternates between periplasmic (outward)-facing, and cytoplasmic (inward)-facing, coupling substrate movement down a concentration gradient (protons in red) in exchange for a substrate moving up a concentration gradient (sugar in black). Once the binding site

is unoccupied, a conformational change occurs, shifting the protein back to outward open and allowing the cycle to repeat. C) The MFS topology was formed by two duplication events, creating a family with an inverted repeat of twelve transmembrane helices (TMH). Yellow-bordered TMH represents original protein; blue-bordered TMH represents protein from first duplication event; black-bordered TMH represents final protein from second duplication event.

This chapter focuses on the sugar transporters of *H. sapiens*. The importance of sugar movement between tissues is demonstrated by the severity of phenotypes caused by glucose transporter disruption. The mammalian GLUT1, also known as SLC2A1 (Solute Carrier Family 2 A1), is a facilitative sugar transporter with a relatively broad substrate specificity, including glucose, galactose and dehydroascorbic acid. Mutations resulting in the GLUT1 deficiency syndrome, a dominant, often de novo mutation, have a variety of phenotypes depending on the severity of the mutation. These can vary in severity from exercised-induced involuntary movements to complex motor disorders, epilepsy and mental retardation (Gras et al 2014). Renal Fanconi syndrome, caused by mutations in the related transporter GLUT2, results in inefficient absorption of sugar and salts from the kidneys, or storage in the liver. This has far-reaching effects on an individual's health from malnutrition disorders such as rickets, to abnormal salt levels in blood serum (Klootwijk et al 2015). Similarly, SGLT1 is a sodium-dependent glucose transporter in the intestinal mucosa of the small intestine. Known mutations disrupt intestinal absorption of glucose and galactose, leading to malnutrition and diarrhoea, causing potentially deadly dehydration (Wright et al 2003).

There are three groups of membrane proteins containing glucose transporters in *H. sapiens*; these are the GLUT (GLUcose Transporters), SGLT (Sodium GLucose Transporter) and SWEET (Sugars Will Eventually Be Exported Transporters) families. As features such as internal symmetry, oligomerisation, substrate binding and alternating access are found in all three classes of transporters, the features of all three groups will be considered in the following section. Special attention will be paid to the H. *sapiens* SWEET transporter and some of its paralogues.

3.1.1 Glucose Transporters (GLUT) of the Major Facilitator Superfamily

The MFS (Major Facilitator Superfamily) proteins are a ubiquitous superfamily of membrane transporters involved in the transport of a range of solutes, including sugars (Mirza et al 2006) phosphates (Pedersen et al 2013), nitrates (Yan et al 2013), antibiotics (Heng et al 2015) and peptides (Doki et al 2013). The GLUT

protein family is the glucose-transporting branch of this superfamily (Mueckler and Thorens 2013). Conforming to most of the superfamily, GLUTs have 12 transmembrane helices, resulting from two separate duplication events that quadrupled the membrane spanning regions (Hvorup and Saier 2002), this is represented in Figure 3.1 C. The amount of information available for individual members varies across the family. For example, the GLUT1 protein was discovered in 1970s and has been extensively characterised, including its crystal structure (Deng et al 2014). Comparatively the GLUT14 protein wasn't identified until 2002 (Wu and Freeze 2002) and it's function is still unknown (Shaghaghi et al 2016). The human GLUT1 was first characterised as a glucose transporter by using radioactive trace experiments in erythrocytes, demonstrating that glucose transport was inhibited by cytochalasin B (Jung and Rampal 1977), a known inhibitor of glucose transporters (Basketter and Widdas 1978). Binding of cytochalasin B to GLUT1 was also shown to be competitively inhibited by the presence of D-glucose but not L-glucose, demonstrating stereospecificity (Jung and Rampal 1977). There are currently 14 known GLUT transporters in humans, these are summarised in Table 3.1, exhibiting a variety of physiological roles, substrate specificity and subcellular localisation.

There have been recent breakthroughs in the structural characterisation of GLUT transporters. These have provided information on substrate interactions and the conformational changes that enable alternating access of the substrate to either side of the membrane. *H. sapiens* GLUT1 has been structurally defined in an inward-facing open conformation in the absence of a substrate (Deng et al 2014) and in the presence of inhibitors (Kapoor et al 2016). *H. sapiens* GLUT3 has been characterised in an occluded state, when bound separately to the substrates glucose or maltose (Deng et al 2015). Meanwhile, an outer-facing-open and inward-facing-open conformation has been structurally defined for the GLUT5 transporters of *R. norvegicus* and *B. taurus*, respectively (Nomura et al 2015).

Figure 3.2 A and Figure 3.2 B show the GLUT transporters open to either the intracellular or extracellular space. The GLUT1 tertiary structure exhibits pseudo symmetry as a result of the aforementioned duplication event. The ribbon structure of Figure 3.2 A shows that six N-terminal transmembrane helices bundle together to form one half of the protein while the 6 C-terminal transmembrane helices formed the other half. This arrangement surrounds a central cavity deemed to be the substrate-binding site. The inward-facing conformation of GLUT1 exhibits contact between transmembrane helices (TMHs) 1 and 7, forming a gate that

obstructs access to the binding pockets from the extracellular side of the protein (Deng et al 2014).

The alternating-access model of GLUT transporters was confirmed by overlaying the outward facing GLUT3 (Deng et al 2015) with the inward facing structure of GLUT 1 (Deng et al 2014). The N-terminal half of the two proteins maintained a rigid position between inward and outward facing, while the C-terminal TMHs 7 and 10 appeared to act as a hinge allowing local rearrangement, a representation of this is shown in Figure 3.2 C. TMH 10 moves away from the central cavity allowing extracellular access to the binding pocket, while TMH 7 shifts inwards to block intracellular access. The TMHs of the N-terminal region form more hydrogen bonds amongst themselves when compared to the C-terminal region; this may explain the difference in manoeuvrability during structural rearrangement. A hydrophobic interaction is present in their counterpart TMH 7 and 10 which may allow the structural rearrangement necessary for the alternate access (Deng et al 2015). Movement of the C-terminus around the rigid N-terminal region has also been proposed for bacterial MFS proteins, with TMHs 7 and 10 also highlighted as key in local rearrangement (Solcan et al 2012). The pepT transporter modelled in this paper is a proton dependent antiporter, not a diffusion facilitator.

Table 3.1 Summary of the SCL2A family including details on proposed substrates, cellular localisation and function in *H. sapiens*

Where experimental evidence proposes multiple characteristics, all are included.

Name	UniProt	Substrate	Localisation	Role
GLUT1 (SCL2A1)	P11166	Glucose (Jung and Rampal 1977), Mannose (Leitch and Carruthers 2009), glucoseamine (Uldry et al 2002), galactose (Ginsburg and Stein 1975), dehydroascorbic acid (Rumsey et al 1997)	Erythrocytes (Jung and Rampal 1977), brain (Yeh et al 2008), blood-brain barrier (Maher et al 1994), fetal trophoblast (Clarson et al 1997), muscle cells (Zhang et al 2016), cardiac muscle (Kraegen et al 1993).	Glucose transport around the body, expression levels thought to be related to energetic requirements. Highly expressed in erythrocytes and the brain they are considered the primary method of glucose acquisition across the blood brain barrier including the glial cells, astrocytes and oligodendrocytes (Yu and Ding 1998).
GLUT2 (SCL2A2)	P11168	Glucoseamine, (Uldry et al 2002), glucose (Johnson et al 1990), fructose (Cheeseman 1993), maltose (Colville et al 1993)	Brain (Arluison et al 2004), Intestine (Affleck et al 2003), Liver, Kidney, pancreas (Thorens et al 1990)	Low affinity glucose sensor allows cells to respond to variations in concentrations one example being insulin release from islets of Langerhans (Johnson et al 1990)
GLUT3 (SCL2A3)	P11169	Dehydroascorbic acid (Rumsey et al 1997), galactose, glucose, maltose (Colville et al 1993)	Brain, testes (Haber et al 1993)	Efficient transporter for highly energetic cells for example grey matter neurons (Haber et al 1993).

GLUT4	P14672	glucoseamine (Uldry et al 2002), glucose (Nishimura et al 1993).	Skeletal muscle (Stuart et al 2006), cardiac muscle (Kraegen et al 1993), adipose tissue (Liu et al 1992)	Internally localised at low insulin levels (Al-Hasani et al 2002), translocates to the plasma membrane upon insulin stimulation enabling glucose uptake (Huang et al 2005).
GLUT5	P22732	Fructose (Burant et al 1992).	Skeletal muscle (Stuart et al 2006), intestine (Davidson et al 1992), spermatozoa (Burant et al 1992).	Important for fructose uptake in the intestine (Davidson et al 1992). Required for differentiation of adipose tissue (Du and Heaney 2012).
GLUT6	Q9UGQ3	Glucose (Doege et al 2000).	Brain, spleen, leukocytes (Doege et al 2000) and skeletal muscle (McMillin 2017)	Unknown function, expressed in vesicles of adipose tissue, unaffected by insulin (Lisinski et al 2001)
GLUT7	Q6PXP3	Glucose and fructose (Li et al 2004).	Small intestine, colon, testis and prostate (Li et al 2004)	Currently unknown. Intestinally localised, but as substrate is uncertain the physiological function has not been assigned (Ebert et al 2017)
GLUT8	Q9NY64	Glucose, fructose, galactose (Ibberson et al 2000) and trehalose (Mayer et al 2016)	mRNA detected brain, testis, heart, skeletal muscle and small intestine (Doege et al 2000)	Currently unknown. The native protein is intracellular so may act as internal transporter (Widmer et al 2005) or respond to hormonal signal to relocated to the plasma membrane (Doege et al 2000). Has

				also been proposed as a trehalose importer (Mayer et al 2016).
GLUT9	Q9NRM0	Uric acid (Vitart et al 2008)	Kidney, liver, heart and chondrocytes (Mobasheri et al 2005)	Thought to function in the disposal of uric acid, a product of purine metabolism, as mutations lead to hyperuricemia (Vitart et al 2008)
GLUT10	O95528	2 – deoxy-D-glucose, D-glucose and D-galactose (Dawson et al 2001)	Heart, brain, lung, liver, skeletal muscle, pancreas, placenta and kidney (Dawson et al 2001)	Interacts with transforming growth factor-β in the formation of the cardio vascular system (Willaert et al 2012)
GLUT11	Q9BYW1	Glucose and fructose (Doege et al 2001)	mRNA transcripts present in lung, thymus, fetal kidney, fetal lung, heart, kidney, liver, pancreas, fetal heart, fetal kidney, bone marrow, lymph node, and tonsil (Sasaki et al 2001)	Currently unknown
GLUT12	Q8TD20	Glucose (Purcell et al 2011)	Skeletal muscle, adipose tissue, and small intestine (Rogers et al 2002)	Possibly functions as an insulin responsive glucose transporter (Purcell et al 2011)
GLUT13	Q96QE2	Myo-inositol (Uldry et al 2001)	Throughout the brain lowest expression neural and glial cells (Uldry et al 2001)	Intracellularly localised, but signal may lead to plasma membrane migration and cause inositol uptake (Uldry et al 2004). Interacts

				with y-secretase, required for membrane protein processing (Lee et al 2011)
GLUT14	Q8TDB8	Unknown	Testis (Wu and Freeze 2002), brain, lung, bone marrow, lymph node and epididymis (Uhlén et al 2015)	Currently unknown. Mutant alleles may have a role in Alzheimer's pathology (Shulman et al 2011)

Substrate-bound structures show the binding pocket and demonstrate protein-ligand interactions. The resolution of Deng et al's (2015) structure of *H. sapiens* GLUT3, bound to glucose at 1.5 Å, indicates residues coordinating protein-substrate binding through hydrogen bonds (Figure 3.2 D) and van-der-Waals interactions (Figure 3.1 E). Six out of the seven residues contributing to glucose binding via hydrogen bonds are located on the C-terminal region of the protein, with only Glutamine 159 contributing from the N-terminal transmembrane regions. The uneven distribution of substrate interactions with the N-terminal and C-terminal halves of the protein is known as asymmetric binding. Hydrophobic residues providing van-der-Waals interactions are more evenly split between the two halves of the protein. The C- and N-terminal regions both coordinate the substrate in the binding pocket.

Figure 3.2 E shows the structure of GLUT1 in the presence of n-nonyl-β-D-glucopyranoside. Although it is not a substrate of GLUT1, D-glucopyranoside is able to occupy the binding pocket, possibly due to the similarity of the head group to glucose (Deng et al 2014). Deng et al (2014) acknowledged the location may be affected by the detergent tail of n-nonyl-β-D-glucopyranoside, however, many of the residues involved in substrate binding are equivalent to those in the structure of glucose-bound GLUT3. Comparing Figure 3.2 D and E, the glutamine residues on transmembrane helix 7 are proposed to hydrogen-bond to the polar groups on carbons 1 and 2 of the glucose or glucoside. Meanwhile, the polar groups of glucose carbons 3 and 4, and carbon 6, hydrogen-bond to an equivalent asparagine on TMH7 and TMH8, respectively. This indicates the D-glucopyranoside is likely to be correctly orientated, however the detergent tail is interrupting any bonds that might occur with carbon 1. Alternatively, the comparably lower resolution of GLUT1 with D-glucopyranoside structure at 3.2 Å may have meant some bonds were less obvious leading to them being missed by the study.

Asymmetric binding between the N- and C-terminal end of proteins is also found in some bacterial members of the MFS family. The xylose transporter xylE was crystallised in the presence of xylose. Eight of the ten H-bonds between xylE and xylose are generated with residues in the C-terminal half of the protein (Sun et al 2012), demonstrating asymmetric substrate-binding. Both xylE and GLUT form more interactions between the substrate and the C-terminal half of the protein. The multidrug transporter mdfA binds to chloramphenicol with eight residues of the N-terminal region and four of the C-terminal region (Heng et al 2015). The N-terminal region of mdfA may be favoured for binding due to the larger size of the substrate

chloramphenicol or as a result of divergence between sugar and xenobiotic transporters.

As previously discussed, mutations in members of the GLUT family can manifest themselves as complex motor disorders and mental retardation. This exhibits their importance in the normal operation of the body. Another important concern in the study of the GLUT family is their contribution to the Warburg Effect, and thus implications for cancer treatments. RNA blotting analysis of digestive system tumours showed upregulation of GLUT1, GLUT2 and GLUT 3 (Yamamoto et al 1990). It was proposed this overexpression of glucose transporters allows increased glucose uptake as part of malignant transformation. Treatment of cancerous lung and epithelial cell lines with the drug dihydroartemisinin (DHA) reduced migration and invasion of the cells (Jiang et al 2016). Flow cytometry was then used to analyse the quantity of Myc-tagged GLUT1 on the cells surface, demonstrating that DHA diminished the presence of GLUT1 in a dose-dependent manner. DHA was also shown to be ineffective in reducing migration and invasion when the GLUT1 gene was knocked out of cancerous cells. This indicates that the GLUT family may be a potential anticancer target.

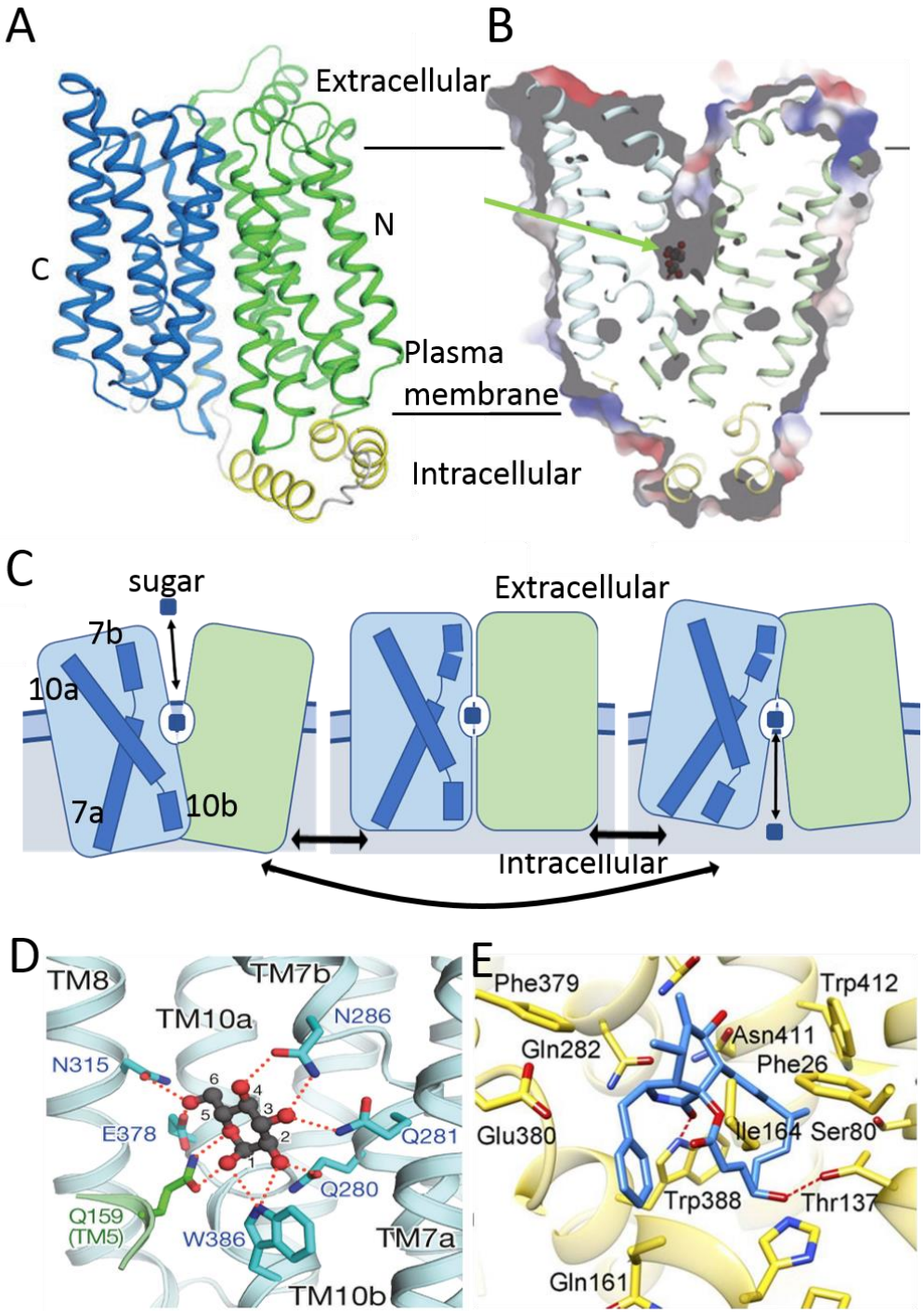


Figure 3.2 Overview of structural information on the GLUT family binding pocket and key residues involved in substrate interaction

A) *H. sapiens* GLUT1 ribbon structure coloured according to terminal domain, TMH 1-6 of the N-terminus in green and TMH 7-12 of the C terminus in blue. B) *H. sapiens* GLUT3 in outward-facing open conformation. Green arrow denotes binding pocket location with glucose molecule bound. A and B are shown relative to the plasma membrane. C) Conformational changes of GLUT1 transmembrane helices (TMH) 7 and 10 during alternating access (all conformational changes can happen in the reverse of described order). When outwards open, TMH10 is in contact with the N-terminal domain. Binding of

the substrate enables the switch to occluded state, as TMH7 forms contact with the N-terminal domain. From the occluded state TMH10 dissociates from the N-terminal domain, forming the inward-open conformation. When the substrate-binding domain is empty, the protein shifts back to outward open and the cycle repeats. D) H. *sapiens* GLUT3 residues hydrogen bonded to D-glucose in the binding pocket. E) H. *sapiens* GLUT3 hydrophobic residues providing Van-der-Waals interactions with D-glucose in the binding pocket. E) *H. sapiens* GLUT1 residues hydrogen-bonded to D-glucopyranoside of n-Nonyl-β-D-Glucopyranoside in binding pocket. Key residues involved in substrate binding are highlighted in D and E. Figures A and E adapted from Deng et al 2014. Figures B and D adapted from Deng et al 2015.

3.1.2 Members of the SGLT family, Sodium-Glucose Linked Transporters

Sodium-glucose symporters were proposed in the 1960s as a method of glucose uptake in the small intestine (Hamilton 2013), although no specific protein was identified. This was followed by studies demonstrating intestinal cell membranes were capable of transporting glucose in a sodium-dependent manner (Hopfer et al 1972; Murer et al 1976), but again this did not identify specific proteins. Hediger et al (1987) produced a cDNA library from rabbit intestinal RNA and injected it into *X. laevis* oocytes thus identifying the genes enabling glucose transport. The size of the proposed protein was equivalent to that of the Fluorescein isothiocyanate-labelled protein previously identified in rabbit brush border membranes (Peerce and Wright 1984). The Na^+/glucose co-transporter was later named Sodium-Glucose Linked Transporter (SGLT), the first member of the Solute Carrier 5 (SLC5) family. The SLC5A family has since been expanded with further members, exhibiting divergence in function and tissue expression. SLC5A family is a member of the solute:sodium symporter superfamily (SSF) (Wright et al 2004), with over 9000 SSF sequences in over 2000 species identified (Finn et al 2014). The group is characterised by the use of sodium to co-transport solutes up a concentration gradient (Wright et al 2004). A schematic diagram of vSGLT is shown in Figure 3.3 A; the group commonly has 14 transmembrane domains, with an inverted repeat made up of two lots of 5 membrane spanning regions forming the protein core (Abramson and Wright 2009).

The SGLTs are secondary transporters utilising alternating access. For instance, the SLC5A1 transports glucose up a concentration gradient in the small intestine, this is powered by transport of Na^+ down a concentration gradient (Gorboulev et al 2012). Structural data comes from prokaryotic SLC5 members as there are currently no refined structures of eukaryotic homologues. A galactose-sodium symporter from V. *parahaemolyticus* (vSGLT) has been crystallised in the inward-occluded state while bound to galactose, providing molecular details on the method of alternating access (Watanabe et al 2010). vSGLT has 32 % identity and 60 % similarity with H. *sapiens* SGLT1 (Faham et al 2008). The overall structure of vSGLT is shown in Figure 3.3, with the substrate-binding residues shown in Figure 3.3 C and 3.3 D (Faham et al 2008). Figure 3.3 B shows the topological arrangement of the proteins transmembrane helices. The two inverted-repeat regions, TMHs 2 - 6 and 7 to 11, form a central region surrounding the galactose and Na^+ substrates. Two sets of five membrane-spanning regions forming an inversely repeated structure is a common feature amongst the Sodium Solute Symporter Family. This tertiary structure is preserved even when sequence homology is limited (Abramson and Wright 2009). The key residues coordinating galactose through hydrogen-bonding, shown in Figure 3.3 C, are contributed by transmembrane helices 2, 3, 7, 8 and 11.

Faham et al (2008) purified mutant vSGLT and reconstituted it into liposomes to test transport of 14C-labelled galactose. Mutation of key residues Gln69, Glu88, Ser91, Asn260, Lys294, Ser365 or Gln428 to alanine, as labelled in Figure 3.3 C, reduced galactose transport by 80 %, demonstrating their importance in galactose-binding. The residues around the binding pocket are proposed to interact with each other above and below the substrate, forming a gate shown in Figure 3.3 D. The intracellular gate is formed of Tyr263 from transmembrane region 7 and lies in plane with the hydrophobic core of galactose. The triad of residues forming the extracellular gate are Met73 from transmembrane region 2, Tyr87 from transmembrane region 3 and Phe424 from transmembrane region 11. This hydrophobic positioning of a substrate within the binding pocket is seen in other families of saccharide transporters, including the GLUT family as described in the previous section (Deng et al 2015).

As a symporter, vSGLT has a binding site for a Na^+ ion. (Faham et al 2008) et al achieved a resolution of 2.7 Å for the structure, which made it difficult to locate the Na^+ ion. Structural alignment was carried out between vSGLT and a 1.65 Å

structure of leuT (Yamashita et al 2005), a sodium symporter of small, hydrophobic amino acids, allowing the prediction of a Na^+ ion binding site, as shown in Figure 3.3 E. The Na+ binding site is important as galactose uptake is coordinated by transport of Na+ ions down a concentration gradient. Na^+ is bound by Ala62, Ile65, Ala361, Ser364 and Ser365. The release of Na^+ from this binding is hypothesised to enable the conformational change from occluded occupied to inwards open (Watanabe et al 2010). As the Na^+ ion vacates this position TMH1 shifts and the residues Ala62 and Ile65 form H-bonds with Ala361, Ser364 and Ser365. This change forms and stabilises the inward open conformation. The movement of TMH1 and its surrounding domains is shown in Figure 3.3 F. The grey sections show the occluded conformation while the coloured sections are the new locations when the transporter is in the inward, open conformation. Figure 3.3 G shows the global realignment of the vSGLT transmembrane regions from inward occluded (grey) to inward open (coloured) enabling the release of galactose.

Currently SSF structures are not available to show every stage of alternating access. However, the SSF family shares core structural inverted repeat domains with NSC1 and NSS families, namely the SFF inverted repeat region shown in Figure 3.3 A (Abramson and Wright 2009). The structure of the NSC1 protein Mhp1 can be used to propose a model for alternating access in vSGLT. In the outwards facing conformation Na^+ binding allows substrate access to the binding site. Following substrate binding the protein adopts an occluded state (Weyand et al 2008). From here the structures of vSGLT provide the final information that Na^+ release from the occluded state allows the structure to open inwards and release the substrate. In this way the symporter is reliant on the two different molecules to complete a transport cycle. As an abundance of Na^+ exists on the extracellular side, this driving force can transport desirable substrates against a concentration gradient.

There is considerable interest in targeting SLC5A family members as biomarkers or drug targets in a range of diseases including: cancer biomarker (Ren et al 2013, Scafoglio et al 2015, Yao et al 2015), atherosclerosis (Neumann et al 2012) and diabetes (Vallianou et al 2016).

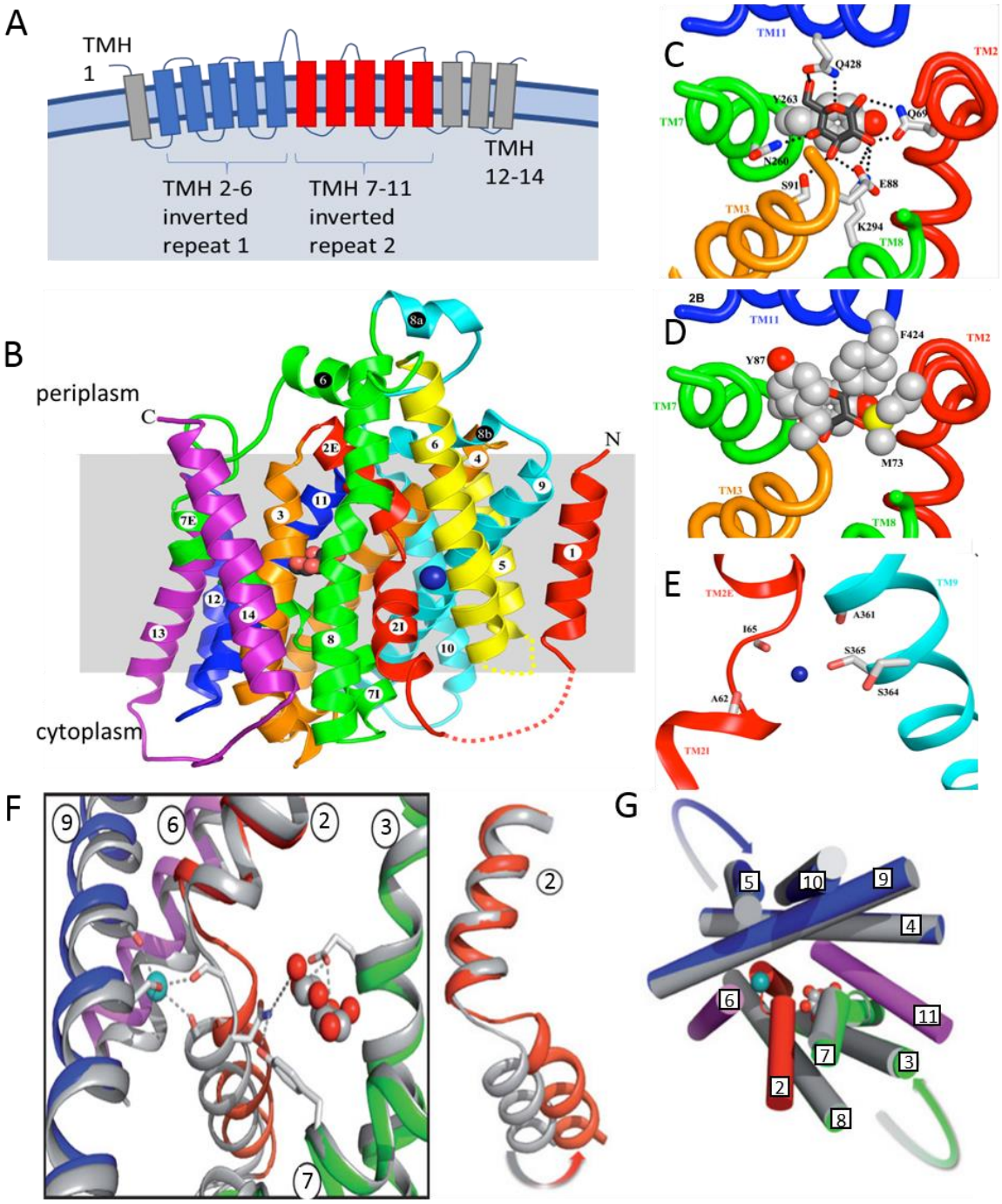


Figure 3.3 The Structure and key features of vSGLT demonstrate the method of alternating access

A) Topological model of vSGLT highlighting the central inverted repeat. B) Ribbon diagram of vSGLT with transmembrane domains numbered, bound galactose displayed as ball and stick model, proposed Na^+ binding sight shown as dark blue sphere. C-D) Key residues of vSGLT in the occluded state bound to galactose. C) Hydrogen bonding between binding pocket residues and galactose. D) Hydrophobic R groups forming the gate occluding the bound galactose. E) Key residues of vSGLT's proposed Na^+ binding site in the occluded site, displayed in presence of Na^+. F-G) (Structures and overlay of the inward-open (coloured) and inward-occluded (grey) conformations, demonstrating the movement of

TMHs between the two states. F) Movement of substrate binding transmembrane regions between inward open and inward occluded confirmation. G) Global movement of transmembrane region between inward-open (coloured) and inward-occluded (grey) conformations, Movement of transmembrane region 2 displayed separately from global model. Figures B-E are adapted from Faham et al (2008). Figures F-G are adapted from Watanabe et al (2010)

3.1.3 SWEET Transporters

The final group of glucose transporters are members of the SLC50 family, or SWEET (**S**ugar **W**ill **E**ventually **E**xport **T**ransporters) transporters. The A. *thaliana* glucose transporter SWEET1 (AtSWEET1) was identified by screening genes in HEK293T cells co-expressing the glucose sensitive reporter FLIPglu600μΔ13V (Chen et al 2010). *A. thaliana* has 17 paralogues of SWEET. When expressed in yeast, AtSWEET4, AtSWEET5, AtSWEET7 and AtSWEET13 enhanced the cells glucose uptake, however, the other 12 SWEETs showed no change. The substrates and physiological role of the SWEET proteins does appear to have diverged. For instance, the Arabidopsis *sweet11 sweet12 sweet15* triple mutant was unable to build up sucrose stores in seeds (Chen et al 2015) and *sweet9* mutants were unable to secrete sucrose from the nectary organ. Reduced expression of AtSWEET17 leads to reduced distribution of fructose from leaves to other plant tissue (Chardon et al 2013). Many of these functions are not required in humans and explain why there is only a single *H. sapiens* homologue.

H. sapiens SWEET, HsSWEET, was previously known as RAG1AP1. A study into the methylation of genes in breast cancer tumours identified HsSWEET as a potential biomarker for diagnosis (Świtnicki et al 2016). Overexpression in cancers may be due to a role in sugar demand, as when over-expressed in *Xenopus* oocytes, HsSWEET caused in increased efflux of 14C labelled glucose (Chen et al 2010). Świtnicki et al (2016) asserted that the increased methylation resulted in upregulation of HsSWEET and may be involved in the Warburg effect, although this data was not provided. This research does not provide enough information to estimate what the physiological role of HsSWEET may be. The goat homologue of SWEET, ChSWEET was expressed in goat mammary gland epithelial cells to investigate the effects this would have on glucose homeostasis (Zhu et al 2015). A fluorophore-based glucose assay indicated efflux of glucose almost tripled in the ChSWEET expressing cells when compared to the controls. This does align with the aforementioned Xenopus studies of HsSWEET, as both suggest it may have a role in efflux of glucose, although neither provide details on tissue localisation or

function. Chen et al. (2010) used a microarray experiment to screen human tissues for HsSWEET expression and concluded it was ubiquitously expressed, with the highest levels in the oviduct epididymis and intestine. The human atlas antibody staining of human tissue (Uhlén et al 2015) indicated that HsSWEET protein was ubiquitously present with particular prevalence in brain, endocrine tissue, lung, liver and gallbladder, pancreas, kidney and urinary bladder. Due to the limited experimental data, a physiological role has yet to be outlined for HsSWEET. It is possible that it performs a similar function as the AtSWEETs but using a single protein that is widely expressed as opposed to multiple paralogues. This would lead to the unanswered question then of why humans still only require one copy while other organisms have multiples copies of SWEET transporters. Until further information is ascertained, it is difficult to assign a function, hence, the work outlined in this chapter aims to help uncover its biological role.

HsSWEET belongs to the PQ loop repeat family prevalent in all kingdoms of life (Yuan and Wang 2013). Other *H. sapiens* PQ loop repeat genes include the cystinosin transporter (CTNS) and PQLC genes. CTNS is internally localised to the lysosome and functions as a proton dependent cysteine transporter. CTNS mutations can lead to the autosomal recessive genetic disorder cystinosis (Taranta et al 2010). For these patients, the defective CTNS transporter is not able to transfer cysteine residues into the cytoplasm. It is the excessive build-up of cysteine within the lysosomes that eventually is the underlying cause of the observed medical problems, which include myopathy, diabetes, pulmonary dysfunction, retinopathy, hypothyroidism and death (Gahl et al 2007). The disease is treatable with a compound called cysteamine, although this only reduces the frequency of symptoms. Cysteamine reacts with cysteine residues to convert them into a positively charged amino acid similar in chemical structure to lysine. Based on its homology to the *S. cerevisiae* lysosomal cationic amino acid transporters Ypq1-3, Jézégou et al (2012) identified PQLC2 as the human lysosomal cationic amino acid transporter. EGFP tagged rat PQLC2 was shown to localise with the lysosomal marker protein LAMP-1. When expressed in Xenopus oocytes a membrane localising PQLC2 mutant facilitated pH dependent transport of radiolabelled Arg, Lys and His (Jézégou et al 2012). In cystinotic fibroblast cells, silencing of PQLC2 with siRNA exacerbated the retention of cellular cysteine and cysteamine-cysteine mixed disulphide (MxD), even when treated with cysteamine. It has been proposed that cysteamine treatment of CTNS allows removal of cysteine from lysosomes in the form of MxD. Jézégou et al (2012) concluded that PQLC2 allows cysteamine to relieve cystinosis. A *C. elegans* homologue of PQLC2 showed the same lysosomal

phenotype when knocked out (Liu et al 2012) demonstrating conservation of the pathway across some sections of the Metazoan kingdom. Analysis of *H. sapiens* open reading frames has also identified the homologues PQLC1 and PQLC3 (Gerhard et al 2004), however, these proteins remain uncharacterised. HsSWEET, like CTNS and PQLC2, is an internally localised transporter that at the molecular level probably operates in a similar fashion. The main difference is that the substrate for HsSWEET appears to be glucose rather than amino acids.

There is currently no structure of an animal SWEET transporter, however, a structural picture of the family is being built up with the use of bacterial and plant homologues. A key difference between eukaryotic and prokaryotic SWEET transporters is that eukaryotic transporters consist of seven transmembrane helices, with an internal duplication of three TMHs, forming a 3-1-3 organisation, while prokaryotic SWEET transporters consist of only three TMHs. For prokaryotes, the active transporter is a dimer that matches the internal duplication in the eukaryotic genes. Therefore, the prokaryotic homologues have been dubbed SemiSWEETs. Like the GLUT transporters, the SWEET transporters move their substrates using a similar alternating access mechanism of transport. The *E. coli* SWEET transporter, EcSemiSWEET, has been characterised in the inward and outward open conformations (Lee et al 2015). The *Thermodesulfocibrio yellowstonii* SWEET TySemiSWEET has been characterised in an occluded state (Wang et al 2014). The eukaryotic *Oryza sativa* (rice) OsSWEET2b has been structurally characterised in the inward open conformation (Tao et al 2015).

Overlaying the bacterial EcSemiSWEET, as shown in Figure 3.4 A and 3.4 B, demonstrates the structural differences between alternating access. Each EcSemiSWEET is a homodimer, in which two triple helix bundles come together forming the tertiary structure. The angle of the PQ loop region between TMH1a and TMH1b rotates by 30 ˚ between the two conformations. While in the inward open conformation, the two triple helix bundles are drawn together at the extracellular end and apart at the intracellular end. The PQ loop is quite central and appears to act as a hinge. As the conformation switches to outward open, the TMH regions are drawn together at the intracellular end and apart at the extracellular end. The outwards open conformation is stabilised by hydrogen bonding between GLN22 of TMH1b and the amide group on the TMH2 backbone. This global structural rearrangement of EcSemiSWEET brings together or separates the residues responsible for gating the protein as shown in Figure 3.4 C. The extracellular gate is formed from Arg57, Tyr53 and Asp59 forming a combination of salt bridges and

hydrogen bonds using residues from each monomer. While the intracellular gate is formed of Phe19, Met39, Tyr40 and Phe43 forming hydrophobic interactions using residues from each monomer.

The gates flank the substrate binding region found in the centre of the protein, this is shown in the occluded state for TySemiSWEET (Figure 3.4 D). This is the intermediate state between inward-open and outward-open in which both extracellular gates are simultaneously closed. Comparison between TySemiSWEET and the *L. biflexa* homologue LbSemiSWEET from (Xu et al 2014) suggests the substrate dependent adaptions. LbSemiSWEET is a glucose transporter (Xu et al 2014) with a substrate binding pocket of 424 $Å^3$ while TySemiSWEET, a sucrose transporter, has a larger substrate binding pocket of 463 $Å^3$ (Wang et al 2014). The architectures of these pockets are largely similar and the disparity in volume is a result of a Phe to Met substitution on TMH2. The Met residue shown in Figure 3.4 D result in a larger substrate binding volume, an adaption for transport of the disaccharide sucrose.

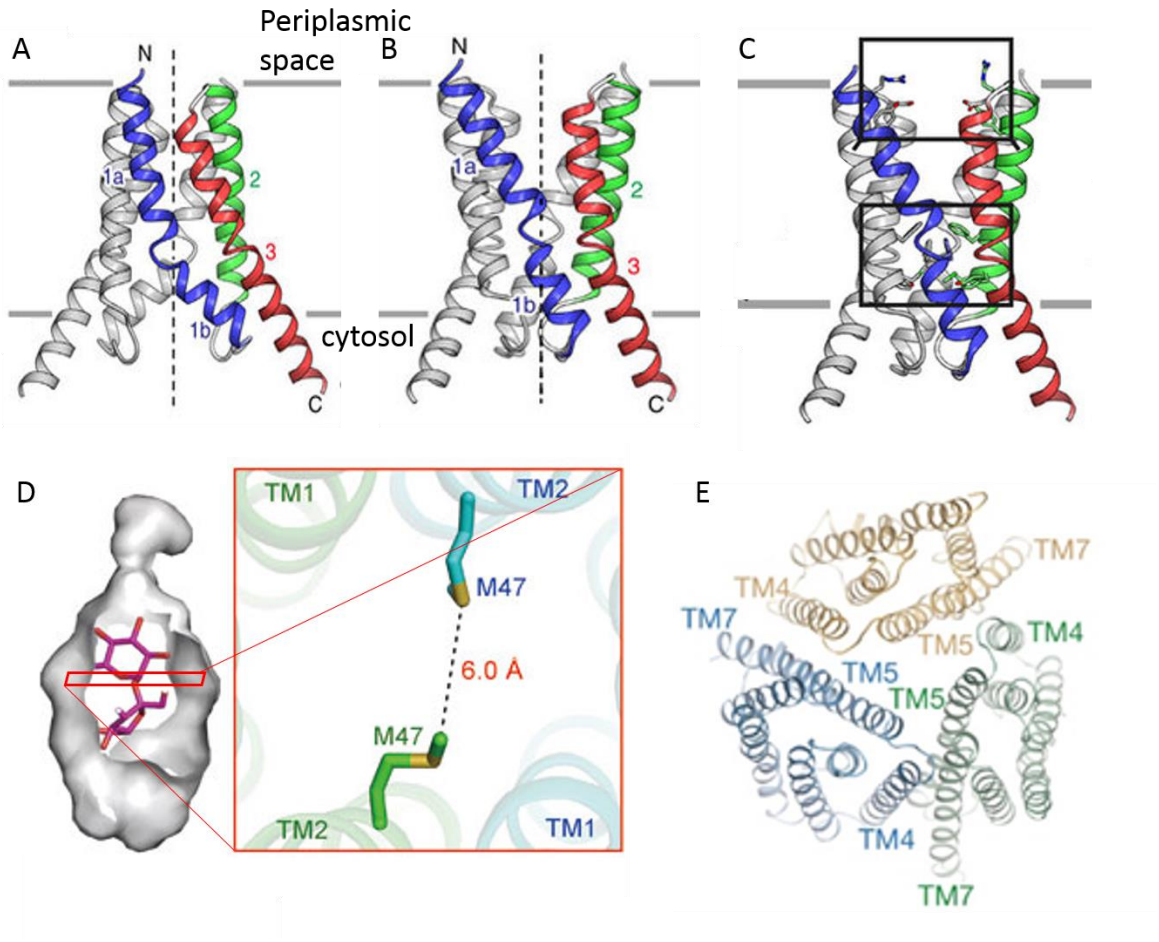


Figure 3.4 The alternating access mechanism of SWEET transporters can be ascertained by overlapping the structures of homologues at different access points

A-C) The movement of TMHs enabling alternating access in the *E. coli* homologue of a SWEET protein, EcSemiSWEET. Arrangement of transmembrane regions for the A) inward open, B) outward open EcSemiSWEET dimer. C) Gating residues highlighted on the outward open conformation. D) Binding site of occluded TySemiSWEET and E) enhanced view of the key residues dictating the binding site size. E) Top down view of OsSWEET trimer topology. The quaternary structure is formed of three OsSWEET monomers. A-C) adapted from Lee et al 2015. D) adapted from Wang et al 2014. E) adapted from Tao et al 2015.

The eukaryotic OsSWEET2b is a glucose transporter closely related to AtSWEET2. Tao et al (2015) showed that tagged OsSWEET2b localised to the vacuolar membrane when expressed in yeast. Transport assays indicated when OsSWEET2b was expressed in HEK293 or reconstituted into liposomes it facilitates glucose uptake. While it transports in the opposite direction to HsSWEET it does use the same substrate. The eukaryotic protein has 7 TMH regions made of two consecutive triple helix bundles each containing a PQ loop, TMH1-3 and TMH5-7 with a linker region in the middle (Xuan et al 2013). The structure of OsSWEET showed that it was organised as a trimer (Figure 3.4 E; Tao et al 2015). When overlaid the triple helix bundles of the EcSemiSWEET dimer match the conformation of the consecutive triple helix bundles of a single OsSWEET protomer. This could mean the OsSWEET trimer is actually three transporters that have formed one trimeric complex. In this structure, the OsSWEET cavity is open cytosolically but closed at the extracellular facial gate. As this is predicted to be localised to the lysosome, it suggests that the closed gate would be on the lysosome's lumen side (Tao et al 2015). The extracellular facial gate consists of hydrogen bonding between Tyr61, Asp190 and Gln132, this parallels the gate formed on the periplasmic face of the characterised SemiSWEETs (Lee et al 2015). An interesting difference in the OsSWEET structure is the asymmetry in the binding site. While the homodimers of SemiSWEET transporters have identical residues in the substrate binding site, there is a difference in the amino acids contributed by triple helix bundle 1 compared to triple helix bundle 2 of OsSWEET transporter. As it is not formed of a homodimer a mutation of the residue in one bundle may adjust function more subtly, allowing for development of greater variation of function. This might go to explain why some species, such as *A. thaliana*, have such an abundance of SWEET proteins (Chen et al 2010).

This research described in this chapter is aimed at producing sufficient quantities of purified HsSWEET protein to allow functional characterisation and potentially crystallographic characterisation. A GFP-tagged version was generated and then used to optimise protein over-expression in a prokaryotic expression system. The development of this method can be found in Chapter 4. Additionally, the intracellular localisation of tagged-HsSWEET in human cell lines was investigated.

Chapter 3

Aims

- To Clone HsSWEET into the pNIC28-Bsa4 and over-express in BL21(DE3) cells.
- To purify HsSWEET using affinity chromatography for use in functional and/or crystallographic studies.
- To generate HsSWEET constructs tagged with GFP and HA for expression in HELA and RPE cells and use confocal microscopy to determine subcellular localisation.
- To clone HsSWEET into the pET28b vector with a GFP tag and express in BL21(DE3) cells.
- To use the GFP-tagged version of HsSWEET for functional characterisation.

3.2 HsSWEET cloning, cellular localisation and expression

3.2.1 Generating the HsSWEET expression systems

Several expression systems were designed with the aim of generating large amounts of over-expressed protein. These are outlined in Table 3.3. A synthetic version of HsSWEET was synthesised by Biomatik, codon optimised for high level protein expression in *E. coli*. In addition, this synthetic gene had all of the cysteine residues mutated to either Val or Ser residues. It was hypothesised that the Cys residues would cause problems during protein purification due to disulphide bond formation as well as potential folding issues hence their removal, this is explored in more detail later in this Section. Based on the solved SWEET structures, none of the mutations are expected to be in functionally significant sites such as gate regions or the selectivity filter. This gene was named HsSWEETCysless and is aligned to WT HsSWEET amino acid sequence below in Figure 3.5 A. The topological representation of HsSWEET a membrane is shown in Figure 3.5 B.

The expression vector selected for the *E. coli* based expression system was pNIC28-Bsa4 (pNIC28-Bsa4 was a gift from Opher Gileadi (Addgene plasmid # 26103)). The final protein construct incorporates an N-terminal hexahistidine tag followed by a TEV protease cleavage site as shown in Figure 3.5 C. Histidine tags have the potential to interfere with the crystallisation process making the TEV cleavage site an important element of this vector.

For the eukaryotic expression systems, two vectors were selected (Table 3.2). Cloning was performed using *E. coli* before transfection into eukaryotic cells. The eukaryotic vectors expressed HsSWEETCysless with reporter tags allowing identification of cellular localisation. The vectors selected contain GFP and HA tags, expressed at the N or C terminus. It was unclear if the size or location of the tag was going to interfere with the results, hence all four options were tested. The HA tag is derived from the human influenza haemagglutinin surface glycoprotein and has the amino acid sequence YPYDVPDYA. This was selected as it is small and may overcome potential issues attributed to the size of GFP.

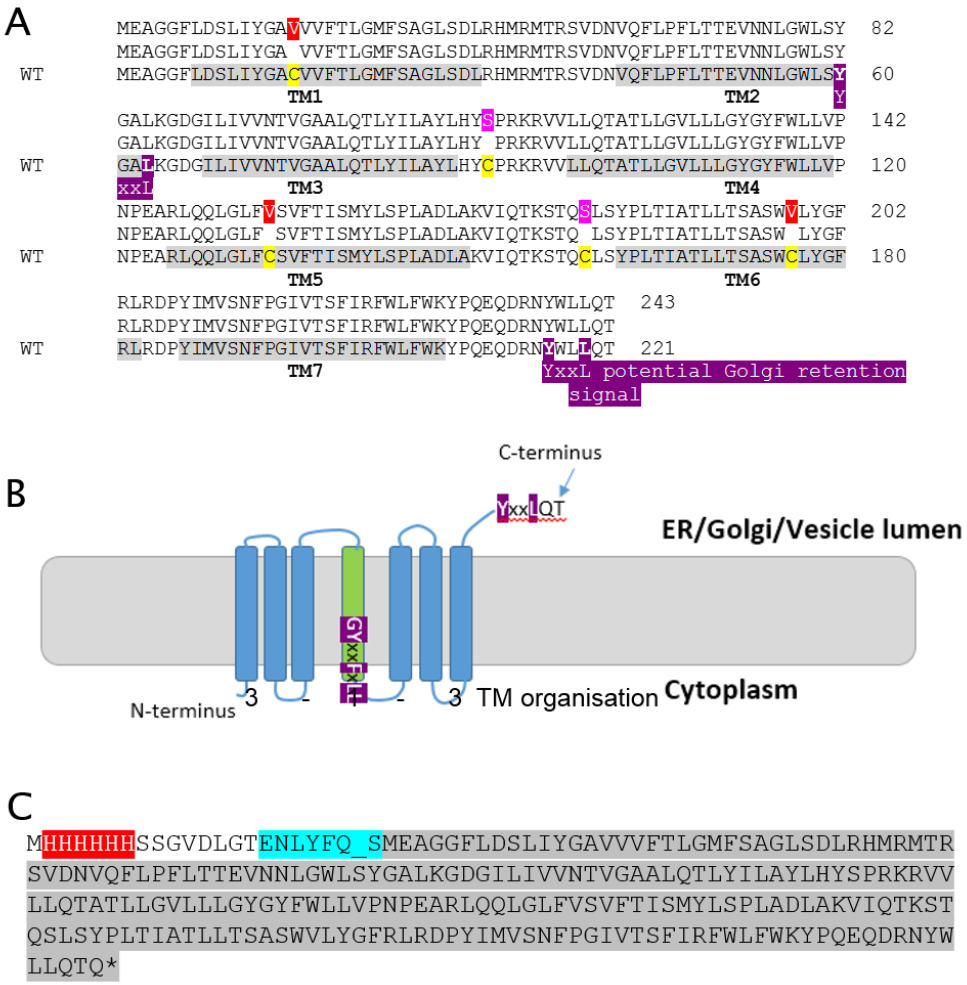


Figure 3.5 Primary sequence and topological structure of HsSWEETCysless

A) Clustal Omega (Sievers et al, 2011) amino acid pairwise sequence alignment of WT HsSWEET and HsSWEETCysless. Grey bars, TMH based on prediction by TOPCONS (Tsirigos et al 2015). Purple bars, Golgi retention signals (YxxL) identified by LocSigB (Negi et al 2015). B) HsSWEET protein topology predicted by TOPCONS with Golgi retention signals. C) The amino acid sequence of the expression construct generated by cloning HsSWEETCysless (grey highlight) into pNIC28-Bsa4 to produce a TEV cleavable (blue highlight; underscore represents the cleavage location) hexahistidine tag (red highlight).

Table 3.2 Protein constructs and vectors used in the HsSWEETCysless localisation experiments.

The two tags, green fluorescent protein (GFP) and the hemagglutinin tag (HA), denoting the primary sequence of the protein and tag.

Construct	Vector
GFP-HsSWEETCysless	pEGFP-C1
HsSWEETCysless-GFP	pEGFP-N1
HA-HsSWEETCysless	pcDNA3.1
HsSWEETCysless- HA	pcDNA3.1

The native HsSWEET sequence contains an odd number of cysteine residues. This is potentially problematic for the crystallisation process. Cysteines may form disulphide bridges in an irregular fashion leading to a disorganised crystal lattice. An irregular lattice will not diffract X-rays giving no structural data. However, cysteine residues may also be important for the protein tertiary structure. Modelling software was used to explore the possibility of replacing cysteine residues. A topological prediction of HsSWEET using the predictive software TOPCONS (Tsirigos et al 2015) is shown in Figure 3.6 with the five cysteines marked. Based on this prediction the topology matches that of structurally defined SWEET proteins from other species.

Structural predictions of HsSWEET were made using the SWISS-MODEL homology modelling software (Figure 3.6 B; Biasini M et al 2014). SWISS-MODEL uses a Basic Local Alignment Search Tool (BLAST) search to identify proteins with similar sequences to HsSWEET from the Protein Data base (PDB), these are shown in Table 3.5. The identified SWISS-MODEL structures are shown in Figure 3.6 B-D. This is shown alongside an equivalent view of the OsSWEET (Tao et al 2015), demonstrating the similarities between the structural models. However, measures of the SWISS-MODELs global quality are largely negative indicating a low probability of accuracy. Even so, it was still felt that the model could be used to give some guidance as to general location of the Cys residues and their potential mutants. The main assumption was that Cys residues in hydrophobic regions could be mutated to Val residues while those in soluble loops could be changed to Ser residues.

Table 3.3 Outline of constructs designed to investigate human SWEET1 protein over-expression in *E. coli*

All vectors were assembled in *E. coli* using PCR products and digested vector. The pNIC vector was used for *E. coli* expression, all other vectors were used in eukaryotic cell line localisation studies.

Construct name	**PCR size (b.p.)**	**(AA's)**	**Vector**	**Tag**	**Construct + tags**
pNIC-HsSWEETCysless	690	226	pNIC28-BSA4	n/a	start - Hexahis - TEV - target protein- STOP
pEGFP-N1-HsSWEETCysless	675	221	pEGFP-N1	C	HindIII – START-HsSWEETCysless – CG – BamHI – GFP - STOP
pDDGFP-HsSWEETCysless c004	675	221	pcDNA3.1	N	HindIII – START – HA – HsSWEETCysless – STOP – BamHI
pDDGFP-HsSWEETCysless c005	675	221	pcDNA3.1	C	HindIII – START HsSWEETCysless – HA – STOP – BamHI
pEGFP-C1-HsSWEETCysless	675	221	pEGFP-C1	N	START – GFP – HindIII – HsSWEETCysless – STOP – BamHI

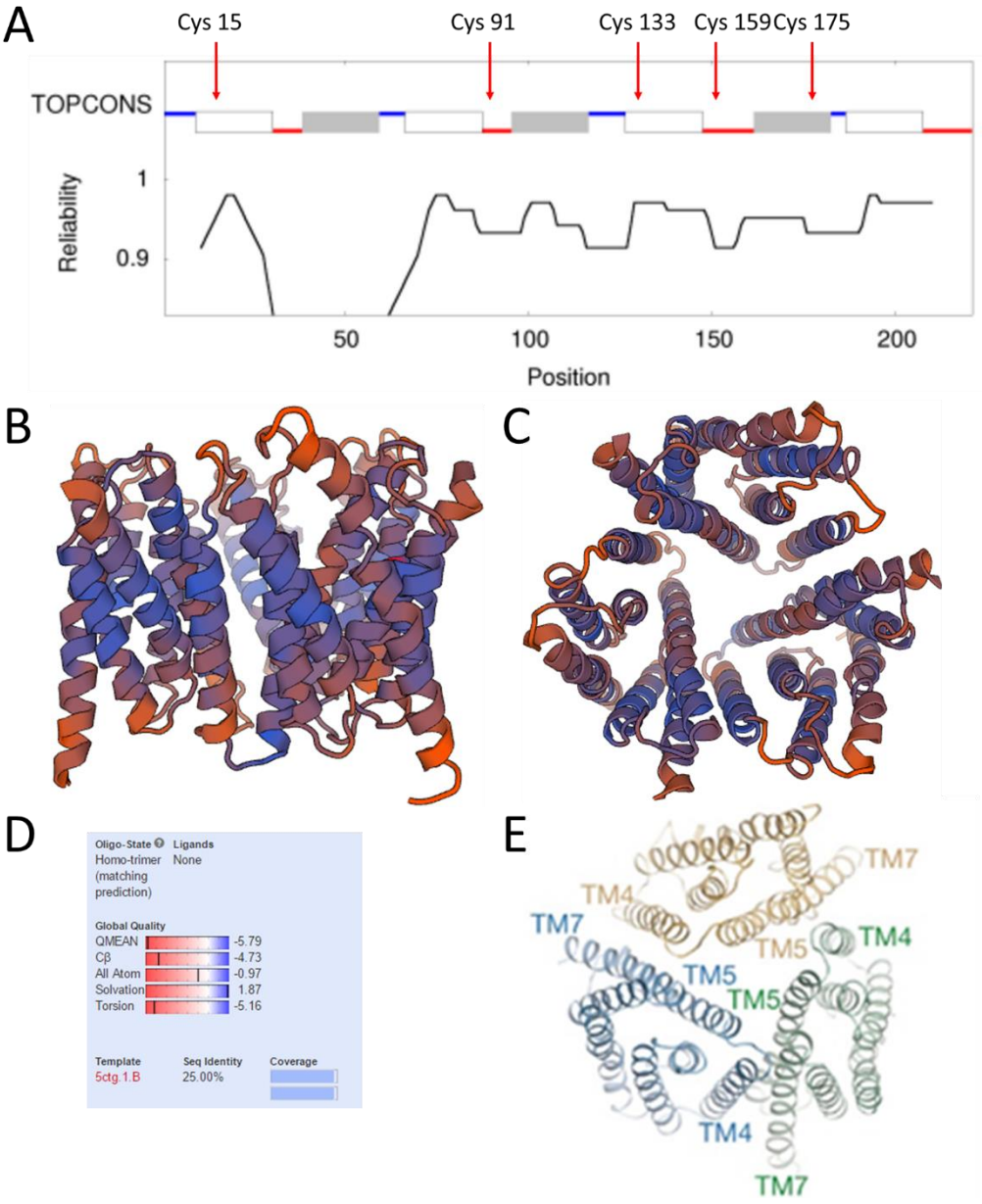


Figure 3.6 Bioinformatics predictions of *H. sapiens* SWEET transporter (SLC50A1)

HsSWEET sequence Gene ID: 55974. A) TOPCONS (Tsirigos et al 2015) transmembrane helix (TMH) predictions for the amino acid sequence of HsSWEET. TMH are shown as white and grey boxes connected by blue intracellular loops or red extracellular loops. Each area is scored on reliability. Red arrows indicate the location of cysteine residues. B-C) SWISS-MODEL (Biasini et al 2014) homology-based structural prediction using the HsSWEET primary sequence. Model is informed by published structures of homologues from *O. sativa, L. biflexa, Vibrio sp, B. japonicum*. D) Global quality scores are largely negative indicating a low probability of accuracy. E) OsSWEET crystal structure adapted from Tao et al (2015). The difference in colour represents separate monomers, numbers indicate TMHs.

Table 3.4 PDB proteins homologous to HsSWEET identified by SWISS-MODEL BLAST

Proteins were identified using a BLAST of the *H. sapiens* SWEET1 primary sequence. Identified structures were used by SWISS-MODEL to predict the structure of *H. sapiens* SWEET.

PDB ID	Species	Reference
5CTG, 5CTH	O. *sativa*	(Tao et al 2015)
5X5M, 4QNC	L. *biflexa serovar Patoc*	(Lee et al 2015)
5X5N, 4QND	*Vibrio sp.*	(Lee et al 2015)
4rng	B. *japonicum*	(Wang et al 2014)

To investigate the probability of the cysteines forming internal disulphide bridges HsSWEET was aligned with the sequence of OsSWEET, shown in Figure 3.7 A. The alignment shows no conservation of cysteine residues between the proteins. The alignment was used with the solved OsSWEET PDB structure and the SWISS Model structure to investigate the possible proximity of cysteine residues. In Figure 3.7 B, each cysteine residue from HsSWEET has been used to mark its aligned residue on OsSWEET. This demonstrates Cys15 and Cys175 may be in adjacent transmembrane helices. However, Cys91 Cys133 and Cys159 are unlikely to be close enough to form disulphide bridges. Due to the lack of proximity or conservation between the structures and potential for interference during crystallisation, the HsSWEET sequence was mutated to remove cysteine residues, shown in Figure 3.8. Mutagenesis was achieved by purchasing a synthesised sequence; this also enabled codon optimisation of the HsSWEET gene for *E. coli* expression.

PCR was used to copy the codon optimised, synthetic HsSWEETCysless gene for cloning into the vectors. PCR products from Table 3.3 are shown in Figures 3.9 A and B. All PCR reactions produced bands of the expected size. Two of the PCR runs produced smaller non-specific bands than expected, however, gel purification was used to isolate PCR product of the appropriate size. These purified DNA fragments were then used in one of two ways. Vectors were cloned using restriction digestion or ligation independent cloning as outlined in Materials and Methods. All constructs were verified by DNA sequencing and confirmed to be correct.

A

```
          10         20         30         40         50
MEAGGFLDSLIYG-ACVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLS
.DSLYDISCFAA.L.GNI.A.AL.LSPVTTFKRILKAK.TERFDG..Y.FSLL.C.IC.W

60         70          80         90          100        110
YG-ALKGDGILIV--VNTVGAALQTLYILAYLHYCP----RKRVVLLQTATLLGVLLLGY
..LPWVA..R.L.AT..GI..VF.LA..CLFIF.ADSRKT.MKIIG.LVLVVC.FA.VSH

     120        130        140        150        160        170
GYFWLLVPNPEARLQQLGLFCSVFTISMYLSPLADLAKVIQTKSTQCLSYPLTIATLLTS
ASVFFF--DQPL.Q.FV.AVSMASL...FA....VMGV..RSE.VEFMPFY.SLS.F.M.

     180        190        200        210        220
ASWCLYGFRLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT
..FA...LL...FF.YFP.GL.LILGAMQLA.YAY.SSN---SLEV.FQ
```

B

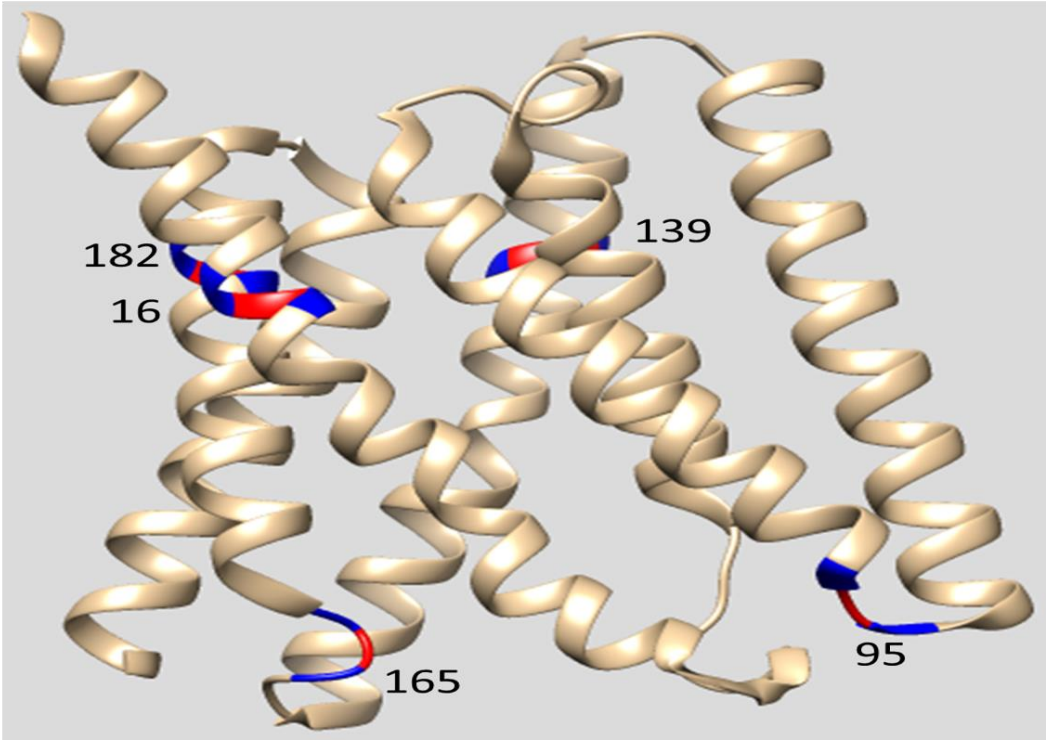


Figure 3.7 Alignment of *H. sapiens* SWEET with *O. sativa* SWEET1 shows little conservation of cysteine residues and may indicate cysteines are not required for tertiary structure

A) Clustal Omega (Sievers et al, 2011) pairwise sequence alignment of *H. sapiens* SWEET Gene ID: 55974 (top sequence) and *O. sativa* SWEET1 Gene ID: 104779417 (bottom sequence). Matching residues are indicated with a period (.) while sequence gaps are represented by a dash (-). B) PDB structure 5CTG showing *O. sativa* SWEET1 transporter made in the 3D modelling software Chimera (Pettersen et al 2004). Residues aligned with *H. sapiens* SWEET cysteines highlighted in red, surrounding residues are highlighted in blue, to demonstrate their relative locations.

```
        10         20         30         40         50         60
MEAGGFLDSLIYGACVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSY
..............V............................................

        70         80         90        100        110        120
GALKGDGILIVVNTVGAALQTLYILAYLHYCPRKRVVLLQTATLLGVLLLGYGYFWLLVP
.............................S..............................

       130        140        150        160        170        180
NPEARLQQLGLFCSVFTISMYLSPLADLAKVIQTKSTQCLSYPLTIATLLTSASWCLYGF
............V........................S................V....

       190        200        210        220
RLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT
.....................................
```

Figure 3.8 Alignment of HsSWEET and HsSWEETCysless highlights location of removed cysteine residues

Pairwise alignment using Clustal Omega (Sievers et al 2011) of *H. sapiens* SWEET Gene ID: 55974 (top sequence) and de novo generated HsSWEETCysless DNA sequence (bottom sequence). Matching residues are indicated with a period (.), mutated residues are represented by single letter amino acid code. Cysteines in transmembrane helices are replace with valine (V), those in cytosolic or extracellular regions are replaced with serine (S).

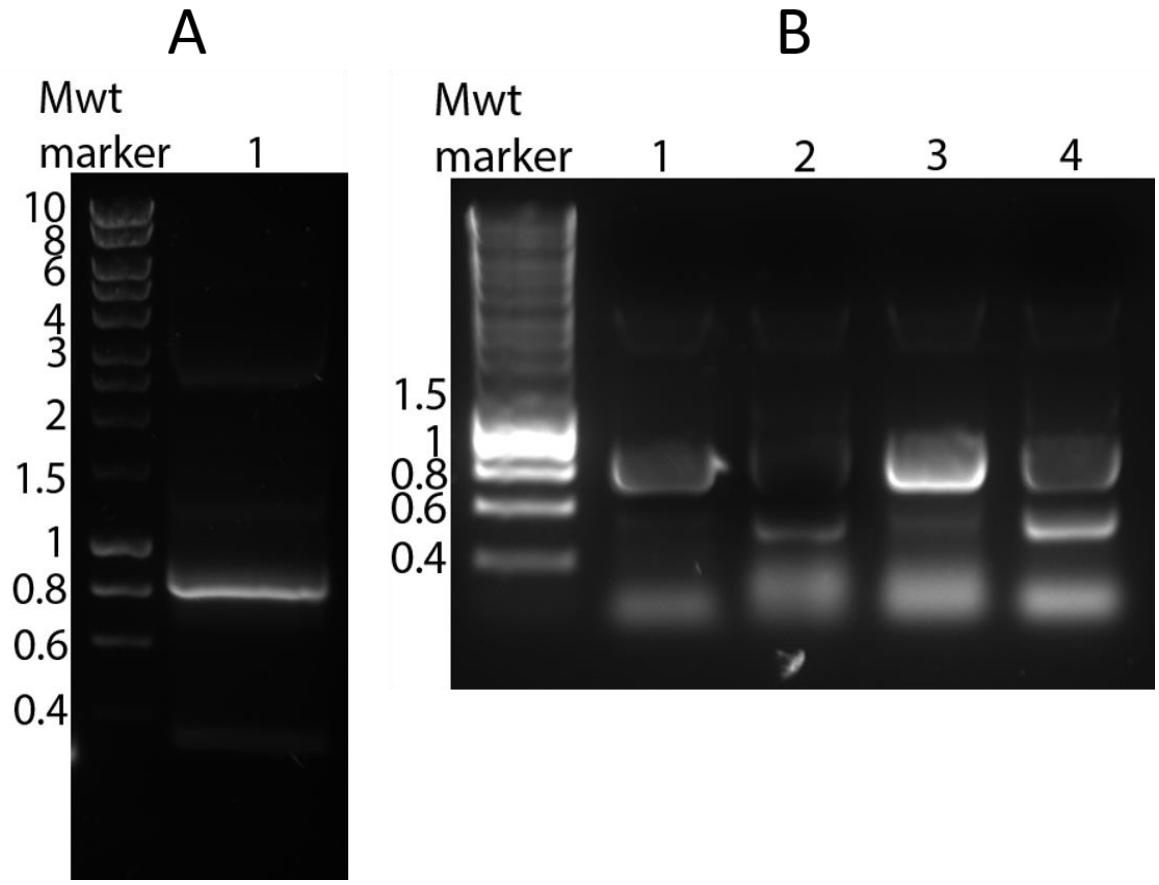


Figure 3.9 Amplification of HsSWEETcysless in preparation of cloning into prokaryotic expression vector and eukaryotic localisation vectors

Amplified from de novo HsSWEETCysless DNA with tags dependent on the intended vector: A) 1. LIC tags for pNIC28-BSA4, (690 bp). B) 1. HINDIII and BAMHI sites for pEGFP-N1-HsSWEETCysless, (675 bp); 2. HINDIII and BAMHI sites for pDDGFP-HsSWEETCysless c004, (675 bp); 3. HINDIII and BAMHI sites for pDDGFP-HsSWEETCysless c005, (675 bp); 5. HINDIII and BAMHI sites for pEGFP-C1-HsSWEETCysless, (675 bp). Molecular weight (Mwt) marker sizes labelled on the left of the gel (kbp).

3.2.2 Localisation of HsSWEETCysless in mammalian cells (collaboration with Dr David Tumbarello, University of Southampton)

Mammalian expression vectors pEGFP-C1, pEGFP-N1 (HsSWEETCysless N- or C-terminal HA-fusions, respectively) and pDDGFP (N- or C-terminal GFP-fusions) were used to investigate protein localisation in Hela and RPE cell lines. An example of the design for the cloning of HsSWEETCysless into the mammalian expression vectors is shown in Appendix A. Vectors were transfected into eukaryotic cell lines seeded on glass coverslips. The HsSWEETCysless intracellular localisation was examined in two different cell lines. This was done to ensure that the observed results were less likely to be cell-type specific and therefore more representative of where it might be located in the main tissues, as outlined above. This work was carried out in collaboration with Dr David Tumbarello (University of Southampton, Biological Sciences). Cells were fixed with formaldehyde then permeabilised. GFP, HA and marker proteins were immunostained with primary antibodies followed by secondary antibodies with associated fluorophores. The mounted cells were imaged using fluorescent microscopy, results are shown in Figures 3.10 and 3.11. Lysosomal-associated membrane protein 1 (LAMP1), is localised to the lysosome and is used as a marker for this internal cellular compartment. Hoechst is a fluorescent stain that binds strongly to A-T rich regions in DNA and is used to define the nucleus.

In the HeLa cells (Figure 3.10), there is strong, punctate staining around the nucleus (confirmed using Hoechst staining (Figure 3.10, top right picture)) which is the same for the GFP or HA tags on both the N- and C-termini of HsSWEETCysless (Figure 3.10, left column). This type of staining is a characteristic of Golgi localised proteins. In addition, in all four cases there is a weaker, diffuse staining throughout the cell that surrounds the nucleus and weakens, as it moves away toward the edge of the cell. The boundaries of the cells are clearly defined by the actin staining (Figure 3.10, bottom two pictures of the middle row). This is indicative of ER staining. Importantly, based on the LAMP1 staining, HsSWEETCysless is not present in the lysosomes. Hence, in HeLa cells, HsSWEETCysless is mainly localised in the Golgi but is also present, in lower levels, in the ER.

In the Retinal pigment epithelium (RPE) cells (Figure 3.11), punctate staining around the nucleus for the GFP-tagged version of HsSWEETCysless is seen. Again, it is the same for the GFP-tagged version, either on the N-terminus (Figure 3.11, bottom row) or the C-terminus (Figure 3.11, first two rows) of HsSWEETCysless. As before, none of the constructs colocalise with the lysosomal LAMP1 while there is weaker

staining around the nucleus that decreases in concentration the further away it goes from the nucleus. In these cells, however, there is additional spherical staining for the N-terminally-tagged version of HsSWEETCysless (Figure 3.11, top two rows). As these do not overlap with the LAMP1 marker (Figure 3.11, top two rows, right hand pictures), it was believed that they represent secretory vesicles. Hence, in RPE cells, HsSWEETCysless is mainly localised in the Golgi but is also present, in lower levels, in the ER and for the N-terminally GFP-tagged version. It is also present in secretory vesicles.

The K-Nearest Neighbour Classifier of PSORT II (Horton and Nakai 1997) finds known motifs within the amino acid sequence of a target protein to predict cellular localisation. PSORTII identified the ER as the most likely location of HsSWEET, a summary of these results is shown in Table 3.5. LocSigB identifies localisation sequences in protein primary structure (Negi et al 2015). LocSigB identified amino acids YGYFWLL at position 111-118 as a Golgi localisation signal. This would fall in the extramembrane loop between TMH 4 and TMH 5, predicted by TOPCONS to consist of amino acids 116 and 127 Figure 3.4. LocSigB also identified the lysosome localisation signal Yx{2}[VILFWCM] 5 times at positions 11-15, 59-63, 111-115, 185-189 and 215-219. The lysosome localisation site Yx{4}LL was identified at positions 111-118, some of these signals are marked on Figure 3.5 B.

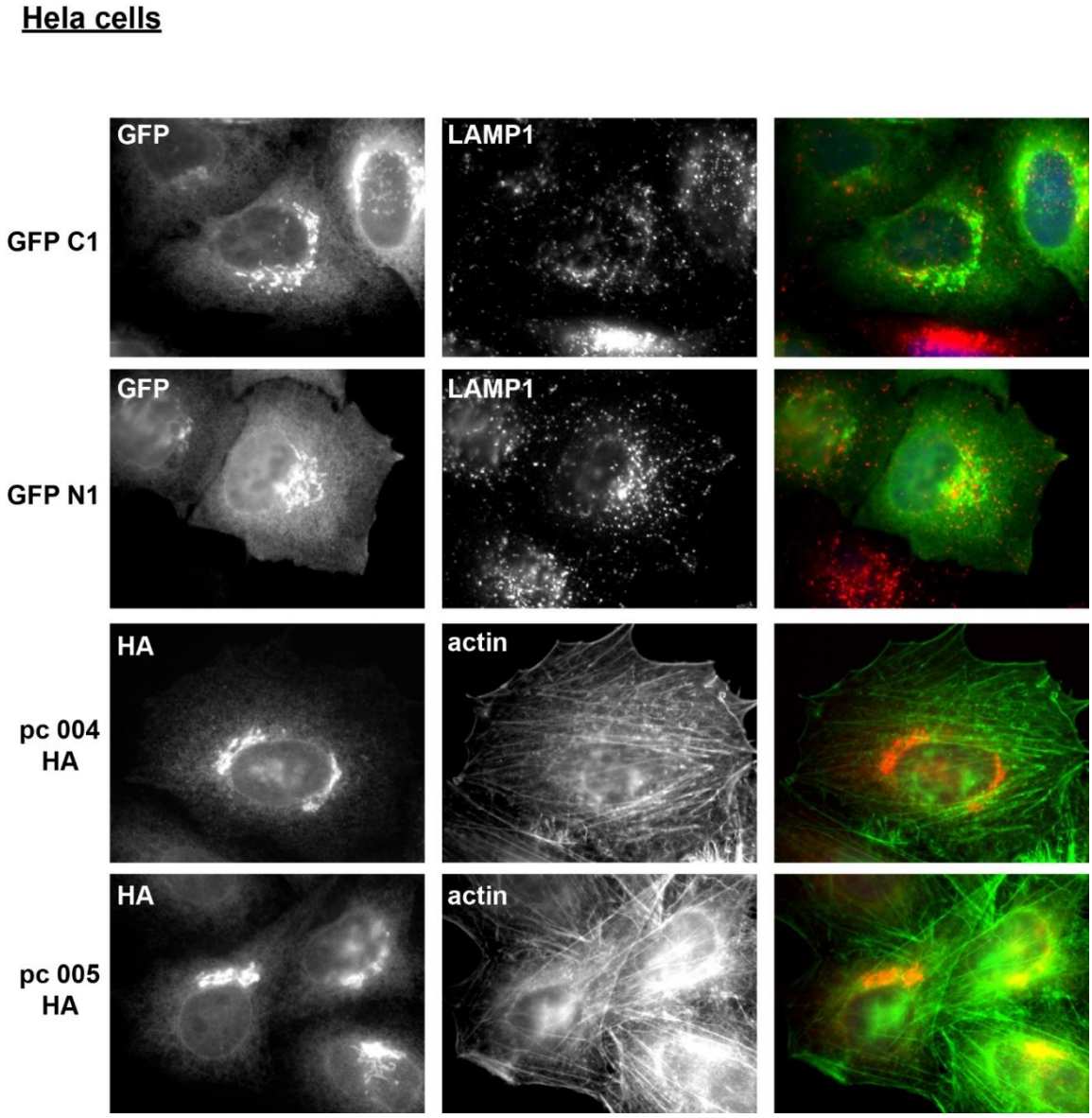


Figure 3.10 HsSWEETCysless localises to an intracellular membrane in HeLa cells

HeLa cells transfected with vectors expressing GFP- (GFP C1 and GFP N1) or HA-tagged (pc 004 HA and pc 005 HA) HsSWEETCysless. HsSWEETCysless tagged N-terminally (GFP C1 and pc 004 HA) and C-terminally tagged (GFP N1 and pc 005 HA). Left column – signal from GFP- or HA-tagged HsSWEETCysless. Middle column – signal from colocalisation marker (LAMP1 or actin, as noted on figure). Right column – false colour overlay of HsSWEET and localisation marker signals. In the top two rows, LAMP1 is coloured red and GFP-tagged HsSWEET is coloured green. In the bottom two rows, actin is coloured green and HA-tagged HsSWEET is coloured red. Nucleus is always blue using Hoechst staining. All proteins visualised using confocal microscopy of immunofluorescent antibodies.

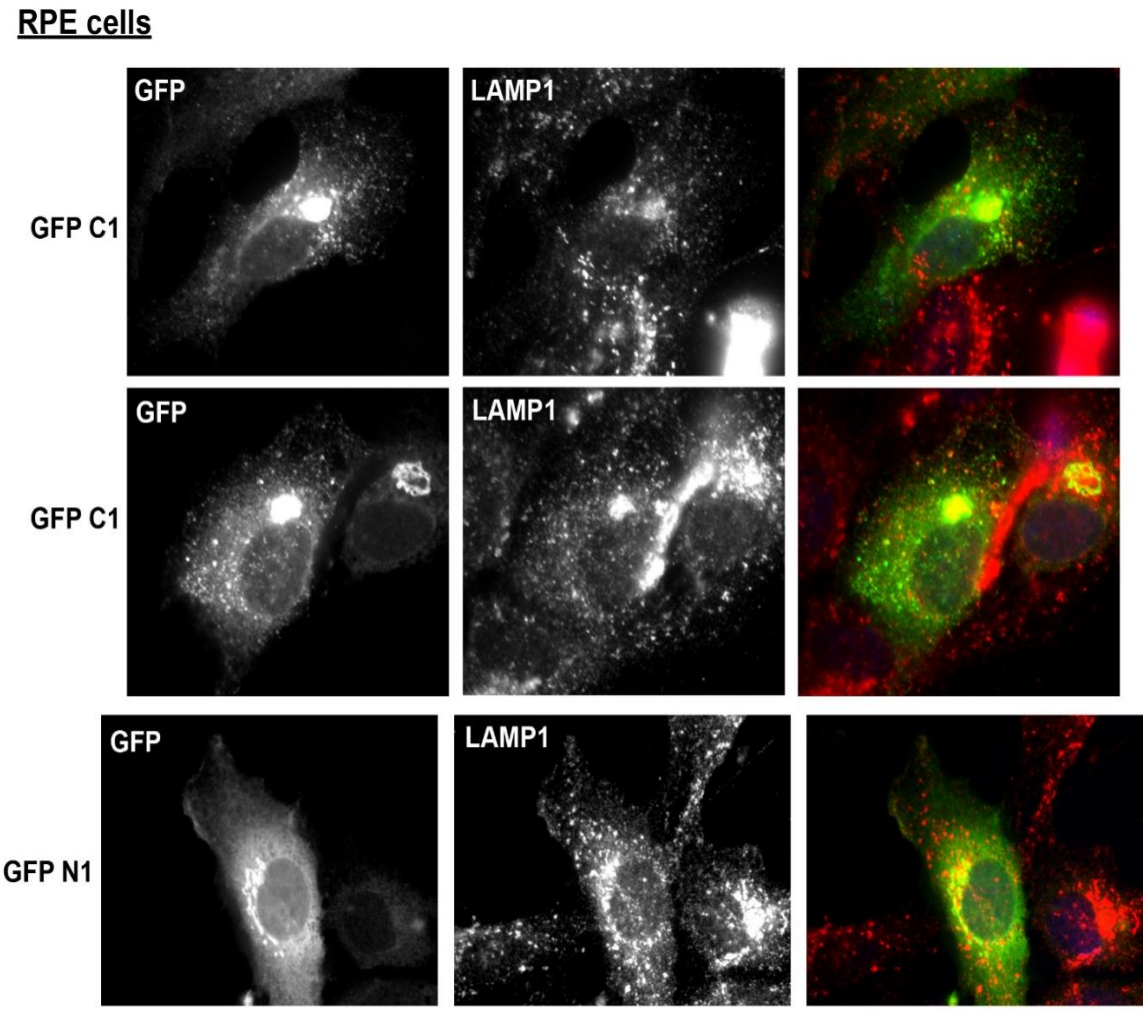


Figure 3.11 HsSWEETCysless localises to an intracellular membrane in RPE cells

RPE (Retinal pigment epithelium) cells transfected with vectors expressing GFP-tagged HsSWEETCysless. HsSWEETCysless tagged N-terminally (GFP C1) and C-terminally tagged (GFP N1). Left column – signal from GFP- HsSWEETCysless. Middle column – signal from colocalisation marker (LAMP1). Right column – false colour overlay of HsSWEET and localisation marker signals. LAMP1 is coloured red and GFP-tagged HsSWEET is coloured green. Nucleus is always blue using Hoechst staining. All proteins visualised using confocal microscopy of immunofluorescent antibodies.

Table 3.5 Results from PSORTII cellular localisation prediction programmes for *H. sapiens* SWEET

Scores are given as proportion of 100 %. ER – endoplasmic reticulum, PS – peroxisome, PM - plasma membrane.

	ER	Vacuole	PM	Mitochondria	Golgi	cytoplasm	PS	Nucleus
PSORT II (%)	33.3	22.2	22.2	11.1	11.1			

Overall, data from both cell lines and with two tags at either end of HsSWEETCysless agree that the main intracellular location for this transporter is in the Golgi. A lower but significant proportion is also present in the ER. Importantly, it was not observed in lysosomes or on the plasma membrane. The N-terminally GFP-tagged version did show localisation to secretory vesicles but as this was not seen elsewhere, it was assumed that the location of the tag was influencing the proteins final destination.

3.3 Over-Expression of HsSWEET using an *E. coli* expression System

The cloned prokaryotic vector pNIC-HsSWEETCysless enables expression of HsSWEET in an *E. coli* expression system. As the vector is a derivative of pET28a, BL21 (DE3) cells were selected for protein over-expression. To date, there is not a definitive over-expression protocol for all target proteins. Therefore, a number of variables must be explored. The next step after over-expression is purification from the membrane. Several variables must also be explored in order to optimise the total amount and purity of target protein.

Key variables to be explored for target protein over-expression include expression temperature, growth media, concentration of inducer (IPTG), total time of induction, additives such as cofactors and *E. coli* strains. In this case, the target protein is an integral membrane protein hence the total expression level is expected to be up to 100-fold lower as compared to soluble proteins. This makes it imperative that optimisation is performed for such targets to maximise the expression level.

The starting conditions for the initial over-expression are described in Materials and Methods Section 2.3.1 and 2.4.5. The first variable tested was temperature. Side by side expressions were set up with induction temperature varying between 15 and 27 °C. Table 3.6 lists some of the characteristics that were initially recorded. Graphing the data (Figure 3.12) shows that there is a linear relationship between temperature and cell density based on OD_{600} with the higher the temperature the lower the cell density. The higher cell density with lower temperatures also coincides with a change in colour of the cell pellets from a light brown colour to white (Table 3.7, Figure 3.13). The implication of these results is that more inclusion bodies are being produced at lower temperatures during HsSWEETCysless over-expression. To check this analysis purification by nickel ion metal affinity chromatography (Ni^{2+}-IMAC) was performed taking advantage of the presence of the hexahistidine tag present on the pNIC-HsSWEETCysless construct. A general Ni^{2+}-IMAC purification protocol is given in Material and Methods section 2.3.2. SDS-PAGE analysis for two purifications at 21 °C and 18 °C are shown in Figures 3.14 A and 3.14 B, respectively.

Table 3.6 Summary of *E. coli* cells post induction of HsSWEETCysless expression at different temperatures

After over-expression, OD_{600} was measure using a Nanodrop and solubilisation of membrane material carried out. Insoluble pellets were analysed by eye before being photographed.

Growth temperature (ºC)	**15**	**18**	**21**	**24**	**27**
Final OD_{600}	4.05	3.66	2.88	2.55	2.28
Membrane pellet colour	White	White	Light brown	Light brown	Light brown

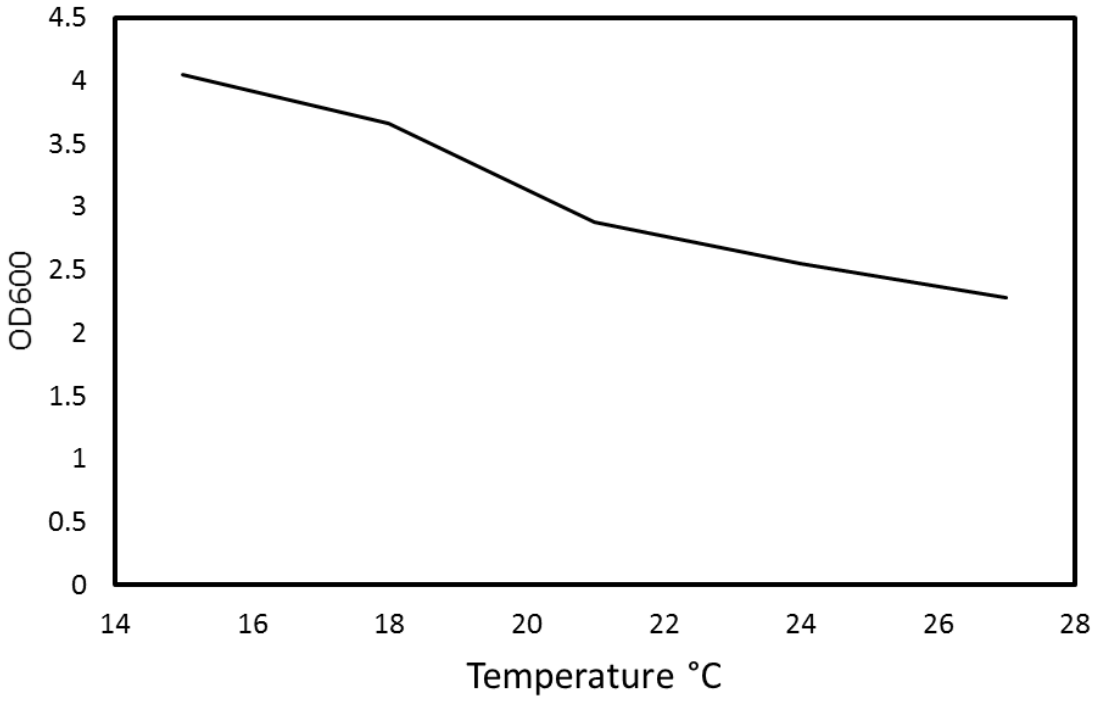


Figure 3.12 Effect of temperature on final OD_{600} of BL21 (DE3) cells expressing HsSWEETCysless

BL21 (DE3) cells transformed with pNIC28HsSWEETCysless expression vector. Expression induced with 0.2 mM IPTG overnight in LB broth. Final cell density (OD_{600}) measured in triplicate.

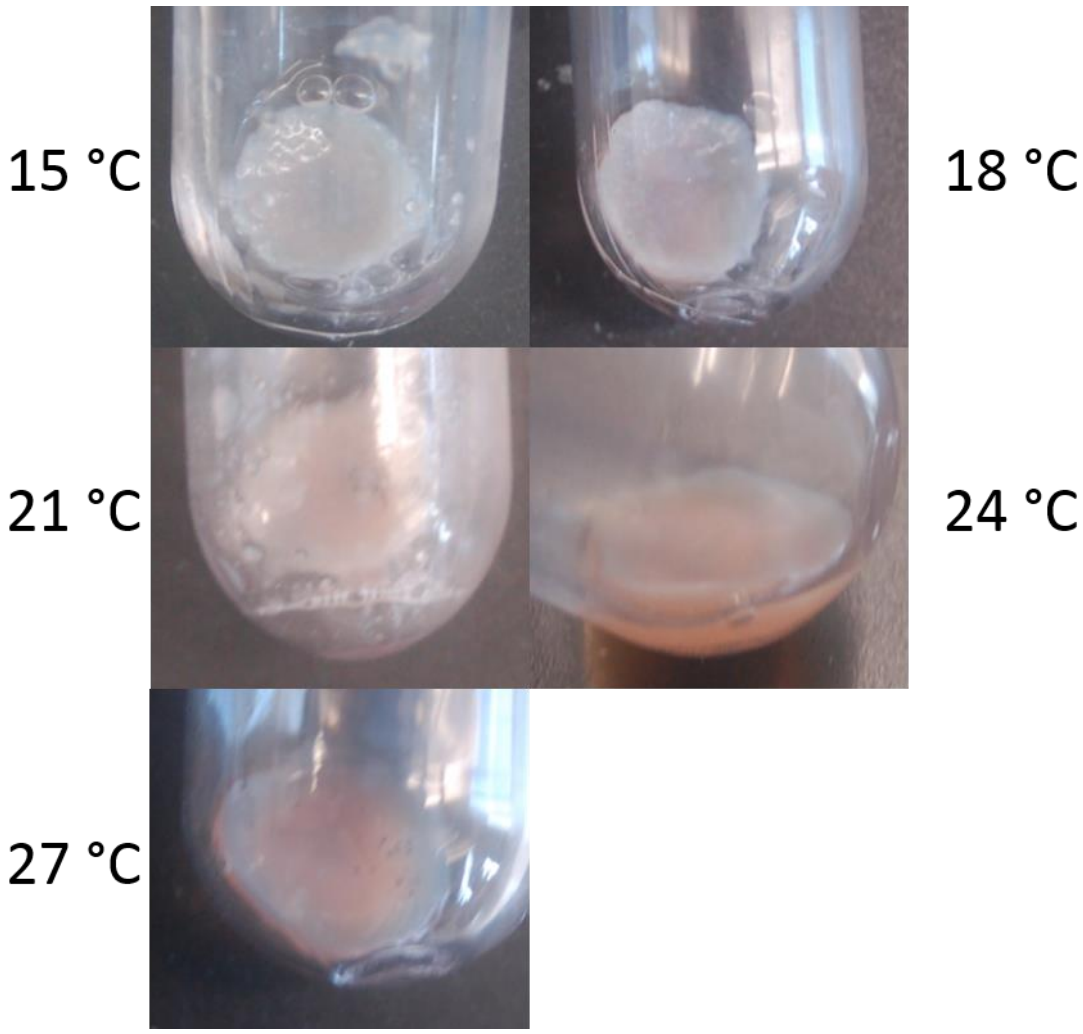


Figure 3.13 The insoluble fraction of membrane material in BL21 (DE3) cells after HsSWEETCysless expression, showing variation that is dependent on expression temperature

BL21 (DE3) cells expressing pNIC28HsSWEETCysless expression vector. Cells were incubated at different temperatures (°C), as labelled. Following overnight expression, cells were lysed and the insoluble fraction isolated with ultracentrifugation.

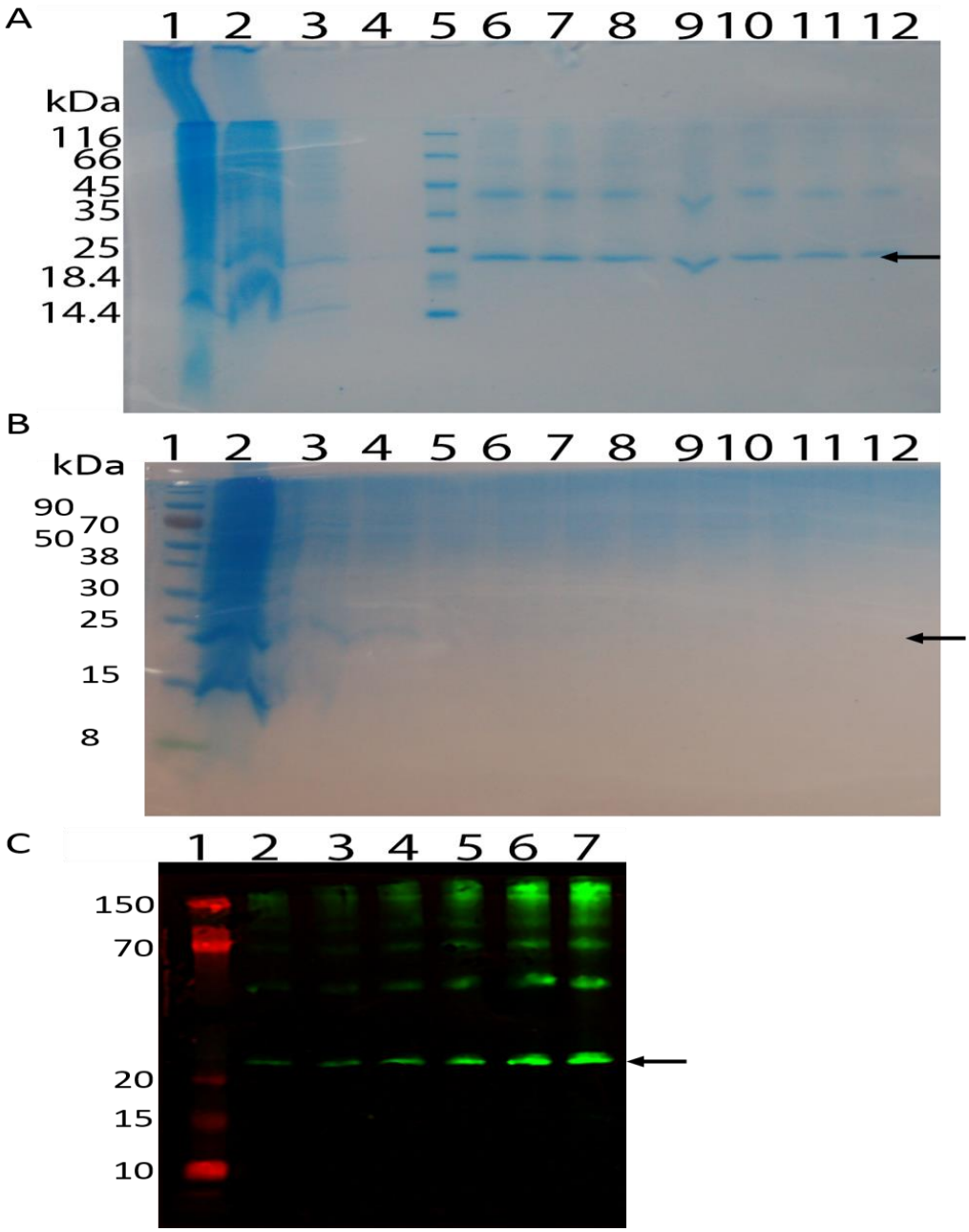


Figure 3.14 Immobilized metal ion affinity chromatography purification of HsSWEETCysless, from over-expression in BL21 (DE3) at 18 ˚C and 21 ˚C

Expression of HsSWEETCysless (24.98 kDa) in BL21 (DE3) followed by purification. Purified protein from the two expression temperatures (18 ˚C and 21 ˚C) was compared using SDS-PAGE. A) Coomassie-stained HsSWEETCysless protein from 21 ˚C expression; 1. cell pellet; 2. flow through; 3. Wash1; 4. Wash2; 5. protein ladder; 6 to 12. elution fractions. B) Coomassie-stained HsSWEETCysless protein from 18 ˚C expression; 1. protein ladder; 2. cell pellet; 3. flow through; 4. Wash1; 5. Wash2; 6 to 12. elution fractions. C) Histidine-tag western blot of HsSWEETCysless protein from 21 ˚C expression; 1. protein ladder; 2 to 7. elution fractions. Black arrow indicates the band expected to be HsSWEETCysless. Protein ladder sizes are labelled on the left of the gel.

The results indicate that the lower the expression temperature the lower the amount of HsSWEETCysless protein that is produced in the membrane. A more intense band at the expected level of HsSWEETCysless is present in the cell pellet for the 18 ˚C protein induction (Figure 3.14 A, lane 2) compared to 21 ˚C (Figure 3.14 B, lane 2). The elution fractions 6 to 13 in Figure 3.15 also demonstrates that very little soluble HsSWEETCysless is produced. This result, along with the previous data, indicate that lower temperatures produces more inclusion bodies and less soluble protein. Therefore, higher temperatures were used as part of the over-expression protocol. From this point on, a 21 ˚C induction temperature was used.

As part of the optimisation of the IMAC purification protocol, a selection of detergents were screened to determine which detergent was most efficient at extracting the protein from the membrane as well as maintaining its stability. Batch purification (Material and Methods Section 2.4.6) was used to assess the effectiveness of the available detergents at solubilising the membrane and releasing HsSWEETCysless. Detergents tested were 3-[N,N-dimethyl(3-palmitoylaminopropyl)ammonio]-propanesulfonate, Brij™ 010, Brij™ 35 and Brij™ 93. Purification was analysed using SDS-PAGE and protein concentration measurements using Nanodrop 280 nm absorbance reading. The results are shown in Table 3.7 and Figure 3.15. These indicated Brij-35 was the most effective detergent for solubilising protein.

Table 3.7 Fractions analysed from batch purification of over-expressed HsSWEETCysless using a variety of detergents

Protein was expressed and purified using Ni^{2+}-IMAC. During the solubilisation stage membrane fraction were separated and screened using four detergents each at a concentration of 40 mM in the lysate, 10 mM in the wash and 5 mM in the elution. Nanodrop 280 nm readings were used as an indicator of protein content. Further analysis was carried out using SDS-PAGE shown in Figure 3.14.

Sample number	**Detergent**	**Fraction**	**Protein concentration mg/mL**
1	n/a	cell lysate	
2	Z-3-14	solubilised material	
3	Z-3-14	wash	
4	Z-3-14	elution	0.2
5	Brij10	solubilised material	
6	Brij10	wash	
7	Brij10	elution	0.36
8	Brij35	solubilised material	
9	Brij35	wash	
10	Brij35	elution	1.17
11	Brij93	wash	
12	Brij93	elution	0.022

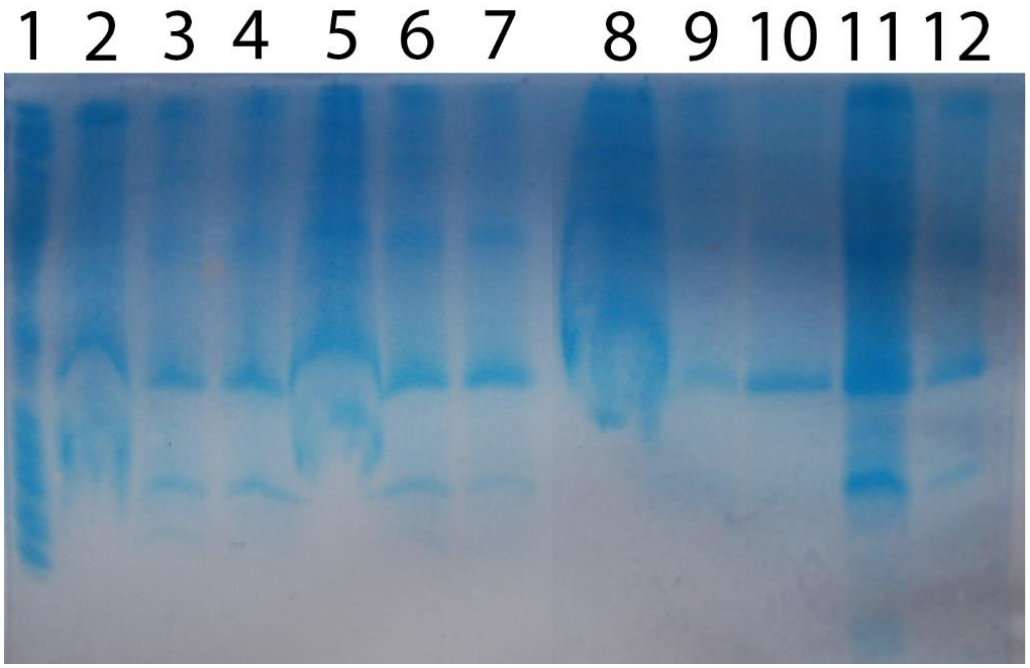


Figure 3.15 Coomassie blue-stained SDS-PAGE gel comparing purified HsSWEETCysless with different detergents

HsSWEETCysless was expressed in BL21 (DE3); insoluble material was isolated and solubilised in different detergents (Zwittergent 3-14, Brij10, Brij35, Brij 93) and purified in batches. 1. Insoluble material; 2. Zwittergent 3-14 solubilised material; 3. Zwittergent 3-14 wash; 4. Zwittergent 3-14 elution; 5. Brij10 solubilised material; 6. Brij10 wash; 7. Brij10 elution; 8. Brij35 solubilised material; 9. Brij35 wash; 10. Brij35 elution; 11. Brij93 wash; 12. Brij93 elution.

Gel filtration was used to analyse the homogeneity of the purified HsSWEET and investigate the presence of any oligomeric species. The HsSWEETCysless gel filtration chromatogram trace is shown in Figure 3.17. the main peak comes out at a volume of 11.7 mL. The void volume peak position must be subtracted to fit the standard curve giving a peak position of 3.3 mL. Inputting this value into the standard curve equation, shown in Figure 3.16, results in a log molecular weight value of 2.10998, or a molecular weight of 128.8 kDa. The average micelle weight of Brij-35 is 48 kDa (Acros Organics, UK). If the average micelle weight is subtracted from the peak position weight of 128.8 the calculated protein weight is 80.8. The protein weight, 80.8, divided by the weight of expressed HsSWEETCysless monomer, 27.8, gives an oligomeric state of 2.9. This indicates the expressed protein has adopted a trimeric state which agrees with the trimer organisation of the eukaryotic rice SWEET structure (Tao et al 2015).

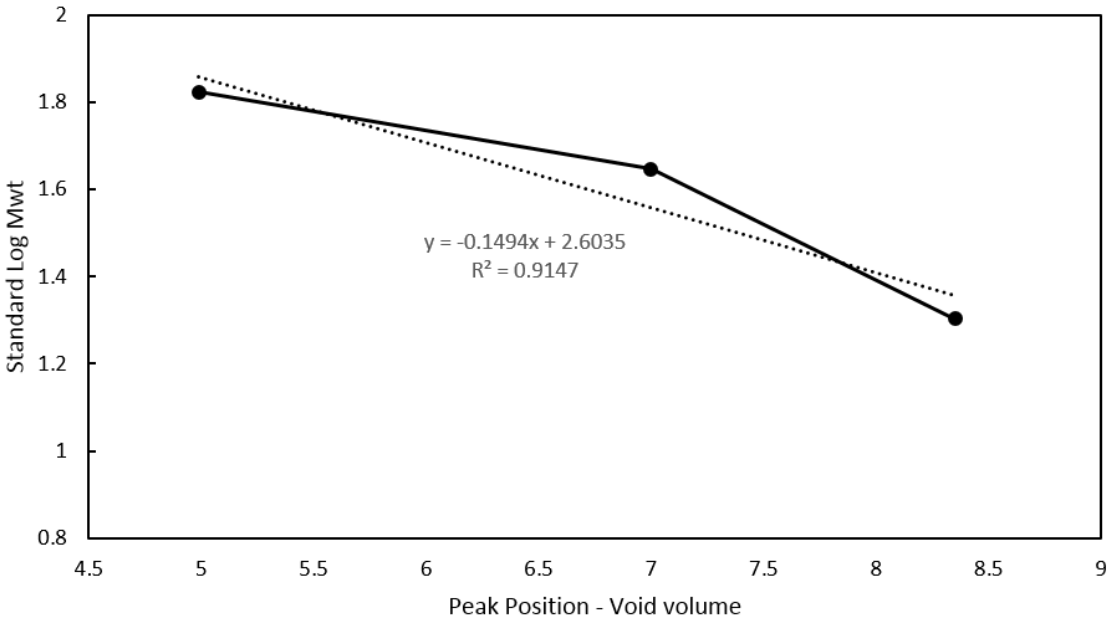


Figure 3.16 Curve of gel filtration using protein standards

Purified protein standards were run on a Superdex 5-150 gel filtration column at 0.3 mL/min in HsSWEETCysless purification buffer. The protein standards used were: Bovine Serum Albumin (66.45 kDa); chicken serum albumin (44.287 kDa); trypsin inhibitor (20.1 kDa). The void volume was calculated using blue dextran. Peak positions were plotted against log Mwt and used to generate a standard curve equation.

The over-expression and purification protocols developed for HsSWEETCysless successfully generated pure transporter of a consistent oligomeric state previously observed for another eukaryotic SWEET transporter (Tao et al 2015). This suggests that the replacement of the endogenous Cys residues did not have a drastic effect on the global structure of the transporter and that HsSWEETCysless would be appropriate for both structural and functional studies.

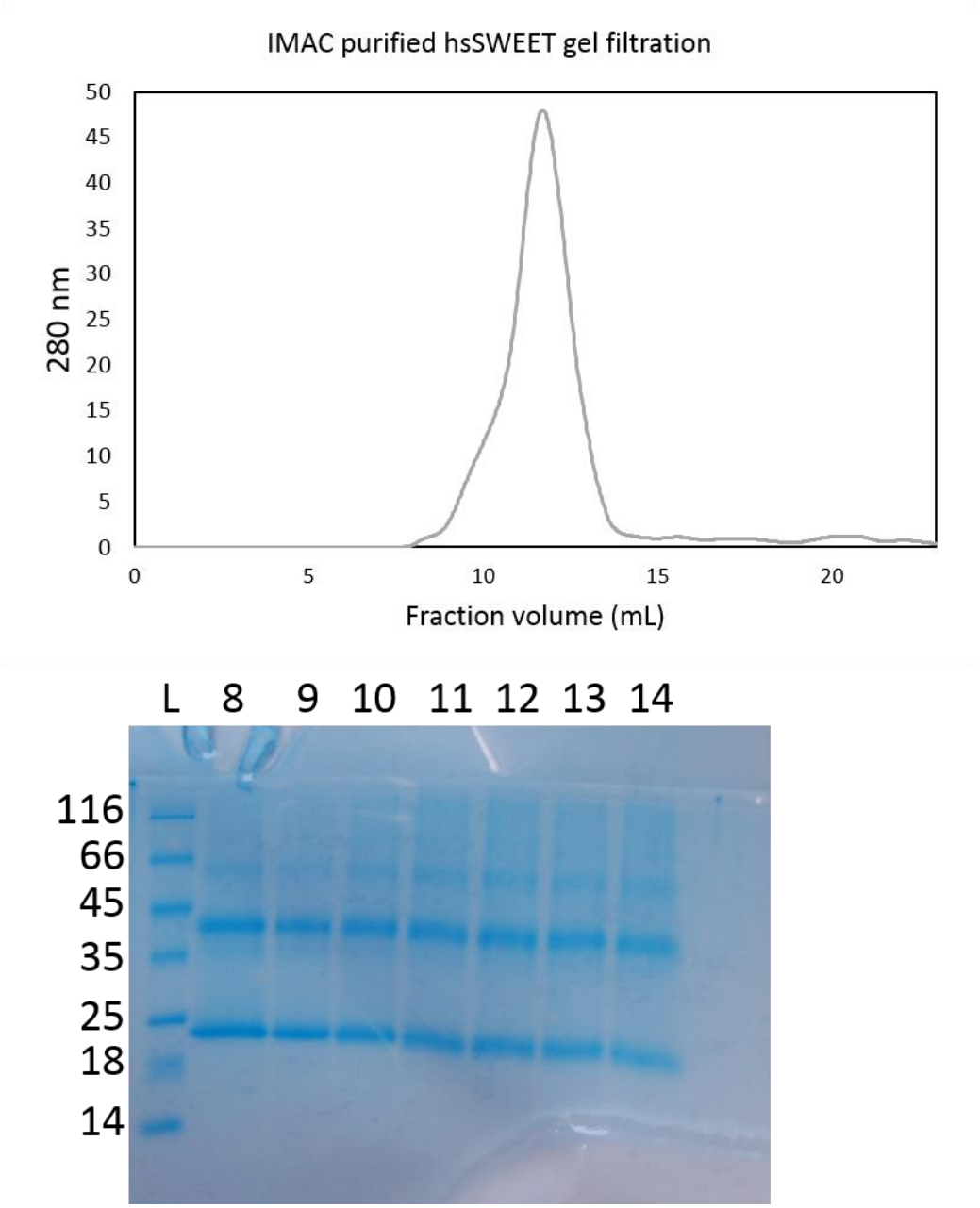


Figure 3.17 Gel filtration of purified HsSWEETCysless

HsSWEETCysless was over-expressed in BL21 (DE3) cells before immobilized with metal ion affinity chromatography (IMAC) purification and gel filtration. A) Gel filtration chromatogram of purified HsSWEET. B) SDS-PAGE analysis of gel purified protein. Gel filtration fractions labelled in figure, 8-14; L, protein ladder with sizes labelled to the left of the gel (kDa); HsSWEET expected size: 24.98 kDa.

3.3.1 Use of GFP-tag for expression optimisation and binding assays with HsSWEET

To better define the purification conditions and for use in a ligand binding assay the HsSWEETCysless transporter was re-cloned into the *E. coli* protein expression vector pET27b with a GFP tag linked to the C-terminus. An outline of the cloning for HsSWEETCysless is given in Appendix B. The same cloning method was also applied to the remaining PQ family members as well as the WT SLC50A1 gene (Table 3.8). Details of the ligand binding assay to be used with the purified GFP-tagged transporter can be found in Chapter 4. If successful, this method will allow screening of the protein against many potential substrates.

Table 3.8 Details of the GFP tagged PQ loop protein expression vectors

Targets generated using PCR and cloned into the pET27b or pGFPSTOP vectors. Restriction digest sites can be utilised for T4 ligase-based construct assembly. Modification Codon optimised, cysteines mutated (cocm).

Target	Species	Modification	Vector	Construct
sfGFP			pET27b	Start-pelB-SalI-sfGFP-HindIII-stop
SLC50A1	H. *sapiens*	None	pGFPSTOP	Start-pelB-EcoRI-SLC50A1-SalI-sfGFP-stop
HsSWEETCysless	H. *sapiens*	cocm	pGFPSTOP	Start-pelB-EcoRI-HsSWEETCysless-SalI-sfGFP-stop
PQLC1	H. *sapiens*	None	pGFPSTOP	Start-pelB-EcoRI-SLC50A1-SalI-sfGFP-stop
PQLC2	H. *sapiens*	None	pGFPSTOP	Start-pelB- EcoRI - SLC50A1-S SalI ALI-sfGFP-stop
PQLC3	H. *sapiens*	None	pGFPSTOP	Start-pelB- EcoRI - SLC50A1- SalI -sfGFP-stop
CTNSCysLess1	H. *sapiens*	cocm	pGFPSTOP	Start-pelB- EcoRI - SLC50A1- SalI -sfGFP-stop
CTNSCysLess2	H. *sapiens*	cocm	pGFPSTOP	Start-pelB- EcoRI - SLC50A1-SalI -sfGFP-stop
CTNSCysLess3	H. *sapiens*	cocm	pGFPSTOP	Start-pelB- EcoRI - SLC50A1-SalI-sfGFP-stop

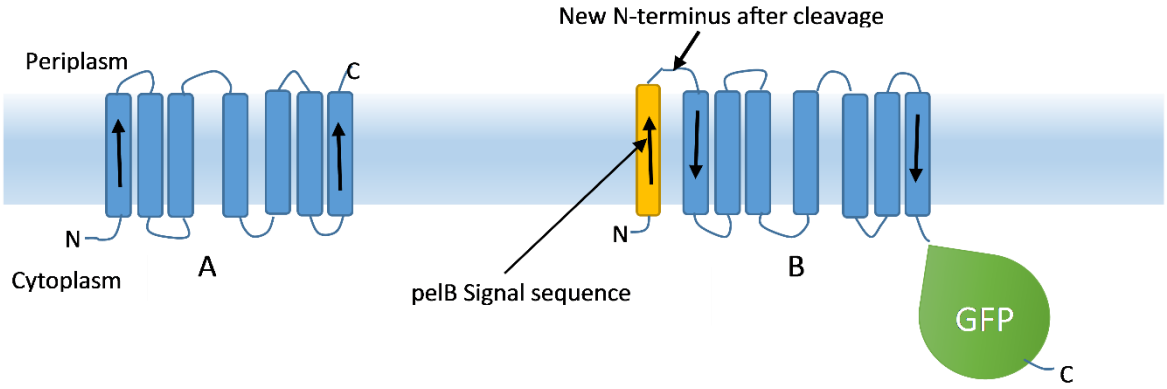


Figure 3.18 Topology model of the GFP-tagged PQ loop protein, expressed using the pGFPSTOP vector

A) wild type PQ loop protein, relative to membrane. The N-terminus is predicted to be cytoplasmic. B) pelB-PQ loop-GFP construct generated using the pGFPSTOP vector, which adds N- and C-terminal pelB signal sequence (orange) and GFP tag (green), respectively. The pelB signal sequence flips the protein's orientation with respect to the membrane, illustrated with thick black arrows.

The pelB sequence present in the pET27b vector will invert the topology of an integral membrane protein in the membrane, as shown in Figure 3.18. This is of interest with the HsSWEETCysless protein as it has already been shown that expression of non-aggregated protein leads to reduction in protein yields. If as predicted this protein does in fact work as a glucose exporter, its production as a functional transporter may be hindering the ability of *E. coli* to grow. In expressing the protein in the reverse direction this could be overcome, allowing for higher cell densities and therefore higher protein expression levels. The success of this technique may be used to as a template expression system for other membrane proteins that would normally hinder *E. coli* growth.

Using the vector H6msfGFP as a template, GFP was copied using PCR from primers with HindIII and SalI restriction sites, the components are shown in Figure 3.19 A. T4 ligase was then used to insert the gene into pET27b, creating pET27b-GFP which we called pGFPSTOP (Appendix B). DNA sequencing confirmed that the GFP sequence was as expected and in frame with the pelB signal sequence. HsSWEETCysless was cloned into pGFPSTOP. The remaining PQ loop transporter genes were produced with PCR but have not yet been inserted into the pGFPSTOP vector (Figure 3.19 C).

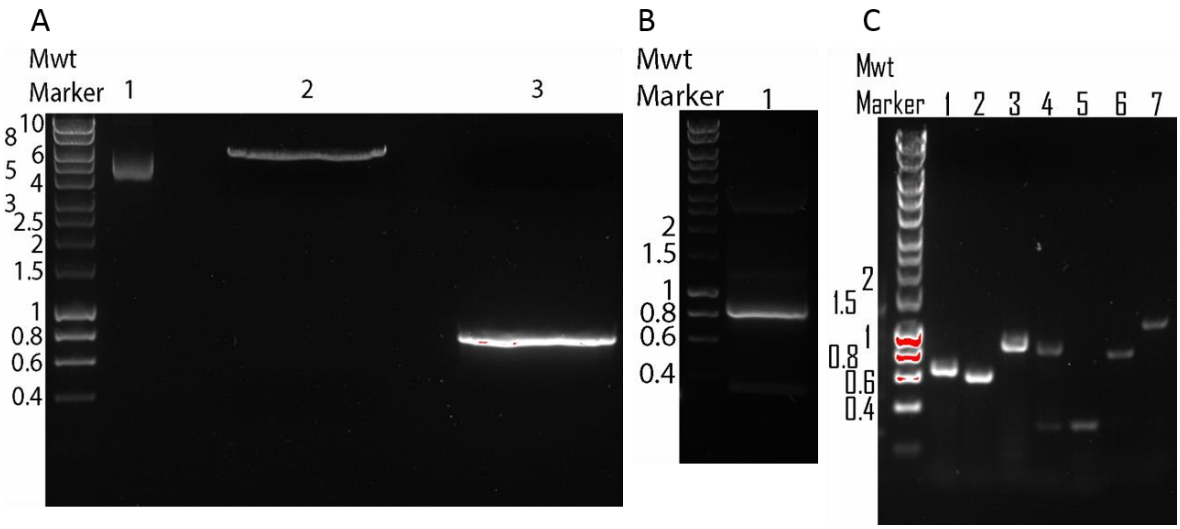


Figure 3.19 Generation of pGFPSTOP vector with PQ loop genes

A) Generation of pGFPSTOP vector; pET27b (lane 2) and GFP (lane 3; 751 bp) were digested with SalI and HindIII before T4 ligation, resulting in pGFPSTOP vector. Lane 1, undigested pET27b. B) Digestion of HsSWEETCysless (679 bp) with EcoRI and SalI before ligation into pGFPSTOP vector. C) Amplification of PQ loop proteins in preparation for digestion and ligation into pGFPSTOP vector. 1. SLC50A1 (751 bp); 2. PQLC3 (622 bp); 3. PQLC2 (895 bp); 4. PQLC1 (829 bp); 5. CTNSCysLess1 (318 bp); 6. CTNSCysLess2 (760 bp); 7. CTNSCysLess3 (1117bp). Molecular marker sizes (kbp) labelled on the left of the gel.

The signal sequence (SS)-HsSWEETCysless-GFP construct was then over-expressed in BL21 (DE3) cells following the previously used expression protocol (Outlined in Section 2.4.6). Over-expression in *E. coli* of SS-HsSWEETCysless-GFP will result in the removal of the signal sequence to generate HsSWEETCysless-GFP. As the GFP is on the C-terminus, it naturally means that the membrane protein section must be made first before GFP. Hence, if, after protein over-expression, GFP fluorescence is observed, it can be assumed that the membrane protein has been made, folded properly and inserted into the membrane. This has been verified in Dr D Doyle's lab by over-expressing several C-terminally GFP-tagged membrane proteins, purifying the membranes and then testing for GFP fluorescence in the membrane and soluble fractions. In all cases, GFP fluorescence was observed only in the membrane fraction (D Doyle personal communications).

As part of the optimisation tests, two medias (LB and TB) were used and three temperatures were tested in the SS-HsSWEETCysless protein over-expression trials (Figure 3.20 A and B). As explained, the total amount of over-expressed target is related to the GFP relative fluorescence units (RFU), making it easier and quicker

than partial purification and examining for over-expressed protein on an SDS-PAGE gel.

For both media, 30 °C trial produced the least amount of protein (Figure 3.20 A). For the two lower temperatures, HsSWEETCysless-GFP levels were higher in TB than LB and the best expression was at 18 °C. For the 18 °C and 23 °C temperature trials, there is a close correlation between total amount of protein produced and cell density (Figure 3.20 B and Table 3.9). In both cases, the higher the cell density the more protein that is produced. This correlation does not exist for the 30 °C trial, again confirming that this temperature is the worst for this trial.

For the H6-TEV-HsSWEETCysless expression trial there was a linear relationship between temperature and cell density (Figure 3.20 B) which does not exist here. This does indicate that flipping of the transporter has an effect on the expression level of the exogenous target protein. It would be interesting to know if the transporter is active in the membrane which may have an influence on the growth properties of the cells. If we assume that flipping the transporter does not have an effect on its ability to transport, then the implications of our data is that there is some directionality to the movement of substrates as the growth and protein production behaves different in both cases. Even though we make many assumptions here the difference in behaviour does imply that the transporter is potentially functional in the membrane.

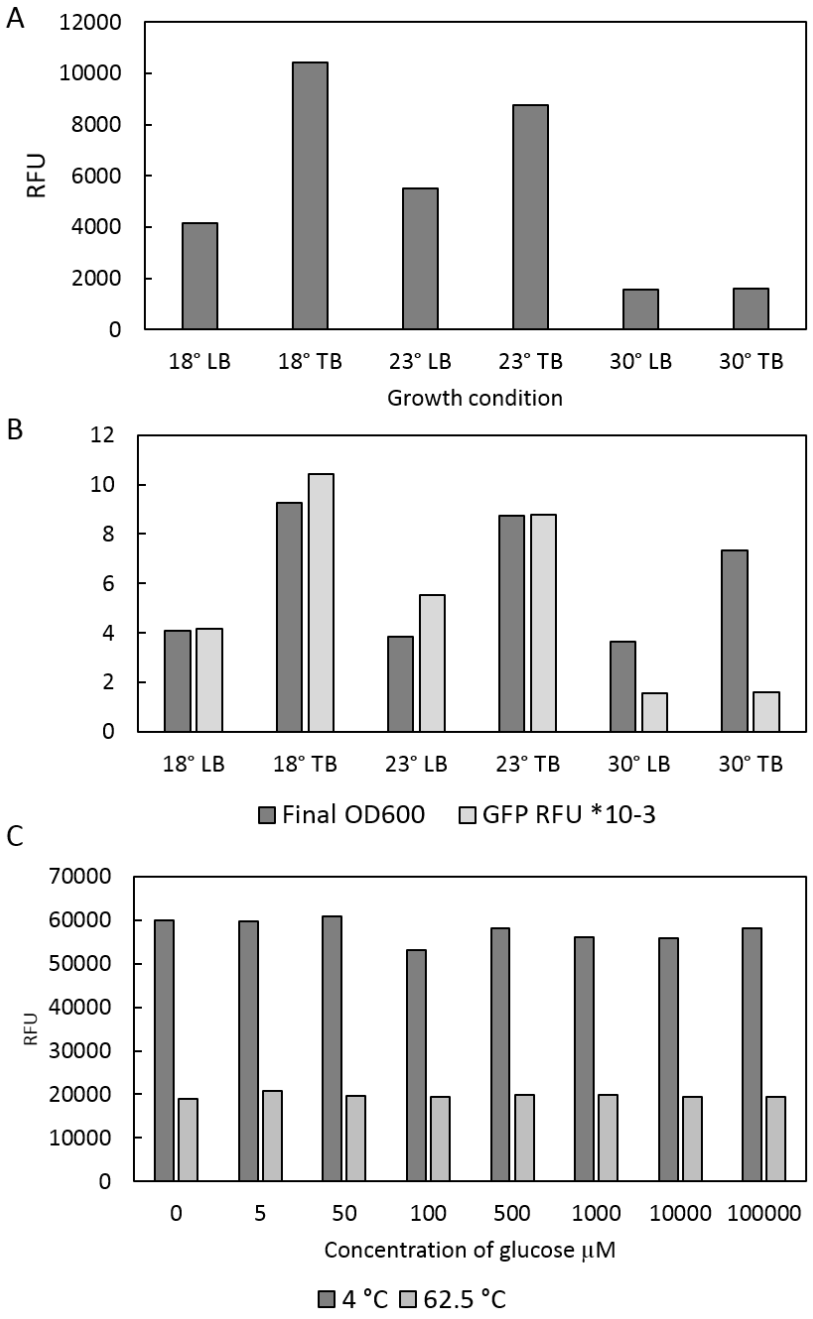


Figure 3.20 Analysis of expressed pet27B HsSWEETCysless-GFP

HsSWEETCysless-GFP was expressed in BL21 (DE3) cells using IPTG induction, at different temperatures (18, 23, 30 ˚C) and using different growth media (LB or TB). A) Comparison of whole cell fluorescence (RFU). B) Relative fluorescence of HsSWEETCysless-GFP (x10 $^{-3}$; pale bar) compared with final cell density (OD_{600}; dark bar) after protein expression. The GFP RFU readings have been uniformly scaled to OD_{600} scale for ease of comparison. C) HsSWEETCysless-GFP T_m stability test. Following expression cells were lysed and the membrane fraction was isolated. The membrane fraction was exposed to 4 ^{o}C (dark bar) or 62 ^{o}C (pale bar); the protein remaining stably in solution was measured using GFP as a reporter. Different concentrations of glucose (µM) were used to investigate substrate binding. For A, B and C, all data points measured in triplicate.

Table 3.9 The final OD_{600} of BL21 (DE3) cells after SS-HsSWEETCysless-GFP over-expression.

Growth temperature	**18 ˚C LB**	**18 ˚C TB**	**23 ˚C LB**	**23 ˚C TB**	**30 ˚C LB**	**30 ˚C TB**
Final OD_{600}	4.08	9.25	3.85	8.73	3.65	7.35

Post solubilisation the effect of glucose on the thermostability of HsSWEETCysless-GFP was tested. For a detailed explanation of the methodology and theory of this test see Chapter 4. In essence, a protein bound to its substrate adopts a different conformation to the unbound protein. The presence of specific substrate has been shown, in some cases, to measurably affect the stability of the protein at different temperatures (Franken et al 2015). The stability of HsSWEETCysless-GFP was measured at 62.5 °C, following heating the protein was centrifuged to remove any temperature denatured protein. The effect of different concentrations of glucose on the HsSWEETCysless-GFP was measured to attempt to identify a binding event, shown in Figure 3.20 C. This shows there is no increase or decrease in the thermostability of HsSWEETCysless-GFP when comparing control to 100 mM glucose; indicating no binding event measurable by this technique.

3.4 Discussion

This chapter aimed to produce purified HsSWEETCysless transporter suitable for functional and structural studies. This work is the starting point for a better characterisation of the human SWEET transporter which lacks detail in the literature, as there is only one publication that positively identifies it as a glucose transporter (Chen et al 2010). This lack of information regarding HsSWEET means that no physiological role has yet been assigned even though it is an important protein based on its ubiquitous expression in *H. sapiens* tissues (Chen et al 2010). Information is available regarding homologues in other organisms. Expression of the *C. hircus* homologue in goat mammary gland epithelial cells leads to increased efflux of glucose, suggesting ChSWEET1 is involve in glucose efflux (Zhu et al 2015). The 17 homologues in *A. thaliana* are thought to have a variety of physiological functions, relating to differing tissue expression and substrate transport (Chen et al 2010).

Compared to the SWEET proteins, GLUT and SGLT transporter characterisation is more advanced and, as such, their significance in human disease is better understood. This has informed treatment design leading to drugs tackling diabetes (Vallianou et al 2016), cancers (Jiang et al 2016), cystinosis (Gahl et al 2007) and cancer markers (Scafoglio et al 2015). Establishing an effective HsSWEET expression system would be an important step in investigating its function, structure and physiological role.

Modelling software was initially used to glean any structural information available from the HsSWEET sequence to aid biochemical processing. The TOPCONS analysis in Figure 3.4 shows a summary of five separate topology modelling programmes. A selection of TMH prediction programmes were in agreement with OCTUPUS (Viklund and Elofsson 2008), POLYPHOBIUS (Käll et al 2005), SCAMPI (Bernsel et al 2008) and SPOCTOPUS (Skwark and Elofsson 2008) all predicting seven transmembrane regions. This topology also matches that of the known OsSWEET structure (Tao et al 2015). The protein sequence alignments, structural models and examining the known SWEET PDB structures allowed us to replace the five endogenous Cys residues with less reactive amino acids. Overall, this was a successful approach as the Cys deletion mutant form was able to be over-expressed and purified as a trimer based on gel filtration analysis (Figure 3.16). Adjusting a protein's codons to fit that of the expression system can be key to successful expression (Šnajder et al 2015). This appears to be the case for the codon

optimised HsSWEETCysless, however, for a proper comparison quantified expression levels for both HsSWEETCysless and WT HsSWEET under the same conditions are required.

3.4.1 HsSWEET targets an intracellular membrane

Chen et al (2010) demonstrated Golgi-targeting of HsSWEET using the C-terminal V5 tag, and a Golgin97 marker (Figure 3.21 J-L). It has been shown that tags can disrupt native localisation of proteins (Kalatzis et al 2001). To reduce the likelihood of artefacts created by tagging, localisation studies should utilise both the C- and N-terminus. As HsSWEET has 7 TMHs this will result in the tags expression on both sides of the membrane. Good practise also dictates the use of multiple tags to avoid any unforeseen interactions with the cell. In this way the effect of a specific tag on protein trafficking will be highlighted.

As part of the characterisation of HsSWEET, its cellular localisation was investigated by expression of HsSWEET-GFP, GFP-HsSWEET, HsSWEET-HA and HA-HsSWEET tagged versions. It is clear from Figure 3.8 and 3.9 that HsSWEET displays punctate expression and targets an intracellular membrane. The use of localisation markers indicates that HsSWEET does not target actin or lysosomal compartments, nor does it target the nuclear envelope. In the absence of any available markers for additional intracellular membranes, a hypothesis can be drawn using comparison with the confocal work of others. This is shown in Figure 3.21. Lysosomal proteins are transported from the Golgi, lysosomal markers were used here to demonstrate this is not the final location of HsSWEET.

In mammalian cell lines, Golgi apparatus appear as punctate spots, asymmetrically located around the cellular nucleus (example shown Figure 3.21A-D). This matches the strongest staining seen by HsSWEET in the expression systems set up, as shown in Figure 3.8 and 3.9. These are more similar to Golgi markers than those of the ER (Figure 3.21E) or endosome markers (Figure 3.21F). There appears to be a disparity in subcellular localisation depending on the localisation of the GFP tag. Although N-terminally tagged HsSWEET (GFP-HsSWEET, Figure 3.8 C1 and Figure 3.9 C1) appears mostly punctate, C-terminally tagged protein appears a bit more diffused across the cell (HsSWEET-GFP, Figure 3.8 N1 and 3.9 N1).

The one drawback for all of these localisation experiments is that the target protein has been over-expressed to levels that are not seen in normal growth. It is possible that this causes misdirection of transporter so that it is seen in additional

compartments such as the ER. A better approach would be to use CRISPR technology to replace the endogenous HsSWEET gene with a GFP-tagged version and repeat the localisation experiments. A simpler experiment would be to remove the potential localisation signal identified by bioinformatics analysis such as in Figure 3.5 B before repeating the experiments. Alternatively, disruption of the Golgi apparatus with the application of the drug brefeldin A (Bershadsky and Futerman 1994) could positively identify this intracellular region as the main target location.

A major unanswered question remains from this data if Golgi-localisation of HsSWEET is correct. Why do we have an intracellular glucose transporter located in the Golgi apparatus? For obvious reasons, most glucose transporters are located on the plasma membrane so that glucose can be either stored or released into the blood stream depending on the metabolic state of the cell and the total energy levels of the organism.

Golgi-localisation of sugar transporters has been reported before, consistent with a role of Golgi-localised sugar transporters facilitating the directed export of sugars from the cell. MmGLUT1 targets the Golgi of mammary glands of lactating mice, declining during weaning of young (Nemeth et al 2000). MmSWEET is also upregulated during lactation, which fits in with its known ability to export sugars in the formation of milk (Chen et al 2010).

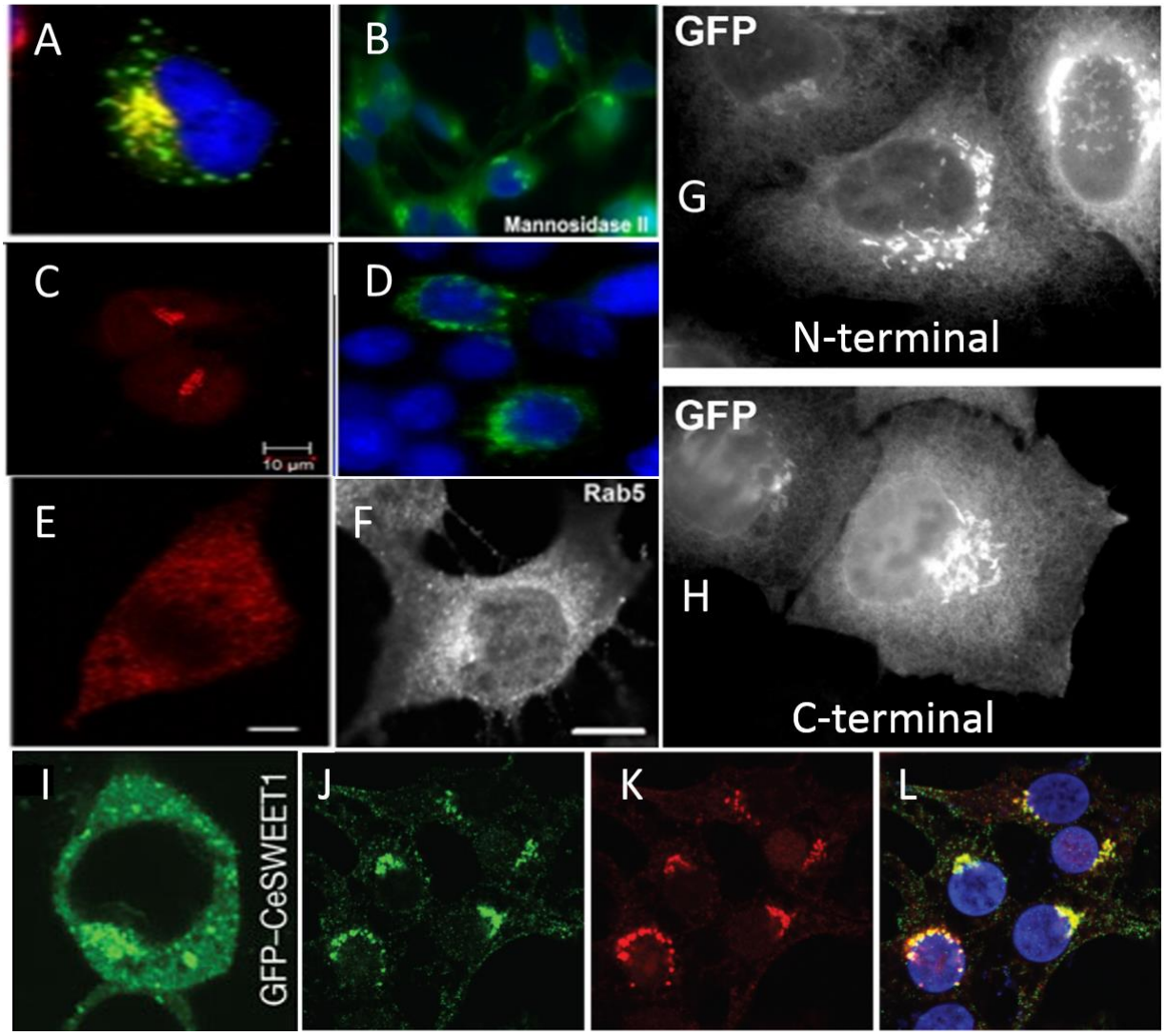


Figure 3.21 HsSWEET localisation from this study and other studies, compared with localisation markers in mammalian cell lines

A-D) Golgi markers tagged with fluorescent proteins demonstrating the shape of the Golgi in mammalian cell lines; A) Golgin in yellow (Liu et al 2007); B) mannosidase II in green (Johswich et al 2009); C) GM130 in red (Konitsiotis et al 2014); D) RFP-N-acetylgalactosaminyltransferase in green (Wang et al 2013); E) ER marker KDEL-RFP (Gorleku et al 2011); F) endosome marker Rab5 (Bu et al 2009). G-H) HsSWEETCysless localisation from this thesis. C-terminally and N-terminally GFP-tagged HsSWEETCysless in HeLa cells demonstrating intracellular membrane localisation. I-L from Chen et al 2010. I) GFP-CeSWEET, no co-localisation data available; J-L) Colocalisation of HsSWEET-GFP (J, green signal) with Golgi-marker Golgin97 (K, red signal); overlap shown in L; nucleus stained with Hoechst 33258 blue signal (Chen et al 2010).

3.4.2 The expression HsSWEETCysless in BL21 (DE3) cells

HsSWEETCysless was effectively produced in the BL21 (DE3) expression system at 21 ºC. At expression temperatures of 18 ºC or lower, the produced protein aggregates and is only present in inclusion bodies. As the native environment of HsSWEET is 37 ºC, it is possible that at lower temperatures it does not fold correctly.

Expression of HsSWEETCysless negatively impacts the BL21 (DE3) cell growth, as demonstrated by the reduced OD_{600}. This effect is also seen when expressing the SET bacterial sugar exporters in Chapter 4.2. When overexpressed in *E. coli*, sugar transporters may reduce the growth potential, simply by lowering internal sugar availability. If this is indeed the case HsSWEETCysless may be exporting glucose.

Gel filtration of purified HsSWEET produced a triplex, homomeric protein matching the oligomeric state of OsSWEET. Each HsSWEET monomer contains all the comparative sections of a complete SemiSWEET channel, suggesting transport should be possible without oligomerisation. Coexpression of native protein with non-functional AtSWEET1 in yeast appears to hinder the functionality of native protein (Xuan et al 2013). This could be a result of allosteric regulation between monomers, so incorporation of a non-functional protein into the oligomer hinders adjacent functional proteins. Enzyme kinetics work on *Haemophilus influenzae* β-carbonic anhydrase indicates 2 amino acid mutations are enough to uncouple an allosteric interaction (Hoffman et al 2015). This may mean not a great deal of change is required in the oligomer to go from allosterically connected to independent, or vice versa. Alternatively, it is possible that while each monomer can form a functional transporter, the intermolecular interactions of the oligomer stabilises protein substrate–interactions, leading to improved efficiency (Fatmi et al 2010).

A pelB-HsSWEET-GFP fusion protein was used to further investigate HsSWEET characteristics. Glucose binding did not increase stability of the fusion protein to a detectable degree. As seen in Chapter 4, the binding of ligands may not cause a detectable, stabilising shift in the protein tertiary structure. It is possible that the results are real or else the sensitivity of the detection method is within the noise level making this approach unsuitable for membrane protein ligand binding studies. The use of the Monolith thermophoresis (see Chapter 4) could be an alternative method to investigate HsSWEET ligand interactions. However, recloning

of a hexahistidine-tagged version is required so that the protein can be isolated before analysis.

Flipping HsSWEET using the pelB sequence appears to rectify the growth limitations triggered by induction of HsSWEET. This indicates HsSWEET is no longer acting as a glucose exporter. This may suggest the protein is not a passive diffuser, as its actions are affected by its directionality. Flipping protein directions with pelB could provide a useful technique for expressing other transport proteins shown to adversely affect cell growth when induced such as for the setC *E. coli* MFS sugar exporters (D Doyle, personal communications).

GFP tagging has improved our understanding of the HsSWEET protein. It was used as a reporter tag for microscopy, demonstrating cellular localisation was internal. GFP tagged HsSWEET in a prokaryotic expression system was able to simply measure the protein yield at different expression temperatures, highlighting its potential for optimisation. The GFP tagged protein produced can now be used with the assay developed in Chapter 4 to investigate substrate binding.

Chapter 4 Development of a Non-specific, GFP-based, Ligand Binding Assay for Integral Membrane Proteins

4.1 The MFS family are key proteins in the development of antimicrobial resistance, a persistent threat to human health.

As discussed in Chapter 3, exchange between a cell and its environment is crucial for normal function. For all of life, cells have to uptake and expel a range of compounds including nutrients, metabolites and xenobiotic substances. To maintain a cell's required equilibrium, much of this transport must be carried out actively against concentration gradients, using primary active or secondary transport. This Chapter is chiefly concerned with the secondary transporters of the Major Facilitator Superfamily (MFS).

MFS transporters are ubiquitous in both the eukaryotic and the prokaryotic domains. An example of a eukaryotic MFS transporter is discussed in Chapter 3, that of the human GLUT transporter. As glucose is a fundamental energy source for many forms of life these transporters are by necessity key players in this process. In humans, generally these essential transporters do not mutate significantly. Some mutations are known and result in diseases such as the mental retardation and epilepsy phenotypes associated with mutant forms of *GLUT1* (Gras et al 2014).

The importance of MFS transporters in prokaryotes has been demonstrated by genomic analysis of prokaryotic transport proteins. This data illustrated that the MFS family members made up to 22 % of all transporters depending on the species (Paulsen et al 1998). One of the best known and best studied MFS transporter is the lactose permease (lacY) (Teather et al 1978; Dornmair 1988; Abramson et al 2003; Chaptal et al 2011). LacY was shown to facilitate the uptake of lactose, using protons as a co-transported substrate (Viitanen et al 1984).

MFS transporters are not only used in nutrient uptake but cover a variety of other important cellular roles. This is reflected by the variation in transport direction, substrate diversity and co-transported molecules (Table 4.1). Some members of the

MFS facilitate cellular efflux of molecules (Edgar and Bibi 1997) while others facilitate passive diffusion into the cell (Kasahara and Hinkle 1977).

Table 4.1 List of MFS transporter whose structures have been determined.

Substrate information is included, including the primary substrate and the direction of movement with respect to the plasma membrane.

Protein - UniProt	Substrate	Substrate direction	Species	Structural conformation	PDD reference
xylE - P0AGF4	xylose transporter	Uptake	*E. coli*	inward open (Quistgaard et al 2013)	4JA4
				substrate bound (Quistgaard et al 2013)	4GBY
emrD - P31442	multiple drugs	Efflux	*E. coli*	inward occluded (Yin et al 2006)	2GFP
mdfA - P0AEY8	multiple drugs	Efflux	*E. coli*	Substrate bound (Heng et al 2015)	4ZOW
yajR - P77726	multiple drugs	Efflux	*E. coli*	outward open (Jiang et al 2013)	3WDO
glpT - P08194	glycerol-3-phosphate	Uptake	*E. coli*	inward open (Huang et al 2003)	1PW4
lacY - P02920	lactose	Uptake	*E. coli*	inward open (Guan et al 2007)	2V8N
				outward occluded (Kumar et al 2015)	4ZYR
fucP - P11551	fucose	Uptake	*E. coli*	outward open (Dang et al 2010)	3O7Q
nark - P10903	nitrate	NO_3 uptake/NO_2 efflux	*E. coli*	inward open/inward bound/ bound occluded (Fukuda et al 2015)	4U4V /4U4T /4U4W
narU - P37758	nitrate	uptake	*E. coli*	occluded inward open (Yan et al 2013)	4IU9
ybgH - P75742	peptide	Uptake	*E. coli*	inward open (Zhao et al 2014)	4Q65
BbFPN - Q6MLJ0	divalent cation	Uptake	*B. bacterio -vorus*	inward/outward open (Taniguchi et al 2015)	5AYO /5AYM
PiPT - A8N031	peptide	Uptake	*P. indica*	inward occluded (Pedersen et al 2013)	4J05
PepT - Q5M4H8	peptide	Uptake	*S. thermop hilus*	inward open (Solcan et al 2012)	4APS

GLUT (Chapter 3)	various	Uniporter	*Various mammalian*	Various see Chapter 3	
NRT1.1 - Q05085	nitrate	Uptake	*A. thaliana*	inward open/nitrate bound (Parker et al 2014)	5A2N/ 5A2O
melB - P30878	melibiose	Uptake	*S. typhimurium*	outward open and occluded (Ethayathulla et al 2014)	4M64

4.1.1 Common Structural Motifs and Secondary Transport Mechanisms in MFS

Despite the variety in function, MFS transporters have common structural elements. 75 % of MFS transporters are made up of 12 transmembrane helices (TMH) with the remaining family members consisting of 14 TMH (Reddy et al 2012). It is largely agreed that TMHs 7-12 originate from an evolutionary duplication events of TMH 1-6, but the formation of the original TMH 1-6 is contentious. Reddy et al (2012) used comparative alignments to predict that the TMH 1-6 originated from triplication of two TMH units. Other sequence comparisons lead to the proposal of a duplication of three TMHs to form TMH 1-6 (Hvorup et al 2002), a hypothesis which was favoured by structural comparison of lacY and fucP motifs (Madej et al 2013). There is possibly more evidence that favours the duplication hypothesis, and duplication events leading to the formation of protein channels are well documented in reviews (Duran and Meiler 2013). While there are examples of trimeric transporters (Tao et al 2015) and genome triplication leading to divergent gene functions (Reyes-Chin-Wo et al 2017), less information is available on triplication events forming a single channel.

The MFS transporter TMH are split into two bundles, TMH 1-6 and TMH 7-12, which are proposed to divide further into two subunits, so a total of four subunits each of three TMH. A central cavity is formed of a helix from each of the four proposed subunits: TMH 1, 4, 7 and 10 (Heng et al 2015). Alignments of 12 and 14 TMH transporters suggest that the two additional helices sit in between the two 6 TMH bundles (Griffith et al 1992).

Structural information for several *E. coli* MFS transporters is now available (Table 4.1). From *E. coli*, there is now structural data for the xylose transporter xylE (Quistgaard et al 2013) and a mutated lactose permease lacY (Kumar et al 2015) in both inwards and outward facing conformations. This enables identification of key

tertiary structure movements in the model of alternating access. The structure and mechanisms are similar to those of the GLUT MFS transporters described in Chapter 3. Figure 4.1 A shows the structure of lacY in the presence of the inhibitor tetramethylrhodamine-5-maleimide. The domains have been highlighted in different colours, the N-terminal THM 1-6 in blue, and the C-terminal TMH 7-12 in red. The two symmetric halves both contribute to form the transport cavity, binding site and gates (Chaptal et al 2011).

An alternating access symporter was outlined in Chapter 3, Figure 3.1. In this example, the binding site is positively coupled. The sugar and proton reinforce binding or exit of each other at the active site. This process can be used to uptake nutrients. The tertiary rearrangements required for this process in *E. coli* xylE are shown in Figures 4.1 B and 4.1 C (Quistgaard et al 2013). The N terminal region stays stationary while the C terminal region shifts to allow for the alternative access. Figure 4.1 C shows outward facing xylE, the arrows indicate the movement undertaken by TMHs to switch to the inward facing conformation. When inward open, the two halves of the protein come together at the periplasmic end with TMH4 and TMH10 in close proximity to one another thus acting as a gate. When in the outward open conformation, the two halves of the protein come together at the cytoplasmic end with TMH1 and TMH7 in close proximity and acting as a gate.

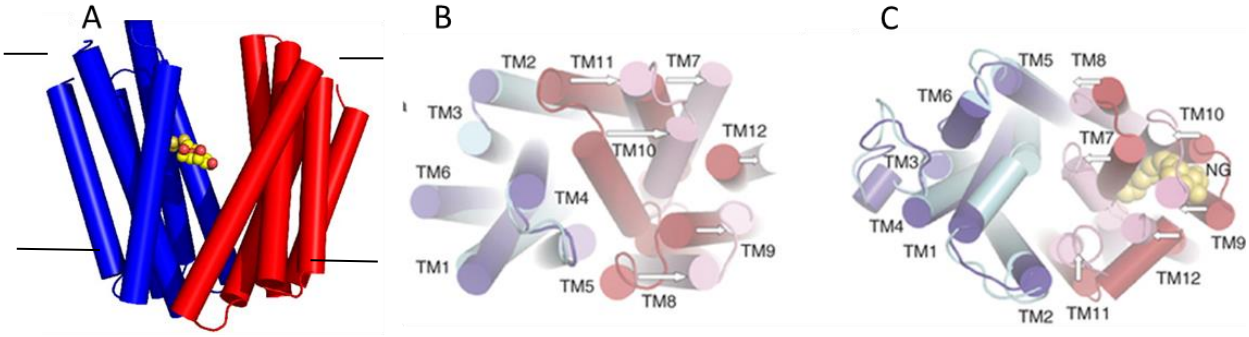


Figure 4.1 Model of alternating access for the *E. coli* Major Facilitator Superfamily (MFS) transporters

A) tertiary arrangement of transmembrane helices (TMHs) of lacY permease, in outward conformation, bound to an inhibitor (methanethiosulfonyl-galactopyranosides). The N-terminal domain is coloured blue while the C-terminal domain is coloured red (adapted from Chaptal et al (2011)). Structure and TMH positions of xylE in inward (B) and outward (C) facing conformations (adapted from Quistgaard et al (2013)). Arrows indicate movement of TMH to the opposite conformation, demonstrating the tertiary movement enabling access to alternate sides of the membrane.

An example of a MFS transporter that effluxes it's substrate is the *E. coli* transporter mdfA. This transporter is an example of an antiporter that uses the interaction between proton motive forces down a gradient into the cell with expulsion of its substrates, such as chloramphenicol, out of the cell, against a concentration gradient. Transport can be electrogenic or electroneutral, meaning it incorporates a net movement of charge such as protons across the membrane, or electroneutral, which relies on the proton gradient for transport of the substrate (Lewinson et al 2003). As an antiporter the two substrates (the proton and substrate) allow structure cycling, however they work in an antagonistic manner. When facing inward, a centrally located amino acid important in substrate binding, Asp34, is protonated but the drug binding site is empty. Upon binding a substrate, such as chloramphenicol, the affinity of the Asp34 for a proton is reduced and the proton exits to the cytoplasm. With the drug present and Asp34 deprotonated, the mdfA structure of rearranges, opening to the periplasm and closing to the cytoplasm. The substrate is then released to the periplasm promoting protonation of Glu26. Interaction between the positive protonation sites and the negative membrane potential of the cytoplasmic side of the plasma membrane causes structural rearrangement to the inward facing conformation. With the protonated mdfA facing inwards the cycle can restart (Heng et al 2015). Key residues for this process are highlighted in Figure 4.2 A.

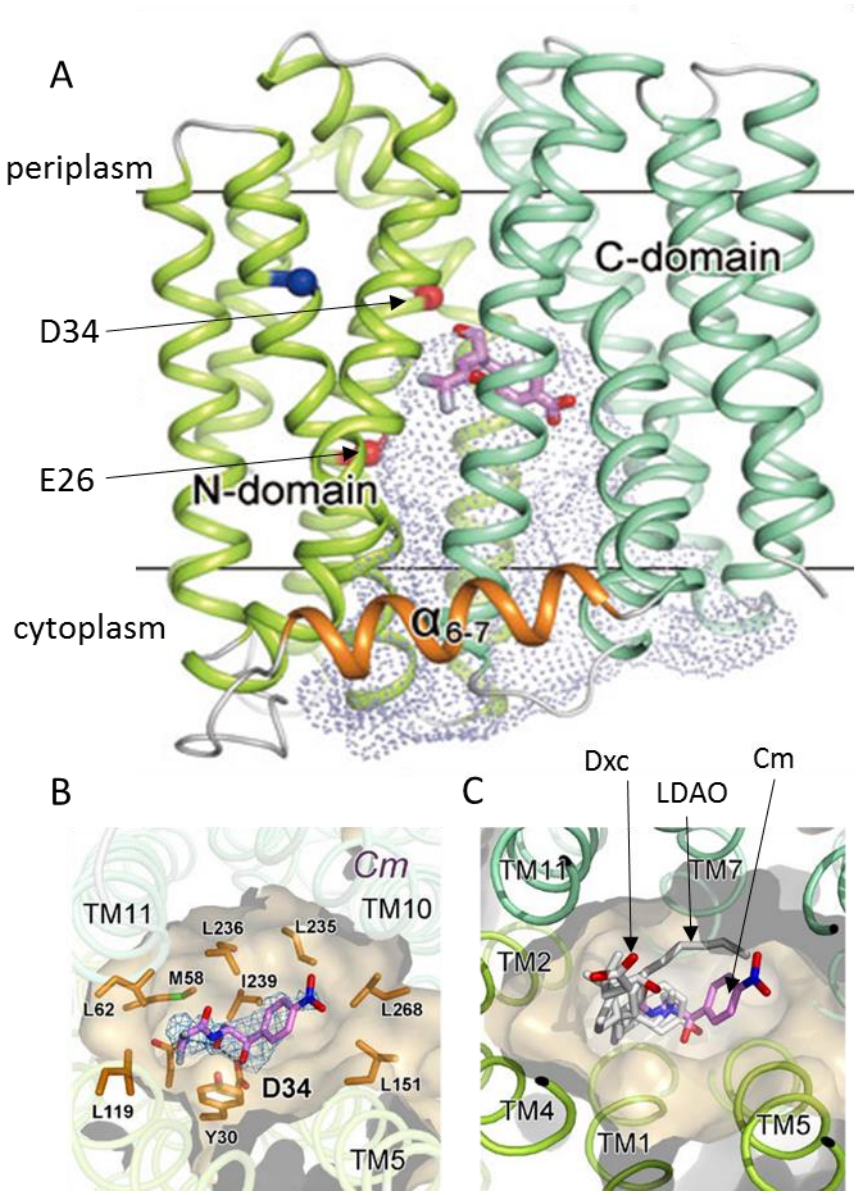


Figure 4.2 Structure of substrate-bound mdfA, with key residues highlighted

Figure adapted from Heng et al (2015). A) Tertiary structure of chloramphenicol-bound mdfA (shown as a ball and stick model); key residues in proton transport are indicated as red spheres. The large internal cavity is highlighted by the blue and white net structure. B) The mdfA chloramphenicol binding site, key residues involved in substrate interaction are coloured orange, showing their position relative to a bound chloramphenicol molecule. C) Binding site of mdfA containing overlays of deoxycholate, LDAO (structurally distinct substrate analogues) and chloramphenicol.

Although many MFS transporters are specific for a particular substrate, others have a wider substrate range (Table 4.2). A number of multidrug transporters, including mdfA, contribute to resistance against several antimicrobials when expressed in *E. coli*, suggesting a binding site capable of targeting different molecules (Nishino et

al 2001). The binding site of mdfA was characterised in the presence of chloramphenicol, N,N-dimethyldodecylamine N-oxide (LDAO) and sodium deoxycholate (Dxc). As highlighted in Figure 4.2 B and 4.2 C, residues Tyr30, Asn33, Asp34 and Leu236 are involved in the binding site of all three substrates, which is flexible and varies depending on the substrate in question. An additional binding site that is distinct from that of chloramphenicol has been proposed for the ion TPP^+ (Lewinson et al 2001). The large cavity, as seen in the mdfA crystal structure, could be sufficiently large to accommodate both molecules (Figure 4.2 A).

4.1.2 Antimicrobial Resistance, Epidemiology and Mechanisms

As highlighted in Table 4.1, MFS transporters have a very wide variety of substrates including nutrient uptake and efflux of harmful substances. The remainder of this chapter will focus on those involved in antimicrobial resistance (Heng 2015, Nishino et al 2001, Paulsen et al 1996) illustrating the importance of this family of transporters.

Many of the antibiotics used clinically are derived from natural products synthesised by bacteria and fungus as a means of defending themselves, or as part of an active offence against other competing organisms. We use them to fight off infections with modern medical approaches such as organ transplantation and chemotherapy heavily reliant on antimicrobial medication. Available data suggests antimicrobial resistance is becoming more prevalent hindering there use in medicine. A 2014 WHO report asked member states to report the frequency that bacteria responsible for common infections exhibited antimicrobial resistance (WHO 2014). In five of the six WHO regions, at least one nation reported that 50 % of *E. coli, K. pneumoniae* and *S. aureus* were resistant to the third-generation antimicrobials, such as cephalosporins, fluoroquinolones and carbapenems. This was an increase in resistance compared to a similar report released in 2011. A comparison of data from 2011 with that from 2014 showed an increase in resistance for antibiotics tested, apart from methicillin resistance S. aureus (ECDC 2014). In accordance with this trend of increasing resistance, a coordinated review into the impact of antimicrobial resistance estimates that 10 million deaths a year will occur as a result of antimicrobial resistance (AMR) by 2050 (O'Neill et al 2016). This is higher than the predicted number of deaths from cancer over the same period.

The well-known *E. coli* is considered to be the work house for molecular biology, protein expression, genetics, etc., as a result of the ease and safety of its propagation. In fact, it is found within the lower intestinal tract of ourselves. However, strains outside of the laboratory and our gastrointestinal tract can and do cause harm. Most concerning is the rise in antibiotic resistant *E. coli* strains that cause urinary tract infections (UTIs). It is estimated that 150 million people a year suffer from UTIs (Gupta et al 2001), with 65% of complicated infections and 75% of uncomplicated infections in the US caused by *E. coli* (Flores-Mireles et al 2015). Additionally, infection can complicate treatment following a kidney transplant due to the need for immune system suppression to prevent donor organ rejection. Of the 49.7% of recipients who contracted an infection, 25.5% of infections were caused by *E. coli* (Bahrami et al 2017). 15.4% of hospital-acquired infections are caused by *E. coli* and related to antimicrobial resistance. This included catheter associated infections, central line blood stream infections, ventilator associated infections and surgical site infections (Weiner et al 2016). A summary of *E. coli* resistance to antimicrobials between 2011 and 2014 is shown in Table 4.2. Other than to carbapenems, an increase in resistance was seen across the time period.

Table 4.2 National Health Safety Network 2014 report on antibiotic resistance in hospital acquired *E. coli* infections (Weiner et al 2016).

Data was collected from reports of 365,490 infections from 4,515 hospitals. Trends show in most cases the proportion of isolates demonstrating resistance increases with time

	Central line blood associated infections resistance (%)		**Catheter associated infections resistance (%)**		**Ventilator associated pneumonias resistance (%)**		**Surgical site associated infections resistance (%)**	
	2011	2014	2011	2014	2011	2012	2011	2014
Extended-spectrum cephalosporin	19.7	22.2	12.9	16.1	15.0	16.7	13.3	15.3
Fluoroquinolone	41.1	49.3	32.7	34.8	38.9	30.8	29.1	30.9
Carbapenems	1.3	1.9	1.2	1.1	1.1	2.2	0.9	0.7
Multi drug resistant	11.1	14.1	5.5	8.0	7.7	9.7	6.1	6.5

Another advantage of focusing on *E. coli* transporters is that they can also provide information on homologous proteins. A report on clinical isolates of *S. aureus* found 17 % had increased expression of the multidrug resistance transporter norA. When only methicillin resistant isolates were considered, 89 % of isolates overexpressed norA (Kosmidis et al 2012). The homologous protein in *E. coli* also provides drug resistance when upregulated (Sun et al 1996).

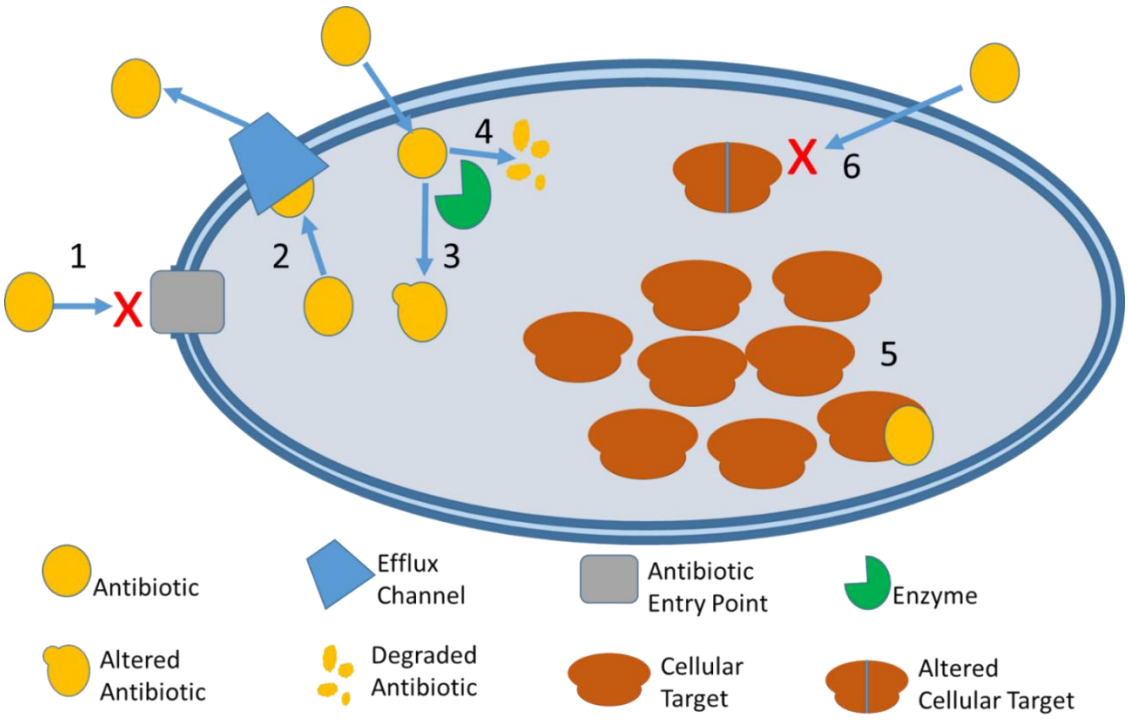


Figure 4.3 Methods of antibiotic resistance acquisition in bacteria

The key indicates important components of the pathway that leads to acquired resistance to antibiotics, as labelled. Blue sphere represents bacterial cell, with bi-layer membrane. 1. Entry of antibiotic blocked by change of or suppression of uptake protein; 2. Antibiotic efflux by specific or broad-spectrum transporters; 3. An enzyme modifies the antibiotic, reducing its ability to interact with a target; 4. Antibiotic enzymatically degraded; 5. Upregulation of antibiotic target means greater concentrations are required to maintain inhibitory affect; 6. Alteration of the target means it no longer interacts with the antibiotic.

Understanding the mechanism of how resistance is acquired is one of the first steps in slowing its progression. Examples of resistance mechanisms are shown in Figure 4.3. One method is to interfere with the uptake of antibiotics at the plasma membrane. For instance, mutations causing alteration of the porin ompC may be used to decrease entry of hydrophilic carbapenem antibiotics (Bajaj et al 2016). Alternatively adjusting the plasma membrane composition can reduce the penetration of hydrophobic antibiotics (Peterson et al 1987).

Bacterial enzymes can also alter or degrade antimicrobial agents. For instance, beta-lactamases provide resistance by hydrolysing the beta lactam ring (Naas et al 2008). It is hypothesised that *K. oxytoca* has been evolving beta lactamases for 100 million years, to achieve resistance against naturally-occurring antimicrobials

(Fevre et al 2005). Alternatively, enzymes can add functional groups to an antibiotic to affect function: acetyltransferases can inactivate antibiotics by the specific addition of an acetyl group (Leslie et al 1988).

Resistance can also be achieved by altering the antimicrobial cellular target. Upregulating the target can decrease antibiotic potency, by increasing the concentration of drug required to inhibit the cell (Palmer et al 2014). Mutations that lead to alteration of cellular proteins may reduce the affinity of an antibiotic and its target (Long et al 2006). The reduced affinity means a greater concentration of the drug is then required to inhibit the cell.

Finally, upregulation of efflux pumps reduces cytoplasmic drug concentration, allowing bacteria to survive at higher antibiotic concentrations (Nishino et al 2001). The MFS transporters have relevant clinical implications with regards to antimicrobial resistance as a result of drug efflux. The importance of individual efflux channels is not simple to unravel. Where there is functional overlap between MFS transporters, it may be necessary to perform double knockout experiments, but this may reveal little information on the capabilities of an individual transporter. For example, when overnight cultures are grown on 100 μg/μL of ethidium bromide, ermE or mdfA single mutants can grow as well as the wild type. However, the ermE and mdfA double mutants did not demonstrate resistance and could not be grown (Tal et al 2009). Schuster et al (2016) grew *E. coli* in media containing sub lethal concentrations of norfloxacin which was then doubled daily to influence the evolution of drug resistant strains. Three strains were assessed; a wild type strain BW25113, a BW25113 acrB knockout (ΔacrB) and BW25113 triple knockout of emrE, mdfA and mdtM (Δ3). Chromosomal DNA was sequenced to detect mutations. Expression of target genes was analysed using qPCR. A wild type strains resistance to 400 μM norfloxacin was achieved within 14-18 days. The Δ3 only achieved a resistance to 12.5 μM norfloxacin across the same timeframe. 5 mutations were seen in all 3 isolates of the wild type strain compared to 12 in all 3 isolates of Δ3 demonstrating the Δ3 strain accumulated mutations faster. Isolates of WT resistant to 400 μM norfloxacin had the emrE, mdfA and mdtM genes knocked out (Datsenko and Wanner 2000). In the absence of these transporters resistance was maintained. Therefore, the transporters were required for the acquisition of the adaptive traits providing high level resistance but were not the key transporters providing the resistance.

4.1.3 Functional characterisation is incomplete

Table 4.3 *E. coli* MFS transporters with inconsistent function assigned by literature

Methodology used is summarised alongside functional assignment.

Protein (UniProt)	Function	Methodology summary
araJ (P23910)	Possible drug exporter (Saier et al 2014)	Based on sequence similarity to other drug transporters.
	Arabinose inducible (Hendrickson et al 1990)	DNase footprinting shows it is downstream of an araC binding site. As this regulator matches araE it is probable it is arabinose inducible.
	Unlikely to function in arabinose uptake (Reeder et al 1991)	Deletion of araJ does not affect ability to grow on low or high concentrations of arabinose. It does not appear to function in arabinose uptake.
Bcr (P28246)	Inhibitory dipeptide efflux (Hayashi et al 2010)	Peptidase deficient strains growth is inhibited by Ala-Gl. Bcr overexpression rescues this phenotype enabled growth on media containing these dipeptides by decreased intracellular concentrations.
	Bicyclomycin efflux (Bentley et al 1993)	Plasmid expression enables growth on higher concentrations of bicyclomycin.
	Sulfonamide (Vedantam et al 1998)	A point mutation was determined to be a trait in *E. coli* adapted to survive on higher concentrations of sulfonamide. Increased expression of bcr increased the minimum inhibitory concentration of sulphonamide.

Protein (UniProt)	Function	Methodology summary
	Multidrug efflux (Nishino et al 2001)	Increases the minimum inhibitory concentration when grown on plates containing tetracyclin, kanamycin, fosfomycin and acriflavine. Bicyclomycin and sulphonamide were not tested.
cynX (P17583)	Cyanate activated (Guilloton et al 2003)	The third gene of an operon activated in response to cyanate, operon named cyn. Its deletion does not affect the ability of *E. coli* to break down cyanate.
	Bromoacetate efflux (Desai et al 2010)	A library of genes was overexpressed in cells grown on bromoactetate to identify those imparting resistance to bromoacetate.
emrB (P0AEJ0)	Multidrug (Lomovskaya et al 1992)	When overexpressed it increases MIC against carbonylcyanide, m-chlorophenylhydrazone and nalidixic acid.
	Bile efflux (Thanassi et al 1997)	Deletion of emrB increases susceptibility when grown in media containing the bile salt carbonyl cyanide m-chlorophenylhydrazone.
	Thiolactomycin efflux (Furukawa et al 1993)	Plasmid overexpression provides resistance when grown in media containing thiolactomycin.
emrY (P52600)	Possible drug transporter (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
	Possible drug transporter (Paulsen et al 1996)	Prediction based on similarity to 4 TMH drug transporter emrE.
emrD (P31442)	Multidrug efflux (Nishino et al 2001)	When overexpressed it increases MIC against benzalkonium and SDS.

Protein (UniProt)	Function	Methodology summary
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in emrD knock out mutants.
	Efflux of H+ uncouplers such as tetrachlorosalicylanilide (Krulwich et al 1990)	When overexpressed it increases MIC of tetrachlorosalicylanilide.
	Provides protection against low energy shock (Naroditskaya et al 1993)	emrD knock out mutants were less able to restore proton motive force in response to uncouplers such as tetrachlorosalicylanilide.
hsrA (P31474)	Predicted multidrug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
	Homocysteine metabolism (Goodrich-Blair et al 2000(	Screening carried out with library to identify genes inhibiting rspA. Overexpression of hsrA lead to accumulation of homocysteine and repressed rspA.
mdfA (P0AEY8)	Multidrug efflux (Edgar et al 1997; Mine et al 1998; Nishino et al 2001; Lewinson et al 2001)	When overexpressed it increases MIC of chloramphenicol, tetracyclin, norfloxacin, doxorubicin, Trimethoprim, acriflavine, ethidium bromide and TPP.
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in mdfA knock out mutants.
	(Lewinson et al 2004) pH homeostasis	Mutant forms of mdfA were designed to hinder transporter function. Wild type E. coli were able to grow at higher pH compared to mutants.

Protein (UniProt)	Function	Methodology summary
	Chloramphenicol efflux (Heng et al 2015)	Structurally defined with X-ray crystallography bound to chloramphenicol.
mdtD (P36554)	Possible drug efflux (Nagakubo et al 2002)	Response regulator BaeR upregulates mdtD conferring resistance to novobiocin and bile salts.
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in mdtD knock out mutants.
mdtG (P25744)	Multidrug efflux (Nishino et al 2001)	When overexpressed it increases MIC against fosfomycin and deoxycholate.
	Norfloxacin resistance (Fabrega et al 2010)	SoxS activator transcriptionally upregulates acrB-tolc and mdtG providing resistance to norfloxacin.
mdtM (P39386)	Multidrug efflux (Nishino et al 2001; Holdsworth et al 2012)	When overexpressed it increases MIC against chloramphenicol, norfloxacin, acriflavine, ethidium bromide and TPP.
	Bile salt efflux (Paul et al 2014)	Knockouts have reduced MIC on bile salt media. Vesicles made following plasmid overexpression of mdtM transport mdtM.
	pH homeostasis (Holdsworth et al 2013)	Knockouts show growth inhibition at pH9-10, lower than the pH required to inhibit wild type.
	AMP uptake (Krizsan et al 2015)	Cells lacking the ABC transporter SbmA and mdtM don't uptake proline rich antimicrobial peptides.
mhpT (P77589)	3-hydroxyphenylpropionic influx (Ferrandez et al 1997 ; Xu et al 2013)	E. coli with mhtP knocked out don't uptake (14C) 3-(3-hydroxyphenyl) propionate. Attempted KO mutation in gene resulted in enhanced uptake.

Protein (UniProt)	Function	Methodology summary
	Arabinose efflux (Koita et al 2013)	Overexpression of mhpT reduced intracellular arabinose concentration.
setA (P31675)	Response to glucose phosphate stress (Sun et al 2011)	Growth in media containing α-methyl glucoside was hindered in setA knock outs. The α-methyl glucoside simulated glucose phosphate stress.
	Sugar efflux (Liu et al 1999; Liu et al 1999)	A library of genes was screened for their ability to reduce growth repression caused by IPTG induced tetA expression. SetA was shown to reduce intracellular IPTG and [14C] labelled lactose.
setC (P31436)	Sugar efflux (Liu et al 1999 ; Liu et al 1999)	A library of genes was screened for their ability to reduce growth repression caused by IPTG induced tetA expression. SetC was shown to reduce intracellular IPTG and [14C] labelled lactose.
	Arabinose efflux (Koita et al 2013)	When setC is overexpressed intracellular arabinose concentration was decreased.
yajR (P77726)	Predicted multidrug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
yeaN (P76242)	Nitroimidazole efflux (Ogasawara et al 2015)	Knock outs where more sensitive when grown in media containing 2-nitromidazole.
	Bromoacetate efflux (Desai et al 2010)	A Library of genes were screened to identify those providing resistance to bromoacetate.

Protein (UniProt)	Function	Methodology summary
ydeE (P31126)	Predicted multidrug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
	Arabinose efflux (Koita et al 2013)	When ydeA is overexpressed intracellular arabinose concentration decreased.
	Inhibitory dipeptide efflux (Hayashi et al 2010)	Peptidase deficient strains growth is inhibited by Ala-Gl. ydeE overexpression enabled *E. coli* growth on media containing these dipeptides by decreased intracellular concentrations.
ydhC (P37597)	Predicted drug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in ydhC knock out mutants.
ydiM (P37597)	unknown (Pao et al 1998)	Sequence similarity.
	Possible drug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
ydiN (P76198)	Unknown (Pao et al 1998)	Sequence too dissimilar to other MFS members predict function.
	Possible shikimic metabolism (Johansson et al 2006)	YdiN is upregulated during shikimic acid metabolism.
yfcJ (P77549)	Unknown (Pao et al 1998)	Sequence too dissimilar to other MFS members predict function.
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in yfcJ knock out mutants.
yhhS (P37621)	Glyphosate export (Staub et al 2012)	Radiolabelled glyphosphate accumulates at a lower rate in BL21 cells expressing yhhS.

Protein (UniProt)	Function	Methodology summary
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in yhhS knock out mutants.
yhjX (P37662)	Unknown (Pao et al 1998)	Sequence too dissimilar to other MFS members predict function.
	Upregulated in response to toxic compounds (Fozo et al 2008)	Identified by microarray analysis of *E. coli* after toxic peptide exposure.
ynfM (P43531)	Predicted multidrug efflux (Nishino et al 2001)	Overexpression did not affect MIC of tested antimicrobials.
	Arabinose efflux (Koita et al 2013)	The concentration of intracellular arabinose rose in ynfM knock out mutants.

One of the impediments of tackling the antimicrobial capabilities of MFS channels is the inconsistency of current information. The differences are addressed in Table 4.3, this highlights the different methods used, to investigate possible sources of discrepancy.

Where multiple functions have been assigned the literature is not always in direct disagreement. Most contradictions are a result of assays focusing on different functions yielding positive results. The protein mdfA, for example, has been assessed using multiple techniques yielding several different suggested functions. Nishino et al (2001) established minimum inhibitory concentrations (MIC) with a screen of antimicrobials in the presence of transporters. When overexpressed mdfA was shown to provide protection against toxicity of the antimicrobials chloramphenicol, tetracyclin, norfloxacin, doxorubicin, trimethoprim, acriflavine, ethidium bromide and TPP+ indicating function as a broad specificity drug exporter. Work on pH tolerance showed that in the absence of wild type mdfA *E. coli* cells are less tolerant to alkaline growth conditions suggesting it may function in cellular homeostasis as well (Lewinson et al 2004). Koita et al (2013) demonstrated that deletions of mdfA led to higher intracellular arabinose concentrations and postulated a role in this monosaccharide export. It is likely

mdfA does function in drug efflux as its active site has been structurally defined bound to chloramphenicol (Heng et al 2015), however, it is not clear why any bacteria would want to export a food source such as arabinose. This does however show the importance of methodology in inferring function and that care should be taken to check presumed functional roles for this and other targets. Even for well-studied transporters, many of the experimental methods are measuring a physiological effect hence there is potential the transporter is not directly interacting with the substrate.

Koita et al (2013) assigned arabinose export function to mdfA and twelve other MFS transporters using gene knock out and over-expression experiments. The identified transporters are setC, cmr (mdfA), ynfM, mdtD, yfcJ, yhhS, emrD, ydhC, ydeA, ybdA, ydeE, mhpT, and kgtP. For the deletion experiments, the intracellular arabinose concentration increased while for the over-expression experiments the internal arabinose concentration decreased. This lead to the conclusion that mdfA and the other transporters function in arabinose efflux. It may not be so simple; carbon metabolism is complex with different energy sources requiring different genes for uptake and metabolism. Binding of cAMP to CRP allosterically increases the CRPs DNA binding affinity; the DNA bound CRP interacts with bacterial RNA polymerase activating catabolite dependent operons (Tutar and Harman 2006). Interfering with a gene may affect the repression/activation of an operon. This interference may therefore have physiological effects other than just increasing or decreasing transport of a single substrate. Where additional experimental data is available, many of those assigned arabinose transporters have other potential roles. When expressed using plasmids a number of these arabinose efflux transporters have been shown to increase the MIC of *E. coli* in the presence of antimicrobial substances for instance emrD (Nishino et al 2001), mdtD (Nagakubo et al 2002), yhhS (Staub et al 2012). In these whole cell experiments, the transporters may again not be directly transporting the tested substrate. It may instead be a case that the physiological changes make these cells less susceptible. The transporters yajR, yfcJ, ydhC and ynfM are predicted antibiotic transporters based on sequence similarity but no substrates have been identified (Nishino et al 2001). Assignment of function should again be approached with caution and ideally the function of each transporter should be experimentally tested (Rahman et al 2017).

Based on this issue, the main aim of this work is to implement a non-specific approach for characterising ligand binding to specific integral membrane proteins.

Non-specific in this context means that any substrate can be tested producing fluorescent or radioactive alterations. This avoidance of specific radio-labelled, etc., substrates would speed up screening and drastically reduce costs. It also avoids the main flaw when using a whole cell non-specific method, such as minimum inhibitory concentrations (MIC), which assumes that the observed effect can only be via the deleted or over-expressed transporter. Specifically, as the MIC method relies on inhibition of *E. coli* growth it is not suitable for identifying substrates that are not toxic, such as sugars.

One of the best non-specific methods used to determine ligand binding is the thermal temperature shift assay. In this approach, the protein is heated in the presence of a fluorescent dye such as SYBR Orange. As the protein unfolds and exposes its hydrophobic core, the SYBR Orange binds and begins to fluoresce. Addition of a substrate that binds to the target protein results in the stabilisation of the tertiary structure, so a higher temperature is needed to begin the unfolding process. This temperature shift can be used to identify ligands capable of binding to the protein. It is a cheap, fast method that has proved to be successful for many targets (Huynh and Partch 2015). The main problem with this approach is that it is only suitable for soluble proteins. The outer protein surface of an integral membrane protein and the dynamic detergent micelles required for solubilisation and stabilisation of the membrane protein outside of a membrane are already hydrophobic, hence, the SYBR Orange fluorescent signal is already high even before protein denaturation/ unfolding begins.

Tagged-sfGFP was used as an indirect means of following target protein in the absence of SYBR Orange. An illustrated outline of the approach is shown in Figure 4.4. To generate a thermoprofile the protein sample was divided and exposed to a temperature gradient. After temperature exposure, the samples were centrifuged to remove any precipitated, insoluble protein. The GFP concentration is then measured. At lower temperatures protein solubility is maintained and the GFP concentration matches that of the 4 ºC control, this is represented by the graphic in Figure 4.4 A. Figure 4.4 B-C represents the temperature that begins to destabilise the protein leading to aggregation, precipitation and GFP signal loss.

To allow comparisons between experiments, the melting temperature (T_m) is utilised. The temperature at which the fluorescence signal is half that of the maximum and minimum GFP fluorescence values is deemed the protein's melting temperature. In Figure 4.4 C the GFP signal is almost completely lost as the GFP-protein has entirely lost solubility. To identify potential substrates a thermoprofile

can be generated in the presence of the substrate and compared to the control. 4.12 D shows a theoretical comparison of a protein with and without substrate. The substrate causes increased stability adjusting the thermoprofile and the T_m.

We will be applying this approach to transporter proteins. If we observe binding based on a positive temperature shift then it is an indication of protein-ligand interactions. The inherent assumption in this approach is that the substrate binds within the conduction pathway and therefore is a specific substrate. This may not be the case as binding can occur through non-specific interactions, binding to allosteric sites as well as binding resulting in inhibition and not transport. Therefore, this approach requires additional experiments to confirm that the identified binding substrate is transported. The approach, however, does provide important information that is potentially relevant to the behaviour of the transporter in the cell's membrane.

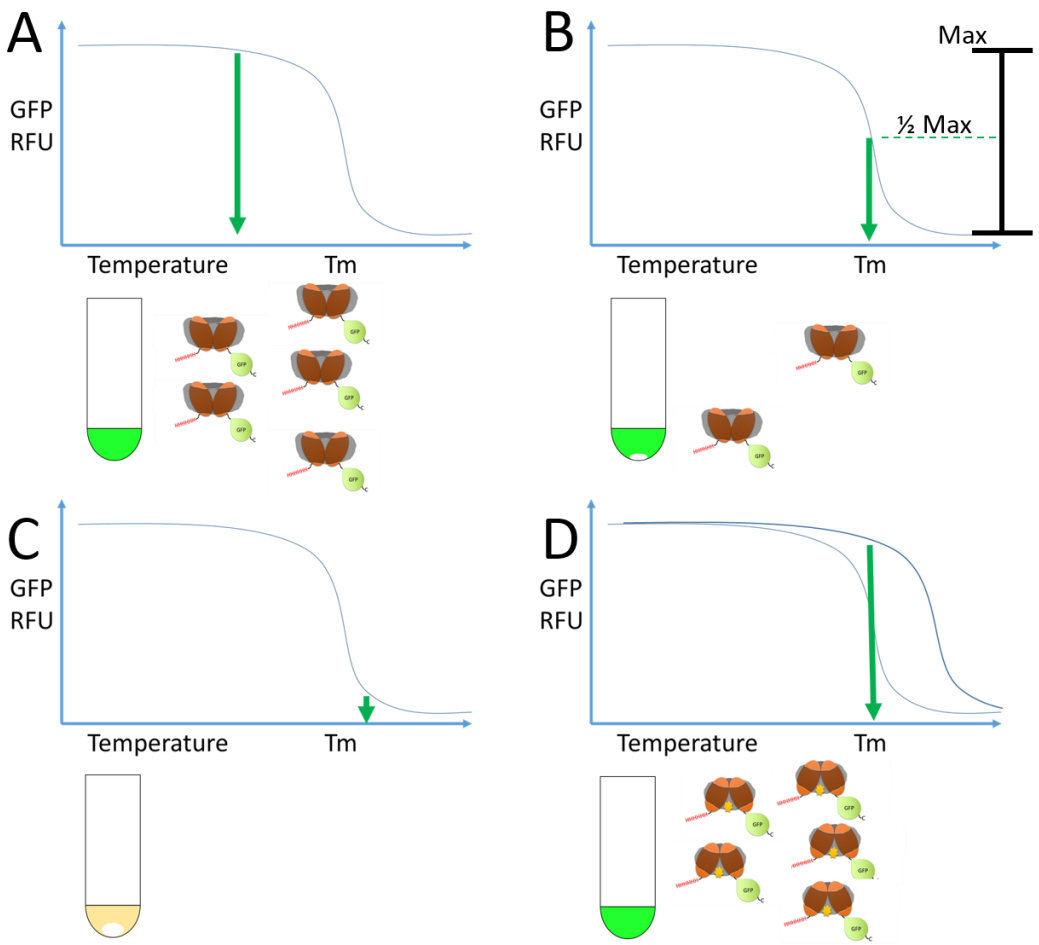


Figure 4.4 The hypothetical thermoprofile assay using a GFP reporter

Graphs show the GFP relative fluorescence units (RFU) plotted against temperature. T_m (melting temperature) is the temperature at which the signal is half that of the maximum RFU. Green arrows illustrate the strength of GFP signal, RFU, at a specific temperature. Test tubes: sample colour represents the amount of GFP-bound protein remaining in solution; greener illustrates more GFP-bound protein in solution. Beside the test tubes are representative pictures of the membrane protein: two halves of the MFS transporter (orange ovals) surrounded by detergent micelle (grey band); GFP (green circle) and histidine (orange tail) tags included. In D, the yellow dot represents bound ligand. A-C) hypothetical thermoprofile assay different temperatures: A, low temperature; B) melting temperature; C) above melting temperature. D) melting temperature with bound ligand. At lower temperatures (A), the protein's solubility, and therefore maximum GFP signal, is maintained. As the temperature increases (B), the protein is destabilised, leading to aggregation and precipitation, resulting in a reduction in GFP fluorescence signal reduction. At temperatures above the T_m, all GFP signal is eventually lost as protein loses solubility. D) The addition of the protein's ligand causes increased stability of the protein, causing increase in GFP signal across the temperature range, including at the T_m temperature.

4.1.4 Chapter aims

- Develop a GFP-tagged method to determine membrane protein/substrate binding with a range of substrates. The method will rely on changes in the target protein and not the substrate so that substrates can be screened.

- Apply the method to the same protein against multiple potential substrates to allow identification of the range of a protein's substrates.

- Focus on antibiotic transporters to define their substrates

4.2 Results

4.2.1 Creation of expression system and expression testing

E. coli MFS transporters were selected because (1) it was assumed that *E. coli* proteins would be easier to over-express in *E. coli;* (2) Many MFS transporters have been characterised so they make ideal control target proteins; (3) they are self-contained and do not require additional proteins to be functional (4) they have a broad range of functionalities, making any information that we discover potentially biologically important. The vector that was selected was H6-msfGFP (a gift from Scott Gradia (Addgene plasmid # 29725)). Insertion of any target protein into this vector will generate a protein with a N-terminal hexahistidine tag and a C-terminal GFP tag. The actual variant of GFP used is called superfolder GFP as it is capable of folding properly and being functional in the periplasm as well as the cytoplasm (Dinh and Bernhardt 2011). This vector is designed to induce expression of inserted target protein in BL21 (DE3) cells using the lactose/IPTG inducible T7 promoter. Additionally, this vector provided two key features, firstly, the presence of a TEV cleavable N-terminal hexahistidine tag. Secondly, a C-terminal GFP tag used as a reporter for expression optimisation and subsequent assay development. Figure 4.5 shows a topological representation of a 12 transmembrane MFS transporter expressed using the H6-msfGFP vector.

Primers were designed to clone the MFS genes from the *E. coli* strain K-12 MG1655 using PCR. This strain is considered to be a representative strain (type strain) for *E. coli* (Keseler et al 2017). Primers were flanked to introduce ligation-independent cloning (LIC) sites (Table 2.19). An outline of the methodology for LIC is given in Materials and Methods, Section 2.1.4.2. Figure 4.6 shows the successful amplification of these products by PCR, using an annealing temperature of 55 ºC. Figure 4.6 H used a range of annealing temperatures; 50 ºC 55 ºC and 60 ºC for each PCR reaction. PCR products were gel purified (Chapter 2, Section 2.1.3) and used in ligation independent cloning. Vectors were isolated after cloning and confirmed with an analytical PCR screen (Figure 4.7 A-B) or analytical restriction digest using the empty H6-msfGFP vector as a control (Figures 4.8, 4.9 and 4.10). XhoI digests the empty vector control to give a band of 6088 base pairs (bp); digested vectors larger than the control were deemed potential candidates. These were sequenced to confirm the presence of the desired PCR product in-frame with the start codon. Table 4.4 shows a summary of the cloning process with

information on products, indicating whether the target product was successfully amplified, cloned or the correct, in-frame sequence confirmed.

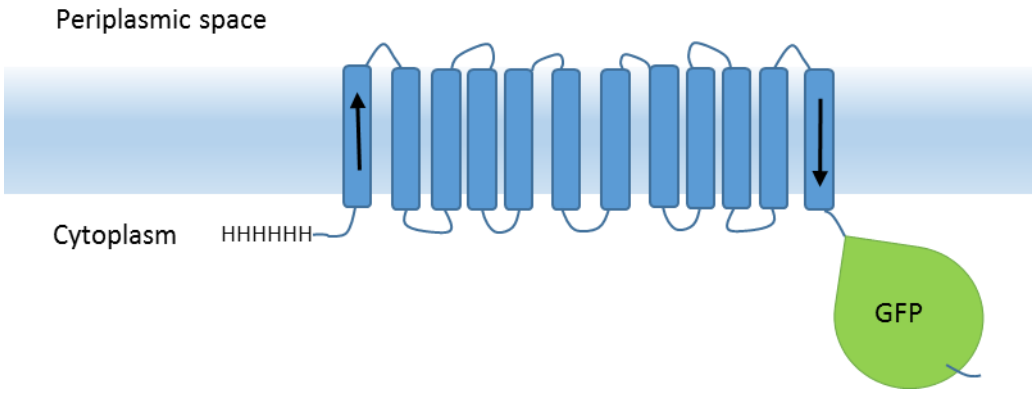


Figure 4.5 Hypothetical topology of MFS transporters, expressed with the H6-msfGFP vector

The blue rectangles crossing the plasma membrane represent transmembrane helices. Relative positions of the GFP (green circle) and histidine tags (HHHHHH) are shown at N- and C-termini of the protein, respectively.

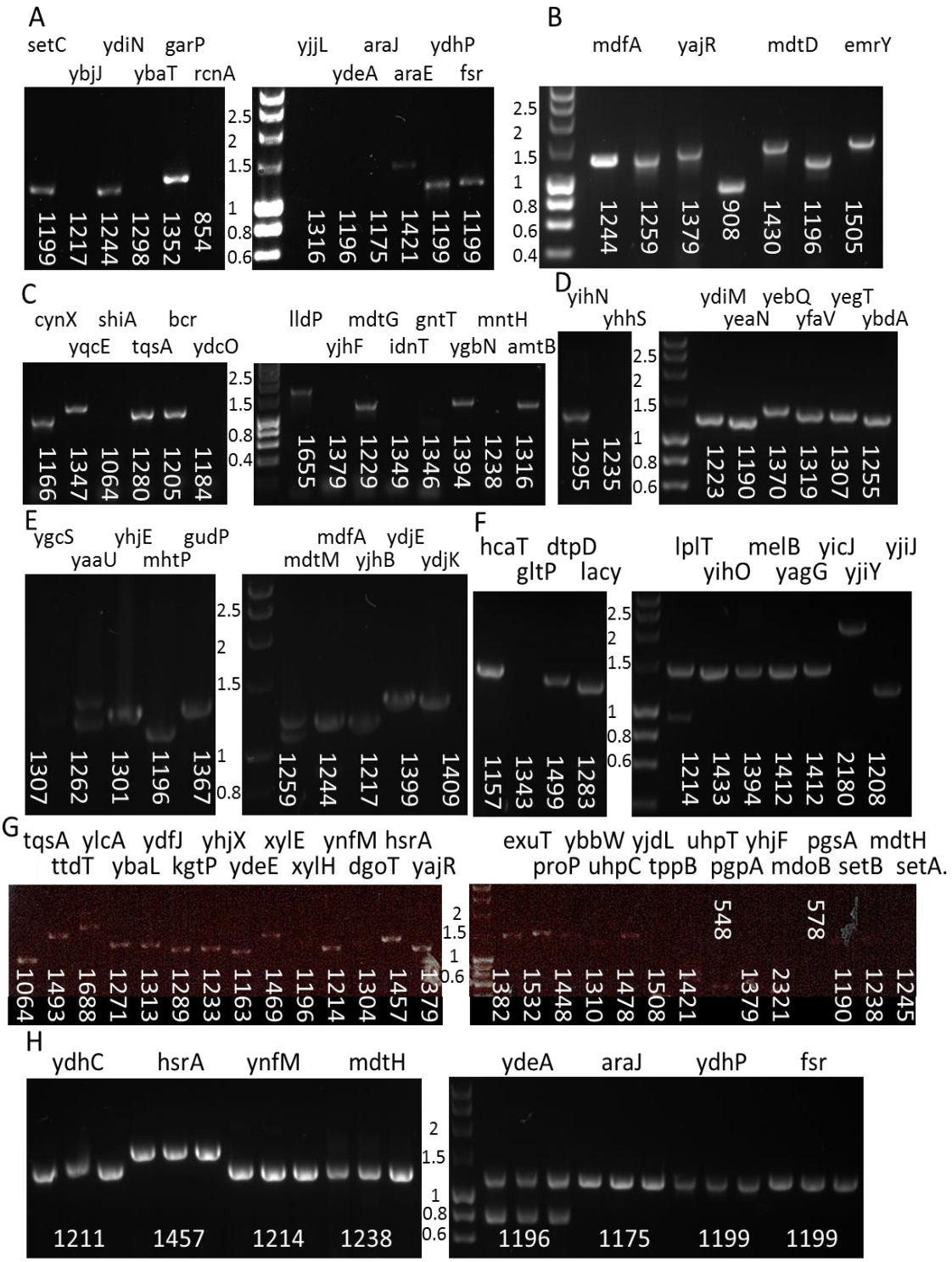


Figure 4.6 PCR amplification of major facilitator superfamily in preparation of cloning into prokaryotic expression vector, H6-msfGFP

Amplified from *E. coli* K12 MG1655 genomic DNA with H6-msfGFP LIC primer tags for cloning. Gene targets labelled above the gel (e.g. setC, ybjJ). Expected PCR product size (bp) displayed perpendicular to or below the target. Molecular weight (Mwt) marker sizes labelled on the left or centre of the gel (kbp).

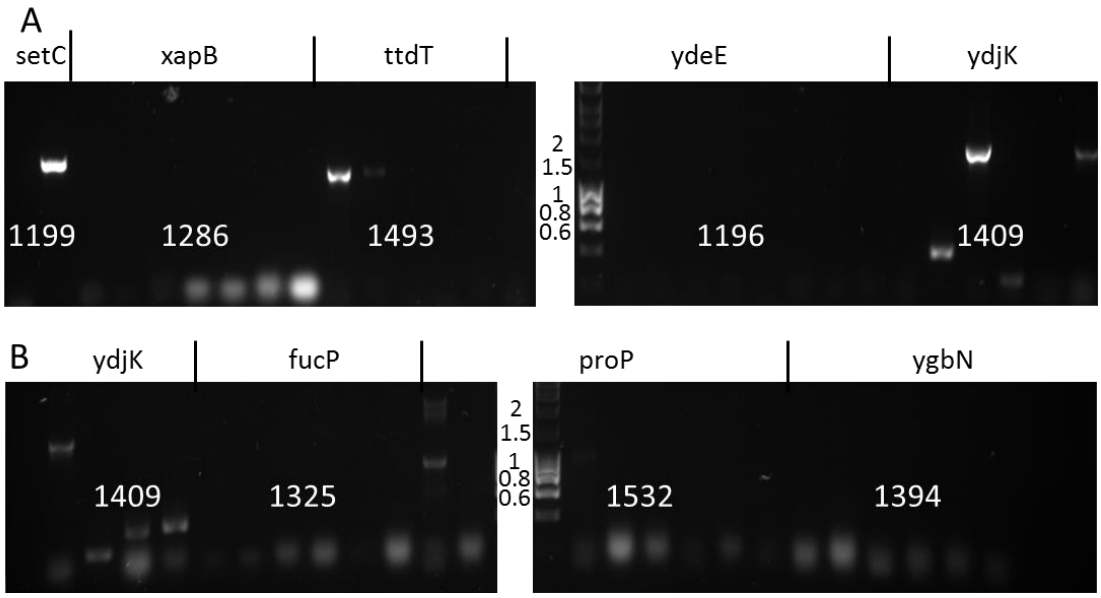


Figure 4.7 Analytical PCR of constructed prokaryotic expression vector, H6-msfGFP, containing major facilitator superfamily proteins

Amplified from constructed H6-MFS-sfGFP expression vectors. Gene targets labelled above the gel (e.g. xapB, ttdT). Expected PCR product size (bp) displayed perpendicular to or below the target. Molecular weight (Mwt) marker sizes labelled in the centre of the gel (kbp).

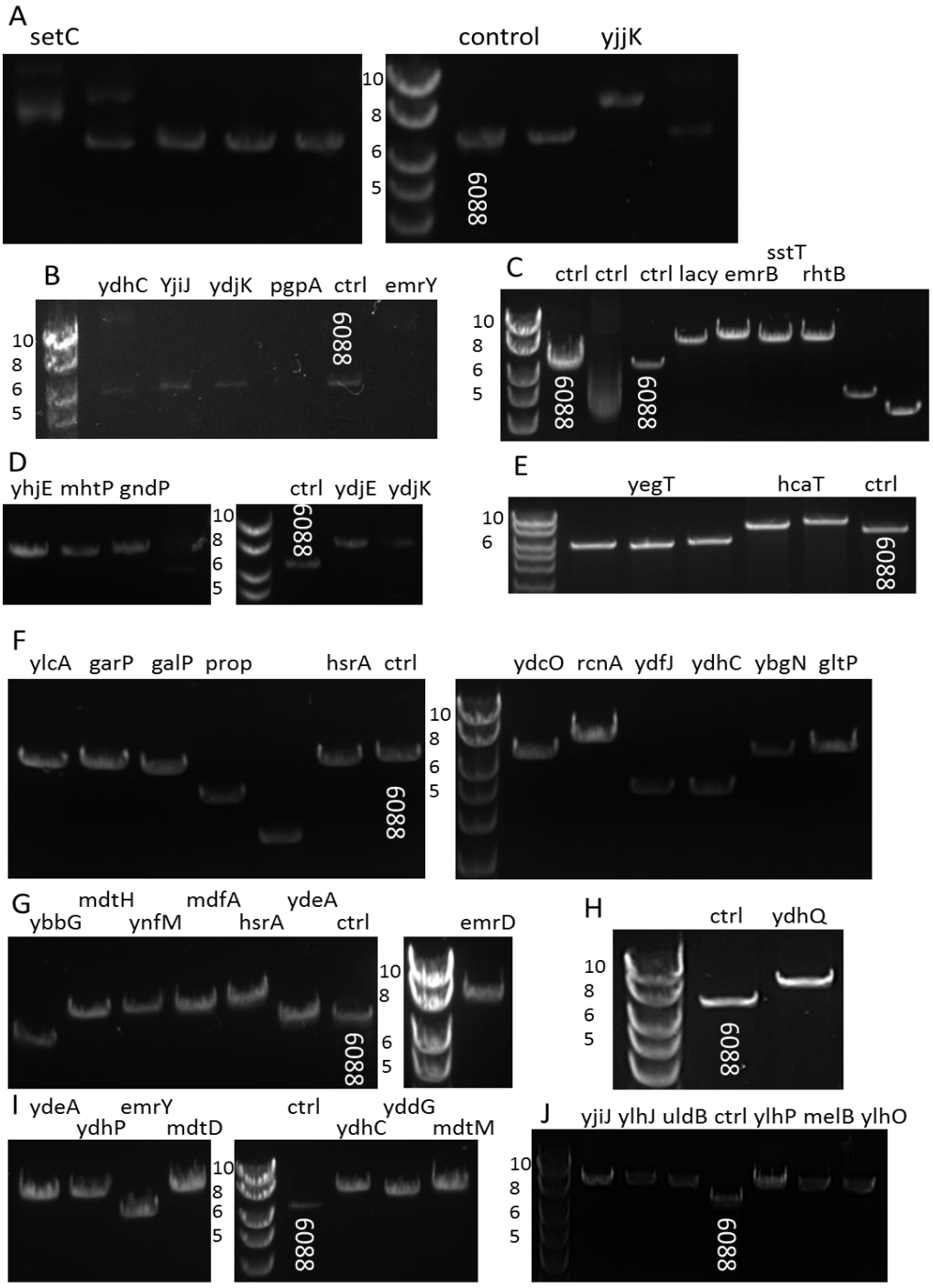


Figure 4.8 Analytical digest of constructed MFS expression vectors

Expression vectors H6-msfGFP, either empty or carrying MFS inserts, with vector backbone digested using XhoI. Empty H6-msfGFP control lanes labelled 'ctrl' and '6088' (predicted size of linearised empty vector is 6088bp). Expression vectors which were the same size as the linearised control were discontinued at this stage; these are the lanes without a label at the top. Expression vectors which were linearised to give a product larger than the control were treated as successful candidates; these are labelled above their lanes with the insert (e.g. setC, yjjK). Molecular weight (Mwt) marker sizes labelled in centre of the gel (kbp).

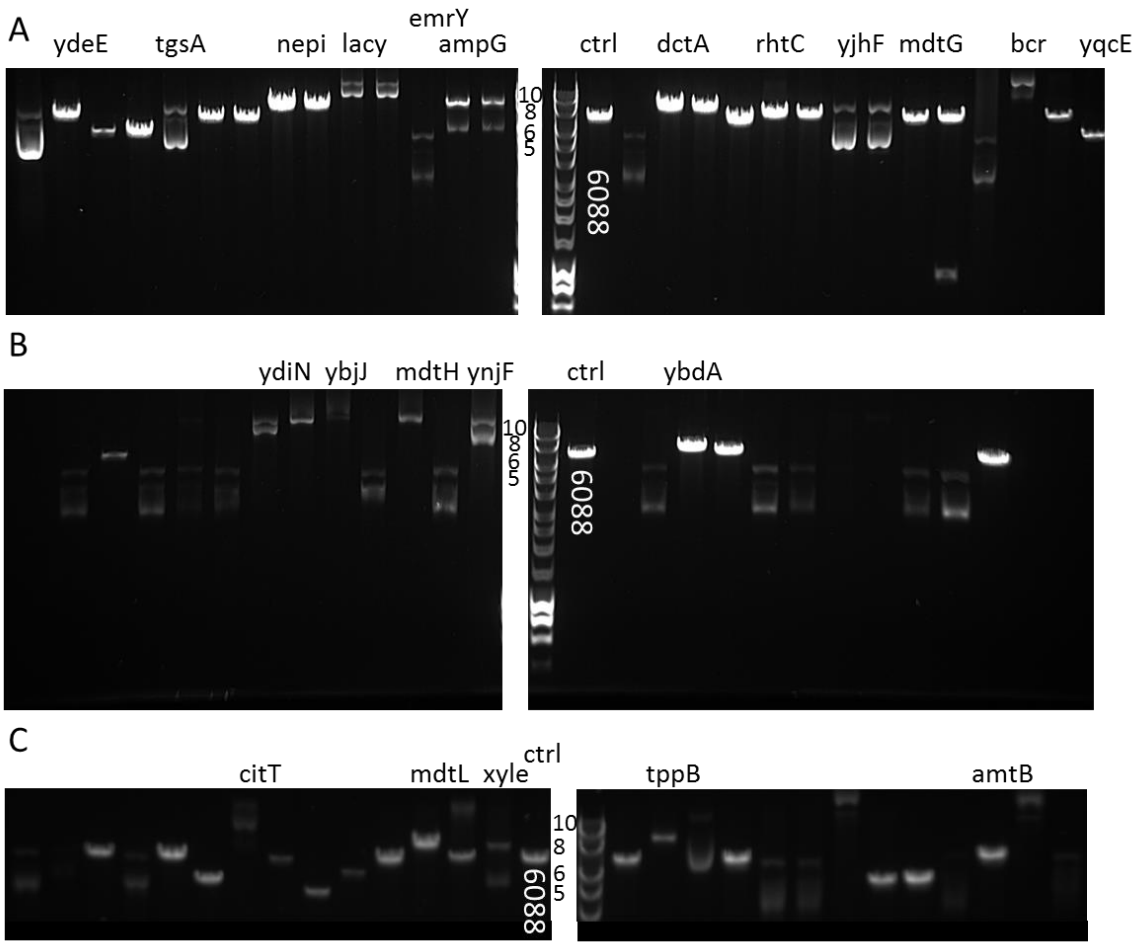


Figure 4.9 Analytical digest of constructed MFS expression vectors

Expression vectors H6-msfGFP, either empty or carrying MFS inserts, with vector backbone digested using XhoI. Empty H6-msfGFP control lanes labelled 'ctrl' and '6088' (predicted size of linearised empty vector is 6088bp). Expression vectors which were the same size as the linearised control were discontinued at this stage; these are the lanes without a label at the top. Expression vectors which were linearised to give a product larger than the control were treated as successful candidates; these are labelled above their lanes with the insert (e.g. ydeE, nepi). Molecular weight (Mwt) marker sizes labelled in centre of the gel (kbp).

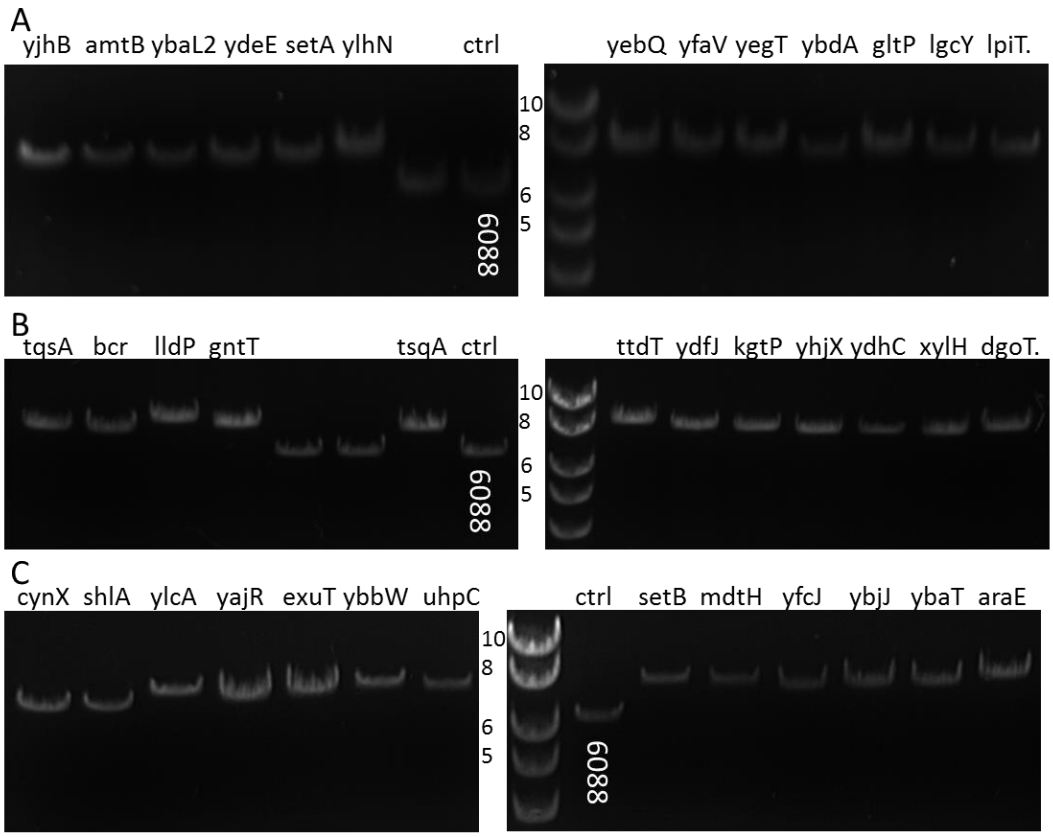


Figure 4.10 Analytical digest of constructed MFS expression vectors

Expression vectors H6-msfGFP, either empty or carrying MFS inserts, with vector backbone digested using XhoI. Empty H6-msfGFP control lanes labelled 'ctrl' and '6088' (predicted size of linearised empty vector is 6088bp). Expression vectors which were the same size as the linearised control were discontinued at this stage; these are the lanes without a label at the top. Expression vectors which were linearised to give a product larger than the control were treated as successful candidates; these are labelled above their lanes with the insert (e.g. yjhB, amtB). Molecular weight (Mwt) marker sizes labelled in centre of the gel (kbp).

Table 4.4 Details of the *E. coli* MFS family members investigated in this project

'X' indicates successful amplification of target, that colonies were produced in T4 cloning or that the sequence was confirmed by sequencing to be correct, with the start codon in-frame with the vector backbone. Blank space indicates unsuccessful step, or a target was not further. Hypothetical function is included for reference.

Target	UniProt	amplified	cloned	Seque-nced	Size (AA's)	Function
ydjE	P38055	X	X	X	451	Sugar efflux
ydjK	P76230	X	X	X	458	Sugar efflux
yhjE	P37643	X	X	X	422	unknown
mhtP	P77589	X	X	X	387	arabinose / 3-hydroxyphenylpropionic
gudP	Q46916	X	X	X	444	glucarate
mdtM	P39386	X	X	X	408	multidrug efflux / bile salt
mdfA	P0AEY8	X	X	X	403	Multidrug efflux
mdtL	P31462	X	X	X	375	Multidrug efflux
cynX	P17583	X	X	X	377	cyanate permease
shiA	P76350	X	X	X	437	shikimate transporter
tqsA	P0AFS5	X	X	X	389	AI-2
bcr	P28246	X	X	X	390	bicyclomycin/cysteine/ sulfonamide
mdtG	P25744	X	X	X	398	multidrug efflux
gntT	P39835	X	X	X	437	3-phenylpropionic acid / fructuronate
ybaL	P39830	X	X	X	424	NADP
amtB	P69681	X	X	X	427	ammonia
ttdT	P39414	X	X	X	486	L-tartrate/succinate
ylcA	P0ACZ8	X	X	X	551	glycolate
ydfJ	P77228	X	X	X	426	unknown
kgtP	P0AEX3	X	X	X	418	alpha-ketoglutarate
yhjX	P37662	X	X	X	396	pyruvate
xylE	P0AGF4	X	X	X	478	xylose

Target	UniProt	amplified	cloned	Seque-nced	Size (AA's)	Function
ydhC	P37597	X	X	X	392	multidrug efflux /arabinose
xylH	P0AGI4	X	X	X	376	xylose
hsrA	P31474	X	X	X	474	multidrug efflux / homocysteine metabolism
dgoT	P0AA76	X	X	X	423	galactonate
ynfM	P43531	X	X	X	393	multidrug efflux /arabinose
yajR	P77726	X	X	X	448	unknown
exuT	P0AA78	X	X	X	449	hexuronate
ydeE	P31126	X	X	X	387	multidrug efflux /arabinose
ybbW	P75712	X	X	X	471	allantoin
uhpC	P09836	X	X	X	425	uhpT regulation
setB	P33026	X	X	X	385	Sugar efflux
mdtH	P69367	X	X	X	401	multidrug efflux
setA	P31675	X	X	X	391	Sugar efflux
yfcJ	P77549	X	X	X	384	arabinose
setC	P31436	X	X	X	388	Sugar efflux
ybaT	P77400	X	X	X	421	amino acid
rcnA	P76425	X	X	X	273	nickel/cobalt resistance
ydeA	P31122	X	X	X	387	Arabinose operon
araE	P0AE24	X	X	X	462	arabinose
fsr	P52067	X	X	X	388	fosmidomycin
nepI	P0ADL1	X	X	X	395	ribonucleoside
fucP	P11551	X	X	X	430	L-fucose
yddG	P46136	X	X	X	291	tryptophan
citT	P0AE74	X	X	X	486	citrate/succinate
mdtD	P36554	X	X	X	465	multidrug efflux /arabinose
dctA	P0A830	X	X	X	417	citrate/succinate
sstT	P0AGE4	X	X	X	408	sodium serine

Target	UniProt	amplified	cloned	Seque-nced	Size (AA's)	Function
rhtC	P0AG38	X	X	X	205	threonine transporter
ampG	P0AE16	X	X	X	489	muropeptide
emrB	P0AEJ0	X	X	X	496	multidrug
emrD	P31442	X	X	X	387	multidrug
yihN	P32135	X	X	X	420	unknown
yebQ	P76269	X	X	X	445	multidrug
yfaV	P76470	X	X	X	428	unknown
yegT	P76417	X	X	X	424	unknown
ybdA	P24077	X	X	X	406	arabinose
hcaT	Q47142	X	X	X	374	phenylpropionic acid
gltP	P21345	X	X	X	436	glutamate
lacY	P02920	X	X	X	416	lactose permease
lplT	P39196	X	X	X	393	lysophospholipid flippase
lldP	P33231	X	X	X	540	lactate metabolism
yjhB	P39352	X	X	X	394	Sialic acid
amtB	P69681	X	X			Ammonia Influx
gntT	P39835	X	X			Gluconate influx
lldP	P33231	X	X			Lactate and glycolate uptake
melB	P02921	x	X			Melbiose uptake
prop	P0C0L7	X	X			Osmoregulator uptake
tppB	P77304	X	X			Dipeptide uptake
ttdT	P39414	X	X			L-tartrate/succinate uptake
ybaL	P39830	X	X			unknown
ybjJ	P75810	X	X			unknown
ydhP	P77389	X	X			unknown
ydiN	P76198	X	X			Possible shikimic metabolism
yjhB	P39352	X	X			unknown
yjhF	P39357	X	X			unknown
yjiJ	P39381	X	X			bromoacetate efflux
yjjK	P0A9W3	X	X			Unknown

Target	UniProt	amplified	cloned	Seque -nced	Size (AA's)	Function
ynjF	P76226	X	X			unknown
yqcE	P77031	X	X			unknown
araJ	P23910	X				Unknown
dtpD	P75742	X				di/tripeptide uptake
emrY	P52600	X				Possible drug transport
garP	P0AA80	X				unknown
yaaU	P31679	X				unknown
yagG	P75683	X				unknown
ydiM	P76197	X				unknown
yeaN	P76242	X				Nitroimidazole/ bromoacetate efflux
yicJ	P31435	X				unknown
yihO	P32136	X				Sulphoquinovose uptake
yjdL	P39276	X				di/tripeptide uptake
yjiY	P39396	X				possible peptide transporter

Following confirmation of successful cloning, expressed MFS transporters were purified using affinity chromatography (Materials and Methods, Section 2.3.1-2.3.2). The cloned transporters are C-terminally GFP tagged, therefore, the GFP fluorescence signal can be used to quantify MFS transporter expression.

In order to quantify expression levels, the GFP signal was calibrated. To do this, the fluorescence of defined concentrations of GFP were measured, generating a standard curve. The standard curve was used to estimate the levels of over-expressed transporter. The GFP was purified using expression of the unaltered H6-msfGFP vectors (Materials and Methods, Section 2.3.7). Purified H6-sfGFP was quantified by measuring 280 nm (Nanodrop). ExPASy ProtParam (Gasteiger et al 2005) defined the molecular weight for H6-sfGFP to be 31.435 kDa and an extinction coefficient of 2602 L mol^{-1} cm^{-1}. These were used to calculate the H^-sfGFP concentration. Figure 4.11 B shows the calibration curve that was generated. Concentrations of MFS transporters could then be calculated using the equation of the standard curve.

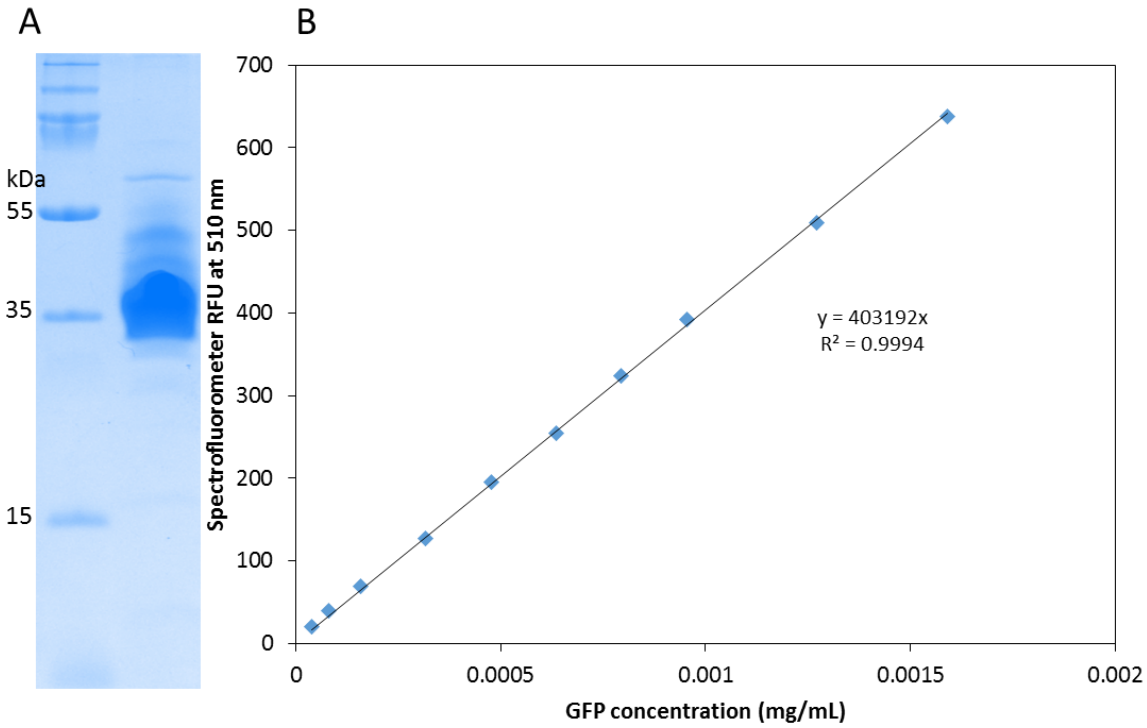


Figure 4.11 Standard curve of purified H6-sfGFP using spectrofluorimeter

H6-sfGFP was expressed in BL21 (DE3) and purified using affinity chromatography. A) Purified protein was analysed using SDS-PAGE (expected kDa 31.435 kDa). Molecular weight marker sizes are shown to the left of gel. B) H6-sfGFP RFU measurements at a range of concentrations were used to generate a standard curve equation. Each value is an average of three measurements.

As discussed in Chapter 3, the exact experimental conditions can significantly influence the efficiency of protein over-expression. Selecting the appropriate conditions is important when optimising protein expression levels. There are many variables that can be tested when screening for the optimal conditions, however, two conditions were selected as the most important. These were growth media (LB, TB or 2YT) and temperature. *E. coli*'s growth rate is highly temperature-dependent. Its natural environments include the gastrointestinal tract and the soil, and as such it must be prepared for growth at a variety of temperatures. MFS transporters have a variety of proposed functions and may, therefore, favour different temperatures or richness of growth media. The expression temperatures selected were 18 °C and 30 °C, the time frame for expression was overnight, ~16 hours.

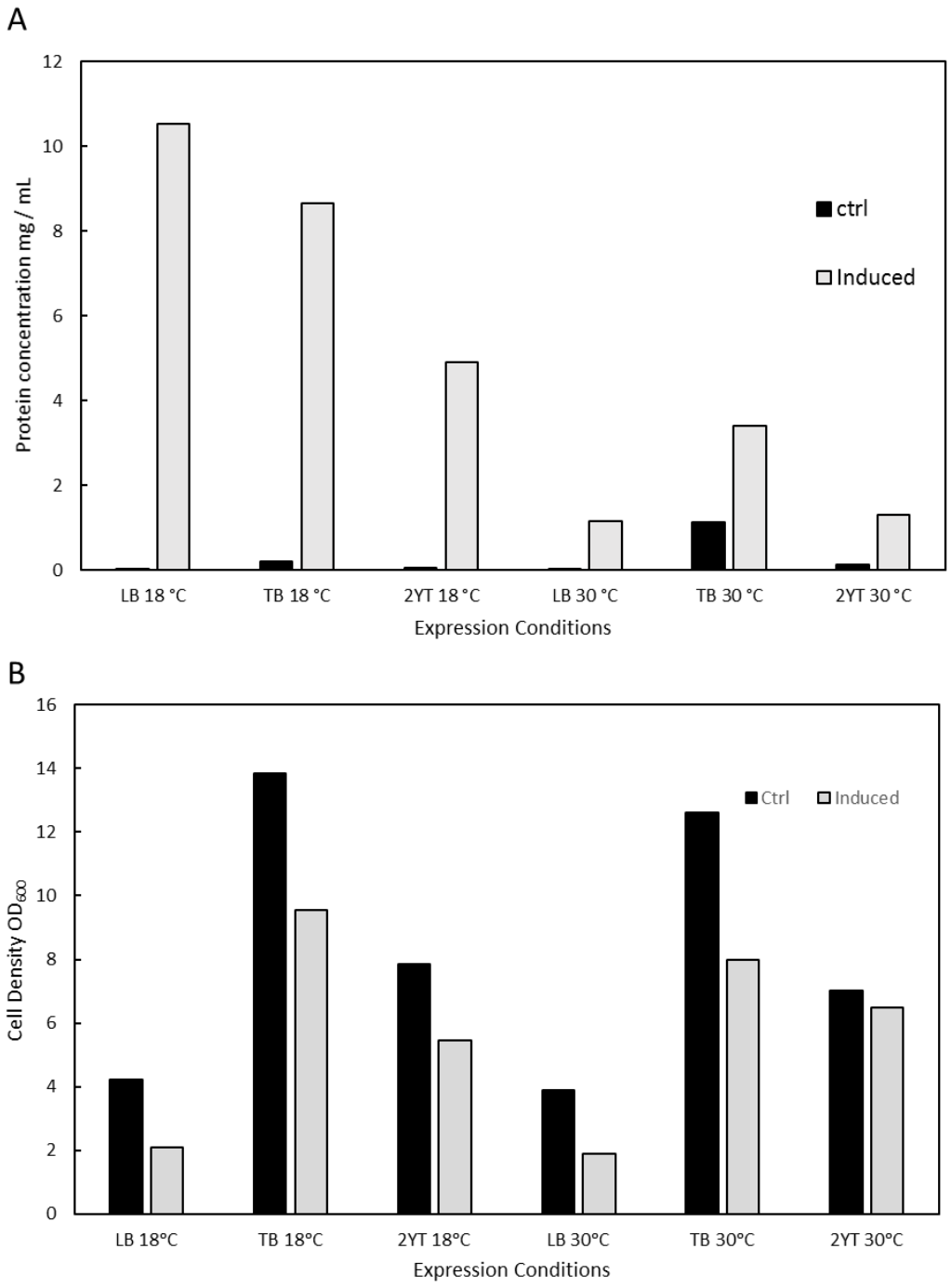
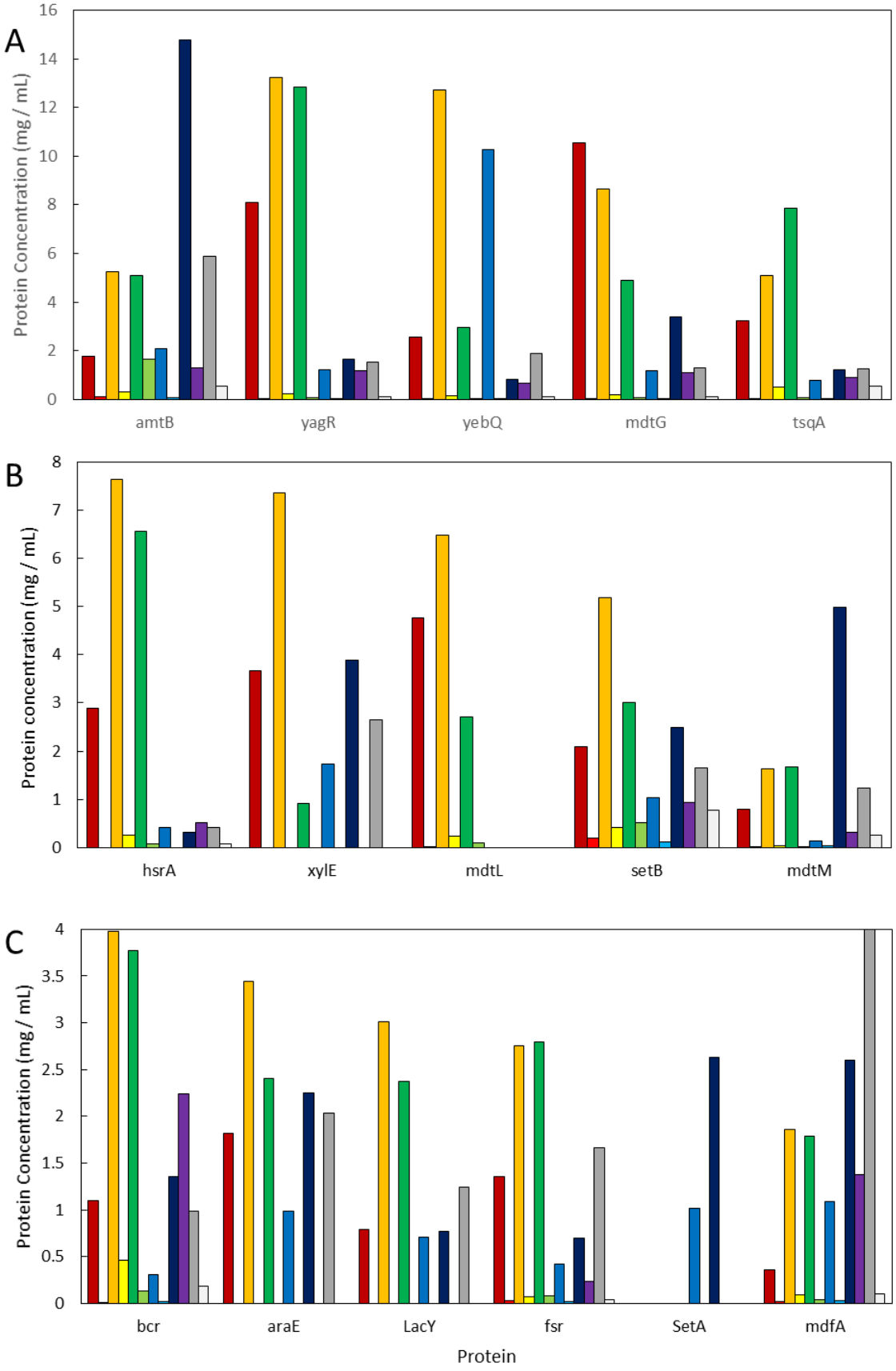


Figure 4.12 Protein concentration and cell density when mdtG is expressed in BL21 (DE3) under different conditions

Overnight protein expression in BL21 (DE3) *E. coli* at two temperatures (18, 30 °C), in three different media (LB, TB and 2YT). Post-expression cell density (A) was measured (OD_{600}) and the cells were pellet by centrifugation. Following cell resuspension, H6-sfGFP fluorescence was measured using a spectrofluorimeter and a standard curve was used to calculate H6-sfGFP concentration in mg/mL (B; standard curve not shown). Each value is an average of three measurements.

Twenty-seven of the successfully cloned MFS transporters were screened for optimal expression conditions. mdtG is shown as a detailed example in Figure 4.12. All conditions produced larger amounts of mtdG protein when induced when compared to the uninduced control. The 18 °C temperature trials were consistently better than 30 °C, with the best protein expression being in LB. Figures 4.13 and 4.14 collate this information for the remaining expressed transporters. Generally, the best growth condition was 18 °C in TB media, but this was not always the case. Table 4.5 summarises the best conditions for each protein tested. Lower than average final cell density indicates overexpression of a protein reduces the cells growth capability. For instance, mdtM shows consistently lower cell densities when induced compared to uninduced. The data for setA and araE is incomplete but final density can be compared to other proteins. When induced both show a greater decrease in final density than other overexpressed MFS transporters. It is possible that the mRNA from these over-expressed targets could have an adverse effect on cell growth. However, we suggest that this indicates the transporter is active within the membrane of the intact cell. As such, over-expression of the transporter causes the transport of an unknown substrate that negatively effects cell growth.


A
Protein Concentration (mg / mL)
16
14
12
10
8
6
4
2
0
amtB
yagR
yebQ
mdtG
tsqA
B
Protein concentration (mg / mL)
8
7
6
5
4
3
2
1
0
hsrA
xylE
mdtL
setB
mdtM
C
Protein Concentration (mg / mL)
4
3.5
3
2.5
2
1.5
1
0.5
0
bcr
araE
LacY
fsr
SetA
mdfA
Protein

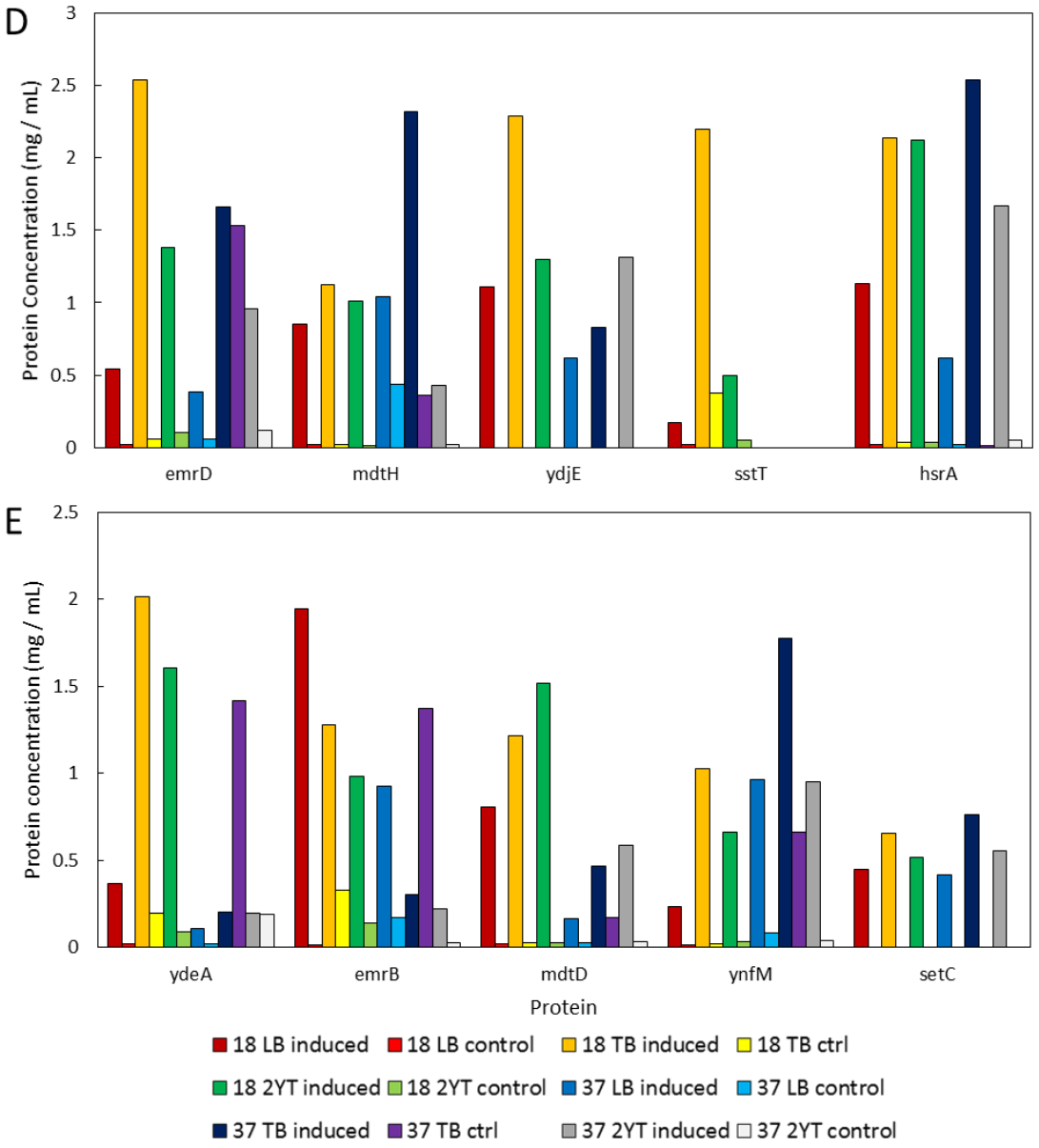
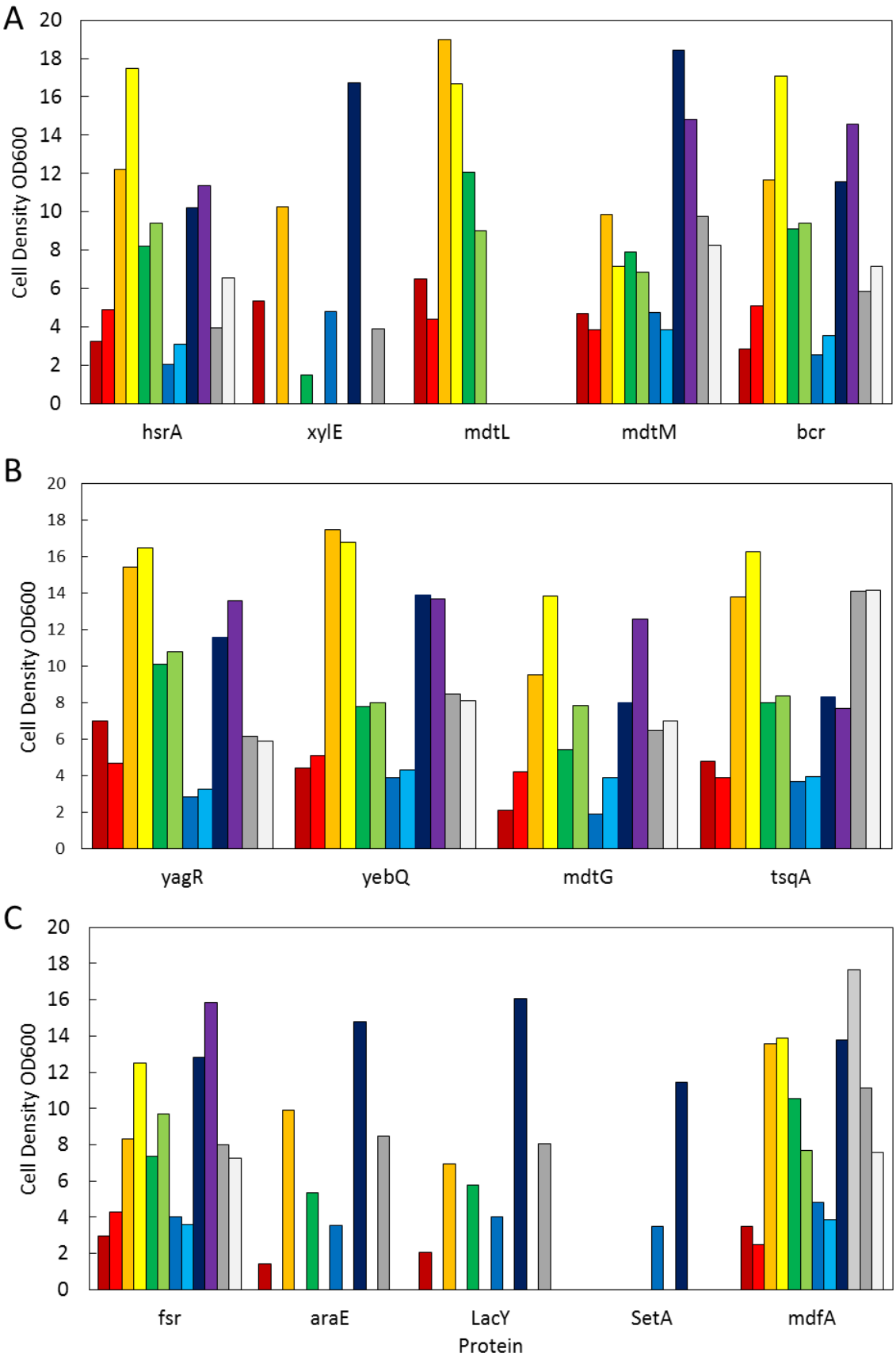


Figure 4.13 Protein concentration when Multi Facilitator Superfamily proteins are expressed in BL21 (DE3) under different conditions.

Overnight protein expression in BL21 (DE3) *E. coli* at two temperatures (18, 30 °C), in three different media (LB, TB and 2YT). Post-expression cells were pellet by centrifugation. Following cell resuspension, H6-sfGFP fluorescence was measured using a spectrofluorimeter and a standard curve was used to calculate H6-sfGFP concentration in mg/mL (standard curve not shown). The expression condition is shown in the key. The protein is indicated below each set of bars. Graphs are ordered from highest average expression to lowest. Each value is an average of three measurements.


A
Cell Density OD600
0
2
4
6
8
10
12
14
16
18
20
hsrA
xylE
mdtL
mdtM
bcr
B
Cell Density OD600
0
2
4
6
8
10
12
14
16
18
20
yagR
yebQ
mdtG
tsqA
C
Cell Density OD600
0
2
4
6
8
10
12
14
16
18
20
fsr
araE
LacY
SetA
mdfA
Protein

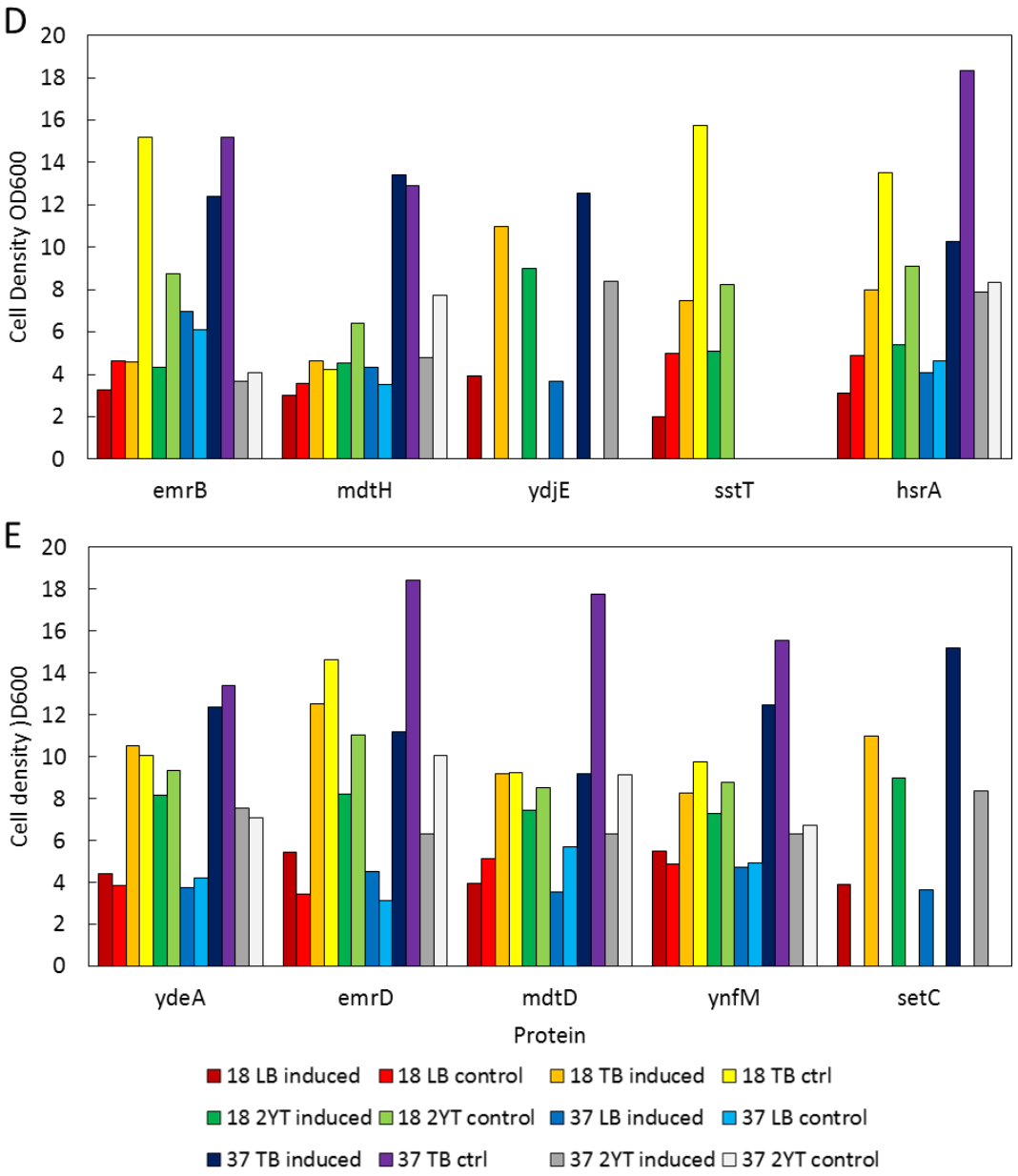


Figure 4.14 Final cell density when Multi Facilitator Superfamily proteins are expressed in BL21 (DE3) under different conditions

Overnight protein expression in BL21 (DE3) *E. coli* at two temperatures (18, 30 ºC), in three different media (LB, TB and 2YT). Post-expression cell density was measured (OD600). The expression condition is shown in the key. The overexpressed protein is indicated below each set of bars. The graphs are ordered from highest average expression to lowest. Each value is an average of three measurements.

Table 4.5 Optimal expression conditions for tested MFS proteins

Optimal expression was deemed to be the condition that produced most protein when expression was assessed using the H6-sfGFP fluorescence.

Expressed protein	Optimum expression condition
amtB	TB 18 ºC
yagR	TB 18 ºC
yebQ	TB 18 ºC
mdtG	LB 18 ºC
tsqA	2YT 18 ºC
hsrA	TB 18 ºC
xylE	TB 18 ºC
mdtL	TB 18 ºC
setB	TB 18 ºC
mdtM	TB 30 ºC
bcr	TB 18 ºC
araE	TB 18 ºC
lacY	TB 18 ºC
fsr	2YT 18 ºC
SetA	TB 30 ºC
mdfA	TB 30 ºC
emrD	TB 18 ºC
mdtH	TB 30 ºC
ydjE	TB 18 ºC
sstT	TB 18 ºC
hsrA	2YT 18 ºC
ydeA	TB 18 ºC
emrB	LB 18 ºC
mdtD	2YT 18 ºC
ynfM	TB 18 ºC
setC	TB 30 ºC

Approximately 62 % of the H6-transporter-sfGFP proteins expressed to their highest levels in TB at 18 °C, ~19 % for TB at 30 °C and ~15 % for 2YT at 18 °C. TB and 2YT are rich growth media, hence, rich media accounted for the best conditions while for the temperature, ~81 % of the conditions were best at 18 °C. Based on these results, the remaining protein production conditions were fixed at 18 °C in TB for ~16 hrs.

4.2.2 Development of the ligand binding assay

As described in the introduction, a major aim of this work was to implement a non-specific approach of characterising integral membrane protein-ligand interactions. The above GFP-tagged transporters were used to investigate the protein's thermal stability in the presence and absence of ligands. Clearly from the above experiments, it can be inferred that the GFP-tagged transporters can be measured in whole cells. Thus, three methods were used, as detailed below, to measure H6-transporter-sfGFP thermostability: one used resuspended cell lysate while the other two methods took advantage of the hexahistidine tag which allowed for the purification of the H6-transporter-sfGFP protein.

4.2.2.1 Method 1: Membrane solubilised transporter-sfGFP

The advantage of having the target transporter GFP-tagged is that the signal from the GFP allows the potential characterisation of the target within a mixture of proteins. Partial purification is expected to increase the sensitivity, however, the approach of attaching GFP to the transporter should provide a signal that stands out sufficiently well as *E. coli* naturally does not contain fluorescent molecules with similar profiles to GFP. Thus, it was hypothesised that a thermoprofile could be generated from the detergent solubilised membrane containing a transporter tagged with GFP, allowing the determination of the T_m melting temperature of the target protein.

As outlined above, the initial step was to over-express the transporter. Cells were then lysed before the soluble protein content was separated from insoluble membranes by high speed centrifugation. The purified insoluble cellular content, which included the integral membrane proteins, was resuspended before solubilisation using a detergent, such as Brij-35. Use of partial purification reduces the cost and length of the procedure. The partially purified material was split and

combined with either the substrate or an equivalent concentration of the substrate buffer and incubated at 4 ˚C. Following incubation, the protein was divided between separate PCR tubes. Each tube was exposed to one temperature from a range covering the thermoprofile. Following temperature exposure, precipitated material was removed by centrifugation and the final concentration of H6-sfGFP measured using spectrofluorimetry. Each data point was carried out in triplicate.

Initial tests compared the profile of the protein under control conditions (absence of substrate) and in the presence of a substrate, as summarised in Table 4.6. As an example, Figure 4.17 A shows the comparative thermoprofile of mdtM, with and without 5 mM chloramphenicol. Previous crystallographic studies have determined chloramphenicol bound mdtM (Alegre 2016). The GFP relative fluorescent units (RFU) indicate the amount of protein in solution which corresponds with protein stability: increasing the temperature initially has no effect on RFU. At a certain temperature the stability decreases dramatically and the H6-mdtM-sfGFP precipitates. As a means of characterising this process and allowing comparison between conditions and targets, the T_m is the temperature at which the fluorescence plot reaches the RFU halfway between the maximum and minimum RFU values. H6-mdtM-sfGFP did show a significant difference between the T_m of protein in the presence of substrate, 39.7 ˚C compared to the control, 47.9 ˚C. However, a repeat with a separate purification did not replicate this result.

The experiment was repeated for H6-setB-sfGFP, H6-lacY-sfGFP, H6-mdfA-sfGFP and H6-araE-sfGFP. There T_m values are shown in Figure 4.17 B. The thermoprofiles for H6-setB-sfGFP, H6-lacY-sfGFP, H6-mdfA-sfGFP and H6-araE-sfGFP in the presence and absence of potential substrates and showed no difference.

It was also possible that the observed results were mostly influenced by H6-sfGFP and not the transporter. If this were the case, then the T_m for H6-sfGFP alone would be similar. To test this, H6-sfGFP was purified and its thermoprofile recorded (Figure 4.15).

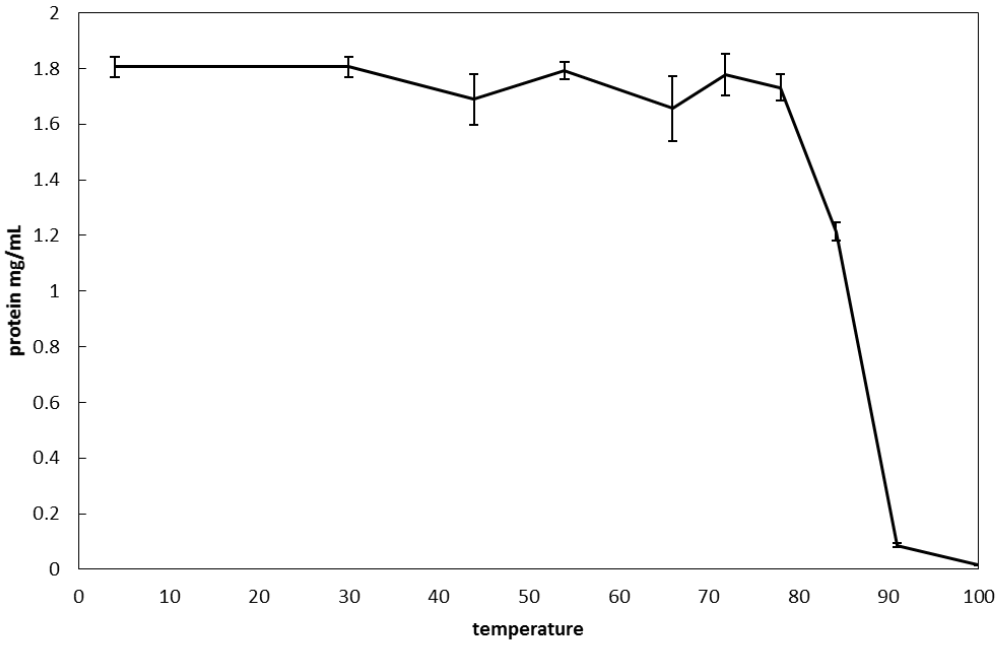


Figure 4.15 Thermoprofile for H6-sfGFP in cellular lysate

H6-sfGFP protein was expressed overnight in BL21 (DE3) at 18 °C. Following expression, cells where lysed and the soluble fraction was isolated from the insoluble fraction by centrifugation. Fractions of whole cell lysate were incubated at designated temperatures, from 4 to 90 °C. Following incubation samples were ultracentrifuged to separate aggregated protein. Following centrifugation, the H6-sfGFP concentration was measured using spectrofluorimetry. Each value is the mean of two measurements with average dispersion from the mean indicated by standard deviation.

The determined T_m for H6-sfGFP of 87 °C was substantially higher than all of the other H6-transporter-sfGFP constructs tested (Figure 4.12 B). The average T_m value for the transporters tested is 56.7 °C. As this value is lower than for H6-sfGFP it implies that it is the effect of temperature on the transporter that is dominating the loss of signal and not the effect of temperature on sfGFP. Clearly, H6-sfGFP is stable up to 80 °C (Figure 4.15). What can also be seen from Figure 4.17 B is that there is a variation in T_m values for each of the H6-transporter-sfGFP constructs, with a T_m of 62.5 °C for xylE at the highest and a value of 39.69 °C for mdtM at the lowest. This is expected as the transporters vary in sequence and, therefore, differ in their thermo-stabilities within a detergent micelle. This data did provide evidence that the approach is able to detect differences between the thermostability of specific transporters.

Based on previously published data, the ligands tested were known substrates. At this stage, it was thought that failure could be due to unsuitable concentrations of the ligand tested, as only one substrate concentration was examined. To investigate this mdfA was tested with a range of substrate concentrations. The target selected was the well characterised H6-mdfA-sfGFP and the substrate chloramphenicol (Heng et al 2015). The thermoprofiles are shown in Figure 4.18 A, and the T_ms plotted in Figure 4.18 B.

These results initially appeared exciting as an increase in chloramphenicol concentration produced an obvious change in T_m. However, instead of a stabilisation of the complex, as seen for soluble protein-ligand binding (Kanbar and Ozdemir 2010; Rupesh et al 2014; Huynh and Partch 2015), the data showed a decrease in stability as indicated by the decreasing T_m values at higher concentrations of chloramphenicol. In this case, as the ligand binds to mdfA, the transporter undergoes a conformational change based on the alternating access model previously described. As it changes conformation, the structure of the detergent micelle surrounding mdfA would change so that the complex is less stable, resulting in protein precipitation. The changes are shown schematically in Figure 4.16.

Looking closely at Figure 4.18, it appears that the maximum negative shift remains stable between 0.5 – 2.5 mM chloramphenicol. If our hypothesis is correct then we could assume that the system has become saturated. There is clearly a rapid decrease in T_m values from 2.5 to 8.5 mM chloramphenicol, the maximum tested. The partial purification method (purification of membranes containing transporter-sfGFP) means that there are a variety of *E. coli* proteins in the preparation and not just H6-mdfA-sfGFP. These other proteins could be interacting with mdfA in the substrate bound state resulting in its precipitation. To remove potential interfering interactions, nickel ion immobilised metal affinity chromatography (Ni^{2+}-IMAC) was used to purify the H6-transporter-sfGFP for further investigations.

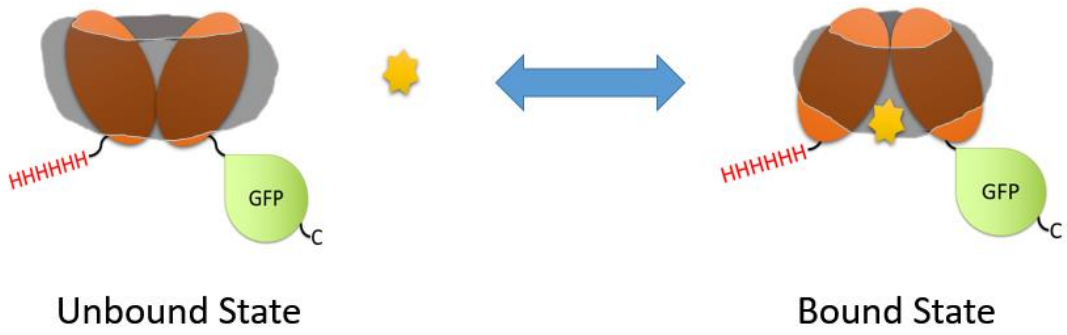


Figure 4.16 Schematic view of MFS transporter in a detergent micelle binding to its substrate

The MFS tranporter (two orange ovals) surrounded by detergent micelle (translucent grey band), with N- and C-terminal histidine (HHHHHH) and sfGFP (green circle), respectively. The substrate (yellow star) is bound and released, causing a large conformational change (blue arrow). This conformational change potentially stabilises or destabilises the system.

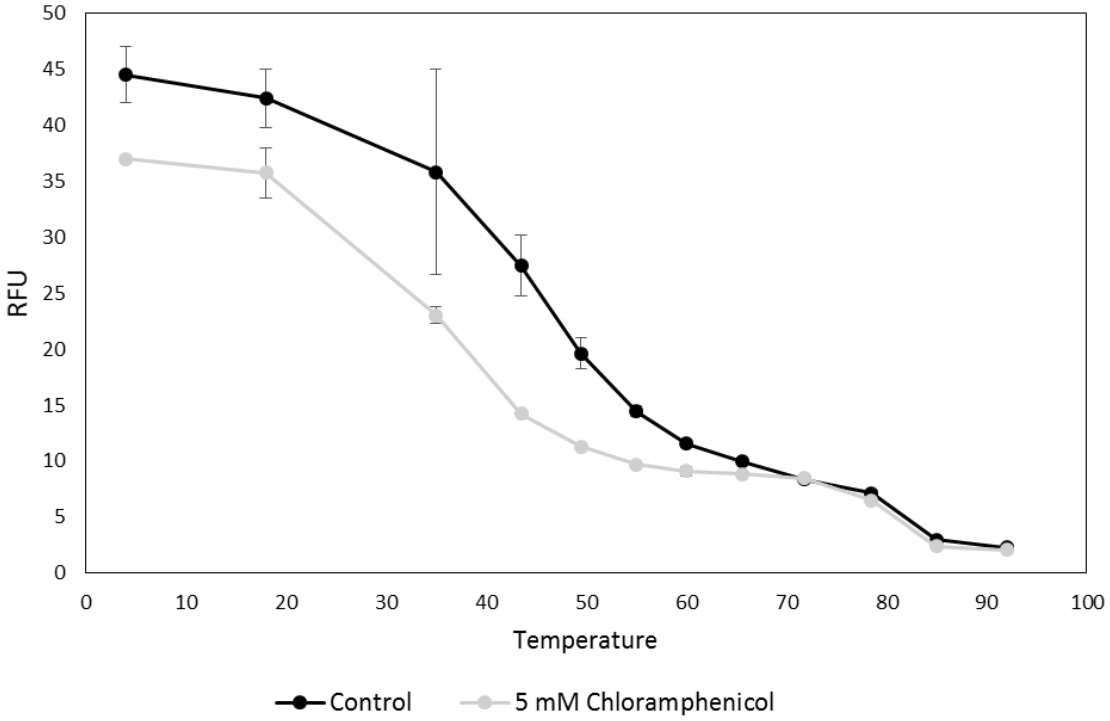
Table 4.6 Proteins expressed and tested for thermostability in the presence of ligands

Constructs were expressed in BL21 (DE3) cells. All MFS transporters are C-terminally sfGFP tagged. Method 1 used partial purification to isolate membrane proteins. Methods 2 and 3 used nickel affinity purification to isolate proteins. Samples were tested with and without noted ligands. sfGFP fluorescence was used to indicate the presence of sfGFP in samples.

Method	Construct	Ligands tested
1	H6-GFP	
1	H6-setB-sfGFP	lactose (0 & 5 mM)
1	H6-lacY-sfGFP	lactose (0 & 5 mM)
1	H6-xylE-sfGFP	sucrose (0 & 5 mM)
1	H6-mdfA-sfGFP	chloramphenicol (0 & 5 mM; 0 – 8.6 mM)
1	H6-araE-sfGFP	arabinose (0 & 5 mM)
1	H6-mdtM-sfGFP	chloramphenicol (0 & 5 mM)
1	H6-ydeA-sfGFP	cAMP (0 & 5 mM)

2	H6-GFP	chloramphenicol (0 - 1 mM)
2	H6-mdfA-sfGFP	chloramphenicol (0 - 1 mM), kanamycin (0 - 1 mM), TPP+ (0 - 1 mM), EtBr (0 - 1 mM), norfloxacin (0 - 1 mM)
2	H6-ydeA-sfGFP	chloramphenicol (0 - 20 mM)
2	H6-mdtM-sfGFP	chloramphenicol (0 - 5 mM)
2	H6-araE-sfGFP	arabinose (0 - 15 mM)
3	H6-mdfA-sfGFP	chloramphenicol (0 - 1 mM)
3	H6-hsrA-sfGFP	cysteine (0 - 1 mM)
3	H6-mdtM-sfGFP	chloramphenicol (0 - 1 mM)
3	H6-ydeA-sfGFP	cAMP (0 - 1 mM)

A



B

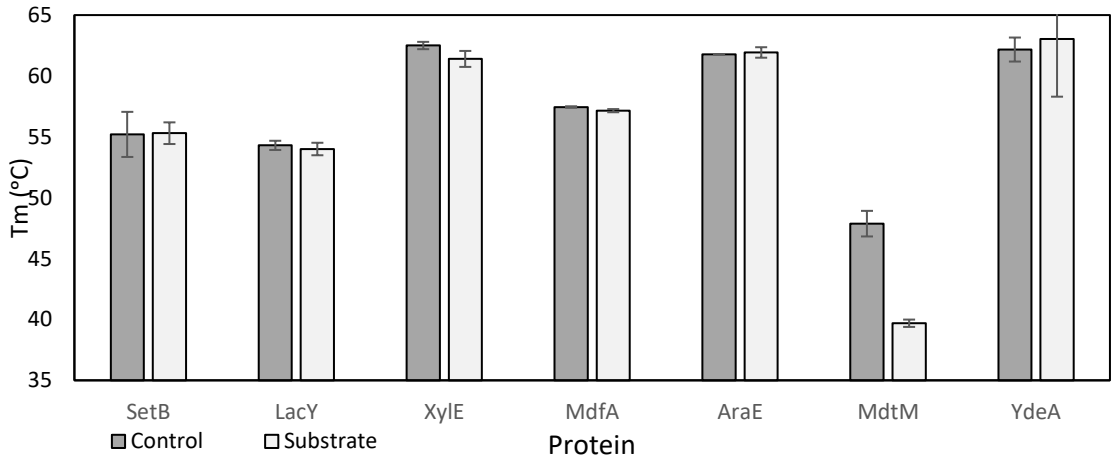
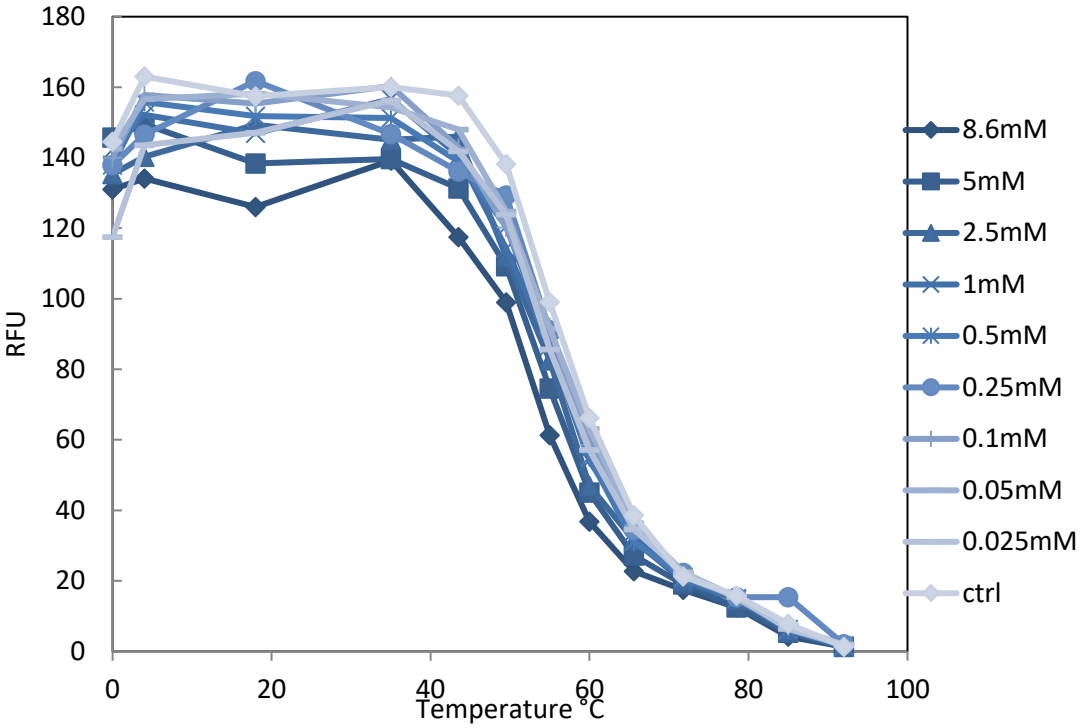


Figure 4.17 Thermoprofile melting temperatures (T_m) of H6-MFS-GFP proteins solubilised in detergent

H6-MFS-GFP proteins were expressed overnight in BL21 (DE3) at 18 °C. Following expression and lysis, the insoluble fraction was isolated by centrifugation; insoluble material was solubilised in detergent micelles. Fractions of whole cell lysate were incubated at a range of temperatures, and ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated protein was measured using spectrofluorimetry. A) Thermoprofile of mdtM in the presence (pale line) and absence

(dark line) of 5mM chloramphenicol; B) Thermoprofile melting temperatures (T_m; temperature at which RFU is half of maximum RFU) of MFS H6-transporter-sfGFPs, with (substrate; pale bar) and without (control; dark bar) 5 mM substrate. Substrates for each MFS are: SetB – lactose, lacY – lactose, xylE – sucrose, mdfA – chloramphenicol, araE – arabinose, mdtM – chloramphenicol, ydeA – cAMP. Each value is the mean of three measurements with average dispersion from the mean indicated by standard deviation.

A



B

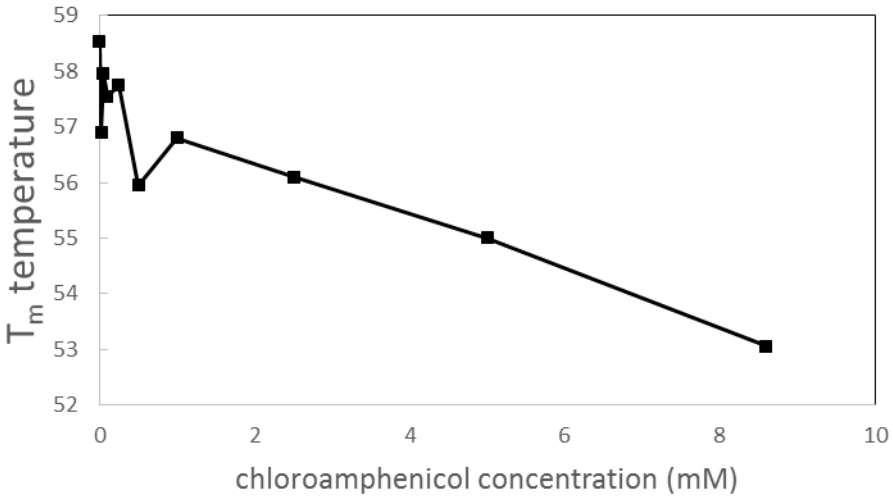


Figure 4.18 Thermoprofiles of detergent solubilised H6-mdfA-GFP protein with chloramphenicol substrate

H6-mdfA-GFP was expressed overnight in BL21 (DE3) at 18 ºC. Following expression and cell lysis, the insoluble fraction was isolated by centrifugation; insoluble material was solubilised in detergent micelles. Fractions of whole cell lysate were incubated at designated temperatures with chloramphenicol concentrations, shown in the key from 8.6 mM (darkest blue) to control with no chloramphenicol (palest blue). Following incubation,

samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. B) comparison thermoprofile melting temperatures (T_m; temperature at which RFU is half of maximum RFU) of H6-mdfA-sfGFP titrated with chloramphenicol. Each value is the mean of three measurements with average dispersion from the mean indicated by standard deviation.

As well as testing the method on transporters with known substrates, other unknown transporter/ligand combinations were tried. Among these were hsrA a MFS member with fourteen predicted TMH. The only publication specifically on this transporter identified that over-expression of hsrA resulted in the intracellular accumulation of homocysteine (Goodrich-Blair and Kolter 2000). This molecule is structurally related to cysteine (Figure 4.19). It is an intermediate in the synthesis of the amino acids cysteine and methionine. In *E. coli,* homocysteine is toxic at high levels (Tuite et al 2005) hence it must be tightly regulated. This could be achieved either through changes in metabolic pathways or efflux of homocysteine out of the cell. We could speculate that hsrA is involved in homocysteine regulation, but it is not likely to act as an efflux transporter as overexpression increases cellular homocysteine concentration (Goodrich-Blair & Kolter, 2000). Due to availability issues, the amino acid cysteine was tested instead of homocysteine. The thermoprofiles for H6-hsrA-sfGFP are presented in Figure 4.19.

Even though the H6-hsrA-sfGFP's thermoprofiles in the presence and absence of cysteine are very similar, there does appear to be a very slight positive in the presence of cysteine. If true, then the binding of the ligand to H6-hsrA-sfGFP stabilises the transporter/detergent/ligand complex in contrast to what was observed for H6-mdfA-sfGFP (Figure 4.18) which was destabilised. This is possible as each transporter will expose and hide different functional surfaces as it switches between conformational states. Depending on the detergent present this could either stabilise or de-stabilise the system.

A sugar transporter was selected for additional tests as these are influx transporters, unlike the antibiotic effluxing mdfA. The target that was selected was the *E. coli* xylose transporter xylE (Quistgaard et al 2013). Figure 4.20 shows the thermoprofiles for H6-xylE-sfGFP in the presence and absence of glucose. Glucose was selected as it is a known inhibitor of xylE (Sun et al 2012). As for hsrA, the addition of the binding ligand (glucose) causes a slight positive shift (red line) in comparison to H6-xylE-sfGFP alone (light blue line). However, as the error bars overlap there is no significant differences between bound and unbound.

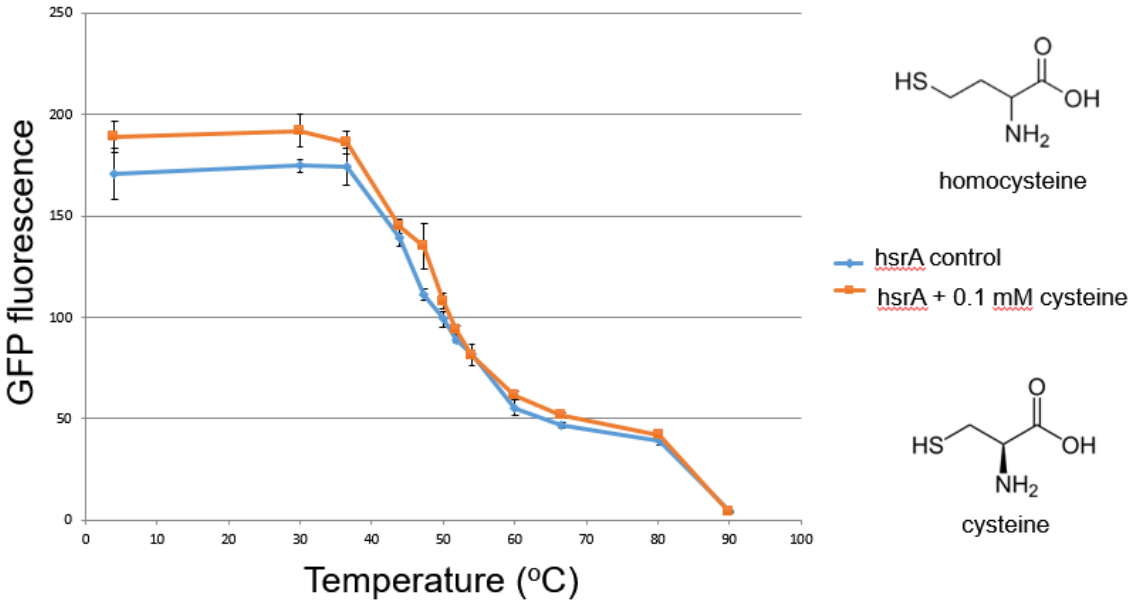


Figure 4.19 Thermoprofiles of H6-hsrA-GFP protein with cysteine, solubilised in detergent

H6-hsrA-GFP was expressed overnight in BL21 (DE3) at 18 ºC. Following expression and cell lysis, the insoluble fraction was isolated by centrifugation; insoluble material was solubilised in detergent micelles. Fractions of whole cell lysate were incubated at designated temperatures with 0 (blue) or 0.01 mM (orange) cysteine. Following incubation, samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. The data was collected and analysed with the help of Chris Ackroyd (Master Student, University of Southampton). Each value is the mean of three measurements with average dispersion from the mean indicated by standard deviation. The chemical structure of cysteine (the substrate utilised) and homocysteine (a previously identified substrate) are shown for comparison.

While testing another transporter for binding to zinc ions, it was noted that mM quantities of Zn^{2+} caused a dramatic destabilisation of the transporter. This is also seen with H6-xylE-sfGFP. Two values indicate this destabilisation: (1) the large negative T_m shift from 62 ˚C for WT to 51.5 ˚C for H6-xylE-sfGFP + Zn^{2+} (Figure 4.20, green line) and (2) a decrease in the total sfGFP fluorescence from over 600 units to below 400 RFU at lower temperatures. As the same amount of H6-transporter-sfGFP was used in both cases, the loss of sfGFP fluorescence must be a result of transporter precipitation. If glucose did bind to the H6-xylE-sfGFP transporter and stabilise it then the addition of glucose should reverse the destabilisation effect of zinc ions. This is seen in Figure 4.20 in which the precipitation is prevented in the presence of glucose (purple line from 30 to 45 ˚C). In addition, the T_m value for H6-xylE-sfGFP in the presence of zinc ions and glucose increases to 54 ˚C. In this case,

the glucose-bound form of xylE reduces the negative effect of zinc ions on the transporter by reversing the destabilisation, as indicated by the change in T_m and inhibition of protein precipitation.

Caution should be exercised with this interpretation as the error bars are large and the concentration of the ligands used are high. In addition, the 40 mM of glucose used would be very high in an environment. The observed protective effect of glucose may be a result of non-specific interactions rather than the specific binding to xylE as noted.

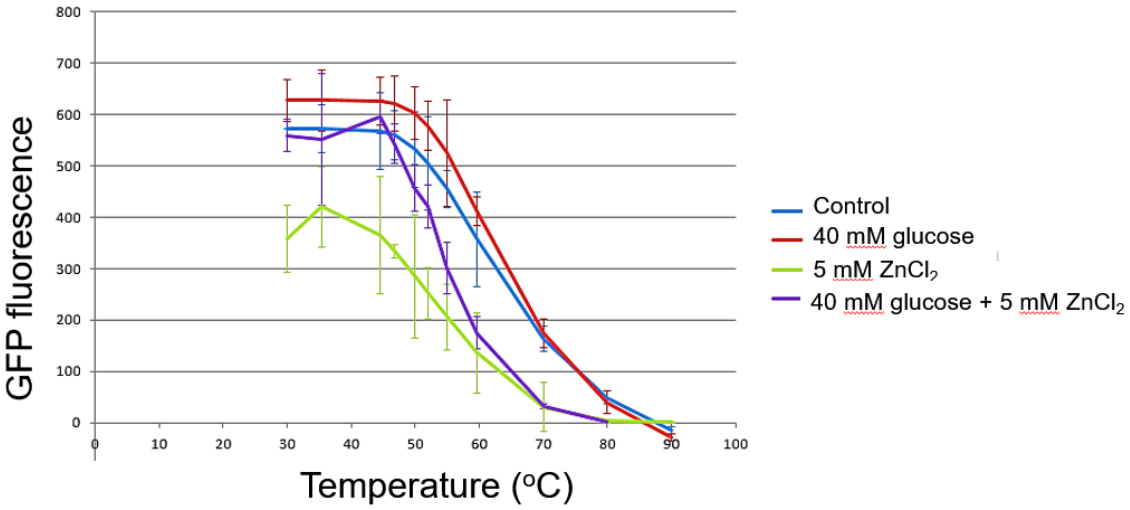


Figure 4.20 Thermoprofiles H6-xylE-GFP protein with glucose and zinc, solubilised in detergent

H6-xylE-sfGFP was expressed overnight in BL21 (DE3) at 18 °C. Following expression and cell lysis, the insoluble fraction was isolated by centrifugation; insoluble material was solubilised in detergent micelles. Fractions of whole cell lysate were incubated at designated temperatures with no substrate (blue), 40 mM glucose (red), 5 mM $ZnCl_2$ (green) or 40 mM glucose and 5 mM $ZnCl_2$ (purple). Following incubation, samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. The data was collected and analysed with the help of Andy Korentang (Master Student, University of Southampton). Each value is the mean of three measurements with average dispersion from the mean indicated by standard deviation. The chemical structure of cysteine (the substrate utilised) and homocysteine (a previously identified substrate) are shown for comparison.

4.2.2.2 Method 2: Ni^{2+}-IMAC purified H6-transporter-sfGFP

Again, H6-transporter-sfGFP was over-expressed using the previously identified optimised conditions. Post-expression, proteins were purified to remove soluble cellular content and membrane proteins unable to interact by Ni^{2+}-IMAC (Materials and Methods, Section 2.3.2). The isolated protein was then used to generate thermoprofiles in the presence of varying concentrations of potential substrates, in the same method (Materials & Methods, Section 2.5.3.1).

Figure 4.22 A and B show the SDS-PAGE and Western blot of purified H6-mdfA-sfGFP used for the thermoprofile assay, indicating the protein has been successfully purified. The purified H6-mdfA-sfGFP is highlighted in both gels. The Western blot reveals break down products below of H6-mdfA-sfGFP. Purification of H6-sfGFP is stable and is not proteolytically broken down, shown by a band of ~ 36 kDa (Figure 4.11). A band of this approximate size is present in (Figure 4.22 B) demonstrating some cleavage of the sfGFP tag from mdfA. The smaller bands that are picked up in the Western blot are either non-specific bands or further breakdown products of mdfA. In all cases these sfGFP-free fragments would not interfere with the measurements as they no longer visible without the sfGFP tag. The solubilised and Ni^{2+}-IMAC purified H6-mdfA-sfGFP was titrated with chloramphenicol (Figure 4.22 C).

A significant difference between detergent solubilised membrane thermoprofile and the Ni^{2+}-IMAC thermoprofile was the change in T_m. For Method 1 (detergent solubilised membranes containing over-expressed H6-mdfA-sfGFP) the T_m value averaged at approximately 57 °C while for Method 2 (Ni^{2+}-IMAC purified H6-mdfA-sfGFP) the T_m values increased to an average of 75 °C. Even though the average T_m value is higher, it is still below that of H6-sfGFP itself by ~ 13 °C. This indicates the results are dominated by the behaviour of the transporter rather than by the attached sfGFP domain. The increased T_m could be explained if other proteins were interacting with H6-mdfA-sfGFP in the partially purified membrane solubilised prep, thus destabilising the complex at lower temperatures.

Importantly, the behaviour of the partially purified H6-mdfA-sfGFP was the same as previously observed i.e., higher chloramphenicol concentrations result in a destabilisation of the system (Figure 4.22 C). This time the data did show saturation, a hall mark of specific binding. Assuming that, for this system, decreasing stability represented changes in ligand saturation it allowed for the construction of a protein/ligand binding curve and the calculation of the binding

constant Kd. Figure 4.21 shows the calculated ligand binding curve that display saturation and allowed for the determination of a binding constant Kd of 50 μM.

The transporter mdfA was selected as a positive control as it is a well-studied protein. As well as determining the *E. coli* mdfA structure in the presence of chloramphenicol, Heng et al, (2015) determined the Kd of mdfA-chloramphenicol as 75 μM. This is in reasonable agreement with our measurement.

Two additional Ni^{2+}-IMAC purified H6-mdfA-sfGFP thermoprofiles titrated against chloramphenicol were recorded (Figure 4.23 A). The Figure shows a change in T_m values for all of three H6-mdfA-sfGFP preparations (#1-3) titrated at pH 8.0. What is clear is that there is large variation in the initial T_m value, the final steady state T_m values and the maximum difference between unbound and fully bound states (mdfA #1 has a reduction in T_m of around 3 ºC between 0.025 mM chloramphenicol and 0.1 mM chloramphenicol while mdfA #2 has only a 1 ºC difference). mdfA #3 T_m changes with chloramphenicol concentration were highly variable. Disappointingly, a consistent change in T_m values was not replicable.

Since the mdfA/chloramphenicol complex structure was determined at pH 6.4 (Heng 2015) and the transport process is known to require protons, it was thought that pH may play a significant role in process. The screen was repeated at the lower pH values of 6.0 and 7.2, with HEPES replacing Tris as part of the purification buffers. Figure 4.23 B shows that, apart from pH 8.0, there is essentially no change in T_m values with increasing chloramphenicol concentrations. At pH 7.2, there is a reduction in T_m above 0.5 mM chloramphenicol, which may suggest a reduction in protein stability, but this was generally an inconsistent result.

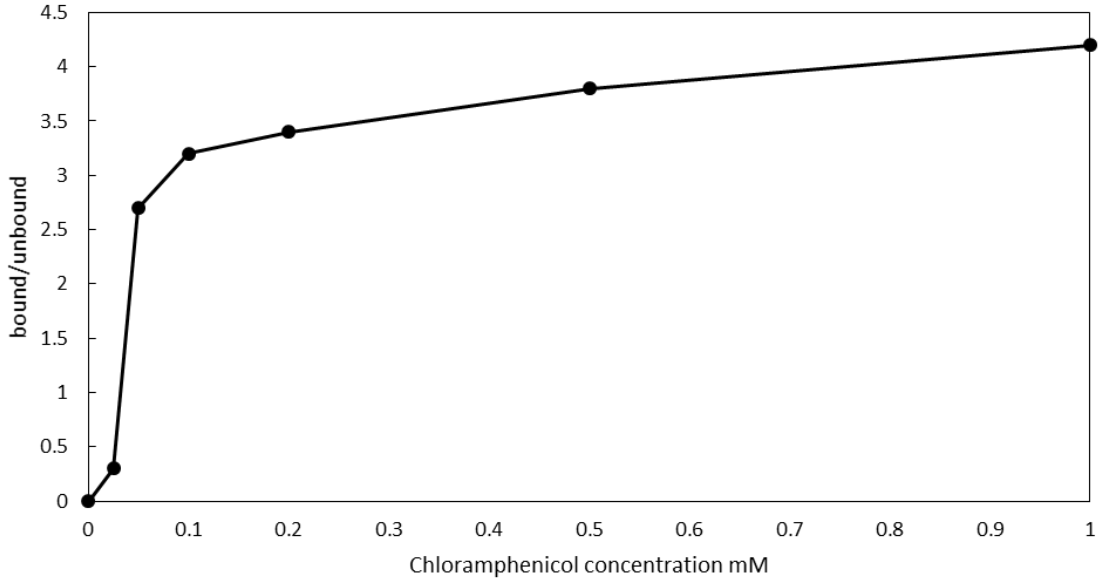


Figure 4.21Binding curve of mdfA with chloramphenicol, generated using thermoprofile T_ms from Ni^{2+}-IMAC purification

H6-mdfA-sfGFP was expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified protein was incubated at designated temperatures with 0, 0.025, 0.05, 0.1, 0.2, 0.5 and 1 mM chloramphenicol. Following incubation, samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. The thermoprofile T_m was calculated as the temperature at which the sfGFP concentration was at half the maximum. A curve was generated by plotting T_m against chloramphenicol concentration. Each value is the mean of three measurements.

To further determine whether the inconsistencies were due to the method being used or the chloramphenicol-binding ability of mdfA, alternative controls were sought: kanamycin, ethidium bromide (Etbr), norfloxacin and tetraphenylphosphonium (TPP+). Radiolabelled TPP+ has been shown to bind to mdfA (Fluman et al 2009). Overexpression of mdfA increases the minimum inhibitory concentration (MIC) of TPP+, norfloxacin and EtBr, which suggests mdfA may be able to transport these substrates. However, its overexpression has no effect on the MIC of kanamycin (Nishino et al 2001) suggesting it is not a mdfA substrate and is therefore a suitable negative control.

Figure 4.23 C shows the thermoprofile of mdfA titrated with each of these substrates. There does not appear to be a correlation between Etbr or norfloxacin concentration and T_m value. However, there does appear to be a reduction in protein stability with increasing TPP+ concentrations, with a maximum reduction in T_m value of around 4 ºC between 0.1 and 0.25 mM TPP+. This could suggest TPP+

binding to H6-mdfA-sfGFP has been detected. As mentioned above, kanamycin is not believed to bind or be transported by mdfA. However, there is an increase in T_m values up to 0.1 mM before it returns to baseline. Although evidence suggests that mdfA does not transport kanamycin (Nishino et al 2001) this does not rule out non-specific binding. The rise and fall of the T_m values, however, suggests that the assay is generally inconsistent. Repeat experiments may have helped clarify this issue, instead, additional transporters were tested against known transport substrates. The idea behind this broad approach was to counteract the issue that the observed inconsistency was not due to the methodology but possibly something specific to do with the transporter mdfA.

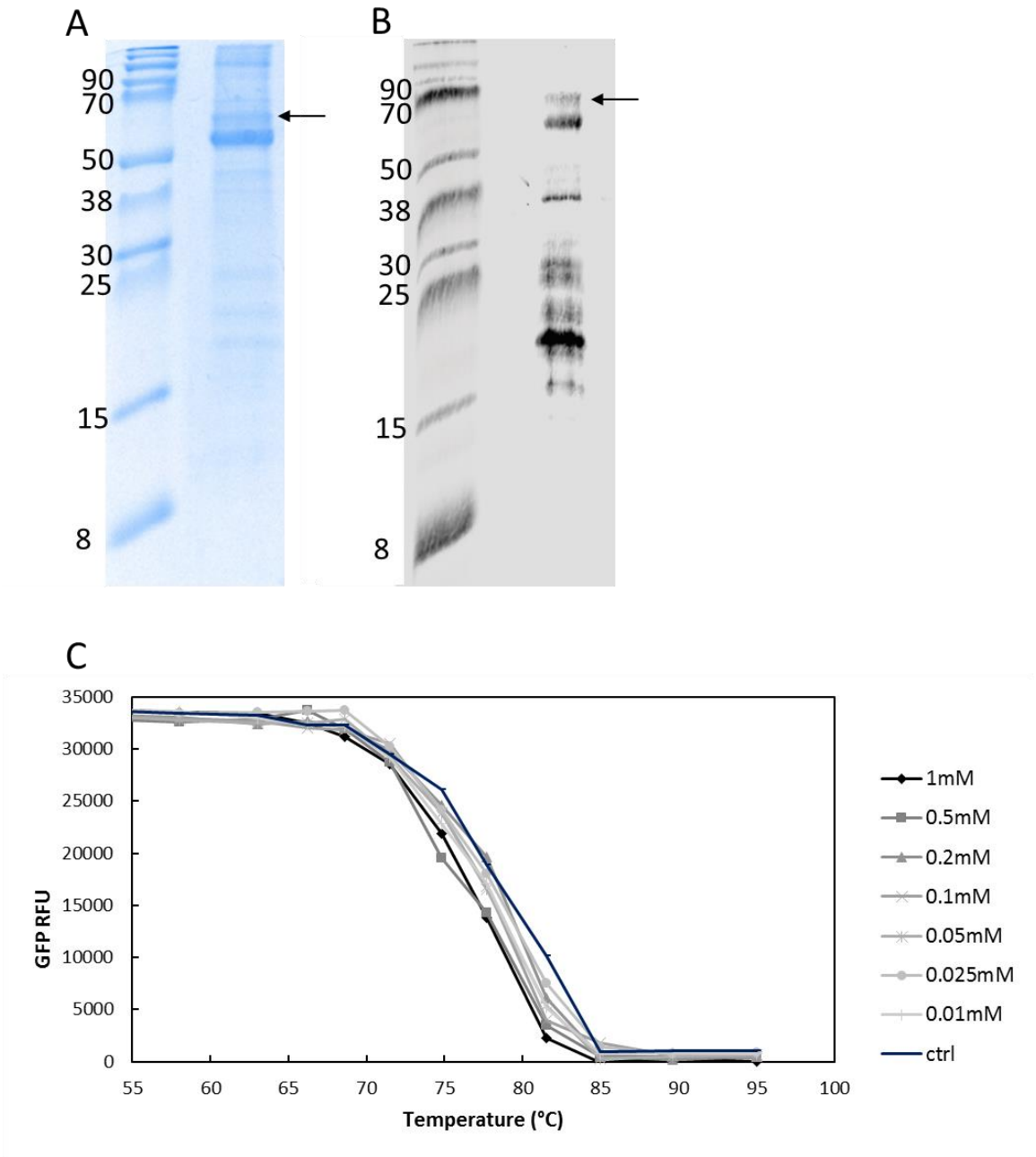


Figure 4.22 Thermoprofiles of Ni^{2+}-IMAC-purified H6-mdfA-GFP protein with chloramphenicol

H6-mdfA-GFP was expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). A-B) Purification was analysed using SDS-PAGE (A) and western blot (B) with primary antibodies for the His-tag. Arrows indicate predicted product of H6-mdfA-GFP (72.2 kDA). Molecular ladder sizes displayed on the left of each gel (kDa). C) Fractions of purified protein were incubated at designated temperatures, with a range of chloramphenicol concentrations, shown in the key, from 0.01 mM (pale grey) to 1 mM (dark grey); control (ctrl) in blue. Following incubation, samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. Each value is the mean of three measurements.

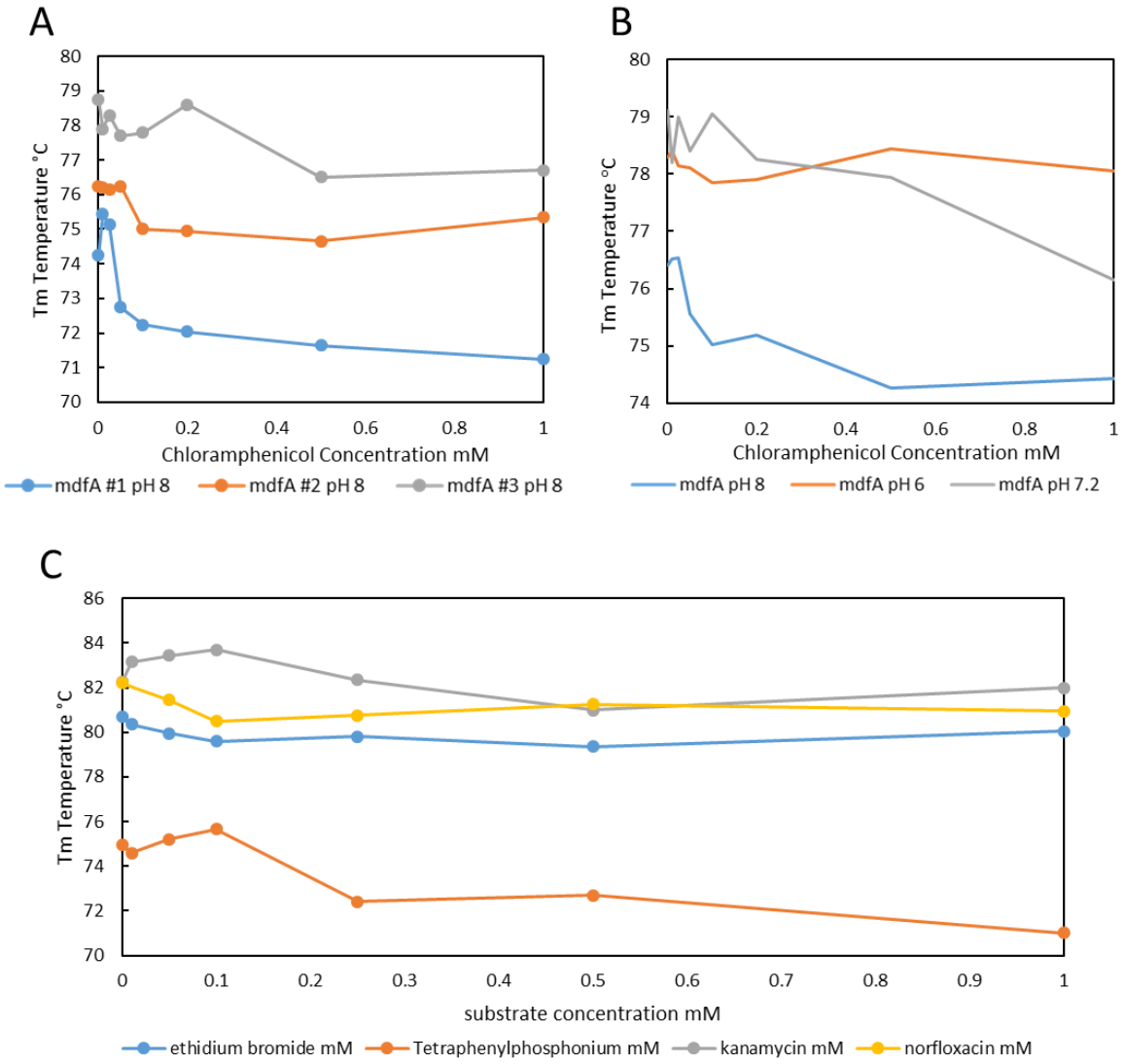


Figure 4.23 Binding curve generated using thermoprofile T_ms from Ni^{2+}-IMAC-purified H6-mdfA-sfGFP, with specific substrates

H6-mdfA-GFP was expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified protein was incubated at designated temperatures with 0, 0.025, 0.05, 0.1, 0.25, 0.5 and 1 mM of either chloramphenicol (A, B) or C) ethidium bromide, tetraphenylphosphonium, kanamycin or norfloxacin. Following incubation, samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. The thermoprofile T_m was calculated as the temperature at which the sfGFP concentration was at half the maximum. A curve was generated by plotting T_m against substrate concentration. A) Thermoprofile T_ms of three independent H6-mdfA-sfGFP preparations at pH 8; B) Thermoprofile T_ms of three independent H6-mdfA-sfGFP preparations, prepared with three buffer pH values; C) Thermoprofile T_ms of purified mdfA

with positive controls: tetraphenylphosphonium, norfloxacin and ethidium bromide and negative control kanamycin. Each value is the mean of three measurements.

In addition to mdfA, the *E. coli* proteins araE, ydeE and mdtM were purified and investigated using Method 2. The substrates used in the thermoprofile assays are summarised in Table 4.6, and the thermoprofiles displayed in Figure 4.24.

The thermoprofile of H6-araE-sfGFP is shown in Figures 4.24 A, B and C, using arabinose as a positive control and chloramphenicol as a negative control. If araE-sfGFP behaved in a similar fashion to H6-mdfA-sfGFP then we would expect that the addition of a binding substrate would result in destabilisation of the transporter/detergent/substrate complex with a resulting decrease in T_m values. A sharp reduction in T_m values is observed for H6-araE-sfGFP in the presence of arabinose, with it remaining constant (saturated) from 0.05 mM to 1 mM arabinose (Figure 4.24 A). This is also seen when using an extended arabinose concentration range (Figure 4.24 B), however, this effect is reversed above 5 mM arabinose as the T_m values begins to increase again. A generous interpretation of these results is that destabilisation of the system is concentration dependent and saturable up to 1 mM arabinose but at higher arabinose concentrations, the non-specific effect of sugars stabilising proteins reverses the destabilisation. However, again the initial T_m values at 0 mM arabinose were inconsistent (Figures 4.24 A showing 81 ˚C and B displaying 76.3 ˚C) and the final arabinose concentration at which they display the maximum negative shift as compared to 0 mM differs (1 mM or 5 mM, Figures 4.24 A and B, respectively). As expected, H6-sfGFP is a negative control when titrated with chloramphenicol which showed no significant changes between T_m with increasing chloramphenicol concentrations.

Neither ydeA nor araE are believed to bind and/or transport chloramphenicol. Testing these two sfGFP-tagged transporters against chloramphenicol displayed negative T_m shifts in both cases (Figure 4.24 C). Both results appear to be linear, possibly indicating non-specific destabilisation of the detergent solubilised transporters by the hydrophobic chloramphenicol.

Chloramphenicol was used as a positive control for mdtM (Alegre et al 2015), which does appear to have a consistent, saturable and reduced T_m with elevated chloramphenicol concentrations (Figure 4.24 D), a hallmark of specific binding. This was a very exciting result, however, the general destabilisation of chloramphenicol as observed in Figure 4.24 C would have to be taken into account.

As, in this case, this method is not able to distinguish between the binding of a specific substrate and non-specific destabilisation. Using soluble substrates such as arabinose also highlighted some of the limitations as the results were inconsistent and jumped between destabilising the system at low concentrations to stabilisation at higher concentrations (Figures 4.24 A and B).

An inverse sigmoid function on the thermoprofiles should be seen to infer a specific interaction between protein and substrate, which would indicate saturation of the binding sites. A linear correlation between protein and substrate concentration indicates a nonspecific interaction. Meanwhile, any thermoprofile which shows no correlation between T_m and substrate concentration suggests no substrate binding is detected. Overall, the mixed findings of Method 2 are summarised in Table 4.7. It is possible that with further optimisation and modifications, TPP+ and chloramphenicol could be used as substrate-binding positive controls for mdfA and mdtM, respectively. To improve the assay, an increased number of temperature points could be taken with multiple repeats per experiment. In summary, the results from Method 2 have been inconsistent, suggesting this method is not appropriate for demonstrating protein-substrate interactions.

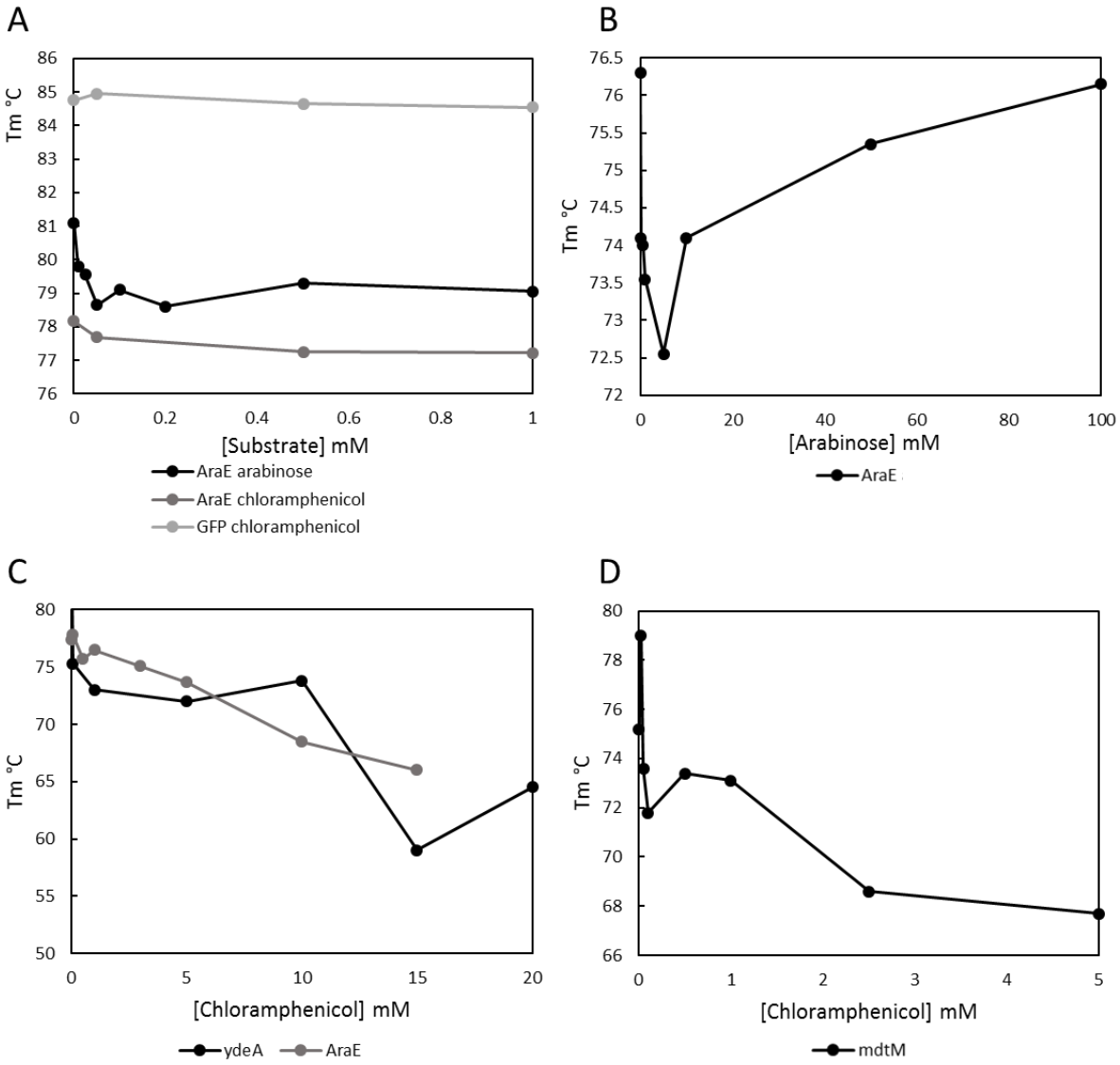


Figure 4.24 Binding curve generated using thermoprofile T_ms from Ni^{2+}-IMAC-purified H6-MFS-sfGFP with specific substrates

H6-MFS-GFP proteins (araE, ydeA, mdtM) and H6-sfGFP were expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified protein was incubated at a range of temperatures with a range of arabinose or chloramphenicol concentrations. Incubation samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using spectrofluorimetry. The thermoprofile T_m was calculated as the temperature at which the sfGFP concentration was at half the maximum. A curve was generated by plotting T_m against substrate concentration. A) Thermoprofile T_ms of H6-araE-sfGFP with arabinose, H6-araE-GFP and H6-sfGFP with chloramphenicol. B) Thermoprofile T_ms of H6-araE-sfGFP with arabinose. C) Thermoprofile T_ms of H6-araE and H6-ydeA with chloramphenicol. D) Thermoprofile T_ms of H6-mdtM-sfGFP with chloramphenicol. Each value is the mean of three measurements.

Table 4.7 Summary of the thermoprofile investigations

Protein	Result with 'positive' control	Result with 'negative' control	Notes
mdfA	Inconclusive with chloramphenicol; possible reduction in stability at pH 7.2 but inconsistent No change in stability with EtBr or norfloxacin Reduction in stability seen with TPP+ at lower concentrations	Decreased stability seen with kanamycin	TPP+ could be used as positive control Kanamycin should be a negative control (no effect) but results may display non-specific effects
araE	Reduction in stability seen with arabinose at lower concentrations which was reversed above 5 mM	Decreased stability seen with chloramphenicol	Chloramphenicol should be a negative control (no effect) but results may display non-specific effects
ydeA	n/a	Decreased stability seen with chloramphenicol	Chloramphenicol should be a negative control (no effect) but results may display non-specific effects
mdtM	Decreased stability seen with chloramphenicol	n/a	Chloramphenicol is a positive control
GFP	GFP was used as a protein control. Substrate tested showed no changes in T_m values		GFP is a negative control

4.2.2.3 Method 3: Ni^{2+}-IMAC purified H6-transporter-sfGFP T_m hold assay

Method 3 attempts to rectify potential issues associated with Method 1 and 2 that were attributed to a lack of reproducibility. In the generation of the previous temperature shift assays, the proteins were held at the fixed temperatures for 3 mins before spinning and recording remaining protein via sfGFP fluorescence. The 3-minute time interval was selected arbitrarily. The T_m value represents the half way temperature between the maximum and minimum stabilised protein. At this temperature the protein is precipitating, it is unstable. If a protein were held at its T_m value for extended periods, the protein would be expected to continue to precipitate. Method 3 exploits this idea by holding the H6-transporter-sfGFP at its known T_m in the presence and absence of substrate. If as previously observed substrates can stabilise or destabilise the system, then this should be observed using this approach.

The sfGFP-tagged transporters were prepared as previously described and titrated against a selection of potential substrates (Table 4.6). The methodology is described in Material and Methods Section 2.5.3.3. In addition to a 4 ˚C control representing a stable sample of the protein, a T_m test point was chosen based on the findings from Method 2. The protein was held at the desired temperature for 15 minutes, before returning to 4 ˚C to prevent further precipitation. Precipitated protein was removed by centrifugation. From here the sfGFP fluorescence was measured using a spectrofluorometric plate reader, to determine whether substrate concentration has any effect on RFU. Method 3 is referred to as the T_m hold assay. Another potential advantage for this approach was a reduction in the workload, allowing for greater replication and which may therefore improve the reliability of the technique.

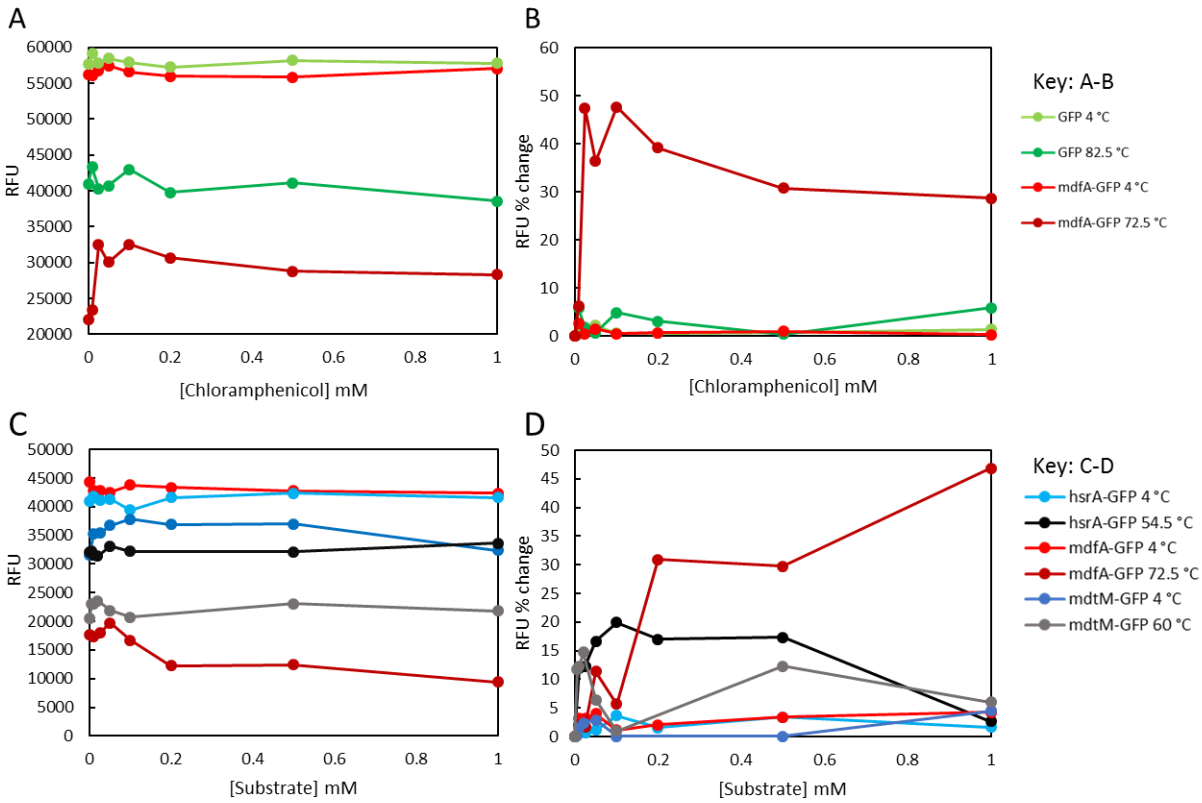


Figure 4.25 Curve generated using T_m hold assay with Ni^{2+}-IMAC-purified H6-MFS-sfGFP and H6-sfGFP, with specific substrates

H6-MFS-GFP proteins (mdfA, hsrA, mdtM) and sfGFP were expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified protein was incubated at either 4ºC or the temperature calculated to be the protein aggregation temperature (indicated in graph key) with a range of substrate concentrations; mdfA and mdtM: chloramphenicol; hsrA: cysteine. Incubation samples were ultracentrifuged to separate aggregated protein. The sfGFP concentration of unaggregated sample was measured using a microplate reader. A and C_ raw RFU values. B and D) percentage change in RFU, compared to value at 0 mM substrate. A-B) Positive control H6-mdfA-sfGFP and negative control H6-sfGFP with chloramphenicol; C-D) Positive controls H6-mdfA-sfGFP and H6-mdtM-sfGFP with chloramphenicol, and H6-hsrA-sfGFP with cysteine. Each value is the mean of nine measurements.

Results of the T_m hold assay are shown in Figure 4.25. Raw RFU values are presented in Figures 4.25 A and C, with the percentage change in RFU shown in Figures 4.25 B and D. The negative control, purified sfGFP, remained relatively unchanged at either temperature with the addition of chloramphenicol. Purified H6-mdfA-sfGFP remained unchanged at the control 4 ºC as expected, but when exposed to 72.5 ºC in the T_m-hold assay, its RFU was greatly increased above 0.05 mM chloramphenicol. The greater change in RFU indicating an increase in stability with the occupation of

the mdfA binding site by chloramphenicol (Figure 4.25 A). This was a surprising result, as previously all other experimental methods indicated that, using this system, chloramphenicol binding destabilised H6-mdfA-sfGFP.

Although this finding was promising, as with Method 2 it could not be repeated. Instead, when heated to the test temperature, an increase in RFU was observed for H6-mdfA-sfGFP below 0.05 mM chloramphenicol, followed by a sharp decrease in RFU above this concentration (Figure 4.25 C & D). Additionally, two other potential positive controls were tested, mdtM with chloramphenicol and hsrA with cysteine. The increase in RFU observed for hsrA with elevated cysteine could provide evidence of hsrA-cysteine binding and increasing H6-hsrA-sfGFP stability, while H6-mdtM-sfGFP did not show such an obvious correlation. This could indicate a stabilising effect of cysteine as a substrate for hsrA, importantly the sigmoidal shape indicates a specific interaction. In all cases, 4 ºC remained a useful negative control, with little change in RFU observed with increased substrate availability.

The opposite trends for mdfA samples observed between the two repeats would indicate both a stabilising and destabilising effect. Therefore, as with previous methods, this shows a lack of consistency. It is possible that the system is in fact working well but that H6-mdfA-sfGFP is misfolded, and therefore incapable of binding its known substrate, chloramphenicol. Alternative methods should be sought to determine protein-substrate binding and demonstrate protein viability. Alternatively, it is possible that the MFS transporters are binding their proposed substrates, but not generating a signal that can be measured with available spectrofluorometric equipment or is being masked by background noise. To test this, the same sample of prepared H6-mdtM-sfGFP was split between eight different wells in a spectrofluorometric plate reader (POLARstar Omega spectrofluorometer plate reader (BMG LABTECH, UK)) and sfGFP fluorescence recorded. RFU for these samples varied by ~14 % (Figure 4.26 A). This compares to the maximum and minimum change in T_m value for H6-mdfA-sfGFP with chloramphenicol titration of ~ 10,000 RFU. This level of background noise added to the issues of pipetting variability may therefore be hiding any real effects of the assay; Method 3 cannot be reliably and accurately performed using the available protocol and equipment.

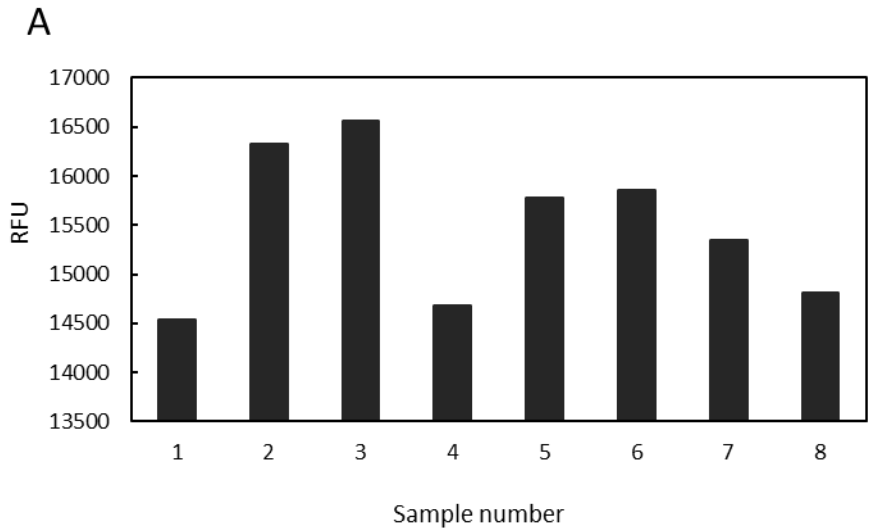


Figure 4.26 Noise tests of POLARstar Omega spectrofluorometer plate reader

mdtM was expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified protein was split between multiple plate wells and measured in parallel using a microplate reader. Each value is a single measurement.

4.2.2.4 Method 4: Substrate binding assays using the Nanotemper Monolith NT.115

Methods 1-3 investigated protein-substrate interactions by measuring fluorescence and the corresponding protein stability via thermoprofile assays or the T_m hold assay. However, although providing some promising results, these methods were not consistent or reproducible.

One potential alternative method used to identify protein-substrate interactions is microscale thermophoresis. This works by setting up a localised temperature gradient to affect the movement of molecules. Heated molecules have greater kinetic energy compared to cooler molecules, meaning molecules from the heated area move away faster than they are replenished. This leads to a localised concentration gradient, with fewer molecules in the warm area compared to the cooler. Fluorescence from a GFP tag can be measured using the Nanotemper Monolith NT.115 to visualise the dynamics of this concentration gradient, in the presence of varying concentrations of a potential substrate. The dynamics of changes in fluorescence correspond to changes in protein kinetics, which can be

translated to protein-substrate binding, Figure 4.27 shows a visual representation of this process.

The addition of a ligand that binds to the target protein changes its mobility properties so long as there is a significant change in the protein. This can have either a positive or negative effect on movement depending on the target protein/substrate complex. As MFS transporters are known to undergo significant conformational changes upon ligand binding, this makes them good candidates for this approach. The added advantage of this methodology is that it is substrate independent as the monitored property is the thermal mobility of the target protein.

Chloramphenicol has been previously characterised as a ligand for mdfA and mdtM (Heng et al, 2015; Algre et al, 2016). mdfA is also known to bind TPP+ (Lewinson et al, 2001). Methods 1-3 in this Chapter used Brij35 as a detergent, but mdfA and mdtM were previously characterised using the detergents DM (n-decyl-β-D-maltopyranoside) and DDM (n-dodecyl-β-D-maltopyranoside), respectively (Lewinson et al, 2001; Heng et al, 2015; Algre et al, 2016). To ensure proper comparisons, this method was set up using the detergents and buffers outlined in the published papers. This is summarised in Table 4.8 along with the substrate tested.

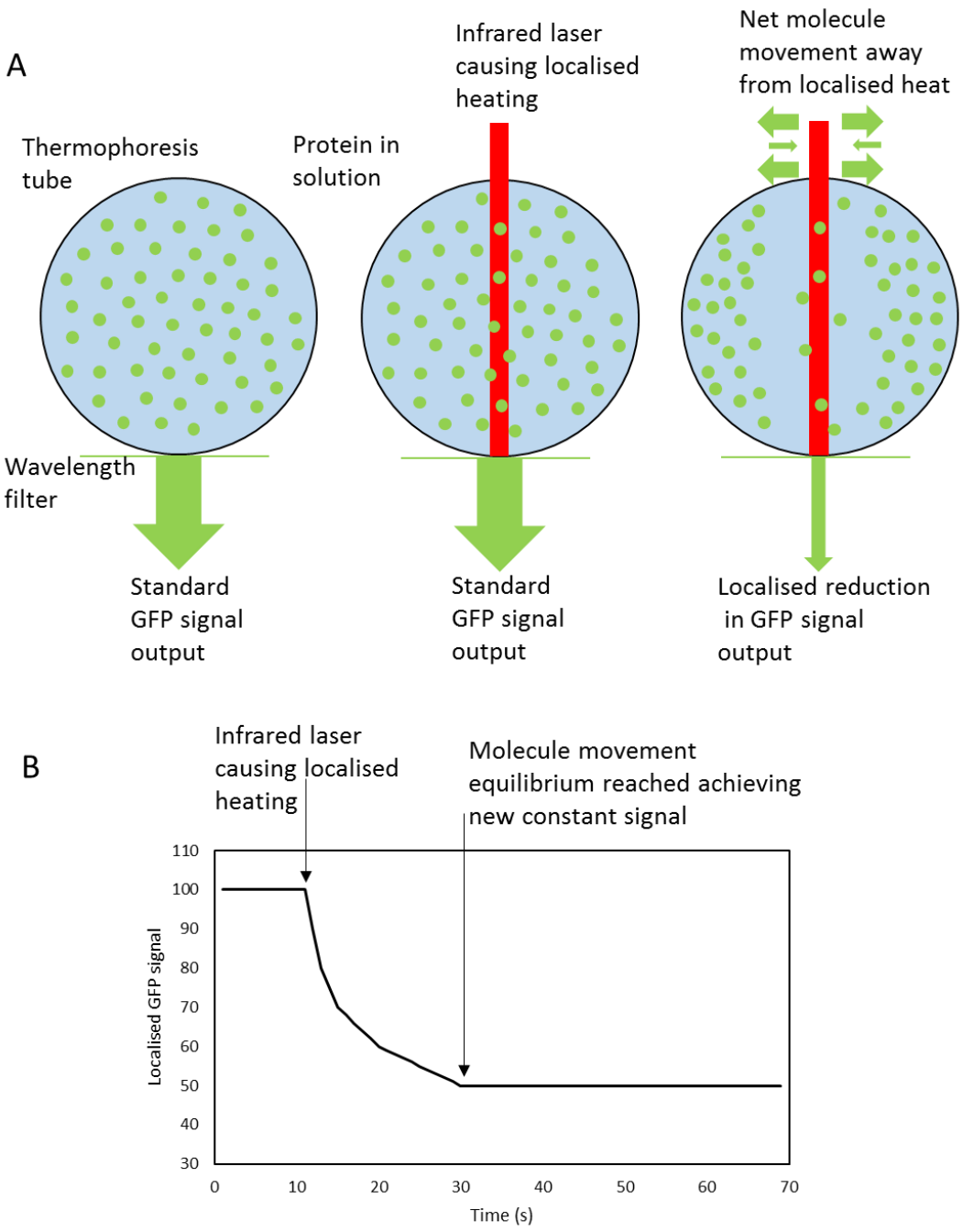


Figure 4.27 Hypothetical output of a microscale thermophoresis experiment

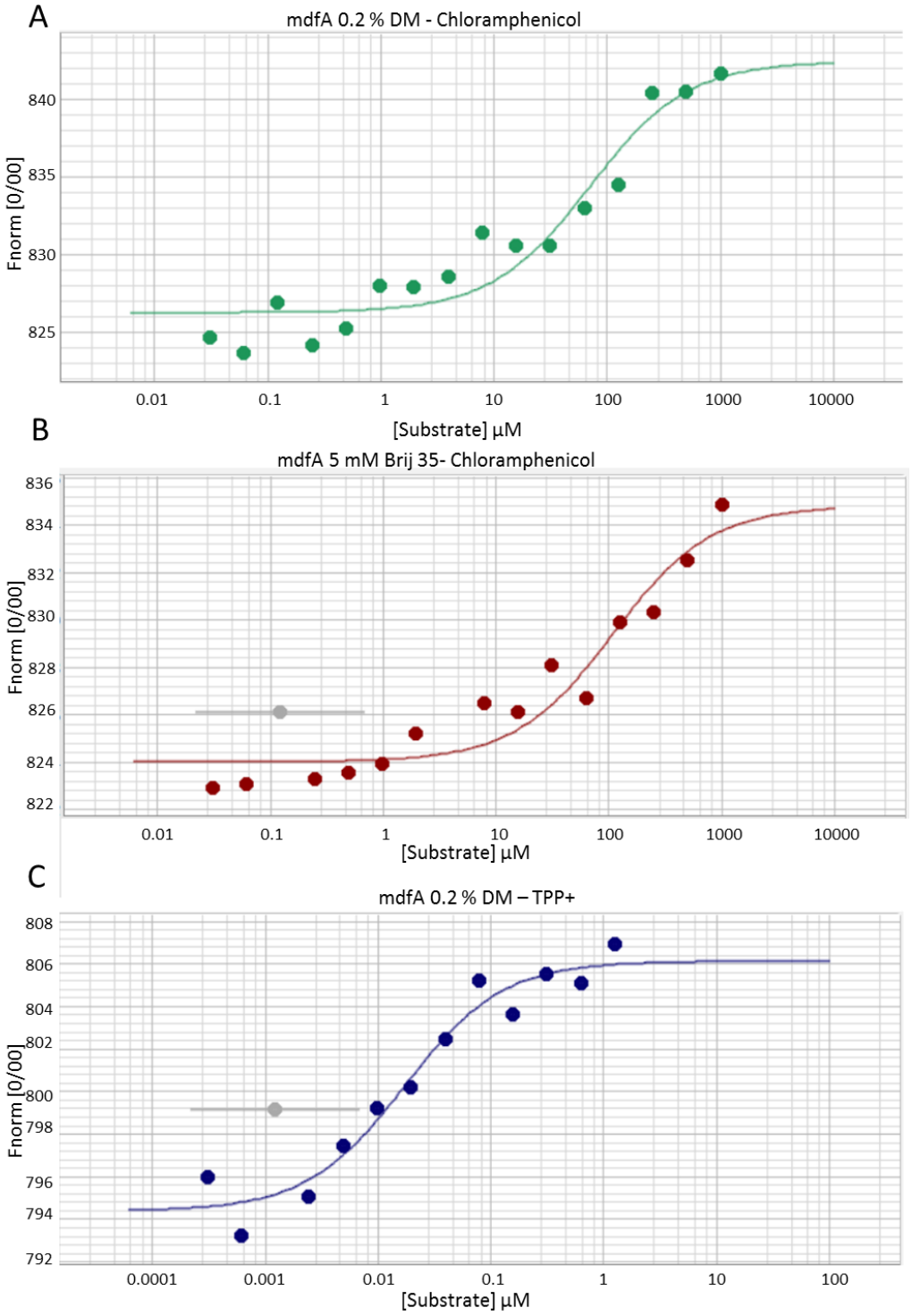
A) Key steps in a microscale thermophoresis experiment. The GFP in a protein sample in a tube is excited and measured through a filter to calculate the concentration of GFP (standard signal output). An infrared laser causes localised heating of the tube sample. Initially this does not affect the signal output, but due to the disparity this causes in kinetic energy, molecules moving away from the heated area are faster than those replacing them, and a net movement away from the localised heat. This leads to a localised reduction in GFP signal output. B) The key steps of the microscale thermophoresis experiment, shown graphically. The GFP signal drops off sharply as the area is heated. Eventually an equilibrium is met producing a constant, reduced signal.

Figure 4.28 shows the effect of substrate concentration on thermophoresis of H6-mdfA-sfGFP (Figure 4.28 A-C, F), H6-mdtM-sfGFP (Figure 4.28 D) and H6-hrsA-sfGFP (Figure 4.28 E). Fnorm is the normalised ratio of localised fluorescence in the heated area, compared with the unheated area, plotted against increased substrate concentration on a logarithmic scale. Each plot follows a similar pattern. At low substrate concentrations, there is insufficient substrate-protein binding to affect protein kinetics on an observable scale. However, as the substrate concentration increases and begins to bind the protein, the kinetics of the sfGFP-tagged protein changes, leading to a quantifiable difference in Fnorm. Eventually, all Fnorm values begin to plateau. This is because all binding sites are saturated, so an increase in substrate concentration has no further effect on protein kinetics. This plateau indicates a specific interaction between substrate and protein.

Table 4.8 List of targets and ligands for the ligand binding determination using the Nanotemper Monolith

The detergent used in the solubilisation and purification process are indicated. Further details on buffers can be found in Materials and Methods (2.5.4).

Protein	**Detergent**	**Fluorescent label**	**Substrate**
mdfA	0.2 % DM	sfGFP-fusion	chloramphenicol
mdfA	5 mM Brij35	sfGFP-fusion	chloramphenicol
mdfA	0.2 % DM	sfGFP-fusion	TPP+
mdtM	0.05 % DDM	sfGFP-fusion	chloramphenicol
hsrA	2 mM DDM	sfGFP-fusion	cysteine


A
mdfA 0.2 % DM - Chloramphenicol
Fnorm [0/00]
840
835
830
825
0.01
0.1
1
10
100
1000
10000
[Substrate] µM
B
mdfA 5 mM Brij 35- Chloramphenicol
Fnorm [0/00]
836
834
832
830
828
826
824
822
0.01
0.1
1
10
100
1000
10000
[Substrate] µM
C
mdfA 0.2 % DM – TPP+
Fnorm [0/00]
808
806
804
802
800
798
796
794
792
0.0001
0.001
0.01
0.1
1
10
100
[Substrate] µM

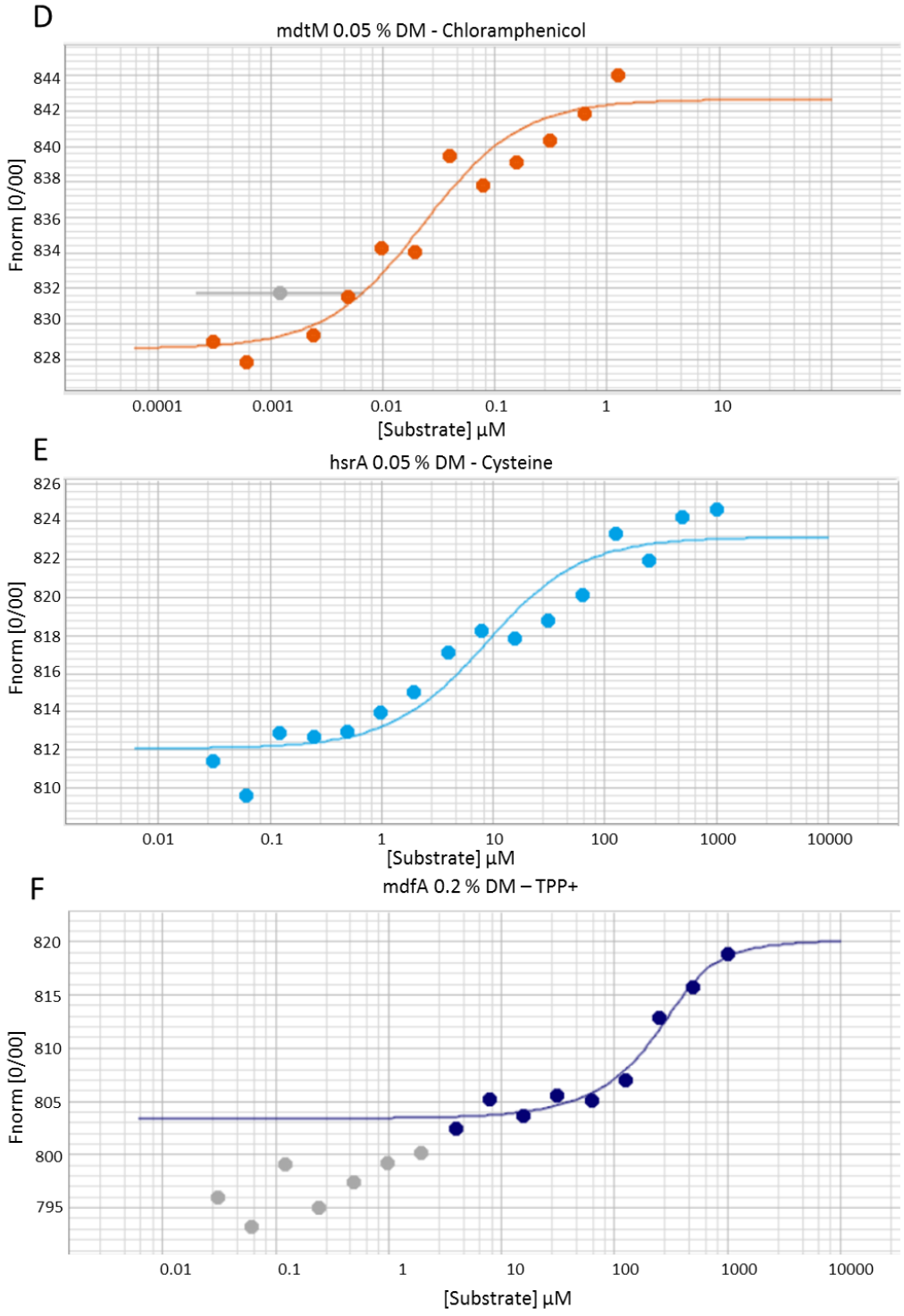


Figure 4.28 Ligand-binding curves demonstrating an interaction between Ni^{2+}-IMAC purified H6-MFS-sfGFP and specific substrates using thermophoresis

H6-MFS-GFP proteins (mdfA, hsrA, mdtM) were expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified samples were split into thermophoresis tubes containing different concentrations (indicated by the x-axis) of substrate (indicated in the graph heading). The Nanotemper Monolith NT.115 was used for thermophoresis experiments and analysis. Grey circles on graphs B, C and D

were points rejected after initial tube scans due to high variance from the mean start value. A) mdfA in 0.2 % DM buffer with chloramphenicol titre 30 nm – 1 mM; B) mdfA in 5 mM brij35 buffer with chloramphenicol titre 30 nm – 1 mM; C) mdfA in 0.2 % DM buffer with TPP+ titre 0.3 nm – 10 µM; D) mdtM in 0.05 % DDM buffer with chloramphenicol titre 0.3 nm – 10 µM; E) hsrA in 2 mM DDM buffer with cysteine titre 30 nm – 1 mM; F) mdfA in 0.2 % DM buffer with TPP+ titre 0.3 nm – 10 µM. Each value was obtained from a single measurement.

Based on the thermophoresis method, all tested samples displayed binding events. If true, this implies that the expression and purification of the H6-MFS-sfGFP transporters produce viable protein for binding assays, suggesting that it is the detection method for Methods 1-3 that was the issue rather than the sample preparation. Binding affinity can be inferred from these values, using Nanotemper Monolith NT.115 software. The determined values were compared to published data, where available (Table 4.10). When using the same detergent as published, the calculated binding affinity displayed comparable values to those published. This was particularly true for H6-mdfA-sfGFP in 0.2 % DM with chloramphenicol, and H6-mdtM-sfGFP in 0.05 % DDM with chloramphenicol. Using the detergent Brij35 for the H6-mdfA-sfGFP with chloramphenicol experiments instead of DM as a detergent produced a binding event (108 µM) that was similar to published data. The difference is not surprising as different detergents have different properties which impacts on the functionality of solubilised membrane proteins. The Kd value for chloramphenicol binding to H6-mdfA-sfGFP in DM (69 µM) as compared to Brij35 (108 µM) may represent a slight conformational change in the binding site. Repeat measurements would allow for a better comparison.

A major finding was that H6-mdfA-sfGFP displayed an additional binding event for the TPP+ ligand at lower concentrations, which has not been previously observed. For our system, two binding affinity events were observed: one with a Kd of 20 nM, Figure 4.28 C, and the second of 0.7 µM, Figure 4.28 Sigal et al (2007) calculated the Kd of TPP+ for mdfA in DDM to be 2.9 µM. This value is closer to the second Kd that was determined but is still significantly different. This may reflect the different detergents that were used: DM in our case and DDM for Sigal et al (2007).

If mdfA truly has two TPP+ binding events, then this could represent either two separate binding events or overlapping sites that change affinity as the transporter changes conformation upon ligand binding. As the observed cavity is large in the mdfA/chloramphenicol crystal structure (Figure 4.2 A), the most likely explanation is that there are two binding sites; one in the same location as the bound

chloramphenicol and the other elsewhere in the cavity. A series of mutations of the suspected amino acids in the cavity with ligand binding studies should be able to identify this new site.

Another major finding comes from the H6-hsrA-sfGFP binding data. This is a MFS transporter with unknown function that has previously been tested for antibiotic efflux, but with no antibiotic ligand determined (Nishino et al 2001). hsrA was selected for testing as it is predicted to have 14 transmembrane helices (TMHs), compared to the usual 12 TMHs of most MFS transporters, including mdfA and mdtM (Daley et al 2005). Cysteine was investigated as a potential substrate, as overexpression of hsrA has been shown to increase accumulation of homocysteine (Goodrich-Blair & Kolter 2000). Thermophoresis measurements defined an interaction or binding event between hsrA and cysteine (Figure 4.26 E). The calculated Kd for this interaction was 8 μM. The intracellular concentration of cysteine found under normal conditions was reported to be 200 μM (Park and Imlay 2003). At this cysteine concentration, it would be expected that, if present in the membrane, a significant proportion of hsrA transporters would be bound to cysteine. This is therefore likely to be a significant biological finding that could involve the regulation of homocysteine in *E. coli*. Caution must be exerted as the binding event observed does not inform us of what the transporter is actually doing. All that the experiment does infer is binding, this may represent binding and transport of cysteine, binding and inhibition of the hsrA transporter or binding that has no effect on the transport process.

To confirm the results and provide more information, a second round of thermophoresis experiments were carried out on the Nanotemper Monolith NT.115 (Biochemistry Dept., Oxford University, thanks to David Staunton for providing access). As well as reconfirming the initial binding event for mdfA as a positive control, an additional aim was to test whether the proposed arabinose exporters kgtP and mdfA (Koita et al 2013) could in fact be influencing the catabolite repression mechanism. Protein was expressed and purified as described previously in Materials and Methods, Section 2.5.4. The samples were transported to Oxford at 4 ˚C. The experiments are summarised in Table 4.9 and plotted graphs are shown in figure 4.29. Binding of mdfA to the positive control chloramphenicol was used to test the reliability of the system prior to investigatory experiments with cAMP. It was intended all samples would also be tested with arabinose.

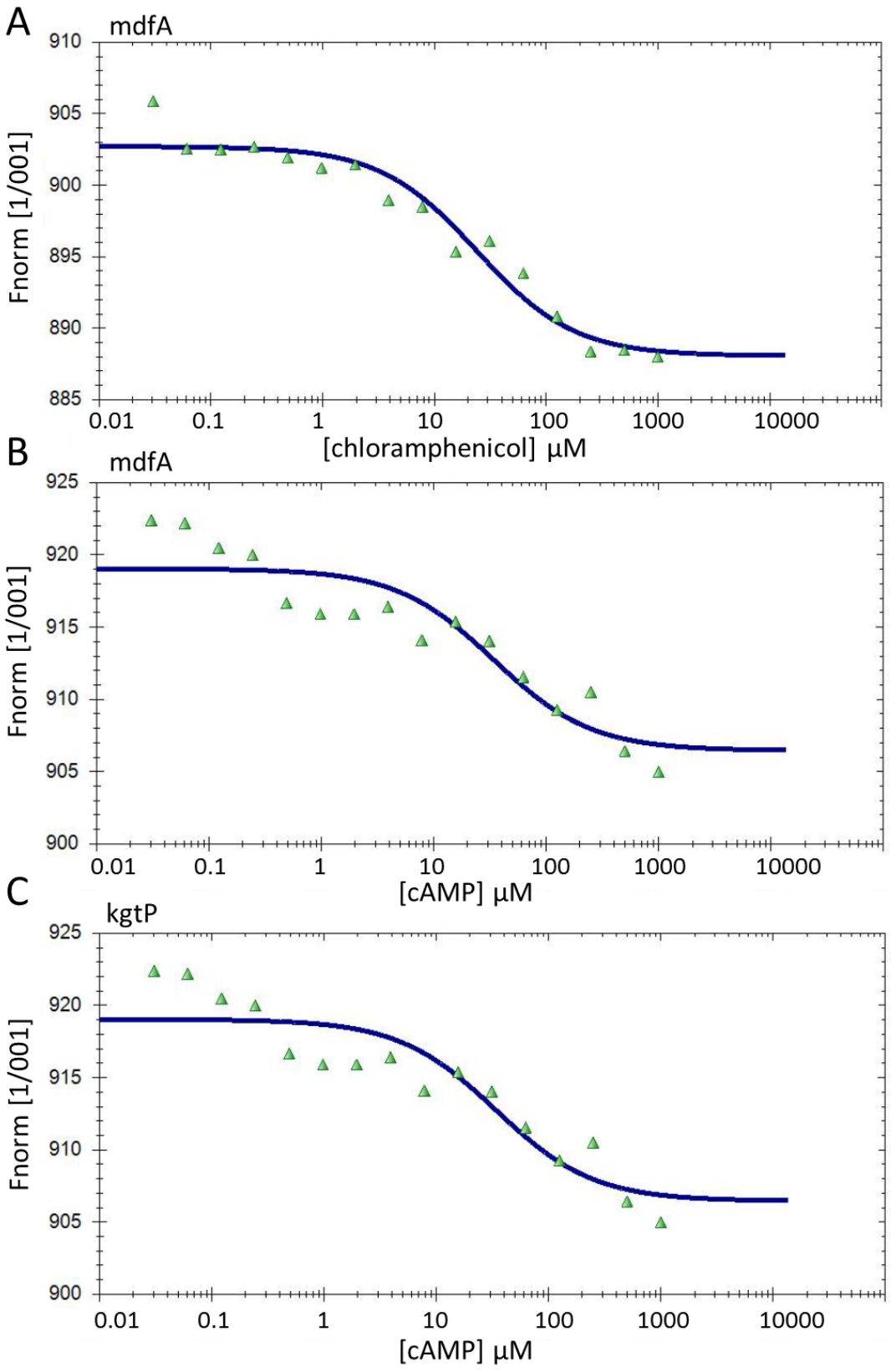


Figure 4.29 Ligand-binding curves demonstrating an interaction between Ni2+-IMAC purified H6-MFS-sfGFP and specific substrates using thermophoresis

H6-MFS-GFP proteins (mdfA, kgtP) were expressed overnight in BL21 (DE3) at 18 ºC and purified using immobilised metal affinity chromatography (IMAC). Purified samples were split into thermophoresis tubes containing different concentrations (indicated by the x-axis) of substrate (indicated in the x-axis title). The Nanotemper Monolith NT.115 was used for thermophoresis experiments and analysis. A) H6-mdfA-sfGFP titrated with chloramphenicol (30 nm – 1 mM). B) H6-mdfA-sfGFP titrated with cAMP (90 nm – 3 mM). C) H6-kgtP-sfGFP titrated with cAMP (30 nm – 1 mM).

Table 4.9 Details of targets and ligands for the ligand binding determination using the Monolith from Nanotemper at Oxford University.

Protein	Fluorescent label	Substrate
mdfA	sfGFP-fusion	chloramphenicol
mdfA	sfGFP-fusion	cAMP
kgtP	sfGFP-fusion	cAMP

The control and investigatory samples all showed binding events, again indicating the expression, purification binding assay was effective. The calculated binding affinities for each transporter/ligand are reported in Table 4.10. The affinity of H6-mdfA-sfGFP for its control, chloramphenicol, in DDM was 139 µM, which is similar to the previous calculated affinities of 69 µM (DM) and 108 µM (Brij35) as well as the published affinity of 75 µM in DM (Heng et al 2015). This indicated that the H6-mdfA-sfGFP was viable and that the method was reproducible.

Interestingly mdfA and kgtP were measured interacting with cAMP, with binding affinities of 24 µM and 34 µM, respectively. This interaction has not previously been reported and supports the theory that these transporters are involved in catabolite repression. Again, as this technique can only infer binding, it is unclear whether cAMP binds and is transported, binds and inhibits or has no effect when bound to the transporters.

The data presented here indicates the Monolith thermophoresis ligand binding approach (Method 4) is suitable as a non-specific approach to begin to characterise integral membrane proteins. Varying transporters, ligands and detergents

produced binding events, indicating the assay is robust under a variety of conditions.

Table 4.10 Summary of ligand binding results using the Nanotemper Monolith with assigned binding Kd values

Protein	**Detergent**	**Ligand**	**Determined Kd**	**Published Kd**	**Reference and detergent used**
mdfA	0.2 % DM	Chloramphenicol	69 µM	75 µM	Heng et al, 2015 (DM)
mdfA	5 mM Brij35	Chloramphenicol	108 µM	75 µM	Heng et al, 2015 (DM)
mdfA	0.2 % DM	TPP+	20 nM and 685 nM	4.7 µM 2.9µM	Lewinson et al, 2001 (DM) Sigal et al 2007 (DDM)
mdtM	0.05 % DDM	Chloramphenicol	23 nM	22 nM	Algre et al, 2016 (DDM)
hsrA	2 mM DDM	Cysteine	8 µM	n/a	n/a
mdfA	2mM DDM	Chloramphenicol	139 µM	75 µM	Heng et al, 2015 (DM)
mdfA	2mM DDM	cAMP	24 µM	n/a	n/a
kgtP	2mM DDM	cAMP	34 µM	n/a	n/a

4.3 Discussion

This chapter aimed to develop a method of determining ligand binding to integral membrane proteins that does not require the substrate to be labelled. The reason behind this was to apply this methodology to characterising the MFS antibiotic efflux transporters from *E. coli*. Functional evidence for many transporters are mostly in agreement, such as the drug transporter mdfA which is thought to efflux a range of antibiotics (Nishino et al 2001; Lewinson et al 2001; Heng et al 2015). However other MFS proteins, such as ydeE, have a variety of proposed roles such as antibiotic and sugar transport (Nishino et al 2001; Hayashi et al 2010; Koita et al 2013). This inconsistency is often due to the focus on phenotypes that result from over- or under-expression studies of the target gene and therefore could be misinterpreted.

While these methods are useful for screening potential substrates, there are some flaws when used to assign the function. For example, proteomics based on knockout studies in proteobacteria and mice, effect the expression levels of hundreds of other proteins (Zhou et al 2013). Therefore, observed phenotypes such as antimicrobial resistance may be a result of these global proteome changes. Other methods are more appropriate for demonstrating binding and transport, for example Lewinson et al (2001) used radiolabelled TPP+ to demonstrate binding to purified mdfA. However, it is expensive and not practical for larger-scale screening experiments.

To tackle these issues a more direct method suitable for large scale screening was tested. The information obtained will be used in conjunction with the current literature, to further inform the function of transport proteins with uncertain or unknown function.

Using a vector with a GFP tag, allows indirect assessment of the target protein. In this way, it is possible to monitor the effects of potential substrates on a protein's thermophoresis (Jerabek-Willemsen et al 2011) or thermostability (Franken et al 2015). To utilize these properties, the H6-msfGFP vector was selected for cloning, which expresses protein under the control of a T7 promoter adding a hexahistidine tag to the N-terminus and sfGFP tag to the C-terminus of the expressed protein.

For this purpose, a screening library was set up using the MFS family, one cause of antimicrobial resistance in *E. coli* (Nishino et al 2001), a clinically relevant bacterial species (Table 4.2; Weiner et al 2016). The trend of increased antimicrobial

resistance has serious implications for the prolonged effectiveness of worldwide clinical healthcare (O'Neil et al 2016). Of the 118 members of the MFS family for which cloning was attempted, 63 members were successfully cloned into the H6-mfsGFP vector using a ligation independent cloning technique. Others were either successfully amplified but not cloned or were not amplified. It was concluded that 63 constructs were sufficient to begin assay development. Successfully cloned constructs were confirmed with sequencing to be in-frame with the N- and C-terminal regions.

C-terminally tagging constructs with sfGFP enabled easy screening of the cloned MFS transporters. Expressed proteins showed variation in overall yields and final cell densities, based on OD_{600} (Figure 4.12-4.14). This variation is likely to be a due to the differences in characteristics between transporters. For example, the requirement of chaperones will affect how quickly some proteins can be expressed and correctly folded (Denoncin et al 2012). Other proteins, such as Lon proteases, regulate cellular function in *E. coli* (Chang et al 2016) and their ability to target an overexpressed protein will affect the final yield. Additionally, the physiological effects of expressing sugar proteins, such as setC, may reduce cellular availability of monosaccharides and result in reduced capacity for growth and expression.

This over-expression screening step identified which conditions, including media type and temperature, would achieve the best yields. This allows customisation of the protein expression protocol to match the target. It is important to maximise yield at this expression step so that sufficient protein is generated for downstream assays. It was concluded that sufficiently high protein expression was achieved for a number of the transporters to allow for assay development. However, to improve expression yields of those transporters that had poor yields, such as setC, it may be possible to further optimise conditions by changing media composition, IPTG concentration, antibiotic concentration, or co-expression with chaperones (Martínez-Alonso et al 2010; Vincentelli et al 2011; Marini et al 2014).

4.3.1 The use of GFP as a reporter for MFS transporter thermal profiles in the development of a substrate binding assay

Different methods were used to design a thermal shift assay utilising the sfGFP protein tag. Previous thermal shift assays use fluorophores responsive to hydrophobic environments, so when proteins are denatured, their hydrophobic core is exposed, and a signal is produced. The impact of a ligand on the stability profile can be assessed to identify substrates (Rupesh et al 2014; Huynh and Partch

2015). However, this technique is unsuitable for membrane proteins, which require a constant hydrophobic environment to maintain solubility. Instead, a C-terminal sfGFP tag was added which acted as the thermal shift reporter. As the MFS transporter denatures at elevated temperatures and begins to aggregate and precipitate, removal of the non-functioning protein is achieved by simple centrifugation. Three different methods were used to optimise this assay: 1) using solubilised membrane fractions from the whole cell lysate 2) using Ni^{2+}-IMAC purified protein and 3) a T_m hold assay.

4.3.1.1 Analysing the effectiveness of solubilised cell content to generate a substrate binding assay using thermoprofiles

Using ultra centrifugation, membranes containing the over-expressed transporter were isolated from all soluble proteins. Membrane proteins were then released by solubilisation of the membrane using detergents. The GFP fluorescence signal could then be used to assess the amount of protein left after a temperature shock of the target protein. The effectiveness of this method was assessed using the MFS proteins, setB, lacY, xylE, mdfA, araE and mdtM. These were good proteins to choose as positive controls for the assay, as binding has already been implicated or demonstrated (Daruwalla et al 1981; Liu et al 1999; Sun et al 2012; Kumar et al 2014; Heng et al 2015; Alegre et al 2016) and it was necessary to test the assay using proteins already known to bind a substrate, before screening uncharacterised proteins.

However, other than mdtM, the addition of the putative substrate did not appear to affect the thermoprofiles of these positive control proteins. It was hypothesised that this may be due to the presence of additional native *E. coli* proteins or an excess of free detergent sequestering the available substrate, essentially lowering the concentration of substrate available to the expressed protein. Alternatively, the high detergent concentrations used to solubilise all of the membranes may have inactivated the transporters. One further possibility was that binding of the ligand did take place, however, this binding event had no effect on the protein's thermostability.

4.3.1.2 Analysing the effectiveness of purified protein to generate a substrate binding assay using thermoprofiles

To rule out the possibility of excess, non-specific sequestering of the substrate molecules, the target transporters were purified using Ni^{2+} immobilised affinity chromatography. The assay for Method 2 titrated the purified protein with a range

of substrate concentrations. Using multiple substrate concentrations can provide information on binding affinities and show if any interaction is specific (Hulme et al 2010). Where effects on the graph are linear this indicates a non-specific interaction between substrate and protein, whereas sigmoidal effects indicate a specific interaction with a binding site that can be saturated (Hulme and Trevethick 2010).

Unlike in Method 1, addition of substrates to purified MFS H6-transporter-sfGFP micelles did appear to have an effect on thermostability of proteins araE and mdfA. However, substrates appeared to have a destabilising effect, which is the opposite of the stabilisation seen in other thermal shift assays (Franken et al 2015). This could not be replicated, however, and thus could simply be a result of system noise. The inability to replicate results of this assay could be caused by a range of reasons, including inactive protein, substrate binding does not affect thermostability, pipetting variation, variable noise levels due to the material of the plates used or that the changes in thermostability is masked by the noise of the system.

4.3.1.3 Analysing the effectiveness of purified protein to generate a substrate binding assay using T_m shift

An issue of Methods 1 and 2 was the lack of replicability, in part due to their lengthy procedures and multiple pipetting steps. To rectify this, Method 3 held proteins at their T_m, meaning fewer temperature point measurements, using a plate reader to read several samples at once. This again failed to improve reproducibility as initial results appeared to show that increased chloramphenicol availability led to an increase in mdfA stability, but it was not possible to repeat this finding. It was concluded that the plate reader, used for Method 3, was not suitable to detect any change in T_m by this method.

4.3.1.4 Analysing the effectiveness of purified protein to generate a substrate binding assay using thermophoresis

To investigate whether the protein was active, an alternate assay was tested. Microscale thermophoresis is summarised in Figure 4.26-4.27, using GFP to observe changes in kinetics and localised protein concentrations, upon heating a sample. This technique has been previously shown to provide consistent, reliable results (Seidel et al 2013). Results shown here indicate this technique is suitable as a non-specific approach in characterising integral membrane proteins, robust with the use of different transporters, ligands, detergents and protein topology.

Microscale thermophoresis was used to calculate the binding affinities of H6-mdtM-sfGFP and H6-mdfA-sfGFP, which were selected as positive controls as their binding affinities had been previously published, and H6-hsrA-sfGFP, which looked promising from previous tests in Method 2 (Table 4.10). The results indicated protein was active and that binding events were able to be recorded. It is therefore possible that protein-substrate binding of these samples does not affect thermostability on a measurable scale with available equipment used in Methods 1-3. This is directly comparable for H6-mdfA-sfGFP purified in Brij-35, which was assessed using the thermophoresis assays in Methods 1-3 and using microscale thermophoresis.

Calculated binding affinities of H6-mdfA-sfGFP with chloramphenicol or TPP+, and H6-mdtM-sfGFP with chloramphenicol were generally comparable to previously published values (Lewinson et al 2001; Heng et al 2015; Algre et al 2016). The less comparable result is H6-mdfA-sfGFP with TPP+, which previously gave a single binding affinity of 4.7 μM or 1.7 μM (Lewinson et al 2001; Sigal et al 2007), but here gave two binding affinities of 20 nM and 685 nM. It is therefore possible that TPP+ can bind to a separate site on mdfA than the one occupied by chloramphenicol (Lewinson et al 2001).

4.3.1.4.1 Purified H6-hsrA-sfGFP is capable of specifically binding to potential substrate/inhibitor cysteine

Microscale thermophoresis indicates an interaction between H6-hsrA-sfGFP and cysteine, with a calculated binding affinity of 8 μM. This interaction has not previously been reported, although it is known that overexpression of hsrA in *E. coli* leads to an accumulation of homocysteine (Goodrich-Blair & Kolter, 2000). Cysteine and homocysteine thiolactone both operate within the same metabolic pathway, with cysteine a product of homocysteine metabolism (Miller et al 2012).

This allows some speculation as to the interaction between cysteine, homocysteine and hsrA. Hypothetically, there could be a negative feedback loop to inhibit conversion of homocysteine to cysteine, when cysteine is imported. Alternatively, cysteine binding to hsrA may inhibit homocysteine export. As the thermophoresis result only demonstrates binding, it is unclear whether cysteine binds to hsrA to be transported, to inhibit the transporter, or whether this would have no effect.

4.3.2 Thermal shift assays were ineffective with MFS transporters despite thermophoresis indicating active protein.

A shift in thermal stability has been previously reported for soluble proteins upon binding their substrate, including complexes with ERK1/2 (Jafari et al 2014) and herpes simplex virus-1 (Rupesh et al 2014). However, a change in thermal stability in the presence of known substrates was not routinely seen with Methods 1-3. Binding was observed when using the alternative method microscale thermophoresis, indicating assays 1- 3 require more work to be useful.

It is unclear whether the absence of a change in thermal stability is characteristic of the MFS family or of membrane proteins in general. The Solubility of a membrane protein depends on inclusion within a micelle, which may be the main factor controlling thermostability rather than substrate binding, explaining why the substrate-based assay was ineffective. A recent circular dichroism study investigated the stability of micelle-bound lacY and xylE when treated with denaturing urea, in the presence of their known substrates, lactose and xylose, respectively. xylE was stabilised by addition of xylose, but lacY was not stabilised by lactose (Harris et al 2017).

This indicates it is possible for substrate binding to provide a stabilising effect on MFS transporters under chemical denaturation method, in this case 8 M urea. However, a change in thermal stability indicating binding was not reliably observed for Methods 1-3 in this assay. Method 4, which observed binding events independently of thermal stability, proved the most robust, non-specific approach in characterising integral membrane proteins. Interestingly thermophoresis has previously been used on whole cell lysate, in a purification similar to that of Method 1 (Khavrutskii et al 2013).

Chapter 5 Protein over-expression using the *E. coli* pET/IPTG system requires genetic adaptation

5.1 Introduction

As previously mentioned, the difficulties associated with membrane protein expression and characterisation have led to their underrepresentation in the published literature. There are many options to be considered in the production of a protein. Careful consideration should be used to select an expression system, and each stage of the production and purification may require careful optimisation.

5.1.1 Selecting a recombinant expression system

There are many options to be considered in the production of a protein. When selecting an expression system, time, costs, facilities, post translational processing, scalability and ethics may need to be considered (Mohajeri et al 2017). Additionally, once established, there is no guarantee the system will be effective. Taking this into consideration, there is no single best expression system that works for all proteins.

E. coli is a key expression system for characterising proteins: as of September 2017, 84.5 % of all protein structures in the Protein Databank (PDB, Berman et al 2000) were purified from *E. coli* BL21 or BL21 (DE3) expression systems. Bacterial expression systems tend to be a simpler, cheaper option compared to eukaryotic systems, due to lower costs of cell lines, growth materials and growth facilities. More elaborate facilities may be required when using alternative expression systems (Gecchele et al 2015). It may be cost effective to attempt *E. coli* expression, followed up by investigation of alternative systems if this proves ineffective. To improve the efficiency of protein purification, methodologies have been developed to screen hundreds of proteins simultaneously, providing information on possible targets for study, this shows with current technologies it is not always possible to select one particular protein to study (Jia et al 2016). Success is much more likely if many homologues are considered and the best option from screening studies selected. Poor expression can be remedied through a number of techniques, these will be considered later on.

Once an expression system has been selected, an appropriate vector can be selected for cloning. An expression vector is a genetic element; most commonly they are plasmids, adapted to drive the production of an encoded protein.

Plasmids are mobile genetic elements naturally occurring within bacterial populations. Under the correct conditions a bacterial cell can uptake a plasmid, known as transformation, and utilise the encoded genes. These often encode genes advantageous in certain situations, such as the presence of antibiotics (Al-Tawfiq et al 2017). For instance, the pBR322 vector includes tetracycline and ampicillin resistance genes allowing either to be used as selection pressures (Bolivar et al 1977). To drive exogeneous protein expression, a vector will contain a promoter site alongside encoded genes. Artificially produced vectors have a cloning site, which will allow insertion of a gene, while the associated promoter site will control the production of mRNA which ultimately results in protein expression. The *E. coli* pET system is controlled by a T7 promoter site, this is shown in Figure 5.1. The native *E. coli* T3 polymerase does not recognise the T7 promoter site so protein expression will not occur. The T7 RNA polymerase is genomically encoded in DE3 strains of *E. coli*. This gene is under the control of the lacUV5 promoter which is recognised by the endogenous T3 RNA polymerase (Studier et al 1986). Under standard growth conditions, the lac repressor blocks the lacUV5 promoter and the T7 promoter so no mRNA coding for the T7 polymerase or the target protein is produced (Figure 5.1 A). In the presence of lactose, or the lactose analogue IPTG, the lac repressor is no longer able to bind to either promoter sites. In the absence of the lac repressor *E. coli* T3 RNA polymerase can transcribe T7 polymerase as shown in Figure 5.1 B. Once produced, T7 polymerase recognises the plasmid promoter site and produces the mRNA of the encoded protein. Figure 5.1 B shows a representation of a protein produced with the tags GFP and poly-histidine. It should be noted that this process is the same for both soluble as well as membrane protein over-expression.

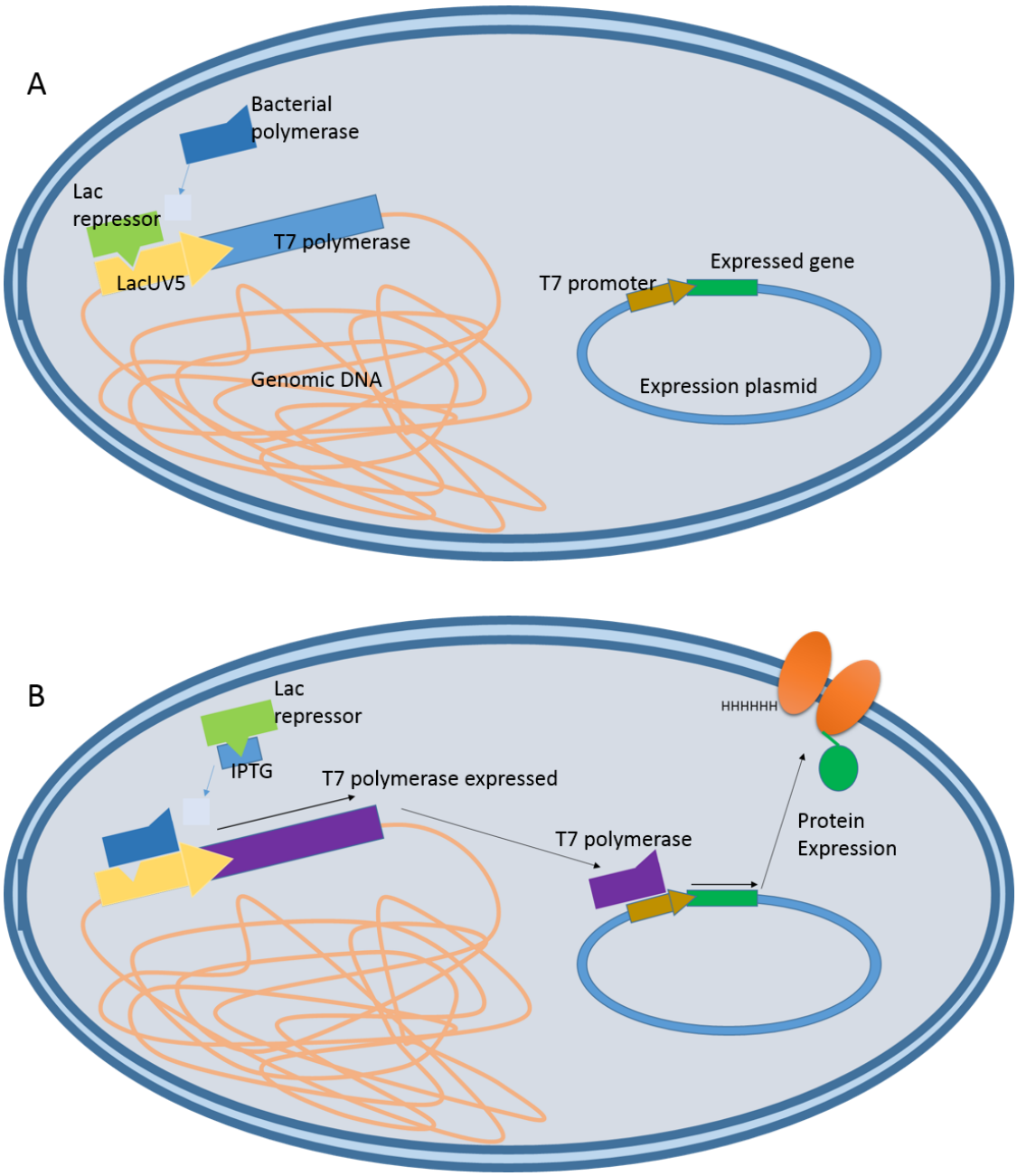


Figure 5.1 The *E. coli* T7 expression system responds to the presence of IPTG by expressing a plasmid-encoded protein

Cellular components are labelled in the figure. A) In the absence of IPTG, the lac repressor protein binds to the LacUV5 promoter site, inhibiting binding of the bacterial T3 RNA polymerase to LacUV5. T7 polymerase is not expressed meaning the plasmid encoded protein is not produced. B) The Lac repressor binds IPTG when it is present, rather than the LacUV5 promoter. This allows the bacterial T3 RNA polymerase to bind LacUV5, leading to expression of the T7 polymerase. T7 polymerase binds to the T7 promoter site on the plasmid. The encoded protein is expressed.

Popular vector examples used in *E. coli* are the pET and pUC expression vectors (Rosano and Ceccarelli 2014). Different vectors can provide additional primary protein structure. This additional sequence can form a tag providing a variety of uses. One of the most commonly used tags is a chain of histidines (His-tag), which has a high affinity to divalent metal cations. Utilising this affinity is helpful when isolating the expressed protein (Lilius et al 1991). Other common examples of tags utilised for purification include FLAG-tags (Schmidt et al 2012), maltose binding protein (MBP) (Lebendiker et al 2011) and Glutathione-S-transferase (GST) (Harper et al 2011). Where tags are inappropriate, antibodies targeting the recombinant protein can be used for purification, in a process called immunoaffinity chromatography (Fitzgerald et al 2011). Purification tags have associated attributes that can be considered during the selection process and reviews on the topic are available (Lichty et al 2005). For example, MBP (Lebendiker et al 2011) and cellulose-binding module (CBM) (Andrade et al 2010) can reduce aggregation, when protein stability is an issue.

To optimise expression and purification, reporter proteins can be used. GFP-tagging can be a useful tool for guiding numerous stages of protein production, with methodologies for GFP-based screening established in *E. coli* (Chae et al 2000) and *S. cerevisiae* (Drew et al 2008). An outline of these techniques is described in Chapter 1.

As the characteristics of proteins are varied, there is no single best method for the expression (Esposito et al 2006). Therefore, it is important to consider all options before cloning, improving the chance of forming an effective expression system.

5.1.2 Strategies for improving protein expression for targets with poor yields

Traditional cloning methods amplify the gene of interest via PCR, before ligation into an expression vector (Büssow et al 2000). *De novo* synthesis is an alternative option (Kosuri and Church 2014) which can eliminate the PCR error rate of long sequences, although this may be unnecessarily costly for short, easy to amplify sequences (McInerney et al 2014). Synthesis can also be used to improve expression yields by codon optimisation: replacing the native amino acid sequence with codons commonly used by the expression system (Webster et al 2016). This can be generated by *de novo* synthesis, as an added bonus this removes the need for mutagenesis experiments if any primary structure alterations are required.

Despite the variety of cloning vectors and expression systems successful protein production is not always possible, limiting characterisation. An analysis looking at research bias of publications found 65 % of kinase papers were on 10 % of all kinases (Edwards et al 2011). Rather than greater importance of these proteins, the bias was attributed to availability of tools such as robust chemical probes, antibodies or inhibitors. Screening techniques identify proteins showing promise for purification, meaning more difficult proteins may be disregarded at a very early stage of investigation. Clearly tackling the difficulties attributed to membrane proteins would open up many poorly investigated avenues of research.

To investigate why some proteins will express in *E. coli* while others will not, Dumon-Seignovert et al (2004) transformed 28 expression vectors each containing a different integral membrane protein (IMP) into different *E. coli* strains. 62% of the proteins could not be propagated in BL21 (DE3) cells, due to leaky expression causing deleterious toxicity, even in the absence of induction.

Expression of a range of membrane proteins have been shown to affect *E. coli* cell growth, pH, cell shape, mRNA and native protein expression (Gubellini et al 2011); as well as upregulate native proteins associated with DNA damage, heat shock, stationary phase and bacteriophage infection (Gill et al 2000). During recombinant protein production, native protein expression and their activity can be reduced by up to 75 % (Wagner et al 2007), including the important ribosomes (Dong et al 1995) and proteins involved in respiration such as cytochrome oxidase. Part of this is due to cytoplasmic aggregates of misfolded proteins, containing recombinant proteins, chaperones and other native IMPs (Wagner et al 2007). The expressed IMP may therefore be physiologically damaging to the expression system, as the cells struggle to maintain themselves while producing a recombinant protein. Gubellini et al (2011) concluded that toxicity is caused by expressed proteins with missense mutations or misfolding of expressed proteins.

5.1.3 Optimisation of protein expression systems

Where protein expression is poor, the expression conditions can be optimised to improve yields. As mentioned previously, there is no universal best expression condition, so screening different conditions is likely to be required. For instance expression of the three homologous human proteins, OCTN1, OCTN2 and OCTN3, requires two vectors (pH6EX3 and pET-21a(+)), two Rosetta cell lines ((DE3)pLyS and DE3) and two induction times (6 and 4 h) for optimal expression. Despite their functional and structural similarities, each required a different combination of

conditions (Indiveri et al 2013). Certain combinations of chaperones can also improve expression of soluble proteins (Marco et al 2007) and some membrane proteins (Nannenga et al 2011), however as with most membrane protein techniques, this is less well established than for soluble counterparts.

Miroux et al (1996) used BL21 (DE3) to look at the toxic effect of protein over-expression. To investigate this a screen of globular and membrane proteins were cloned into pET vectors and transformed into cells grown on media containing IPTG. They found that even in the presence of an empty vector, induction using IPTG lethally hampered growth with most cells failing to colonise plates. Cells tolerant to the toxic effect of oxoglutarate-malate transport protein expression were selected for further experimentation. Plasmid isolated from these cells was shown to still generate the toxic phenotype indicating the change was likely to be genomic. Using non-selective media to remove the plasmid they generated the strain C41 (DE3). The toxic effect of protein expression was compared between C41 (DE3) and BL21 (DE3). The toxicity was reduced or removed from some proteins, however the effect did not work for all those screened. The process was repeated with C41 (DE3), mutants showing reduced toxicity were called C43 (DE3). Both strains are available for purchase and can be used where protein toxicity is hampering growth in BL21 (DE3).

Schlegel et al (2015) used sequencing to identify the mutations reducing IPTG induced toxicity. These were then analysed with allelic replacement and showed mutation in the lacUV5 site controlling T7 RNA polymerase was critical for this phenotype. The promoter region is weaker, producing less T7 polymerase upon induction and alleviating the toxic effect of IPTG induction. Baumgarten et al (2017) showed an alternative mutation in the RNA polymerase site weakens binding of the T7 polymerase to the promoter site, reducing the toxic effect of IPTG induction. Screening with expression vectors showed that this improves the production for some proteins. This supports the theory that the interaction of the T7 RNA polymerase with its promoter site and not just T7 polymerase production is required to produce a toxic effect. Both mutants reduce the interaction of T7 polymerase with its promoter site in the expression vector. The pLysS and Lemo expression systems work in the same way, utilising an inhibitor of T7 polymerase. None have addressed whether this improved expression could be replicated by reducing the concentration of IPTG used for induction.

Table 5.1 Some *E. coli* based protein expression systems developed from the strain BL21

All cell lines cannot utilise galactose, have inactive Lon and OmpT proteases, dcm methylase, hsdS restriction modification subunit. This enables the use of plasmid expression vectors and reduces proteolysis. All (DE3) strains contain genomic copy of the T7 RNA polymerase controlled by lacUV5 promoter.

Cell strain	**Characteristic**	**Utilised expression promoters**	**Reference**
BL21		Native *E. coli*	Jeong et al (2015)
BL21 (DE3)	Produces T7 RNA polymerase in the presence of lactose or IPTG.	Native *E. coli* and T7	Studier and Moffatt (1986)
Rosetta (DE3)	T7 polymerase. Contains rare tRNA codons poorly represented in *E. coli*.	Native and T7	(Dabrowski et al 2003)
BL21 (DE3)pLysS	T7 polymerase. Expresses T7 lysozyme reducing background T7 polymerase.	Native and T7	Dubendorff et al (1991)
Lemo21 (DE3)	T7 polymerase. Expresses T7 lysozyme controlled by rhamnose promoter.	Native and T7	Schlegel et al (2012)
KRX	T7 polymerase, rhaBAD promoter in the place of lacUV5	Native and T7	(Hartnett et al 2006)

C41 (DE3)	T7 polymerase controlled by weak LacUV5 promoter.	Native and T7	Miroux et al (1996)
C43 (DE3)	T7 polymerase controlled by weak LacUV5 promoter.	Native and T7	Miroux et al (1996)
Mt56 (DE3)	Mutant T7 polymerase with weaker binding to promoter sequence.	Native and T7	Baumgarten et al (2017)

Induction-based screening methods have the potential to improve protein production (Massey-Gendel et al 2009). In this method, before protein production is fully implemented, mutant *E. coli* are isolated using agar containing IPTG and used in protein production. Kwon et al (2015) identified the lac repressor gene of the lac operon as an additional mutant hotspot for conferring resistance to IPTG, with a reduced affinity for the inducer molecule in resistant mutants. As with previously identified mutants *E. coli*'s response would match that of a lower IPTG concentration. Expression of proteins considered to be non-toxic may be applying a greater strain on the system than previously considered. This was shown by expression of the Designed Ankyrin Repeat Protein (DARPin) in *E. coli* (Kawe et al 2009). DARPin was selected for its soluble and monomeric characteristics, also, as it targets human growth factors it was considered unlikely to specifically bind to the expression system's native proteins. Despite these attributes, a selection pressure was exerted on the expressing cell lines, leading to inactivation of the expression vector promoter sequence.

Production and purification of soluble as well as membrane proteins is still clearly a complex process and achieving the desired expression yield is difficult. Improving our understanding of this process may lead to gains in protein production. As different proteins required different conditions, one set of rules may never be applicable to all, but more rigorous initial screening techniques could still be applied. The GFP tag can work as a visual reporter for the expression of proteins in BL21 (DE3) cells. By plating cells on agar, different colonies can be selected based

on their variation in expression levels using visible green colouration. This could be a simple way to improve expression levels of proteins in the future.

5.1.4 Aims of chapter

- To transform BL21 (DE3) cells with expression vectors containing target proteins that are GFP tagged for use as a visual screening technique to isolate colonies with improved resistance to IPTG-induction toxicity. Expression vectors were generated as described in Chapter 4.
- To use microbiological techniques to characterise the physiology behind the improved toxicity resistance in mutant BL21 (DE3) lines.

5.2 Results

5.2.1 IPTG induced protein expression has toxic effects on *E. coli* carrying plasmids with T7 promoter sites

Chapter 4 outlined the construction of vectors either expressing membrane proteins tagged with GFP or expressing GFP alone. This chapter will express these constructs in *E. coli* to investigate the effect of recombinant protein expression under the control of a T7 promoter on cellular toxicity. T7 RNA polymerase is expressed when IPTG removes lacI inhibition of the LacUV5 promoter site. In theory, greater IPTG concentrations should further uninhibit T7 polymerase expression, driving greater expression of the protein of interest via the T7 promoter. This is summarised in Figure 5.1, as described in the introduction.

To verify that IPTG itself is not toxic to the cells, four different *E. coli* cell lines KRX, BL21 (DE3), LEMO (DE3) and Mach1, three protein expression cell lines and one for DNA manipulation (Mach1)) were plated on LB agar with increasing concentrations of IPTG (Figure 5.2). To ensure that no protein was expressed the cells did not contain expression vectors.

As expected, none of the cell lines were visibly affected by IPTG up to 1 mM. The control, Mach1, has no genomic copy of T7 RNA polymerase. Each of the other cell lines have. The other three *E. coli* strains differ in their inducible promoters. This indicates that there is no toxic effect of IPTG by itself.

Next the IPTG concentration screen was repeated but this time with the H_6-sfGFP expression vector, containing either GFP alone or with a range of integral membrane proteins, transformed into BL21 (DE3). The expressed target proteins include H_6-sfGFP (soluble superfolder variant of GFP), H_6-amtB-sfGFP (an ammonium transporter), H_6-bcr-sfGFP (antibiotic efflux transporter), H_6-mdfA-sfGFP (antibiotic efflux transporter), H_6-mdtL-sfGFP (antibiotic efflux transporter), H_6-mdtG-sfGFP (antibiotic efflux transporter) and H_6-setB-sfGFP (sugar influx transporter). The selection includes soluble and membrane proteins with various families and functions.

Successful transformants were grown to an OD_{600} of 0.1 and then streaked on LB agar with 50 µg/mL kanamycin as well as varying concentrations of IPTG (Figure 5.3). Membrane proteins are known to express to lower levels than soluble proteins hence the plates were left in the incubator for 72 hours; ensuring that changes in

the colouration of the colonies would be visible by eye. The underlying assumption in this experiment is that the more mRNA of the target that is produced the higher the final amount of target protein that is synthesised (Crick 1970). We also know that for membrane proteins, localisation and proper folding requires a ribosome containing a signal sequence to bind to the membrane embedded translocon (White and von Heijne 2008). Hence, compared to soluble proteins, IMP production, at a minimum, follows the process outline of DNA → mRNA → ribosome → translocon → final product. Clearly there are a finite number of translocons within the inner membrane of a bacterial cell hence it would be expected that there is an optimum concentration of mRNA that allows target protein production while not swamping the essential native protein production processes. To try and find the optimum concentration of target mRNA, a series of IPTG concentrations (0.01, 0.05, 0.1, 0.5 and 1.0 mM) were screened. In addition, we made the assumption that there is a natural variation of target protein expression ability within the population of *E. coli* that would allow for a range of expression levels.

Based on the above thinking, the initial expectation was that there would be a variation in green coloured-colonies as some would express the GFP-tagged protein better than others. It was expected there would be an optimum IPTG concentration for each protein, published concentrations include IPTG at 0.1 mM (Casagrande et al 2009) 0.5 mM (Heng et al 2015) and 2 mM (Kumar et al 2015). Figure 5.3 shows an illustration of the predicted outcome for this experiment.

Figure 5.4 shows the actual results. In the absence of IPTG, all transformants grew equally well, with no expression of target protein, based on the lack of green colouration (Fig 5.4B). Adding even a small amount of IPTG (0.01 mM; Fig. 5.4C) induced the expression of proteins based on the visible signal of soluble GFP. There is also a slight green tinge for the colonies containing membrane proteins indicating expression. This is to be expected as GFP is a soluble protein that can be expressed to very high levels in *E. coli* (Moraes et al 2014) as compared to the other membrane proteins. Expression of target proteins is seen in the remaining IPTG concentrations; however, the biggest surprise was the dramatic decrease in the number of colonies. Instead of a lawn of cells growing in each case, fewer, single colonies could be seen with the numbers dropping as low as two viable colonies at the highest 1 mM IPTG concentration (Figure 5.4G). A closer look at the colonies formed, demonstrates that there are at least four phenotypic colonies (Figure 5.5): 1) large, uniform and colourless; 2) large, uniform and slightly green; 3) large, uniform and green; 4) small punctate and green.

From this experiment, we see that the higher the concentration of IPTG used, the fewer viable colonies produced. As well as variation of expression levels, based on the intensity of green colouration for each colony, there is also a variation in colony appearance. Those colonies that were colourless were assumed to be a result of the total inhibition of target protein production and not endogenous protein production. The most significant result was that soluble proteins and membrane proteins behaved in the same way in terms of viable colony numbers and colony phenotypes hence the results are not protein-type dependent. The implication from our results is that IPTG is not toxic but that the process of target protein over-expression using the pET system is toxic to *E. coli*. This is surprising, as previously it is believed that the expression of native proteins or inert foreign proteins (Jensen 2012) should have little effect on the cell viability.

Since the introduction of the pET expression system, over 30 years ago (Studier and Moffatt 1986), it has been assumed that the main causes for a lack of target protein expression was the inability of the target protein to fold properly, protease cleavage or toxicity of the target protein. Here instead, we show that it is the act of over-expressing the target protein that is toxic to *E. coli*. If we assume that during protein over-expression using the pET system, so much mRNA of both the T7 RNA polymerase and the target protein is produced that *E. coli*'s protein expression system is swamped, making it impossible to produce all of its essential genes. In such a case the cell will die which is what we observe. In order for the cell to survive the heavy burden on the protein expression system must be alleviated, the cells must decrease the production of the exogeneous target protein. This can be achieved by decreasing the amount of target protein mRNA produced via mutations at the promoter sites to lower the affinity for the endogenous T3 RNA polymerase, at the lacUV5 promoter site and/or mutations at the T7 promoter site on the pET vector (Figure 5.1). If this is correct, then the colonies that survive and grow have to be mutants. In addition, as seen in Figure 5.4 D, the concentration of IPTG that selects for these mutant forms is 0.05 mM and above based on the appearance of single, large colonies and the absence of a lawn. From now on, we will consider that IPTG concentration of 0.01 mM are effectively wild type, unmutated *E. coli* while all colonies produced by protein over-expression above 0.05 mM IPTG as mutants.

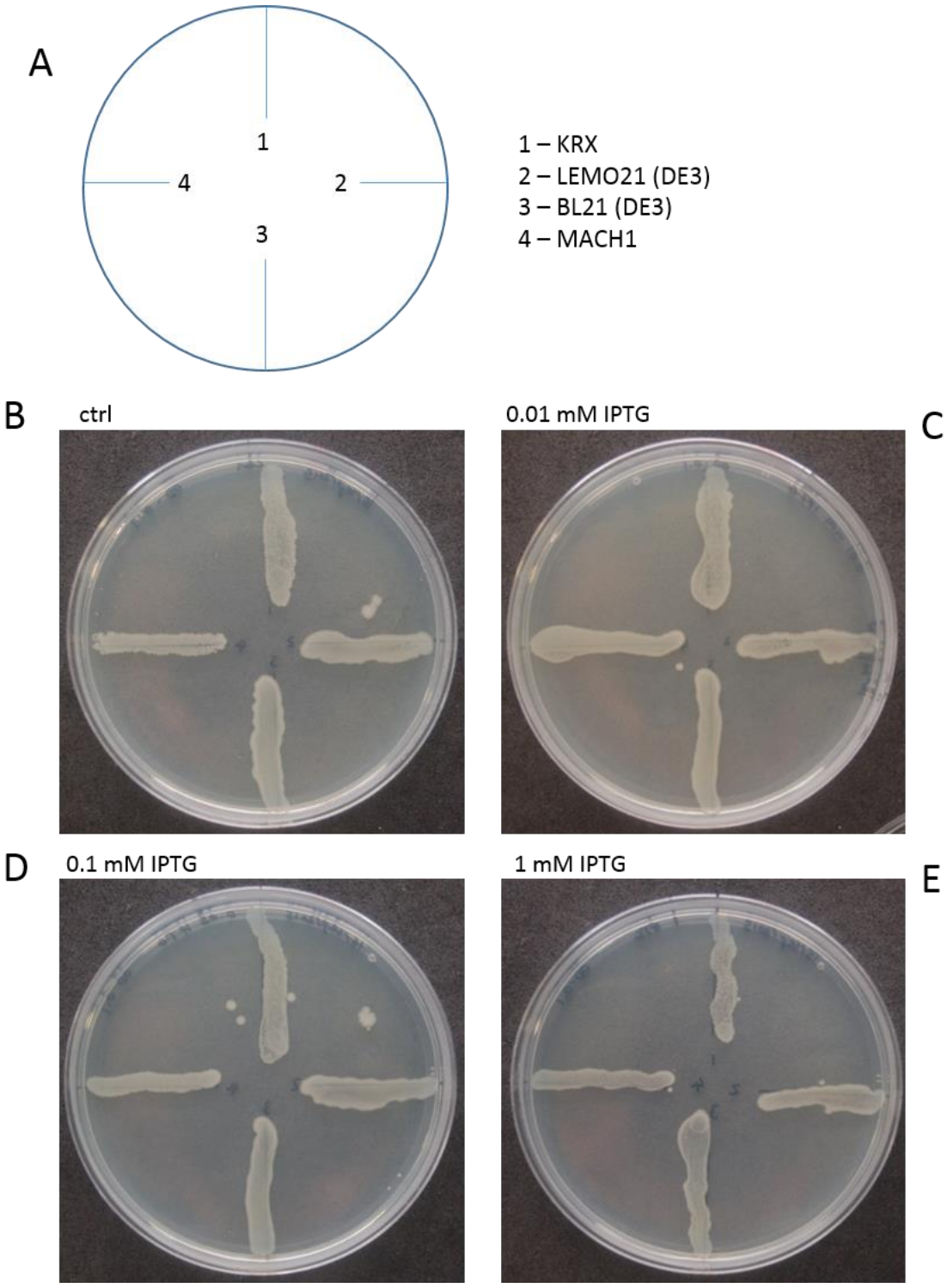


Figure 5.2 Common *E. coli* expression systems grown in the presence of IPTG do not exhibit toxicity in the absence of expression vectors

Cell lines KRX, LEMO21 (DE3), BL21 (DE3) and MACH1 were grown in LB broth to an OD_{600} of 0.1 and spread on plates containing varying concentration of IPTG, as indicated in the figure. Cell plates were incubated at 37 ^{o}C for 16 hours. A) position of cell lines on plates; B – E) LB agar plates with different IPTG concentrations. Ctrl; 0 IPTG added.

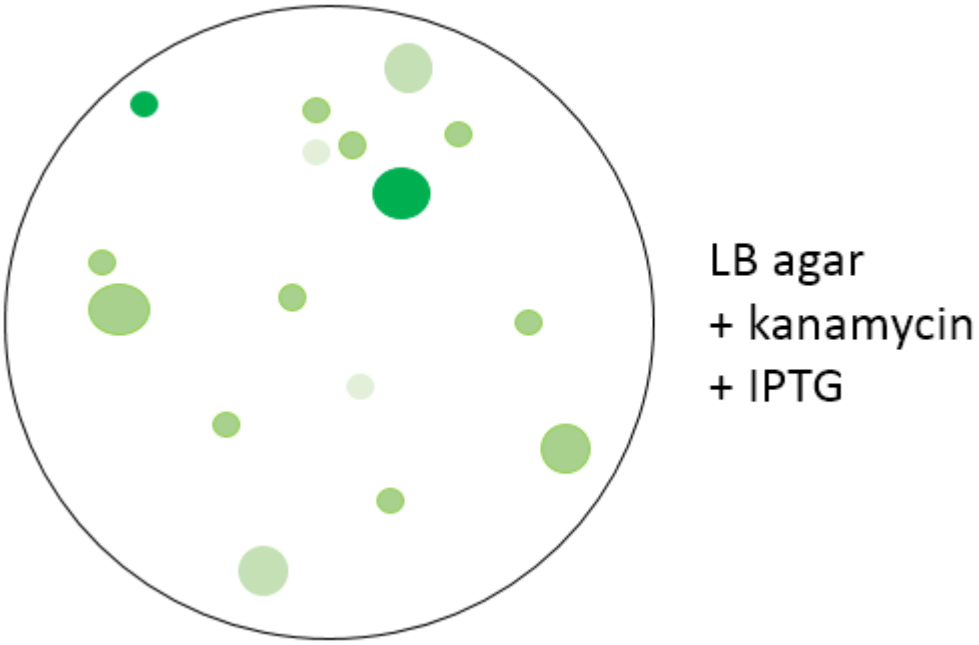


Figure 5.3 Predicted outcome from the expression experiment

E.coli is transformed with MFS expression vector and grown in LB broth to an OD_{600} of 0.1. The cell culture is spread on LB agar plates containing kanamycin and IPTG and incubated at 37 °C for 16 hours. Each green coloured sphere represents a single colony. Visual analysis is used to identify high expressing colonies: darker or more intense green colonies suggest more GFP-tagged protein has been produced. The largest and most intensely green colonies will be selected for further investigation.

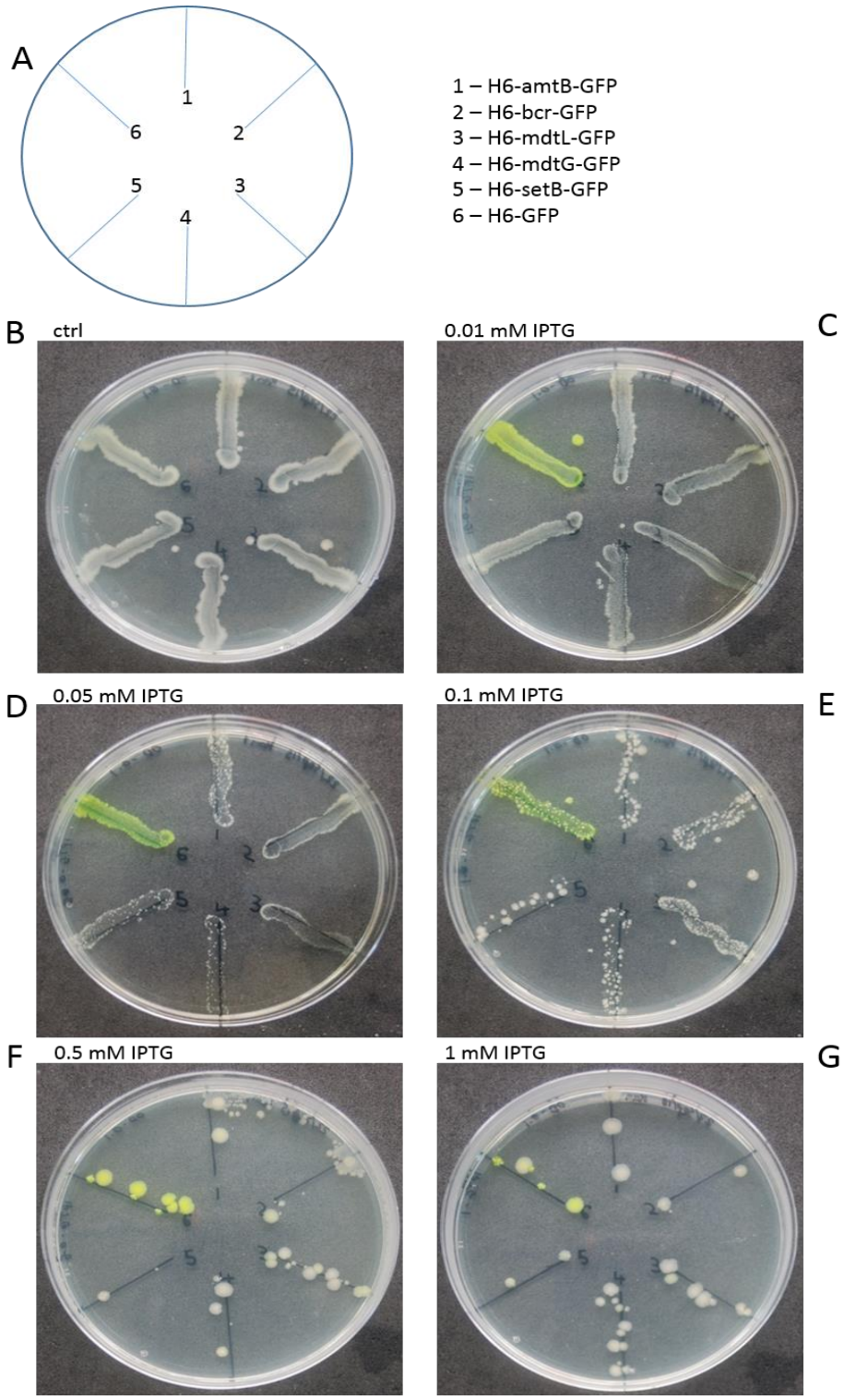


Figure 5.4 BL21 (DE3) *E. coli* carrying H6-MFS-sfGFP expression vectors, grown on LB agar with varying concentrations of IPTG

BL21 (DE3) *E. coli* was transformed with MFS expression vectors and grown in LB broth to an OD600 of 0.1. The cell culture was spread on LB agar plates containing IPTG at a range of concentrations and incubated at 37 °. Visual analysis was then used to identify intensely green colonies as highly expressing colonies. A) position of vectorson agar plates. B – G) LB agar plates with 50 µg/mL kanamycin contain varying concentrations of IPTG, as labelled. Ctrl; 0 IPTG added.

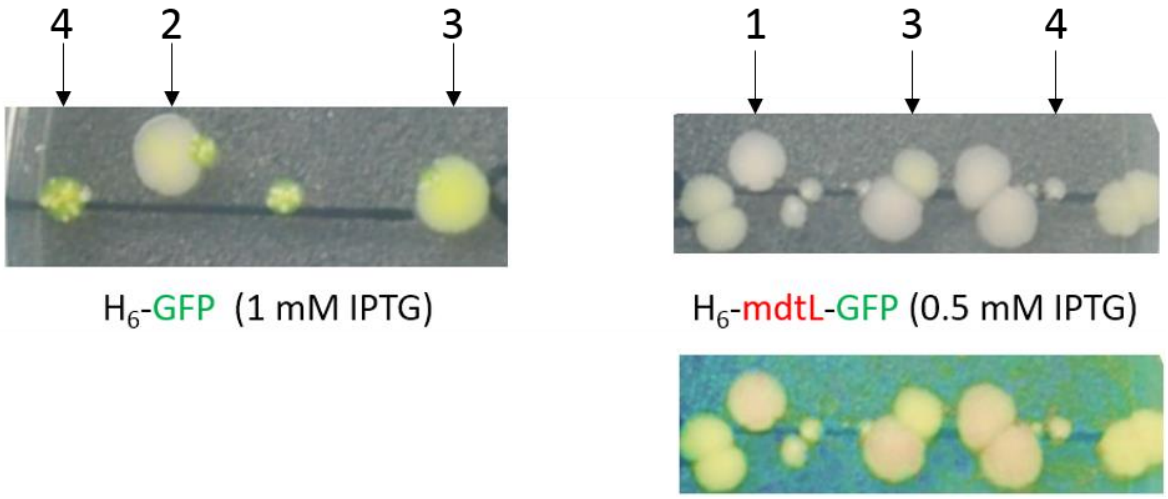


Figure 5.5 Enhanced view of *E. coli* colony phenotypes carrying either the expression vectors H_6GFP or H_6-mdtL-sfGFP, when grown under elevated IPTG concentrations

E. coli BL21 (DE3) was transformed with either H6-GFP (left) or H6-mdtL-GFP (right) expression vectors and grown in LB broth to an OD_{600} of 0.1. The cell culture was spread on LB agar plates containing either 0.5 or 1 mM IPTG, as labelled, and incubated at 37 ºC. The numbers correspond to the different colony phenotypes identified: 1) large, uniform and colourless; 2) large, uniform and slightly green; 3) large, uniform and green; 4) small, punctate and green. The picture on the right is also shown as a colour-enhanced version (lower right) to highlight the difference between the green and non-green colonies.

If indeed mutants are selected for during protein over-expression, then the properties of these cells should be different from the wild-type cells. To test this, the over-expression experiment was repeated, and individual colonies were selected from a range of IPTG concentrations (Figure 5.6). The four colonies included both soluble and membrane proteins and were:

A. 0.1 mM IPTG; H_6-setB-sfGFP (sugar influx transporter); colourless and uniform

B. 0.5 mM IPTG; H_6-bcr-sfGFP (antibiotic efflux transporter); greenish and uniform

C. 0.5 mM IPTG H_6-sfGFP; green and uniform

D. 1.0 mM IPTG; H_6-sfGFP; green and uniform

As before (Figure 5.4), the higher the IPTG concentration the fewer viable colonies were present. These four colonies were then propagated overnight in the absence of IPTG, this was to allow the cells to potentially revert if adaptations to high IPTG exposure were metabolic rather than genetic. On the following day, the four colonies were plated on LB agar with increasing concentrations of IPTG. A control of H_6-sfGFP transformed into fresh BL21 (DE3) cells was included representing the WT. The expectation is that the WT cells would behave as before with decreasing cell viability at increased IPTG concentrations. While if the adaptions in the IPTG resistance colonies lower target protein expression down to a viable level they should be able to cope with the higher IPTG concentrations.

Figure 5.7 clearly show that the mutants are insensitive to the concentration of IPTG used unlike the control WT cells which decrease in number of viable colonies with increasing IPTG concentration. In addition, the WT cells display the same variety of phenotypes as previously described (Figure 5.8). Based on the fact that all of the mutants were visibly similar, the initial concentration of IPTG that the cells were exposed to has no effect on cell growth rates hence mutations exist even at 0.1 mM IPTG. Again, the results were independent of protein type as both membrane proteins and the soluble protein behaviours were identical.

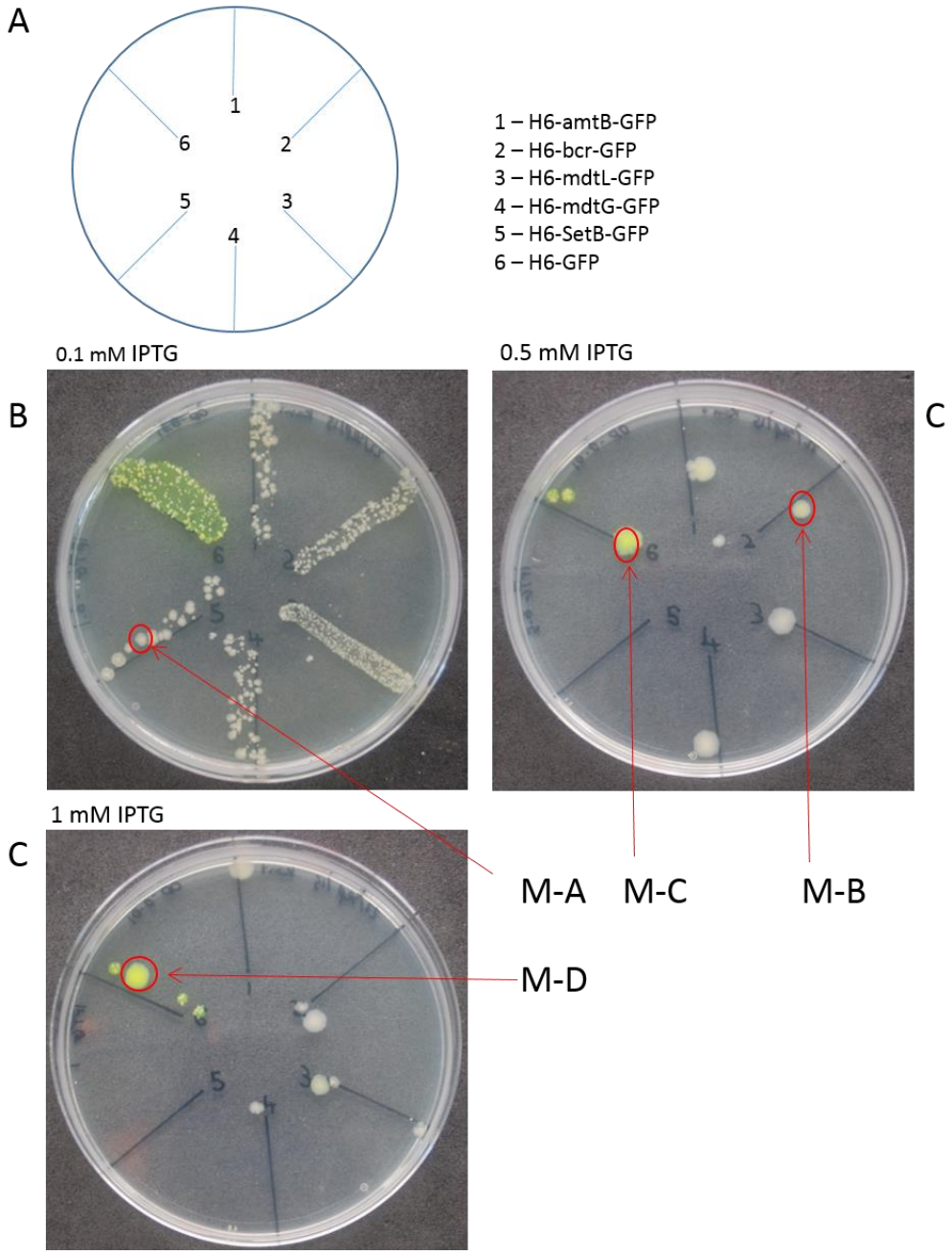


Figure 5.6 Identification and isolation of BL21 (DE3) *E. coli* with adaptions to IPTG toxicity, when expressing H6-GFP and of H6-MFS-sfGFP

E. coli BL21 (DE3) were transformed with MFS expression vectors and grown in LB broth to an OD_{600} of 0.1. The cell culture was spread on LB agar plates containing IPTG at a range of concentrations and incubated at 37 °C. Visual analysis was then used to identify colonies with adaptation to high IPTG, with the green colour demonstrating protein expression. Adapted colonies are labelled mutant A (M-A) to mutant (M-D): colony mutant A - H_6-setB-sfGFP; colony mutant B - H_6-bcr-sfGFP, colony mutant C - H_6-sfGFP and colony mutant D - H_6-sfGFP. A) Layout of vectors on agar plates. B – D) LB agar plates contain 50 µg/mL kanamycin and varying concentrations of IPTG. Concentrations are labelled above the plates.

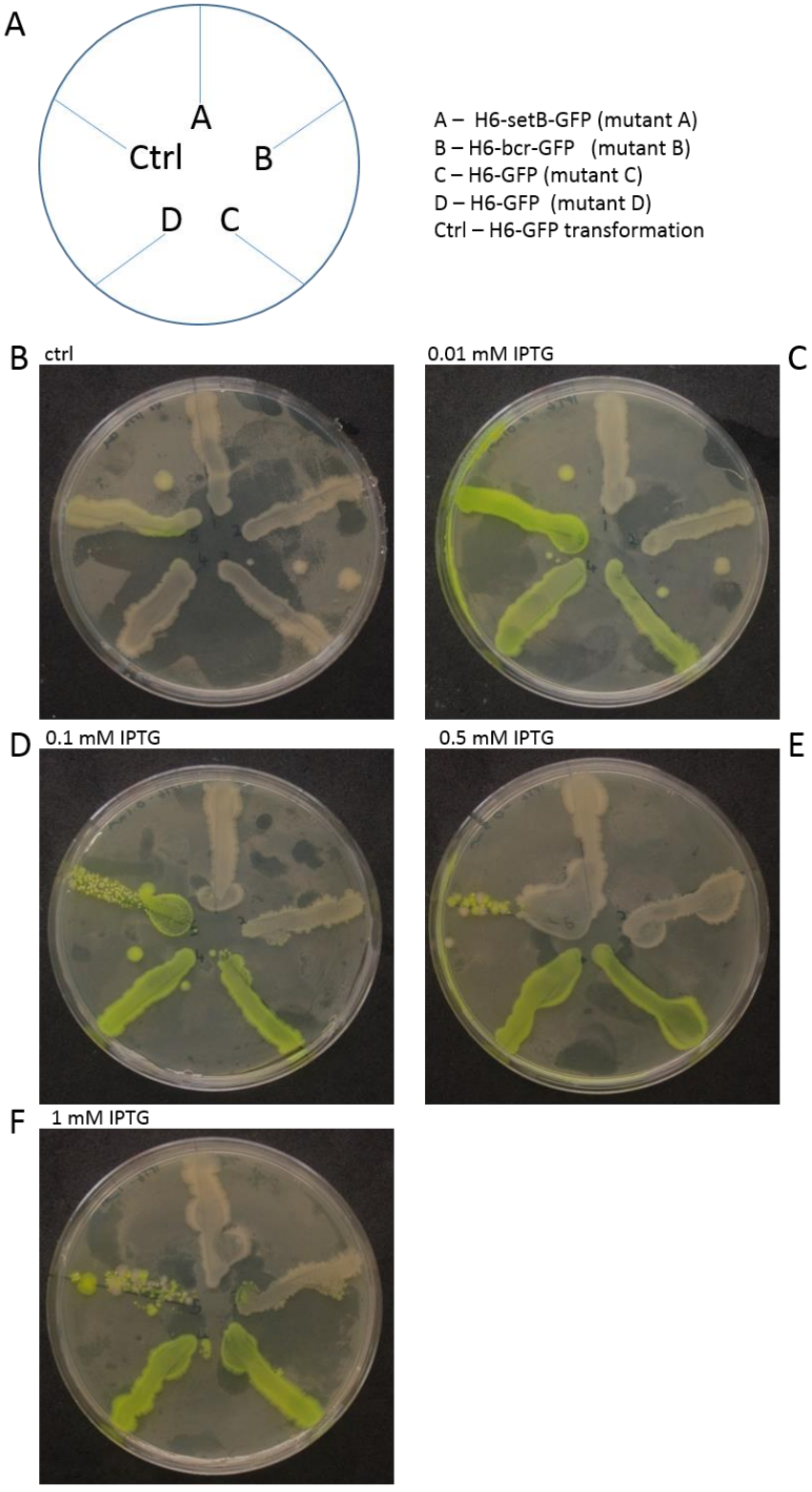


Figure 5.7 BL21 (DE3) *E. coli* cell adaptions to high concentrations of IPTG persist when cells are replated

Mutant colonies A-D, containing MFS expression vectors, were selected for their ability to grow on plates with high IPTG concentrations. All colonies were grown to an OD600 of

0.1 and spread on plates containing various IPTG concentrations. Freshly transformed *E. coli*, with no previous exposure to IPTG, was used as a control (position 5). Following incubation at 37 ˚C plates were visually analysed. A) Layout of vectors on agar plates. B – F) LB agar plates containing varying concentrations of IPTG. Concentrations are labelled above the plates. Ctrl, 0 mM IPTG added.

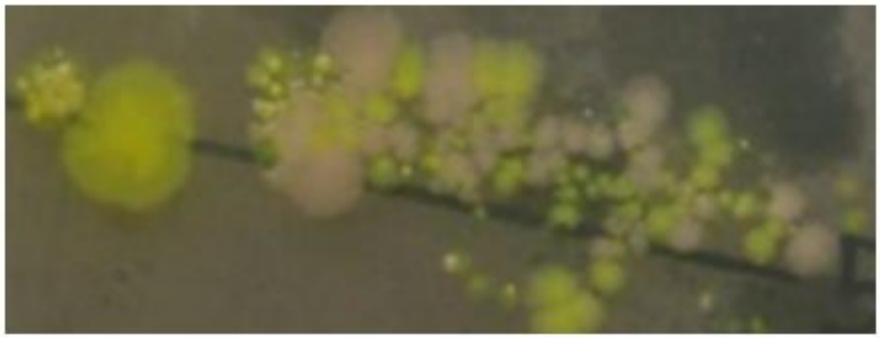


Figure 5.8 Enhanced view of BL21 (DE3) cells transformed with H6-sfGFP and grown in the presence of 1 mM IPTG

E. coli were transformed with the H6-sfGFP expression vectors and grown in LB broth to an OD600 of 0.1. The cell culture was spread on LB agar plates containing 1mM IPTG and incubated at 37 ºC for 72 hours. Visual analysis was then used to identify adapted colonies. This is an enhanced image of Figure 5.7F.

Based on the previous results, the standard practice of inducing target protein expression with the addition of IPTG when the cell density reaches mid-exponential phase will generate a selection of mutations that express proteins to varying degrees. To ensure that we were not studying a mixed population of cells we first isolated similar mutants based on their phenotypes. As such, we selected six mutants (target protein over-expressed using 1 mM IPTG) and two non-mutants (target protein over-expressed using 0.01 mM IPTG) to grow and express in liquid LB medium. The non-mutants behaved like wild type cells and hence were considered as controls. As part of the experiment, two growth phases that did not have IPTG were included to exclude the driving force for exogeneous protein production and allow for potential reversal if metabolic changes were the source of adaptations. The targets selected included another membrane protein, mdfA which is an antibiotic efflux transporter, and the superfolder variant of GFP, again allowing comparisons between a soluble and membrane protein. A new membrane protein was selected to demonstrate that the effect of mutant selection during protein over-expression was again not protein dependent.

Figure 5.9 show the plates that were used to select for the mutants and non-mutants, this was referred to as Round 1. As plates C and D contained only 0.01 mM IPTG, a lawn of cells formed as there was no inhibition of colony formation. We will refer to these samples as non-mutants (NM) and believe that they are equivalent to WT. These samples are not from a single colony but a swab of the sample. The phenotypic characteristics of the eight samples are shown in Table 5.2. These experiments were carried out by Daniel Noel, a University of Southampton Master student under my supervision.

Table 5.2 Sample details used to test persistence of mutations in liquid medium

NAME	Target	[IPTG] (mM)	Description
C1	H_6-sfGFP	1	Large colony, uniform, no green
C2	H_6-sfGFP	1	Medium sized colony, uniform, green
C3	H_6-sfGFP	1	Medium sized colony, punctate, no green
C4	H_6-mdfA-sfGFP	1	Medium sized colony, uniform, no green
C5	H_6-mdfA-sfGFP	1	Medium sized colony, uniform, slightly green
C6	H_6-mdfA-sfGFP	1	Very small colony, uniform, no green
NM1	H_6-sfGFP	0.01	Uniform lawn of cell, very green
NM2	H_6-mdfA-sfGFP	0.01	Uniform lawn of cells, green

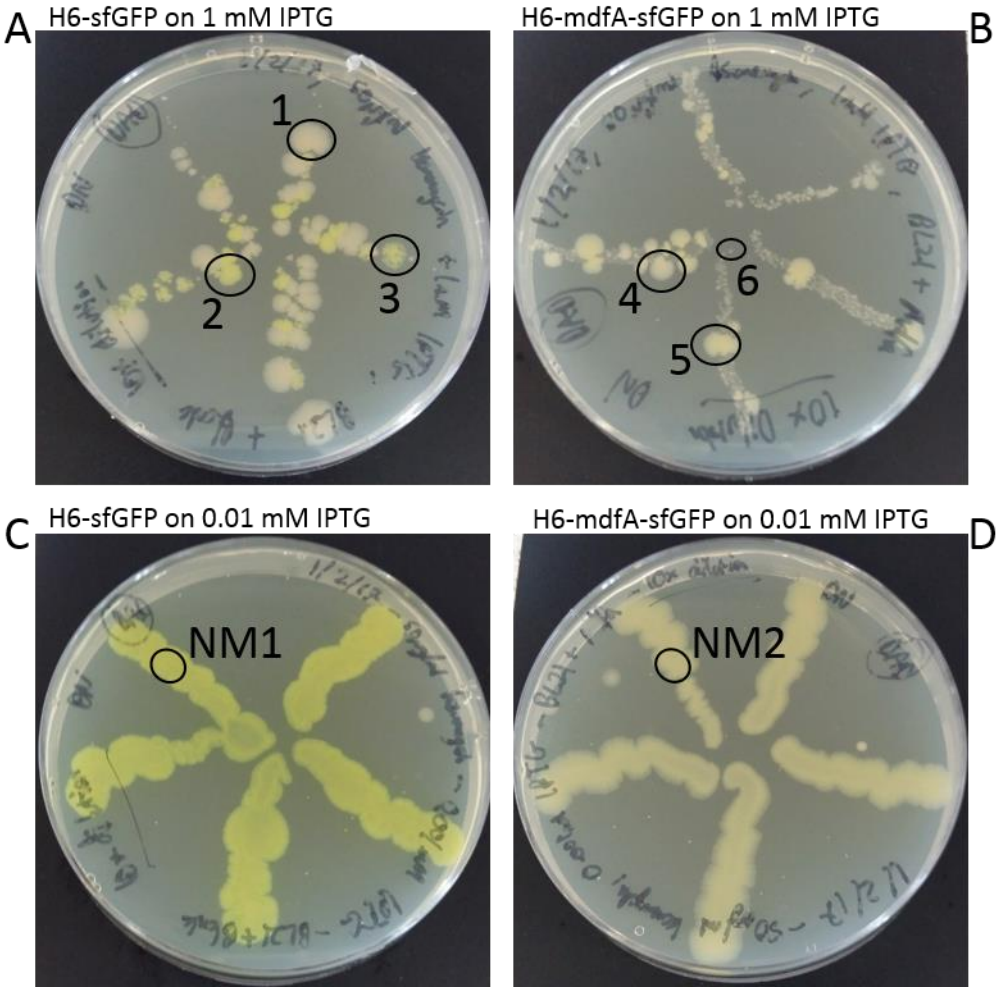


Figure 5.9 Round 1 isolation of BL21 (DE3) E. coli with adaptions to IPTG toxicity, when expressing H6-GFP and of H6-MFS-sfGFP

E. coli BL21 (DE3) were transformed with MFS expression vectors and grown in LB broth to an OD_{600} of 0.1. The cell culture was spread on LB agar plates containing 50 µg/mL kanamycin and either 0.01 or 1 mM IPTG and incubated at 37 ºC for 72 hours. Visual analysis was then used to identify colonies which had adapted to grow at 1 mM IPTG, labelled 1-6. Colonies 1-3 are adapted colonies carrying H_6-sfGFP; colonies 4-6 are adapted colonies carrying H_6-mdfA-sfGFP. NM1 and NM2 are used as controls (NM, not mutant).

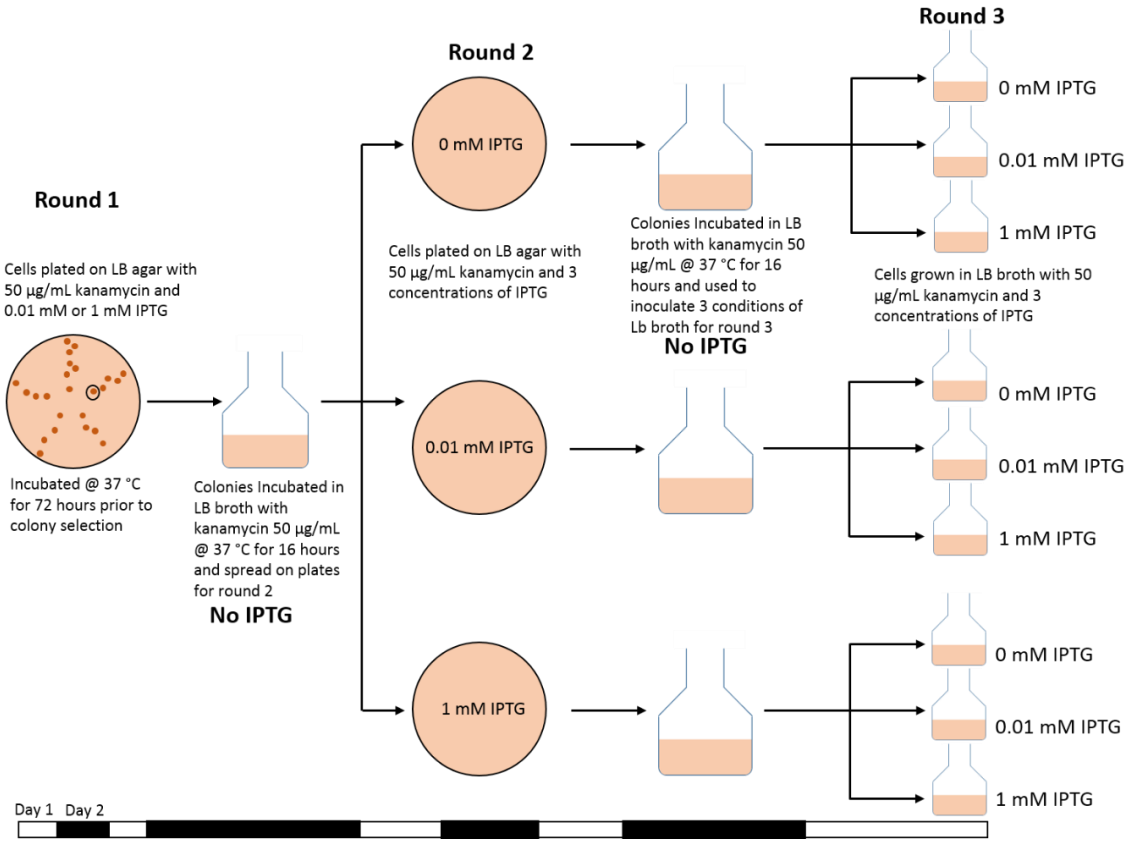


Figure 5.10 Overview of experimental design to test persistence of adaptations for growth on high concentrations of IPTG in liquid LB medium

Round 1(Figure 5.9): *E. coli* were transformed with MFS expression vectors and cultured on LB agar plates containing 50 µg/mL kanamycin and either 0.01 or 1 mM IPTG. Following incubation at 37 ºC, visual analysis was used to identify adapted colonies (colonies 1-6 from Figure 5.9). Round 2 (Figure 5.11-5.18): adapted cells are grown to an OD600 of 0.1 and streaked on plates containing 0, 0.01 or 1 mM IPTG. Plates were incubated at 37 ºC then visually analysed. Round 3 (Figure 5.11-5.18): adapted cells from round 2 were grown overnight in LB broth in the absence of IPTG, then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. The cell culture GFP concentration and density were monitored for 24 hours.

Following round 1 (Fig 5.10), the selected colony was grown overnight in 5 mL of LB plus antibiotic without any IPTG present. The sample was then diluted to OD_{600} of 0.1 before streaking out on three separate LB agar plates containing 50 µg/mL kanamycin and either 0, 0.01 or 1.0 mM IPTG and grown overnight at 37 °C (Round 2). A single colony from these plates was then used to inoculate another 5 mL of LB plus 50 µg/mL kanamycin without any IPTG present and grown overnight at 37 °C. This growth was used to inoculate 3 growth flasks of 40 mL of LB broth to a final cell density of OD_{600} 0.01. Each flask contained 50 µg/mL kanamycin and either 0, 0.01 or 1.0 mM IPTG (Round 3). Over the course of the

next 24 hrs 1 mL samples were taken every 2 hours for the first 8 hours and then a final reading after 24 hrs. For each of these samples the OD_{600} as an indication of growth was recorded as well as the GFP fluorescence which is an indication of the total amount of target protein over-expressed.

Figures 5.11 shows the results for NM1, the non-mutant control expressing the soluble sfGFP protein. If this is not a mutation, then we would expect that the number of colonies decreases with increasing IPTG concentration. The initial plating of the cells was not even, hence it is hard to draw a conclusion from this, even though there are more defined colonies for the 1 mM LB agar plate. However, there is a definite green colouration for the 0.01 mM IPTG plate which is lacking in the higher 1.0 mM IPTG plate. This is because the cells are expected to express protein at 0.01 mM IPTG, but mutations have occurred at 1.0 mM IPTG and these mutations inhibit the production of target protein. Moving onto the growth curves. For the 0 mM and 0.01 mM tests (Fig 5.11, top graphs) both cells grow rapidly as expected, unlike the 1.0 mM IPTG. If as previously noted the 1.0 mM IPTG inhibits cell growth for almost all of the cells, apart from a few mutants which will require a certain amount of time to catch up in terms of cell density and growth rates. This is reflected in the 1 mM growth curve. For protein production, the 0 mM IPTG produces no protein as expected while the 0.01 mM IPTG test produces over 1200 GFP relative fluorescence units (RFU). The expression level for 1 mM test, produces an intermediate curve which would be the case if there is a mix of target protein producing cells and non-producing cells as well as having fewer cells over the same timeframe as indicated in the grow curves.

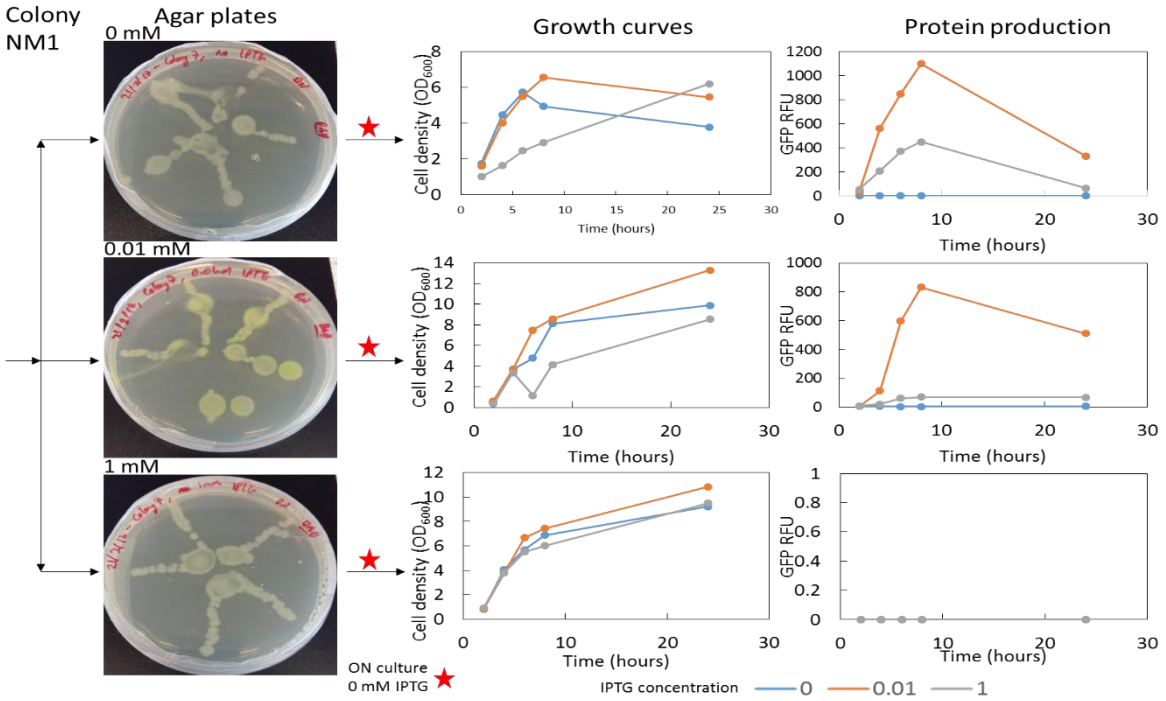


Figure 5.11 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has not adapted to high IPTG: starter colony NM1, carrying H_6-sfGFP

Starter colony NM1 was transformed with H_6-sfGFP and grown on 0.01 mM IPTG agar for 72 hours (Fig 5.9). NM1 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ('agar plates'). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements

The initial 0.01 mM IPTG exposure (Fig 5.11, middle section), shows green colonies that have a similar profile as described for the 0 mM IPTG exposure as expected for essentially WT cells. This time both the maximum sfGFP expressions are decreased for the 0.01 and 1.0 mM IPTG samples in comparison to the previous 0 mM initial IPTG exposure.

Significantly for the 1.0 mM initial IPTG exposure (Fig 5.11, bottom section), the white colonies, which are non-expressing mutants, do not express protein at any IPTG concentration while the cells are adopted to grow well at any IPTG concentration hence the three growth curves are very similar.

The results for the second control, NM2 non-mutant for transformant with the IMP H_6-mdfA-sfGFP are shown in Figure 5.12. The overall results follow a similar pattern as NM1 (Fig 5.11) with cell viability decreasing with increasing IPTG concentration on the LB agar plates, generally cell growth for the 0 and 0.01 mM IPTG samples being similar while the 1 mM IPTG growth curves takes longer to reach the maximum growth. Protein production also occurs for the 0.01 and 1 mM tests following the initial 0 and 0.01 mM IPTG exposures, albeit, approximately two orders of magnitude lower than for the soluble sfGFP protein. Differences between NM2 and NM1 samples are seen in the growth curve for the 1 mM IPTG and the protein production for the 0.01 mM IPTG test following 1 mM initial exposure (Figure 5.12 and 5.11, bottom sections). Repeat experiments would have to be performed to see if these were outliers or actual results. However, overall, both controls display a lack of adaptations, with higher IPTG hindering growth in liquid medium.

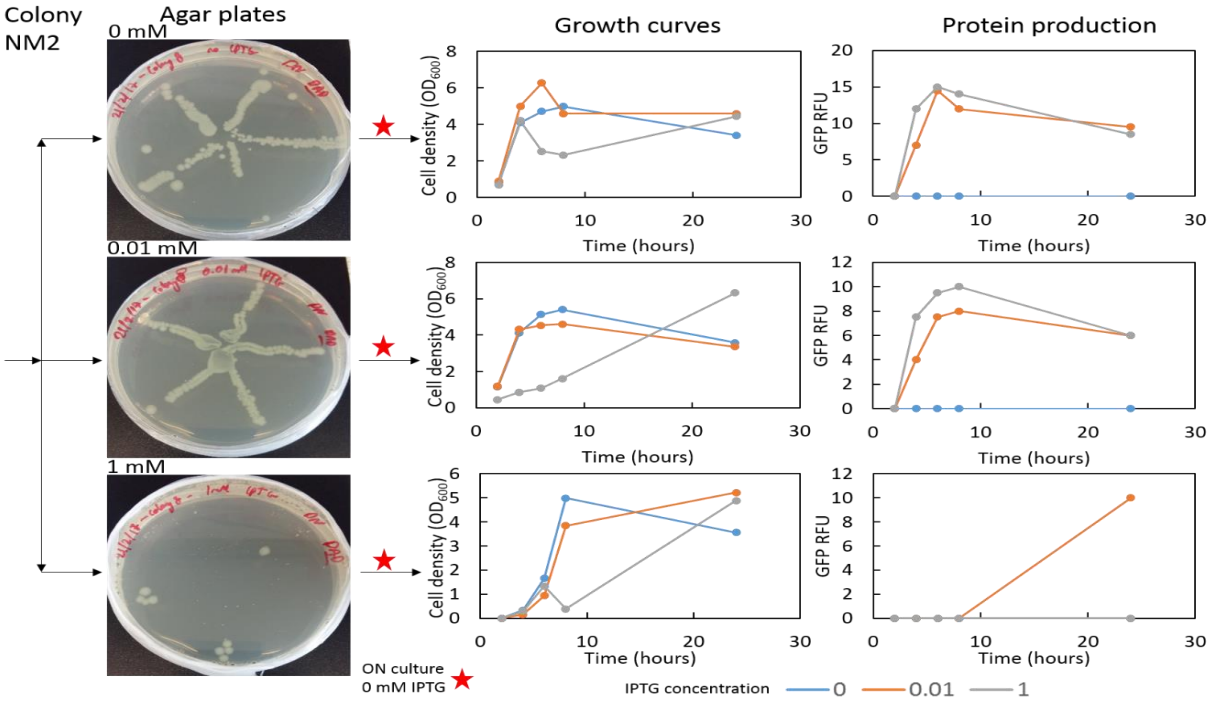


Figure 5.12 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has not adapted to high IPTG: starter colony NM2, carrying H_6-mdfA-sfGFP

Starter colony NM2 was transformed with H_6-mdfA-sfGFP and grown on 0.01 mM IPTG agar for 72 hours (Fig 5.9). NM2 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

Moving onto the mutants, the results for the soluble sfGFP and mdfA from uniform, non-coloured colonies are shown in Figure 5.13 and 5.16, respectively.

As expected for both target proteins, the cells grow equally as well at all IPTG concentrations on LB agar, the growth curves are nearly identical, and protein is not produced at any initial IPTG exposure or at any IPTG concentration tested.

Another group of mutants displayed uniform colonies that were coloured green, implying that these mutants could express protein. The results for these groups of sfGFP and mdfA mutants are shown in Figures 5.14 and 5.17, respectively.

Generally, these results agree with our expectations that these mutants are unaffected by increasing IPTG concentrations either on LB agar or in their growth curves while this time they are able to express target protein either soluble or membrane proteins in the presence of IPTG. It is unclear why no mdfA protein was expressed for the initial 1.0 mM IPTG exposure at any IPTG concentration (Fig 5.17, bottom section, right hand graph).

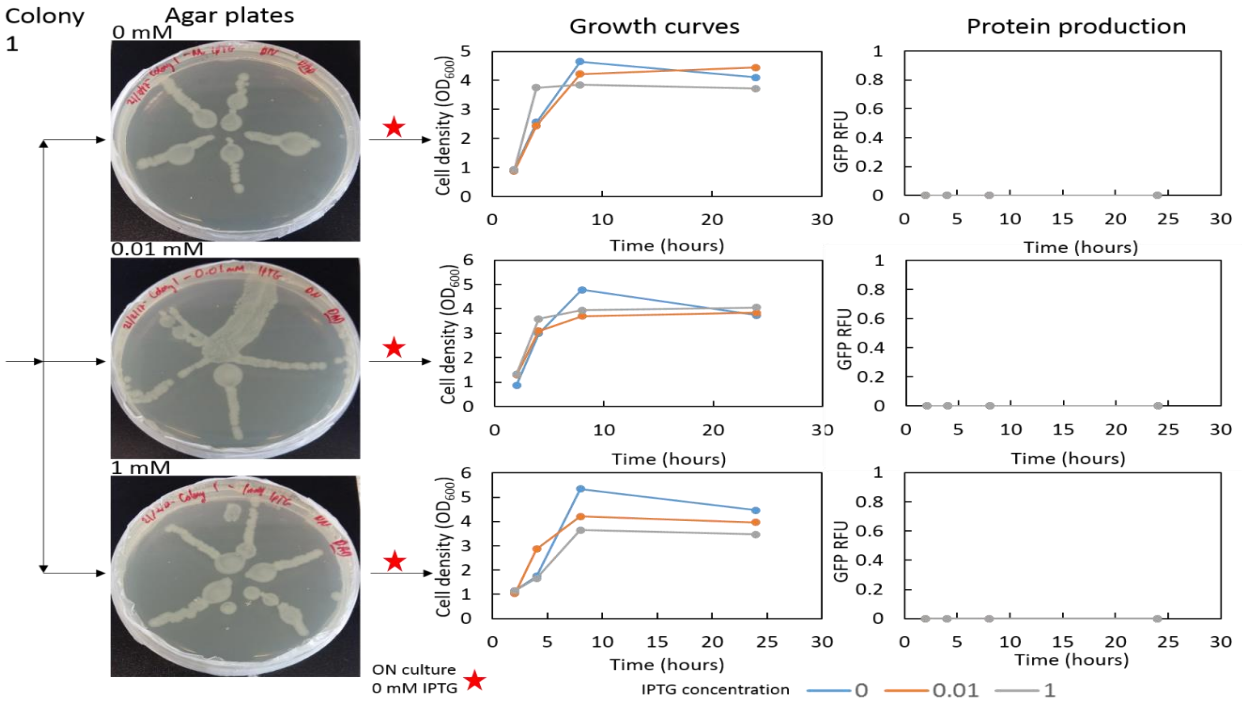


Figure 5.13 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 1, carrying H_6-sfGFP

Starter colony 1 was transformed with H_6-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 1 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

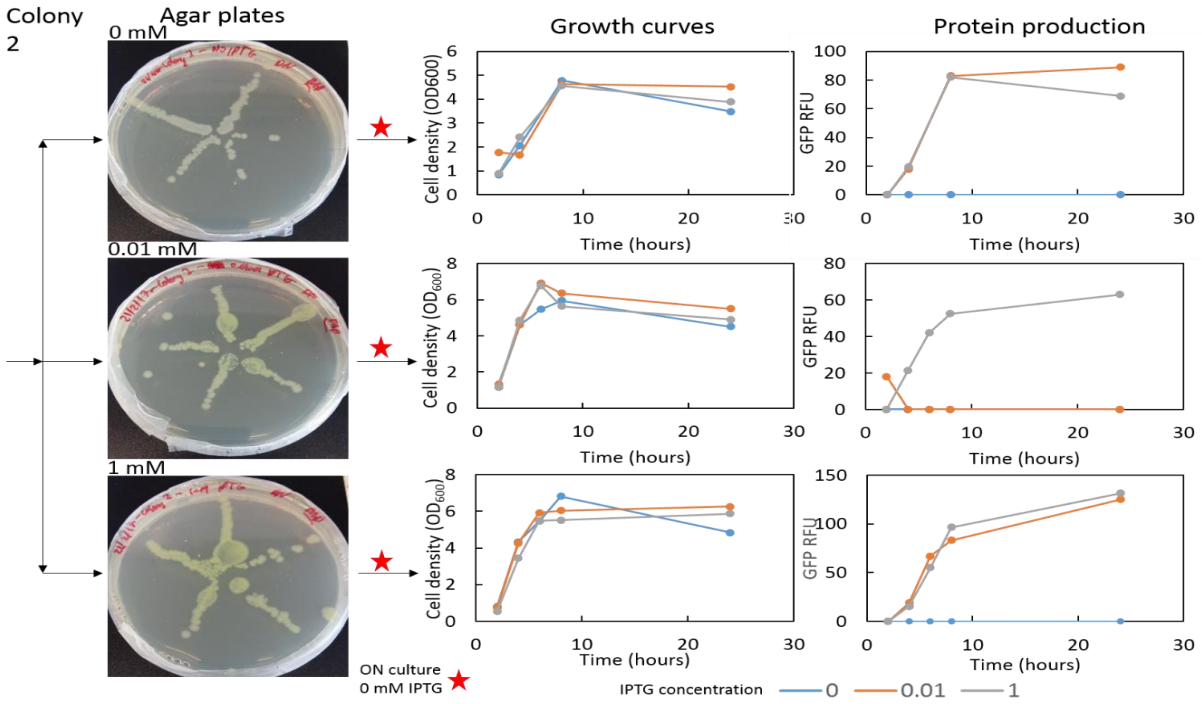


Figure 5.14 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 4, carrying H_6-mdfA-sfGFP

Starter colony 4 was transformed with H_6-mdfA-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 4 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

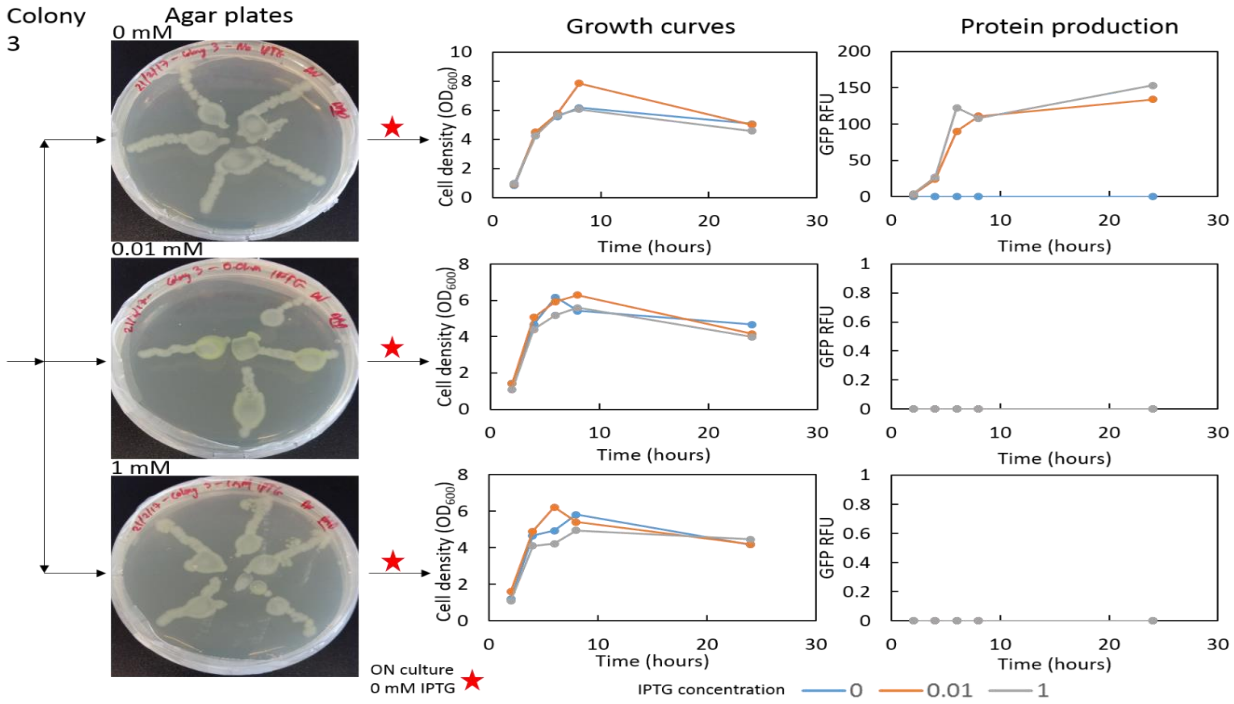


Figure 5.15 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 2, carrying H_6-sfGFP.

Starter colony 2 was transformed with H_6-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 2 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

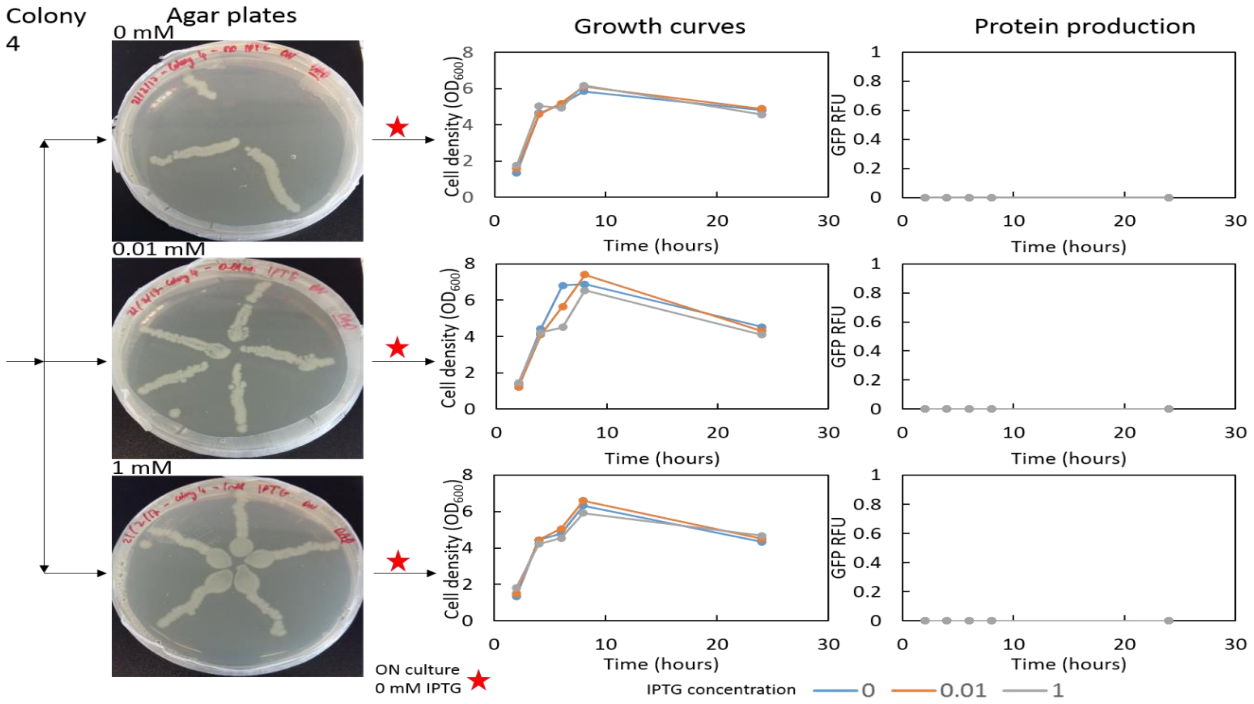


Figure 5.16 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 5, carrying H_6-mdfA-sfGFP

Starter colony 5 was transformed with H_6-mdfA-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 5 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

The results for colony 3 mutant (H6-sfGFP, medium sized colony, punctate, no green) are shown in Figure 5.15. The similar growth curves in all experiments implies mutations as previously noted. However, the 0.01 mM IPTG LB agar plate (Fig 5.15, middle section) displays both green and white colouration, clearly indicating a mix of colonies. The clear colony was selected for growth; expression and protein production curves both correspond to a non-target protein expressing mutant. Based on this LB agar plate result it was felt that the data could not be relied upon as it was a mix of colonies.

Figure 5.18 shows the results for the last mutant, colony 6. This mutant produced data comparable to the previous data, leaving unanswered questions such as why was it able to produce target protein when the original colony was colourless? Why were the 0.01 IPTG initial exposure growth curves not similar if this is a mutant (Fig 5.18, middle section)? If it is a mutant, then why are there almost no colonies on the 1.0 mM IPTG LB agar plate? The underlying reasons for these results are unexplained. A summary of the findings are shown in Table 5.3.

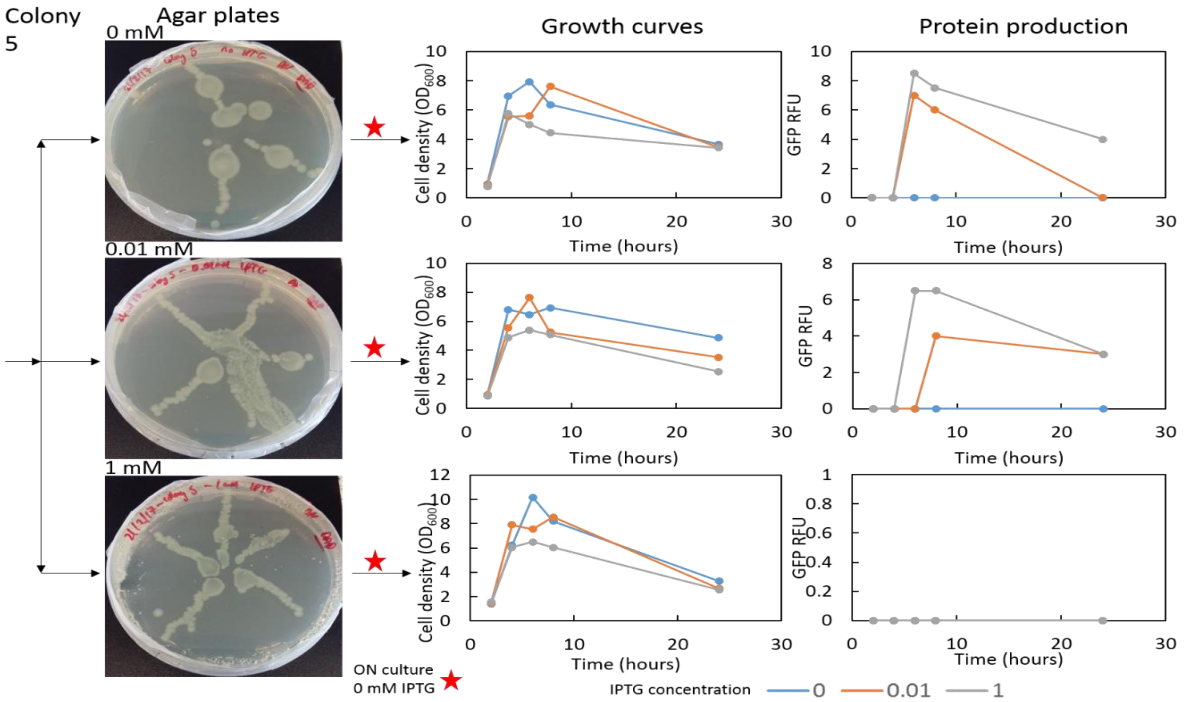


Figure 5.17 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 3, carrying H_6-sfGFP

Starter colony 3 was transformed with H_6-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 3 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

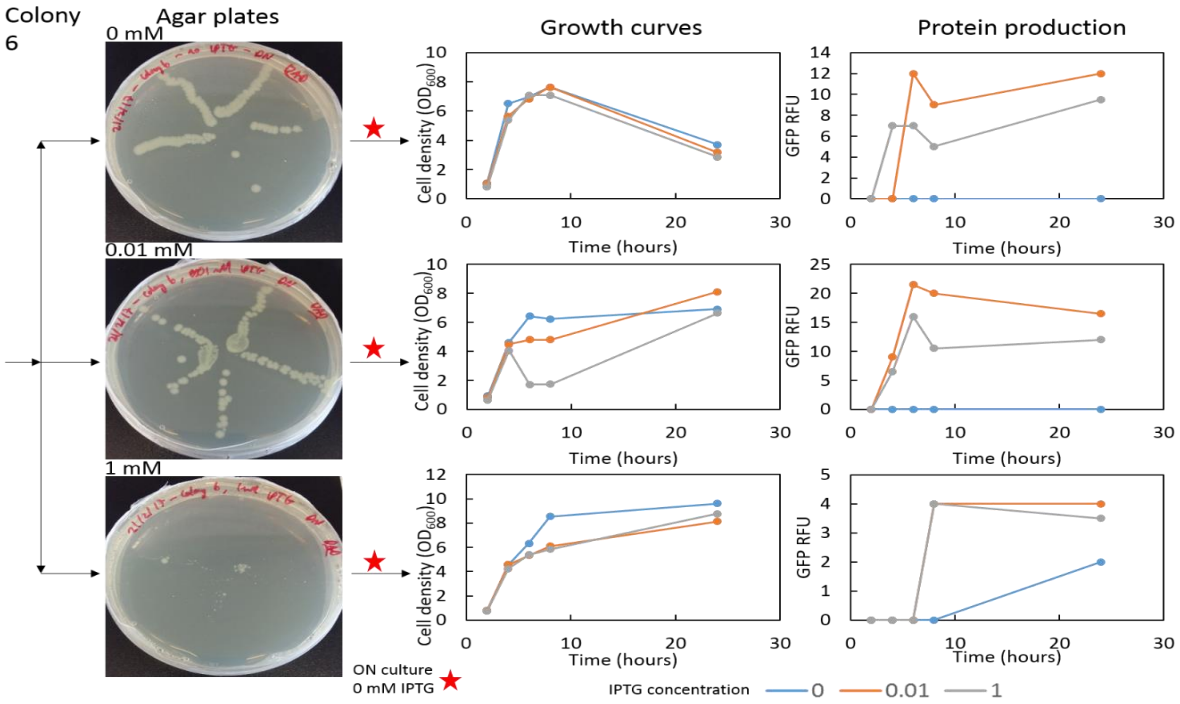


Figure 5.18 Investigation into the persistence of *E. coli* BL21 (DE3) adaptions to IPTG when expressing a colony which has adapted to high IPTG: starter colony 6, carrying H_6-mdfA-sfGFP

Starter colony 6 was transformed with H_6-mdfA-sfGFP and grown on 1 mM IPTG agar for 72 hours (Fig 5.9). Colony 6 was incubated in the absence of IPTG then transferred to new plates containing either 0, 0.01 or 1 mM IPTG ("agar plates"). Plated cells were grown overnight in LB broth containing no IPTG (step represented by the red star) and then transferred to LB broth containing either 0, 0.01 or 1 mM IPTG. Cell culture OD_{600} (growth curves) and GFP concentration (protein production) were monitored at intervals for 24 hours. All graph data points are the mean of three measurements.

Table 5.3 Summary of colonies selected for investigation into persistence of adaptions to high IPTG concentrations

Cell lines displaying R2 toxicity were mostly unable to grow on the 1 mM IPTG concentration agar plate in round 2 experiments.

Colony	Construct	R1 IPTG concentr-ation	GFP produc-ing	R2 toxicity	Time (hr) taken to reach max OD_{600} in 1 mM IPTG at R3	Highest GFP RFU reading
NM1	H6-sfGFP	0.01 mM	Yes	Yes	10 - 24	1200
NM2	H6-mdfA-sfGFP	0.01 mM	Yes	Yes	10 - 24	16
1	H6-sfGFP	1 mM	No	No	8	0
2	H6-sfGFP	1 mM	Yes	No	6 - 8	140
3	H6-sfGFP	1 mM	Yes	No	6 - 8	160
4	H6-mdfA-sfGFP	1 mM	No	No	8	0
5	H6-mdfA-sfGFP	1 mM	Yes	No	6 - 8	9
6	H6-mdfA-sfGFP	1 mM	Yes	Yes	6 - 24	23

As an initial characterisation, we could say that there are at least two types of similar mutations between the soluble sfGFP and the membrane protein H6-mdfA-sfGFP. This is based on the characteristics outlined in Table 5.4. Hence, colonies 1 (H6-sfGFP) and 4 (H6-mdfA-sfGFP), which we will call Mutant X, are likely to have similar mutations while colonies 2 (H6-sfGFP) and 5 (H6-mdfA-sfGFP), which we will call Mutant Y, are also similar to one another but significantly different as compared to Mutant X. Mutant X completely knocks out target protein expression while Mutant Y allows for some protein expression at levels significantly lower than the non-mutant samples (Table 5.3).

Table 5.4 Highlighting similar phenotypes between the soluble sfGFP and mdfA membrane protein

Colony number (mutant variant)	**Plasmid vector**	**Growth impediment at IPTG concentration (mM)**			**Protein production at IPTG concentration (mM)**		
		0.00	0.01	1.00	0.00	0.01	1.00
NM1	H_6-sfGFP	No	No	**Yes**	No	**Yes**	**Yes**
NM2	H_6-mdfA-sfGFP	No	No	**Yes**	No	**Yes**	**Yes**
1 (X)	H_6-sfGFP	No	No	No	No	No	No
4 (X)	H_6-mdfA-sfGFP	No	No	No	No	No	No
2 (Y)	H_6-sfGFP	No	No	No	No	**Yes**	**Yes**
5 (Y)	H_6-mdfA-sfGFP	No	No	No	No	**Yes**	**Yes**
3	H_6-sfGFP	No	No	No	No	**Yes**	No
6	H_6-mdfA-sfGFP	No	No	**Yes**	No	**Yes**	No

Finally, to confirm if the mutations are localised to the genome or the plasmid DNA, plasmids were extracted from the colonies and retransformed into fresh competent BL21 (DE3) cells. If a mutation occurs in the genome and not in the plasmid, removal of the plasmid from a mutant cell line and transforming back into WT should produce WT behaviour i.e., decreasing colonies with increasing IPTG concentrations and phenotype variants. On the other hand, if a mutation occurs in the plasmid and not in the genome, a similar experiment should produce colonies that are unaffected by increasing IPTG concentrations. Figure 5.19 shows the results from this experiment.

All retransformants behaved like WT, hence, the mutations are all expected to reside within the genomic DNA. Interestingly, a non-target protein producing mutant was formed for colony 1 at 1 mM IPTG, demonstrating the selection pressure is causing a repetition of adaptations.

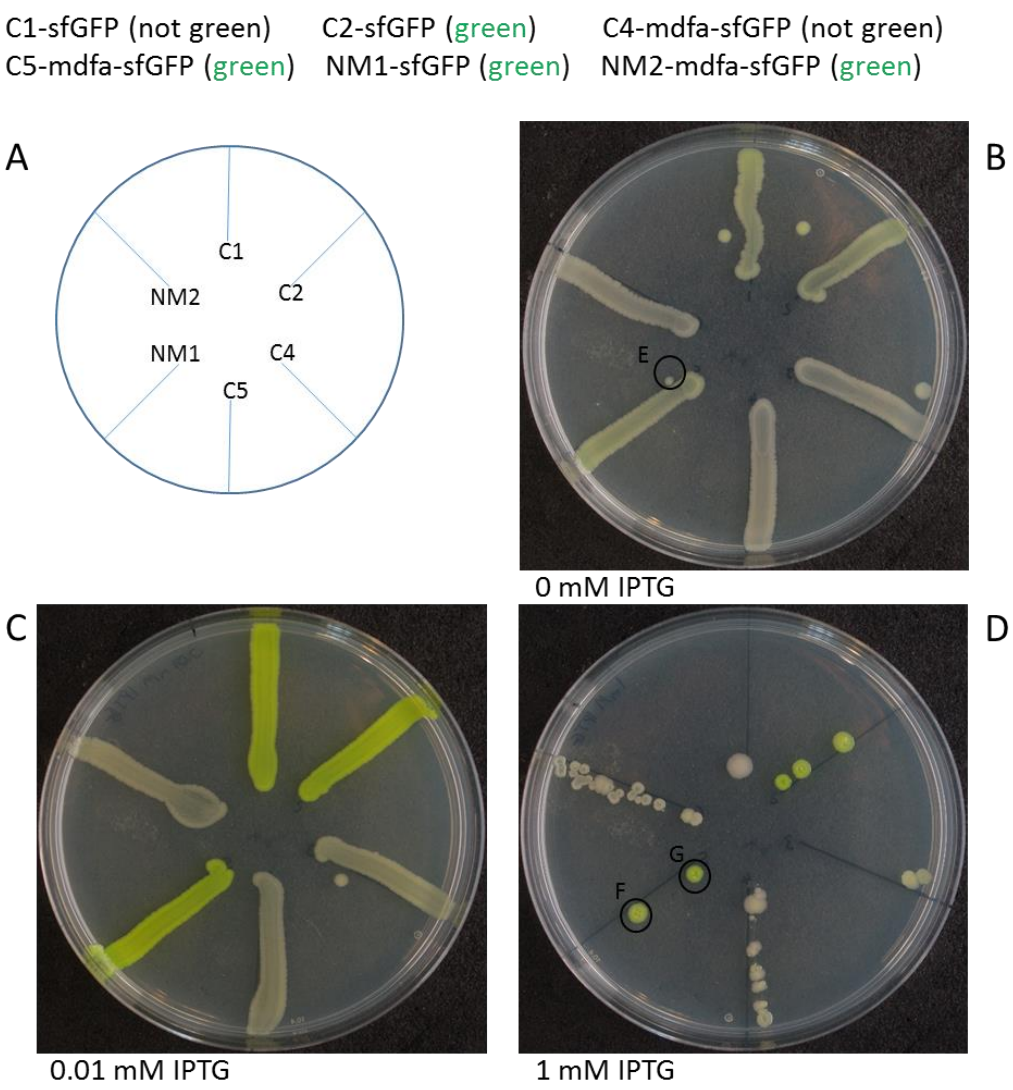


Figure 5.19 Extraction of plasmids from adapted cell lines indicated the resistance to high IPTG concentrations were genomic

Plasmids were extracted from transformed cell lines which had developed an adaptation to high levels of IPTG. These plasmids were then transformedinto WT BL21 (DE3). Newly transformed cells were grown in LB media and then plated onto LB agar plates with 0, 0.01 or 1 mM IPTG concentrations as indicated in the figure. A) position of strains on plate, corresponding to key at top of figure, e.g. C2 is transformed with the sfGFP plasmid from adapted colony C2.

To investigate the location of any genomic mutations leading to the observed adaptations, genomic DNA was extracted from cell lines in preparation for sequencing. Genomic DNA was run on a gel alongside WT BL21 (DE3) genomic DNA. Figure 5.20 indicates successful genomic preparations have been isolated for cell lines 1, 2, 4, 5, NM1 and NM2. It was not possible to perform genomic sequencing within the timeframe of this project. Circle colonies were also selected for further testing, see Section 5.2.2.

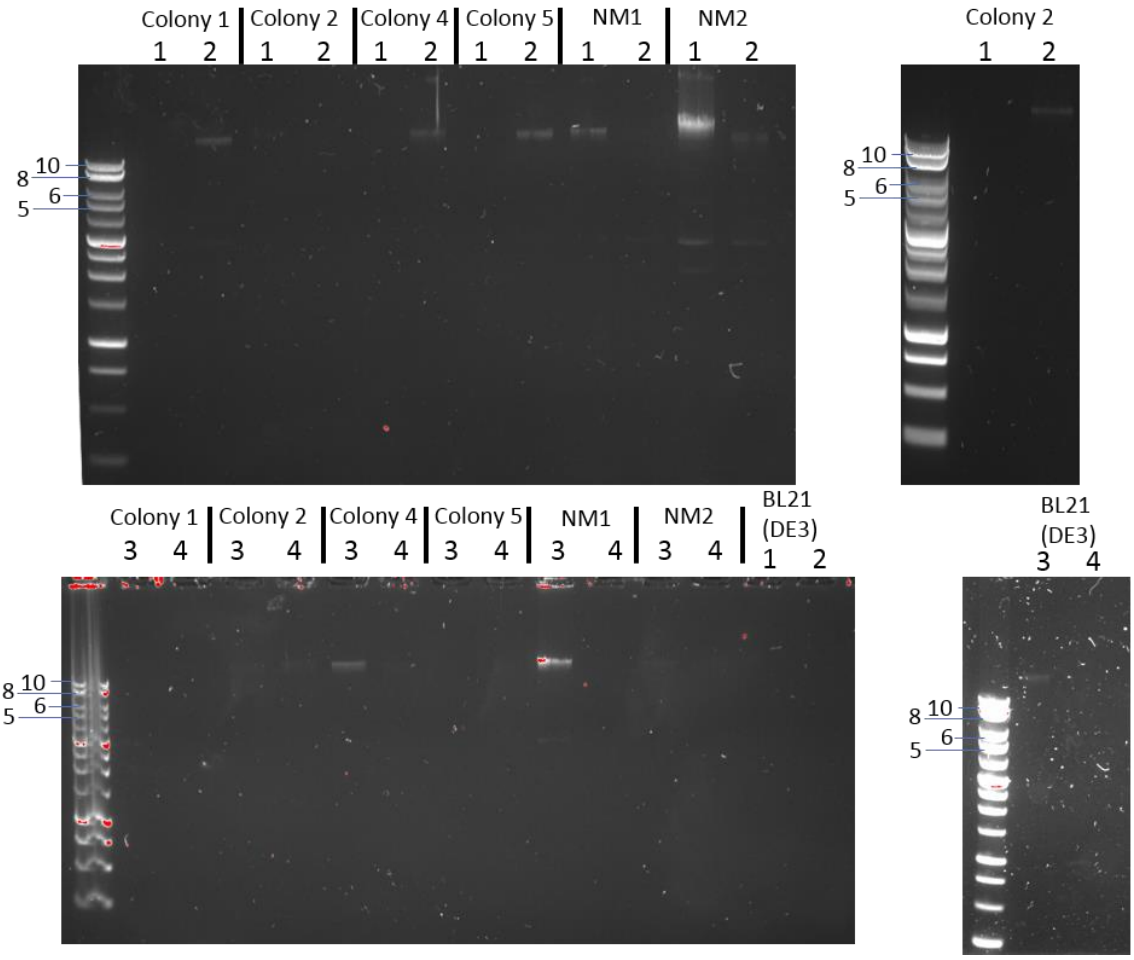


Figure 5.20 Genomic DNA extracted from cell lines displaying persistent adaptions to high concentrations of IPTG

Genomic DNA was extracted from cell lines which displayed adaptations to high IPTG concentrations, after overnight growth from frozen glycerol stock. Genomic DNA was analysed using gel electrophoresis. Multiple samples were extracted from some of the colonies (samples 1-4). WT BL21 (DE3) was included as a control. Molecular weight marker in left lane of gel, with sizes (kbp) labelled to the left of the gel.

5.2.2 Genomic adaptions to high concentrations of IPTG may not improve the yield of expression systems

To investigate the expression capabilities of cells with adaptations, the cells transformed with the plasmid from colony NM1, H_6-sfGFP were tested for expression. Cell lines are circled in Figure 5.19. Colony E has had no exposure to IPTG. Colonies F and G are adapted cell lines grown on 1 mM IPTG, both display uniform GFP expression. Colonies E, F and G were used to inoculate 50 mL of LB broth in a 250 mL flask to a final OD_{600} of 0.005 and grown at 37 ºC, 150 rpm. Upon reaching an OD_{600} of 0.8 growths were inoculated with 0.01, 0.2 or 1 mM IPTG and expression left overnight at 18 ºC. Cell cultures were pelleted and visually

analysed, shown in Figure 5.21. A green pellet indicates GFP is produced at high levels. The Colony which had not developed adaptations to grow on high IPTG levels appeared most green, indicating they produced the most GFP. 0.01 and 0.2 mM IPTG induction was most effective, with visibly reduced expression using 1 mM induction. Neither of the cell lines adapted on 1 mM IPTG growth expressed visible amounts of protein. The adaptions allowing these mutant colonies to survive on 1 mM IPTG agar plates appear to hinder protein expression.

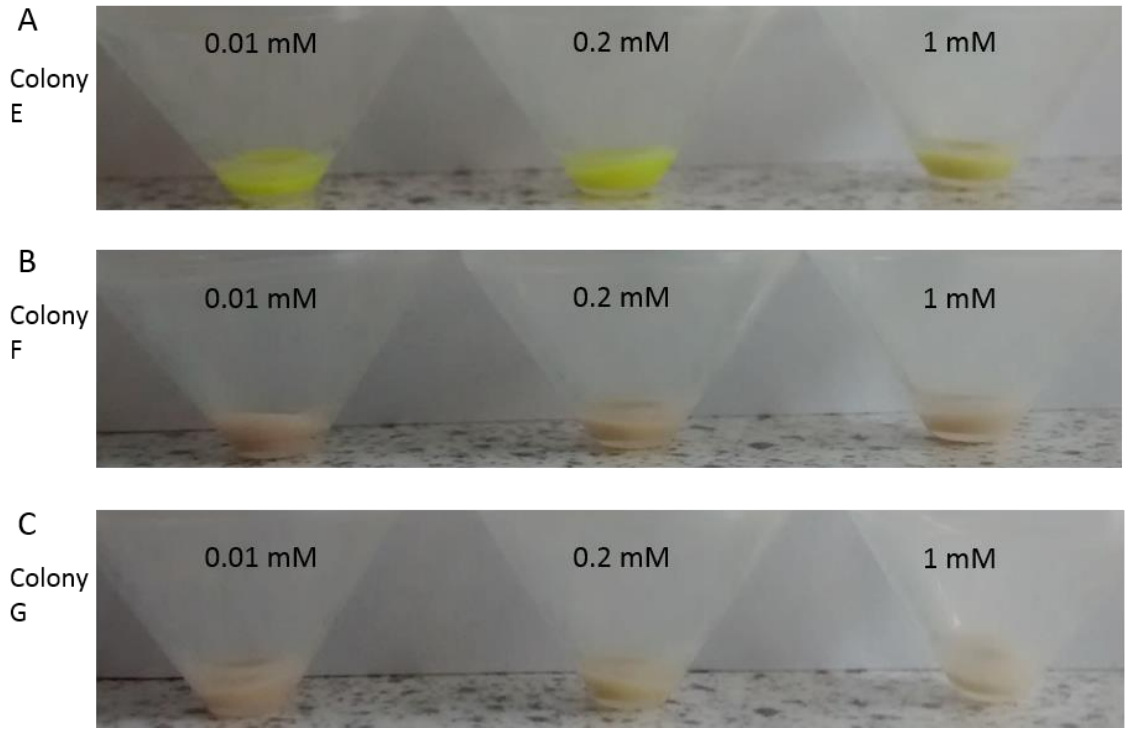


Figure 5.21 Mutant cell lines with persistent adaptations to high IPTG concentrations may have reduced protein expression levels.

A) Colony E: BL21 (DE3) cells transformed with H6-sfGFP and grown on 0 mM IPTG. B) Colony F and and C) colony G are BL21 (DE3) cells transformed with H6-sfGFP and grown on 1 mM IPTG. Cells were grown in LB broth OD_{600} of 0.8 then induced with either 0.01, 0.2 or 1 mM IPTG, as labelled. Cells were harvested with centrifugation.

5.3 Discussion

This chapter aimed to investigate cellular toxicity after expression of membrane proteins, using sfGFP as a reporter. Around 25 % of genes encode membrane proteins (Sundararaj et al 2004; Fagerberg et al 2010), yet, they make up <2 % of the characterised structures (Kozma et al 2013). The production and purification of membrane proteins is associated with additional difficulties when compared to soluble proteins (Lin and Guidotti 2009) and may account for the disparity of information between the protein types. The most popular protein expression system is *E. coli* (Rosano and Ceccarelli 2014), but the observed toxic effects of protein over-expression makes it unsuitable for characterisation of many proteins (Dumon-Seignovert et al 2004). Cell lines with a greater resistance to these toxic effects have been developed under high IPTG conditions (Miroux et al 1996), in spite of their availability for over twenty years membrane proteins are still under-characterised relative to soluble proteins.

GFP was used as a reporter, to investigate the adaptations in *E. coli* to express toxic membrane proteins when induced by IPTG. Plasmids expressing either GFP alone or a range of membrane proteins tagged with GFP were transformed into BL21 (DE3) cells. GFP was used as a reporter to screen for colonies adapted to persist under high IPTG conditions, which could correspond to colonies adapted to express membrane proteins to relatively high levels.

5.3.1 Expression of T7 polymerase does not negatively impact cell viability

Initial control tests investigated the effect of T7 polymerase production in *E. coli*. The BL21 (DE3) cell line has a genomic copy of T7 polymerase under the control of a lacUV5 promoter, which can be used to express proteins under the control of a T7 polymerase promoter (Studier and Moffatt 1986). BL21 and BL21 (DE3) are the expression systems for around 84.5 % of protein data bank structures (Berman et al 2000). They are also the basis for creation of several adapted expression systems such as LEMO21 (DE3) (Schlegel et al 2012) and KRX (Miladi et al 2013) which also express T7 polymerase when exposed to lactose or IPTG. The cell line Mach1 does not have a copy of T7 polymerase and was used as a control.

BL21 (DE3), LEMO21 (DE3), KRX and Mach1 were grown on plates supplemented with up to 1 mM IPTG (Fig 5.2), or up to 43.8 mM lactose (data not shown). There was no difference in growth between the control Mach1 cell lines and those

encoding T7 polymerase. This indicates that the production of T7 polymerase alone does not cause a toxic effect. This is consistent with previous findings, which have reported toxicity to be caused by problems with the recombinant protein overexpression, rather than problems with T7 polymerase (Wagner et al 2007; Gubellini et al 2011).

5.3.2 Most cells carrying expression plasmids are lethally impacted by IPTG

As T7 polymerase did not induce cellular toxicity, the next step was to investigate how toxicity was affected by the presence of plasmids containing T7-inducible protein. Protein would not be expressed when cells were grown on 0 mM IPTG, but would be induced at higher levels of IPTG, upon induction of T7 polymerase. Expression vectors were generated in Chapter 4 and carry native *E. coli* MFS proteins tagged with H6 and sfGFP at the N- and C-terminus, respectively (H_6-MFS-sfGFP) or H_6-tagged sfGFP alone (H_6-sfGFP).

No cellular toxicity was observed when grown on LB agar plates without IPTG supplementation. However, lethality was observed when both H_6-sfGFP and H_6-IMPs-sfGFP were exposed to IPTG, in a dose dependent manner. There did appear to be a slight variation in toxicity depending on the vector insert, with amtB-GFP, mdtG-GFP and setB-GFP expression causing slightly greater toxicity than bcr-GFP, mdtG-GFP and GFP expression (Fig 5.4). However, this would require more repeats to draw reliable conclusions. Induction of these proteins causing cellular toxicity corresponds with previous findings, which suggest toxicity is caused by misfolding and aggregation of overexpressed protein, or interruption of cellular machinery by overexpressed protein (Wagner et al 2007; Gubellini et al 2011).

Most previous work has highlighted the toxic effect of non-native protein expression in *E. coli* (Dumon-Seignovert et al 2004). Here, overexpression-related toxicity is demonstrated with the native MFS proteins. However, expression of GFP also leads to cellular toxicity, with reduced growth at higher IPTG concentrations, albeit not to the same extent as expression of bcr-GFP, setB-GFP or amtB-GFP (Fig 5.4). This indicates GFP is not inert in the expression system if it is causing toxic effects when expressed. However, it may be interesting to investigate whether it is the vector itself causing this toxic effect, as empty, non-GFP carrying pET vectors have also been shown to be toxic to BL21 (DE3) cells when exposed to IPTG (Miroux et al, 1996). Significantly, our data points to a previously unrecognised mechanism of toxicity, which, is the actual process of over-expression is damaging. We

speculate that this is based on too much target protein mRNA overwhelming the endogenous protein production facilities.

5.3.3 Cell lines can adapt to grow on high IPTG concentrations

Although growth is clearly inhibited when cell lines are grown on high levels of IPTG, certain colonies are still able to grow up to concentrations of 1 mM IPTG, suggesting they may be adapted to avoid the toxic effects of protein over-expression. There are two visible variations in these adapted colonies. While some appear green, indicating GFP is present and the inserted protein is still being expressed, others are light brown, suggesting complete inhibition or termination of exogeneous protein expression. These adaptions were shown to be persistent by alternating growth with and without IPTG (Fig 5.11-5.18). The most likely explanation for the adaptions effects is genomic mutations that are passed down through successive generations even in the absence of the selection pressure.

Where adaptations have led to a reduction in recombinant protein expression, it is possible the adaptation has interrupted the interaction between the T7 RNA polymerase and the T7 promoter site in the vector. Our data indicates that the most likely mutation sites are genomic as plasmids stocks extracted from mutants were shown to match behaviour of fresh plasmid stocks (Fig 5.19). Genomic DNA was successful isolated and genomic sequencing is the next appropriate step to investigate the phenomenon but was not feasible in the timeframe of this study. It would be interesting to identify any changes in the genome of adapted colonies and determine whether they are related to the T7 polymerase or affect elsewhere in the genome. Previous genomic adaptations to expression toxicity highlight four areas that are altered; the plasmid promoter region for T7 polymerase (Simmons and Yansura 1996), the promoter recognition site of T7 polymerase (Baumgarten et al 2017), the lactose binding site of the lac repressor protein (Kwon et al 2015) and the T3 promoter site controlling T7 expression (Schlegel et al 2015).

5.3.4 The possibility of using GFP as a reporter for cellular toxicity

A main aim of this chapter was to investigate the expression capabilities of adapted colonies. The GFP reporter clearly indicated whether cultures with adaptations were expressing protein. The effect this would have on protein yield was explored visually (Fig. 5.21): adapted colonies were induced with IPTG and compared to non-adapted BL21 (DE3), with the former showing a dramatic reduction in GFP expression levels. This is consistent with predictions that the interaction between

T7 RNA polymerase and the T7 promoter of the plasmid have been interrupted. Therefore, selecting cells that can grow on high IPTG may not be an effective method for high level protein expression and expression-induced toxicity may not be necessary to produce protein. If by developing a protocol for achieving high levels of expressed protein based on our findings, potential improvements may be achieved by screening lower concentrations of IPTG as part of a normal protein over-expression protocol. This is made simple with a GFP tag but could be possible using Western blot analysis where a visible tag is not appropriate.

Massey-Gendel et al (2009) suggested specifically screening for cell lines producing high protein expression when approaching a new protein. This was achieved by expressing their protein fused with a sequence providing antibiotic resistance, selecting for highly expressing cell lines. As this experiment utilised the pBAD plasmid, a native *E. coli* polymerase it avoids the toxic effects of overexpression using T7 polymerase (Massey-Gendel et al 2009). This demonstrates that is could be possible to select for highly expressing cells as a method for improving cellular yields. GFP would provide a measurable means of doing this. Proteins tagged with GFP have been shown to express comparatively poorly and previous work has shown removing this tag can improve protein expression levels (Hammon et al 2009). This could mean it is possible to develop a system using GFP as a reporter for its effectiveness, but that adapted expression systems may be more effective once the GFP tag has been removed.

A key point may be that as GFP does not appear to be lethal to cells, therefore, at lower concentrations of IPTG there is no benefit to adaptive mutations. Additionally, adaptations to IPTG-induced protein expression lead to a drastic decrease in protein yield. However, some proteins have been shown to be so toxic even leaky expression is lethal (Dumon-Seignovert et al 2004). Where this is this case, adapted strains could reduce lethality and improve protein expression, but there is little evidence for this being successful. As mentioned in the introduction, C41 and C43 are cell lines that could be used for this purpose. However, there are few examples of protein structures being solved that utilise the C41 or C43 expression systems (He et al 2014). Although a handful of published papers have used C41 and C43 systems for expression, purification and some initial diffraction or circular dichroism experiments (Smith and Walker 2003; Terakado et al 2010; Tait and Straus 2011; Dey and Ramachandran 2014; Pandey et al 2014). While it is possible to express toxic proteins in these cell lines, it is currently easier to investigate a homologue that expresses more effectively, meaning the system is

not widely used. This may be due to the drastic reductions in protein yield in these adapted cell lines, as seen in this chapter.

5.3.5 The generation of mutation

In our series of experiments, we demonstrated that mutants were viable when over-expressing target proteins using IPTG concentrations above 0.1 mM IPTG. In the beginning, we assumed that these mutants were naturally generated within the growing *E. coli* population due to the non-perfect DNA copying mechanism that exists in this organism. What we cannot tell from our experiments is whether there are natural variations present within the population of cells or if the act of target protein over-expression increases the mutation rate thus resulting in the generation of these mutants. It would be interesting to uncover these details as it might resemble some of the mechanism by which bacteria become resistant to toxic substances such as antimicrobials.

5.3.6 In conclusion

As shown in this chapter, protein over-expression using IPTG concentrations greater than 0.1 mM IPTG selects for mutants that either inhibit protein production. This has a significant impact on health and safety issues related to laboratory safety as all laboratories state in their safety protocols that they are not producing genetically modified organisms. Our results clearly contradict this. Significantly, if we look at the latest manual for protein over-expression using the *E. coli* pET system (Novagen 2005) it states that '… for "plain" T7 promoter vectors, use 200 µL IPTG for a final concentration of 0.4 mM'. At this IPTG concentration the recommended protocol is actively selecting for mutants.

This chapter has highlighted the use of GFP as a way to speed up characterisation of cell lines adapted to the toxic effects of target protein over-expression. It provides a simple, quicker way to screen expression and identify cell lines which were no longer able to express recombinant protein. There may be some adaptations capable of balancing a reduction in toxicity with a useful protein yield, and GFP could be used to screen adapted cell lines for these cell lines. An alternative, directed approach could utilise the same vectors but with intentional mutations in the regions featuring possible changes once the full range of adaptive mutation are known.

Chapter 6 Discussion

This thesis is centred around the use of GFP protein to answer questions on protein localisation, protein function and cellular response to stimuli. Each of these processes are key to the role proteins play in cellular physiology. The progress of sequencing technology has provided vast amounts of data identifying novel proteins (Wiemann et al 2001), with over 60 million sequences now contained in the UniProt database. However, less than 1 % of UniProt protein sequences have been manually annotated with functional information extracted from the literature (The UniProt Consortium 2017). Due to the magnitude of this task, the majority have only been analysed computationally. Comparison with previously characterised proteins can be used to predict the function of novel proteins from their sequence. The accuracy of this technique is dependent on the similarity between the query and known proteins (Loewenstein et al 2009). Computational predictions can be informative; however, they still need to be verified, through the individual characterisation of novel proteins and development of new methods to speed up investigations. This is particularly pressing for membrane proteins as they can make up to a quarter of the genome (Sundararaj et al 2004), yet, they make up <2 % of the characterised structures (Kozma et al 2013). The metabolic state and therefore function of any cell depends on what comes in and what goes out of the cell. From this perspective, it is the gateways through the cells that define the role of the cell thus the membrane proteins are central to what sort of cell it is. As exciting as integral membranes are there are drawbacks when studying them such as difficulties with protein production, purification and characterisation (Moraes et al 2014). The lack of information on the functional roles for many integral membrane proteins and the physical difficulties when dealing with them has led to their under-representation in the literature.

To tackle several of these goals concurrently, the protein tag GFP was selected for its relatively inert nature, proven ability in localisation studies (Zimmer 2002) and potential in the development of functional studies.

6.1 GFP-tagged protein enables investigations into protein function, including substrate interactions and subcellular localisation

This thesis has demonstrated the diversity of GFP during the functional characterisation of proteins. A GFP-tag provided the framework for investigations into subcellular localisation and protein-substrate interactions, both are key determinants of protein function.

Subcellular localisation, can be investigated with techniques such as fluorescence-tagged proteins (Vilardaga et al 2002; Meissner et al 2011) and immunohistochemistry (Liu 2002). GFP-tagged HsSWEET was used to identify subcellular localisation in Chapter 3, providing insight into how it may function. In Chapter 4, thermophoresis successfully utilised GFP fluorescence to calculate binding affinities of MFS protein-substrate interactions (Wienken et al 2010).

6.2 The GFP-tagged HsSWEET protein is localises to an intracellular membrane

Chapter 3 examined the role of HsSWEET, a novel human sugar transporter from the SWEET family. The SWEET proteins are a relatively recently identified class of sugar transporters. In *A. thaliana,* they are necessary for the movement of sugars between specific tissues (Chen et al 2010). Expression of the human SWEET in yeast led to a reduction in cellular glucose concentration, possibly implying a role in glucose efflux (Chen et al 2010).

Confocal microscopy was used to investigate the subcellular localisation of GFP-tagged HsSWEET, a key piece of information when considering protein function. Fluorescent signal was detected at an intracellular membrane, but not at the plasma membrane. The protein was localised at a punctate and motile organelle, close to the nucleus, which is most likely to be the Golgi. No overlap was detected between HsSWEET and lysosomal, actin or nucleus markers. Additional intracellular organelle markers could be used to identify any colocalisation, to further determine the targeted subcellular organelle.

The Human Protein Atlas project showed high levels of HsSWEET expression at the lower intestine and mammary glands (Uhlen et al 2015). It is possible SWEET is important in the redistribution of sugars, a key role in both tissues. Intestinal cells

could utilise SWEET as a method of moving absorbed sugars to the blood, while in oviduct tissue SWEET could be used in the production of nutrient rich milk. An alternative sugar transporter, GLUT1, has been shown to localise to the Golgi prior to weaning, possibly indicating a role in lactose formation (Nemeth et al 2000). The HsSWEET protein may have a similar role in the preparation of excretory vesicles either for lactation or intestinal absorption. Chen et al (2010) showed a decrease in cytosolic glucose upon upregulation of HsSWEET in yeast, however, they did not investigate protein localisation, so the direction of movement is uncertain.

The functional role of HsSWEET in healthy cells is still undefined, however, it has also been flagged as a biomarker due to its upregulation in breast cancers (Świtnicki et al 2016). The role it plays in disease pathology is also uncertain. As a putative sugar transporter, SWEET is most like to be involved in the increased nutrient requirements of cancer cells. GLUT1 transporter inhibitors are under investigation as novel drugs for the treatment of cancer. The thermophoresis method developed in Chapter 4 could be used to investigate inhibitors with potential as cancer therapeutics targeting HsSWEET.

If HsSWEET targets the Golgi, determining exactly which compartment it localises to could provide useful functional information. In the cis Golgi network (CGN), sugar transporters are required for polysaccharide biosynthesis and glycosylation (Neckelmann and Orellana 1998). Whereas, in the trans Golgi network (TGN) sugar transporters would be required for secretory vesicles preparation (Banfield 2011). Identifying HsSWEET Golgi localisation could be useful in distinguishing between sugar transport for processing of proteins or preparation of secretory vesicles. Antibodies for GM130 and galT can be used to identify the CGN (Nakamura et al 1995) and TGN (Schaub et al 2006) respectively and should be used in future localisation investigations.

As well as cellular localisation, this thesis has established two additional systems which could be used to further characterise HsSWEET. The thermophoresis method can now be used to screen for potential substrates to a GFP-tagged HsSWEET. Meanwhile, the prokaryotic expression system can be used to undertake crystallisation trials and begin the task of structural characterisation. The combination of localisation, substrate binding and structural information will advance the functional characterisation of this unique protein in the human genome.

6.3 GFP-tagged protein can be used to improve the effectiveness of current protein purification systems

This thesis took two approaches to improving the efficiency of protein purification systems: optimising the existing BL21 (DE3) expression systems and investigating ways to adapt BL21 (DE3) to improve expression capabilities. These approaches aim to reduce the impact of the first bottle neck in membrane protein characterisation, expression (Moraes et al 2014). All proteins have different properties, meaning expression conditions may require optimisation to achieve a functional yield (Löw et al 2012). Even with optimisation, recombinant protein expression may not be possible. To navigate this issue, many homologues can be screened until a suitable target protein is found; an effective screen may require almost 100 homologues (Brunner et al 2014). Improving our ability to express membrane proteins could reduce the impact of the first hurdle in the methodology of structural and functional characterisation. With improved characterisation, key proteins in human health can be used in therapeutic development for tackling important issues, such as cancer (Shibuya et al 2015) and antimicrobial resistance (Fiamegos et al 2011). The MFS protein family was selected for investigation, as it is already a relatively well studied group (Guan et al 2007; Quistgaard et al 2013; Fukuda et al 2015; Heng et al 2015; Kumar et al 2015), they are a functionally diverse, can be expressed in the simple *E. coli* expression system, its members include key antimicrobial resistance efflux proteins (Shuster et al 2016) and yet there are still many functionally uncharacterised proteins.

A screen of GFP-tagged MFS transporters was established to enable simple monitoring of expression yields. Through this, different proteins and expression conditions could be directly compared. A portion of these constructs were also used to investigate the physiological effects of protein expression on the BL21 (DE3) cell line. The original focus of this work was aimed at identifying any BL21 (DE3) variants that demonstrated improved over-expression of MFS transporters.

6.3.1 Expression libraries tagged with GFP can be used to inform functional characterisation

In Chapter 4, the H6-msfGFP vector was used in the creation of a protein expression library. Sixty-three members of the MFS family were cloned (Table 4.5), many of which were uncharacterised past their primary structure (Table 4.2). The goal was to produce high yields of protein using the BL21 (DE3) expression system. BL21

(DE3) and its derivatives are the most commonly used *E. coli* recombinant expression system in the world (Jia and Jeon 2016). Due to high production rates, simplicity and low cost.

During untagged protein expression, there are few simple indicators of protein yield. Techniques such as western blotting can be used to monitor protein expression (Psakis et al 2007). However, the additional steps required to take measurements make this method unsuitable for large scale screening. The GFP-tagged MFS transporter makes it a simple method to monitor protein over-expression. A spectrofluorimeter enables quick measurements, a vast improvement to the alternative of Western blotting. The simplicity of the process made it suitable for screening many conditions in parallel. Yields were compared between the different proteins and expression conditions. This identified the optimal conditions for expression of controls used in this report. It also highlighted well-expressed proteins for use in investigatory binding assays, such as novel substrates of mdfA discussed in section 7.2.3. The characterisation techniques required a suitable quantity of sample, this would have been much more time consuming to achieve without the GFP-tag.

The constructed library of MFS transporters provided an efficient method to begin this project. Data generated through expression tests was essential for the success of protein functional studies, providing optimised conditions for protein production.

6.3.2 The T7 expression system severely impacts the BL21 (DE3) cell line

Chapter 5 investigated the adaptations of *E. coli* to the toxicity of protein over-expression using the pET system. *E. coli* is the most popular expression system, with 84.5 % of Protein Data Bank (PDB) structures having used the expression systems BL21 or BL21 (DE3) (Berman et al 2000). However, the production of protein under the control of the T7 promoter can cause lethal stress (Gubellini et al 2011; Wagner et al 2007). Depending on the protein, this toxicity may be too great for the propagation of cells (Dumon-Seignovert et al 2004). This chapter aimed to investigate the toxic effect of recombinant proteins, by expressing GFP and native MFS membrane proteins tagged with GFP, under the control of the T7 promoter.

As alluded to in the General Introduction, the inability to produce a recombinant protein inhibits further study, causing another bottle neck in characterisation

(Moraes et al 2014). In response to this, variant *E. coli* expression systems have been generated, able to resist IPTG-induced toxicity, these include C41 (DE3), C43 (DE3) (Miroux et al 1996) and Mt56 (DE3) (Baumgarten et al 2017). However, there are very few examples of protein characterisation using these generated cell lines. Some researchers recommend screening the BL21 (DE3) expression system for high-yielding isolates prior to an investigation (Massey-Gendel et al 2009). However, the proposed methods are time consuming and are rarely considered in the published literature. Screening a library of homologues for a well-expressed option is a much more frequent approach (Brunner et al 2014). This has limitations, and investigations into generating better expression systems would be valuable for protein biology.

The Sigma-Aldrich guidelines advise induction of protein expression using an IPTG concentration between 0.25 and 2 mM IPTG. IPTG in the absence of a recombinant expression vector had no effect on cell lines. However, the toxic effects of the T7 expression system on BL21 (DE3) was shown to be much greater than anticipated. As shown in Chapter 5, very few colonies can survive exposure to these IPTG concentrations when expressing recombinant protein. The next step would be to determine the number of colony forming units at the various IPTG concentrations. In our experiments we partially standardised the amount added to the plates by using the same cell densities (OD_{600} of 0.1) and a fixed volume (the amount of cells that a loop could contain). With this rough technique at 1 mM IPTG in several instances only two colonies survived. This implies that almost all cells died or were not able to grow at this IPTG concentration. The key to this approach was the fact that colonies producing GFP and those that could not were easy to identify. In other words, GFP fluorescence was a quick way to identify colonies with adaptations to IPTG exposure, while maintaining protein expression. Further growth demonstrated that the few surviving colonies had developed heritable changes, indicating alterations to their DNA.

The T7 expression system requires an interaction between genomic and plasmid DNA. To investigate whether the genomic or plasmid DNA harboured adaptations to IPTG, the plasmid DNA was isolated from the cell lines resistant to IPTG toxicity and re-transformed into non-resistant cell lines. Newly transformed cell lines remained sensitive to IPTG and recombinant extracted plasmids were shown to be highly toxic upon induction. This indicated adaptations were genomic rather than plasmid based. Next generation sequencing will be used to investigate the changes in these mutant strains that protect the cells against IPTG induction toxicity when

using the pET system. It is anticipated that the lacUV5 promoter site of T7 RNA polymerase will be targeted. Identifying which genomic mutations provide increased resistance may help to develop cell lines which are more suitable for expression of membrane proteins.

Genomic sequencing of BL21 (DE3) cells adapted to high IPTG concentrations has so far identified three regions influencing expression induced toxicity: the promoter region for T7 polymerase (Schlegel et al 2015), the promoter recognition site of T7 polymerase (Baumgarten et al 2017), and the lactose binding site of the lac repressor protein (Kwon et al 2015). These mutations either reduce the cells ability to produce T7 polymerase in response to IPTG or reduce the activity of T7 polymerase on the expression vector. Alternative adaptations may sedate T7 induced expression by targeting RNA processing or translation mechanisms. Identifying these alternatives could be used to improve the *E. coli* expression system.

Based on the data presented in this thesis, use of the pET/IPTG system should be more carefully considered during protein production, commonly used concentrations may be causing lethal expression levels. Screening of IPTG concentrations could be utilised during the development of expression protocols. Tuned expressions with lower concentrations of IPTG may be useful in improving the yields of more difficult proteins.

6.4 Thermophoresis with GFP-tagged MFS transporters can now be used to characterise the superfamily

Chapter 4 utilised GFP in the development of a ligand binding assay, aiming to identify substrates for the *E. coli* MFS transporters. MFS transporters have a diverse range of substrates, from carbohydrates, metabolites to antimicrobials (Nishino et al 2001; Kumar et al 2015). They are considered a clinically relevant target, as they are required for the development of antimicrobial resistance in *E. coli* (Shuster et al 2016). The library constructed in this thesis consists of 63 family members, including many proteins with disputed or uncharacterised functions. As in Section 6.1, the GFP tag was used as a reporter to identify highly expressed proteins from this library of expression vectors. Current methods demonstrating protein-substrate interactions may not be suitable for screening and vice versa. A simple assay utilising the thermal motion properties of proteins was implemented to

screen protein substrates. The GFP linked to the C-terminus of the transporter was used as the signal to measure the thermal induced motions.

6.4.1 Thermophoresis assays with the Monolith NT.115 is a useful tool for screening specific protein-substrate interactions

Initial expression tests with the library of MFS transporters established expression conditions for targets of interest; this is described in Section 6.1. Certain proteins gave higher expression yields than others, for example mdtG-sfGFP, mdfA-sfGFP and hsrA-sfGFP. When establishing the thermophoresis assays in Chapter 4, targets were chosen which consisted of well-expressed control proteins with functional information available, and well-expressed but previously uncharacterised proteins. Chapter 4 identified the appropriate expression conditions to use in developing the ligand binding assay.

The control targets, mdfA and mdtM, have published substrate binding affinities, both have been shown to bind to chloramphenicol, while mdfA has additionally been shown to bind to TPP+ (Holdsworth 2013; Fluman et al 2014; Heng et al 2015). This information is required in assay development, as by comparison with measured affinities their accuracy can be verified. The thermophoresis technique established in Chapter 4.2.2.4 calculated substrate binding affinities for both mdfA-sfGFP and mdtM-sfGFP, with values comparable to the published literature. This demonstrates accurate measurements of GFP-tagged MFS-substrate interactions using thermophoresis. This method shows protein-substrate interactions and is suitable for screening.

Following the verification of controls, thermophoresis was used to investigate novel substrates of mdfA-sfGFP and hsrA-sfGFP. MdfA-sfGFP was shown to bind to cAMP. An important secondary messenger, cAMP is responsible for catabolite repression. This may indicate mdfA-sfGFP has a role in gene regulation. HsrA-sfGFP was shown to bind to cysteine. Upregulation of hsrA has been shown to cause accumulation of homocysteine thiolactone, a regulator of bacterial response to stress (Goodrich-Blair and Kolter 2000). The interaction between cysteine and hsrA may have a role in bacterial survival strategies.

Collectively, these findings demonstrate the potential for thermophoresis in screening the MFS family for substrate binding. Current methods available for the characterisation of protein-substrate interactions tend to be either indirect, thus incapable of demonstrating binding, or direct but not suitable for screening

purposes due to costs or labour intensity. For example, the role of a protein can be indirectly investigated by measuring the physiological effects of upregulating or knocking out a gene of interest. Nishino et al (2001) demonstrated upregulation of some MFS transporters provided protection from antimicrobial substances in growth media. One advantage of this method is the ability to screen the developed cell line with multiple substrates. However, further experimentation is required to demonstrate an interaction between the protein and its substrate. Additionally, the manipulation of one gene can have knock on effects, disrupting expression levels of numerous other proteins within the proteome (Strader et al 2011). The observed phenotypic effect, such as decreased susceptibility to a toxin, may therefore be caused by this disruption rather than altering the gene of interest.

Methods that demonstrate the protein-substrate interaction include isothermal calorimetry (ITC; Heng et al 2015), fluorescence quenching (Holdsworth 2013) and substrate radiolabelling (Fluman et al 2014). As well as calculating binding affinities, substrate-protein measurements can be used to investigate the interaction of inhibitors, a useful method for identifying therapeutic compounds. However, these methods are frequently less suitable for screening substrates; for instance, isotopic labelling of a substrate is expensive, rendering large-scale screening impractical.

Contrastingly, thermophoresis of GFP-tagged transporters is shown here to be suitable for screening, while providing accurate results regarding specific substrate binding. The substrates do not require expensive or laborious treatment, meaning many can be screened. Additional investigation of interactions with protein lysate could further simplify the technique, by removing steps in protein preparation and shortening the protocol.

Future work should involve applying this technique to investigate other functionally uncharacterised membrane proteins.

6.4.2 Thermophoresis with the generated MFS library could be used to investigate novel antimicrobials

The thermophoresis assay paired with the library of MFS transporters could potentially investigate novel inhibitors of MFS, a key player in bacterial resistance to antibiotics. Some MFS transporters have been shown to efflux antimicrobials, decreasing bacterial susceptibility (Nishino et al 2001). This efflux is an important mechanism for the bacterial cell to reduce the cellular concentration of toxins,

enabling development of highly efficient drug-resistance adaptations (Shuster et al 2016), examples of these mechanisms are shown in Figure 4.3. Antimicrobial resistance is becoming increasingly prevalent, threatening many aspects of modern medicine (O'Neill et al 2015).

Performing thermophoresis on the expression library of MFS transporters generated in Chapter 4 could potentially be used in the development of therapeutics. Ideal substrates when developing therapeutics would interact with the MFS transport region to inhibit the efflux pathway, without undergoing transportation. Substrates able to inhibit sections of the MFS family members could potentially lead to new drugs for use alongside normal antibiotic treatment. Where MFS-mediated efflux is inhibited, it would be much harder for bacteria to establish resistance pathways (Shuster et al 2016). The ability of efflux pump inhibitors (EPIs) to inhibit resistance mechanisms is currently being explored in Gram positive bacteria such as *S. aureus* (Handzlik et al 2013). Development of EPIs for gram negative bacteria could positively impact the issue of antimicrobial resistance (Mahmood et al 2016). It would be interesting to explore the practicality of this technique with clinically relevant MFS targets.

6.5 mdfA may have a role in *E. coli* catabolite repression

The thermophoresis assay developed in this thesis measured an interaction between mdfA-sfGFP and cAMP, possibly indicating a transported substrate. In *E. coli*, cellular cAMP is key in a process known as catabolite repression, controlling which energy source is preferentially metabolised. When glucose is available, cAMP concentration is kept low, repressing the production of enzymes for alternative sugars. In the absence of glucose, cellular cAMP increases and allosterically activates the catabolite activator Protein (CAP); this is required for transcription of operons for alternate energy sources such as lactose (Kremling et al 2015). If mdfA controls cellular cAMP concentration, it may be regulating expression of other genes, enabling *E. coli* to adapt to different environmental stimuli.

The MFS family, including mdfA, have been shown to impact cellular metabolism. When upregulated, they cause a reduction in intracellular arabinose, which previously has been attributed to sugar efflux (Koita et al 2012). However, this work did not demonstrate protein-arabinose binding. It is therefore possible these MFS transporters impact sugar concentrations by regulating other transporters or enzymes, and not direct arabinose efflux. If mdfA does indeed transport cAMP, it

may be contributing to downregulation of genes for arabinose uptake, or upregulation of genes for arabinose metabolism.

Some of the MFS transporters suggested as arabinose transporters, such as mhpT and mdtD, are located within an operon (Koita et al 2012). If these are in fact cAMP transporters they may help to regulate expression of the operon they are part of. This would ensure the operon could be inactivated when a more favourable catabolite becomes available. To test this theory, the thirteen MFS transporters identified by Koita et al (2012) should be tested for cAMP binding using the thermophoresis assay developed in this thesis.

The novel protein-substrate interaction between mdfA and cAMP highlights the potential for the thermophoresis assay to identify new avenues of scientific investigation, providing information on previously unanswered questions.

6.6 GFP-tagging is a useful tool, but complications make it unsuitable for some techniques

GFP is often described as an inert and effective reporter tag in prokaryotic (Feilmeier et al 2000) and eukaryotic (Kain et al 1995) expression systems. This thesis has shown it to be an effective tool for optimising expression systems and characterising the function of proteins. Using these methods, GFP can be used to help tackle the deficit of characterised membrane proteins.

However, when scrutinised, recombinant GFP expression can interfere with cellular metabolism (Li et al 2013), protein localisation (Palmer and Freeman 2004) and protein function, interrupting the ability to interact with ligands (Skube et al 2011). Additionally, in some cases, GFP tags must be removed to achieve a practical expression yield (Hammon et al 2009).

It is possible that tagging certain MFS transporters with GFP may have negatively impacted their expression. Several MFS transporters have been structurally characterised, including lacY (Abramson et al 2003), xylE (Sun et al 2012), mdfA (Heng et al 2015) and emrD (Yin et al 2006). Comparison of protein purification showed limited yields have been achieved with the expression library created for this report (Fluman et al 2009, Alegre and Law 2015, Dornmair 1988). These issues mirror those faced by Hammon et al (2009), they found to achieve appropriate yields for structural characterisation, proteins had to be expressed without a GFP tag.

For the purposes of this thesis, GFP-tags provided a useful tool, enabling expression screening, cellular localisation and functional investigation. Each of these would have been more difficult in the absence of GFP. This work combats the underrepresentation of characterised membrane protein by investigating individual proteins and improving analysis methods.

6.7 Final conclusions and future directions

This thesis aimed to tackle the underrepresentation of characterised membrane proteins. GFP tags were used to investigate the expression capabilities of recombinant systems and the function of uncharacterised proteins.

GFP tagging was shown to be an effective method for expressed protein yield optimisation, allowing the comparison of many conditions and proteins. This is much more labour efficient than alternative methods enabling larger screens to be undertaken. The tag also enabled the investigation of BL21 (DE3) adaptations to exogenous protein over-expression using the pET system. Difficulties expressing proteins is one of the first hurdles in characterisation, a feature that is particularly pronounced in membrane proteins, impairing characterisation and leading to their underrepresentation. Improved understanding of the mechanism behind recombinant protein-induced toxicity could be used to optimise our current expression systems, improving the ability to characterise proteins. Membrane proteins are related to many facets of disease, including genetic disorders, microbial infections and cancers. An improved understanding of membrane protein function could be translated into clinical applications and the design of new therapeutics. It will be interesting to see if the same toxicity is observed in other *E. coli* over-expression systems that do not use the T7 promoter pET system such as those of the pQE vector range.

Thermophoresis was shown to be an effective method of characterising substrate interactions for MFS transporters. This could be used to establish the function of uncharacterised proteins improving our understanding of cellular physiology. A simple method for screening substrates also has interesting implications for the future development of therapeutic agents. Proteins with a role in diseases such as cancers and antimicrobial resistance can be screened for novel inhibitory substances. This could be used to improve human health by providing new treatments.

This thesis tackles several of the bottlenecks associated with membrane protein characterisation potentially opening new areas of research as these tools are further exploited to reveal not just the functionality of one protein but the biology of whole organisms.

Appendices

Appendix A Cloning HsSWEET WT into pEGFP-C1 for generation of an N-term EGFP construct

CONSTRUCT NAME: hSWEET-c006

General construct will look like

START – GFP – C – HindIII – CC – hSWEET – STOP – EcoRI

PCR Primers :

>Primer 1 **hSWEET-f004** : -- Size : 11 + 21 = 32 nt -- Tm initial/Tm final : 69.1 / 80.2

ctcaagcttccATGGAGGCGGGCGGCTTTCTG (overhang-- HindIII – CC –START- hSWEET)

>Primer 2 **hSWEET-r004**: -- Size : 11 + 22 = 33 nt -- Tm initial/Tm final : 64. / 73.9

cg**gaattctta**GGTTTGCAGGAGCCAGTAGTTC (hSWEET – STOP – EcoRI -overhang)

Using hSWEET-WT-c001 as template DNA produces

NheI **CspAI**

BspOI **PinAI**

AsuNHI **AgeI**

BmtI **BshTI**

| | |

G CAG AGC TGG TTT AGT GAA CCG TCA GAT CCG CTA GCG CTA CCG GTC GCC ACC ATG GTG AGC AAG GGC GAG < 630

Q S W F S E P S D P L A L P V A T M V S K G E

C GTC TCG ACC AAA TCA CTT GGC AGT CTA GGC GAT CGC GAT GGC CAG CGG TGG TAC CAC TCG TTC CCG CTC

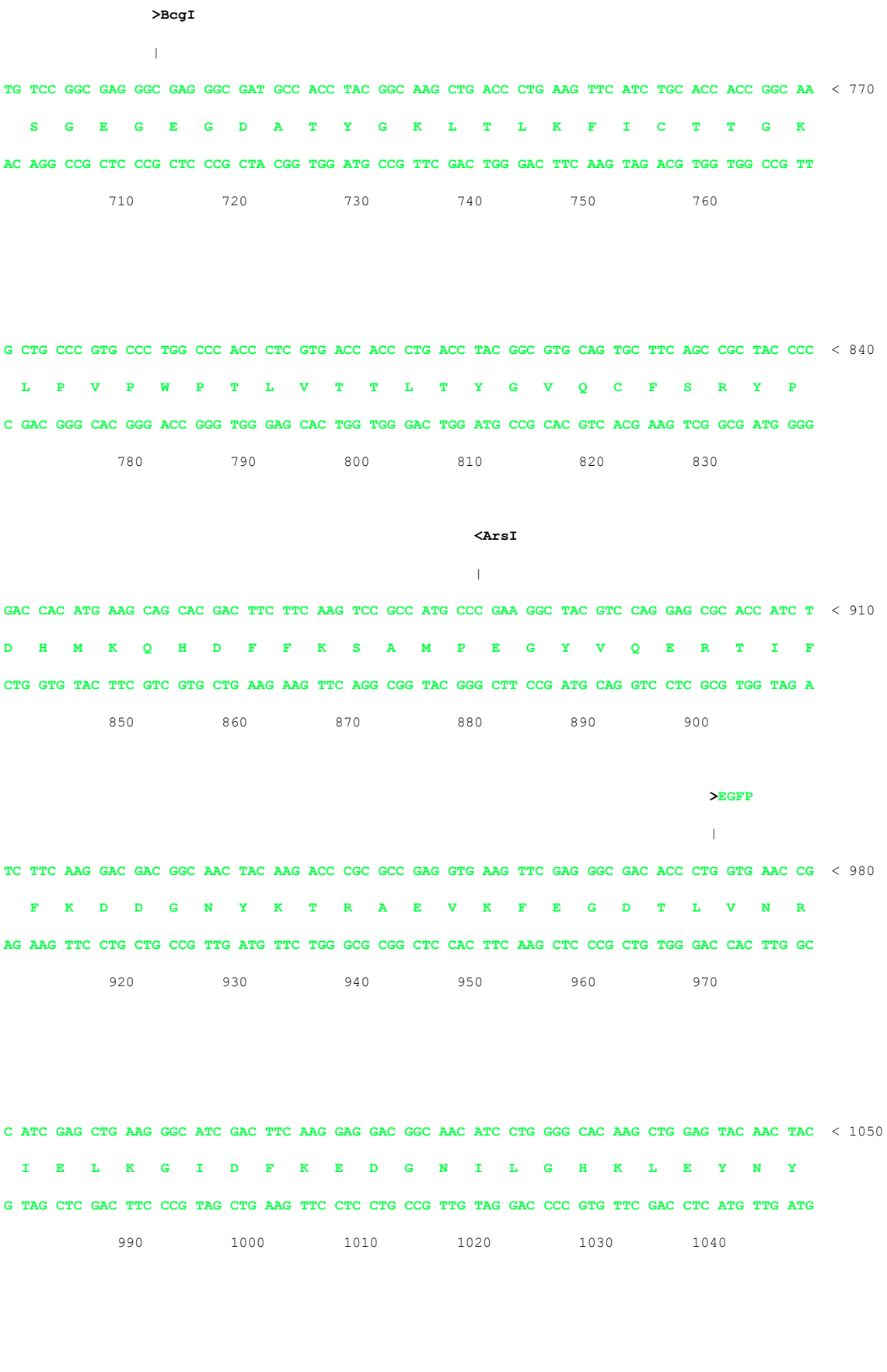
570 580 590 600 610 620

GAG CTG TTC ACC GGG GTG GTG CCC ATC CTG GTC GAG CTG GAC GGC GAC GTA AAC GGC CAC AAG TTC AGC G < 700

E L F T G V V P I L V E L D G D V N G H K F S V

CTC GAC AAG TGG CCC CAC CAC GGG TAG GAC CAG CTC GAC CTG CCG CTG CAT TTG CCG GTG TTC AAG TCG C

640 650 660 670 680 690


>BcgI
TG TCC GGC GAG GGC GAG GGC GAT GCC ACC TAC GGC AAG CTG ACC CTG AAG TTC ATC TGC ACC ACC GGC AA < 770
S G E G E G D A T Y G K L T L K F I C T T G K
AC AGG CCG CTC CCG CTC CCG CTA CGG TGG ATG CCG TTC GAC TGG GAC TTC AAG TAG ACG TGG TGG CCG TT
710 720 730 740 750 760
G CTG CCC GTG CCC TGG CCC ACC CTC GTG ACC ACC CTG ACC TAC GGC GTG CAG TGC TTC AGC CGC TAC CCC < 840
L P V P W P T L V T T L T Y G V Q C F S R Y P
C GAC GGG CAC GGG ACC GGG TGG GAG CAC TGG TGG GAC TGG ATG CCG CAC GTC ACG AAG TCG GCG ATG GGG
780 790 800 810 820 830
<ArsI
GAC CAC ATG AAG CAG CAC GAC TTC TTC AAG TCC GCC ATG CCC GAA GGC TAC GTC CAG GAG CGC ACC ATC T < 910
D H M K Q H D F F K S A M P E G Y V Q E R T I F
CTG GTG TAC TTC GTC GTG CTG AAG AAG TTC AGG CGG TAC GGG CTT CCG ATG CAG GTC CTC GCG TGG TAG A
850 860 870 880 890 900
>EGFP
TC TTC AAG GAC GAC GGC AAC TAC AAG ACC CGC GCC GAG GTG AAG TTC GAG GGC GAC ACC CTG GTG AAC CG < 980
F K D D G N Y K T R A E V K F E G D T L V N R
AG AAG TTC CTG CTG CCG TTG ATG TTC TGG GCG CGG CTC CAC TTC AAG CTC CCG CTG TGG GAC CAC TTG GC
920 930 940 950 960 970
C ATC GAG CTG AAG GGC ATC GAC TTC AAG GAG GAC GGC AAC ATC CTG GGG CAC AAG CTG GAG TAC AAC TAC < 1050
I E L K G I D F K E D G N I L G H K L E Y N Y
G TAG CTC GAC TTC CCG TAG CTG AAG TTC CTC CTG CCG TTG TAG GAC CCC GTG TTC GAC CTC ATG TTG ATG
990 1000 1010 1020 1030 1040
AAC AGC CAC AAC GTC TAT ATC ATG GCC GAC AAG CAG AAG AAC GGC ATC AAG GTG AAC TTC AAG ATC CGC C < 1120
N S H N V Y I M A D K Q K N G I K V N F K I R H
TTG TCG GTG TTG CAG ATA TAG TAC CGG CTG TTC GTC TTC TTG CCG TAG TTC CAC TTG AAG TTC TAG GCG G

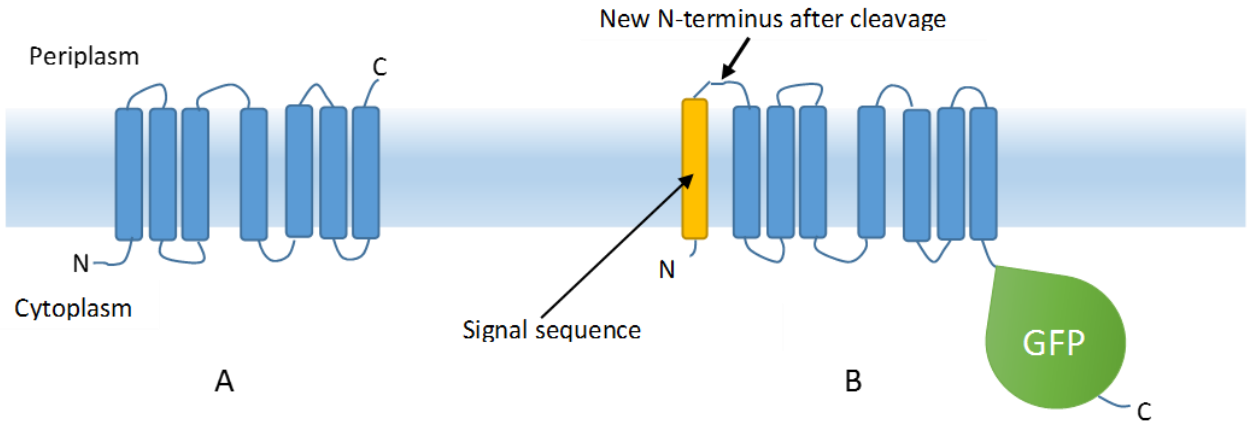
```
            1060         1070         1080         1090         1100         1110

          >RpaB5I
          |
AC AAC ATC GAG GAC GGC AGC GTG CAG CTC GCC GAC CAC TAC CAG CAG AAC ACC CCC ATC GGC GAC GGC CC  < 1190
   N   I   E   D   G   S   V   Q   L   A   D   H   Y   Q   Q   N   T   P   I   G   D   G   P
TG TTG TAG CTC CTG CCG TCG CAC GTC GAG CGG CTG GTG ATG GTC GTC TTG TGG GGG TAG CCG CTG CCG GG
            1130         1140         1150         1160         1170         1180

C GTG CTG CTG CCC GAC AAC CAC TAC CTG AGC ACC CAG TCC GCC CTG AGC AAA GAC CCC AAC GAG AAG CGC  < 1260
  V   L   L   P   D   N   H   Y   L   S   T   Q   S   A   L   S   K   D   P   N   E   K   R
G CAC GAC GAC GGG CTG TTG GTG ATG GAC TCG TGG GTC AGG CGG GAC TCG TTT CTG GGG TTG CTC TTC GCG
            1200         1210         1220         1230         1240         1250

                                                                                  Aor13HI
                                                                                  BspMII
                                                                                  AccIII
                                                                                  Kpn2I
                                                                                  MroI
                                                                      BsrGI       BspEI
                                                                      Bsp1407I    Bsp13I
                                                                      BstAUI      BseAI
                                                                      |           |
GAT CAC ATG GTC CTG CTG GAG TTC GTG ACC GCC GCC GGG ATC ACT CTC GGC ATG GAC GAG CTG TAC AAG T  < 1330
D   H   M   V   L   L   E   F   V   T   A   A   G   I   T   L   G   M   D   E   L   Y   K   S
CTA GTG TAC CAG GAC GAC CTC AAG CAC TGG CGG CGG CCC TAG TGA GAG CCG TAC CTG CTC GAC ATG TTC A
            1270         1280         1290         1300         1310         1320

                    Psp124BI
                    SacI
                    SstI
                    Eco53kI
                    EcoICRI
                    >UcoMSI
                StrI
                Sfr274I
                PaeR7I
                SciI
                SlaI
```

Appendix A

```
                XhoI        HindIII
           BglII    Ecl136II                                          <TstI
           |    |   |       |                                         |
CC GGA CTC AGA TCT CGA GCT CAa gct tcc GAG GCG GGC GGC TTT CTG GAC TCG CTC ATT TAC GGA GCA TG  < 1400
   G   L   R   S   R   A   Q   A   S   E   A   G   G   F   L   D   S   L   I   Y   G   A   C
GG CCT GAG TCT AGA GCT CGA GTt cga agg CTC CGC CCG CCG AAA GAC CTG AGC GAG TAA ATG CCT CGT AC
           1340        1350        1360        1370        1380        1390

C GTG GTC TTC ACC CTT GGC ATG TTC TCC GCC GGC CTC TCG GAC CTC AGG CAC ATG CGA ATG ACC CGG AGT  < 1470
  V   V   F   T   L   G   M   F   S   A   G   L   S   D   L   R   H   M   R   M   T   R   S
G CAC CAG AAG TGG GAA CCG TAC AAG AGG CGG CCG GAG AGC CTG GAG TCC GTG TAC GCT TAC TGG GCC TCA
            1410        1420        1430        1440        1450        1460

GTG GAC AAC GTC CAG TTC CTG CCC TTT CTC ACC ACG GAA GTC AAC AAC CTG GGC TGG CTG AGT TAT GGG G  < 1540
V   D   N   V   Q   F   L   P   F   L   T   T   E   V   N   N   L   G   W   L   S   Y   G   A
CAC CTG TTG CAG GTC AAG GAC GGG AAA GAG TGG TGC CTT CAG TTG TTG GAC CCG ACC GAC TCA ATA CCC C
            1480        1490        1500        1510        1520        1530

            <BsmBI
            <Esp3I
            |
CT TTG AAG GGA GAC GGG ATC CTC ATC GTC GTC AAC ACA GTG GGT GCT GCG CTT CAG ACC CTG TAT ATC TT  < 1610
   L   K   G   D   G   I   L   I   V   V   N   T   V   G   A   A   L   Q   T   L   Y   I   L
GA AAC TTC CCT CTG CCC TAG GAG TAG CAG CAG TTG TGT CAC CCA CGA CGC GAA GTC TGG GAC ATA TAG AA
            1550        1560        1570        1580        1590        1600

                                                            PspPPI
                                                            Psp5II
                                                            PpuMI
                                                            |
G GCA TAT CTG CAT TAC TGC CCT CGG AAG CGT GTT GTG CTC CTA CAG ACT GCA ACC CTG CTA GGG GTC CTT  < 1680
  A   Y   L   H   Y   C   P   R   K   R   V   V   L   L   Q   T   A   T   L   L   G   V   L
C CGT ATA GAC GTA ATG ACG GGA GCC TTC GCA CAA CAC GAG GAT GTC TGA CGT TGG GAC GAT CCC CAG GAA
            1620        1630        1640        1650        1660        1670

            >insert from PCR(hSWEET WT.xdna)
            |
```

```
CTC CTG GGT TAT GGC TAC TTT TGG CTC CTG GTA CCC AAC CCT GAG GCC CGG CTT CAG CAG TTG GGC CTC T  < 1750
L   L   G   Y   G   Y   F   W   L   L   V   P   N   P   E   A   R   L   Q   Q   L   G   L   F
GAG GAC CCA ATA CCG ATG AAA ACC GAG GAC CAT GGG TTG GGA CTC CGG GCC GAA GTC GTC AAC CCG GAG A
            1690        1700        1710        1720        1730        1740

TC TGC AGT GTC TTC ACC ATC AGC ATG TAC CTC TCA CCA CTG GCT GAC TTG GCT AAG GTG ATT CAA ACT AA  < 1820
  C   S   V   F   T   I   S   M   Y   L   S   P   L   A   D   L   A   K   V   I   Q   T   K
AG ACG TCA CAG AAG TGG TAG TCG TAC ATG GAG AGT GGT GAC CGA CTG AAC CGA TTC CAC TAA GTT TGA TT
            1760        1770        1780        1790        1800        1810

A TCA ACC CAA TGT CTC TCC TAC CCA CTC ACC ATT GCT ACC CTT CTC ACC TCT GCC TCC TGG TGC CTC TAT  < 1890
  S   T   Q   C   L   S   Y   P   L   T   I   A   T   L   L   T   S   A   S   W   C   L   Y
T AGT TGG GTT ACA GAG AGG ATG GGT GAG TGG TAA CGA TGG GAA GAG TGG AGA CGG AGG ACC ACG GAG ATA
            1830        1840        1850        1860        1870        1880

GGG TTT CGA CTC AGA GAT CCC TAT ATC ATG GTG TCC AAC TTT CCA GGA ATC GTC ACC AGC TTT ATC CGC T  < 1960
G   F   R   L   R   D   P   Y   I   M   V   S   N   F   P   G   I   V   T   S   F   I   R   F
CCC AAA GCT GAG TCT CTA GGG ATA TAG TAC CAC AGG TTG AAA GGT CCT TAG CAG TGG TCG AAA TAG GCG A
            1900        1910        1920        1930        1940        1950

                                                                    EcoRI
                                                                    |
TC TGG CTT TTC TGG AAG TAC CCC CAG GAG CAA GAC AGG AAC TAC TGG CTC CTG CAA ACC taa gaa ttC TG  < 2030
  W   L   F   W   K   Y   P   Q   E   Q   D   R   N   Y   W   L   L   Q   T   *   E   F   C
AG ACC GAA AAG ACC TTC ATG GGG GTC CTC GTT CTG TCC TTG ATG ACC GAG GAC GTT TGG att ctt aaG AC
            1970        1980        1990        2000        2010        2020
```

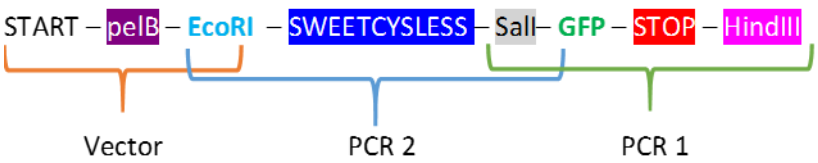
Appendix B Cloning of HsSWEETCYSLESS into pET27b via pGFPSTOP



Topology of the HsSWEETCysless transporter. Due to having an odd number of TM helices, the C-terminus will be located in the periplasmic space (picture A) which does not allow for the proper folding of GFP, if present. To change the orientation, a signal sequence is added to the N-terminus. The knock on effect is that the C-terminus will now be located in the cytoplasmic space (picture B) allowing for the proper folding of the GFP domain.

Version 3 uses EcoRI , HindIII and SalI as none of these cut PQLC1, PQLC2, PQLC3 or SLC50A1.

The construct that will be designed is



Two steps – First make pET27b with GFP-STOP inserted. This can be used for all of the other PQ transporters. This vector will be called pGFPSTOP. Then:

STEP 1 – Cleave pGFPSTOP with EcoRI and SalI

STEP 2 – Digest SWEETCYSLESS PCR fragment with EcoRI and SalI

STEP 3 – Ligate together the main digested fragments from STEPS 1 and 2.

6.8 Cloning details – pET27b conversion to pGFPSTOP

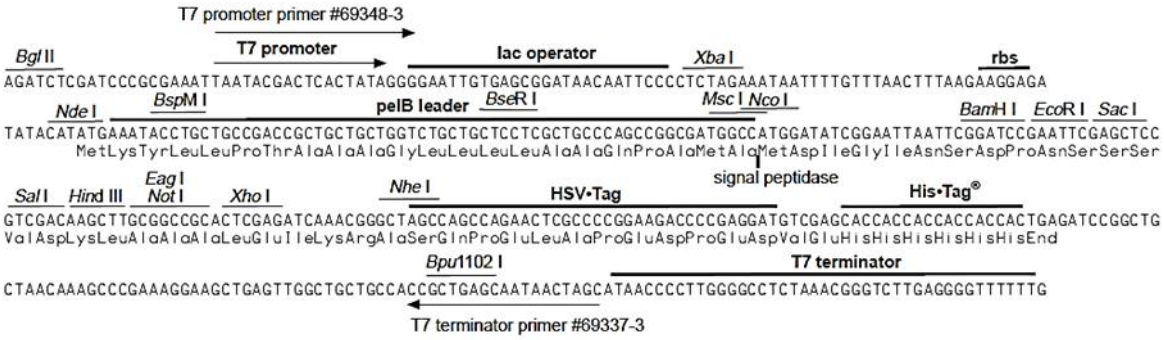


Table 7.1 shows if the group of PQ transporter genes can be cleaved or not with the restriction sites present in the MCS of pET27b.

	BamH I	Eag I	EcoR I	HindII I	Nco I	Nhe I	Not I	Sac I	Sal I	Xho I
hSWTCysLess	X	X	X	X	X	X	X	X	X	X
CTNS	X	X	X	X	X	X	X	X	X	X
SLC50A1	Yes	X	X	X	X	X	X	X	X	X
PQLC1	X	Yes	X	X	Yes	X	Yes	Yes	X	X
PQLC2	X	X	X	X	X	X	X	X	X	Yes
PQLC3	X	X	X	X	X	Yes	X	X	X	X

Appendix B

The three restriction endonucleases that do not have a site present within the genes of the PQ transporters are highlighted in green. Therefore, these can be used for all genes. Each construct will be generated as follows.

EcoRI – target gene – SalI – GFP – STOP – HindIII

which when cloned into pET27b will produce the following protein:

START Met – pelB signal seq – EcoRI a/a's – PQ transporter – SalI a/a's – GFP – STOP

STEP 1 – Cleave pET27b with SalI and HindIII

STEP 2 – Digest PCR fragment GFP – STOP with SalI and HindIII

STEP 3 – Ligate together main fragments from STEPS 1 and 2

6.9 Cloning details for vector pGFPSTOP

GFP template DNA and amino acid translation sequences

ATTGGTAGTGGGAGCAACGGCAGCAGCGGATCCGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGT

CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGA

CCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCG

CACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCA

TCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAAC

GTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGT

GCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA

CCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC

ACTCTCGGCATGGACGAGCTGTACAAGTAG

IGSGSNGSSGSVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC
FSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVY
ITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG
MDELYK*

Blastp search against PDB database

Chain A, Structure Of S65t Y66f R96a Gfp Variant In Precursor State

Sequence ID: pdb|2HFC|ALength: 239Number of Matches: 1

Related Information

Structure-3D structure displays

Range 1: 2 to 239GenPeptGraphics Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Method	Identities	Positives	Gaps
474 bits(1221)	2e-172	Compositional matrix adjust.	228/238(96%)	231/238(97%)	0/238(0%)

```
Query  12   VSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTL  71
            SKGEELFTGVVPILVELDGDVNGHKFSV GEGEGDAT GKLTLKFICTTGKLPVPWPTL
Sbjct  2    ASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL  61

Query  72   VTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLV  131
            VTTLT+GVQCFSRYPDHMKQHDFFKSAMPEGYVQE TISFKDDG YKTRAEVKFEGDTLV
Sbjct  62   VTTLTFGVQCFSRYPDHMKQHDFFKSAMPEGYVQEATISFKDDGNYKTRAEVKFEGDTLV  121

Query  132  NRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLAD  191
            NRIELKGIDFKEDGNILGHKLEYN+NSHNVYITADKQKNGIKANFKIRHN+EDGSVQLAD
Sbjct  122  NRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIEDGSVQLAD  181

Query  192  HYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK  249
            HYQQNTPIGDGPVLLPDNHYLSTQS LSKDPNEKRDHMVLLEFVTAAGIT GMDELYK
Sbjct  182  HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK  239
```

6.9.1 Primers for GFP-STOP

```
>Primer 1 : SalI-GFP-f001 -- Size : 10 + 21 = 31 nt -- Tm initial/Tm final : 65.2
/ 78.9

ctccgtcgacAGCAAGGGCGAGGAGCTGTTC
```

Appendix B

>Primer 2 : GFP-HindIII-r001 -- Size : 11 + 19 = 30 nt -- Tm initial/Tm final : 57.3 / 73.4

tccgcaagcttCTACTTGTACAGCTCGTCC

Using these primers with H6msfGFP as template produces the following PCR fragment:

ctccgtcgacAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCA
TCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTT
CAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAG
CGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
CTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGAACTTCAAGATCCGC
CACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCG
TGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCA
CATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAGaagctt
gcgga

which translates to:

VDSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI
KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*

Highlighted in gray are the amino acids produced due to the SalI site, in green is the GFP domains and in red is the stop codon. In the above DNA sequence, the HindIII site is highlighted in pink.

6.10 Cloning details – hSWEETCysLess (hSWTCysLess) into pGFPSTOP

DNA sequence and translation

ATGGAAGCTGGTGGTTTCCTGGACTCTCTGATCTACGGTGCTGTTGTTGTTTTCACCCT
GGGTATGTTCTCTGCTGGTCTGTCTGACCTGCGTCACATGCGTATGACCCGTTCTGTT
GACAACGTTCAGTTCCTGCCGTTCCTGACCACCGAAGTTAACAACCTGGGTTGGCTGT
CTTACGGTGCTCTGAAAGGTGACGGTATCCTGATCGTTGTTAACACCGTTGGTGCTGC
TCTGCAGACCCTGTACATCCTGGCTTACCTGCACTACTCTCCGCGTAAACGTGTTGTT

CTGCTGCAGACCGCTACCCTGCTGGGTGTTCTGCTGCTGGGTTACGGTTACTTCTGGC
TGCTGGTTCCGAACCCGGAAGCTCGTCTGCAGCAGCTGGGTCTGTTCGTTTCTGTTTT
CACCATCTCTATGTACCTGTCTCCGCTGGCTGACCTGGCTAAAGTTATCCAGACCAAA
TCTACCCAGTCTCTGTCTTACCCGCTGACCATCGCTACCCTGCTGACCTCTGCTTCTT
GGGTTCTGTACGGTTTCCGTCTGCGTGACCCGTACATCATGGTTTCTAACTTCCCGGG
CATAGTTACCAGCTTCATCCGTTTCTGGCTGTTCTGGAAATACCCGCAGGAACAGGAC
CGTAACTACTGGCTGCTGCAGACC

MEAGGFLDSLIYGAVVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSYGALKGDGILIVVNTVGAALQT
LYILAYLHYSPRKRVVLLQTATLLGVLLLGYGYFWLLVPNPEARLQQLGLFVSVFTISMYLSPLADLAKVIQTKSTQSLSY
PLTIATLLTSASWVLYGFRLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT

Blastp search against human genome

sugar transporter SWEET1 isoform a [Homo sapiens]

Sequence ID: ref|NP_061333.2|Length: 221Number of Matches: 1

6.10.1.1 **See 3 more title(s)**

Related Information

Gene-associated gene details

Map Viewer-aligned genomic context

Identical Proteins-Identical proteins to NP_061333.2

Range 1: 1 to 221GenPeptGraphics Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Method	Identities	Positives	Gaps
437 bits(1124)	1e-157	Compositional matrix adjust.	216/221(98%)	216/221(97%)	0/221(0%)

```
Query  1    MEAGGFLDSLIYGAVVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSY  60
            MEAGGFLDSLIYGA VVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSY
Sbjct  1    MEAGGFLDSLIYGACVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSY  60

Query  61   GALKGDGILIVVNTVGAALQTLYILAYLHYSPRKRVVLLQTATLLGVLLLGYGYFWLLVP  120
            GALKGDGILIVVNTVGAALQTLYILAYLHY PRKRVVLLQTATLLGVLLLGYGYFWLLVP
Sbjct  61   GALKGDGILIVVNTVGAALQTLYILAYLHYCPRKRVVLLQTATLLGVLLLGYGYFWLLVP  120

Query  121  NPEARLQQLGLFVSVFTISMYLSPLADLAKVIQTKSTQSLSYPLTIATLLTSASWVLYGF  180
            NPEARLQQLGLF SVFTISMYLSPLADLAKVIQTKSTQ LSYPLTIATLLTSASW LYGF
Sbjct  121  NPEARLQQLGLFCSVFTISMYLSPLADLAKVIQTKSTQCLSYPLTIATLLTSASWCLYGF  180

Query  181  RLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT  221
            RLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT
```

Sbjct 181 RLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQT 221

6.10.2 Primers for hSWTCysLess

Forward primer has an additional C base added after the EcoRI site to get it back into frame.

>Primer 1 : EcoRI -hSWTCysLess-f001 -- Size : 9 + 21 = 30 nt -- Tm initial/Tm final : 63.3 / 74.8

cggaattccGAAGCTGGTGGTTTCCTGGAC

>Primer 2 : hSWTCysLess-SalI-r001 -- Size : 10 + 19 = 29 nt -- Tm initial/Tm final : 61.6 / 76.1

gcttgtcgacGGTCTGCAGCAGCCAGTAG

Produces

cggaattccGAAGCTGGTGGTTTCCTGGACTCTCTGATCTACGGTGCTGTTGTTGTTTTCACCCTGGGTATGTTCTCTGC

TGGTCTGTCTGACCTGCGTCACATGCGTATGACCCGTTCTGTTGACAACGTTCAGTTCCTGCCGTTCCTGACCACCGAAG

TTAACAACCTGGGTTGGCTGTCTTACGGTGCTCTGAAAGGTGACGGTATCCTGATCGTTGTTAACACCGTTGGTGCTGCT

CTGCAGACCCTGTACATCCTGGCTTACCTGCACTACTCTCCGCGTAAACGTGTTGTTCTGCTGCAGACCGCTACCCTGCT

GGGTGTTCTGCTGCTGGGTTACGGTTACTTCTGGCTGCTGGTTCCGAACCCGGAAGCTCGTCTGCAGCAGCTGGGTCTGT

TCGTTTCTGTTTTCACCATCTCTATGTACCTGTCTCCGCTGGCTGACCTGGCTAAAGTTATCCAGACCAAATCTACCCAG

TCTCTGTCTTACCCGCTGACCATCGCTACCCTGCTGACCTCTGCTTCTTGGGTTCTGTACGGTTTCCGTCTGCGTGACCC

GTACATCATGGTTTCTAACTTCCCGGGCATAGTTACCAGCTTCATCCGTTTCTGGCTGTTCTGGAAATACCCGCAGGAAC

AGGACCGTAACTACTGGCTGCTGCAGACCgtcgacaagc

This translates to

RNSEAGGFLDSLIYGAVVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLTTEVNNLGWLSYGALKGDGILIVVNTVGAAL

QTLYILAYLHYSPRKRVVLLQTATLLGVLLLGYGYFWLLVPNPEARLQQLGLFVSVFTISMYLSPLADLAKVIQTKSTQSL

SYPLTIATLLTSASWVLYGFRLRDPYIMVSNFPGIVTSFIRFWLFWKYPQEQDRNYWLLQTVDKX

Highlighted in blue is SWEETCysLess and in gray are the additional amino acids from the SalI site.

6.11 Final Construct

Ligation of the following double digested samples, pGFPSTOP (EcoRI and SalI) and hsSWEETCysless (EcoRI and SalI), produced the following construct:

AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTT

TGTTTAACTTTAAGAAGGAGATATACATATGAAATACCTGCTGCCGACCGCTGCTGCTGGTCTGCTGCTCCTCGCTGCCC

AGCCGGCGATGGCCATGGATATCGGAATTAATTCGGATCCGaattccGAAGCTGGTGGTTTCCTGGACTCTCTGATCTAC

GGTGCTGTTGTTGTTTTCACCCTGGGTATGTTCTCTGCTGGTCTGTCTGACCTGCGTCACATGCGTATGACCCGTTCTGT

TGACAACGTTCAGTTCCTGCCGTTCCTGACCACCGAAGTTAACAACCTGGGTTGGCTGTCTTACGGTGCTCTGAAAGGTG

ACGGTATCCTGATCGTTGTTAACACCGTTGGTGCTGCTCTGCAGACCCTGTACATCCTGGCTTACCTGCACTACTCTCCG

CGTAAACGTGTTGTTCTGCTGCAGACCGCTACCCTGCTGGGTGTTCTGCTGCTGGGTTACGGTTACTTCTGGCTGCTGGT

TCCGAACCCGGAAGCTCGTCTGCAGCAGCTGGGTCTGTTCGTTTCTGTTTTCACCATCTCTATGTACCTGTCTCCGCTGG

CTGACCTGGCTAAAGTTATCCAGACCAAATCTACCCAGTCTCTGTCTTACCCGCTGACCATCGCTACCCTGCTGACCTCT

GCTTCTTGGGTTCTGTACGGTTTCCGTCTGCGTGACCCGTACATCATGGTTTCTAACTTCCCGGGCATAGTTACCAGCTT

CATCCGTTTCTGGCTGTTCTGGAAATACCCGCAGGAACAGGACCGTAACTACTGGCTGCTGCAGACCgtcgaCAGCAAGG

GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGCGC

GGCGAGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG

GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCT

TCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTCCTTCAAGGACGACGGCACCTACAAGACCCGCGCC

GAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT

GGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACGTCTATATCACGGCCGACAAGCAGAAGAACGGCATCAAGGCGA

ACTTCAAGATCCGCCACAACGTCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGAC

GGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATCA

CATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAGAAGCTTGCGGCCG

CACTCGAGATCAAACGGGCTAGCCAGCCAGAACTCGCCCCGGAAGACCCCGAGGATGTCGAGCACCACCACCACCACCAC

TGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCT

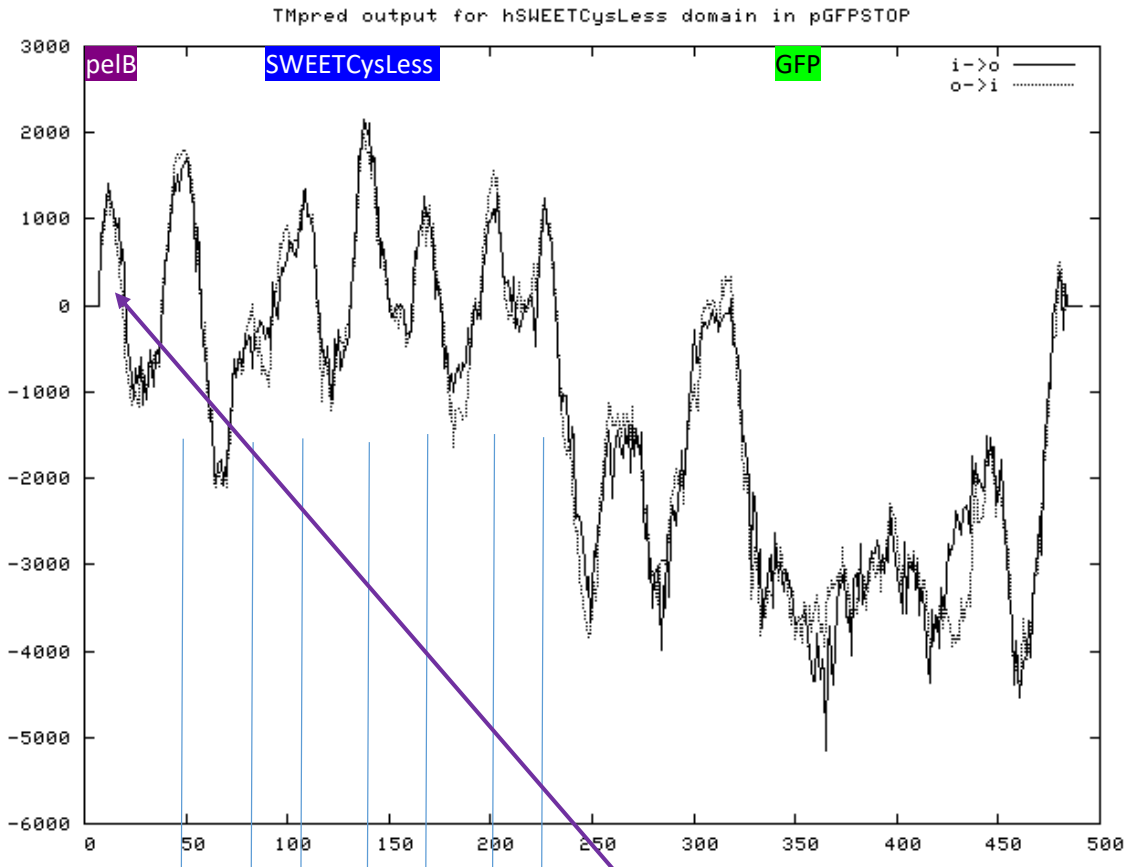
TGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG

which translates to

MKYLLPTAAAGLLLLAAQPAMAMDIGINSDPNSEAGGFLDSLIYGAVVVFTLGMFSAGLSDLRHMRMTRSVDNVQFLPFLT
TEVNNLGWLSYGALKGDGILIVVNTVGAALQTLYILAYLHYSPRKRVVLLQTATLLGVLLLGYGYFWLLVPNPEARLQQLG
LFVSVFTISMYLSPLADLAKVIQTKSTQSLSYPLTIATLLTSASWVLYGFRLRDPYIMVSNFPGIVTSFIRFWLFWKYPQE
QDRNYWLLQTVDSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC
FSRYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVY
ITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLG
MDELYK*

pelB signal sequence - linker - EcoRI - SWEETCysLess - SalI - GFP - STOP - HindIII

The presence of the signal sequence will place the linker on the periplasmic side of the membrane, with the remaining 7 TM helices resulting in the C-terminal GFP being located in the cytoplasmic side.



This composite structure is a mixture of the signal sequence pelB and the SWEETCysLess protein which by itself looks like this

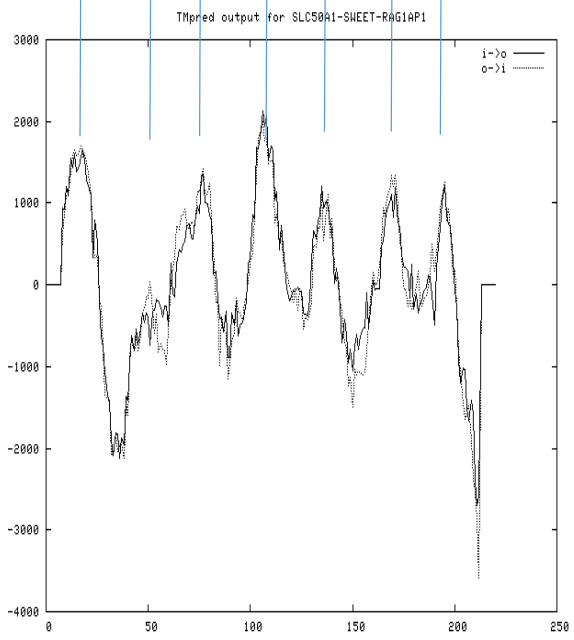


Table 7.2 Primer sequences used to amplify *E. coli* MFS genes for expression vector construction

The forward sequence TACTTCCAATCCAATGCA and the reverse sequence CTCCCACTACCAATGCC, highlighted in bold, correspond to the ligation independent cloning site. All other sequences are complementary to the gene of interest. The primer names correspond to their gene targets in *E. coli* K-12 MG1655 genomic DNA. All sequences are listed 5´ to 3´. Custom DNA oligos purchased from Invitrogen, UK.

Primer name	Sequence
ampG-f001	**TACTTCCAATCCAATGCA**TCCAGTCAATATTTACGTATT
ampG-r001	**CTCCCACTACCAATGCC**CAGATGCGTTTTTCGTAGCGC
amtB-f001	**TACTTCCAATCCAATGCA**AAGATAGCGACGATAAAAACT
amtB-r001	**CTCCCACTACCAATGCC**CGCGTTATAGGCATTCTCGCC
araE-f001	**TACTTCCAATCCAATGCA**ACGCCACGTTCTTTGCGGGAT
araE-r001	**CTCCCACTACCAATGCC**GACGCCGATATTTCTCAACTT
araJ-f001	**TACTTCCAATCCAATGCA**AAAAAAGTCATTTTATCTTTG
araJ-r001	**CTCCCACTACCAATGCC**CTGGCGCTTATAGCGACCATA
bcr-f001	**TACTTCCAATCCAATGCA**TCGTCGTTTGCTATTGTTTTT
bcr-r001	**CTCCCACTACCAATGCC**CCGTTTTTTCGGCCGACTGGC
citT-f001	**TACTTCCAATCCAATGCA**TCTTTAGCAAAAGATAATATA
citT-r001	**CTCCCACTACCAATGCC**GTTCCACATGGCGAGAATCGG
cynX-f001	**TACTTCCAATCCAATGCA**CTGCTGGTACTGGTGCTGATT
cynX-r001	**CTCCCACTACCAATGCC**CTGCGGAAAACGTACTGGTGC
dctA-f001	**TACTTCCAATCCAATGCA**AAAACCTCTCTGTTTAAAAGC
dctA-r001	**CTCCCACTACCAATGCC**CGCACGATTATTCAGCACATC
dcuA-f001	**TACTTCCAATCCAATGCA**CTAGTTGTAGAACTCATCATA
dcuA-r001	**CTCCCACTACCAATGCC**CAGCATGAAGCTACCCAGCAC
dcuB-f001	**TACTTCCAATCCAATGCA**TTATTTACTATCCAACTTATC
dcuB-r001	**CTCCCACTACCAATGCC**TAAGAACCCGTACATCGCGGC
dcuC-f001	**TACTTCCAATCCAATGCA**CTGACATTCATTGAGCTCCTT
dcuC-r001	**CTCCCACTACCAATGCC**CTTGCCTGTGACCGCTGCTGC
dcuS-f001	**TACTTCCAATCCAATGCA**AGACATTCATTGCCCTACCGC
dcuS-r001	**CTCCCACTACCAATGCC**TCTGTTCGACCTCTCCCCGTC

Primer name	Sequence
dgoT-f001	**TACTTCCAATCCAATGCA**GCAAAGCCGGGGCGTCGGCGT
dgoT-r001	**CTCCCACTACCAATGCC**GCCAACGCGCTTCACATCGCC
dtpD-f001	**TACTTCCAATCCAATGCA**TCACAGCCGCGCGCTATTTAC
dtpD-r001	**CTCCCACTACCAATGCC**AGACTCCAGCGCCAGCGCGCG
eamB-f001	**TACTTCCAATCCAATGCA**ACACCGACCCTTTTAAGTGCT
eamB-r001	**CTCCCACTACCAATGCC**ATAGAAAATGCGTACCGCGCA
emrB-f001	**TACTTCCAATCCAATGCA**CAAAAACCGCTGGAAGGCGCG
emrB-r001	**CTCCCACTACCAATGCC**TTTAGCAAACCACACCAGCCC
emrD-f001	**TACTTCCAATCCAATGCA**AAAAGGCAAAGAAACGTCAAT
emrD-r001	**CTCCCACTACCAATGCC**CGACATCCGCGTCGCCAGCGG
emrY-f001	**TACTTCCAATCCAATGCA**GGGACGTTATGGTGCGTCACT
emrY-r001	**CTCCCACTACCAATGCC**TTTCGCAAACCAAACCAAAAC
exuT-f001	**TACTTCCAATCCAATGCA**CGGTTTTTTTTCGGTTACCCG
exut-r001	**CTCCCACTACCAATGCC**CGGTTTGTTCTGCAACACGGT
fsr-f001	**TACTTCCAATCCAATGCA**ACCAAGGCCCGAACATCGTTT
fsr-r001	**CTCCCACTACCAATGCC**TTTATGCCGGTTATCAGGCAG
fucP-f001	**TACTTCCAATCCAATGCA**CAGAGTTACCGTGCGGTAGAT
fucP-r001	**CTCCCACTACCAATGCC**GTTAGTTGCCGTTTGAGAACG
galP-f001	**TACTTCCAATCCAATGCA**AAACAGGGGCGGTCAAACAAG
galP-r001	**CTCCCACTACCAATGCC**TATTTCGCGCAGTTTACGACC
garP-f001	**TACTTCCAATCCAATGCA**GTTGACGAAAAAAAGAAAGGC
garP-r001	**CTCCCACTACCAATGCC**TTTCTGCAATTCCATACGTTT
gltP-f001	**TACTTCCAATCCAATGCA**AAAAATATAAAATTCAGCCTG
gltP-r001	**CTCCCACTACCAATGCC**TTGATCCGCAGTTTTATCAAA
gntT-f001	**TACTTCCAATCCAATGCA**CCATTAGTCATTGTTGCTATC
gntT-r001	**CTCCCACTACCAATGCC**AATCACCATATTCAGCAGCAG
gudP-f001	**TACTTCCAATCCAATGCA**AGTTCTTTAAGTCAGGCTGCG
gudP-r001	**CTCCCACTACCAATGCC**TTTCAACTCGATACGCTTGAT
hcaT-f001	**TACTTCCAATCCAATGCA**GTTTTGCAATCCACGCGCTGG
hcaT-r001	**CTCCCACTACCAATGCC**AACTTTCGGGCGCAAAAACAT
hsrA-f001	**TACTTCCAATCCAATGCA**AGCGATAAAAAGAAGCGCAGT

Appendix B

Primer name	Sequence
hsrA-r001	**CTCCCACTACCAATGCC**CTCCGATTCTGATGGAACGCG
idnT-f001	**TACTTCCAATCCAATGCA**CCATTAATCATTATTGCGGCA
idnT-r001	**CTCCCACTACCAATGCC**GTGCAATACGGCGTTAATGGC
kgtP-f001	**TACTTCCAATCCAATGCA**AGTGATACTCGTCGCCGCATT
kgtP-r001	**CTCCCACTACCAATGCC**AAGACGCATCCCCTTCCCTTT
lacY-f001	**TACTTCCAATCCAATGCA**TACTATTTAAAAAACACAAAC
lacY-r001	**CTCCCACTACCAATGCC**AGCGACTTCATTCACCTGACG
lgt-f001	**TACTTCCAATCCAATGCA**CCGGAGTTTGATCCGGTCATT
lgt-r001	**CTCCCACTACCAATGCC**GGAAACGTGTTGCTGTGGGCT
lldP-f001	**TACTTCCAATCCAATGCA**GGGAATATCTGGCTTTCCAGT
lldP-r001	**CTCCCACTACCAATGCC**AGGAATCATCCACGTTAAGAC
lplT-f001	**TACTTCCAATCCAATGCA**GTGCACACTAACACTTCGTTG
lplT-r001	**CTCCCACTACCAATGCC**ATGACGGCGCTGCCAGATCCA
mdfA-f001	**TACTTCCAATCCAATGCA**CAAAATAAATTAGCTTCCGGT
mdfA-r001	**CTCCCACTACCAATGCC**CATCTGTTTATCTTTTAAAAA
mdoB-f001	**TACTTCCAATCCAATGCA**TCAGAACTACTCTCTTTCGCC
mdoB-r001	**CTCCCACTACCAATGCC**CCCTTCACGTTCTACCACTTT
mdtD-f001	**TACTTCCAATCCAATGCA**AGCACCCGTTGGCAATTGTGG
mdtD-r001	**CTCCCACTACCAATGCC**TTGCGCGCTCCTTTTTCGCCG
mdtG-f001	**TACTTCCAATCCAATGCA**GACACCCCTATAAACTGGAAA
mdtG-r001	**CTCCCACTACCAATGCC**GGGTATTCGACGACGACGTAG
mdtH-f001	**TACTTCCAATCCAATGCA**TCCCGCGTGTCGCAGGCGAGG
mdtH-r001	**CTCCCACTACCAATGCC**GGCGTCGCGTTCAAGCAAACG
mdtL-f001	**TACTTCCAATCCAATGCA**TCCCGCTTTTTGATTTGTAGT
mdtL-r001	**CTCCCACTACCAATGCC**CGCGACGAACATAATCAGCAA

Primer name	Sequence
mdtM-f001	**TACTTCCAATCCAATGCA**CGTTTTTTTACCCGCCATGCC
mdtM-r001	**CTCCCACTACCAATGCC**CTGCTCCTCCACTAGCTCGGC
melB-f001	**TACTTCCAATCCAATGCA**AGCATTTCAATGACTACAAAA
melB-r001	**CTCCCACTACCAATGCC**TACTTTGCGATATTTATCCAG
mhtP-f001	**TACTTCCAATCCAATGCA**TCATCCCGCCTGATGCTGACC
mhtP-r001	**CTCCCACTACCAATGCC**TCGTGATCTCCGGCTCATCAA
mntH-f001	**TACTTCCAATCCAATGCA**CGCGTTGAGAGTAGCAGCGGA
mntH-r001	**CTCCCACTACCAATGCC**CAATCCCAGCGCCGTCCCCAC
nanT-f001	**TACTTCCAATCCAATGCA**ACCCAGAATATCCCGTGGTAT
nanT-r001	**CTCCCACTACCAATGCC**GTCGATAGCGTCATGAGTACG
narK-f001	**TACTTCCAATCCAATGCA**ACTGGAGCTGTCATTACAGAT
narK-r001	**CTCCCACTACCAATGCC**TTTTTTAGAATGCCGACCATA
narU-f001	**TACTTCCAATCCAATGCA**AAAAATAGTCGTTATCTTTTG
narU-r001	**CTCCCACTACCAATGCC**TTTTTGGCTGAACTTCCGCCG
nepI-f001	**TACTTCCAATCCAATGCA**AGTGAATTTATTGCCGAAAAC
nepI-r001	**CTCCCACTACCAATGCC**GGATTTCTTCATTTTCACCTT
nupG-f001	**TACTTCCAATCCAATGCA**AATCTTAAGCTGCAGCTGAAA
nupG-r001	**CTCCCACTACCAATGCC**GTGGCTAACCGTCTGTGTGCC
pgpA-f001	**TACTTCCAATCCAATGCA**ACCATTTTGCCACGCCATAAA
pgpA-r001	**CTCCCACTACCAATGCC**CGACAGAATACCCAGCGGCCA
pgpB-f001	**TACTTCCAATCCAATGCA**CGTTCGATTGCCAGACGTACC
pgpB-r001	**CTCCCACTACCAATGCC**ACTTTCTTGTTCTCGTTGCGC
pgsA-f001	**TACTTCCAATCCAATGCA**CAATTTAATATCCCTACGTTG
pgsA-r001	**CTCCCACTACCAATGCC**CTGATCAAGCAAATCTGCACG
proP-f001	**TACTTCCAATCCAATGCA**CTGAAAAGGAAAAAAGTAAAA
proP-r001	**CTCCCACTACCAATGCC**TTCATCAATTCGCGGATGTTG
rcnA-f001	**TACTTCCAATCCAATGCA**ACCGAATTTACAACTCTTCTT
rcnA-r001	**CTCCCACTACCAATGCC**TCGCATTATGCCCATGAAGCC
rhtB-f001	**TACTTCCAATCCAATGCA**ACCTTAGAATGGTGGTTTGCC
rhtB-r001	**CTCCCACTACCAATGCC**CGCATGCCTCGCCGATGCTAA

Primer name	Sequence
rhtC-f001	**TACTTCCAATCCAATGCA**TTGATGTTATTTCTCACCGTC
rhtC-r001	**CTCCCACTACCAATGCC**CCGCGAAATAATCAAATGAAT
setA-f001	**TACTTCCAATCCAATGCA**ATCTGGATAATGACGATGGCT
setA-r001	**CTCCCACTACCAATGCC**AACGTCTTTAACCTTTGCGGT
setB-f001	**TACTTCCAATCCAATGCA**AGCGCGAAATCGTTTGACCTG
setB-r001	**CTCCCACTACCAATGCC**AACATCTTTAATCCGCAGTAA
setC-f001	**TACTTCCAATCCAATGCA**ACTCCATCAAAAATACTTGAT
setC-f001	**CTCCCACTACCAATGCC**AATATCTTTAATAAACAGCAG
shiA-f001	**TACTTCCAATCCAATGCA**GACTCCACGCTCATCTCCACT
shiA-r001	**CTCCCACTACCAATGCC**AGCGCGTTGACTGTCTTTCAT
sstT-f001	**TACTTCCAATCCAATGCA**CCGGGGCTATTCCGGCGTCTG
sstT-r001	**CTCCCACTACCAATGCC**ATTACGCAGGGCGCTATTTGC
tgsA-f001	**TACTTCCAATCCAATGCA**AATCGCATCAAGCTCACATGG
tgsA-r001	**CTCCCACTACCAATGCC**ATGAGAGGTCAGGGTGTTATG
tppB-f001	**TACTTCCAATCCAATGCA**GAAAGCGTCAGTTTGAACGCT
tppB-r001	**CTCCCACTACCAATGCC**CGCTACGGCTGCTTTCGCCGC
tqsA-r001	**CTCCCACTACCAATGCC**CTCTTTATTGAGATCGCTTAA
tsqA-f001	**TACTTCCAATCCAATGCA**GCAAAGCCGATCATCACGCTC
ttdT-f001	**TACTTCCAATCCAATGCA**AAACCTTCCACTGAATGGTGG
ttdT-r001	**CTCCCACTACCAATGCC**AAGCAACACCACGGGCATCCA
uhpC-f001	**TACTTCCAATCCAATGCA**ACTGATAAATATGAAATTGAT
uhpC-r001	**CTCCCACTACCAATGCC**CGCTTCGCGCGGTGTCTGGGC
uhpT-f001	**TACTTCCAATCCAATGCA**CTGGCTTTCTTAAACCAGGTT
uhpT-r001	**CTCCCACTACCAATGCC**TGCCACTGTCAACTGCTGAAT
uidB-f001	**TACTTCCAATCCAATGCA**AATCAACAACTCTCCTGGCGC
uidB-r001	**CTCCCACTACCAATGCC**ATTAGTGATATCGCTGATTAA
xapB-f001	**TACTTCCAATCCAATGCA**AGCATCGCGATGCGCTTAAAG
xapB-r001	**CTCCCACTACCAATGCC**ATGAGTCACCGCTCGATGCTT
xylE-f001	**TACTTCCAATCCAATGCA**AGTTATATATTTTCGATTACC
xylE-r001	**CTCCCACTACCAATGCC**TGTTTTCTTCGTTTCCGGTTC
xylH-f001	**TACTTCCAATCCAATGCA**GGCTTCTCCGGGCTGAAATCA
xylH-r001	**CTCCCACTACCAATGCC**AGAACGGCGTTTGGTTGCGGA
yaaU-f001	**TACTTCCAATCCAATGCA**TCTATTCACCGCCGCATTTTG
yaaU-r001	**CTCCCACTACCAATGCC**CGGGGCAAACGCTACGGAAAT

Primer name	Sequence
yagG-f001	**TACTTCCAATCCAATGCA**ACGCAATTAACCATGAAAGAC
yagG-r001	**CTCCCACTACCAATGCC**ATGGGATGCGGCTGTCGCGGC
yahN-f001	**TACTTCCAATCCAATGCA**ATGCAGTTAGTTCACTTATTT
yahN-r001	**CTCCCACTACCAATGCC**CCGCTGCGTCACCCCTTCGTA
yajR-f001	**TACTTCCAATCCAATGCA**ACGCCAGGTGAGAGGCGCGCG
yajR-r001	**CTCCCACTACCAATGCC**TGCCTGACGAATTGCCTGTTC
ybaL-f001	**TACTTCCAATCCAATGCA**CATCACGCCACCCCGCTTATC
ybaL-r001	**CTCCCACTACCAATGCC**TGGCGTTTCCAGCAGTTCCAG
ybaL-r002	**CTCCCACTACCAATGCC**TGGGATCTGCTTCTCTTCTTC
ybaT-c001	**TACTTCCAATCCAATGCA**AACAAACCTCTCGGTCTATGG
ybaT-r001	**CTCCCACTACCAATGCC**TACGGTTTTATTGCGCTTCAT
ybbW-f001	**TACTTCCAATCCAATGCA**GAACATCAGAGAAAACTATTC
ybbW-r001	**CTCCCACTACCAATGCC**TGTACGTTTCTTTAATAAGGC
ybdA-f001	**TACTTCCAATCCAATGCA**AATAAACAATCCTGGCTGCTT
ybdA-r001	**CTCCCACTACCAATGCC**CGTCTGGCGAAAATGTCGCAA
ybgN-f001	**TACTTCCAATCCAATGCA**TCCACAATTACATTGTTATGC
ybgN-r001	**CTCCCACTACCAATGCC**AATTACCGCCCATACGCACCA
ybiR-f001	**TACTTCCAATCCAATGCA**AGCCTGCCTTTTTTACGCACG
ybiR-r001	**CTCCCACTACCAATGCC**GTTGGCCGGGAGTATAACTAA
ybjJ-f001	**TACTTCCAATCCAATGCA**ACCGTAAATTCTTCACGTAAT
ybjJ-r001	**CTCCCACTACCAATGCC**CGTTTTGGTATCGGGTTTGGC
ycaD-f001	**TACTTCCAATCCAATGCA**TCCACGTATACCCAGCCTGTC
ycaD-r001	**CTCCCACTACCAATGCC**CACGTGAGCAACGGGTTTCGG
ydcO-f001	**TACTTCCAATCCAATGCA**CCCACGCTACTGGCGGGGTTT
ydcO-r001	**CTCCCACTACCAATGCC**ATATCGGTTTCTGTCAGCGAT
yddG-f001	**TACTTCCAATCCAATGCA**ACACGACAAAAAGCAACGCTC
yddG-r001	**CTCCCACTACCAATGCC**ACGACGTGTCGCCAGCCAGCA
ydeA-f001	**TACTTCCAATCCAATGCA**ACTGTTTCCCGCAAAGTGGCG
ydeA-r001	**CTCCCACTACCAATGCC**GAGTGTCACTGGCCAGCGGCG
ydeE-f001	**TACTTCCAATCCAATGCA**AACTTATCCCTACGACGCTCT
ydeE-r001	**CTCCCACTACCAATGCC**CGGTCTTGCTCGAATCCCTTT
ydfJ-f001	**TACTTCCAATCCAATGCA**GATTTCCAGTTATATTCGCTC

Appendix B

Primer name	Sequence
ydfJ-r001	**CTCCCACTACCAATGCC**CAGACTTCTGGAAGGTTGCGC
ydhC-f001	ACTTCCAATCCAATGCACCTGGGAAAAGATTTTTAGTC
ydhC-r001	**CTCCCACTACCAATGCC**ATTGCCATGATTCTGGCAGCC
ydhP-f001	**TACTTCCAATCCAATGCA**AAAATTAACTATCCGTTGCTG
ydhP-r001	**CTCCCACTACCAATGCC**GCTGTTAGCAACGCAAACTGT
ydiM-f001	**TACTTCCAATCCAATGCA**TTCCCTACCGCACTTGGGTTG
ydiM-r001	**CTCCCACTACCAATGCC**AGCGACATTTTCCTTTAGCGA
ydiN-f001	**TACTTCCAATCCAATGCA**TTCAGCACGCCATTTATCCTG
ydiN-r001	**CTCCCACTACCAATGCC**TTTACGCTCGCCAAACCGCAC
ydjE-f001	**TACTTCCAATCCAATGCA**GAACAATATGATCAAATTGGC
ydjE-r001	**CTCCCACTACCAATGCC**ATTCACCTCAGAAATCTCTTC
ydjK-f001	**TACTTCCAATCCAATGCA**GAACAGATAACAAAACCGCAT
ydjK-r001	**CTCCCACTACCAATGCC**TTTATTGGCTACTGCATCAAT
ydjY-f001	**TACTTCCAATCCAATGCA**TTCGCTGAATACGGGGTTCTG
ydjY-r001	**CTCCCACTACCAATGCC**GGATTGCAGCGTCGCCAGTCG
yeaN-f001	**TACTTCCAATCCAATGCA**GGCAAAAACAGGATTGTCCTT
yeaN-r001	**CTCCCACTACCAATGCC**GCGAATTTCTTTGTCTCTCCC
yebQ-f001	**TACTTCCAATCCAATGCA**CAGCGATACGGTGCGATATTA
yebQ-r001	**CTCCCACTACCAATGCC**TGCCCTGGATCGTGGCTGAGT
yegT-f001	**TACTTCCAATCCAATGCA**AAAACAACAGCAAAGCTGTCG
yegT-r001	**CTCCCACTACCAATGCC**TTTAACTTCCCCTTGTGTCAA
yfaV-f001	**TACTTCCAATCCAATGCA**AGCACCGCTTTGCTTGACGCC
yfaV-r001	**CTCCCACTACCAATGCC**ATGATGTGCCACGTCGGTCTG
yfcJ-f001	**TACTTCCAATCCAATGCA**GAAACACGATCTTCTGCCAAT
yfcJ-r001	**CTCCCACTACCAATGCC**CCGACGAAATGACAGTATCGT
ygcS-f001	**TACTTCCAATCCAATGCA**CCGCTTAACCGTTTTCACTGC
ygcS-r001	**CTCCCACTACCAATGCC**TACATTTCCCGCCGCCACCAG
yhhS-f001	**TACTTCCAATCCAATGCA**CCCGTAGCCGAACCCGCGCTA
yhhS-r001	**CTCCCACTACCAATGCC**TGAGGCGGCCTCAGGGACGTG
yhjE-f001	**TACTTCCAATCCAATGCA**TCGCGTAATAAAGTCCTTGTC
yhjE-r001C	TCCCACTACCAATGCCCAACGACTGATGTCGCGTCTC
yhjX-f001	**TACTTCCAATCCAATGCA**CAGCGTACCCGCTGGCTGACA
yhjX-r001	**CTCCCACTACCAATGCC**AAGGGAGCCATGCGCCTCACG

Primer name	Sequence
yicJ-f001	**TACTTCCAATCCAATGCA**AAGAGTGAAGTGTTGTCCGTT
yicJ-f001	**CTCCCACTACCAATGCC**GTTCTGCACTTCTTGAGAGGT
yihN-f001	**TACTTCCAATCCAATGCA**CTCACGAAAAAGAAATGGGCG
yihN-r001	**CTCCCACTACCAATGCC**AGCTTCAGCCGCCAGCATCTC
yihO-f001	**TACTTCCAATCCAATGCA**TCTGACCATAATCCACTGACA
yihO-r001	**CTCCCACTACCAATGCC**TAACGTTACGGAAGCCGTTTT
yihP-f001	**TACTTCCAATCCAATGCA**GGCAAAGGGAGAACATCGATG
yihP-r001	**CTCCCACTACCAATGCC**CGCCGTGCGTTTACGGGCTTC
yjdL-f001	**TACTTCCAATCCAATGCA**TCACAGCCGCGCGCGATATAC
yjdL-r001	**CTCCCACTACCAATGCC**ATCGTTGCTCTCCTGTATCAT
yjhB-f001	**TACTTCCAATCCAATGCA**CCACAACGGAAAGCTCTTTTT
yjhB-r001	**CTCCCACTACCAATGCC**TTTAGCCACGGATAGTTTATA
yjhF-f001	**TACTTCCAATCCAATGCA**CCACTAATTATCGTTGTGGCA
yjhF-r001	**CTCCCACTACCAATGCC**TAAAACTAAGCTGGCGAGCAG
yjiJ-f001	**TACTTCCAATCCAATGCA**CCTTCGTCCACGCATCCCGTA
yjiJ-r001	**CTCCCACTACCAATGCC**TGAAGTGACCACTTGTAGCTT
yjiY-f001	**TACTTCCAATCCAATGCA**GATACTAAAAAGATATTCAAG
yjiY-r001	**CTCCCACTACCAATGCC**GTGGTGCGAAGAGATCTTCAC
yjjL-f001	**TACTTCCAATCCAATGCA**CCAGATGGATTAGTTCAACGC
yjjL-r001	**CTCCCACTACCAATGCC**TTTACGTGGGTCGTTGATCGG
ylcA-f001	**TACTTCCAATCCAATGCA**CCGATGGGAGGACTGGGGCTA
ylcA-r001	**CTCCCACTACCAATGCC**CGAGACTAACATCCCGGTAAA
ynfM-f001	**TACTTCCAATCCAATGCA**CAATTTATTAAACGCGGTACG
ynfM-r001	**CTCCCACTACCAATGCC**GGCGTGCAGACGACGATGCAA
ynjF-f001	**TACTTCCAATCCAATGCA**CTAGACCGCCATCTTCATCCC
ynjF-r001	**CTCCCACTACCAATGCC**CTGACGCTGGAGCGACTTCAG
yqcE-f001	**TACTTCCAATCCAATGCA**TCATATCGCCGTTGGATAACC
yqcE-r001	**CTCCCACTACCAATGCC**GGGGGCGCTATCAGCAGTACG

Appendix B

Table 7.3 Primer sequences used to amplify the HsSWEET and PQ loop genes

Lower case sequence corresponds to the ligation independent cloning site. Upper case sequence is complementary to the gene of interest. All sequences are listed 5´ to 3´. Custom DNA oligos purchased from Eurofins, UK.

Name	Sequence
HsSWEETCysless_f001	tacttccaatccatgGAAGCTGGTGGTTTCCTGGAC
HsSWEETCysless_r001	tatccacctttactgGGTCTGCAGCAGCCAGTAGTTACGG
HINDIII-HsSWEETCysLess-f002	ctcaagcttATGGAAGCTGGTGGTTTCCTG
HsSWEETCysLess-BAMHI-r002	gccaggatccttaGGTCTGCAGCAGCCAGTAG
HINDIII-HsSWEETCysLess-f003	ctcaagcttATGTACCCATACGATGTTCCAGATTACGCTGAAGCT GGTGGTTTCCTGGAC
HsSWEETCysLess-BAMHI-r003	gccaggatccGAGGTCTGCAGCAGCCAGTAGTTAC
HINDIII-HsSWEETCysLess-f004	ctcaagcttCGATGGAAGCTGGTGGTTTCCTG
HsSWEETCysLess-BAMHI-r004	cgggatccttaAGCGTAATCTGGAACATCGTATGGGTAggtctgcag cagccagtagttac
SalI-GFP-f001	ctccgtcgacAGCAAGGGCGAGGAGCTG
GFP-HindIII-r001	tccgcaagcttCTACTTGTACAGCTCGTCC
EcoRI-SLC50A1-f001	cggaattccGAGGCGGGCGGCTTTCTGGA
SLC50A1-SalI-r001	gcttgtcgacGGTTTGCAGGAGCCAGTAG
EcoRI-HsSWTCysLess-f001	cggaattccGAAGCTGGTGGTTTCCTGG
HsSWTCysLess-SalI-r001	gcttgtcgacGGTCTGCAGCAGCCAGTAG

EcoRI-PQLC1-f001	cggaattccGAGGCCGAGGGCCTGGACTG
PQLC1-SalI-r001	gcttgtcgacGAGGGCCTTGGTGCCAGTG
EcoRI-mPQLC2-f001	cggaattccGTCTGGAGGACACTGGGCGC
mPQLC2-SalI-r001	gcttgtcgacGCTGGGGAGGAGGGGCTCT
EcoRI-PQLC3-f001	cggaattccGAGGCGGCGCTGCTGGGGCT
PQLC3-SalI-r001	gcttgtcgacTTCAGCCTTTATAGCGGTC
EcoRI-CTNS-f001	cggaattccTCTGTTTCTCTGACCGTTCCG
EcoRI-CTNS-f002	cggaattccTCTGCTATCTCTATCATCAAC
NcoI-CTNS-f001	atggccatgggtTCTGTTTCTCTGACCGTTC
CTNS-SalI-r001	gcttgtcgacGTTCAGCTGGTCGTAACCCGG
CTNS-NheI-r001	gctggctagcAGAAGAACGGATAACCAGG

List of References

Abe, T., Sakaue-Sawano, A., Kiyonari, H., Shioi, G., Inoue, K., Horiuchi, T., Nakao, K., Miyawaki, A., Aizawa, S. & Fujimori, T., 2013. Visualization of cell cycle in mouse embryos with Fucci2 reporter directed by Rosa26 promoter. Development, 140, 237-246.

Abramson, J. & Wright, E. M., 2009. Structure and function of Na+-symporters with inverted repeats. Current Opinion in Structural Biology, 19, 425-32.

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S., 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. Science, 301, 610-615.

Affleck, J. A., Helliwell, P. A., Kellett, G. L., 2003. Immunocytochemical detection of GLUT2 at the rat Intestinal brush-border membrane. The Journal of Histochemistry and Cytochemistry, 51, 1567-74.

Agmon, N., 2005. Proton pathways in green fluorescence protein. Biophysical Journal, 88, 2452-2461.

Ahad, F. & Ganie, S. A., 2010. Iodine, iodine metabolism and iodine deficiency disorders revisited. Indian Journal of Endocrinology and Metabolism, 14, 13-17.

Alegre, K. O. & Law, C. J., 2015. Purification of a multidrug resistance transporter for crystallization studies. Antibiotics, 4, 113-135.

Al-Hasani, H., Kunamneni, R. K., Dawson, K., Hinck, C. S., Müller-Wieland, D. & Cushman, S. W., 2002. Roles of the N- and C-termini of GLUT4 in endocytosis. Journal of Cell Science, 115, 131-140.

Al-Tawfiq, J. A., Laxminarayan, R. & Mendelson, M., 2017. How should we respond to the emergence of plasmid-mediated colistin resistance in humans and animals?. International Journal of Infectious Diseases, 54, 77-84.

Amir Shaghaghi, M., Murphy, B. & Eck, P., 2016. The SLC2A14 gene: genomic locus, tissue expression, splice variants and subcellular localization of the protein. Biochemistry and Cell Biology, 94, 331-335.

Amsterdam, A., Lin, S., Moss, L. G. & Hopkins, N., 1996. Requirements for green fluorescent protein detection in transgenic zebrafish embryos. Gene, 173, 99-103.

Anandan, A. & Vrielink, A., 2016. Detergents in membrane protein purification and crystallisation. Advances in Experimental Medicine and Biology, 922, 13-28.

Anderle, P., Sengstag, T., Mutch, D. M., Rumbo, M., Praz, V., Mansourian, R., Delorenzi, M., Williamson, G. & Roberts, M. A., 2005. Changes in the

transcriptional profile of transporters in the intestine along the anterior-posterior and crypt-villus axes. BMC Genomics, 6, 69-86.

Andrade, F. K., Moreira, S. M., Domingues, L. & Gama, F. M., 2010. Improving the affinity of fibroblasts for bacterial cellulose using carbohydrate-binding modules fused to RGD. Journal of Biomedical Materials Research, 92, 9-17.

Andresen, M., Stiel, A. C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M. C., Hell, S. W. & Jakobs, S., 2007. Structural basis for reversible photoswitching in Dronpa. Proceedings of the National Academy of Sciences of the United States of America, 104, 13005-13009.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L. S., 2004. UniProt: the universal protein knowledgebase. Nucleic Acids Research, 32, 115-119.

Arluison, M., Quignon, M., Nguyen, P., Thorens, B., Leloup, C. & Penicaud, L., 2004. Distribution and anatomical localization of the glucose transporter 2 (GLUT2) in the adult rat brain—an immunohistochemical study. Journal of Chemical Neuroanatomy, 28, 117-136.

Backmark, A. E., Olivier, N., Snijder, A., Gordon, E., Dekker, N. & Ferguson, A. D., 2013. Fluorescent probe for high-throughput screening of membrane protein expression. Protein Science, 22, 1124-1132.

Baghirova, S., Hughes, B. G., Hendzel, M. J. & Schulz, R., 2015. Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. Methods, 7;2, 440-445.

Bahrami, A., Shams, S. F., Eidgahi, E. S., Lotfi, Z., Sheikhi, M. & Shakeri, S., 2017. Epidemiology of infectious complications in renal allograft recipients in the first year after transplant. Experimental and Clinical Transplantation, 15, 631-635.

Bajaj, H., Scorciapino, M. A., Moynié, L., Page, M. G., Naismith, J. H., Ceccarelli, M. & Winterhalter, M., 2016. Molecular basis of filtering carbapenems by porins from β-lactam-resistant clinical strains of *Escherichia coli*. The Journal of Biological Chemistry, 291, 2837-2847.

Balamurugan, K., Ortiz, A. & Said, H. M., 2003. Biotin uptake by human intestinal and liver epithelial cells: role of the SMVT system. American Journal of Physiology, 285, 73-77.

Banfield, D. K., 2011. Mechanisms of protein retention in the Golgi. Cold Spring Harbor Perspectives in Biology, 3, a005264.

Basketter, D. A. & Widdas, W. F., 1978. Asymmetry of the hexose transfer system in human erythrocytes. Comparison of the effects of cytochalasin B, phloretin and maltose as competitive inhibitors. The Journal of Physiology, 278, 389-401

Baumgarten, T., Schlegel, S., Wagner, S., Löw, M., Eriksson, J., Bonde, I., Herrgård, M. J., Heipieper, H. J., Nørholm, M. H., Slotboom, D. J. & de Gier, J. W., 2017. Isolation and characterization of the *E. coli* membrane protein production strain Mutant56 (DE3). Scientific Reports, 7, 45089.

Beis, K., 2015. Structural basis for the mechanism of ABC transporters. Biochemical Society Transactions, 43, 889-893.

Bentley, J., Hyatt, L. S., Ainley, K., Parish, J. H., Herbert, R. B. & White, G. R., 1993. Cloning and sequence analysis of an *Escherichia coli* gene conferring bicyclomycin resistance. Gene, 127, 117-120.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E., 2000. The Protein Data Bank. Nucleic Acids Research, 28, 235-242.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., Von Heijne, G. & Elofsson, A., 2008. Prediction of membrane-protein topology from first principles. Proceedings of the National Academy of Sciences of the United States of America, 105, 7177-7181.

Berry, G. T., Wu, S., Buccafusca, R., Ren, J., Gonzales, L. W., Ballard. P. L., Golden, J. A., Stevens, M. J. & Greer, J. J., 2003. Loss of murine Na+/myo-Inositol cotransporter leads to brain myo-Inositol depletion and central apnea. The Journal of Biological Chemistry, 278, 18297-182302.

Bershadsky, A. D & Futerman, A. H., 1994. Disruption of the Golgi apparatus by brefeldin A blocks cell polarization and inhibits directed cell migration. Proceedings of the National Academy of Sciences of the United States of America, 91, 5686-5689.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo, Cassarino, T., Bertoni, M., Bordoli, L. & Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 42, 252-258.

Bird, L. E., Rada, H., Verma, A., Gasper, R., Birch, J., Jennions, M., Löwe, J., Moraes, I. & Owens, R. J., 2015. Green fluorescent protein-based expression screening of membrane proteins in *Escherichia coli*. Journal of Visualised Experiments, 6, 52357.

Blodgett, D. M., Graybill C & Carruthers A., 2008. Analysis of glucose transporter topology and structural dynamics. The Journal of Biological Chemistry, 283, 36416-36424.

Bogdanov, A. M., Mishin, A. S., Yampolsky, I. V., Belousov, V. V., Chudakov, D. M., Subach, F. V., Verkhusha, V. V., Lukyanov, S. & Lukyanov, K. A., 2009. Green fluorescent proteins are light-induced electron donors. Nature Chemical Biology, 5, 459-461.

Bogliolo, M. & Surrallés, J., 2015. Fanconi anemia: a model disease for studies on human genetics and advanced therapeutics. Current Opinion in Genetics & Development, 33, 32-40.

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H. & Falkow, S., 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene, 2, 95-113.

Borirak, O., Bekker, M. & Hellingwerf, K. J., 2014. Molecular physiology of the dynamic regulation of carbon catabolite repression in *Escherichia coli*. Microbiology, 160, 1214-1223.

Bou-Abdallah, F., Chasteen, N. D. & Lesser, M. P., 2006. Quenching of superoxide radicals by green fluorescent protein. Biochimica et Biophysica Acta, 1760, 1690-1695.

Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M. & Remington, S. J., 1997. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. Proceedings of the National Academy of Sciences of the United States of America, 94, 2306-2311.

Britton, R. A., Lin, D. C. & Grossman, A. D., 1998. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. Genes & Development, 12, 1254-1259.

Brooks, C. L., Morrison, M. & Joanne Lemieux, M., 2013. Rapid expression screening of eukaryotic membrane proteins in Pichia pastoris. Protein Science, 22, 425-433.

Brown, P. R., Odet, F., Bortner, C. D. & Eddy, E. M., 2014. Reporter Mice Express Green Fluorescent Protein at Initiation of Meiosis in Spermatocytes. Genesis, 52, 976-984.

Brunner, J. D., Lim, N. K., Schenck, S., Duerst, A. & Dutzler, R., 2014. X-ray structure of a calcium-activated TMEM16 lipid scramblase. Nature, 516, 207-212.

Bu, W., Chou, A. M., Lim, K. B., Sudhaharan, T. & Ahmed, S., 2009. The Toca-1-N-WASP complex links filopodial formation to endocytosis. The Journal of Biological Chemistry, 284, 11622-11636.

Burant, C. F., Takeda, J., Brot-Laroche, E., Bell, G. I. & Davidson, N. O., 1992. Fructose transporter in human spermatozoa and small intestine is GLUT5. The Journal of Biological Chemistry, 267, 14523-14526.

Büssow, K., Nordhoff, E., Lübbert, C., Lehrach, H. & Walter, G., 2000. A human cDNA library for high-throughput protein expression screening. Genomics, 65, 1-8.

Cai, B., Chen, X., Liu, F., Li, J., Gu, L., Liu, J. R. & Liu, J., 2014. A cell-based functional assay using a green fluorescent protein-based calcium indicator dCys-GCaMP. Assay and Drug Development Technologies, 12, 342-351.

Callaghan, R., Luk, F. & Bebawy, M., 2014. Inhibition of the multidrug resistance P-glycoprotein: time for a change of strategy. Drug Metabolism and Disposition, 42, 623-631.

Campelo, A. B., Rodríguez, A. & Martínez, B., 2010. Use of green fluorescent protein to monitor cell envelope stress in Lactococcus lactis. Applied and Environmental Microbiology, 76, 978-981.

Canbolat, M. F., Gera, N., Tang, C., Monian, B., Rao, B. M., Pourdeyhimi, B. & Khan, S. A., 2013. Preservation of cell viability and protein conformation on immobilization within nanofibers via electrospinning functionalized yeast. ACS Applied Materials & Interfaces, 5, 9349-9354.

Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S., 2008. Overcoming the challenges of membrane protein crystallography. Current Opinion in Structural Biology, 18, 581-586.

Casagrande, F., Harder, D., Schenk, A., Meury, M., Ucurum, Z., Engel, A., Weitz, D., Daniel, H. & Fotiadis, D., 2009. Projection structure of DtpD (YbgH), a prokaryotic member of the peptide transporter family. Journal of Molecular Biology, 394, 708-717.

Casaite, V., Bruzyte, S., Bukauskas, V., Setkus, A., Morozova-Roche, L. A. & Meskys, R., 2009. Expression and purification of active recombinant equine lysozyme in *Escherichia coli*. Protein Engineering, Design & Selection, 22, 649-654.

Chae, H. J., Delisa, M. P., Cha, H. J., Weigand, W. A., Rao, G. & Bentley, W. E., 2000. Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. Biotechnology and Bioengineering, 69, 275-285.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C., 1994. Green fluorescent protein as a marker for gene expression. Science, 263, 802-805.

Chang, C. Y., Hu, H. T., Tsai, C. H. & Wu, W. F., 2016. The degradation of RcsA by ClpYQ (HslUV) protease in *Escherichia coli*. Microbiological Research, 184, 42-50.

Chaptal, V., Kwon, S., Sawaya, M. R., Guan, L., Kaback, H. R. & Abramson, J., 2011. Crystal structure of lactose permease in complex with an affinity inactivator yields

unique insight into sugar recognition. Proceedings of the National Academy of Sciences of the United States of America, 108, 9361-9366.
Chardon, F., Bedu, M., Calenge, F., Klemens, P. A., Spinner, L., Clement, G., Chietera, G., Léran, S., Ferrand, M., Lacombe, B., Loudet, O., Dinant, S., Bellini, C., Neuhaus, H. E., Daniel-Vedele, F. & Krapp, A., 2013. Leaf fructose content Is controlled by the vacuolar transporter SWEET17 in *Arabidopsis*. Current Biology, 23, 697-702.
Chaudhary, S., Pak, J. E., Pedersen, B. P., Bang, L. J., Zhang, L. B., Ngaw, S. M., Green, R. G., Sharma, V., & Stroud, R. M., 2011. Efficient expression screening of human membrane proteins in transiently transfected Human Embryonic Kidney 293S cells. Methods, 55, 273-280.
Cheesemann, C. I., 1993. GLUT2 is the transporter for fructose across the rat intestinal basolateral membrane. Gastroenterology, 105, 1050-1056.
Chen, H., Shaffer, P. L., Huang, X. & Rose, P. E., 2013. Rapid screening of membrane protein expression in transiently transfected insect cells. Protein Expression and Purification, 88, 134-142.
Chen, L. Q., Lin, I. W., Qu, X. Q., Sosso, D., McFarlane, H. E., Londoño, A., Samuels, A. L. & Frommer, W. B., 2015. A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. The Plant Cell, 27, 607-619.
Chen, L., Hou, B, Lalonde, S., Takanaga, H, Hartung, M. L, Qu, X., Guo, W., Kim, J., Underwood, W., Chaudhuri, B., Chermak, D., Antony, G., White, F. F., Somerville, S. C., Mudgett, M. B, 2 & Frommer, W. B, 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. Nature, 468, 527-532.
Chen, S., Hong, Y., Liu, Y., Liu, J., Leung, C. W., Li, M., Kwok, R. T., Zhao, E., Lam, J. W., Yu, Y. & Tang, B. Z., 2013. Full-range intracellular ph sensing by an aggregation-induced emission-active two-channel ratiometric fluorogen. Journal of the American Chemical Society, 135, 4926-4929.
Chen, T. W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K. & Kim, D. S., 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature, 499, 295-300.
Chitsaz, M. & Brown, M. H., 2017. The role played by drug efflux pumps in bacterial multidrug resistance. Essays in Biochemistry, 61, 127-139.
Christie, J. M., 2007. Phototropin blue-light receptors. Annual Review of Plant Biology, 58, 21-45.

Christie, J. M., Hitomi, K., Arvai, A. S., Hartfield, K. A., Mettlen, M., Pratt, A. J., Tainer, J. A. & Getzoff, E. D., 2012. Structural tuning of the fluorescent protein iLOV for improved photostability. The Journal of Biological Chemistry, 287, 22295-22304.

Clarson, L. H., Glazier, J. D., Sides, M. K. & Sibley, C. P., 1997. Expression of the facilitated glucose transporters (GLUT1 and GLUT3) by a choriocarcinoma cell line (JAr) and cytotrophoblast cells in culture. Placenta, 18, 333-339.

Coady, M. J., Chang, M. H., Charron, F. M., Plata, C., Wallendorff, B., Sah, J. F., Markowitz, S. D., Romero, M. F. & Lapointe, J. Y., 2004. The human tumour suppressor gene SLC5A8 expresses a Na+–monocarboxylate cotransporter. The Journal of Physiology, 557, 719-731.

Coady, M. J., Wallendorff, B., Gagnon, D. G. & Lapointe, J. Y., 2002. Identification of a novel Na+/myo-inositol cotransporter. The Journal of Biological Chemistry, 277, 35219-35224.

Cody, C. W., Prasher, D. C., Westler, W. M., Prendergast, F. G. & Ward, W. W., 1993. Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein. Biochemistry, 32, 1212-1218.

Colville, C. A., Seatter, M. J., Jess, T. J., Gould, G. W. & Thomas, H. M., 1993. Kinetic analysis of the liver-type (GLUT2) and brain-type (GLUT3) glucose transporters in Xenopus oocytes: substrate specificities and effects of transport inhibitors. The Biochemical Journal, 290, 701-706.

Conti, E., Franks, N. P. & Brick, P., 1996. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure, 4, 287-298.

Cormack, B., 1998. Green fluorescent protein as a reporter of transcription and protein localization in fungi. Current Opinion in Microbiology, 1, 406-410.

Coutard, B., Gagnaire, M., Guilhon, A. A., Berro, M., Canaan, S. & Bignon, C., 2008. Green fluorescent protein and factorial approach: An effective parnership for screening the soluble expression of recombinant proteins in *Escherichia coli*. Protein Expression and Purification, 61, 184-190.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P., 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. Nature Biotechnology, 14, 315-319.

Crick, F., 1970. Central dogma of molecular biology. Nature, 227, 561-563.

Cui, B., Zhang, L., Song, Y., Wei, J., Li, C., Wang, T., Wang, Y., Zhao, T. & Shen, X., 2014. Engineering an enhanced, thermostable, monomeric bacterial luciferase gene as a reporter in plant protoplasts. PLoS One, 9, e107885.

Cura, A. J. & Carruthers, A., 2012. The role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism and homeostasis. Comprehensive Physiology, 2, 863-914.

Dabrowski, S. & Kiaer Ahring, B., 2003. Cloning, expression, and purification of the His6-tagged hyper-thermostable dUTPase from *Pyrococcus woesei* in *Escherichia coli*: application in PCR. Protein Expression and Purification, 31, 72-78.

Dai, G., Yu, H., Kruse, M., Traynor-Kaplan, A. & Hille, B., 2016. Osmoregulatory inositol transporter SMIT1 modulates electrical activity by adjusting PI(4,5)P2 levels. Proceedings of the National Academy of Sciences of the United States of America, 113, 3290-3299.

Dai, Z., Chung, S. K., Miao, D., Lau, K. S., Chan, A. W. & Kung A.W., 2011. Sodium/myo-inositol cotransporter 1 and myo-inositol are essential for osteogenesis and bone formation. Journal of Bone and Mineral Research, 26, 582-90.

Daley, D. O., Rapp, M., Granseth, E., Melén, K., Drew, D. & von Heijne, G., 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. Science, 308, 1321-1323.

Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J. & Yan, N., 2010. Structure of a fucose transporter in an outward-open conformation. Nature, 467, 734-738.

Daruwalla, K. R., Paxton, A. T. & Henderson, P. J., 1981. Energization of the transport systems for arabinose and comparison with galactose transport in *Escherichia coli*. The Biochemical Journal, 200, 611-627.

Datsenko, K. A. & Wanner, B. L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America, 97, 6640-6645.

Davenport, A. P. & Kuc, R. E., 2012. Cellular localization of receptors using antibodies visualized by light and dual labeling confocal microscopy. Methods in Molecular Biology, 897, 239-260.

Davidson, N. O., Hausman, A. M., Ifkovits, C. A., Buse, J. B., Gould, G. W., Burant, C. F. & Bell, G. I., 1992. Human intestinal glucose transporter expression and localization of GLUT5. The American Journal of Physiology, 262, 795-800.

Dawson, P. A., Mychaleckyj, J. C., Fossey, S. C., Mihic, S. J., Craddock, A. L. & Bowden, D. W., 2001. Sequence and functional analysis of GLUT10: a glucose transporter in the type 2 diabetes-linked region of chromosome 20q12-13.1. Molecular Genetics and Metabolism, 74, 186-199.

De Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B., 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnology, 7, e32.

De Marco, A., Vigh, L., Diamant, S. & Goloubinoff, P., 2005. Native folding of aggregation-prone recombinant proteins in *Escherichia coli* by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. Cell Stress & Chaperones, 10, 329-339.

De Wet, J. R., Wood, K. V., Helinski, D. R. & DeLuca, M., 1985. Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America, 82, 7870-7873

Del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. & Díaz-Orejas, R., 1998. Replication and control of circular bacterial plasmids. Microbiology and Molecular Biology Reviews, 62, 434-464.

Delaere, F., Duchampt, A., Mounien, L., Seyer, P., Duraffourd, C., Zitoun, C., Thorens, B. & Mithieux, G., 2012. The role of sodium-coupled glucose co-transporter 3 in the satiety effect of portal glucose sensing. Molecular Metabolism, 2, 47-53.

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S. & Yan, N., 2015. Molecular basis of ligand recognition and transport by glucose transporters. Nature, 526, 391-396.

Deng, D., Xu, C., Sun, P., Wu, J., Yan, C., Hu, M., Yan, N., 2014. Crystal structure of the human glucose transporter GLUT1. Nature, 510, 121-125.

Denoncin, K., Schwalm, J., Vertommen, D., Silhavy, T. J. & Collet, J. F., 2012. Dissecting the *Escherichia coli* periplasmic chaperone network using differential proteomics. Proteomics, 12, 1391-1401.

Desai, K. K. & Miller, B. G., 2010. Recruitment of genes and enzymes conferring resistance to the nonnatural toxin bromoacetate. Proceedings of the National Academy of Sciences of the United States of America, 107, 17968-17973.

Dey, A. & Ramachandran, R., 2014. Cloning, overexpression, purification and preliminary X-ray analysis of a feast/famine regulatory protein (Rv2779c) from Mycobacterium tuberculosis H37Rv. Acta Crystallographica, 70, 97-100.

Diez-Sampedro, A., Hirayama, B. A., Osswald, C., Gorboulev, V., Baumgarten, K., Volk, C., Wright, E. M., Koepsell, H., 2003. A glucose sensor hiding in a family of transporters. Proceedings of the National Academy of Sciences of the United States of America, 100, 11753-11758.

Dinh, T. & Bernhardt, T. G., 2011. Using Superfolder green fluorescent protein for periplasmic protein localization studies. Journal of Bacteriology, 193, 4984-4987.
Doege, H., Bocianski, A., Joost, H. G. & Schürmann, A., 2000. Activity and genomic organization of human glucose transporter 9 (GLUT9), a novel member of the family of sugar-transport facilitators predominantly expressed in brain and leucocytes. The Biochemical Journal, 350, 771-776.
Doege, H., Bocianski, A., Scheepers, A., Axer, H., Eckel, J., Joost, H. G. & Schürmann, A., 2001. Characterization of human glucose transporter (GLUT) 11 (encoded by SLC2A11), a novel sugar-transport facilitator specifically expressed in heart and skeletal muscle. The Biochemical Journal, 359, 443-449.
Doege, H., Schürmann, A., Bahrenberg, G., Brauers, A. & Joost, H. G., 2000. GLUT8, a novel member of the sugar transport facilitator family with glucose transport activity. The Journal of Biological Chemistry, 275, 16275-16280.
Doki, S., Kato, H. E., Solcan, N., Iwaki, M., Koyama, M., Hattori, M., Iwase, N., Tsukazaki, T., Sugita, Y., Kandori, H., Newstead, S., Ishitani, R. & Nureki, O., 2013. Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. Proceedings of the National Academy of Sciences of the United States of America, 110, 11343-11348.
Dong, H., Nilsson, L. & Kurland, C. G., 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. Journal of Bacteriology, 177, 1497-1504.
Dornmair, K., 1988. An improved purification of lactose permease. FEBS Letters, 235, 35-39.
Dove, S., 2004. Scleractinian corals with photoprotective host pigments are hypersensitive to thermal bleaching. Marine Ecology Progress Series, 272, 99-116.
Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G. & Iwata, S., 2008. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. Nature Protocols, 3, 708-798.
Du, L. & Heaney, A. P., 2012. Regulation of adipose differentiation by fructose and GluT5. Molecular Endocrinology, 26, 1773-1782.
Dubendorff, J. W. & Studier, F. W., 1991. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. Journal of Molecular Biology, 219, 45-59.
Dumon-Seignovert, L., Cariot, G. & 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, 203-206.

Duran, A. M. & Meiler, J., 2013. Inverted topologies in membrane proteins: a mini-review. Computational and Structural Biotechnology Journal, 8, e201308004.

Ebert, K., Ludwig, M., Geillinger, K. E., Schoberth, G. C., Essenwanger, J., Stolz, J., Daniel, H. & Witt, H., 2017. Reassessment of GLUT7 and GLUT9 as putative fructose and glucose transporters. The Journal of Membrane Biology, 250, 171-182.

Edgar, R. & Bibi, E., 1997. MdfA, an *Escherichia coli* multidrug resistance protein with an extraordinarily broad spectrum of drug recognition. Journal of Bacteriology, 179, 2274-2280.

Edwards, A. M., Isserlin, R., Bader, G. D., Frye, S. V., Willson, T. M. & Yu, F. H., 2011. Too many roads not taken. Nature, 470, 163-165.

Elion, E. A., 2007. Detection of protein-protein interactions by coprecipitation. Current Protocols in Immunology, 20, e10.1002

Elliott, G., McGrath, J. & Crockett-Torabi, E., 2000. Green fluorescent protein A novel viability assay for cryobiological applications. Cryobiology, 40, 360-369.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K., 2010. Features and development of Coot. Acta Crystallographica, 66, 486-501.

Epel, B. L., Padgett, H. S., Heinlein, M. & Beachy, R. N., 1996. Plant virus movement protein dynamics probed with a GFP-protein fusion. Gene, 173, 75-79.

Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P., 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. Analytical Biochemistry, 357, 289-298.

Esposito, D. & Chatterjee, D. K., 2006. Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology, 17, 353-358.

Ethayathulla, A. S., Yousef, M. S., Amin, A., Leblanc, G., Kaback, H. R. & Guan, L., 2014. Structure-based mechanism for Na+/melibiose symport by MelB. Nature Communications, 5, e3009.

European Centre for Disease Prevention and Control, ECHPC, 2015. Summary of the latest data on antibiotic resistance in the European Union. Antimicrobial Resistance Surveillance in Europe, 2015, e2900/23549.

Fàbrega, A., Martin, R. G., Rosner, J. L., Tavio, M. M. & Vila, J., 2010. Constitutive SoxS expression in a fluoroquinolone-resistant strain with a truncated SoxR protein and identification of a new member of the marA-soxS-rob regulon, mdtG. Antimicrobial Agents and Chemotherapy, 54, 1218-1225.

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L., 2010. Prediction of the human membrane proteome. Proteomics, 10, 1141-1149.

Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M. & Abramson, J., 2008. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. Science, 321, 810-814.

Fantin, V. R., St-Pierre, J. & Leder, P., 2006. Attenuation of LDH-A expression uncovers a link between glycolysis, mitochondrial physiology, and tumor maintenance. Cancer Cell, 9, 425-434.

Fatmi, M. Q. & Chang, C. E., 2010. The role of oligomerization and cooperative regulation in protein function: The Case of Tryptophan Synthase. PLoS Computational Biology, 6, 1000994-1001008

Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J., 2000. Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. Journal of Bacteriology, 182, 4068-4076.

Ferrández, A., Garciá, J. L. & Díaz, E., 1997. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of *Escherichia coli* K-12. Journal of Bacteriology, 179, 2573-2581.

Fevre, C., Jbel, M., Passet, V., Weill, F. X., Grimont, P. A. & Brisse, S., 2005. Six groups of the OXY β-Lactamase evolved over millions of years in *Klebsiella oxytoca*. Antimicrobial Agents and Chemotherapy, 49, 3453-3462.

Fiamegos, Y. C., Kastritis, P. L., Exarchou, V., Han, H., Bonvin, A. M., Vervoort, J., Lewis, K., Hamblin, M. R. & Tegos, G. P., 2011. Antimicrobial and efflux pump inhibitory activity of caffeoylquinic acids from *Artemisia absinthium* against gram-positive pathogenic bacteria. PLoS One, 6, e18127.

Field, S. F., Bulina, M. Y., Kelmanson, I. V., Bielawski, J. P. & Matz, M. V., 2006. Adaptive evolution of multicolored fluorescent proteins in reef-building corals. Journal of Molecular Evolution, 62, 332-339.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J. & Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Research, 42, 222-230.

Fitzgerald, J., Leonard, P., Darcy, E. & O'Kennedy, R., 2011. Immunoaffinity chromatography. Methods in Molecular Biology, 681, 35-59.

Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J., 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nature Reviews Microbiology, 13, 269-284.

Fluman, N., Adler, J., Rotenberg, S. A., Brown, M. H. & Bibi, E., 2014. Export of a single drug molecule in two transport cycles by a multidrug efflux pump. Nature Communications, 5, e4615.

Fluman, N., Cohen-Karni, D., Weiss, T. & Bibi, E., 2009. A promiscuous conformational switch in the secondary multidrug transporter MdfA. The Journal of Biological Chemistry, 284, 32296-32304.

Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y., Mendieta, K. S., Jones, K. L., Ocampo, A., Rudd, K. E. & Storz, G., 2008. Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. Molecular Microbiology, 70, 1076-1093.

Franken, H., Mathieson, T., Childs, D., Sweetman, G. M., Werner, T., Tögel, I., Doce, C., Gade, S., Bantscheff, M., Drewes, G., Reinhard, F. B., Huber, W. & Savitski, M. M., 2015. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nature Protocols, 10, 1567-1593.

Fukuda, M., Takeda, H., Kato, H. E., Doki, S., Ito, K., Maturana, A. D., Ishitani, R. & Nureki, O., 2015. Structural basis for dynamic mechanism of nitrate/nitrite antiport by NarK. Nature Communications, 6, e7097.

Fukuzawa, T., Fukazawa, M., Ueda, O., Shimada, H., Kito, A., Kakefuda, M., Kawase, Y., Wada, N. A., Goto, C., Fukushima, N., Jishage, K., Honda, K., King, G. L., Kawabe, Y., 2013. SGLT5 reabsorbs fructose in the kidney but its deficiency paradoxically exacerbates hepatic steatosis induced by fructose. PLoS One, 8, 56681-56691

Futai, M., Nakanishi-Matsui, M., Okamoto, H., Sekiya, M. & Nakamoto, R. K., 2012. Rotational catalysis in proton pumping ATPases: from *E. coli* F-ATPase to mammalian V-ATPase. Biochimica et Biophysica Acta, 1817, 1711-1721.

Gahl, W. A., Balog, J. Z. & Kleta, R., 2007. Nephropathic cystinosis in adults: natural history and effects of oral cysteamine therapy. Annals of Internal Medicine, 21, 242-250.

Gecchele, E., Merlin, M., Brozzetti, A., Falorni, A., Pezzotti, M. & Avesani, L., 2015. A comparative analysis of recombinant protein expression in different biofactories: bacteria, insect cells and plant systems. Journal of Visualised Experiments, 97, e52459.

Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., Klein, S. L., Old, S., Rasooly, R., Good, P., Guyer, M., Peck, A. M., Derge, J. G., Lipman, D., Collins, F. S., Jang, W., Sherry, S., Feolo, M., Misquitta, L., Lee, E., Rotmistrovsky, K., Greenhut, S. F., Schaefer, C. F., Buetow, K., Bonner, T. I.,

Haussler, D., Kent, J., Kiekhaus, M., Furey, T., Brent, M., Prange, C., Schreiber, K., Shapiro, N., Bhat, N. K., Hopkins, R. F., Hsie, F., Driscoll, T., Soares, M. B., Casavant, T. L., Scheetz, T. E., Brown-stein, M. J., Usdin, T. B., Toshiyuki, S., Carninci, P., Piao, Y., Dudekula, D. B., Ko, M. S., Kawakami, K., Suzuki, Y., Sugano, S., Gruber, C. E., Smith, M. R., Simmons, B., Moore, T., Waterman, R., Johnson, S. L., Ruan, Y., Wei, C. L., Mathavan, S., Gunaratne, P. H., Wu, J., Garcia, A. M., Hulyk, S. W., Fuh, E., Yuan, Y., Sneed, A., Kowis, C., Hodgson, A., Muzny, D. M., McPherson, J., Gibbs, R. A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madari, A., Young, A. C., Wetherby, K. D., Granite, S. J., Kwong, P. N., Brinkley, C. P., Pearson, R. L., Bouffard, G. G., Blakesly, R. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Griffith, M., Griffith, O. L., Krzywinski, M. I., Liao, N., Morin, R., Palmquist, D., Petrescu, A. S., Skalska, U., Smailus, D. E., Stott, J. M., Schnerch, A., Schein, J. E., Jones, S. J., Holt, R. A., Baross, A., Marra, M. A., Clifton, S., Makowski, K. A., Bosak, S., Malek, J. & MGC Project Team, 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). Genome Research, 14, 2121-2127.

Gill, R. T., Valdes, J. J. & Bentley, W. E., 2000. A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. Metabolic Engineering, 2, 178-189.

Gilman, A. G., 1970. A protein binding assay for adenosine 3′:5′-cyclic monophosphate. Proceedings of the National Academy of Sciences of the United States of America, 67, 305-312.

Ginsburg, H. & Stein, W. D., 1975. Zero-trans and infinite-cis uptake of galactose in human erythrocytes. Biochimica et Biophysica Acta, 382, 353-368.

Göbl, C., Resch, M., Strickland, M., Hartlmüller, C., Viertler, M., Tjandra, N. & Madl, T., 2016. increasing the chemical-shift dispersion of unstructured proteins with a covalent lanthanide shift reagent. Angewandte Chemie, 55, 14847-14851.

Goodrich-Blair, H. & Kolter, R., 2000. Homocysteine thiolactone is a positive effector of sigma(S) levels in *Escherichia coli*. FEMS Microbiology Letters, 185, 117-121.

Gopal, E., Umapathy, N. S., Martin, P. M., Ananth, S., Gnana-Prakasam, J. P., Becker, H., Wagner, C. A., Ganapathy, V. & Prasad, P. D., 2007. Cloning and functional characterization of human SMCT2 (SLC5A12) and expression pattern of the transporter in kidney. Biochimica et Biophysica Acta, 1768, 2690-2697.

Gopalakrishnan, A. S., Chen, Y. C., Temkin, M. & Dowhan, W., 1986. Structure and expression of the gene locus encoding the phosphatidylglycerophosphate synthase of *Escherichia coli*. The Journal of Biological Chemistry, 261, 1329-1338.

Gorboulev, V., Schürmann, A., Vallon, V., Kipp, H., Jaschke, A., Klessen, D., Friedrich, A., Scherneck, S., Rieg, T., Cunard, R., Veyhl-Wichmann, M., Srinivasan, A., Balen, D., Breljak, D., Rexhepaj, R., Parker, H. E., Gribble, F. M., Reimann, F., Lang, F., Wiese, S., Sabolic, I., Sendtner, M. & Koepsell, H., 2012. Na(+)-D-glucose cotransporter SGLT1 is pivotal for intestinal glucose absorption and glucose-dependent incretin secretion. Diabetes, 61, 187-196.

Gorleku, O. A., Barns, A. M., Prescott, G. R., Greaves, J. & Chamberlain, L. H., 2011. Endoplasmic reticulum localization of DHHC palmitoyltransferases mediated by lysine-based sorting signals. The Journal of Biological Chemistry, 286, 39573-39584.

Gözü, H. I., Yavuzer, D., Kaya, H., Vural, S., Sargın, H., Gezen, C., Sargın, M. & Akalın, S., 2010. Alterations of NIS expression in functioning thyroid nodules. Journal of Ear, Nose and Throat, 20, 285-292.

Gras, D., Roze, E., Caillet, S., Méneret, A., Doummar, D., Billette de Villemeur, T., Vidailhet, M. & Mochel, F., 2014. GLUT1 deficiency syndrome: an update. Revue Neurologique, 170, 91-99.

Green, A. A. & McElroy, W. D., 1956. Crystalline firefly luciferase. Biochimica et Biophysica Acta, 20, 170-176.

Green, J., Stapleton, M. R., Smith, L. J., Artymiuk, P. J., Kahramanoglou, C., Hunt, D. M. & Buxton, R. S., 2014. Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. Current Opinion in Microbiology, 18, 1-7.

Grempler, R., Augustin, R., Froehner, S., Hildebrandt, T., Simon, E., Mark, M. & Eickelmann, P., 2012. Functional characterisation of human SGLT-5 as a novel kidney-specific sodium-dependent sugar transporter. FEBS Letters, 586, 248-253.

Griffith, J. K., Baker, M. E., Rouch, D. A., Page, M. G., Skurray, R. A., Paulsen, I. T., Chater, K. F., Baldwin, S. A. & Henderson, P. J., 1992. Membrane transport proteins: implications of sequence comparisons. Current Opinion in Cell Biology, 4, 684-695.

Groszmann, M., Osborn, H. L. & Evans, J. R., 2017. Carbon dioxide and water transport through plant aquaporins. Plant Cell & Environment, 40, 938-961.

Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R., 2007. Structural determination of wild-type lactose permease. Proceedings of the National Academy of Sciences of the United States of America, 104, 15294-15298.

Gubellini, F., Verdon, G., Karpowich, N. K., Luff, J. D., Boël, G., Gauthier, N., Handelman, S. K., Ades, S. E. & Hunt, J. F., 2011. physiological response to membrane protein overexpression in *E. coli*. Molecular & Cellular Proteomics, 10, e007930.

Guilloton, M. B., Lamblin, A. F., Kozliak, E. I., Gerami-Nejad, M., Tu, C., Silverman, D., Anderson, P. M. & Fuchs, J. A., 1993. A physiological role for cyanate-induced carbonic anhydrase in *Escherichia coli*. Journal of Bacteriology, 175, 1443-1451.

Gupta, K., Hooton, T. M. & Stamm, W. E., 2001. Increasing antimicrobial resistance and the management of uncomplicated community-acquired urinary tract infections. Annals of Internal Medicine, 135, 41-50.

Gurav, A., Sivaprakasam, S., Bhutia, Y. D., Boettger, T., Singh, N. & Ganapathy, V., 2015. Slc5a8, a Na+-coupled high-affinity transporter for short-chain fatty acids, is a conditional tumor suppressor in colon that protects against colitis and colon cancer under low-fiber dietary conditions. The Biochemical Journal, 469, 267-278.

Gurevich, V. V. & Gurevich, E. V., 2008. GPCR monomers and oligomers: it takes all kinds. Trends in Neurosciences, 31, 74-81.

Gurskaya, N. G., Verkhusha, V. V., Shcheglov, A. S., Staroverov, D. B., Chepurnykh, T. V., Fradkov, A. F., Lukyanov, S. & Lukyanov, K. A., 2006. Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. Nature Biotechnology, 24, 461-465.

Haber, R. S., Weinstein, S. P., O'Boyle, E. & Morgello, S., 1993. Tissue distribution of the human GLUT3 glucose transporter. Endocrinology, 132, 2538-2543.

Hager, K., Hazama, A., Kwon, H. M., Loo, D. D., Handler, J. S. & Wright, E. M., 1995. Kinetics and specificity of the renal Na+/myo-inositol cotransporter expressed in *Xenopus* oocytes. The Journal of Membrane Biology, 143, 103-113.

Hamilton, K. L., 2013. Robert K. Crane - Na+-glucose cotransporter to cure?. Frontiers in Physiology, 4, e53.

Hammon, J., Palanivelu, D. V., Chen, J., Patel, C. & Minor, D. L., 2009. A green fluorescent protein screen for identification of well-expressed membrane proteins from a cohort of extremophilic organisms. Protein Sciences, 18, 121-133.

Handzlik, J., Matys, A. & Kieć-Kononowicz, K., 2013. Recent advances in multi-drug resistance (MDR) efflux pump inhibitors of gram-positive bacteria *S. aureus*. Antibiotics, 2, 28-45.

Harper, S. & Speicher, D. W., 2011. Purification of proteins fused to glutathione S-tranferase. Methods in Molecular Biology, 681, 259-280.

Harris, N. J., Findlay, H. E., Sanders, M. R., Kedzierski, M., Dos Santos, Á. & Booth, P. J., 2017. Comparative stability of Major Facilitator Superfamily transport proteins. European Biophysics Journal, 46, 655-663.

Hartnett, J., Gracyalny, J., & Slater, M. R., 2006. The single step (KRX) competent cells: efficient cloning and high protein yields. Promega, 94, 27-30.

Hastings, J. W., 1983. Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems. Journal of Molecular Evolution, 19, 309-321.

Hayashi, M., Tabata, K., Yagasaki, M. & Yonetani, Y., 2010. Effect of multidrug-efflux transporter genes on dipeptide resistance and overproduction in *Escherichia coli*. FEMS Microbiology Letters, 304, 12-19.

He, Y., Wang, K. & Yan, N., 2014. The recombinant expression systems for structure determination of eukaryotic membrane proteins. Protein & Cell, 5, 658-672.

Hediger, M. A., Coady, M. J., Ikeda, T. S. & Wright, E. M., 1987. Expression cloning and cDNA sequencing of the Na+/glucose co-transporter. Nature, 330, 379-381.

Heim, R., Prasher, D. C. & Tsien, R. Y., 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proceedings of the National Academy of Sciences of the United States of America, 91, 12501-12504.

Hendrickson, W., Stoner, C. & Schleif, R., 1990. Characterization of the *Escherichia coli* araFGH and araJ promoters. Journal of Molecular Biology, 215, 497-510.

Heng, J., Zhao, Y., Liu, M., Liu, Y., Fan, J., Wang, X., Zhao, Y. & Zhang, X. C., 2015. Substrate-bound structure of the *E. coli* multidrug resistance transporter MdfA. Cell Research, 25, 1060-1073.

Heuer, H., Abdo, Z. & Smalla, K., 2008. Patchy distribution of flexible genetic elements in bacterial populations mediates robustness to environmental uncertainty. FEMS Microbiology Ecology, 65, 361-371.

Hoffmann, K. M., Million-Perez, H. R., Merkhofer, R., Nicholson, H. & Rowlett, R. S., 2015. Allosteric reversion of *Haemophilus influenzae* β-carbonic anhydrase via a proline shift. Biochemistry, 54, 598-611.

Hofmann, T., Schaefer, M., Schultz, G. & Gudermann, T., 2002. Subunit composition of mammalian transient receptor potential channels in living cells. Proceedings of the National Academy of Sciences of the United States of America, 99, 7461-7466.

Holdsworth, S. R. & Law, C. J., 2012. Functional and biochemical characterisation of the *Escherichia coli* major facilitator superfamily multidrug transporter MdtM. Biochimie, 94, 1334-1346.

Holdsworth, S. R. & Law, C. J., 2013. Multidrug resistance protein MdtM adds to the repertoire of antiporters involved in alkaline pH homeostasis in *Escherichia coli*. BMC Microbiology, 13, e113.

Hoogenkamp, M. A., Crielaard, W. & Krom, B. P., 2015. Uses and limitations of green fluorescent protein as a viability marker in *Enterococcus faecalis* An observational investigation. Journal of Microbiological Methods, 115, 57-63.

Hopfer, U., Nelson, K., Perrotto, J. & Isselbacher, K. J., 1973. Glucose transport in isolated brush border membrane from rat small intestine. The Journal of Biological Chemistry, 248, 25-32.

Horton, P. & Nakai, K., 1997. Better prediction of protein cellular localization sites with the k nearest neighbors classifier. International Conference on Intelligent Systems for Molecular Biology, 5, 147-152.

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. & Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Research, 35, 585-587.

Hosokawa, M. & Thorens, B., 2002. Glucose release from GLUT2-null hepatocytes: characterization of a major and a minor pathway. American Journal of Physiology-Endocrinology and Metabolism, 282, 794-801.

Hu, F. B., 2002. Dietary pattern analysis: a new direction in nutritional epidemiology. Current Opinion in Lipidology, 13, 3-9.

Huang, P., Altshuller, Y. M., Hou, J. C., Pessin, J. E. & Frohman, M. A., 2005. Insulin-stimulated plasma membrane fusion of Glut4 glucose transporter-containing vesicles is regulated by phospholipase D1. Molecular Biology of the Cell, 16, 2614-2623.

Huang, W. Y., Aramburu, J., Douglas, P. S. & Izumo, S., 2000. Transgenic expression of green fluorescence protein can cause dilated cardiomyopathy. Nature Medicines, 6, 482-483.

Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N., 2003. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. Science, 301, 616-620.

Hulme, E. C. & Trevethick, M. A., 2010. Ligand binding assays at equilibrium: validation and interpretation. British Journal of Pharmacology, 161, 1219-1237.

Hummel, C. S., Lu, C., Loo, D. D., Hirayama, B. A., Voss, A. A. & Wright, E. M., 2011. Glucose transport by human renal Na+/d-glucose cotransporters SGLT1 and SGLT2. American Journal of Physiology - Cell Physiology, 300, 14-21.

Huynh, K. & Partch, C. L., 2015. Analysis of protein stability and ligand interactions by thermal shift assay. Current Protocols in Protein Science, 79.28.9, 1-14

Hvorup, R. N. & Saier, M. H., 2002. Sequence similarity between the channel-forming domains of voltage-gated ion channel proteins and the C-terminal domains of secondary carriers of the major facilitator superfamily. Microbiology, 148, 3760-3762.

Hynes, T. R., Hughes, T. E. & Berlot, C. H., 2004. Cellular localization of GFP-tagged α subunits. Methods in Molecular Biology, 237, 233-246.

Iancu, C. V., Zamoon, J., Woo, S. B., Aleshin, A. & Choe, J. Y., 2013. Crystal structure of a glucose/H+ symporter and its mechanism of action. Proceedings of the National Academy of Sciences of the United States of America, 110, 17862-17867.

Ibberson, M., Uldry, M. & Thorens, B., 2000. GLUTX1, a novel mammalian glucose transporter expressed in the central nervous system and insulin-sensitive tissues. The Journal of Biological Chemistry, 275, 4607-4612.

Ihara, M., Takano, Y. & Yamashita, A., 2014. General flexible nature of the cytosolic regions of fungal transient receptor potential (TRP) channels, revealed by expression screening using GFP-fusion techniques. Protein Science, 23, 923-931.

Indiveri, C., Galluccio, M., Scalise, M. & Pochini, L., 2013. Strategies of bacterial over expression of membrane transporters relevant in human health: the successful case of the three members of OCTN subfamily. Molecular Biotechnology, 54, 724-736.

Iwamoto, H., Blakely, R. D. & De Felice, L. J., 2006. Na+, Cl-, and pH dependence of the human choline transporter (hCHT) in *Xenopus* oocytes: the proton inactivation hypothesis of hCHT in synaptic vesicles. The Journal of Neuroscience, 26, 9851-9859.

Jackson, B. J. & Kennedy, E. P., 1983. The biosynthesis of membrane-derived oligosaccharides. A membrane-bound phosphoglycerol transferase. The Journal of Biological Chemistry, 258, 2394-2398.

Jafari, R., Almqvist, H., Axelsson, H., Ignatushchenko, M., Lundbäck, T., Nordlund, P. & Martinez Molina, D., 2014. The cellular thermal shift assay for evaluating drug target interactions in cells. Nature Protocols, 9, 2100-2122.

Jang, H. H., Kim, D. H., Ahn, T. & Yun, C. H., 2010. Functional and conformational modulation of human cytochrome P450 1B1 by anionic phospholipids. Archives of Biochemistry and Biophysics, 493, 143-150.

Jayakumar, A., Kang, Y., Henderson, Y., Mitsudo, K., Liu, X., Briggs, K., Wang, M., Frederick, M. J., El-Naggar, A. K., Bebök, Z. & Clayman, G. L., 2005. Consequences of C-terminal domains and N-terminal signal peptide deletions on LEKTI secretion, stability, and subcellular distribution. Archives of Biochemistry and Biophysics, 435, 89-102.

Jensen, E. C., 2012. Use of fluorescent probes: their effect on cell biology and limitations. Anatomical Record, 295, 2031-2036.

Jeong, H., Kim, H. J. & Lee, S. J., 2015. Complete genome sequence of *Escherichia coli* strain BL21. Genome Announcements, 3, e00134-15.

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. & Duhr, S., 2011. Molecular interaction studies using microscale thermophoresis. Assay and Drug Development Technologies, 9, 342-353.

Jézégou, A., Llinares, E., Anne, C., Kieffer-Jaquinod, S., O'Regan, S., Aupetit, J., Chabli, A., Sagné, C., Debacker, C., Chadefaux-Vekemans, B., Journet, A., André, B. & Gasnier, B., 2012. Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. Proceedings of the National Academy of Sciences in the United States of America, 109, 3434-3443.

Jia, B. & Jeon, C. O., 2016. High-throughput recombinant protein expression in *Escherichia coli*: current status and future perspectives. Open Biology, 6, e160196.

Jiang, D., Zhao, Y., Wang, X., Fan, J., Heng, J., Liu, X., Feng, W., Kang, X., Huang, B., Liu, J. & Zhang, X. C., 2013. Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. Proceedings of the National Academy of Sciences of the United States of America, 110, 14664-14669.

Jiang, J., Geng, G., Yu, X., Liu, H., Gao, J., An, H., Cai, C., Li, N., Shen, D., Wu, X., Zheng, L., Mi, Y. & Yang, S., 2016. Repurposing the anti-malarial drug dihydroartemisinin suppresses metastasis of non-small-cell lung cancer via inhibiting NF-κB/GLUT1 axis. Oncotarget, 7, 87271-87283.

Jin, J., Guffanti, A. A., Beck, C. & Krulwich, T. A., 2001. Twelve-transmembrane-segment (TMS) version (ΔTMS VII-VIII) of the 14-TMS Tet(L) antibiotic resistance protein retains monovalent cation transport modes but lacks tetracycline efflux capacity. Journal of Bacteriology, 183, 2667-2671.

Johansso, L. & Lidén, G., 2006. Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* W3110 grown under carbon- and phosphate-limited conditions. Journal of Biotechnology, 126, 528-45.

Johnson, F. H., Shimomura, O., Saiga, Y., Gershman, L. C., Reynolds, G. T. & Waters, J. R.,1962. Quantum efficiency of Cypridinaluminescence, with a note on that of *Aequorea*. Journal of Cellular Physiology, 60, 85-103.

Johnson, J. H., Newgard, C. B., Milburn, J. L., Lodish, H. F. & Thorens, B., 1990. The high Km glucose transporter of islets of langerhans is functionally similar to the low affinity transporter of liver and has an identical primary sequence. The Journal of Biological Sciences, 265, 6548-6551.

Johswich, A., Kraft, B., Wuhrer, M., Berger, M., Deelder, A. M., Hokke, C. H., Gerardy-Schahn, R. & Bakker, H., 2009. Golgi targeting of *Drosophila melanogaster* β4GalNAcTB requires a DHHC protein family–related protein as a pilot. The Journal of Cell Biology, 184, 173-183.

Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H. & Johnsson, K., 2003. Directed evolution of O6-alkylguanine-dna alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo. Chemistry & Biology, 10, 313-317.

Jung, C. Y. & Rampal, A. L., 1977. Cytochalasin B binding sites and glucose transport carrier in human erythrocyte ghosts. The Journal of Biological Chemistry, 252, 5456-5463.

Kain, S. R., Adams, M., Kondepudi, A., Yang, T. T., Ward, W. W. & Kitts, P., 1995. Green fluorescent protein as a reporter of gene expression and protein localization. BioTechniques, 19, 650-655.

Kalatzis, V., Cherqui, S., Antignac, C. & Gasnier, B., 2001. Cystinosin, the protein defective in cystinosis, is a H(+)-driven lysosomal cystine transporter. The EMBO Journal, 20, 5940-5949.

Käll, L., Krogh, A. & Sonnhammer, E. L., 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics, 21, 251-257.

Kalyanaraman, B. & Zielonka, J., 2017. Green fluorescent proteins induce oxidative stress in cells: A worrisome new wrinkle in the application of the GFP reporter system to biological systems?. Redox Biology, 12, 755-757.

Kanbar, B. & Ozdemir, E., 2010. Thermal stability of carbonic anhydrase immobilized within polyurethane foam. Biotechnology Progress, 26, 1474-1480.

Kanno, A., Ozawa, T. & Umezawa, Y., 2011. Detection of protein-protein interactions in bacteria by GFP-fragment reconstitution. Methods of Molecular Biology, 705, 251-258.

Kapoor, K., Finer-Moore, J. S., Pedersen, B. P., Caboni, L., Waight, A., Hillig, R. C., Bringmann, P., Heisler, I., Müller, T., Siebeneicher, H. & Stroud, R. M., 2016. Mechanism of inhibition of human glucose transporter GLUT1 is conserved between cytochalasin B and phenylalanine amides. Proceedings of the National Academy of Sciences in the United States of America, 113, 4711-4716.

Kasahara, M. & Hinkle, P. C., 1977. Reconstitution and purification of the D-glucose transporter from human erythrocytes. The Journal of Biological Chemistry, 252, 7384-7390.

Kawe, M., Horn, U. & Plückthun, A., 2009. Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors. Microbial Cell factories, 8, e10.1186.

Kere, J., 2006. Overview of the SLC26 family and associated diseases. Novartis Foundation Symposium, 273, 2-11.

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I. & Karp, P. D., 2016. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Research, 45, 543-550.

Khavrutskii, L., Yeh, J., Timofeeva, O., Tarasov, S. G., Pritt, S., Stefanisko, K. & Tarasova, N., 2013. Protein purification-free method of binding affinity determination by microscale thermophoresis. Journal of Visualized Experiments, 78, e10.3791.

Kim, Y. G., Park, B., Ahn, J. O., Jung, J. K., Lee, H. W. & Lee, E. G., 2012. New cell line development for antibody-producing Chinese hamster ovary cells using split green fluorescent protein. BMC Biotechnology, 12:24, e1472-6750.

Klootwijk, E. D., Reichold, M., Unwin, R. J., Kleta, R., Warth, R. & Bockenhauer, D., 2015. Renal Fanconi syndrome: taking a proximal look at the nephron. Nephrology, Dialysis, Transplantation, 30, 1456-60.

Koita, K. & Rao, C. V., 2012. Identification and analysis of the putative pentose sugar efflux transporters in *Escherichia coli*. PLoS One, 7, e0043700.

Konitsiotis, A. D., Chang, S. C., Jovanović, B., Ciepla, P., Masumoto, N., Palmer, C. P., Tate, E. W., Couchman, J. R. & Magee, A. I., 2014. Attenuation of hedgehog acyltransferase-catalyzed sonic hedgehog palmitoylation causes reduced

signaling, proliferation and invasiveness of human carcinoma cells. PLoS One, 9, e0089899.

Kosmidis, C., Schindler, B. D., Jacinto, P. L., Patel, D., Bains, K., Seo, S. M. & Kaatz, G. W., 2012. Expression of multidrug resistance efflux pump genes in clinical and environmental isolates of *Staphylococcus aureus*. International Journal of Antimicrobial Agents, 40, 204-209.

Kosuri, S. & Church, G. M., 2014. Large-scale de novo DNA synthesis: technologies and applications. Nature Methods, 11, 499-507.

Nature Methods, 11, 499-507.

Kozma, D., Simon, I. & Tusnády, G. E., 2013. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. Nucleic Acids Research, 41, 524-529.

Kraegen, E. W., Sowden, J. A., Halstead, M. B., Clark, P. W., Rodnick, K. J., Chisholm, D. J. & James, D. E., 1993. Glucose transporters and in vivo glucose uptake in skeletal and cardiac muscle: fasting, insulin stimulation and immunoisolation studies of GLUT1 and GLUT4. The Biochemical Journal, 295, 287-293.

Kremling, A., Geiselmann, J., Ropers, D. & de Jong, H., 2015. Understanding carbon catabolite repression in *Escherichia coli* using quantitative models. Trends in Microbiology, 23, 99-109.

Krizsan, A., Knappe, D. & Hoffmann, R., 2015. Influence of the yjiL-mdtM gene cluster on the antibacterial activity of proline-rich antimicrobial peptides overcoming *Escherichia coli* resistance induced by the missing SbmA transporter system. Antimicrobial Agents and Chemotherapy, 59, 5992-5998.

Krulwich, T. A., Quirk, P. G. & Guffanti, A. A., 1990. Uncoupler-resistant mutants of bacteria. Microbiological Reviews, 54, 52-65.

Kumagai, A., Ando, R., Miyatake, H., Greimel, P., Kobayashi, T., Hirabayashi, Y., Shimogori, T. & Miyawaki, A., 2013. A bilirubin-inducible fluorescent protein from eel muscle. Cell, 153, 1602-1611.

Kumar, A., Beloglazova, N., Bundalovic-Torma, C., Phanse, S., Deineko, V., Gagarinova, A., Musso, G., Vlasblom, J., Lemak, S., Hooshyar, M., Minic, Z., Wagih, O., Mosca, R., Aloy, P., Golshani, A., Parkinson, J., Emili, A., Yakunin, A. F. & Babu, M., 2016. Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in *Escherichia coli*. Cell Reports, 14, 648-661.

Kumar, H., Finer-Moore, J. S., Kaback, H. R. & Stroud, R. M., 2015. Structure of LacY with an α-substituted galactoside: Connecting the binding site to the protonation site. Proceedings of the National Academy of Sciences of the United States of America, 112, 9004-9009.

Kumar, H., Kasho, V., Smirnova, I., Finer-Moore, J. S., Kaback, H. R. & Stroud, R. M., 2014. Structure of sugar-bound LacY. Proceedings of the National Academy of Sciences of the United States of America, 111, 1784-1788.

Kwon, S. K., Kim, S. K., Lee, D. H. & Kim, J. F., 2015. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. Scientific Reports, 5, e16076.

Landete, J. M., Peirotén, Á., Rodríguez, E., Margolles, A., Medina, M. & Arqués, J. L., 2014. Anaerobic green fluorescent protein as a marker of *Bifidobacterium* strains. International Journal of Food Microbiology, 175, 6-13.

Lebendiker, M. & Danieli, T., 2011. Purification of proteins fused to maltose-binding protein. Methods in Molecular Biology, 681, 281-293.

Lee, Y., Nishizawa, T., Yamashita, K., Ishitani, R. & Nureki, O., 2015. Structural basis for the facilitative diffusion mechanism by SemiSWEET transporter. Nature communications, 6, 6112.

Legrain, P., Aebersold, R., Archakov, A., Bairoch, A., Bala, K., Beretta, L., Bergeron, J., Borchers, C. H., Corthals, G. L., Costello, C. E., Deutsch, E. W., Domon, B., Hancock, W., He, F., Hochstrasser, D., Marko-Varga, G., Salekdeh, G. H., Sechi, S., Snyder, M., Srivastava, S., Uhlén, M., Wu, C. H., Yamamoto, T., Paik, Y. K. & Omenn, G. S., 2011. The human proteome project: current state and future direction. Molecular & Cellular Proteomics, 10, e009993.

Leitch, J. M. & Carruthers, A., 2009. α- and β-monosaccharide transport in human erythrocytes. American Journal of Physiology - Cell physiology, 296, 151-161.

Lenartowicz, M., Krzeptowski, W., Lipiński, P., Grzmil, P., Starzyński, R., Pierzchała, O. & Møller, L. B., 2015. Mottled mice and non-mammalian models of menkes disease. Frontiers in Molecular Neuroscience, 8, e00072

Llères, D., Swift, S. & Lamond, A. I., 2007. Detecting protein-protein interactions in vivo with FRET using multiphoton fluorescence lifetime imaging microscopy (FLIM). Current Protocols in Cytometry, 12, e12.10.

Leslie, A. G., Moody, P. C. & Shaw, W. V., 1988. Structure of chloramphenicol acetyltransferase at 1.75-A resolution. Proceedings of the National Academy of Sciences of the United States of America, 85, 4133-4137.

Lewinson, O. & Bibi, E., 2001. Evidence for simultaneous binding of dissimilar substrates by the *Escherichia coli* multidrug transporter MdfA. Biochemistry, 40, 12612-12618.

Lewinson, O., Adler, J., Poelarends, G. J., Mazurkiewicz, P., Driessen, A. J. & Bibi, E., 2003. The *Escherichia coli* multidrug transporter MdfA catalyzes both

electrogenic and electroneutral transport reactions. Proceedings of the National Academy of Sciences of the United States of America, 100, 1667-1672.
Lewinson, O., Padan, E. & Bibi, E., 2004. Alkalitolerance: a biological function for a multidrug transporter in pH homeostasis. Proceedings of the National Academy of Sciences of the United States of America, 101, 14073-14078.
Li, H., Wei, H., Wang, Y., Tang, H. & Wang, Y., 2013. Enhanced green fluorescent protein transgenic expression in vivo is not biologically inert. Journal of Proteome Research, 12, 3801-3808.
Li, Q., Manolescu, A., Ritzel, M., Yao, S., Slugoski, M., Young, J. D., Chen, X. Z. & Cheeseman, C. I., 2014. Cloning and functional characterization of the human GLUT7 isoform SLC2A7 from the small intestine. American Journal of Physiology - Gastrointestinal and Liver Physiology, 287, 236-242
Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J. & Tan, S., 2005. Comparison of affinity tags for protein purification. Protein Expression and Purification, 41, 98-105.
Lilius, G., Persson, M., Bülow, L. & Mosbach, K., 1991. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. European Journal of Biochemistry, 198, 499-504.
Lin, S. H. & Guidotti, G., 2009. Purification of membrane proteins. Methods in Enzymology, 463, 619-629.
Lippincott-Schwartz, J. & Smith, C. L., 1997. Insights into secretory and endocytic membrane traffic using green fluorescent protein chimeras. Current Opinion in Neurobiology, 7, 631-639.
Lisinski, I., Schürmann, A., Joost, H. G., Cushman, S. W. & Al-Hasani, H., 2001. Targeting of GLUT6 (formerly GLUT9) and GLUT8 in rat adipose cells. The Biochemical Journal, 358, 517-522.
Liu, B., Du, H., Rutkowski, R., Gartner, A. & Wang, X., 2012. LAAT-1 is the lysosomal lysine/arginine transporter that maintains amino acid homeostasis. Science, 337, 351-354.
Liu, J. Y., Miller, P. F., Gosink, M. & Olson, E. R., 1999. The identification of a new family of sugar efflux pumps in *Escherichia coli*. Molecular Microbiology, 31, 1845-1851.
Liu, J. Y., Miller, P. F., Willard, J. & Olson, E. R., 1999. Functional and biochemical characterization of *Escherichia coli* sugar efflux transporters. The Journal of Biological Chemistry, 274, 22977-22984.
Liu, M. L., Olson, A. L., Moye-Rowley, W. S., Buse, J. B., Bell, G. I. & Pessin, J. E., 1992. Expression and regulation of the human GLUT4/muscle-fat facilitative

glucose transporter gene in transgenic mice. The Journal of Biological Chemistry, 267, 11673-11676.
Liu, Q., Zhou, J., Daiger, S. P., Farber, D. B., Heckenlively, J. R., Smith, J. E., Sullivan, L. S., Zuo, J., Milam, A. H. & Pierce, E. A., 2002. Identification and subcellular localization of the RP1 protein in human and mouse photoreceptors. Investigative Ophthalmology & Visual Science, 43, 22-32.
Liu, X. & Zheng, X. F., 2007. Endoplasmic reticulum and Golgi localization sequences for mammalian target of rapamycin. Molecular Biology of the Cell, 18, 1073-1082.
Lo, C. A., Kays, I., Emran, F., Lin, T. J., Cvetkovska, V. & Chen, B. E., 2015. Quantification of protein levels in single living cells. Cell Reports, 13, 2634-2644.
Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linial, M., Orengo, C., Thornton, J. & Tramontano, A., 2009. Protein function annotation by homology-based inference. Genome Biology, 10, e207.
Lomovskaya, O. & Lewis, K., 1992. Emr, an *Escherichia coli* locus for multidrug resistance. Proceedings of the National Academy of Sciences of the United States of America, 89, 8938-8942.
Long, K. S., Poehlsgaard, J., Kehrenberg, C., Schwarz, S. & Vester, B., 2006. The Cfr rRNA methyltransferase confers resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A antibiotics. Antimicrobial Agents and Chemotherapy, 50, 2500-2505.
López-Barradas, A., González-Cid, T., Vázquez, N., Gavi-Maza, M., Reyes-Camacho, A., Velázquez-Villegas, L. A., Ramírez, V., Zandi-Nejad, K., Mount, D. B., Torres, N., Tovar, A. R., Romero, M. F., Gamba, G. & Plata, C., 2016. Insulin and SGK1 reduce the function of Na+/monocarboxylate transporter 1 (SMCT1/SLC5A8). American Journal of Physiology - Cell Physiology, 311, 720-734.
Löw, C., Jegerschöld, C., Kovermann, M., Moberg, P. & Nordlund, P., 2012. Optimisation of over-expression in *E. coli* and biophysical characterisation of human membrane protein synaptogyrin 1. PLoS One, 7, e0038244.
Lu, F., Li, S., Jiang, Y., Jiang, J., Fan, H., Lu, G., Deng, D., Dang, S., Zhang, X., Wang, J. & Yan, N., 2011. Structure and mechanism of the uracil transporter UraA. Nature, 472, 243-246.
Lu, Y. H., Guan, Z., Zhao, J. & Raetz, C. R., 2011. Three phosphatidylglycerol-phosphate phosphatases in the inner membrane of *Escherichia coli*. The Journal of Biological Chemistry, 286, 5506-5518.

Ludin, B., Doll, T., Meili, R., Kaech, S. & Matus, A., 1996. Application of novel vectors for GFP-tagging of proteins to study microtubule-associated proteins. Gene, 173, 107-111.

Madej, M. G., Dang, S., Yan, N. & Kaback, H. R., 2013. Evolutionary mix-and-match with MFS transporters. Proceedings of the National Academy of Sciences of the United States of America, 110, 5870-5874.

Maher, F., Vannucci, S. J. & Simpson, I. A., 1994. Glucose transporter proteins in brain. FASEB journal, 8, 1003-1011.

Mahmood, H. Y., Jamshidi, S., Sutton, J. M. & Rahman, K. M., 2016. Current advances in developing inhibitors of bacterial multidrug efflux pumps. Current Medicinal Chemistry, 23, 1062-1081.

Mao, F., Dam, P., Chou, J., Olman, V. & Xu, Y., 2009. DOOR: a Database for prOkaryotic OpeRons. Nucleic Acids Research, 37, 459-463.

Marini, G., Luchese, M. D., Argondizzo, A. P., de Góes, A. C., Galler, R., Alves, T. L., Medeiros, M. A. & Larentis, A. L., 2014. Experimental design approach in recombinant protein expression: determining medium composition and induction conditions for expression of pneumolysin from *Streptococcus pneumoniae* in *Escherichia coli* and preliminary purification process. BMC Biotechnology, 14, e1472-6750.

Martin, P. M., Dun, Y., Mysona, B., Ananth, S., Roon, P., Smith, S. B. & Ganapathy, V., 2007. Expression of the sodium-coupled monocarboxylate transporters SMCT1 (SLC5A8) and SMCT2 (SLC5A12) in retina. Investigative Opthalmology & Visual Science, 48, 3356-3363.

Martínez-Alonso, M., García-Fruitós, E., Ferrer-Miralles, N., Rinas, U. & Villaverde, A., 2010. Side effects of chaperone gene co-expression in recombinant protein production. Microbial Cell Factories, 9, e1475-2859.

Massey-Gendel, E., Zhao, A., Boulting, G., Kim, H. Y., Balamotis, M. A., Seligman, L. M., Nakamoto, R. K. & Bowie, J. U., 2009. Genetic selection system for improving recombinant membrane protein expression in *E. coli*. Protein Science, 18, 372-383.

Masumoto, N., Lanyon-Hogg, T., Rodgers, U. R., Konitsiotis, A. D., Magee, A. I. & Tate, E. W., 2015. Membrane bound O-acyltransferases and their inhibitors. Biochemical Society Transactions, 43, 246-252.

Mayer, A. L., Higgins, C. B., Heitmeier, M. R., Kraft, T. E., Qian, X., Crowley, J. R., Hyrc, K. L., Beatty, W. L., Yarasheski, K. E.., Hruz, P. W. & DeBosch, B. J., 2016. SLC2A8 (GLUT8) is a mammalian trehalose transporter required for trehalose-induced autophagy. Scientific Reports, 6, e38586.

McInerney, P., Adams, P. & Hadi, M. Z., 2014. Error rate comparison during polymerase chain reaction by DNA polymerase. Molecular Biology International, 2014, e287430.

McMillin, S. L., Schmidt, D. L., Kahn, B. B. & Witczak, C. A., 2017. GLUT4 is not necessary for overload-induced glucose uptake or hypertrophic growth in mouse skeletal muscle. Diabetes, 66, 1491-1500.

Meissner, B., Rogalski, T., Viveiros, R., Warner, A., Plastino, L., Lorch, A., Granger, L., Segalat, L. & Moerman, D. G., 2011. Determining the sub-cellular localization of proteins within *Caenorhabditis elegans* body wall muscle. PLoS One, 6, e19937

Miladi, B., Dridi, C., El Marjou, A., Boeuf, G., Bouallagui, H., Dufour, F., Di Martino, P. & Elm'selmi, A., 2013. An improved strategy for easy process monitoring and advanced purification of recombinant proteins. Molecular Biotechnology, 55, 227-235.

Miller, J. W., Beresford, S. A., Neuhouser, M. L., Cheng, T. Y., Song, X., Brown, E. C., Zheng, Y., Rodriguez, B., Green, R. & Ulrich, C. M., 2013. Homocysteine, cysteine, and risk of incident colorectal cancer in the Women's Health Initiative observational cohort. The American Journal of Clinical Nutrition, 97, 827-834.

Miroux, B. & 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, 289-298.

Mirza, O., Guan, L., Verner, G., Iwata, S. & Kaback, H. R., 2006. Structural evidence for induced fit and a mechanism for sugar/H+ symport in LacY. The EMBO Journal, 25, 1177-1183.

Mishin, A. S., Belousov, V. V., Solntsev, K. M. & Lukyanov, K. A., 2015. Novel uses of fluorescent proteins. Current Opinion in Chemical Biology, 27, 1-9

Mobasheri, A., Dobson, H., Mason, S. L., Cullingham, F., Shakibaei, M., Moley, J. F. & Moley, K. H., 2005. Expression of the GLUT1 and GLUT9 facilitative glucose transporters in embryonic chondroblasts and mature chondrocytes in ovine articular cartilage. Cell Biology International, 29, 249-260.

Mohajeri, A., Sanaei, S., Kiafar, F., Fattahi, A, Khalili, M. & Zarghami, N., 2017. The challenges of recombinant endostatin in clinical application: focus on the different expression systems and molecular bioengineering. Advanced Pharmaceutical Bulletin, 7, 21-34.

Mollwitz, B., Brunk, E., Schmitt, S., Pojer, F., Bannwarth, M., Schiltz, M., Rothlisberger, U. & Johnsson, K., 2012. Directed evolution of the suicide protein O^6-alkylguanine-DNA alkyltransferase for increased reactivity results in an alkylated protein with exceptional stability. Biochemistry, 51, 986-994.

Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D., 2014. Membrane protein structure determination - The next generation. Biochimica et Biophysica Acta, 1838, 78-87.

Mueckler, M. & Thorens, B., 2013. The SLC2 (GLUT) family of membrane transporters. Molecular Aspects of Medicine, 34, 121-138.

Mukherjee, A., Walker, J., Weyant, K. B. & Schroeder, C. M., 2013. Characterization of flavin-based fluorescent proteins: an emerging class of fluorescent reporters. PLoS One, 8, e64753.

Murer, H., Hopfer, U. & Kinne, R., 1976. Sodium/proton antiport in brush-border-membrane vesicles isolated from rat small intestine and kidney. The Biochemical Journal, 154, 597-604.

Murray, J. W., Maghlaoui, K. & Barber, J., 2007. The structure of allophycocyanin from *Thermosynechococcus elongatus* at 3.5 Å resolution. Acta Crystallographica, 63, 998-1002

Naas, T., Poirel, L. & Nordmann, P., 2008. Minor extended-spectrum β-lactamases. Clinical Microbiology and Infection, 14, 42-52.

Nagakubo, S., Nishino, K., Hirata, T. & Yamaguchi, A., 2002. The Putative response regulator BaeR stimulates multidrug resistance of *Escherichia coli* via a novel multidrug exporter system, MdtABC. Journal of Bacteriology, 184, 4161-4167.

Nagamori, S., Smirnova, I. N. & Kaback, H. R., 2004. Role of YidC in folding of polytopic membrane proteins. The Journal of Cell Biology, 165, 53-62.

Nakai, K. & Horton, P., 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends in Biochemical Sciences, 24, 34-36.

Nakamura, N., Rabouille, C., Watson, R., Nilsson, T., Hui, N., Slusarewicz, P., Kreis, T. E. & Warren, G., 1995. Characterization of a cis-Golgi matrix protein, GM130. The Journal of Cell Biology, 131, 1715-1726.

Nannenga, B. L. & Baneyx, F., 2011. Reprogramming chaperone pathways to improve membrane protein expression in *Escherichia coli*. Protein science, 20, 1411-1420.

Neckelmann, G. & Orellana, A., 1998. Metabolism of uridine 5′-diphosphate-glucose in Golgi vesicles from pea stems. Plant Physiology, 117, 1007-1014.

Negi, S., Pandey, S., Srinivasan, S. M., Mohammed, A. & Guda, C., 2015. LocSigDB: a database of protein localization signals. Database, 2015, ebav003.

Neher, E. & Sakmann, B., 1976. Single-channel currents recorded from membrane of denervated frog muscle fibres. Nature, 260, 799-802.

Nemeth, B. A., Tsang, S. W., Geske, R. S. & Haney, P. M., 2000. Golgi targeting of the GLUT1 glucose transporter in lactating mouse mammary gland. Pediatric Research, 47, 444-450.

Neumann, S. A., Linder, K. J., Muldoon, M. F., Sutton-Tyrrell, K., Kline, C., Shrader, C. J., Lawrence, E. C., Ferrell, R. E. & Manuck, S. B., 2012. Polymorphic variation in choline transporter gene (CHT1) is associated with early, subclinical measures of carotid atherosclerosis in humans. The International Journal of Cardiovascular Imaging, 28, 243-250.

Newman, J., Peat, T. S., Richard, R., Kan, L., Swanson, P. E., Affholter, J. A., Holmes, I. H., Schindler, J. F., Unkefer, C. J. & Terwilliger, T. C., 1999. Haloalkane dehalogenases: structure of a rhodococcus enzyme. Biochemistry, 38, 16105-16114

Newton, H. E., 1957. A history of luminescence from the earliest times until 1900. The American Philosophical Society, 50, 68-69.

Nishino, K. & Yamaguchi, A., 2001. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. Journal of Bacteriology, 183, 5803-5812.

Nomura, N., Verdon, G., Kang, H. J., Shimamura, T., Nomura, Y., Sonoda, Y., Hussien, S. A., Qureshi, A. A., Coincon, M., Sato, Y., Abe, H., Nakada-Nakura, Y., Hino, T., Arakawa, T., Kusano-Arai, O., Iwanari, H., Murata, T., Kobayashi, T., Hamakubo, T., Kasahara, M., Iwata, S. & Drew, D., 2015. Structure and mechanism of the mammalian fructose transporter GLUT5. Nature, 526, 397-401.

Novagen, 2005. pet system manual, 11th edition.

O neill, J. & World Health Organisation, 2014. Antimicrobial resistance: global report on surveillance 2014. WHO, 2014, ISBN: 978 92 4 156474 8.

Obergrussberger, A., Stölzle-Feix, S., Becker, N., Brüggemann, A., Fertig, N. & Möller, C., 2015. Novel screening techniques for ion channel targeting drugs. Channels, 9, 367-375.

Ogasawara, H., Ohe, S. & Ishihama, A., 2015. Role of transcription factor NimR (YeaM) in sensitivity control of *Escherichia coli* to 2-nitroimidazole. FEMS Microbiology Letters, 362, 1-8.

O'Neill, J. & World health organisation, 2016. The review on antimicrobial resistance. World Health Organisation, 2016.

O'Neill, J. & World health organisation, 2016.Tackling drug-resistance infections globally: Final report and recommendations. World Health Organisation, 2016.

Ong, W. J., Alvarez, S., Leroux, I. E., Shahid, R. S., Samma, A. A., Peshkepija, P., Morgan, A. L., Mulcahy, S. & Zimmer, M., 2011. Function and structure of GFP-like proteins in the protein data bank. Molecular BioSystems, 7, 984-992.

Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J., 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. Science, 273, 1392-1395.

Ota, S., Taimatsu, K., Yanagi, K., Namiki, T., Ohga, R., Higashijima, S. I. & Kawahara, A., 2016. Functional visualization and disruption of targeted genes using CRISPR/Cas9-mediated eGFP reporter integration in zebrafish. Scientific Reports, 6, e34991.

Ow, D. W., Jacobs, J. D. & Howell, S. H., 1987. Functional regions of the cauliflower mosaic virus 35S RNA promoter determined by use of the firefly luciferase gene as a reporter of promoter activity. Proceedings of the National Academy of Sciences of the United States of America, 84, 4870-4874.

Palmer, A. C. & Kishony, R., 2014. Opposing effects of target overexpression reveal drug mechanisms. Nature communications, 5, e4296.

Palmer, E. & Freeman, T., 2004. Investigation into the use of C- and N-terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays. Comparative and Functional Genomics, 5, 342-353.

Palmgren, M. G. & Nissen, P., 2011. P-Type ATPases. Annual Review of Biophysics, 40, 243-266.

Pandey, A., Sarker, M., Liu, X. Q. & Rainey, J. K., 2014. Small expression tags enhance bacterial expression of the first three transmembrane segments of the apelin receptor. Biochemistry and Cell Biology, 92, 269-278.

Pao, S. S., Paulsen, I. T. & Saier, M. H., 1998. Major facilitator superfamily. Microbiology and Molecular Biology Reviews, 62, 1-34.

Parikh, V., St Peters, M., Blakely, R. D. & Sarter, M., 2013. The presynaptic choline transporter imposes limits on sustained cortical acetylcholine release and attention. The Journal of Neuroscience, 33, 2326-2337.

Park, H. J., Lee, S. W. & Han, S. W., 2014. Proteomic and functional analyses of a novel porin-like protein in *Xanthomonas oryzae pv. Oryzae*. Journal of Microbiology, 52, 1030-1035.

Park, J. Y., Dus, M., Kim, S., Abu, F., Kanai, M. I., Rudy, B. & Suh, G. S. B., 2016. *Drosophila* SLC5A11 mediates hunger by regulating K+ channel activity. Current Biology, 26, 1965-1974.

Park, S. & Imlay, J. A., 2003. High levels of intracellular cysteine promote oxidative dna damage by driving the fenton reaction. Journal of Bacteriology, 185, 1942-1950.

Park, S. H., Yang, C., Opella, S. J. & Mueller, L. J., 2013. Resolution and measurement of heteronuclear dipolar couplings of a noncrystalline protein

immobilized in a biological supramolecular assembly by proton-detected MAS solid-state NMR spectroscopy. Journal of Magnetic Resonance, 237, 164-168.
Parker, J. L. & Newstead, S., 2014. Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. Nature, 507, 68-72.
Patel, V., Sun, G., Dickman, M., Khuu, P. & Teng, J. M., 2015. Treatment of keratitis-ichthyosis- deafness (KID) syndrome in children: a case report and review of the literature. Dermatologic Therapy, 28, 89-93.
Paul, S., Alegre, K. O., Holdsworth, S. R., Rice, M., Brown, J. A., McVeigh, P., Kelly, S. M. & Law, C. J., 2014. A single-component multidrug transporter of the major facilitator superfamily is part of a network that protects *Escherichia coli* from bile salt stress. Molecular Microbiology, 92, 872-884.
Paulsen, I. T., Brown, M. H. & Skurray, R. A., 1996. Proton-dependent multidrug efflux systems. Microbiological Reviews, 60, 575-608.
Paulsen, I. T., Sliwinski, M. K. & Saier, M. H., 1998. Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities. Journal of Molecular Biology, 277, 573-592.
Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S., 2006. Engineering and characterization of a superfolder green fluorescent protein. Nature Biotechnology, 24, 79-88.
Pedersen, B. P., Kumar, H., Waight, A. B., Risenmay, A. J., Roe-Zurz, Z., Chau, B. H., Schlessinger, A., Bonomi, M., Harries, W., Sali, A., Johri, A. K. & Stroud, R. M., 2013. Crystal structure of a eukaryotic phosphate transporter. Nature, 496, 533-536.
Peerce, B. E. & Wright, E. M., 1984. Sodium-induced conformational changes in the glucose transporter of intestinal brush borders. The Journal of Biological Chemistry, 259, 14105-14112.
Peterson, A. A., Fesik, S. W. & McGroarty, E. J., 1987. Decreased binding of antibiotics to lipopolysaccharides from polymyxin-resistant strains of *Escherichia coli* and *Salmonella typhimurium*. Antimicrobial Agents and Chemotherapy, 31, 230-237.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of Computational Chemistry, 25, 1605-1612.
Phillips, D., 2016. A lifetime in photochemistry; some ultrafast measurements on singlet states. Proceedings: Mathematical, Physical, and Engineering Sciences, 472, e2190.

Pina, A. S., Dias, A. M., Ustok, F. I., El Khoury, G., Fernandes, C. S., Branco, R. J., Lowe, C. R. & Roque, A. C., 2015. Mild and cost-effective green fluorescent protein purification employing small synthetic ligands. Journal of Chromatography, 1418, 83-93.

Pos, K. M., Dimroth, P. & Bott, M., 1998. The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. Journal of Bacteriology, 180, 4160-4165.

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J., 1992. Primary structure of the *Aequorea victoria* green-fluorescent protein. Gene, 111, 229-233.

Psakis, G., Nitschkowski, S., Holz, C., Kress, D., Maestre-Reyna, M., Polaczek, J., Illing, G. & Essen, L. O., 2007. Expression screening of integral membrane proteins from Helicobacter pylori 26695. Protein science, 16, 2667-2676.

Purcell, S. H., Aerni-Flessner, L. B., Willcockson, A. R., Diggs-Andrews, K. A., Fisher, S. J. & Moley, K. H., 2011. Improved insulin sensitivity by GLUT12 overexpression in mice. Diabetes, 60, 1478-1482.

Quistgaard, E. M., Löw, C., Guettou, F. & Nordlund, P., 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nature Reviews Molecular Cell Biology, 17, 123-132.

Quistgaard, E. M., Löw, C., Moberg, P., Trésaugues, L. & Nordlund, P., 2013. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. Nature Structural & Molecular Biology, 20, 766-768.

Rahman, T., Yarnall, B. & Doyle, D. A., 2017. Efflux drug transporters at the forefront of antimicrobial resistance. European biophysics journal, 46, 647-653.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. & Saier, M. H., 2012. The major facilitator superfamily (MFS) revisited. The FEBS Journal, 279, 2022-2035.

Reid, B. G. & Flynn, G. C., 1997. Chromophore formation in green fluorescent protein. Biochemistry, 36, 6786-6791.

Remington, S. J., 2011. Green fluorescent protein: a perspective. Protein Science, 20, 1509-1519.

Ren, J., Bollu, L. R., Su, F., Gao, G., Xu, L., Huang, W. C., Hung, M. C. & Weihua, Z., 2013. EGFR-SGLT1 interaction does not respond to EGFR modulators, but inhibition of SGLT1 sensitizes prostate cancer cells to EGFR tyrosine kinase inhibitors. The Prostate, 73, 1453-1461.

Ren, Q. & Paulsen, I. T., 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Computational Biology, 1, e27.

Reyes-Chin-Wo, S., Wang, Z., Yang, X., Kozik, A., Arikit, S., Song, C., Xia, L., Froenicke, L., Lavelle, D. O., Truco, M. J., Xia, R., Zhu, S., Xu, C., Xu, H., Xu, X., Cox, K., Korf, I., Meyers, B. C. & Michelmore R. W., 2017. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. Nature Communications, 8, e14953.

Rhee, J. M., Pirity, M. K., Lackan, C. S., Long, J. Z., Kondoh, G., Takeda, J. & Hadjantonakis, A. K., 2006. In vivo imaging and differential localization of lipid-modified GFP-variant fusions in embryonic stem cells and mice. Genesis, 44, 202-218.

Riss, T. L. & Moravec, R. A., 2004. Use of multiple assay end points to investigate the effects of incubation time, dose of toxin, and plating density in cell-based cytotoxicity assays. Assay and Drug Development Technologies, 2, 51-62.

Rodrigue, A., Effantin, G. & Mandrand-Berthelot, M. A., 2005. Identification of rcnA (yohM), a nickel and cobalt resistance gene in *Escherichia coli*. Journal of Bacteriology, 187, 2912-2916.

Rodriguez, A. M., Perron, B., Lacroix, L., Caillou, B., Leblanc, G., Schlumberger, M., Bidart, J. M. & Pourcher, T., 2002. Identification and characterization of a putative human iodide transporter located at the apical membrane of thyrocytes. The Journal of Clinical Endocrinology and Metabolism, 87, 3500-3503.

Rodriguez, E. A., Tran, G. N., Gross, L. A., Crisp, J. L., Shu, X., Lin, J. Y. & Tsien, R. Y., 2016. A far-red fluorescent protein evolved from a *cyanobacterial phycobiliprotein*. Nature Methods, 13, 763-769.

Rodríguez-Banqueri, A., Kowalczyk, L., Palacín, M. & Vázquez-Ibar, J. L., 2012. Assessment of membrane protein expression and stability using a split green fluorescent protein reporter. Analytical Biochemistry, 423, 7-14.

Rogers, S., Macheda, M. L., Docherty, S. E., Carty, M. D., Henderson, M. A., Soeller, W. C., Gibbs, E. M., James, D. E. & Best, J. D., 2002. Identification of a novel glucose transporter-like protein - GLUT-12. American Journal of Physiology. Endocrinology and Metabolism, 282, 733-738.

Rosano, G. L. & Ceccarelli, E. A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. Frontiers in Microbiology, 5, e172.

Rosochacki, S. J. & Matejczyk, M., 2002. Green fluorescent protein as a molecular marker in microbiology. Acta Microbiologica Polonica, 51, 205-216.

Roth, M. S., Latz, M. I., Goericke, R. & Deheyn, D. D., 2010. Green fluorescent protein regulation in the coral *Acropora yongei* during photoacclimation. The Journal of Experimental Biology, 213, 3644-3655.

Rudner, D. Z. & Losick, R., 2010. Protein subcellular localization in bacteria. Cold Spring Harbor Perspectives in Biology, 2, e307.

Rui, L., 2014. Energy metabolism in the liver. Comprehensive Physiology, 4, 177-197.

Rumsey, S. C., Kwon, O., Xu, G. W., Burant, C. F., Simpson, I. & Levine, M., 1997. Glucose transporter isoforms GLUT1 and GLUT3 transport dehydroascorbic acid. The Journal of Biological Chemistry, 272, 18982-18989.

Rupesh, K. R., Smith, A. & Boehmer, P. E., 2014. Ligand induced stabilization of the melting temperature of the HSV-1 single-strand DNA binding protein using the thermal shift assay. Biochemical and Biophysical Research Communications, 454, 604-608.

Saier, M. H., 2016. The bacterial phosphotransferase system: new frontiers 50 years after its discovery. Journal of Molecular Microbiology and Biotechnology, 25, 73-78.

Saier, M. H., Reddy, V. S., Tamang, D. G. & Västermark, A., 2014. The transporter classification database. Nucleic Acids Research, 42, 251-258.

Sakaue-Sawano, A., Kobayashi, T., Ohtawa, K. & Miyawaki, A., 2011. Drug-induced cell cycle modulation leading to cell cycle arrest, nuclear mis-segregation, or endoreplication. BMC Cell Biology, 12, e1471-2121.

Salih, A., Larkum, A., Cox, G., Kühl, M. & Hoegh-Guldberg, O., 2000. Fluorescent pigments in corals are photoprotective. Nature, 408, 850-853.

Saltiel, A. R., 2001. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes. Cell, 104, 517-529.

Sankaran, K. & Wu, H. C., 1994. Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. The Journal of Biological Chemistry, 269, 19701-19706.

Sasaki, T., Minoshima, S., Shiohama, A., Shintani, A., Shimizu, A., Asakawa, S., Kawasaki, K. & Shimizu, N., 2001. Molecular cloning of a member of the facilitative glucose transporter gene family GLUT11 (SLC2A11) and identification of transcription variants. Biochemical and Biophysical Research Communications, 289, 1218-1224.

Savitsky, P., Bray, J., Cooper, C. D., Marsden, B. D., Mahajan, P., Burgess-Brown, N. A. & Gileadi, O., 2010. High-throughput production of human proteins for crystallization: the SGC experience. Journal of Structural Biology, 172, 3-13.

Scafoglio, C., Hirayama, B. A., Kepe, V., Liu, J., Ghezzi, C., Satyamurthy, N., Moatamed, N. A., Huang, J., Koepsell, H., Barrio, J. R. & Wright, E. M., 2015. Functional expression of sodium-glucose transporters in cancer. Proceedings of the National Academy of Sciences of the United States of America, 112, 4111-4119

Schaub, B. E., Berger, B., Berger, E. G. & Rohrer, J., 2006. Transition of galactosyltransferase 1 from trans-Golgi cisterna to the trans-Golgi network is signal mediated. Molecular Biology of the Cell, 17, 5153-5162,

Schlegel, S., Genevaux, P. & de Gier, J. W., 2015. De-convoluting the genetic adaptations of *E. coli* C41 (DE3) in real-time reveals how alleviating protein production stress improves yields. Cell Reports, 10, e00175-8

Schlegel, S., Klepsch, M., Gialama, D., Wickström, D., Slotboom, D. J. & De Gier, J. W., 2010. Revolutionizing membrane protein overexpression in bacteria. Microbial Biotechnology, 3, 403-411.

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D. J. & de Gier, J. W., 2012. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). Journal of Molecular Biology, 423, 648-659.

Schmidt, P. M., Sparrow, L. G., Attwood, R. M., Xiao, X., Adams, T. E. & McKimm-Breschkin, J. L., 2012. Taking down the FLAG! How insect cell expression challenges an established tag-system. PLoS One, 7, e37779.

Scott, M. S., Calafell, S. J., Thomas, D. Y. & Hallett, M. T., 2005. Refining protein subcellular localization. PloS Computational Biology, 1, e10066.

Seddon, A. M., Curnow, P. & Booth, P. J., 2004. Membrane proteins, lipids and detergents: not just a soap opera. Biochimica et Biophysica Acta, 1666, 105-117.

Seidel, S. A., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D. & Duhr, S., 2013. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods, 59, 301-315.

Shi, Y., 2013. Common Folds and transport mechanisms of secondary active transporters. Annual Reviews of Biophysics, 42, 51-72.

Shibuya, K., Okada, M., Suzuki, S., Seino, M., Seino, S., Takeda, H. & Kitanaka, C., 2015. Targeting the facilitative glucose transporter GLUT1 inhibits the self-renewal and tumor-initiating capacity of cancer stem cells. Oncotarget, 6, 651-661.

Shimomura, O., Johnson, F. H. & Saiga, Y., 1962. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. Journal of cellular and comparative physiology, 59, 223-239.

Shulman, J. M., Chipendo, P., Chibnik, L. B., Aubin, C., Tran, D., Keenan, B. T., Kramer, P. L., Schneider, J. A., Bennett, D. A., Feany, M. B. & De Jager, P. L., 2011. Functional screening of alzheimer pathology genome-wide association signals in *Drosophila*. American Journal of Human Genetics, 88, 232-238.

Shuster, Y., Steiner-Mordoch, S., Alon Cudkowicz, N. & Schuldiner, S., 2016. A transporter interactome is essential for the acquisition of antimicrobial resistance to antibiotics. PLoS One, 11, e0152917.

Siegel, G. J., Agranoff, B. W., Albers, R. W., Fisher, S. K. & Uhle, M. D, 1999. Basic Neurochemistry: Molecular, Cellular and Medical Aspects. 6th edition. Philadelphia. Lippincott-Raven.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems in Biology, 7, e539.

Sigal, N., Lewinson, O., Wolf, S. G. & Bibi, E., 2007. *E. coli* multidrug transporter MdfA is a monomer. Biochemistry, 46, 5200-5208.

Simmons, L. C. & Yansura, D. G., 1996. Translational level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. Nature Biotechnology, 14, 629-634.

Sinicropi, A., Andruniow, T., Ferré, N., Basosi, R. & Olivucci, M., 2005. Properties of the emitting state of the green fluorescent protein resolved at the CASPT2//CASSCF/CHARMM level. Journal of American Chemical Society, 127, 11534-11535.

Skube, S. B., Chaverri, J. M. & Goodson, H. V., 2010. Effect of GFP tags on the localization of EB1 and EB1 fragments in vivo. Cytoskeleton, 67, 1-12.

Smirnova, D. V. & Ugarova, N. N., 2017. Firefly luciferase-based fusion proteins and their applications in bioanalysis. Photochemistry and Photobiology, 93, 436-447.

Smith, S. M., 2011. Strategies for the purification of membrane proteins. Methods in Molecular Biology, 681, 485-496.

Smith, V. R. & Walker, J. E., 2003. Purification and folding of recombinant bovine oxoglutarate/malate carrier by immobilized metal-ion affinity chromatography. Protein Expression and Purification, 29, 209-216.

Šnajder, M., Mihelič, M., Turk, D. & Ulrih, N. P., 2015. Codon optimisation is key for pernisine expression in *Escherichia coli*. PloS One, 10, e0123288

Soboleski, M. R., Oaks, J. & Halford, W. P., 2005. Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. FASEB Journal, 19, 440-442.

Solcan, N., Kwok, J., Fowler, P. W., Cameron, A. D., Drew, D., Iwata, S. & Newstead, S., 2012. Alternating access mechanism in the POT family of oligopeptide transporters. The EMBO Journal, 31, 3411-3421.

Sonani, R. R., Rastogi, R. P., Patel, R. & Madamwar, D., 2016. Recent advances in production, purification and applications of phycobiliproteins. World Journal of Biological Chemistry, 7, 100-109.

Spooner, P. J., Rutherford, N. G., Watts, A. & Henderson, P. J., 1994. NMR observation of substrate in the binding site of an active sugar-H+ symport protein in native membranes. Proceedings of the National Academy of Sciences of the United States of America, 91, 3877-3881.

Srinivas, S. R., Gopal, E., Zhuang, L., Itagaki, S., Martin, P. M., Fei, Y. J., Ganapathy, V. & Prasad, P. D., 2005. Cloning and functional identification of slc5a12 as a sodium-coupled low-affinity transporter for monocarboxylates (SMCT2). The Biochemical Journal, 392, 655-664.

Staub, J. M., Brand, L., Tran, M., Kong, Y. & Rogers, S. G., 2012. Bacterial glyphosate resistance conferred by overexpression of an *E. coli* membrane efflux transporter. Journal of Industrial Microbiology & Biotechnology, 39, 641-647.

Stevens, F. J. & Wu, T. T., 1976. Growth on D-lyxose of a mutant strain of *Escherichia coli* K12 using a novel isomerase and enzymes related to D-xylase metabolism. Journal of General Microbiology, 97, 257-265.

Strader, M. B., Costantino, N., Elkins, C. A., Chen, C. Y., Patel, I., Makusky, A. J., Choy, J. S., Court, D. L., Markey, S. P. & Kowalak, J. A., 2011. A proteomic and transcriptomic approach reveals new insight into β-methylthiolation of E*scherichia coli* ribosomal protein S12. Molecular & Cellular Proteomics, 10, e005199.

Stuart, C. A., Yin, D., Howell, M. E., Dykes, R. J., Laffan, J. J. & Ferrando, A. A., 2006. Hexose transporter mRNAs for GLUT4, GLUT5, and GLUT12 predominate in human muscle. American Journal of Physiology - Endocrinology and metabolism, 291, 1067-1073.

Studier, F. W. & Moffatt, B. A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology, 189, 113-130.

Sun, L., Sreedharan, S., Plummer, K. & Fisher, L. M., 1996. NorA plasmid resistance to fluoroquinolones: role of copy number and norA frameshift mutations. Antimicrobial Agents and Chemotherapy, 40, 1665-1669.

Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y. & Yan, N., 2012. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. Nature, 490, 361-366.

Sun, Y. & Vanderpool, C. K., 2011. Regulation and function of *Escherichia coli* sugar efflux transporter A (SetA) during glucose-phosphate stress. Journal of Bacteriology, 193, 143-153.

Sun, Y. L., Patel, A., Kumar, P. & Chen, Z. S., 2012. Role of ABC transporters in cancer chemotherapy. Chinese Journal of Cancer, 31, 51-57.

Sundararaj, S., Guo, A., Habibi-Nazhad, B., Rouani, M., Stothard, P., Ellison, M. & Wishart, D. S., 2004. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*. Nucleic Acids Research, 32, 293-295.

Suzuki, K., Bose, P., Leong-Quong, R. Y., Fujita, D. J. & Riabowol, K., 2010. REAP: A two minute cell fractionation method. BMC Research Notes, 3, e294.

Świtnicki, M. P., Juul, M., Madsen, T., Sørensen, K. D. & Pedersen, J. S., 2016. PINCAGE: probabilistic integration of cancer genomics data for perturbed gene identification and sample classification. Bioinformatics, 32, 1353-1365.

Tait, A. R. & Straus, S. K., 2011. Overexpression and purification of U24 from human herpesvirus type-6 in *E. coli*: unconventional use of oxidizing environments with a maltose binding protein-hexahistine dual tag to enhance membrane protein yield. Microbial Cell Factories, 10, e51.

Tal, N. & Schuldiner, S., 2009. A coordinated network of transporters with overlapping specificities provides a robust survival strategy. Proceedings of the National Academy of Sciences of the United States of America, 106, 9051-9056.

Tanaka, K. J., Song, S., Mason, K. & Pinkett, H. W., 2017. Selective substrate uptake: therole of ATP-binding cassette (ABC) importers in pathogenesis. Biochimica et Biophysica Acta, 1860, 868-877

Taniguchi, R., Kato, H. E., Font, J., Deshpande, C. N., Wada, M., Ito, K., Ishitani, R., Jormakka, M. & Nureki, O., 2015. Outward- and inward-facing structures of a putative bacterial transition-metal transporter with homology to ferroportin. Nature Communications, 6, e8545.

Tantama, M., Hung, Y. P. & Yellen, G., 2011. Imaging intracellular pH in live cells with a genetically encoded red fluorescent protein sensor. Journal of the American Chemical Society, 133, 10034-10037.

Tantama, M., Martínez-François, J. R., Mongeon, R. & Yellen, G., 2013. Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. Nature Communications, 4, e2550.

Tao, Y., Cheung, L. S., Li, S., Eom, J. S., Chen, L. Q., Xu, Y., Perry, K., Frommer, W. B. & Feng, L., 2015. Structure of a eukaryotic SWEET transporter in a homotrimeric complex. Nature, 527, 259-263.

Taranta, A., Wilmer, M. J., van den Heuvel, L. P., Bencivenga, P., Bellomo, F., Levtchenko, E. N. & Emma, F., 2010. Analysis of CTNS gene transcripts in nephropathic cystinosis. Pediatric Nephrology, 25, 1263-1267.

Tazawa, S., Yamato, T., Fujikura, H., Hiratochi, M., Itoh, F., Tomae, M., Takemura, Y., Maruyama, H., Sugiyama, T., Wakamatsu, A., Isogai, T. & Isaji, M., 2005. SLC5A9/SGLT4, a new Na+-dependent glucose transporter, is an essential transporter for mannose, 1,5-anhydro-D-glucitol, and fructose. Life Sciences, 76, 1039-1050.

Teather, R. M., Müller-Hill, B., Abrutsch, U., Aichele, G. & Overath, P., 1978. Amplification of the lactose carrier protein in *Escherichia coli* using a plasmid vector. Molecular & General Genetics, 159, 239-248.

Tegos, G. P., Haynes, M., Strouse, J. J., Khan, M. M., Bologa, C. G., Oprea, T. I. & Sklar, L. A., 2011. Microbial efflux pump inhibition: tactics and strategies. Current Pharmaceutical Design, 17, 1291-1302.

Teilum, K., Kunze, M. B., Erlendsson, S. & Kragelund, B. B., 2017. (S)Pinning down protein interactions by NMR. Protein Science, 26, 436-451.

Terakado, K., Kodan, A., Nakano, H., Kimura, Y., Ueda, K., Nakatsu, T. & Kato, H., 2010. Deleting two C-terminal alpha-helices is effective to crystallize the bacterial ABC transporter *Escherichia coli* MsbA complexed with AMP-PNP. Acta Crystallographica, 66, 319-323.

Thanassi, D. G., Cheng, L. W. & Nikaido, H., 1997. Active efflux of bile salts by Escherichia coli. Journal of Bacteriology, 179, 2512-2518.

Thangaraju, M., Ananth, S., Martin, P. M., Roon, P., Smith, S. B., Sterneck, E., Prasad, P. D. & Ganapathy, V., 2006. c/ebpdelta null mouse as a model for the double knock-out of slc5a8 and slc5a12 in kidney. The Journal Of Biological Chemistry, 281, 26769-26773.

Thorens, B., Cheng, Z. Q., Brown, D. & Lodish, H. F., 1990. Liver glucose transporter: a basolateral protein in hepatocytes and intestine and kidney cells.The American Journal of Physiology, 259, 279-285.

Thorn, K., 2017. Genetically encoded fluorescent tags. Molecular biology of the cell, 28, 848-857.

Toddo, S., Söderström, B., Palombo, I., von Heijne, G., Nørholm, M. H. & Daley, D. O., 2012. Application of split-green fluorescent protein for topology mapping membrane proteins in *Escherichia coli*. Protein Science, 21, 1571-1576.

Tsien, R. Y., 1998. The green fluorescent protein. Annual Review of Biochemistry, 67, 509-544.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A., 2015. The TOPCONS web server for combined membrane protein topology and signal peptide prediction. Nucleic Acids Research, 43, 401-407.

Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A., 2015. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Research, 43, 401-407.

Tuite, N. L., Fraser, K. R. & O'byrne, C. P., 2005. Homocysteine toxicity in *Escherichia coli* is caused by a perturbation of branched-chain amino acid biosynthesis. Journal of Bacteriology, 187, 4362-4371.

Tutar, Y. & Harman, J. G., 2006. Effect of salt bridge on transcription activation of CRP-dependent lactose operon in *Escherichia coli*. Archives of Biochemistry and Biophysics, 453, 217-223.

Uchida, Y., Ito, K., Ohtsuki, S., Kubo, Y., Suzuki, T. & Terasaki, T., 2015. Major involvement of Na+-dependent multivitamin transporter (SLC5A6/SMVT) in uptake of biotin and pantothenic acid by human brain capillary endothelial cells. Journal of Neurochemistry, 134, 97-112.

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. & Pontén, F., 2015. Proteomics. Tissue-based map of the human proteome. Science, 347, e1260419.

Uldry, M., Ibberson, M., Horisberger, J. D., Chatton, J. Y., Riederer, B. M. & Thorens, B., 2001. Identification of a mammalian H+-myo-inositol symporter expressed predominantly in the brain. The EMBO Journal, 20, 4467-4477.

Uldry, M., Ibberson, M., Hosokawa, M. & Thorens, B., 2002. GLUT2 is a high affinity glucosamine transporter. FEBS Letters, 524, 199-203.

Uldry, M., Steiner, P., Zurich, M. G., Béguin, P., Hirling, H., Dolci, W. & Thorens, B., 2004. Regulated exocytosis of an H+/myo-inositol symporter at synapses and growth cones. The EMBO Journal, 23, 531-540.

Vallianou, N. G., Geladari, E. & Kazazis, C. E., 2016. SGLT-2 inhibitors: Their pleiotropic properties. Diabetes & Metabolic Syndrome, 11, 311-315.

Vedantam, G., Guay, G. G., Austria, N. E., Doktor, S. Z. & Nichols, B. P., 1998. Characterization of mutations contributing to sulfathiazole resistance in *Escherichia coli*. Antimicrobial Agents and Chemotherapy, 42, 88-93.

Veenstra, M., Lanza, S., Hirayama, B. A., Turk, E. & Wright, E. M., 2004. Local conformational changes in the Vibrio Na+/galactose cotransporter. Biochemistry, 43, 3620-3627.

Viitanen, P., Garcia, M. L. & Kaback, H. R., 1984. Purified reconstituted lac carrier protein from *Escherichia coli* is fully functional influx. Proceedings of the National Academy of Sciences of the United States of America, 81, 1629-1633.

Viklund, H. & Elofsson, A., 2008. Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics, 24, 1662-1668.

Viklund, H., Bernsel, A., Skwark, M. & Elofsson, A., 2008. A combined predictor of signal peptides and membrane protein topology. Bioinformatics, 24, 2928-2929

Vilardaga, J. P., Krasel, C., Chauvin, S., Bambino, T., Lohse, M. J. & Nissenson, R. A., 2002. Internalization determinants of the parathyroid hormone receptor differentially regulate beta-arrestin/receptor association. The Journal of Biological Chemistry, 277, 8121-8129.

Vincentelli, R., Cimino, A., Geerlof, A., Kubo, A., Satou, Y. & Cambillau, C., 2011. High-throughput protein expression screening and purification in *Escherichia coli*. Methods, 55, 65-72.

Vitart, V., Rudan, I., Hayward, C., Gray, N. K., Floyd, J., Palmer, C. N., Knott, S. A., Kolcic, I., Polasek, O., Graessler, J., Wilson, J. F., Marinaki, A., Riches, P. L., Shu, X., Janicijevic, B., Smolej-Narancic, N., Gorgoni, B., Morgan, J., Campbell, S., Biloglav, Z., Barac-Lauc, L., Pericic, M., Klaric, I. M., Zgaga, L., Skaric-Juric, T., Wild, S. H., Richardson, W. A., Hohenstein, P., Kimber, C. H., Tenesa, A., Donnelly, L. A., Fairbanks, L. D., Aringer, M., McKeigue, P. M., Ralston, S. H., Morris, A. D., Rudan, P., Hastie, N. D., Campbell, H. & Wright, A. F., 2008. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. Nature Genetics, 40, 437-442.

Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A., Wagner, C. S., Nord, O., Nygren, P. A., van Wijk, K. J. & de Gier, J. W., 2007. Consequences of membrane protein overexpression in *Escherichia coli*. Molecular and Cellular Proteomics, 6, 1527-1550.

Wang, D. N., Safferling, M., Lemieux, M. J., Griffith, H., Chen, Y. & Li, X. D., 2003. Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. Biochimica et Biophysica Acta, 1610, 23-36.

Wang, J., Yan, C., Li, Y., Hirata, K., Yamamoto, M., Yan, N. & Hu, Q., 2014. Crystal structure of a bacterial homologue of SWEET transporters. Cell Research, 24, 1486-1489.

Wang, L., Quan, C., Liu, B., Xu, Y., Zhao, P., Xiong, W. & Fan, S., 2013. Green Fluorescent Protein (GFP)-Based Overexpression Screening and Characterization of AgrC, a Receptor Protein of Quorum Sensing in *Staphylococcus aureus*. International Journal of Molecular Sciences, 14, 18470-18487.

Wang, R., Xiang, S., Zhang, Y., Chen, Q., Zhong, Y. & Wang, S., 2014. Development of a functional antibody by using a green fluorescent protein frame as the template. Applied and Environmental Microbiology, 80, 4126-4137.

Wang, S. & Hazelrigg, T., 1994. Implications for bcd mRNA localization from spatial distribution of exu protein in *drosophila* oogenesis. Nature, 369, 400-403.

Wang, S. K., Aref, P., Hu, Y., Milkovich, R. N., Simmer, J. P., El-Khateeb, M., Daggag, H., Baqain, Z. H. & Hu, J. C., 2013. FAM20A Mutations Can Cause Enamel-Renal Syndrome (ERS). PLoS Genetics, 9, e1003302.

Ward, W. W. & Bokman, S. H., 1982. Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of renatured protein. Biochemistry, 21, 4535-4540.

Ward, W. W. & Cormier, M. J., 1979. An energy transfer protein in coelenterate bioluminescence. Characterization of the *Renilla* green-fluorescent protein. The Journal of Biological Chemistry, 254, 781-788.

Ward, W. W., Cody, C. W., Hart, R. C. & Cormier, M. J., 1980. Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green-fluorescent proteins. Photochemistry and Photobiology, 31, 611-615.

Watanabe, A., Choe, S., Chaptal, V., Rosenberg, J. M., Wright, E. M., Grabe, M. & Abramson, J., 2010. The mechanism of sodium and substrate release from the binding pocket of vSGLT. Nature, 468, 988-991.

Weber, W., Helms, V., McCammon A. J. & Langhoff, P. W., 1999. Shedding light on the dark and weakly fluorescent states of green fluorescent proteins. Proceedings of the National Academy of Sciences of the United States of America 96, 6177-6182.

Webster, G. R., Teh, A. Y. & Ma, J. K., 2017. Synthetic gene design - the rationale for codon optimization and implications for molecular pharming in plants. Biotechnology and Bioengineering, 114, 492-502.

Weinberg, A., Song, L. Y., Wilkening, C., Sevin, A., Blais, B., Louzao, R., Stein, D., Defechereux, P., Durand, D., Riedel, E., Raftery, N., Jesser, R., Brown, B., Keller, M. F., Dickover, R., McFarland, E., Fenton, T. & Pediatric ACTG Cryopreservation Working Group, 2009. Optimization and limitations of use of cryopreserved peripheral blood mononuclear cells for functional and phenotypic T-cell characterization. Clinical and Vaccine Immunology, 16, 1176-1186.

Weiner, L. M., Webb, A. K., Limbago, B., Dudeck, M. A., Patel, J., Kallen, A. J., Edwards, J. R. & Sievert, D. M., 2016. Antimicrobial-resistant pathogens associated with healthcare-associated infections: summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. Infection Control and Hospital Epidemiology, 37, 1288-1301.

Weyand, S., Shimamura, T., Yajima, S., Suzuki, S., Mirza, O., Krusong, K., Carpenter, E. P., Rutherford, N. G., Hadden, J. M., O'Reilly, J., Ma, P., Saidijam, M., Patching, S. G., Hope, R. J., Norbertczak, H. T., Roach, P. C., Iwata, S., Henderson, P. J. & Cameron, A. D., 2008. Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. Science, 322, 709-713.

White, J. A., Hart, R. J. & Fry, J. C., 1986. An evaluation of the Waters Pico-Tag system for the amino-acid analysis of food materials. The Journal of Automatic Chemistry, 8, 170-177.

White, S. H. & von Heijne, G., 2008. How translocons select transmembrane helices. Annual Reviews of Biophysics, 37, 23-42.

Widmer, M., Uldry, M. & Thorens, B., 2005. GLUT8 subcellular localization and absence of translocation to the plasma membrane in PC12 cells and hippocampal neurons. Endocrinology, 146, 4727-4736.

Wiedenmann, J., Vallone, B., Renzi, F., Nienhaus, K., Ivanchenko, S., Röcker, C. & Nienhaus, G. U., 2005. Red fluorescent protein eqFP611 and its genetically engineered dimeric variants. Journal of Biomedical Optics, 10, e14003.

Wiemann, S., Weil, B., Wellenreuther, R., Gassenhuber, J., Glassl, S., Ansorge, W., Böcher, M., Blöcker, H., Bauersachs, S., Blum, H., Lauber, J., Düsterhöft, A., Beyer, A., Köhrer, K., Strack, N., Mewes, H. W., Ottenwälder, B., Obermaier, B., Tampe, J., Heubner, D., Wambutt, R., Korn, B., Klein, M. & Poustka, A., 2001. Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. Genome Research, 11, 422-435.

Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S., 2010. Protein-binding assays in biological liquids using microscale thermophoresis. Nature Communications, 1, e100

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F., 1999. Protein identification and analysis tools on the ExPASy server. Methods in Molecular Biology, 112, 531-552.

Willaert, A., Khatri, S., Callewaert, B. L., Coucke, P. J., Crosby, S. D., Lee, J. G., Davis, E. C., Shiva, S., Tsang, M., De Paepe, A. & Urban, Z., 2012. GLUT10 is required for the development of the cardiovascular system and the notochord and connects mitochondrial function to TGFβ signalling. Human Molecular Genetics, 21, 1248-1259.

Wisedchaisri, G., Park, M. S., Iadanza, M. G., Zheng, H. & Gonen, T., 2014. Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. Nature Communications, 5, e4521

Wright, E. M., Loo, D. D., Hirayama, B. A. & Turk, E., 2004. Surprising versatility of Na+-glucose cotransporters: SLC5. Physiology, 19, 370-376.

Wright, E. M., Martín, M. G. & Turk, E., 2003. Intestinal absorption in health and disease-sugars. Best Practice & Research - Clinical Gastroenterology, 17, 943-956.

Wu, X. & Freeze, H. H., 2002. GLUT14, a duplicon of GLUT3, is specifically expressed in testis as alternative splice forms. Genomics, 80, 553-557.

Xu, Y., Chen, B., Chao, H. & Zhou, N. Y., 2013. mhpT encodes an active transporter involved in 3-(3-hydroxyphenyl)propionate catabolism by *Escherichia coli* K-12. Applied and Environmental Microbiology, 79, 6362-6368.

Xu, Y., Tao, Y., Cheung, L. S., Fan, C., Chen, L. Q., Xu, S., Perry, K., Frommer, W. B. & Feng, L., 2014. Structures of bacterial homologues of SWEET transporters in two distinct conformations. Nature, 515, 448-452.

Xuan, Y. H., Hu, Y. B., Chen, L. Q., Sosso, D., Ducat, D. C., Hou, B. H. & Frommer, W. B., 2013. Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. Proceedings of the National Academy of Sciences of the United States of America, 110, 3685-3695.

Yamamoto, T., Seino, Y., Fukumoto, H., Koh, G., Yano, H., Inagaki, N., Yamada, Y., Inoue, K., Manabe, T. & Imura, H., 1990. Over-expression of facilitative glucose transporter genes in human cancer. Biochemical and Biophysical Research Communications, 170, 223-230.

Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E., 2005. Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature, 437, 215-223.

Yan, H., Huang, W., Yan, C., Gong, X., Jiang, S., Zhao, Y., Wang, J. & Shi, Y., 2013. Structure and mechanism of a nitrate transporter. Cell Reports, 28, 716-723.

Yang, F., Moss, L. G. & Phillips, G. N., 1996. The molecular structure of green fluorescent protein. Nature Biotechnology, 14, 1246-1251.

Yang, Z., Kirton, H. M., MacDougall, D. A., Boyle, J. P., Deuchars, J., Frater, B., Ponnambalam, S., Hardy, M. E., White, E., Calaghan, S. C., Peers, C. & Steele, D. S., 2015. The Golgi apparatus is a functionally distinct Ca2+ store regulated by PKA and Epac branches of the β1-adrenergic signaling pathway. Science Signaling, 8, e398.

Yao, C., Pan, Y., Li, Y., Xu, X., Lin, Y., Wang, W. & Wang, S., 2015. Effect of sodium/iodide symporter (NIS)-mediated radioiodine therapy on estrogen receptor-negative breast cancer. Oncology Reports, 34, 59-66.

Yeh, W. L., Lin, C. J. & Fu, W. M., 2008. Enhancement of glucose transporter expression of brain endothelial cells by vascular endothelial growth factor derived from glioma exposed to hypoxia. Molecular Pharmacology, 73, 170-177.

Yin, Y., He, X., Szewczyk, P., Nguyen, T. & Chang, G., 2006. Structure of the multidrug transporter EmrD from *Escherichia coli*. Science, 312, 741-744.

Yu, D., Gustafson, W. C., Han, C., Lafaye, C., Noirclerc-Savoye, M., Ge, W. P., Thayer, D. A., Huang, H., Kornberg, T. B., Royant, A., Jan, L. Y., Jan, Y. N., Weiss, W. A. & Shu, X., 2014. An improved monomeric infrared fluorescent protein for neuronal and tumor brain imaging. Nature Communications, 5, e3626.

Yu, S. & Ding, W. G., 1998. The 45 kDa form of glucose transporter 1 (GLUT1) is localized in oligodendrocyte and astrocyte but not in microglia in the rat brain. Brain Research, 797, 65-72.

Yuan, M. & Wang, S., 2013. Rice MtN3/Saliva/SWEET family genes and their homologs in cellular organisms. Molecular Plant, 6, 665-674.

Zehnpfennig, B., Wiriyasermkul, P., Carlson, D. A. & Quick, M., 2015. Interaction of α-lipoic acid with the human Na+/multivitamin transporter (hSMVT). The Journal of Biological Chemistry, 290, 16372-16382.

Zeng, W., Lee, M. G., Yan, M., Diaz, J., Benjamin, I., Marino, C. R., Kopito, R., Freedman, S., Cotton, C., Muallem, S. & Thomas, P., 1997. Immuno and functional characterization of CFTR in submandibular and pancreatic acinar and duct cells. The American Journal of Physiology, 273, 442-455.

Zhang, S., Yang, Q., Ren, M., Qiao, S., He, P., Li, D. & Zeng, X., 2016. Effects of isoleucine on glucose uptake through the enhancement of muscular membrane concentrations of GLUT1 and GLUT4 and intestinal membrane concentrations of Na+/glucose co-transporter 1 (SGLT-1) and GLUT2. The British Journal of Nutrition, 116, 593-602.

Zhao, Y., Mao, G., Liu, M., Zhang, L., Wang, X. & Zhang, X. C., 2014. Crystal structure of the *E. coli* peptide transporter YbgH. Structure, 22, 1152-1160.
Zhong, Y., Yu, H., Wang, X., Lu, Y. & Wang, T., 2011. Towards a novel efficient T-DNA-based mutagenesis and screening system using green fluorescent. Molecular Biology Reports, 38, 4145-4151.
Zhou, Y., Waanders, L. F., Holmseth, S., Guo, C., Berger, U. V., Li, Y., Lehre, A. C., Lehre, K. P. & Danbolt, N. C., 2014. Proteome analysis and conditional deletion of the EAAT2 glutamate transporter provide evidence against a role of EAAT2 in pancreatic insulin secretion in mice. The Journal of Biological Chemistry, 289, 1329-1344.
Zhu, L. Q., Bao, Z. K., Hu, W. W., Lin, J., Yang, Q. & Yu, Q. H., 2015. Cloning and functional analysis of goat SWEET1. Genetics and molecular research, 14, 17124-17133.
Zhu, Y. & Lin, E. C., 1986. An evolvant of *Escherichia coli* that employs the L-fucose pathway also for growth on L-galactose and D-arabinose. Journal of Molecular Evolution, 23, 259-266.
Zimmer, M., 2002. Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. Chemical Reviews, 102, 759-781.