Faculty of Science & Technology

Assessment of the proton translocation domain of the glucuronide transporter (GusB) -

The construction of XylE-GusB chimeras

A dissertation submitted as part of the requirement for the BSc Biological Sciences

Lydia Kristen Doherty

June 2018

Abstract

The glucuronide transporter, GusB of *Escherichia coli* is critical for its survival within the gastrointestinal tract of vertebrates. A wide range of β-D-glucuronides, the major detoxification products of glucuronidation, are scavenged by *E. coli* as a carbon source. The aglycone moieties released from *E. coli* enter hepatic circulation, which can impact human health. Within the major facilitator protein superfamily, GusB is predicted to share 12-transmembrane helices (TM) and has been proven to utilise proton (H^+) motive force for transport (also known as a H^+-linked symporter). However, little is understood about how GusB transports a wide range of glucuronides with different aglycone moieties. Understanding its mechanism of molecular recognition to transport various glucuronides is of interest in this research. The first 6-TM helices of XylE and MelB transporters have been defined by three-dimensional structural studies as the site of H^+ translocation. Since GusB belongs to the same family, it is logical to assume that its first 6-TM helices could also be responsible for H^+ recognition. Therefore, swapping the first 6-TM helices of the XylE protein with that of GusB has been conducted in the hope that GusB is still functional. This research created five variations of XylE-GusB chimeras, joined at different locations. In doing so, relevant DNA fragments were amplified by PCR from genomic DNA and a *gusB* containing plasmid, respectively, for their fusion. The chimera mutants were constructed by molecular cloning and *E. coli* transformation, and identified by glucuronide transport assays. 71 of 563 mutants have been extracted and characterised so far by restriction mapping and DNA sequencing. Of these, one clone contained the mutant *xylE-gusB* chimera, whereas the majority were vector background. However, this mutant did not show transport activity for X-Gluc. Further characterisation of this clone is needed. In addition, other mutant *xylE-gusB* chimera containing transformants need to be identified and characterised in the future.

Acknowledgements

I would like to acknowledge Dr Wei-Jun Liang for all of his support, guidance and freedom throughout this project. I would like to thank my fellow laboratory workers Qi, Christine, Camilla and Stephen for passing on their knowledge, and for their patience, support and charisma within the lab. I would like to thank my family for their encouragement throughout this process. Finally, I would like to thank my partner, James, for his continuous support, time, and invaluable discussions and suggestions whilst writing this dissertation.

Word count: 11,994 (see Appendix 1)

Contents

List of Figures

List of Tables

Abbreviations

XylE - D-xylose proton symporter

GusA - β-glucuronidase

GusB - Glucuronide transporter

MFS - Major facilitator superfamily

TM - Transmembrane

PCR - Polymerase chain reaction

LacY - Lactose permease

CAP - Catabolite activator protein

cAMP - Cyclic adenosine monophosphate

IPTG - Isopropyl β-D-1-thiogalactopyranoside

X-Gluc - 5-bromo-4-chloro-3-indolyl glucuronide

UDP-GlcA - Uridine 5'-diphospho-glucuronic acid

UGT - UDP-glucuronosyltransferase

OAT - Organic anion transport

MATE - Multidrug and toxic compound extrusion

Lac operon - Lactose operon

Gus operon - Glucuronide operon

H^+ - Proton

N-domain - Amino domain

C-domain - Carboxyl domain

GFP - Green fluorescent protein

MCS - Multiple cloning site

1. Introduction

Escherichia coli (*E. coli*) reside symbiotically within the gastrointestinal tract of all known vertebrates. Understanding their behaviour in the gastrointestinal system is important to human health. Studies have demonstrated the importance of commensal gut flora in protection against epithelial cell injury (Rakoff-Nahoum et al. 2004), regulation of energy harvesting and fat storage (Bäckhed et al. 2004), and regulation of intestinal angiogenesis (Stappenbeck et al. 2002). However, the large intestine is a nutrient deficient environment since a high level of absorption occurs within the small intestine (Wallace et al. 2010). Thus, *E. coli* has evolved the ability to scavenge carbon sources other than glucose or lactose. One such example is glucuronides, which are transported into *E. coli* by the membrane-bound glucuronide transporter, GusB (Liang 1992; Liang et al. 2005). Once inside the cell, they are utilised as a carbon source (Wilson et al. 1992; Roberts et al. 2002). The abundance of glucuronides in the human gastrointestinal tract and the lack of other carbon sources (Wallace et al. 2010) means glucuronide scavenging is a vital mechanism for *E. coli* survival.

1.1 Glucuronidation

Glucuronides are mainly present in the gut and urinary tract and are synthesised by glucuronidation, a major detoxification pathway in mammals and vertebrates. This phase II biotransformation reaction conjugates endogenous compounds and xenobiotics with glucuronic acid to yield glucuronide molecules (Dutton and Storey 1954; Storey and Dutton 1955; Sun et al. 2015). This increases their hydrophilicity and promotes their excretion via bile or urine. Examples of endogenous toxins include bilirubin, steroid hormones and fatty acids (Schmid 1956; Levvy 1956; Billing et al. 1957; Lester and Klein 1966; Tephly and Burchell 1990; Bélanger et al. 1998; Jude et al. 2001; Sun et al. 2015; Kallionpää et al. 2015). Examples of xenobiotics include paracetamol (Nelson and Morioka 1963),

catechol-O-methyltransferase inhibitors (Lautala et al. 2000), and environmental carcinogens such as 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (Hoffmann et al. 1979; Ren et al. 2000). The target molecules are conjugated with glucuronic acid, where the substrate uridine 5'-diphospho-glucuronic acid (UDP-GlcA) acts as a glycosyl donor (Dutton and Storey 1953; Storey and Dutton 1954). UDP-GlcA is synthesised from the oxidation of a UDP-glucose molecule (Dutton and Storey 1953; Strominger et al. 1954), which is catalysed by UDP-glucose 6-dehydrogenase (EC 1.1.1.22) (Kalckar et al. 1956; Egger et al. 2010). Since glucose is vital for normal brain function and mammalian survival, it can be readily obtained from the diet or glycogen stores. Its physiological prevalence dictates ease of UDP-GlcA synthesis and infers the importance of the glucuronidation pathway.

1.1.1 Human UDP-glucuronosyltransferases - Synthesis of glucuronide molecules

UDP-glucuronosyltransferase (UGT) enzymes (EC 2.4.1.17) are membrane-bound proteins which reside predominantly within the lumen of the endoplasmic reticulum (Bossuyt and Blanckaert 1997; Radominska-Pandya et al. 1999) and are expressed in many tissues. Many isomers of UGTs exist and are encoded by a superfamily, comprising five gene families and six gene subfamilies (Tephly and Burchell 1990; Meech and Mackenzie 1997; Guillemette 2003; Rowland et al. 2013; Kallionpää et al. 2015; Sun et al. 2015). UGTs are most abundant within the hepatic tissue, however, they are also present in extrahepatic tissues such as the kidneys (Parquet et al. 1988), small intestine (Pacifici et al. 1986; Peters et al. 1991; Kallionpää et al. 2015) and lungs (Ren et al. 2000; Zheng et al. 2002).

UGT enzymes catalyse the conjugation of glucuronic acid with exogenous and endogenous substrates bearing suitable functional groups. Since there is wide structural diversity for substrate specificity, many reagents of different functional groups (such as thiol, amine, hydroxyl, carboxyl or carbonyl groups) undergo

glucuronidation (Levvy 1956; Axelrod et al. 1958; Williams 1963; Dutton and Illing 1972; Tukey and Strassburg 2000; Kaivosaari et al. 2011). During conjugation, UDP-GlcA acts as a glycosyl donor. A nucleophilic substitution reaction occurs between the glucuronosyl group of UDP-GlcA and the functional group of the substrate, forming a covalent linkage (Miners and Mackenzie 1991; Tukey and Strassburg 2000; Rowland et al. 2013). The net products from this are β-D-glucuronide, uridine 5'-diphosphate, [and water].

UGT isoenzymes can exhibit distinct and overlapping substrate specificities (Tephly and Burchell 1990; Lautala et al. 2000; Zheng et al. 2002; Sun et al. 2015; Kallionpää et al. 2015). Many UGTs are capable of conjugating multiple substrates, whereas some UGT isoenzymes are known to selectively act on substrates, such as the human bilirubin UGT (Carbone and Grodsky 1957; Black and Billing 1969; Black et al. 1970; Tephly and Burchell 1990; Zheng et al. 2002; Kallionpää et al. 2015). Other UGT isoforms with broader substrate specificities can be competitively inhibited by conflicting substrates (Arias et al. 1964; Sun et al. 2015), and exposure to certain chemicals can induce UGT activity (Burchell and Coughtrie 1989; Magdalou et al. 1993). Further, UGT polymorphisms can alter enzyme function and lead to the onset of disease (Wennberg et al. 2006; Fretzayas et al. 2012; Kallionpää et al. 2015). Population differences in UGT polymorphisms can also have clinical implications (Lin et al. 2017; Alkharfy et al. 2017).

Synthesised β-D-glucuronides are transported throughout the body via a number of membrane-bound proteins. These proteins function either through ATP-dependent transport, such as the ATP-dependent multidrug resistance proteins (Loe et al. 1996; König et al. 1999; Wu et al. 2012; Mao and Unadkat 2015), or ion-driven secondary active transport, such as organic anion transporter (OAT) proteins and the multidrug and toxic compound extrusion (MATE) proteins (Kuroda and Tsuchiya 2009; Roth et al. 2012; Nigam et al. 2015). Once the β-D-glucuronides have entered or been synthesised within the hepatic tissues,

they can be excreted into the intestines where *E. coli* utilise them as a carbon source.

1.2 The *gus* operon of *E. coli*

1.2.1 The *gus* operon

The *gus* operon enables *E. coli* to scavenge and process glucuronides. In prokaryotes, genomic DNA is commonly organised into operons. Operons are clusters of related genes controlled by a single promoter which allows them to be co-transcribed. The *gus* operon consists of a regulatory gene (*gusR*), a promoter region, an operator region, and three structural genes (*gusA*, *gusB* and *gusC*). The *gusR* gene encodes the regulatory protein, GusR. The structural genes, located downstream of the operator, encode the proteins responsible for glucuronide absorption and metabolism (Wilson et al. 1992). G*usA* is located at 36.5 minutes on the *E. coli* genome map (Blattner et al. 1997; National Library of Medicine (US), National Center for Biotechnology Information (NCBI) 2017). It encodes the intracellular hydrolase, β-glucuronidase (Blanco et al. 1982; Jefferson et al. 1986). This enzyme cleaves β-D-glucuronides within the cytoplasm of *E. coli* to D-glucuronic acid (glycone) and an aglycone. G*usB* resides downstream from *gusA* and encodes the membrane-bound glucuronide-specific transporter, GusB (Liang 1992; Liang et al. 2005). GusB transports β-D-glucuronide molecules into the *E. coli* cell via secondary active transport. *GusC* resides downstream from *gusB* and encodes the outer membrane-associated protein, GusC, which was demonstrated to enhance β-D-glucuronide uptake (Liang et al. 2005). Although limited information is currently known about *gus* operon regulation, information from other well-studied operons can provide a platform for research.

1.2.2 Regulation of the *gus* operon - Utilising the lactose (*lac*) operon

The *lac* operon has been extensively studied (Jacob et al. 1960; Beckwith 1967; Davies and Jacob 1968; Harwood and Peterkofsky 1975; Kolkhof 1992; Lawson et al. 2004; Popovych et al. 2009; Marbach and Bettenbrock 2012; Leonard et al. 2015; Bidart et al. 2018), and can act as a platform for understanding the *gus* operon. The structural genes of the *lac* operon and the *gus* operon share the common function of sugar acquisition and metabolism (Jacob et al. 1960; Beckwith 1967; Blanco et al. 1982; Liang et al. 2005). Therefore, it is logical to consider that they might be regulated in similar ways.

The *lac* operon is negatively regulated by the binding of the regulatory protein, encoded by *lacI* (Jacob et al. 1960). Similarly, GusR negatively regulates *gusA* expression (Novel and Novel 1976a; Novel and Novel 1976b; Blanco et al. 1985). Another regulatory protein, encoded by the *uxuR* gene from a distinct operon, has also been demonstrated to weakly bind to the *gus* operator region (Novel and Novel 1976a; Novel and Novel 1976b). Since *gusB* and *gusC* reside downstream from *gusA* and are governed by the same promoter (Liang et al. 2005), the expression of *gusB* and *gusC* are also likely co-regulated by these repressors in a similar fashion to the *lac* repressor.

In an environment where lactose is present, allolactose binds and induces a conformational change in the *lac* repressor (Monad et al. 1963; Monad et al. 1965). This initiates its release and allows *lac* operon expression. Similarly, in the *gus* operon the substrates β-D-glucuronide and mannonic amide antagonise GusR and UxuR, respectively (Novel et al. 1974; Novel and Novel 1976b). Unlike the *lac* operon, the *gus* operon is governed by two separate repressor proteins. This may be attributed to the increased importance of the *gus* operon because only β-D-glucuronides are readily available within the gastrointestinal tract (Wallace et al. 2010). This cross-regulation of repressor proteins involved in the same catabolic pathway allows for the presence of a single inducer substrate to facilitate expression of multiple operons, which improves the efficiency of substrate

utilisation by the hexuronide-hexuronate pathway (Mandrand-Berthelot et al. 2004).

The *lac* operon is regulated through catabolite repression (Mandelstam 1960; Mandelstam 1961; Mandelstam 1962; Nakada and Magasanik 1964; Loomis and Magasanik 1965; Loomis and Magasanik 1967). The rate is dependent upon the binding of a cyclic adenosine monophosphate (cAMP)-dependent catabolite activator protein (CAP) with a sequence upstream of the promoter. When cAMP is available, it activates CAP which interacts with RNA polymerase to drive gene expression (Zubay et al. 1970; Kolb et al. 1993; Lawson et al. 2004; Popovych et al. 2009). However, when glucose is present cAMP levels are low (Makman and Sutherland 1965; Harwood and Peterkofsky 1975) and CAP cannot be activated (Pastan and Perlman 1970). Catabolite repression is also suspected to regulate the *gus* operon: *E. coli* grown in glucose medium was observed to synthesise significantly less β-glucuronidase (Blanco et al. 1985). The addition of cAMP to these cells removed this repression (Blanco et al. 1985). In addition, a CAP binding site was identified upstream of *gusA* (Jefferson et al. 1986). The exact mechanism of repression is not fully understood in the *gus* operon, however, it is logical to assume that it is similar to that of the *lac* operon.

Further, the *lac* operon is regulated by inducer exclusion (Winkler and Wilson 1967; Nelson et al. 1983). The presence of glucose inactivates lactose permease, which prevents the transport of lactose into the cell, thus the *lac* repressor cannot be inactivated (Nelson et al. 1983; Hogema et al. 1999). Inducer exclusion has not yet been discovered to regulate the *gus* operon, however, many operons are known to be regulated in this way (Nelson et al. 1983; Titgemeyer et al. 1994). Thus, the intracellular presence of glucose may also inhibit the transport of β-D-glucuronides into the cell.

These findings, along with inferences made from the *lac* operon, demonstrate that the structural genes of the *gus* operon undergo tight positive and negative regulation.

1.3 The β-glucuronidase enzyme, GusA

Acquired β-D-glucuronides, which are biosynthesised during glucuronidation (Section 1.1), are hydrolysed by *E. coli* in order to survive. The intracellular β-glucuronidase enzyme, GusA (EC 3.2.1.31), catalyses the cleavage of β-D-glucuronides to D-glucuronic acid (glycone) and an aglycone (Doyle et al. 1955; Blanco et al. 1982; Wilson et al. 1992). The glycone is metabolised by *E. coli* for energy, whereas the aglycone moiety is 'recycled' back into the host's gastrointestinal tract in a process known as enterohepatic circulation (Figure 1.1) (Wilson et al. 1992; Roberts et al. 2002). During enterohepatic circulation the aglycone moieties are reabsorbed into the bloodstream. This process re-circulates harmful aglycones and has been suggested to be a prime factor in the etiology of colon cancers (Takada et al. 1982; Kim and Jin 2001; Humblot et al. 2007; Arthur and Jobin 2011). However, carefully designed drugs could exploit this process to increase their half-lives, decreasing effective doses. Overall, GusA is responsible for cleaving β-D-glucuronides and thus is vital for *E. coli* energy acquisition.

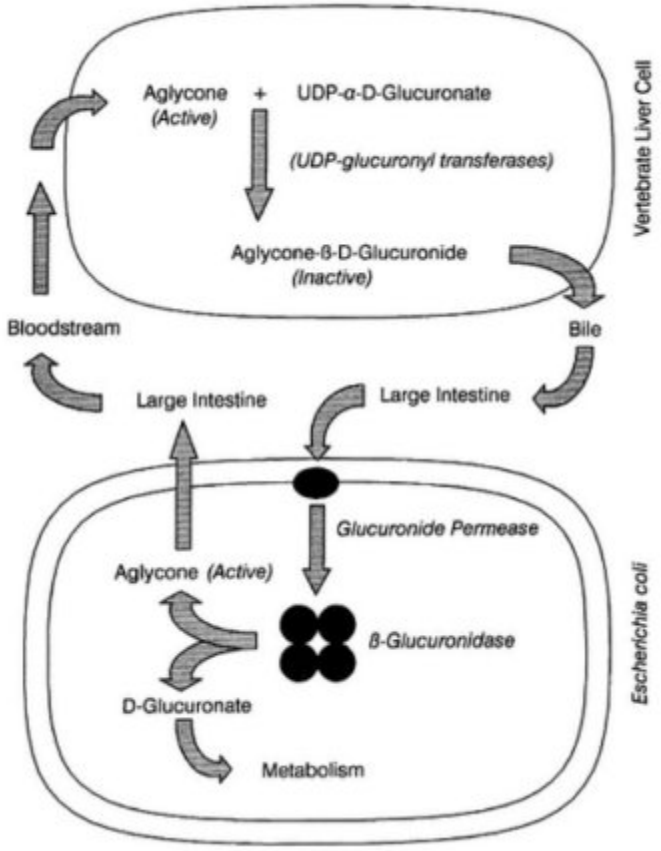

Figure 1.1: Enterohepatic circulation between vertebrates and *E. coli* (Wilson et al. 1992 p.9). β-D-glucuronides are biosynthesised within the liver of vertebrates. β-D-glucuronides are excreted within the bile to the large intestine where *E. coli* reside. The *E. coli* glucuronide transporter, GusB recognises and transports β-D-Glucuronides into the cell. β-Glucuronidase hydrolyses the β-D-glucuronide to D-glucuronic acid and an aglycone moiety. The D-glucuronic acid part is metabolised by *E. coli* and the aglycone is 'recycled' back into the gastrointestinal tract and reabsorbed into the bloodstream. The aglycone is then re-conjugated within the liver.

GusA has also been developed as a reporter gene system (Jefferson et al. 1986), which is especially useful in higher plants since they do not naturally possess β-glucuronidases (Jefferson et al. 1987; Peng et al. 1995; Basu et al. 2004; Conte and Walker 2012; Zhao and Sun 2015). The successful expression of the *gusA*-gene fusion in transgenic plants can be detected using chromogenic glucuronides such as 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc). Here, expression of the *gusA*-gene fusion results in the cleavage of X-Gluc, consequently producing a visible blue precipitate. Due to extensive analyses of the

human β-glucuronidase enzyme, many substrates are commercially available for enzymatic assay (Wilson et al. 1992; Basu et al. 2004; Zhao and Sun 2015). Further, GusA is extremely versatile and tolerant to N- terminus fusions (Jefferson et al. 1987; Jefferson et al. 1989). Understanding the function of GusA facilitated its development as a reporter gene system.

1.4 The glucuronide transporter, GusB

The glucuronide transporter, GusB, is an integral membrane-bound protein responsible for β-D-glucuronide transport into *E. coli*. GusB resides within the inner membrane and consists of 457 amino acids (Figure 1.2) (Liang 1992; Liang et al. 2005). The predicted secondary structure of GusB is 12-TM α-helices, in which the C- and N- termini reside within the cytosol (Liang 1992; Ishii 2010). GusB utilises H^+ motive force to transport β-D-glucuronides across the membrane (Liang 1992; Liang et al. 2005). Despite the large molecular variability seen between the aglycone part of β-D-glucuronides, GusB is able to recognise a wide range of substrates. However, the sites of molecular recognition and H^+ coupling within the transporter are currently unknown. Structural determination studies of GusB suggest that it exists in a pentameric state (Ishii 2010; Ishii 2013). However, further analyses at higher resolutions are required to validate this. Understanding the structure of GusB would provide insight into the sites of molecular recognition.

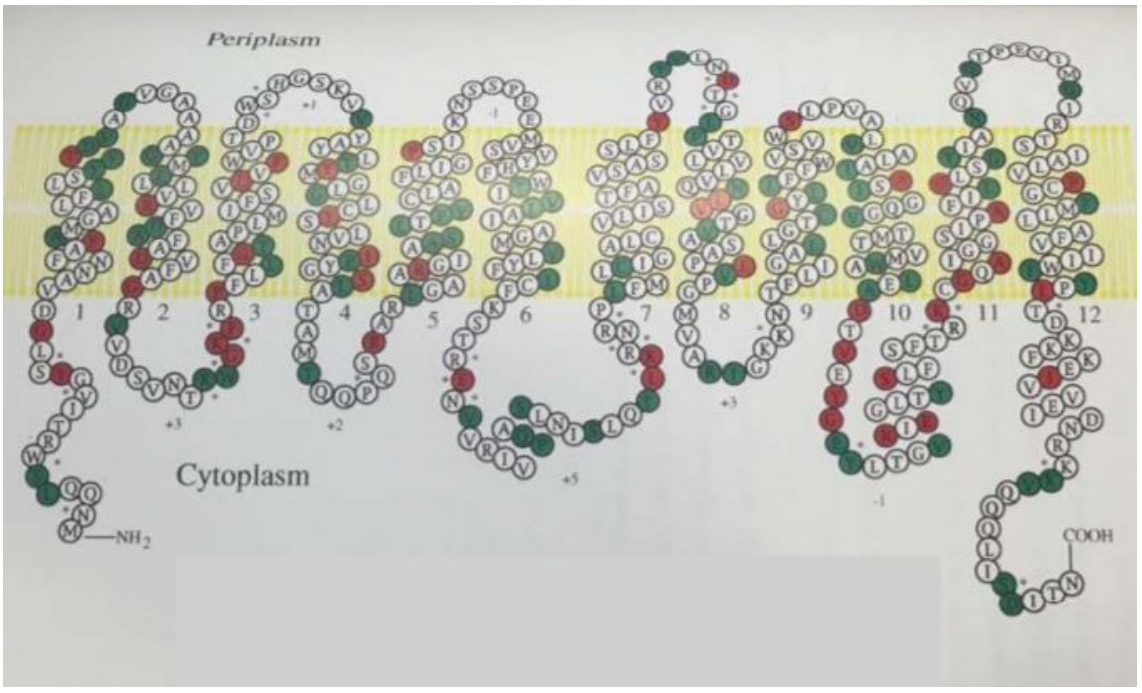

Figure 1.2: Model of the *E. coli* glucuronide transporter, GusB. The predicted secondary structure topology is illustrated by 12-transmembrane helices and the intracellular N- and C- termini. The amino acid residues are labelled. The green residues are conserved between GusB and the *E. coli* MelB protein. The red residues are conserved amongst GusB, the *Streptococcus thermophilus* LacS protein, and *E. coli* MelB (Liang 1992).

1.5 The major facilitator superfamily

GusB belongs to the major facilitator superfamily (MFS), the largest known secondary transporter family which is found ubiquitously across all three kingdoms (Marger and Saier 1993; Pao et al. 1998; Yan 2015). MFS proteins are crucial in many physiological processes and are responsible for recognising and transporting a variety of substrates including amino acids, sugars, lipids, drugs, and ions by uniport, symport, or antiport (Henderson and Maiden 1990; Marger and Saier 1993; Paulsen and Skurray 1994; Goffeau et al. 1997; Saier et al. 1999; Lorca et al. 2007; Chen et al. 2008).

The majority of MFS proteins exhibit a structural topology of 12-TM α-helices, with the N- and C- termini residing in the cytosol (Henderson and Maiden 1990; Marger and Saier 1993; Reddy et al. 2012). Despite low sequence similarities, and distinct substrate specificities and transport coupling mechanisms, all MFS proteins share

a common structural fold (Figure 1.3) (Huang et al. 2003; Abramson et al. 2003; Abramson et al. 2004; Yan 2013; Yan 2015). The MFS fold is comprised of two pseudo-symmetrical domains (N- and C-), each consisting of a 6-TM α-helical bundle (Pao et al. 1998; Saier et al. 1999; Reddy et al. 2012). Each 6-helix bundle contains two 3-helix inverted-topology repeats (Radestock and Forrest 2011; Yan 2013; Yan 2015). The two 6-helix domains are thought to form a hydrophobic cavity for substrate binding and translocation (Henderson and Maiden 1990; Sun et al. 2012; Yan 2013). The alternating access model is widely accepted as the general transport mechanism for MFS proteins, where cycles of conformational changes in the transporter expose the substrate binding site to different sides of the membrane (Jardetzky 1966; Yan 2013; Shi 2013).

Protein crystallisation in different conformational states is required to fully understand the mechanism of transport, molecular recognition and co-transport coupling within these transporters (Quistgaard et al. 2016). Structures of some MFS proteins, such as the D-xylose proton symporter (XylE) (Quistgaard et al. 2013; Wisedchaisri et al. 2014) and lactose permease (LacY) (Abramson et al. 2003; Kumar et al. 2014), have been achieved in multiple conformational states. However, MFS transporters are difficult to crystallise due to their partial hydrophobicity and lack of stability when removed from the membrane (Grisshammer and Tate 1995; Carpenter et al. 2008; Kang et al. 2013; Birch et al. 2018). Therefore, insight into the three-dimensional structures of transporters by biochemical and genetic characterisation, as well as by crystallisation, are of importance.

Within the MFS, GusB belongs to the glycoside-pentoside-hexuronide:cation symporter (GPH) family (Poolman et al. 1996). Proteins within the GPH family are homologous, and GusB was demonstrated to share protein homology (which infers similar protein folding patterns) to the MelB protein subfamily (Figure 1.2) (Poolman et al. 1996). A recent crystal structure of MelB in *Salmonella typhimurium* suggests that its cation binding site resides within the first 6-TMs (Ethayathulla et al. 2014). This is similar to XylE, which has been extensively

studied in terms of structural configurations. Therefore, extrapolation of knowledge from well-studied MFS proteins, such as XylE, can provide a platform to study the mechanism of molecular recognition and co-transport coupling in GusB.

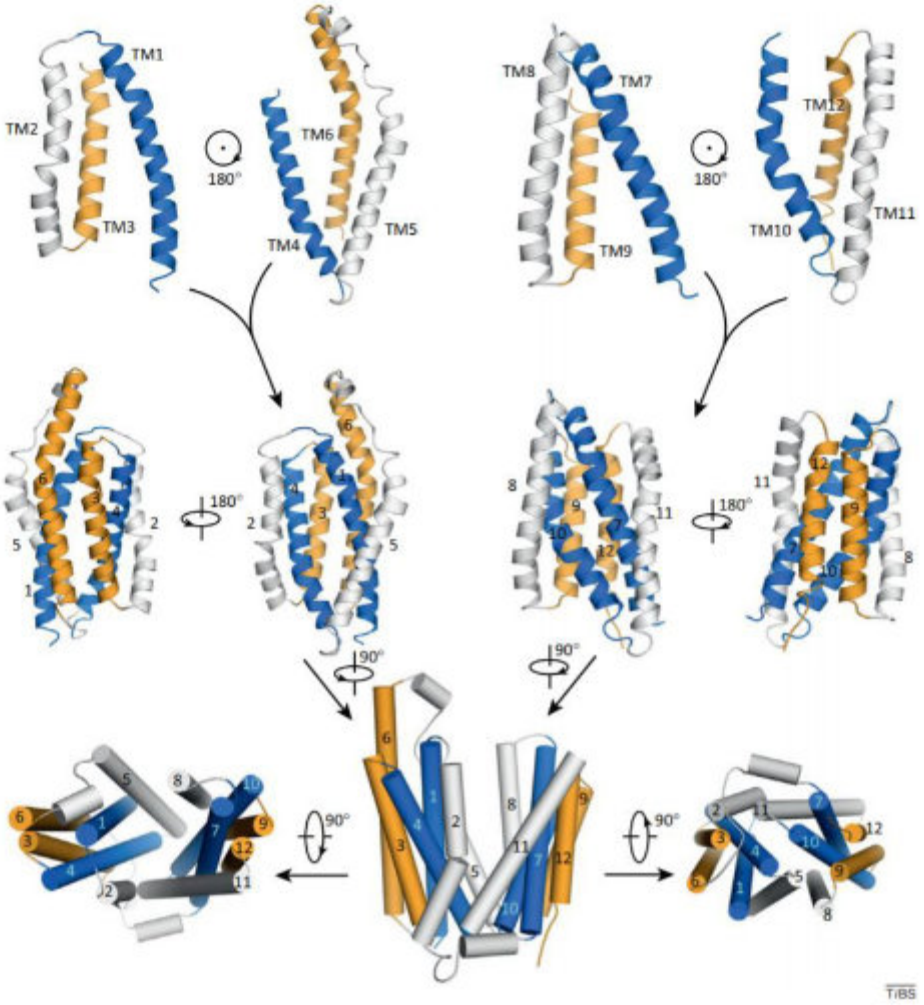

Figure 1.3: A depiction of the major facilitator superfamily (MFS) structural fold of a typical protein comprised of 12-transmembrane (TM) α-helices. The top row represents four 3-TM α-helices. Two inverted 3-TM helix repeats comprise one 6-helix domain (middle row). The two 6-helix domains exhibit pseudo-symmetry in a plane that is perpendicular to the lipid membrane (bottom row). The bottom row represents how both domains comprise a protein of 12-TM α-helices. The corresponding TMs in each 3-TM unit are coloured the same for clarity (Yan 2013).

1.6 The D-xylose proton symporter, XylE

The *E. coli* D-xylose proton symporter (XylE) has been structurally characterised. XylE is an integral membrane protein which is responsible for transporting D-xylose molecules into the cell (Lam et al. 1980). XylE is encoded by the *xylE* gene, which resides separately and downstream of the *xyl* operon (Davis et al. 1984). The *xyl* operon consists of the regulatory gene, *xylR*, and the structural genes *xylA* and *xylB*, which assist in xylose catabolism (Bachmann 1983). XylE consists of 491 amino acids and follows the typical structural topology of MFS proteins: it is comprised of 12-TM α-helices with the N- and C- termini residing within the cytosol (Figure 1.4) (Davis and Henderson 1987; Sun et al. 2012; Quistgaard et al. 2013; Wisedchaisri et al. 2014). The N- and C- domains are pseudo-symmetrical and are connected through an intracellular helical linker (Sun et al. 2012; Wisedchaisri et al. 2014). XylE was the first MFS protein to be structurally characterised in multiple transport conformations which include both outward- and inward- facing conformations (Sun et al. 2012; Quistgaard et al. 2013; Wisedchaisri et al. 2014).

Structural analysis of XylE revealed the transition from the outward- to inward-facing conformation mainly results from structural rearrangement within the C-domain, which undergoes significant local rearrangement during the transport cycle (Wisedchaisri et al. 2014). This non-symmetric conformational change may be attributed to the location of residues responsible for substrate recognition and coordination, which mainly reside in the C-domain (Figure 1.4). Further, structural analysis has revealed that the intracellular helical linkers stabilise XylE and block the substrate translocation pathway until the inward-facing conformation is achieved (Wisedchaisri et al. 2014). Functionality assays of the intracellular helical linkers IC2 and IC3 have demonstrated their importance, where complete removal of the linkers abolished transport activity and mutagenesis substantially reduced transport activity (Wisedchaisri et al. 2014). This indicates that the length and sequence identity of the helical linkers are important for XylE function.

Since XylE has been structurally characterised by achieving multiple crystals, there appears to be no further research into its mechanism of transport or sites of molecular recognition. Much of this research was performed to provide a framework for structurally understanding the human glucose transporters (GLUT), which are XylE orthologs (Yan et al. 2015). Crystals of GLUT transporters have been achieved more recently, which has allowed for their structural determination (Deng et al. 2015; Deng and Yan 2016; Deng and Yan 2018). This could explain the lack of recent XylE structural studies.

1.6.1 Substrate recognition

Structural observations of XylE have revealed that the translocation pathway for D-xylose occurs between the N- and C- domains (Sun et al. 2012). Substrate binding and coordination occur predominantly within the C-domain, where the substrate is coordinated by polar and aromatic amino acids (Figure 1.4) (Sun et al. 2012; Quistgaard et al. 2013). Biochemical characterisation of these residues by site-directed mutagenesis either completely abolished or significantly impaired transport activity (Sun et al. 2012).

To facilitate D-xylose movement through the protein, residues Gln168 and Trp392 (which reside closer to the cytoplasm) form new hydrogen bonds with the substrate (Wisedchaisri et al. 2014). Structural analysis revealed that the majority of residues that comprise the substrate-binding site disperse once in the inward-facing open conformation: Tyr298 remains in position to block the translocation pathway back to the periplasm, whereas Trp392 disperses in a hinge-like movement, which disrupts D-xylose binding and provides access to the cytoplasm (Quistgaard et al. 2013; Wisedchaisri et al. 2014). The dispersion of the substrate-binding residues disrupts substrate-residue interactions which likely weakens the affinity for the substrate and results in its release to the cytoplasm.

1.6.2 Proton coupling sites

Wisedchaisri et al. (2014) performed structural and chemical analyses, and molecular dynamics simulations, which revealed amino acid residues within the N-domain involved with H^+ coupling. These residues are namely Asp27 in TM1 and Glu206 in TM6 (Figure 1.4). They also stated that Arg133 in TM4 interacts with Asp27 via salt bridge and hydrogen bonds. Site-directed mutagenesis confirmed that mutations of Asp27 completely abolished transport function, while Glu206 and Arg133 mutations reduced transport function (Wisedchaisri et al. 2014).

Further, Wisedchaisri et al. (2014) demonstrated that the protonation of Asp27 is vital for conformational changes of XylE. They showed that when deprotonated, Asp27 forms tight interactions with both Arg133 and Glu206 in the periplasmic N-subdomain. These interactions stabilise XylE in the outward-facing conformation and may suppress conformational changes, reducing the rate of transport (Wisedchaisri et al. 2014). In addition, protonation of Asp27 weakens these interactions, facilitating the transition from outward- to inward- facing conformations and allowing the substrate and H^+ to translocate to the cytoplasm (Wisedchaisri et al. 2014). Once the H^+ and D-xylose have dissociated into the cytoplasm, interactions between Asp27 and Arg133, and Asp27 and Glu206, re-establish and facilitate the transition back to the outward-facing conformation (Wisedchaisri et al. 2014). Since Asp27 is structurally distant from the sugar-binding site, D-xylose binding is predicted to be H^+ independent (Wisedchaisri et al. 2014). To summarise, H^+ binding initiates conformational changes and substrate translocation in XylE, and the residues responsible for catalysing this reside within the N-domain.

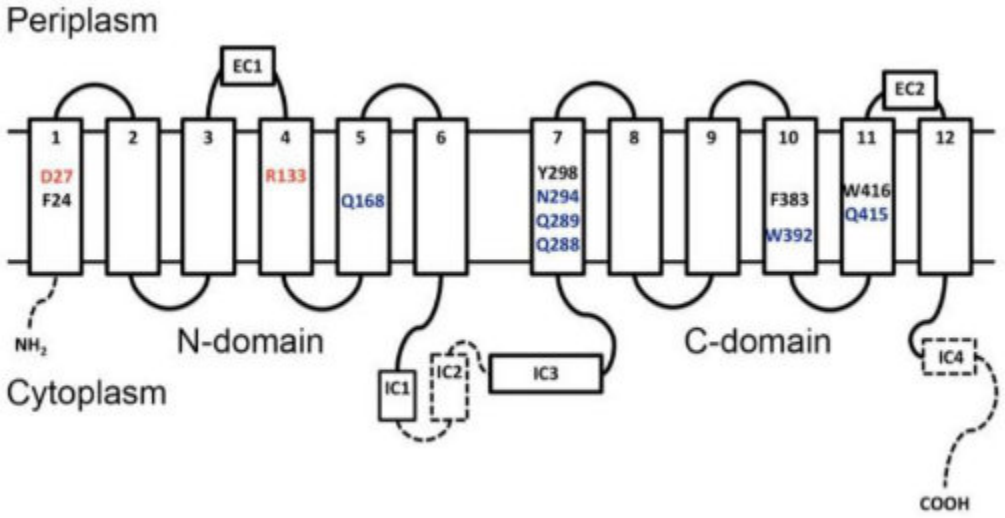

Figure 1.4: Topology of the *E. coli* D-xylose symporter, XylE based on a crystal structure in the inward-facing open conformation. The 12-transmembrane (TM) helices are depicted as rectangles. The N-domain is comprised of TM helices 1-6. The C-domain is comprised of TM helices 7-12. The domains are connected through a helical linker comprised of the three soluble intracellular helices: IC1, IC2 and IC3. The extracellular helical linkers EC1 and EC2 connect TM helices 3 and 4, and 11 and 12, respectively. Dashed rectangles represent areas of partial disorder due to the inward-facing open conformation of the crystal structure. Residues responsible for sugar binding in the outward-facing conformation are labelled blue (interactions through hydrogen bonds) and black (interactions through van der Waals interactions). Resides involved in proton coupling transport are labelled red (Wisedchaisri et al. 2014).

1.7 Rational

Little is known about the three-dimensional structure of GusB. Thus the mechanism of transport, molecular recognition, and H^+ coupling and translocation are not well understood. Findings from other similar, well-studied MFS proteins could help elucidate the mechanisms of transport of GusB. This could have an impact on drug design, drug and disease detection, and advance reporter gene systems in transgenic plants.

Understanding the mechanism of molecular recognition in GusB could guide future drug design. Drugs could be designed to be recognised by GusB, and consequently undergo enterohepatic circulation (Wilson et al. 1992; Roberts et al.

2002). This would increase their half-life and reduce their effective dose. A secondary effect of this preferential recognition could be reduced toxin re-circulation, due to GusB saturation. Therefore, understanding the three-dimensional structure of GusB could advance healthcare.

In addition, functional characterisation of GusB could lead to its development as a commercial reporter gene. GusA has already been developed as a reporter gene in higher plants (Jefferson et al. 1986; Peng et al. 1995; Basu et al. 2004; Conte and Walker 2012; Zhao and Sun 2015). However, to test *gusA* expression, transgenic plants have to undergo enzymatic assays within the laboratory which is expensive and wasteful (Peng et al. 1995; Basu et al. 2004; Conte and Walker 2012; Zhao and Sun 2015). This could be improved with a chimeric *gusA*-*gusB* reporter system that would allow enzymatic assays to be performed within the field: Chromogenic β-D-glucuronides could be administered via a non-invasive method, for example by spraying a liquid. GusB would transport the β-D-glucuronide into the plant cell and GusA would cleave it, giving visual confirmation of gene expression. Therefore, better understanding of the structure of GusB could enable engineering of a *gusA*-*gusB* commercial reporter gene system.

Further, structural understanding of GusB could facilitate the development of biosensors for drug and toxin detection. Glucuronidation of drugs and toxins promotes their excretion (Dutton and Storey 1954; Storey and Dutton 1955; Sun et al. 2015). If GusB could be genetically engineered to recognise glucuronides of interest, transport assays could quantify these substrates in urine and excrement samples. Fluctuations within levels of specific glucuronides could indicate the onset of disease (Wennberg et al. 2006; Fretzayas et al. 2012). Further, this system could be utilised for drug testing.

GusB transports via H^+ symportation and is able to recognise a wide range of β-D-glucuronides (Liang 1992; Liang et al. 2005). Many members of the MFS function by H^+ symportation, despite the unique substrate specificities of each

transporter. The first 6-TM helices of XylE have been defined as the site of H^+ translocation by three-dimensional structural studies and biochemical analyses (Sun et al. 2012; Quistgaard et al. 2013; Wisedchaisri et al. 2014). Since H^+ symportation and structural homology is shared between many MFS transporters (Figure 1.3), the molecular mechanism of H^+ translocation may also be shared. This may be the case for GusB: analysis of a crystal structure of *S. typhimurium*
MelB, which belongs to the same family and shares sequence homology with GusB (Poolman et al. 1996), also suggests that the cation binding occurs within the first 6-TM (Ethayathulla et al. 2014). To test this hypothesis, the first 6-TMs of XylE were fused with the last 6-TMs of GusB. Retained transport activity of GusB would confirm that the first 6-TMs are structurally homologous between the two proteins and advance understanding of the structure of GusB. Since sequence homology between XylE and GusB is low (see Appendix 4) (Henderson and Maiden 1990), this would also infer that the proteins convergently evolved to share structural homology based on similar selection pressures from the environment. Collectively, this knowledge could also further understanding of other MFS proteins.

1.8 Strategy

To investigate the H^+ translocation domain of GusB, domain swapping between XylE and GusB was employed. Structural knowledge from XylE infers that these residues are located primarily within the N-domain. To test whether this was also true for GusB, the first half of the *xylE* gene (encoding the N-domain) was fused with the second half of the *gusB* gene (encoding the C-domain). This was achieved by utilising the polymerase chain reaction (PCR) (Figure 1.5) (Hobert 2002; Bryksin and Matsumura 2010). The x*ylE-gusB* chimera was cloned into the plasmid expression vector pTTQ18 (Figure 1.6). *E. coli* strain MC1061 were transformed with the recombinant DNA. The chimeric XylE-GusB protein was synthesised through the expression of the chimeric *xylE-gusB* gene from the pTTQ18 vector. Chromogenic glucuronides which are transported into the cell are

cleaved by intracellular β-glucuronidases. Therefore, enzymatic assays can be used to assess the function of the membrane transporter when *gusA* expression and β-glucuronidase activity remains constant (Liang et al. 2005).

E. coli strain MC1061 was used for cloning as the genomic *gusB* gene encodes a GusB protein which is only partially functional, due to a Pro100Leu mutation (Liang et al. 2005). However, strain MC1061 genomic *gusA* encodes a functional β-glucuronidase, allowing for enzymatic assays of chromogenic substrates to be employed.

The plasmid vector pTTQ18, first developed by Stark (1987), is a double-stranded DNA of 4563bp (Figure 1.6). Several features make this vector attractive for molecular cloning. Firstly, pTTQ18 contains the pU18 replication origin which drives its high-copy number (500-700) within the cell. In addition, the vector contains a moderately strong promoter (*tac*) to drive transcription, and a strong repressor to prevent transcription in the absence of an inducer (e.g IPTG). The vector also contains a multiple cloning site which is rich with restriction enzyme recognition sequences.

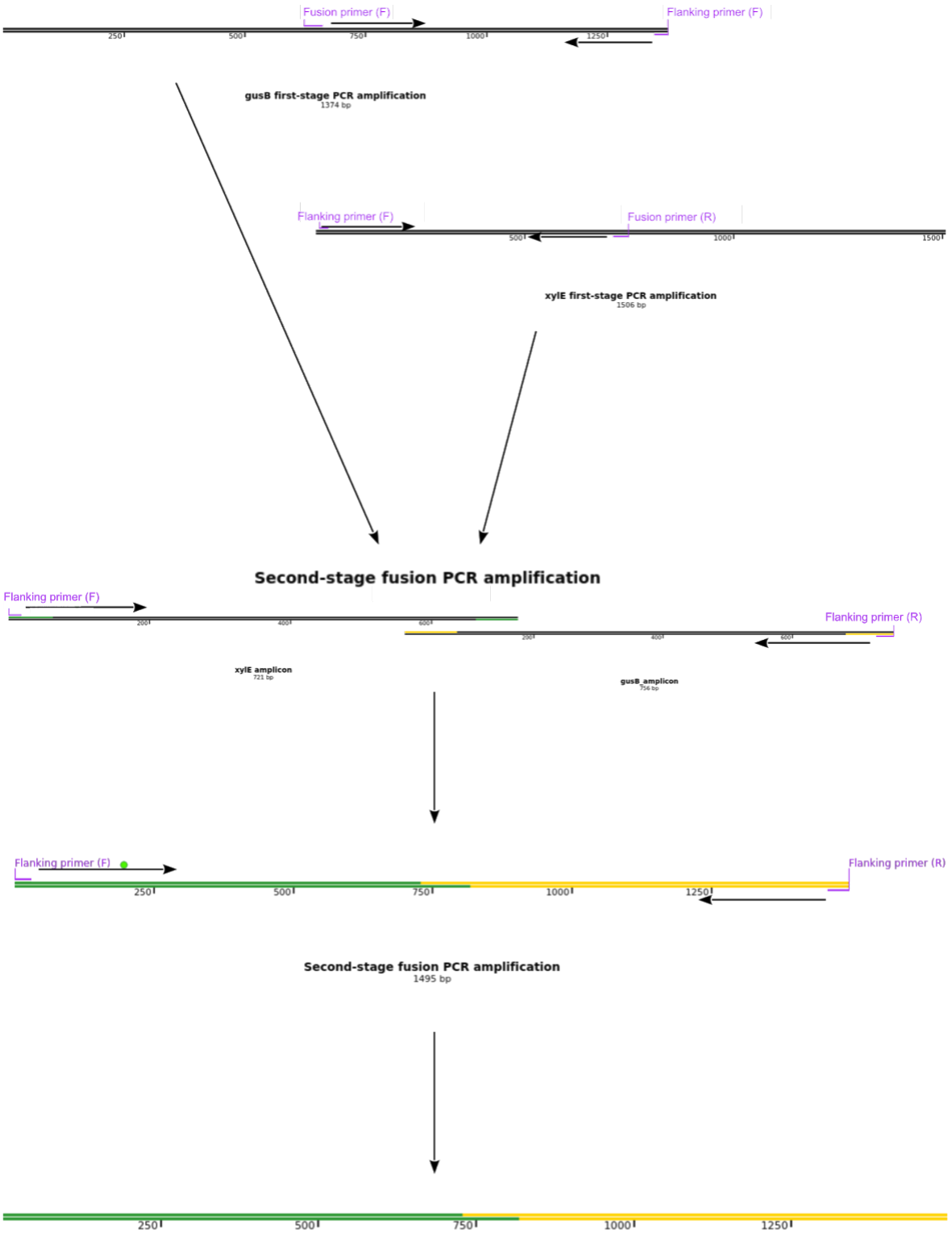

Figure 1.5: An outline of the two-step PCR strategy used within this project. The templates for first-stage PCR were *xylE* and *gusB*. First-stage PCR produced amplicons of the separate domains of *xylE* and *gusB*, respectively. The first-stage amplicons possess overlapping complementary regions which were introduced through primer design. For second-stage fusion PCR the two separate amplicons were mixed, along with flanking primers. The fusion amplicons comprise the first 6-TMs of *xylE* and the last 6-TMs of *gusB* (Hobert 2002; Bryksin and Matsumura 2010).

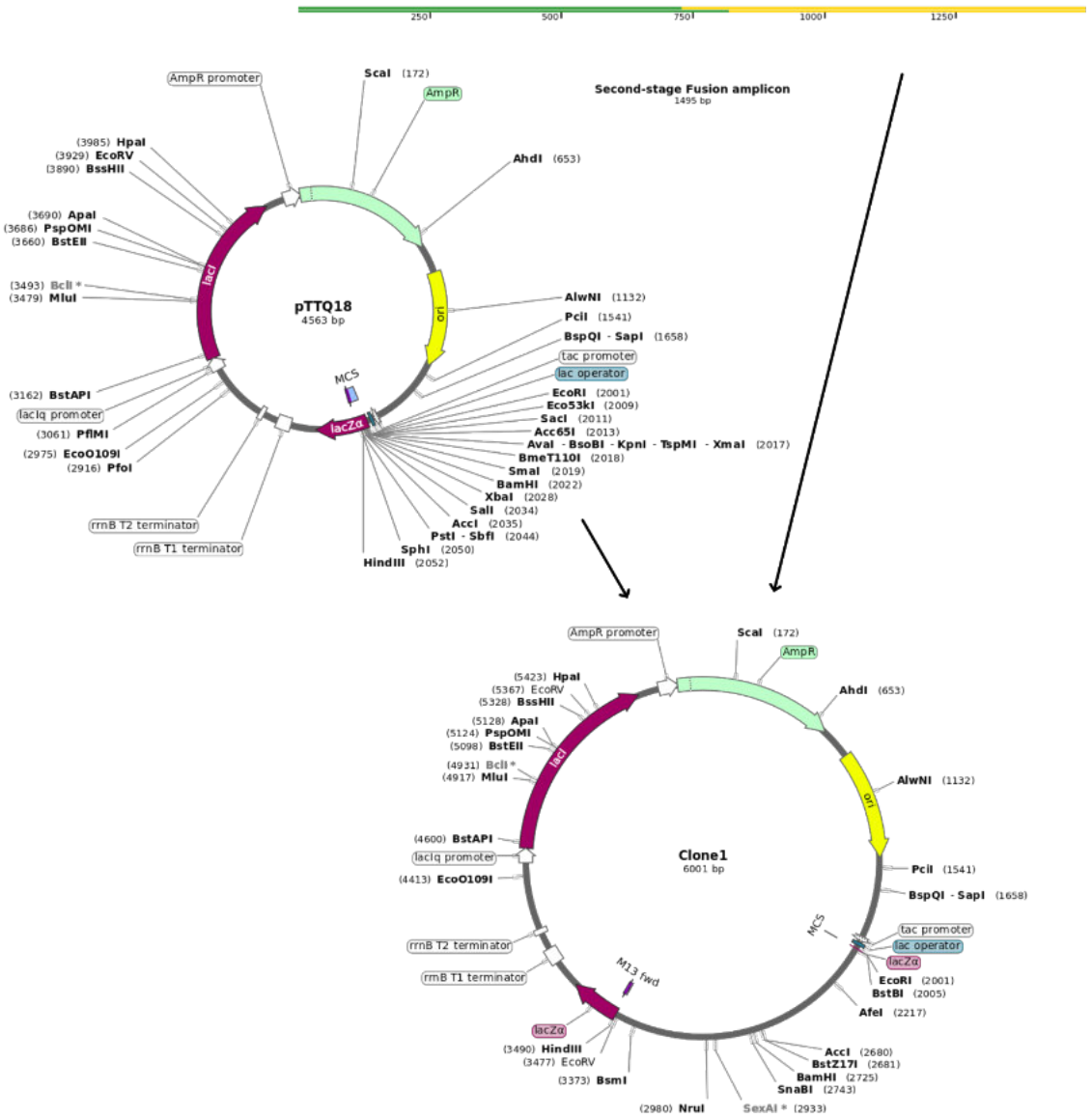

Figure 1.6: An outline of the cloning strategy used within this project. The plasmid pTTQ18 was used as a vector for cloning. Restriction enzymes *Eco*RI and *Hind*III were used to double digest the vector and the second-stage fusion PCR chimeric *xylE-gusB* amplicons, respectively. This process created compatible sticky ends for ligation. *E. coli* strain MC1061 were transformed with the recombinant plasmid DNA.

1.9 Aim and Objectives

1.9.1 Aim

The aim of this project is to locate the H^+ translocation domain of GusB by domain swapping with XylE.

1.9.2 Objectives

The objectives of this project are:

1. Design PCR primers to allow the amplification and fusion of the separate domains of *xylE* and *gusB*.
2. Extract and purify plasmid DNA pTTQ18 (expression vector) and pMJB33 (containing *gusB*), and genomic DNA (containing *xylE*).
3. Separately amplify the N-domain of *xylE* and the C-domain of *gusB* using PCR.
4. Fuse together the amplicons of *xylE* and *gusB* using fusion PCR techniques.
5. Clone the chimeric *xylE-gusB* fusion amplicons into the expression vector pTTQ18.
6. Assess the functionality of the chimeric XylE-GusB protein through enzymatic assays of *gusA* using the chromogenic substrate, X-Gluc.
7. Check the sequence of the chimeric XylE-GusB protein.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals, enzymes and kits

The chemicals and reagents, enzymes, and biological kits used in this project are listed in Table 2.1, Table 2.2, and Table 2.3, respectively.

Table 2.1: The chemicals and reagents used throughout this project. The company from which they were sourced and the catalogue number are provided.

Chemical/Reagent	Source	Catalogue number
Sodium chloride	Fisher Scientific	S271-500
Tryptone	Fisher Scientific	BPE1421-500
Yeast extract	Fisher Scientific	BP1422-500
Agar	Fisher Scientific	BP1423-500
Ampicillin (100mg/ml)	Sigma-Aldrich	A5354
Agarose	Fisher Scientific	BP1356-100
SYBR® safe gel stain	Invitrogen	S33102
5X Green GoTaq® Flexi Buffer	Promega	M7801-M891A
1kb DNA ladder	Promega	G5711
25mM magnesium chloride solution	Promega	M7801-A351B
5X Colorless GoTaq® Flexi Buffer	Promega	M7801-M890A
10X dNTPs	Fisher Scientific	R0192
Sepharose® CL-6B	Sigma-Aldrich	CL6B200
0.1mm Glass beads	Bournemouth University W.J. Liang	-
Tris base	Fisher Scientific	BP152-500
1M Hydrochloric acid	Bournemouth University	Stock
Disodium EDTA	Sigma-Aldrich	2854-15
1M Glacial acetic acid	Bournemouth University	Stock
Glycerol	Sigma-Aldrich	G6279-500
Calcium chloride	Fisher Scientific	AC349615000
10X CutSmart® Buffer	New England Biolabs	B7204S
BSA (10mg/ml)	New England Biolabs	B900IS
T4 DNA ligase 10X buffer	New England Biolabs	B0202S
Isopropyl β-D-1-thiogalactopyranoside	Sigma-Aldrich	I5502
Sodium carbonate	BDH laboratory reagents	30121
5-Bromo-4-chloro-3-indolyl β-D-glucuronide sodium salt	Sigma-Aldrich	B5285-25

Table 2.2: The enzymes used throughout this project. The company from which they were sourced and the catalogue number are provided.

Enzymes	Source	Catalogue number
GoTaq® G2 Flexi DNA Polymerase	Promega	M7801-M780A
Restriction endonuclease *Eco*RI-HF®	New England Biolabs	R3101S
Restriction endonuclease *Hind*III-HF®	New England Biolabs	R3104S
Restriction endonuclease *Sph*I-HF®	New England Biolabs	R3182S
Restriction endonuclease *Eco*RV-HF®	New England Biolabs	R3195S
Restriction endonuclease *Bam*HI	New England Biolabs	R0136S
Restriction endonuclease *Dpn*I	New England Biolabs	R0176S
T4 DNA ligase	New England Biolabs	M0202S
Shrimp alkaline phosphatase (rSAP)	New England Biolabs	M0371S

Table 2.3: The biological kits used throughout this project. The company from which they were sourced and the catalogue number are provided.

Commercial Biological Kits	Source	Catalogue number
QIAprep Spin Miniprep Kit (250)	Qiagen	27106
DNeasy Blood & Tissue Kit (50)	Qiagen	69504
QIAquick PCR Purification Kit (250)	Qiagen	28106
QIAquick Gel Extraction Kit (50)	Qiagen	28704

2.1.2 Equipment

The following specialist equipment was used for the completion of this project: Astell PP65 autoclave; PeqSTAR thermal cycler; Thermo Scientific Nanodrop 2000; Safe Imager 2.0 Introven blue light illuminator; Grant JB Nova water bath; Varian 50 Probe UV spectrophotometer; Bio-Rad PowerPac electrophoresis power supply; Bio-Rad ChemiDoc MP Imaging system; IKA KS 4000i Incubator Shaker; Heraeus Biofuge Pico Centrifuge; and refrigerated Eppendorf Centrifuge 5430R.

2.1.3 Solutions

All solutions, buffers and media were prepared using sterile, deionised water (dH_20). Solutions were mixed using a magnetic stirrer hot plate and sterile magnets and subsequently autoclaved. The constituents and methods for making up routinely used solutions within this project are listed in Appendix 5.

2.1.4 Media

Preparation of LB media and LB agar plates

To ensure growth of bacteria, nutrient-rich Luria-Bertani (LB) media was prepared according to Maniatis et al. (1982). The constituents of these media are listed in Table 2.4.

Table 2.4: Constituents for 1 litre of LB media and LB agar media, respectively. Reagents were dissolved in 1 litre of dH_20.

LB media	
Reagent	Weight (g)
Tryptone	10
Yeast extract	5
Sodium chloride	10

LB agar media	
Reagent	Weight (g)
Tryptone	10
Yeast extract	5
Sodium chloride	10
Agar	15

For both LB media and LB agar media the constituents (Table 2.4) were dissolved in 1 litre of dH_20. The solutions were autoclaved and left to cool. Ampicillin was added at a concentration of 1 µl per 1 ml, where required.

Agar plates were prepared by pouring ~25 ml of the autoclaved agar solution into a sterile petri dish. The plates were left to solidify on the laboratory bench side, before being placed in a 37°C incubator to dry. The dried plates and cooled LB media were stored at 4°C in the fridge.

2.1.5 Bacterial strains and plasmids

The *E. coli* strain and plasmids used in this project are as follows:

1. *E. coli* strain MC1061; F- Δ(ara-leu)7697 [araD139]B/r Δ(codB-lacI)3 galK16 galE15 λ- e14- mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2(r-m+); Sourced from W.J. Liang, Bournemouth University.
2. pTTQ18, *E. coli* strain MC1061; sourced from W.J. Liang, Bournemouth University (Stark 1987).
3. pMJB33, *E. coli* strain MC1061; sourced from W.J. Liang, Bournemouth University.

2.1.6 Primers

Primers were designed to fuse the first 6-TM helices of XylE with the last 6-TM helices of GusB (Table 2.5). *GusB* from pMJB33, *E. coli* strain MC1061 and *xylE* from *E. coli* strain MC1061 were used as templates for primer design. Five sets of primers were designed to fuse the N-domain of XylE and the C-domain of GusB from 5 different positions. Additional primers were designed to fuse the *gfp* gene which encodes the green fluorescent protein (GFP) between these domains (Appendix 9). However, due to time constraints they were not used within this project.

Table 2.5: Primer design for XylE-GusB fusion. Primers were sourced from Eurofins Genomics. The 5'-end of the flanking primers featured a restriction site for either *Eco*RI or *Hind*III. To facilitate PCR fusion, a nucleotide bridge was incorporated between the *xylE* and *gusB* sequences.

Primer name	Primer sequence	Annealing temperatur e (°C)	Author
XyleF	5'-ATACTGAATTCGAATGGTCTAAGGCAGG TCTGA- 3'	66.3	LD*
XyleGus bSR1	5'-TACCACATTTGCGGATCCTCCGCCGCGC GACAT-3'	67.8	LD
XyleGus bSR2	5'-GTTTCAGGGTTGCGGATCCTCCGGTTTT GCGGC-3'	67.8	LD
XyleGus bSR3	5'-GCTGCGCAACTGCGGATCCTCCCAGGA TACCTT-3'	67.8	LD
XyleGus bSR4	5'-CAGACTGATTGCGGATCCTCCCGTGTTG CCCATAAT-3'	67.2	LD
XyleGus bSR5	5'-GATATTCAATGATGCGGATCCTCCCATA ATTTTGCGCAG-3'	65.2	LD
XyleGus bSF1	5'-ATGTCGCGCGGCGGAGGATCCGCAAAT GTGGTA-3'	67.8	LD
XyleGus bSF2	5'-GCCGCAAAACCGGAGGATCCGCAACCC TGAAAC-3'	67.8	LD
XyleGus bSF3	5'-AAGGTATCCTGGGAGGATCCGCAGTTG CGCAGC-3'	67.8	LD
XyleGus bSF4	5'-ATTATGGGCAACACGGGAGGATCCGCA ATCAGTCTG-3'	67.2	LD
XyleGus bSF5	5'-CTGCGCAAAATTATGGGAGGATCCGCAT CATTGAATATC-3'	65.2	LD
GusbR	5'-ATACTAAGCTTTTAATTAGTGATATCGCT GATTAATTGC-3'	69.1	LD

*LD = Lydia Doherty

Key: Random nucleotides *Eco*RI restriction site *Hind*III restriction site *xylE* sequence *gusB* sequence Amino acids forming the nucleotide bridge *Bam*HI Restriction site within the nucleotide bridge

2.1.7 Rehydrating and diluting primers

The primers initially exist in a lyophilized state. Therefore, it was necessary to rehydrate them so that they could be utilised for PCR. The primers were dissolved in the appropriate volumes of $T_{10}E_1$, as specified by Eurofin Genomics. The concentration of the rehydrated primers was 100µM. To prepare the primers for PCR (10µM), a 1 in 10 dilution using dH_20 was performed.

2.2 Methods

2.2.1 Primer design

Several variables were considered when designing primers. These included primer length, sequence composition, melting temperatures, GC content and optimisation for PCR fusion and cloning. See Appendix 8 for the complete list of factors which were considered when designing primers.

2.2.2 Isolation of bacterial colonies

Dry agar plates were streaked with *E. coli* strains and incubated at 37°C overnight. This allowed single bacterial colonies of each strain to be isolated. Plates were streaked with the following strains and plasmids which were isolated from 'deep' stock sample: strain MC1061, pTTQ18 (strain MC1061), and pMJB33 (strain MC1061).

To streak the plates an inoculation loop was sterilised by heating over a hot flame until the loop glowed red-hot. After waiting for ~15 seconds for the loop to cool, it was placed into the 'deep' *E. coli* sample. The pattern of streaking was: three streaks in one direction, sterilisation of the loop, three streaks approximately 120° to and slightly overlapping the previous section. This was repeated in order to obtain streaks of varying bacterial growth, as illustrated in Figure 2.1.

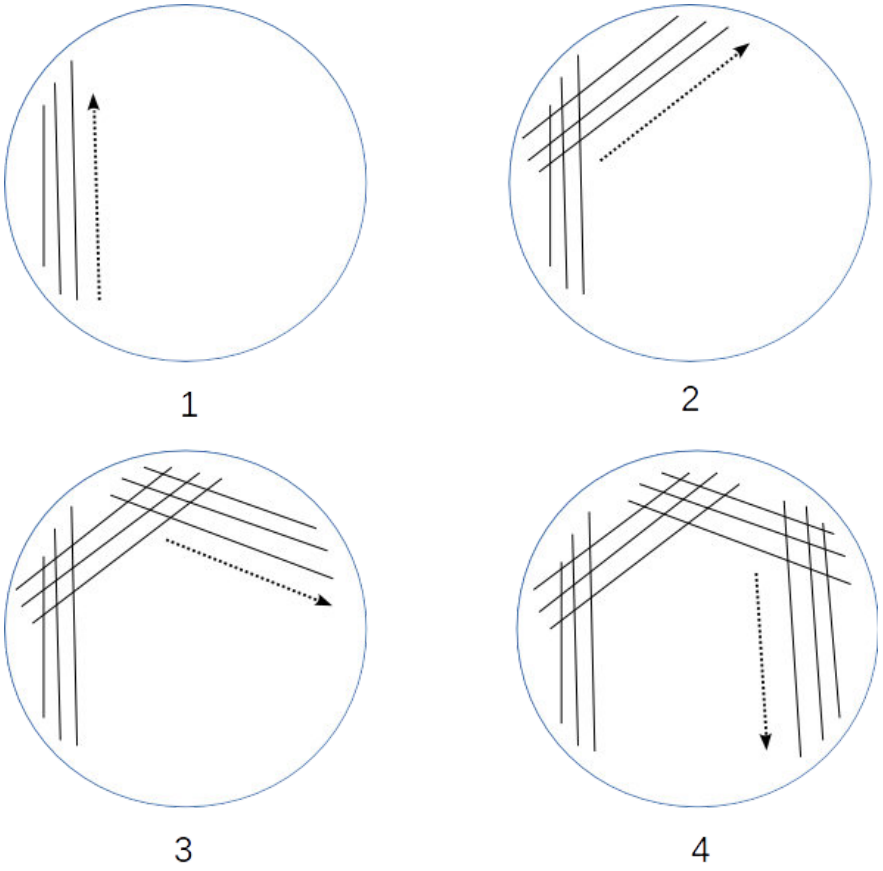

Figure 2.1: The method used to streak agar plates for overnight cell culture. The plates were streaked with an inoculation loop which was dipped into the *E. coli* stock sample prior to stage 1. The inoculation loop was sterilised between each step. At each step the inoculation loop was streaked in the direction indicated by the dotted arrow. This process created streaks of varying bacterial growth to allow for single colonies to be easily extracted.

2.2.3 Bacterial subculture and inoculation

50 ml of LB media was poured into a sterile conical flask. For pMJB33 and pTTQ18, ampicillin was added to the conical flask at a concentration of 1 µl per 1 ml. Ampicillin was not added for *E. coli* strain MC1061. An inoculation loop was sterilised by heating over a hot flame until the loop glowed red-hot. After waiting for ~15 seconds for the loop to cool, a single bacterial colony was extracted from the previously incubated agar plate. The loop was dipped into the LB media and twirled to ensure transfer of the cells. To maintain sterile conditions, this process

was conducted next to an open flame. Unless otherwise stated, samples were incubated overnight at 37°C with vigorous shaking at 250 rpm.

2.2.4 Plasmid DNA extraction

Plasmid DNA was extracted from inoculated bacterial cells using the Qiagen QIAprep Spin Miniprep kit (catalogue number 27106) (Putaala et al. 2000). Within this procedure the manufacturer's protocol was followed, and 1 ml of the inoculated bacterial culture was used as starting volume.

2.2.5 Genomic DNA extraction

Cells were grown overnight to approximately high-log phase (OD_{600nm} 1.0), and 1 ml of the inoculated bacterial culture was used as starting volume. Genomic DNA was extracted from inoculated bacterial cells using the Qiagen DNeasy® Blood & Tissue kit (catalogue number 69504), following the manufacturer's protocol (Huang et al. 2009). The protocol for pretreatment of Gram-Negative Bacteria was also followed in order to maximise the yield of DNA extracted from *E. coli* strain MC1061.

2.2.6 NanoDrop 2000 Spectrophotometer

The NanoDrop 2000 Spectrophotometer was routinely used to estimate the concentration and purity of nucleic acid samples. The loading platforms were cleaned and blanked using 1 µl dH_20, prior to loading 1 µl of the DNA sample. The purity of samples was estimated using the OD_{260nm}/OD_{280nm} ratio. A value between 1.8 and 2.0 was accepted as pure for DNA. A value lower than 1.8 may indicate the presence of proteins, phenols and other contaminants (Maniatis et al. 1982). An optical density value of 1.0 at OD_{260nm} was considered a concentration of 50µg per 1 ml, for double stranded DNA.

2.2.7 Agarose gel electrophoresis

Agarose gel electrophoresis was used to estimate the identity and concentration of nucleic acids within this project, based on molecular weight and intensity, respectively. The co-electrophoresis of a DNA ladder (catalogue number G5711), which comprised of bands of known molecular weight and size, with the samples in this project facilitated this.

Agarose gels were prepared by dissolving the desired amount of agarose powder (catalogue number BP1356-100) in the desired volume of 1X TAE buffer. This was achieved by heating in a microwave oven for bursts of 30 seconds. The concentration of agarose gel used in this project was 1.6% to allow for adequate separation of DNA fragments. The melted agarose solution was left to stand until it had cooled to approximately 55°C. SYBR® safe gel stain (catalogue number S33102) was added in a concentration of 5 µl per 45 ml of agarose solution and poured into a clean gel tray fitted with a plastic comb. Once the gel had solidified, the comb was removed to create loading wells, and then the gel was placed in an appropriately sized boat and immersed in 1X TAE buffer. To allow visualisation, the DNA samples were mixed with 5X Green GoTaq® Flexi Buffer (catalogue number M7801 - M890A) (ratio of 4:1) prior to loading. The DNA marker was co-electrophoresed on the same gel, in which a current of 70V was passed through the gel for 50 minutes.

Gels were visualised using the ChemiDoc MP Imaging system to assess DNA separation. To facilitate visualisation, the colours on the DNA gels were inverted.

2.2.8 Polymerase Chain Reaction (PCR)

The PCR was invented by Kary Mullis in 1983 (Saiki et al. 1985; Mullis 1990) and is used to repeatedly synthesise a region of DNA which is selected for by primers. In this project, PCR was routinely performed using the peqSTAR thermal cycler, in

which a two-stage protocol was implemented (Hobert 2002; Bryksin and Matsumura 2010). The settings used for PCR are described in Table 2.8. The components for each stage of the reaction are described in Tables 2.6 and 2.7.

In the first-stage PCR, amplicons of *xylE* and *gusB* were generated in separate reactions. Amplicons of the N-domain of *xylE* were generated whereby genomic DNA acted as a template for synthesis, and the forward flanking primer (XyleF) and its respective reverse fusion primer, were utilised. Amplicons of the C-domain of *gusB* were generated whereby plasmid DNA (pMJB33) acted as a template for synthesis, and the reverse flanking primer (GusbR) and its respective forward fusion primer, were utilised. Amplicon identity from the first-stage PCR was assessed by agarose gel electrophoresis (Section 2.2.7) for the presence of a band of expected molecular weight.

Following purification of the PCR products (Section 2.2.9), the samples underwent second-stage fusion PCR. In this process, the separate amplicons of *xylE* and *GusB* were mixed, along with the flanking primers (XyleF and GusbR) in order to synthesise and amplify the fused gene product (Figure 1.5). These fusion products consisted of the N-domain of *xylE* and the C-domain of *gusB*. Amplicon identity from the second-stage fusion PCR was checked by agarose gel electrophoresis (Section 2.2.7) for the presence of a band of expected molecular weight.

Table 2.6: The components of first-stage PCR. Components were kept on ice and added in sequential order to 0.2 ml thin wall PCR tubes.

Component	Volume (µl)	Final Concentration
dH_20	28.5	-
5X Colorless GoTaq® Flexi Buffer	10.0	1X
25 mM $MgCl_2$	2.0	1.0 mM
1 mM dNTP	1.0	0.02 mM
10 µM Forward primer	1.0	0.20 µM
10 µM Reverse primer	1.0	0.20 µM
Template DNA (25 ng/µl)	6.0	3 ng/µl
Taq polymerase (5 units/µl)	0.5	2.5 units/µl
Total	50.0	-

Table 2.7: The components of second-stage fusion PCR. Components were kept on ice and added in sequential order to 0.2 ml thin wall PCR tubes.

Component	Volume (µl)	Final Concentration
dH_20	26.5	-
5X Colorless GoTaq® Flexi Buffer	10.0	1X
25 mM $MgCl_2$	2.0	1.0 mM
1 mM dNTP	1.0	0.02 mM
10 µM Forward primer	1.0	0.20 µM
10 µM Reverse primer	1.0	0.20 µM
xylE amplicons (25 ng/µl)	4.0	2 ng/µl
gusB amplicons (25 ng/µl)	4.0	2 ng/µl
Taq polymerase (5 units/µl)	0.5	2.5 units/µl
Total	50.0	-

Table 2.8: The PCR settings used within this project.

Step	Temperature (°C)	Time (minutes)	Purpose
1	95	5	Denaturation
2*	94	1	Denaturation
3*	60	0.5	Annealing
4*	72	0.75 / 1.5**	Polymerisation
5	72	4	Polymerisation

*1 cycle was considered the completion of steps 2 to 4. These steps were consecutively repeated and 35 cycles were completed in total. For re-amplification of second-stage fusion products, only 25 cycles were completed.

**For first-stage PCR the polymerisation step was 0.75 minutes. For second-stage fusion PCR this duration was increased to 1.5 minutes.

2.2.9 Purification of PCR products

Samples from the first-stage PCR were treated with 1 µl of restriction endonuclease *Dpn*I (catalogue number R0176S) and 5.5 µl 10X CutSmart® Buffer (catalogue number B7204S) and incubated at 37°C for 60 minutes. Methylated template DNA was digested to ensure that only PCR amplicons were available for second-stage fusion PCR (Geier and Modrich 1979; Barras and Marinus 1989). *Dpn*I was then heat inactivated at 80°C for 20 minutes. The treated samples were checked by agarose gel electrophoresis (Section 2.2.7), in which the purity of samples was also visually assessed. The treated samples from the first-stage PCR were deemed pure enough to progress with second-stage fusion PCR. The samples from second-stage fusion PCR were purified using the QIAquick PCR Purification Kit (catalogue number 28106) (Kwok et al. 2012). This process was performed according to the manufacturer's protocol, with the exception of the last step. To ensure that the maximum volume of DNA was obtained from the process, the elution buffer provided was replaced with freshly prepared Tris-HCl buffer (Appendix 5).

2.2.10 Restriction digestion

Restriction digestions were performed to facilitate cloning and to characterise the cloned recombinant DNA. Prior to digestion, the total nucleic acid concentrations of samples was estimated using spectrophotometry (Section 2.2.6) and gel electrophoresis (Section 2.2.7). Restriction mapping of the substrates was performed in order to select the restriction sites and calculate expected fragment sizes (See Appendix 6). The volume of restriction enzyme required was calculated based on the manufacturer's unit definition for a known DNA molecule. Molar concentrations were calculated and extrapolated to the DNA substrates within this project. High fidelity restriction endonucleases were used, where available. These steps were performed to prevent star activity and over digestion.

The components of digestion reactions are listed in Table 2.9.

2.2.10.1 Digestion to facilitate cloning

To facilitate cloning, pTTQ18 and the PCR fusion amplicons were digested with the same restriction endonucleases, *Eco*RI (catalogue number R3101S) and *Hind*III (catalogue number R3104S). This generated compatible sticky ends between the two molecules to facilitate ligation.

In addition to this, a single digestion with *Sph*I was performed on the ligated recombinant samples, prior to transformation. This ensured that any self-ligated vectors which did not contain the PCR fusion product were digested, and thus not able to establish during transformation.

2.2.10.2 Successive single digestions of pTTQ18

Successive single digestions with *Eco*RI and *Hind*III were performed on pTTQ18. Both restriction enzymes are 100% active in CutSmart® Buffer (catalogue number B7204S), thus a double digestion was possible. However, to confirm the activity of

each restriction enzyme, successive single digestions were performed. Half of the samples were digested with *Eco*RI and half of the samples were digested with *Hind*III. Samples were incubated at 37°C for 90 minutes, prior to heat inactivation at 80°C for 20 minutes. Digestion success was confirmed through gel electrophoresis (Section 2.2.7), prior to treatment with the alternate restriction enzyme. Since the restriction sites for *Eco*RI and *Hind*III reside within the multiple cloning site of pTTQ18 and are separated by only 21 nucleotide bases (see Appendix 6), a double digestion would not have provided a high enough discrimination power to visualise the success of both enzymes through gel electrophoresis.

2.2.10.3 Double digestion of PCR fusion amplicons

A double digestion with *Eco*RI and *Hind*III was carried out on the PCR fusion amplicons. This was possible since both enzymes are 100% active in CutSmart® Buffer. Samples were incubated at 37°C for 90 minutes, prior to heat inactivation at 80°C for 20 minutes. Digestion success was confirmed through gel electrophoresis (Section 2.2.7).

2.2.10.4 Digestion of ligated samples

A single digestion with *Sph*I (catalogue number R3182S) was performed on the ligated recombinant DNA samples (Section 2.2.13). The restriction endonuclease *Sph*I recognises a single restriction site within the multiple cloning site of pTTQ18 (see Appendix 6). This site exists between the restriction sites of *Eco*RI and *Hind*III, therefore pTTQ18 vectors which had religated without the insert would be digested and prevented from establishing during transformation. The PCR fusion amplicons were assessed to ensure that they did not feature *Sph*I restriction sites (see Appendix 6). Samples were incubated at 37°C for 90 minutes, prior to heat inactivation at 65°C for 20 minutes. Digestion success was confirmed through gel electrophoresis (Section 2.2.7).

Table 2.9: The components for the digestion of: pTTQ18 vector, PCR fusion amplicon insert, ligated recombinant DNA samples, and cloned DNA samples. All reactions were made up to 50 µl with dH_20.

	Final concentration			
Components	pTTQ18	PCR fusion amplicons	Ligated recombinant DNA samples	Cloned DNA samples
10X CutSmart® Buffer	1X	1X	1X	1X
BSA (10 mg/ml)	0.2 µg/µl	0.2 µg/µl	0.2 µg/µl	0.2 µg/µl
pTTQ18 (30 ng/µl)	9.0 ng/µl	-	-	-
PCR fusion product (10 ng/µl)	-	2.0 ng/µl	-	-
Ligated recombinant DNA samples (5 ng/µl)	-	-	2.0 ng/µl	-
Cloned DNA samples (30 ng/µl)	-	-	-	2.1 ng/µl
*Eco*RI (20 units/µl)	0.2 units/µl	0.2 units/µl	-	-
*Hind*III (20 units/µl)	0.2 units/µl	0.2 units/µl	-	-
*Sph*I (20 units/µl)	-	-	0.04 units/µl	-
*Eco*RV (20 units/µl)	-	-	-	1.3 units/µl
*Bam*HI (20 units/µl)	-	-	-	0.2 units/µl

2.2.10.5 Digestion to characterise the transformants

Restriction digestions were performed on the extracted cloned plasmids (Section 2.2.16). This allowed the cloned DNA molecules to be manipulated and compared to expected fragment sizes (see Appendix 6). Along with DNA sequencing, this facilitated identification of the cloned DNA molecules. Restriction endonucleases *Eco*RV and *Bam*HI were used. Three fragments of 3357, 1890 and 752 nucleotide base pairs were expected for a successfully cloned vector harbouring a *xylE-gusB* chimera (see Appendix 6). Samples were incubated at 37°C for 90 minutes, prior

to heat inactivation at 65°C for 20 minutes. Results were visualised by gel electrophoresis (Section 2.2.7).

2.2.11 Shrimp Alkaline Phosphatase (rSAP) treatment of pTTQ18

The digested pTTQ18 samples were treated with 1 µl rSAP (concentration 10 units/µl, catalogue number M0371S) and 4.5 µl of 10X CutSmart® Buffer (catalogue number B7204S), and incubated at 37°C for 30 minutes. rSAP catalyses the dephosphorylation of the 5'- and 3'- ends of the linearised pTTQ18 vector to prevent religation (Olsen et al. 1991; Nilsen et al. 2001). The samples were then incubated at 65°C for 5 minutes to heat inactivate rSAP enzymes.

2.2.12 Spin-column Chromatography of pTTQ18

The digested pTTQ18 samples were purified via spin-column chromatography. Prior to this step, the pTTQ18 samples contain the linearised, dephosphorylated vector and the multiple cloning site which was digested from the plasmid (Section 2.2.10.2). This step ensured the digested multiple cloning site was removed to further prevent religation.

Spin columns were prepared according to Maniatis et al. (1982). 0.6 ml Eppendorfs were pierced with a 25 gauge sterile needle and placed within a 1.5 ml Eppendorf. 100 µl of autoclaved glass beads was pipetted into the 0.6 ml Eppendorfs. 750 µl of the washed sepharose beads (catalogue number CL6B200) (see Appendix 5) was pipetted on top of the glass beads. To prepare the column matrix and remove residual flow-through, the spin columns were centrifuged at 9000 rpm for 10 minutes. The flow through was discarded and the 0.6 ml spin columns were placed into new, sterile 1.5 ml Eppendorfs.

The digested pTTQ18 samples were added to the prepared columns, and the columns were centrifuged at 9000 rpm for 10 minutes. The flow through from this process was the purified pTTQ18 sample.

2.2.13 Ligation

Ligation reactions were set up in sterile 1.5 ml Eppendorfs, with a total reaction volume of 20 µl (Table 2.10). Molar ratios of pTTQ18 vector to PCR fusion amplicon insert were varied to encourage ligation. Controls were performed to assess the efficacy of the ligation process (C2), and to assess the efficacy of rSAP (C1) (Section 2.2.11). The PCR fusion amplicon inserts were absent from both controls; C1 contained the linearised pTTQ18 vector and T4 DNA ligase, C2 contained only the linearised pTTQ18 vector.

Ligation reactions were incubated at 18°C for 16 hours, followed by a further incubation at 4°C for 30 minutes. Samples were then incubated at 65°C for 10 minutes to heat inactivate T4 DNA ligase.

Table 2.10: The components for the ligation of the pTTQ18 vector and the PCR fusion amplicon insert. Components were kept on ice and added in sequential order to 1.5 ml Eppendorfs. Four ratios of vector to insert were performed to encourage ligation (L1-L4). Two controls were used to verify the efficacy of ligation (C2) and rSAP (C1).

Components	L1 (µl)	L2 (µl)	L3 (µl)	L4 (µl)	C1 (µl)	C2 (µl)
dH_20	13.5	11.5	7.5	13.0	16.5	17.5
10X T4 DNA ligase buffer	2.0	2.0	2.0	2.0	2.0	2.0
pTTQ18 (Vector) (3 ng/µl)	0.5	0.5	0.5	1.0	0.5	0.5
PCR fusion amplicon (Insert) (0.5 ng/µl)	3.0	6.0	9.0	3.0	-	-
T4 DNA ligase	1.0	1.0	1.0	1.0	1.0	-
Total	20.0	20.0	20.0	20.0	20.0	20.0
Vector:Insert	1:1	2:1	3:1	1:2	-	-

2.2.14 Preparation of competent cells

E. coli strain MC1061 was cultured on ampicillin-free agar plates (Section 2.2.2), subcultured, and inoculated overnight (Section 2.2.3). 1 ml of this culture was subsampled into an autoclaved conical flask, and 49 ml of fresh LB media was added. The subsampled culture was inoculated for 1.25 hours at 37°C with vigorous shaking at 250 rpm to mid-log phase (OD_{600nm} 0.2-0.3). Cultures were chilled on ice to arrest further cell growth.

1 ml of the cells were placed into a sterile 1.5 ml Eppendorf and harvested by centrifugation at 8000 rpm for 4 minutes. The supernatant was discarded and the cells were resuspended in 500 µl of ice-cold sterile calcium solution (50 mM $CaCl_2$, 10 mM Tris-HCl, pH 8.0). Samples were placed in an ice bath for 15 minutes and centrifuged at 8000 rpm for a further 4 minutes. The supernatant was discarded and the cells were resuspended in 66 µl of ice-cold sterile calcium solution. The competent cells were kept on slushy ice.

2.2.15 Transformation

10 µl of each ligated sample (Section 2.2.13) was added into an Eppendorf of competent cells on slushy ice (Section 2.2.14). The samples were initially incubated on slushy ice for 1 hour. The samples were then heat shocked at 42°C for 90 seconds in a water bath and immediately returned to slushy ice for 2 minutes. 330 µl of LB media was added to each sample, before incubating the cells at 37°C with vigorous shaking at 250 rpm, for 1 hour. 300 µl of the incubated cells were plated onto dried ampicillin-containing LB agar plates (Section 2.1.4) and spread with glass beads to ensure even distribution. The plates were incubated at 37°C overnight. The following day transformants were examined under a blue light using Safe Imager Invitrogen, and stored in the fridge at 4°C.

2.2.16 Transformant characterisation and storage

Colonies from transformation were extracted and cultured overnight. Three single colonies from each plate were extracted and inoculated according to Section 2.2.3.

‘Deep’ samples of this cultured medium were created by combining 300 µl of culture with 150 µl of glycerol (G6279-500) and 150 µl of dH_20. ‘Deep’ samples were stored at -80°C in preparation for functionality tests (Section 2.2.17).

Plasmid DNA was extracted according to Section 2.2.4. Samples were characterised through restriction digestion (Section 2.2.10.5).

2.2.17 Enzymatic functionality tests

The functionality of the clones was assessed through enzymatic assays of intracellular β-glucuronidase enzymes with X-Gluc (5-Bromo-4-chloro-3-indolyl β-D-glucuronide sodium salt, catalogue number B5285-25).

‘Deep’ samples were cultured on ampicillin-free agar plates (Section 2.2.2), subcultured, and inoculated overnight (Section 2.2.3). 1 µl of each culture was subsampled into an autoclaved Eppendorf with 49 µl of fresh LB media and 1 mM of Isopropyl β-D-1-thiogalactopyranoside (IPTG). The addition of IPTG induced transcription of the *xylE-gusB* chimera by removing the repression on the *taq* promoter. The subsampled cultures were inoculated for 2 hours at 37°C with vigorous shaking at 250 rpm to approximately mid-log phase. 1 mM of X-Gluc was added to each sample, and the samples were incubated at 37°C with vigorous shaking at 250 rpm for a further 20 minutes. The samples were visualised, where a blue precipitate indicated X-Gluc transport and cleavage. The samples were re-incubated and observed every 20 minutes for a total of 80 minutes. Positive control pMJB33 and negative control pTTQ18 acted as visual references for colour intensity.

2.2.18 DNA Sequencing

The plasmid DNA from the seven clones which revealed positive X-Gluc transportation was extracted and sent to GENEWIZ for DNA sequencing. The DNA was extracted as described in section 2.2.16, to ensure the same samples were characterised by restriction digestion and sequencing.

3. Results

3.1 Visualisation and assessment of purity of extracted DNA

Genomic DNA from strain MC1061 and plasmid DNA from pTTQ18 and pMJB33 were extracted as described in Sections 2.2.5 and 2.2.4. The integrity of the extracted plasmids was visualised by gel electrophoresis (Section 2.2.7) (Figure 3.1), however, genomic DNA is too large to be visualised in this way. The concentration and purity of all DNA samples were assessed by spectrophotometry

(Section 2.2.6) (Table 3.1). The DNA samples extracted were of concentrations between 19.5 ng/µl and 173.4 ng/µl (Appendix 11). The DNA samples considered the best quality were diluted (see Section 2.2.6), where necessary, to provide a more accurate estimation of purity and concentration (Table 3.1). DNA samples used as template for PCR were diluted as recommended by New England BioLabs (Table 3.1).

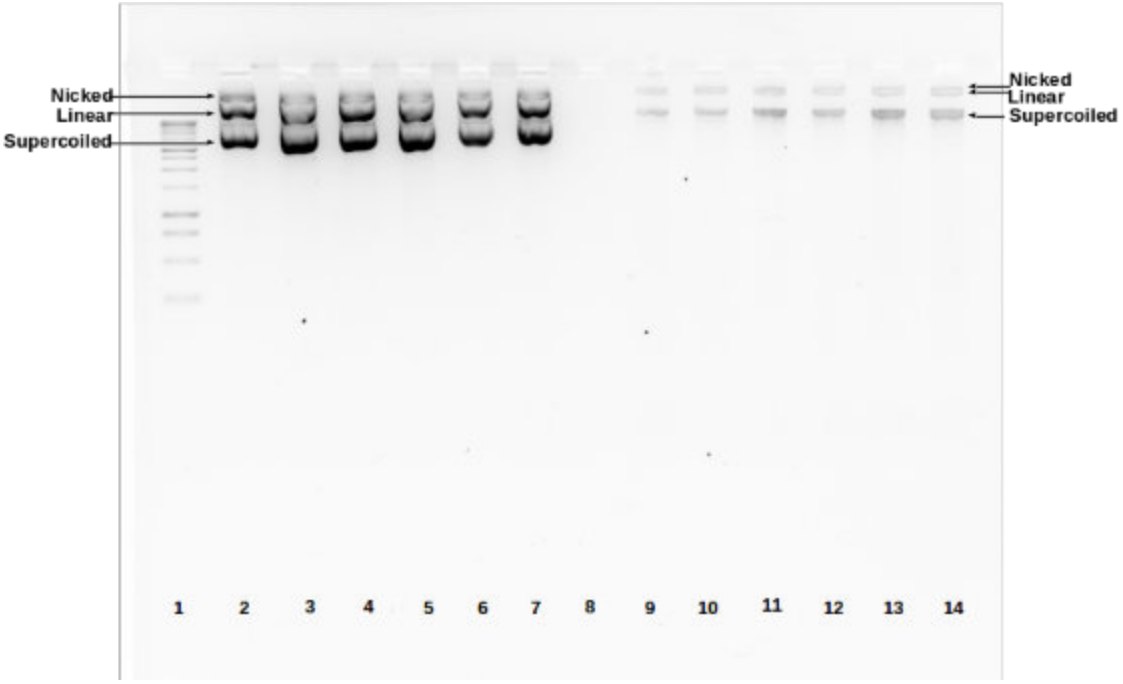

Figure 3.1: Evaluation of plasmid DNA integrity extracted from *E. coli* strain MC1061. Lane 1, DNA ladder (catalogue number G5711); Lanes 2-7, plasmid pTTQ18; Lanes 9-14, plasmid pMJB33. All plasmids appear to be good quality and are present in nicked, linear and supercoiled states. The colours within the DNA gel were inverted for clarity and to save ink.

Table 3.1: The concentrations and purities of extracted DNA samples used within this project.

Sample	Concentration (ng/µl)	OD_{260nm}	OD_{260nm}/OD_{280nm} Ratio
MC1061	23.3	0.466	1.90
MC1061*	5.9	0.218	1.84
pMJB33	22.1	0.443	1.96
pMJB33*	5.4	0.208	1.84
pTTQ18	122.2	2.444	1.86
pTTQ18**	31.8	0.635	1.80

*DNA samples used as template for PCR were diluted by a factor of 4 to ~5ng, as recommended by New England BioLabs.
**The DNA sample used as cloning vector was diluted by a factor of 4 to provide a more accurate estimation of purity and concentration.

3.2 First-stage PCR - amplification of *gusB* and *xylE*

PCR reactions were performed to amplify the separate domains of *gusB* and *xylE* (Section 2.2.8). Amplicon identity was assessed by gel electrophoresis (Section 2.2.7); successful results from first-stage PCR are shown in Figure 3.2. Many PCR reactions were performed prior to these results which were unsuccessful (See Appendix 11). Several troubleshoots were performed in order to obtain successful amplifications for all samples.

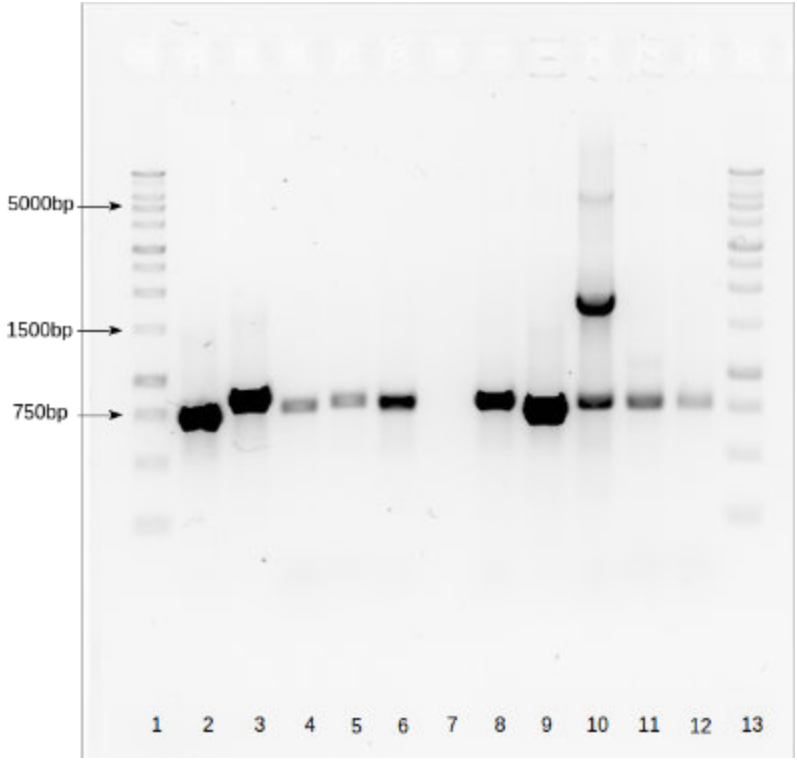

Figure 3.2: Evaluation of amplicon identity from first-stage PCR. Lane 1 and 13, DNA ladder (catalogue number G5711); Lanes 2-6, *xylE* amplicons (molecular weights 693bp, 798bp, 720bp, 741bp and 732bp, respectively); Lanes 8-12, *gusB* amplicons (molecular weights 756bp, 708bp, 741bp, 720bp and 729bp, respectively). Two undefined bands feature in lane 10 at positions ~1700bp and ~5500bp. One faint undefined band features in lane 11 at position ~1200bp. These bands represent non-specific binding of primers. The colours within the DNA gel were inverted for clarity and to save ink.

3.3 Second-stage fusion PCR - fusion of *gusB* and *xylE* amplicons

Second-stage fusion PCR reactions were performed to fuse the separate domains of *gusB* and *xylE* (Section 2.2.8). Amplicon identity was assessed by gel electrophoresis (Section 2.2.7); successful results from second-stage fusion PCR are shown in Figure 3.3 and Figure 3.4, respectively. Figure 3.3 represents the first successful fusion PCR reaction, however, the majority of samples were lost when they were unsuccessfully excised via gel extraction (See Appendix 11). The small volumes of fusion samples remaining acted as template for re-amplification

(Figure 3.4). Figure 3.4 contains a mixed sample of undefined, non-specific amplifications, unfused amplicons and the fusion products.

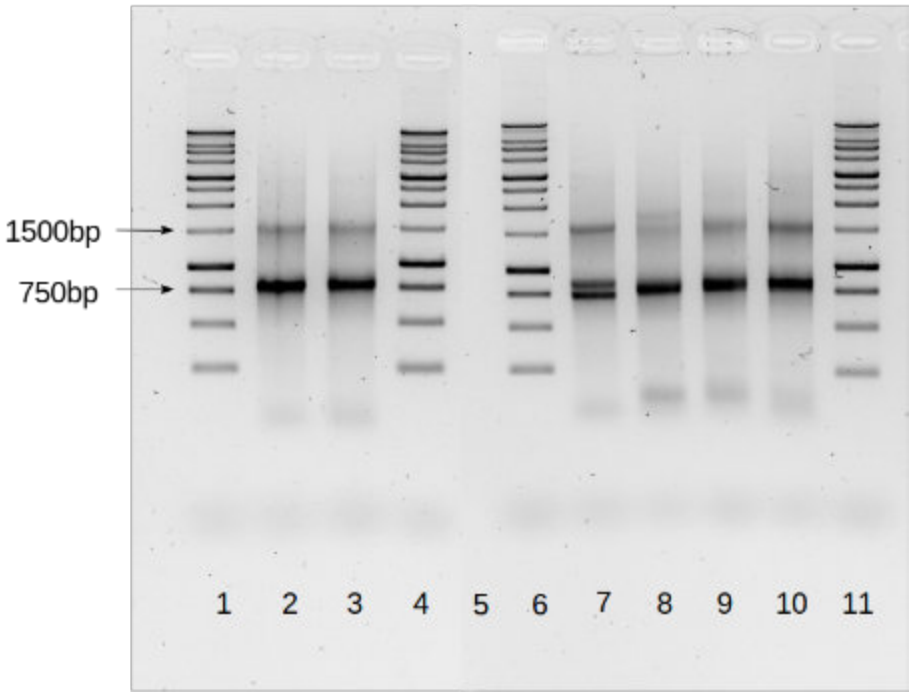

Figure 3.3: Evaluation of amplicon identity from second-stage fusion PCR. Lanes 1, 4, 6 and 11, DNA ladder (catalogue number G5711); Lanes 2 and 3, fusion #1 (sample was re-loaded as lane 2 was pierced); Lane 7, fusion #2; Lane 8, fusion #3; Lane 9, fusion #4; Lane 10, fusion #5. All samples contain a mixture of fused (~1500bp) and non-fused DNA products (~750bp). Bands representing non-specific binding of primers are visible within lane 7 at position ~800np and within lane 8 at position ~1800bp. The colours within the DNA gel were inverted for clarity and to save ink.

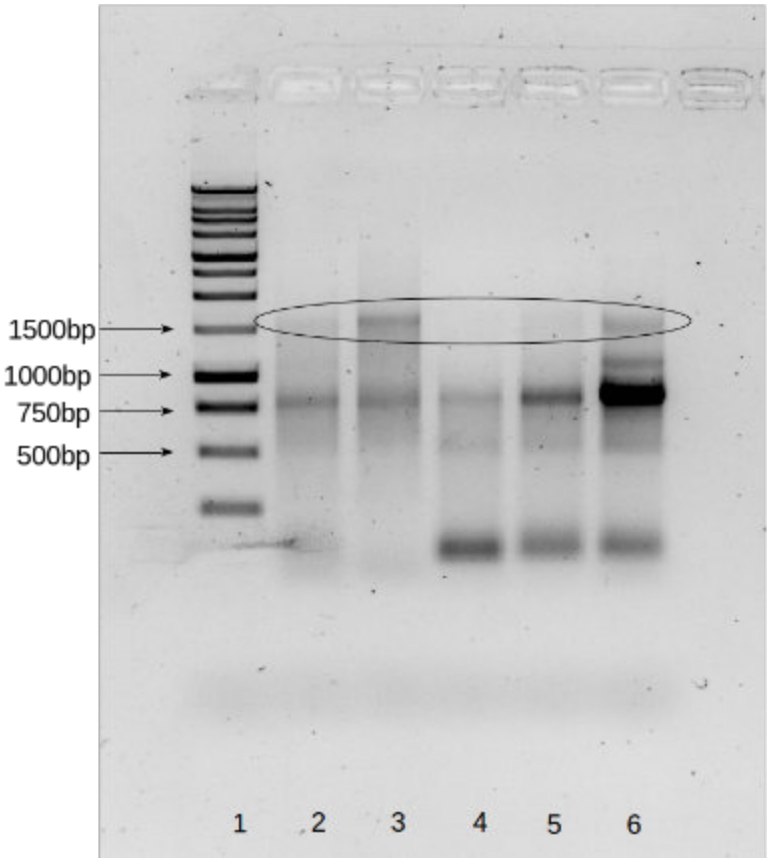

Figure 3.4: Evaluation of amplicon identity from second-stage fusion PCR after re-amplification. Lane 1, DNA ladder (catalogue number G5711); Lane 2, fusion #1; Lane 3, fusion #2; Lane 4, fusion #3 (although not visible in this figure, a very faint band was visible in the original visualisation); Lane 5, fusion #4; Lane 6, fusion #5. The fusion products have been circled. All samples contain a mixture of fused (~1500bp) and non-fused DNA products (~750bp). In addition, all samples contain non-specific amplifications. The colours within the DNA gel were inverted for clarity and to save ink.

3.4 Purification of second-stage PCR fusion products

To facilitate cloning, the second-stage fusion PCR samples were purified as described in Section 2.2.9. The results from PCR purification were assessed by gel electrophoresis (Section 2.2.7), and are shown in Figure 3.5. Purification removed a large proportion of the smearing and non-specific amplification products seen in Figure 3.4. After purification, the samples contained one

undefined fragment of approximately 200bp, unfused amplicons, and the fusion products.

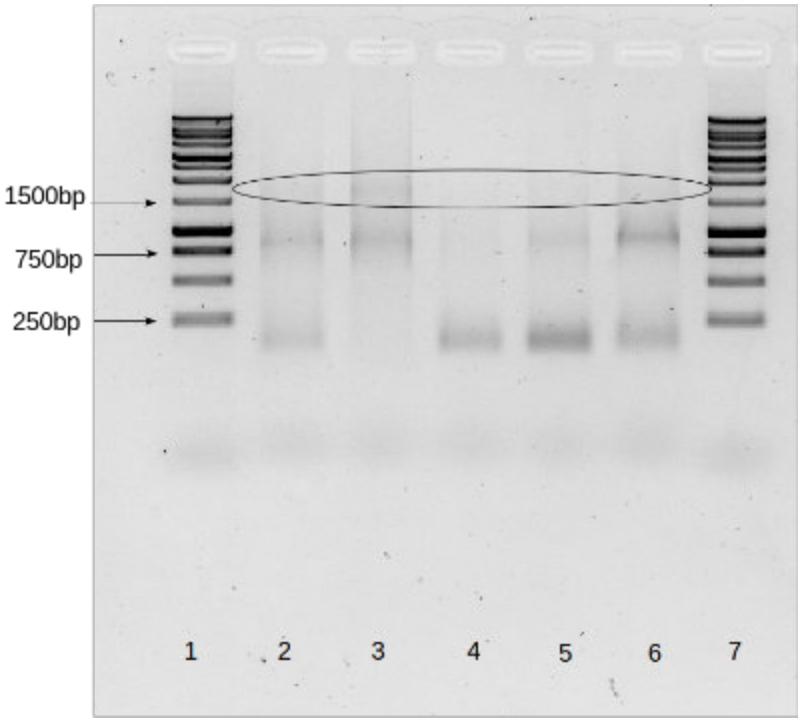

Figure 3.5: Evaluation of PCR purification on second-stage fusion PCR amplicons. Lanes 1 and 7, DNA ladder (catalogue number G5711); Lane 2, fusion #1; Lane 3, fusion #2; Lane 4, fusion #3; Lane 5, fusion #4; Lane 6, fusion #5. Although difficult to see in this figure, the fusion product bands were visible in the original visualisation, and have been circled. All samples contained a mixture of fused (~1500bp) and non-fused DNA products (~750bp), and an undefined fragment at position ~200bp. The PCR purification step was successful as it removed the majority of non-specific amplification products present within the samples. The colours within the DNA gel were inverted for clarity and to save ink.

3.5 Restriction digestion of pTTQ18

The plasmid pTTQ18 was digested with restriction enzymes *Eco*RI and *Hind*III as described in Section 2.2.10.2. Successive single digestions were conducted to ensure the activity of each enzyme. The results from the first single digestion were assessed by gel electrophoresis (Section 2.2.7), and are shown in Figure 3.6.

Single digestions with both enzymes were successful, as evident from the appropriately sized bands at position ~4500bp. The slight juxtaposition between the undigested DNA and digested samples represents the tight packing of supercoiled DNA: less friction is created and thus the samples can move through the agarose gel at a faster rate. A successive single digestion with the alternate enzyme was then completed for each sample. To limit sample waste, the results from the successive digestion were assumed to be adequate and were not visualised by gel electrophoresis.

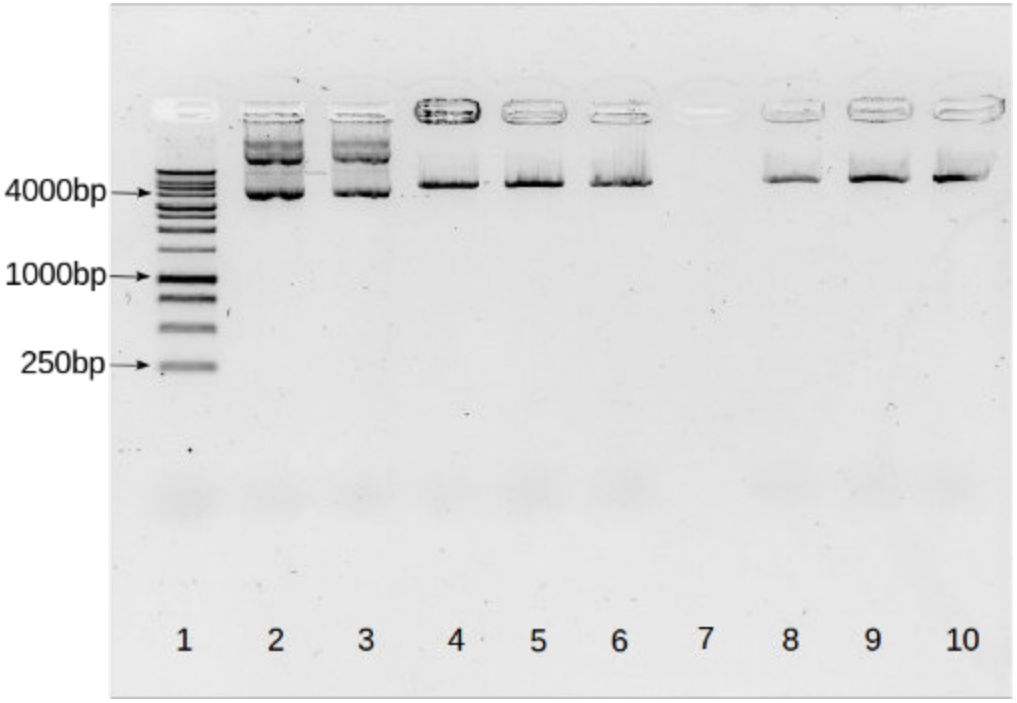

Figure 3.6: Evaluation of restriction digested pTTQ18 with *Eco*RI and *Hind*III. Lane 1, DNA ladder (catalogue number G5711); Lanes 2 and 3, undigested plasmid pTTQ18 as negative control; Lanes 4-6, digested pTTQ18 with *Eco*RI; Lanes 8-10, digested pTTQ18 with *Hind*III. The colours within the DNA gel were inverted for clarity and to save ink.

3.6 Restriction digestion of second-stage PCR fusion products

The fusion products underwent a double digestion with restriction enzymes *Eco*RI and *Hind*III as described in Section 2.2.10.3. The results were assessed by gel electrophoresis (Section 2.2.7), and are shown in Figure 3.7. The results

demonstrate that the PCR fusion products were not degraded, however, gel electrophoresis cannot determine whether the double digestion was successful (Figure 3.7). Since the lengths of DNA digested from the PCR fusion products are ~5bp they are not large enough to be visualised. Digestion was completed twice for each fusion product in order to maximise the amount of insert DNA available for ligation.

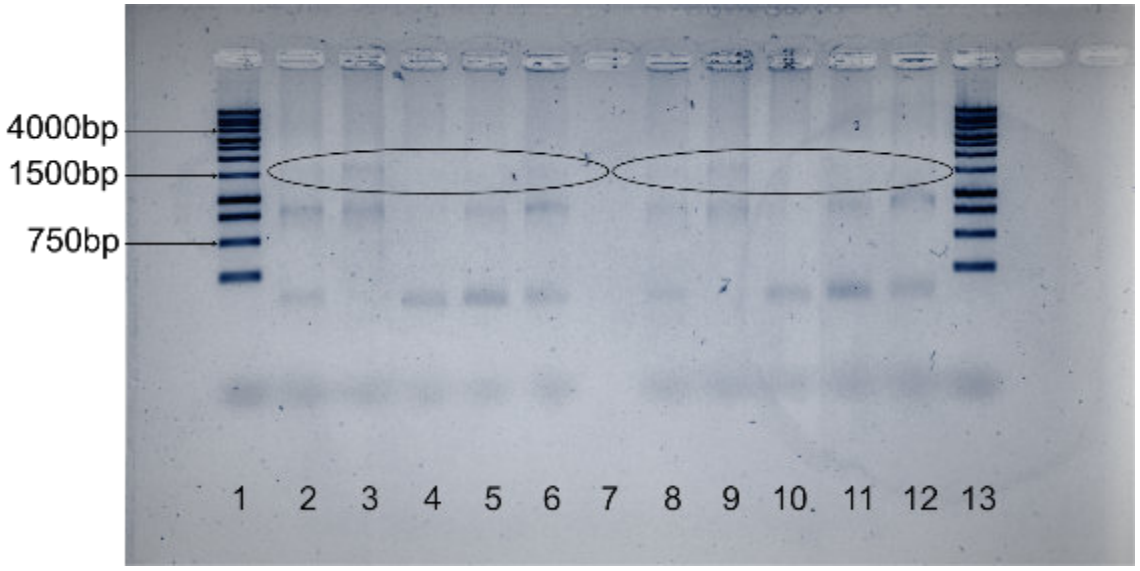

Figure 3.7: Evaluation of *Eco*RI and *Hind*III double digests of second-stage PCR fusion amplicons. Lanes 1 and 13, DNA ladder (catalogue number G5711); Lanes 2 and 8, double-digested fusion amplicon #1; Lanes 3 and 9, double-digested fusion amplicon #2; Lanes 4 and 10, double-digested fusion amplicon #3; Lanes 5 and 11, double-digested fusion amplicon #4; Lanes 6 and 12, double-digested fusion amplicon #5. Although difficult to see here, the digested fusion product bands were visible in the original visualisation and have been circled. The colours within the DNA gel were inverted for clarity and to save ink.

3.7 Ligation of PCR fusion product and pTTQ18

The digested pTTQ18 vector and digested fusion amplicons were ligated overnight as described in Section 2.2.13. Samples were further subjected to a single digestion with *Sph*I (Section 2.2.10.4). The results from ligation and subsequent digestion were assessed by gel electrophoresis (Section 2.2.7), and are shown in Figure 3.8. Ligation was successful, as evident from the faint bands immediately below the loading wells of the gel. Since the amount of DNA present in these

samples was very low, it was predicted that it may be difficult to visualise the bands by gel electrophoresis. The predicted molecular weight of the recombinant DNA product was ~6000bp, however, the bands present from this reaction are extremely high. According to the manufacturer, the increased specificity of high fidelity *Sph*I may increase DNA binding and cause the enzyme to remain attached during electrophoresis.

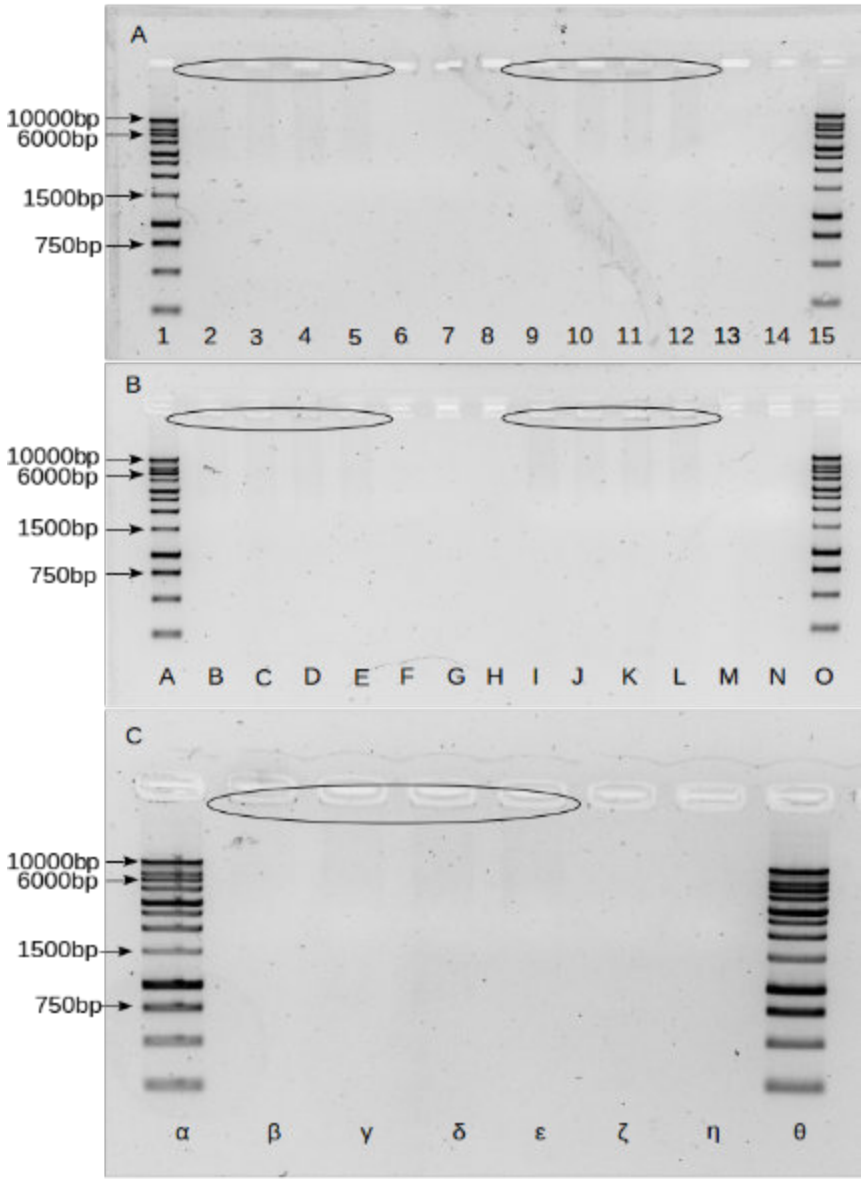

Figure 3.8: Evaluation of ligation of pTTQ18 vector and PCR fusion amplicons. Lanes 1, 15, A, O, α and θ, DNA ladder (catalogue number G5711); Lanes 2-5, varying ligation ratios for vector and *xylE-gusB* chimera #1; Lanes 6 and 7, control ligations for *xylE-gusB* chimera #1; Lanes 9-12, varying ligation ratios for vector and *xylE-gusB* chimera #2; Lanes 13 and 14, control ligations for *xylE-gusB* chimera #2; Lanes B-E, varying ligation ratios for vector and *xylE-gusB* chimera #3; Lanes F and G, control ligations for *xylE-gusB* chimera #3; Lanes I-L, varying ligation ratios for vector and *xylE-gusB* chimera #4; Lanes M and N, control ligations for *xylE-gusB* chimera #4; Lanes β-ε, varying ligation ratios for vector and *xylE-gusB* chimera #5; Lanes ζ and η, control ligations for *xylE-gusB* chimera #5. The faint, high molecular weight bands immediately below the loading wells of the gel represent ligated samples and have been circled for clarity. The colours within the DNA gel were inverted for clarity and to save ink.

3.8 Transformation

Ligated samples were introduced into *E. coli* (strain MC1061) as described in Section 2.2.15. All plates incubated with the ligated recombinant DNA samples showed well-established colonies, suggesting successful transformation (Figure 3.9; Figure 3.10; Figure 3.11; Figure 3.12; Figure 3.13). Collectively, 563 colonies were present on the plates and available for selection. Control transformants (plates C1 and C2) for each clone showed limited growth, relative to the ligated recombinant DNA plates (L1-L4). C1 (Vector with DNA ligase) showed a higher growth frequency than C2 (Vector without DNA ligase). One colony was present within a C2 plate (Figure 3.9). Although all ligation ratios of vector to insert provided well-established bacterial colonies, plates of ratio 2:1 gave the highest growth frequencies, for all samples.

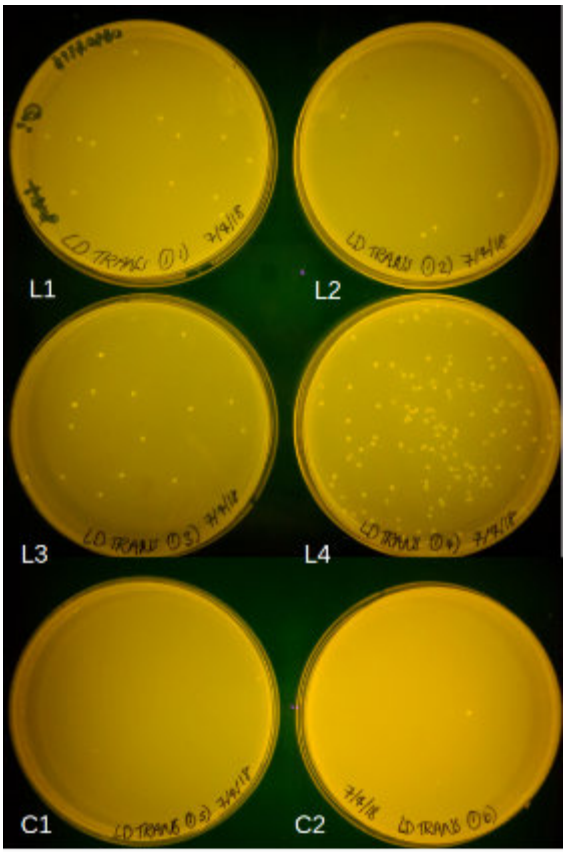

Figure 3.9: Transformed bacterial colonies (*E. coli* strain MC1061) with the chimeric *xylE-gusB*, variation #1. Plates labelled L1-L4 represent varying ligation ratios of pTTQ18 vector to PCR fusion amplicon insert, structural variation #1. Sample L1, ligation ratio 1:1; Sample L2, ligation ratio 1:2; Sample L3, ligation ratio 1:3; Sample L4, ligation ratio 2:1. Plates labelled C1 and C2 represent controls.

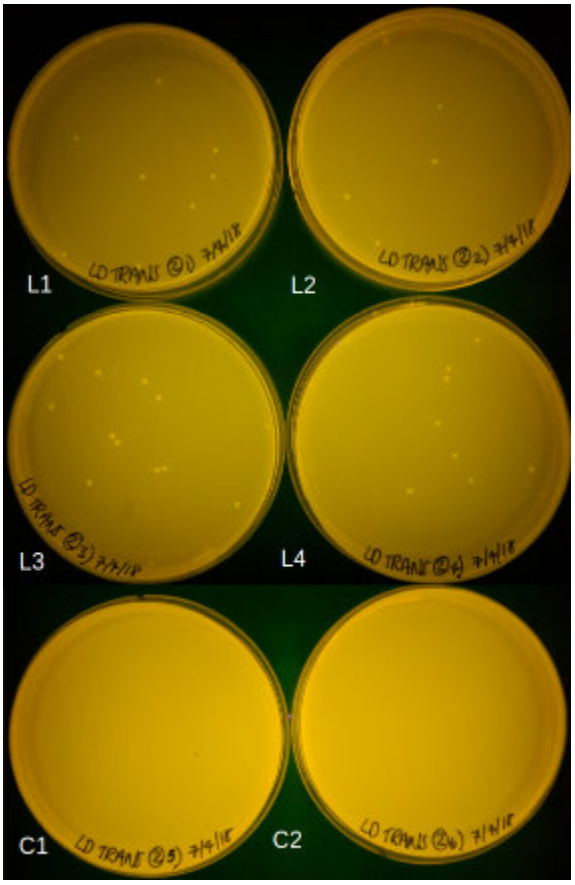

Figure 3.10: Transformed bacterial colonies (*E. coli* strain MC1061) with the chimeric *xylE-gusB*, variation #2. Plates labelled L1-L4 represent varying ligation ratios of pTTQ18 vector to PCR fusion amplicon insert, structural variation #2. Sample L1, ligation ratio 1:1; Sample L2, ligation ratio 1:2; Sample L3, ligation ratio 1:3; Sample L4, ligation ratio 2:1. Plates labelled C1 and C2 represent controls.

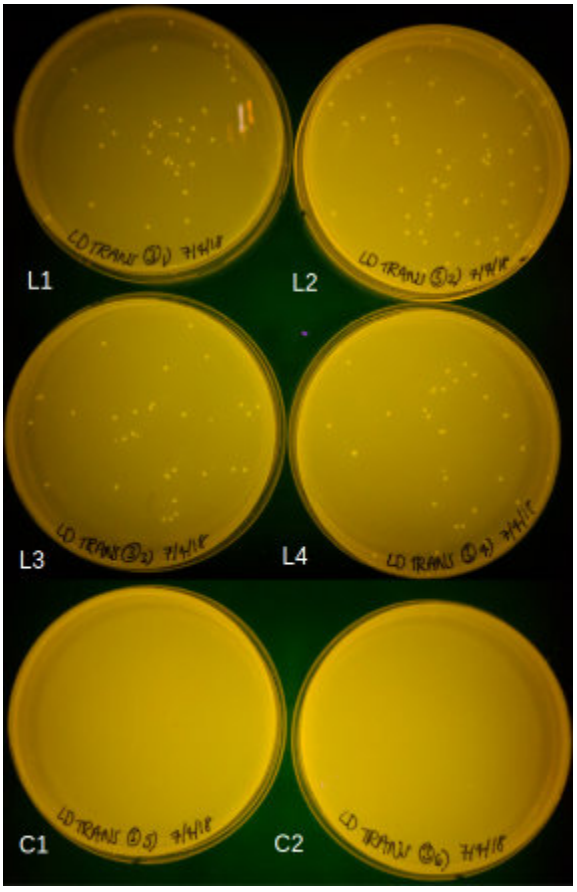

Figure 3.11: Transformed bacterial colonies (*E. coli* strain MC1061) with the chimeric *xylE-gusB*, variation #3. Plates labelled L1-L4 represent varying ligation ratios of pTTQ18 vector to PCR fusion amplicon insert, structural variation #3. Sample L1, ligation ratio 1:1; Sample L2, ligation ratio 1:2; Sample L3, ligation ratio 1:3; Sample L4, ligation ratio 2:1. Plates labelled C1 and C2 represent controls.

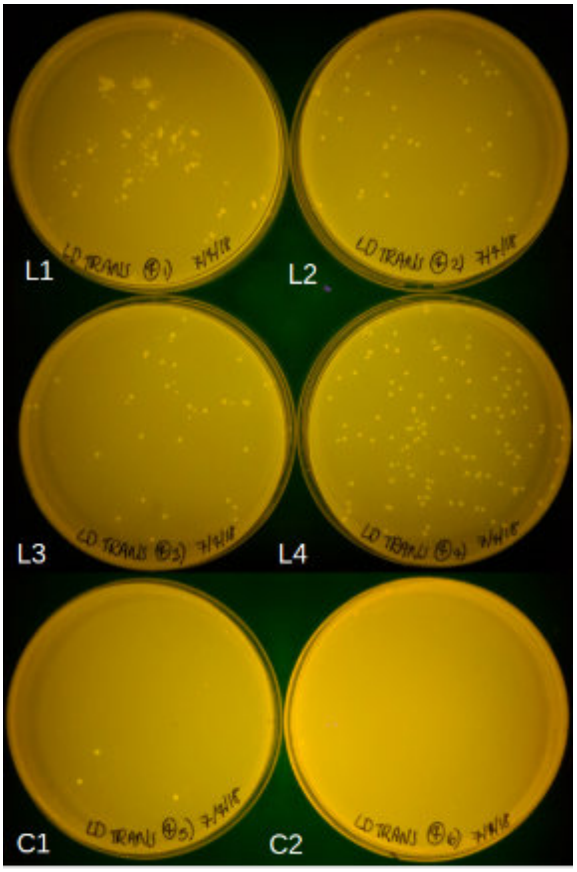

Figure 3.12: Transformed bacterial colonies (*E. coli* strain MC1061) with the chimeric *xylE-gusB*, variation #4. Plates labelled L1-L4 represent varying ligation ratios of pTTQ18 vector to PCR fusion amplicon insert, structural variation #4. Sample L1, ligation ratio 1:1; Sample L2, ligation ratio 1:2; Sample L3, ligation ratio 1:3; Sample L4, ligation ratio 2:1. Plates labelled C1 and C2 represent controls.

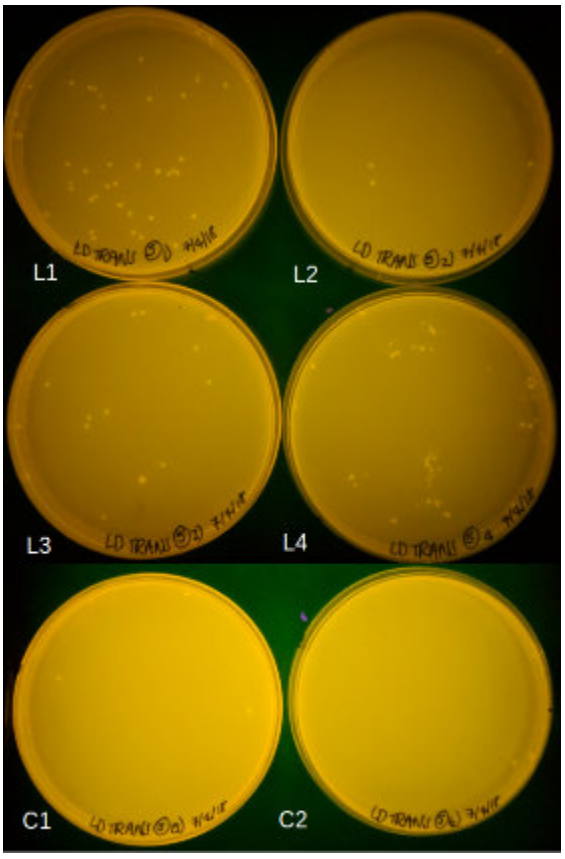

Figure 3.13: Transformed bacterial colonies (*E. coli* strain MC1061) with the chimeric *xylE-gusB*, variation #5. Plates labelled L1-L4 represent varying ligation ratios of pTTQ18 vector to PCR fusion amplicon insert, structurally variation #5. Sample L1, ligation ratio 1:1; Sample L2, ligation ratio 1:2; Sample L3, ligation ratio 1:3; Sample L4, ligation ratio 2:1. Plates labelled C1 and C2 represent controls.

3.9 Enzymatic assay of β-glucuronidase with X-Gluc

Colonies were extracted and assessed for transport functionality, as described in Sections 2.2.16 and 2.2.17. At 20 minutes, transport activity was slightly visible for clone 24C. No activity was seen for any of the other clones or positive controls (Figure 3.14). At 40 minutes, transport activity was observed for clone 24C, and for all positive controls (Figure 3.15). At this time, the intensity of the blue precipitate for clone 24C was greater than the positive controls (Figure 3.15). At 60 minutes, transport activity was observed for clones 24C, 31B, 32C, 33C and 51A, and for all

positive controls (Figure 3.16). At 80 minutes, transport activity was observed for clones 24C, 31B, 32C, 33C, 51A, and 51C, and for all positive controls (Figure 3.17). As the time course progressed, the intensity of blue precipitates became greater for the clones which exhibited transport activity (Figure 3.14; Figure 3.15; Figure 3.16; Figure 3.17).

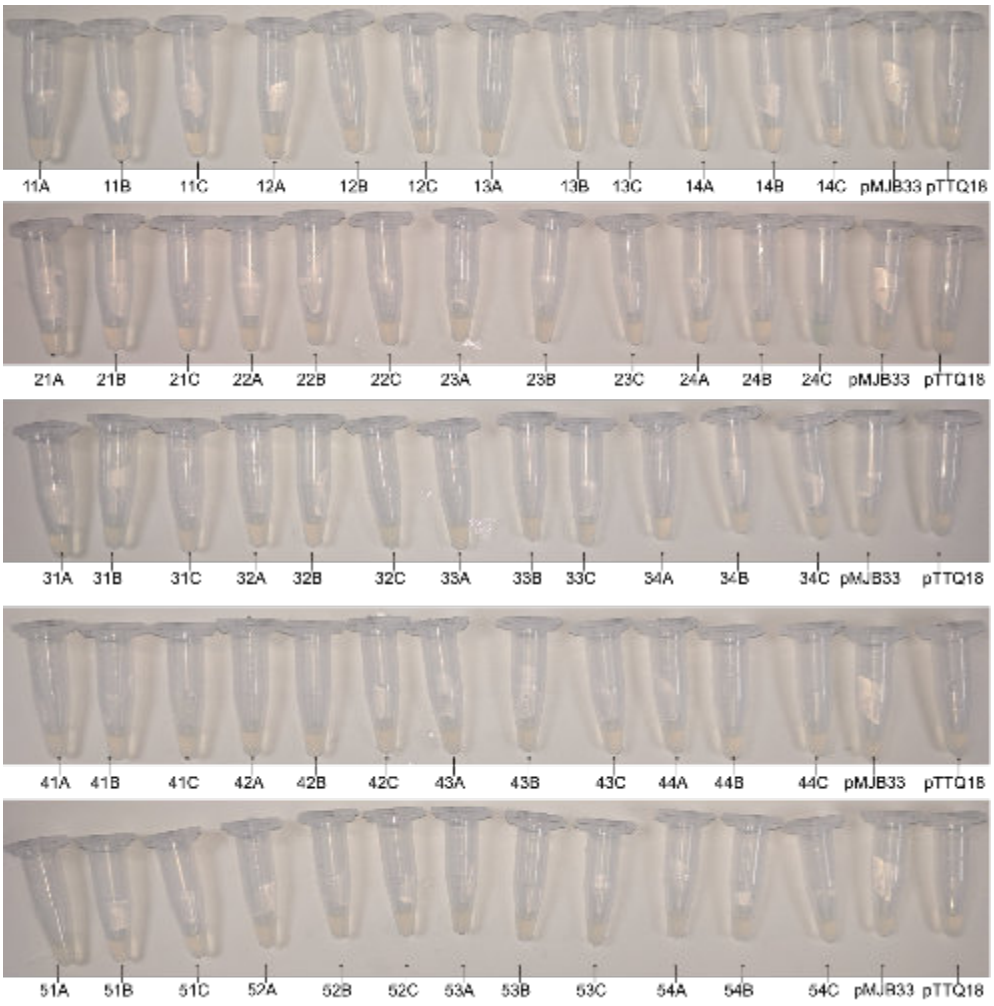

Figure 3.14: Results from the enzymatic assay of β-glucuronidase with X-Gluc in *E. coli* after a 20 minute incubation period. Rows 1-5 contain clones which potentially harbour a chimeric XylE-GusB of structural variations #1-#5, respectively. Positive control, pMJB33 and negative control, pTTQ18 were included within each row for visual reference. Slight transport activity was visible for clone 24C, as evident by the blue precipitate.

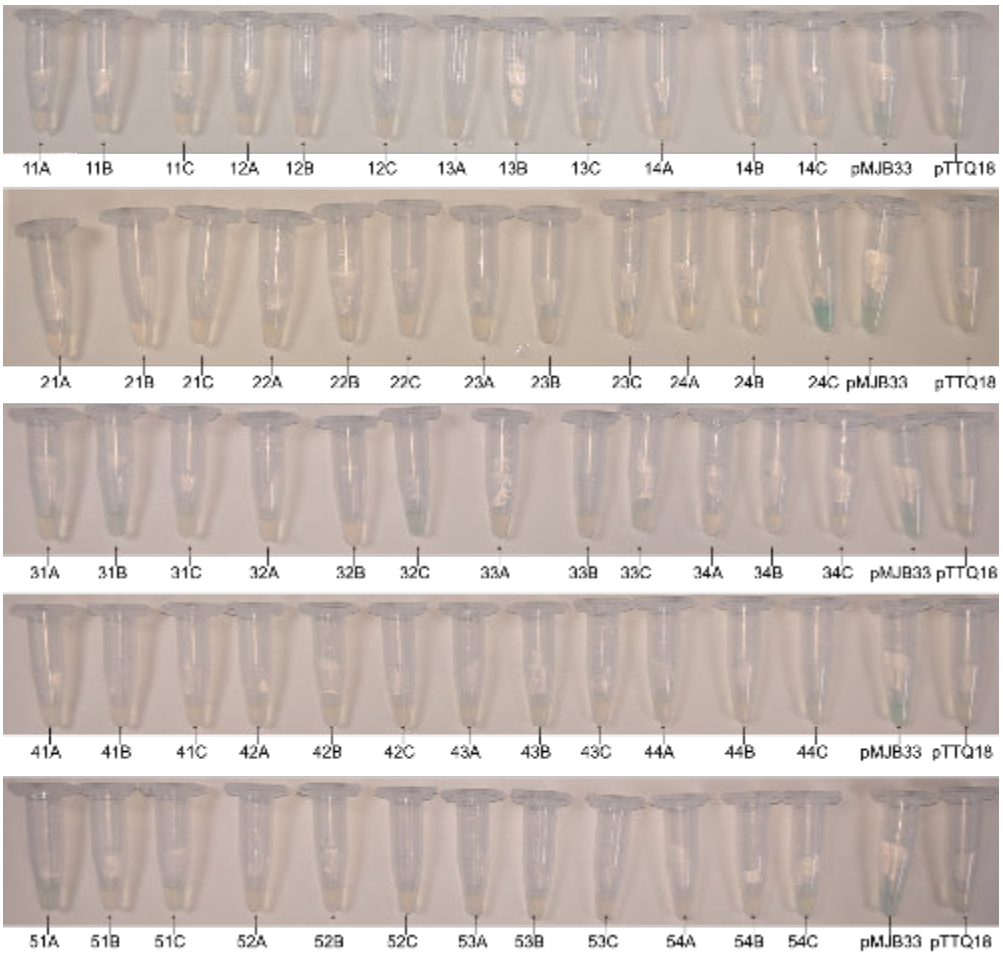

Figure 3.15: Results from the enzymatic assay of β-glucuronidase with X-Gluc in *E. coli* after a 40 minute incubation period. Rows 1-5 contain clones which potentially harbour a chimeric XylE-GusB of structural variations #1-#5, respectively. Positive control, pMJB33 and negative control, pTTQ18 were included within each row for visual reference. Transport activity was visible for clone 24C and all positive controls, as evident by the blue precipitates.

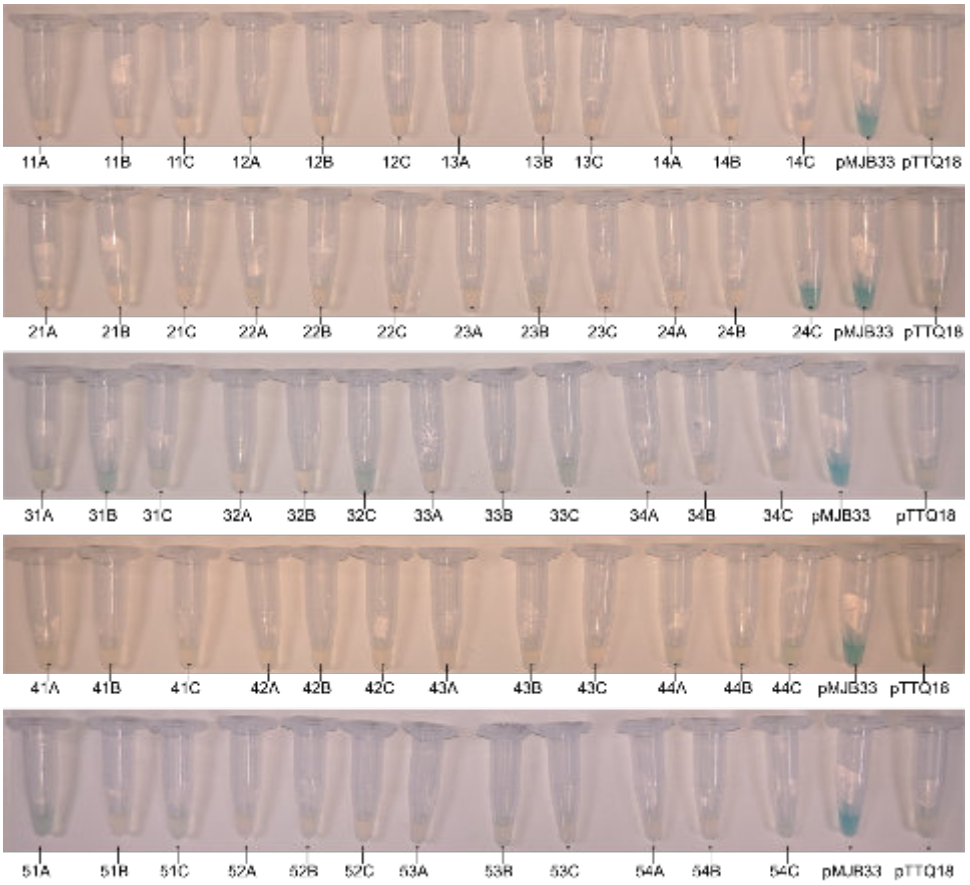

Figure 3.16: Results from the enzymatic assay of β-glucuronidase with X-Gluc in *E. coli* after a 60 minute incubation period. Rows 1-5 contain clones which potentially harbour a chimeric XylE-GusB of structural variations #1-#5, respectively. Positive control, pMJB33 and negative control, pTTQ18 were included within each row for visual reference. Transport activity was visible for clone 24C, 31B, 32C, 33C, 51A and all positive controls, as evident by the blue precipitates.

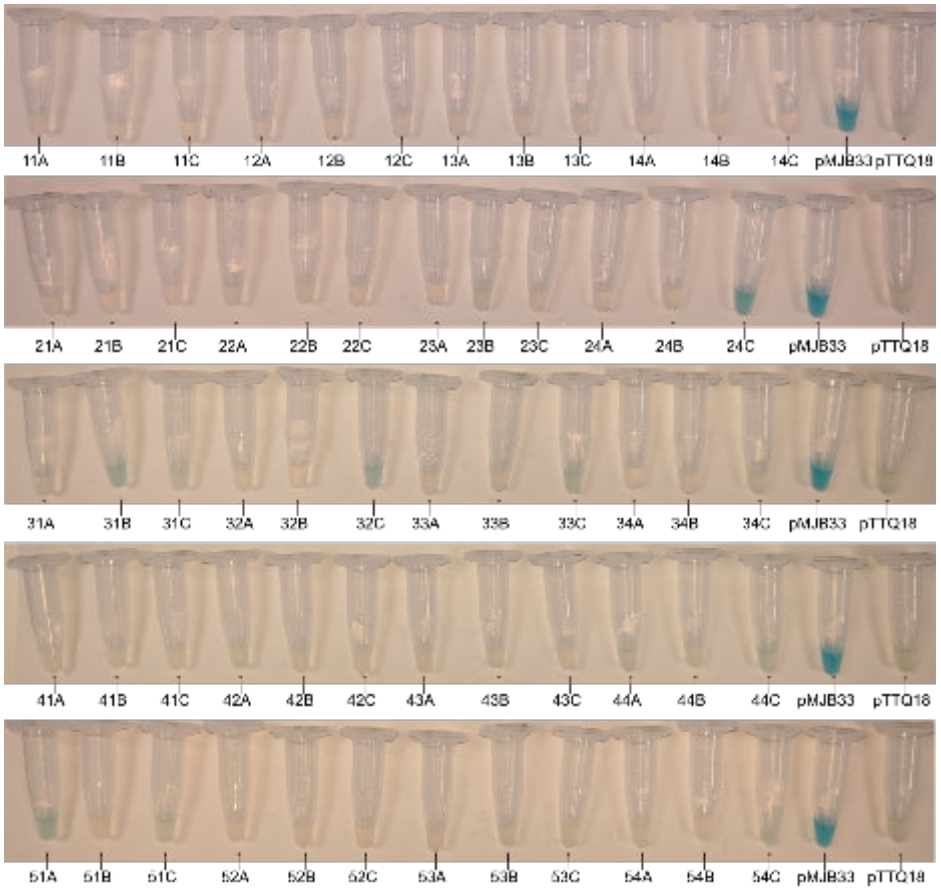

Figure 3.17: Results from the enzymatic assay of β-glucuronidase with X-Gluc in *E. coli* after an 80 minute incubation period. Rows 1-5 contain clones which potentially harbour a chimeric XylE-GusB of structural variations #1-#5, respectively. Positive control, pMJB33 and negative control, pTTQ18 were included within each row for visual reference. Transport activity was visible for clone 24C, 31B, 32C, 33C, 51A, 51C and all positive controls, as evident by the blue precipitates.

3.10 Characterisation of clones by restriction digestion

Clones were extracted and characterised by restriction mapping, as described in Sections 2.2.16 and 2.2.10.5. Overall, one clone was characterised as containing the chimeric *xylE-gusB* insert (Figure 3.19). The size of the three visualised bands matched the expected fragment sizes of a recombinant DNA with the *xylE-gusB* insert, as predicted through restriction mapping (Figure 3.19; Appendix 6). Of the 71 clones extracted, the majority were characterised as harbouring the background vector, pTTQ18 (Figure 3.18; Figure 3.19; Figure 3.20; Figure 3.21;

Figure 3.22). Some of the pTTQ18 vectors were not completely digested, and thus a mixture of uncut, partially cut and cut pTTQ18 can be seen (Figure 3.18; Figure 3.19; Figure 3.20; Figure 3.21; Figure 3.22). Four of the clones were characterised as harbouring the template plasmid, pMJB33, and four of the clones did not appear to contain plasmid DNA.

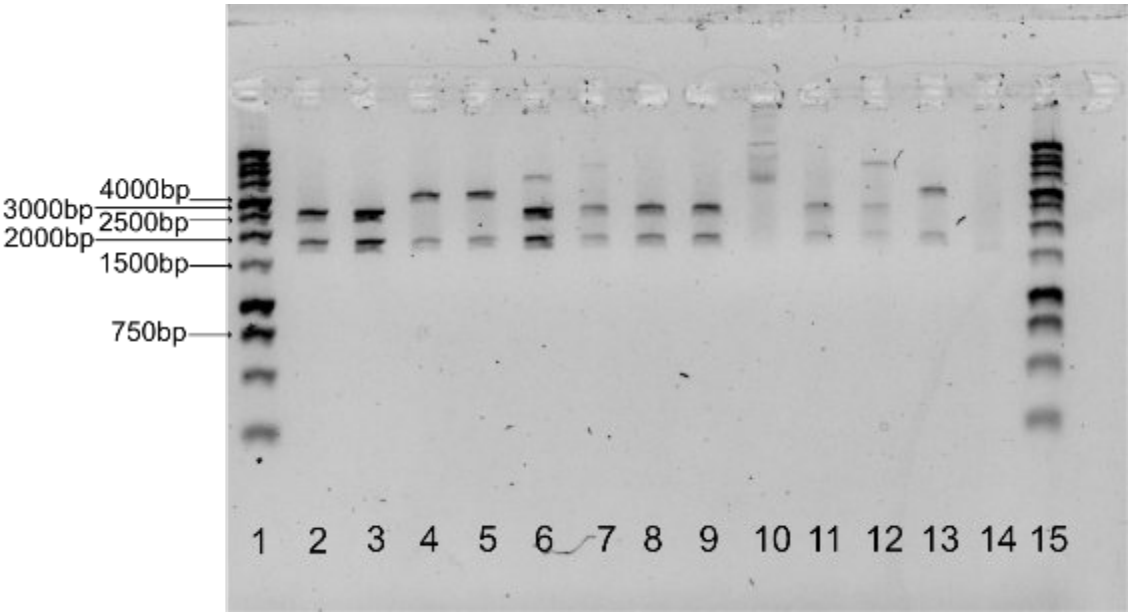

Figure 3.18: Restriction mapping of chimeric *xylE-gusB* mutants, structural variation #1, with *Eco*RV and *Bam*HI. Lanes 1 and 15, DNA ladder (catalogue number G5711); Lanes 2, 3, 8, 9 and 10, fully digested vector pTTQ18; Lanes 6, 7 and 12, digested vector pTTQ18 and uncut supercoiled pTTQ18; Lane 10, uncut nicked pTTQ18; Lanes 4, 5 and 13, plasmid pMJB33; Lane 14, absent of plasmid DNA. The colours within the DNA gel were inverted for clarity and to save ink.

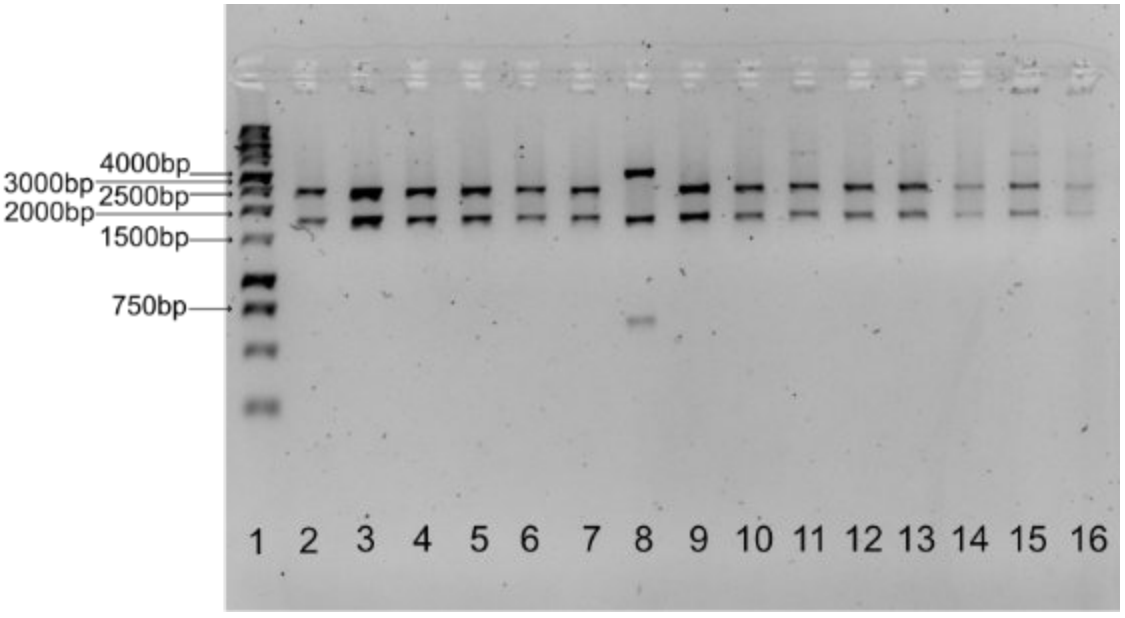

Figure 3.19: Restriction mapping of chimeric *xylE-gusB* mutants, structural variation #2, with *Eco*RV and *Bam*HI. Lane 1, DNA ladder (catalogue number G5711); Lanes 2, 3, 4, 5, 6, 7, 9, 10, 12, 13 and 14, fully digested vector pTTQ18; Lanes 11 and 15, digested vector pTTQ18 and uncut supercoiled pTTQ18; Lane 8, recombinant DNA with the *xylE-gusB* chimera. The colours within the DNA gel were inverted for clarity and to save ink.

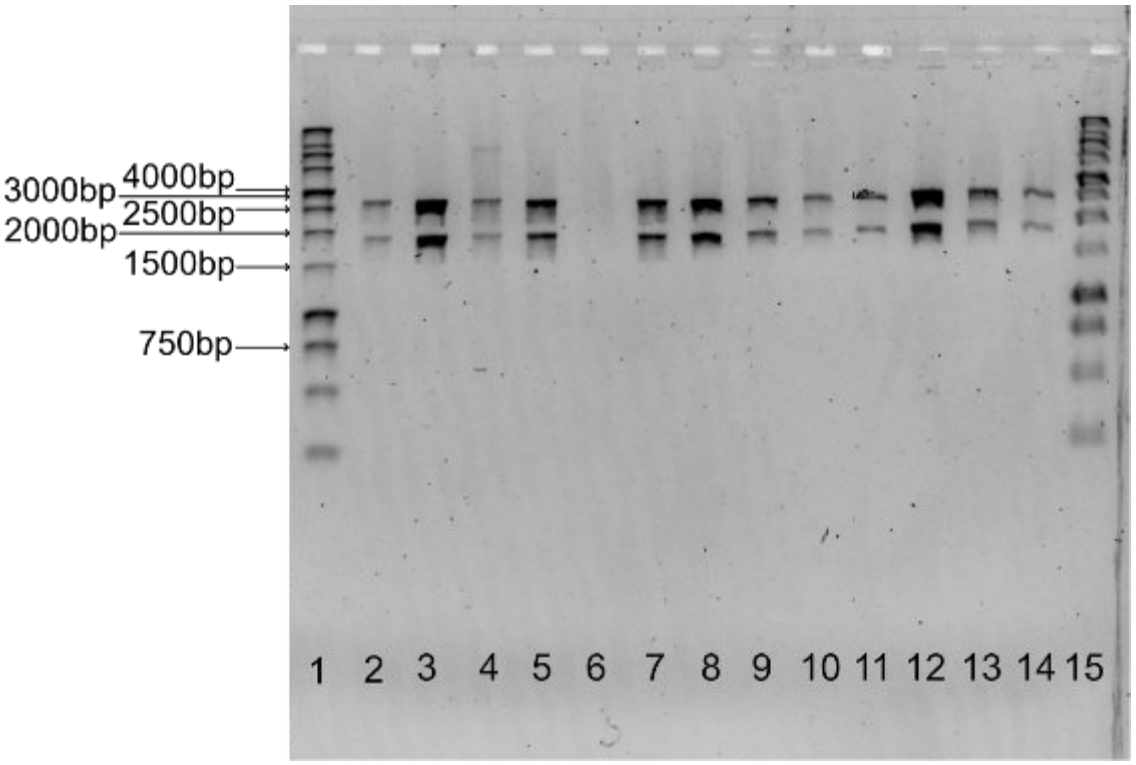

Figure 3.20: Restriction mapping of chimeric *xylE-gusB* mutants, structural variation #3, with *Eco*RV and *Bam*HI. Lanes 1 and 15, DNA ladder (catalogue number G5711); Lanes 2, 3, 5, 7, 8, 9 10, 11, 12, 13 and 14, fully digested vector pTTQ18; Lane 4, digested vector pTTQ18 and uncut supercoiled pTTQ18; Lane 6, absent of plasmid DNA. The colours within the DNA gel were inverted for clarity and to save ink.

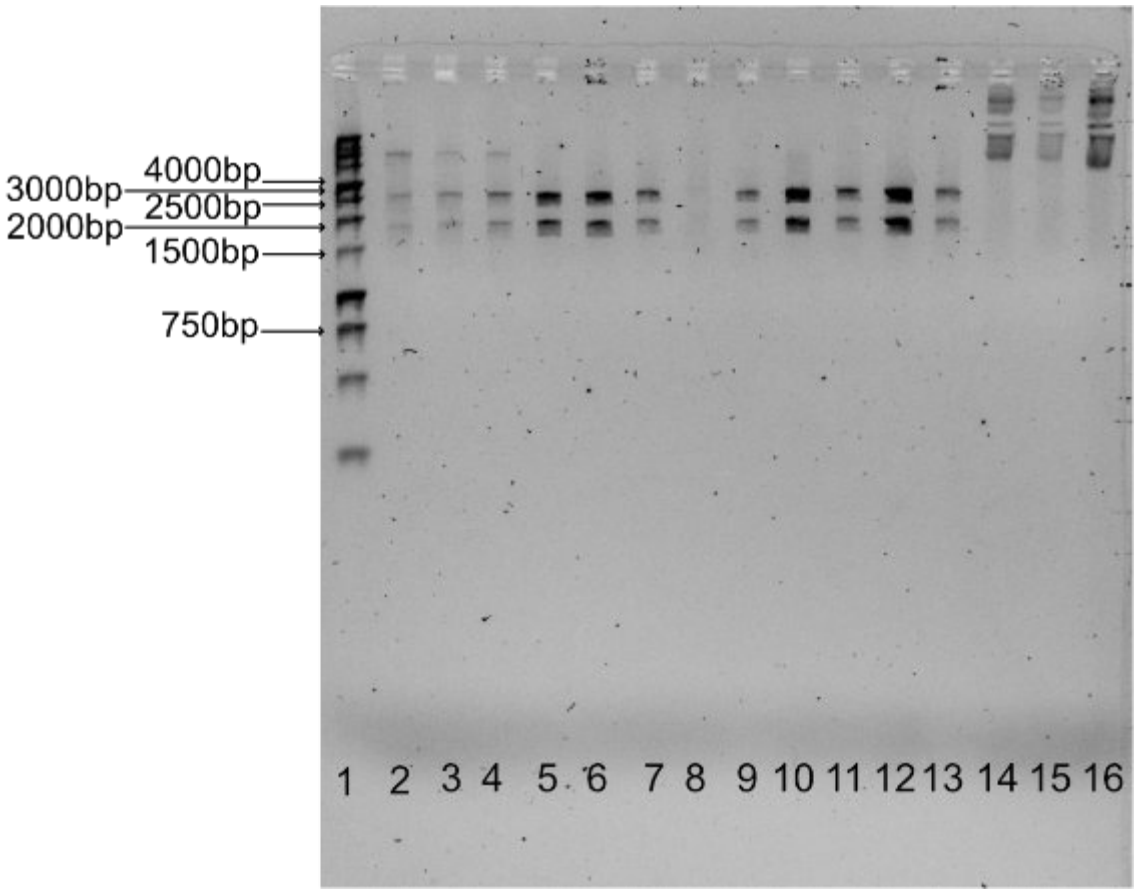

Figure 3.21: Restriction mapping of chimeric *xylE-gusB* mutants, structural variation #4, with *Eco*RV and *Bam*HI. Lane 1, DNA ladder (catalogue number G5711); Lanes 2, 3 and 4, digested vector pTTQ18 and uncut supercoiled pTTQ18; Lanes 5, 6, 7, 9 ,10, 11, 12 and 13, fully digested vector pTTQ18; Lanes 14, 15 and 16, uncut nicked pTTQ18; Lane 8, absent of plasmid DNA. The colours within the DNA gel were inverted for clarity and to save ink.

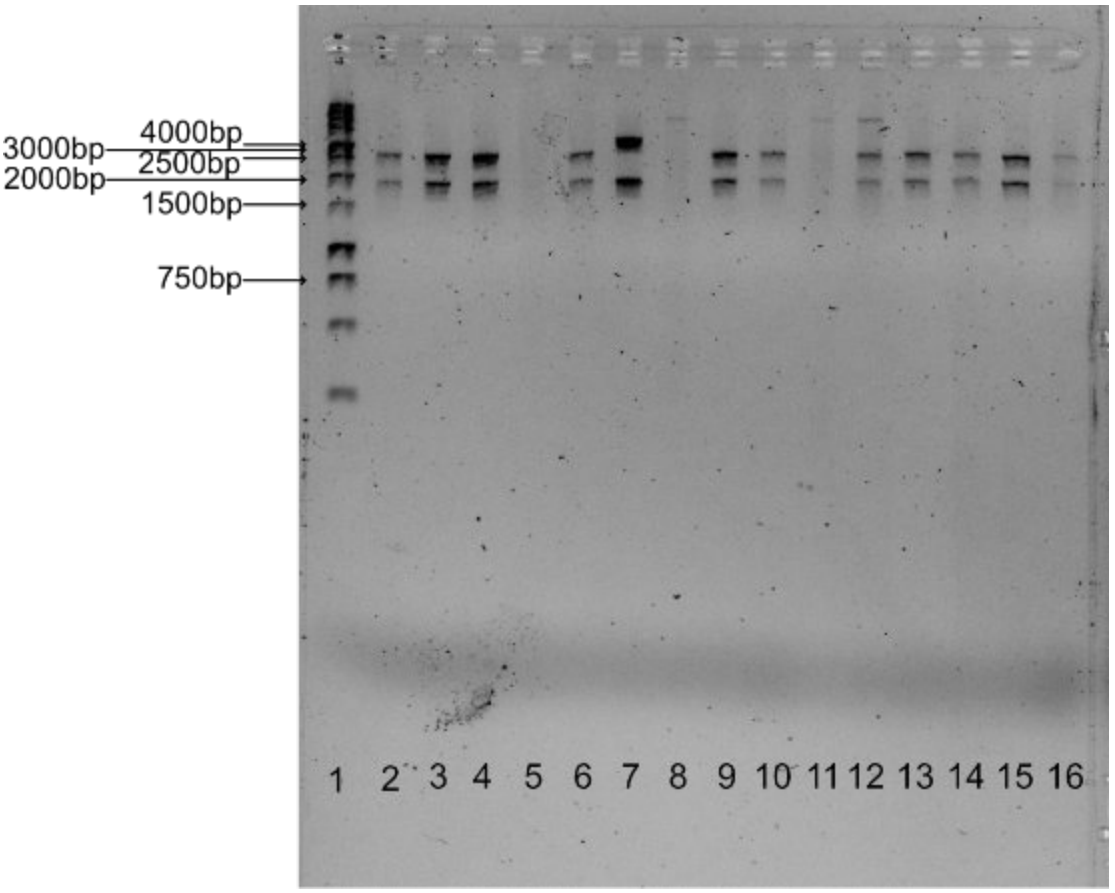

Figure 3.22: Restriction mapping of chimeric *xylE-gusB* mutants, structural variation #5, with *Eco*RV and *Bam*HI. Lane 1, DNA ladder (catalogue number G5711); Lanes 2, 3, 4, 6, 9, 10, 13, 14, 15 and 16, fully digested vector pTTQ18; Lane 12, digested vector pTTQ18 and uncut supercoiled pTTQ18; Lanes 8 and 11, uncut supercoiled pTTQ18; Lane 7, plasmid pMJB33; Lane 5, absent of plasmid DNA. The colours within the DNA gel were inverted for clarity and to save ink.

3.11 DNA sequencing

The plasmid DNA extracted from samples which exhibited transport activity, as evident from the visualised blue precipitate, were sent for sequencing. DNA sequencing revealed that these plasmids were the empty pTTQ18 vector. Despite not harbouring the chimeric *xylE-gusB* insert, DNA sequencing analysis revealed that the samples were of good quality with few ambiguities (See Appendix 10).

4. Discussion

The glucuronide transporter, GusB of *Escherichia coli* is critical for its survival within the gastrointestinal tract of vertebrates. *E. coli* scavenge a wide range of β-D-glucuronides, the major detoxification products of glucuronidation, as a carbon source (Liang 1992; Liang et al. 2005). After cleavage, aglycone moieties are released from *E. coli* and enter hepatic circulation, which can impact human health (Kim and Jin 2001; Rakoff-Nahoum et al. 2004; Arthur and Jobin 2011).

Within the MFS, GusB is predicted to share 12-TM helices and has been proven to utilise H^+ motive force for transport (Liang 1992; Liang et al. 2005). However, little is understood about its mechanisms of molecular recognition. Structural characterisation of the MFS H^+ symporter, XylE, has defined the first 6-TMs as the sites of H^+ coupling and translocation (Sun et al. 2012; Quistgaard et al. 2013; Wisedchaisri et al. 2014). As GusB belongs to the same family, it is logical to assume that its first 6-TM helices could also be responsible for H^+ translocation. To test this hypothesis, the first 6-TMs of XylE were fused with the last 6-TMs of GusB. Retained transport functionality would confirm that the first 6-TMs of GusB are responsible for H^+ translocation, and assist in further understanding its mechanisms of molecular recognition. This could be applied within the fields of healthcare and biotechnology.

Within this project, six out of the seven objectives were successfully met. However, at this time the aim cannot fully be answered. The N-domain of *xylE* and the C-domain of *gusB* were amplified and fused by PCR. The chimeric *xylE-gusB* mutant was constructed by molecular cloning and *E. coli* transformation. After transformation, many colonies were present for all 5 variations of the *xylE-gusB* chimera (Figure 3.9; Figure 3.10; Figure 3.11; Figure 3.12; Figure 3.13). Three clones were selected from each plate for characterisation by enzymatic assay and restriction mapping. However, only one of the extracted colonies was characterised as containing the *xylE-gusB* chimera (clone 23A). This can be

attributed to the high amount of background pTTQ18 and pMJB33 plasmids. Steps were taken throughout this project to favour the *xylE-gusB* containing recombinant DNA, however, they were not 100% effective.

Functional analysis of extracted colonies revealed increased glucuronide transport activity in several samples (Figure 3.14; Figure 3.15; Figure 3.16; Figure 3.17). DNA sequencing of these samples revealed the background pTTQ18 vector. pTTQ18 does not contain a *gusB* gene, however, this transport activity could be attributed to the genomic *gusB* gene. The genomic GusB protein in laboratory strain *E. coli* is normally less functional than in wild-type *E. coli* (Liang et al. 2005). However, X-Gluc is a large molecule which takes a long time to be transported across the membrane and cleaved by β-glucuronidases, and, due to its hydrophobic properties, it is also able to diffuse through the lipid bilayer given enough time (Figure 3.14) (Personal communication with Dr W.J. Liang, June 2018). Therefore, differences in the rates of transport between laboratory strain genomic GusB and wild-type GusB may be negligible during assays with X-Gluc. Furthermore, the amount of *E. coli* cells within each culture was not quantified or standardised prior to functionality tests, therefore increased transport in pTTQ18 is probably the result of a higher cell count, rather than a higher rate of transport. Restriction mapping characterisation of clone 23A infers that it contains the
xylE-gusB chimera (Figure 3.19). However, it did not exhibit any transport function by enzymatic assay of X-Gluc. These results need to be validated with other glucuronide molecules to ensure that the genetic manipulation did abolish glucuronide transport and rule out the possibility that it reduced glucuronide substrate specificity. At this time, it can be suggested that the domain swapping between XylE and GusB abolished X-Gluc transport ability.

This reduction in apparent transport ability could be attributed to multiple factors. Firstly, the loss of functionality of the XylE-GusB chimera could be due to PCR generated sequence mutations. 35 cycles were undertaken for PCR which is higher than the standard maximum of 30 cycles employed for molecular cloning (Maniatis et al. 1982). Since *Taq* DNA polymerase does not possess proofreading

capabilities (Huang et al. 1992; Kunkel 1992; Cline et al. 1996), the chances of PCR generated mutations increases with the number of cycles. A change in the DNA sequence could change the amino acid sequence to residues which are dissimilar in terms size, and side-chain hydropathy and charge, which could cause alterations in protein folding and function. Therefore, the presence of PCR generated mutations in the *xylE-gusB* chimera should be investigated with DNA sequencing.

Once this has been confirmed, other structural reasons for the lack of functionality should be investigated. One variation of the XylE-GusB chimera was successfully extracted and characterised, however, five different chimeras were designed, each fused at a different position within the domain linker. Clone 23A was fused at a position in the sequence directly after IC3 of XylE, and at a point within the domain linker of GusB which resides close to the C-domain. Previous functional assays of XylE have highlighted the importance of both the sequence identity and length of the intracellular helical linkers IC2 and IC3 for transport function (Figure 1.4) (Wisedchaisri et al. 2014). This fusion may have disrupted the stability of the protein, which is ordinarily provided by the GusB intracellular linker, and therefore abolished transport ability (Wisedchaisri et al. 2014). In addition, as the domain linker between this chimera is predominantly composed of XylE residues, important residues for molecular recognition which ordinarily reside within the domain linker of GusB may have been excised. Further, the fusion of two domains belonging to separate proteins may require extra flexibility to retain functionality. Therefore, the apparent loss of transport ability may be due to domain rigidity and the inability to change conformational states.

It cannot be ignored that the lack of observed transport activity might suggest that the first 6-TMs of GusB are not responsible for H^+ translocation like they are in XylE. Further functional characterisation of the XylE-GusB chimeras should be conducted to explore molecular recognition further.

4.1 Critical evaluation of the project

Primers were designed to amplify the separate domains of *xylE* and *gusB*, and fuse the separate amplicons together. This was successful, as evident from the generation of amplicons of desired lengths from both stages of PCR (Figure 3.2; Figure 3.3; Figure 3.4). However, non-specific amplification also occurred during both stages, which can be attributed to the non-specific annealing of primers. The high GC content of the *xylE* DNA sequence made it difficult to design primers of adequate length and sequence composition, whilst ensuring that their annealing temperatures remained within 5°C of each other. The primer sequence compositions were assessed to ensure that internal secondary structures wouldn't form, therefore the non-specific primer binding could be attributed to the lengths of fusion primers.

Primers are typically recommended to be between 18-30 nucleotides in length (Roux 1995; European Molecular Biology Laboratory 2018). Although the fusion primers were between 33-39 nucleotides in length, the sequences of complementarity to either *xylE* or *gusB* were between 9-15 nucleotides in length. Difficulty of fusion primer annealment is evident in the results from second-stage fusion PCR: the separate amplicons of *xylE* and *gusB* appear to have been preferentially amplified (signified by brighter bands) over the chimeric *xylE-gusB* amplicon (Figure 3.3; Figure 3.4). Although this decreased the yield of chimeric *xylE-gusB* amplicons, it was not detrimental to the project. The second-stage PCR fusion samples were purified which removed some of the non-specific amplicons (Figure 3.5), and the chimeric *xylE-gusB* fusion products were selected for by their compatible sticky ends created from restriction digestion.

The chimeric *xylE-gusB* amplicons were successfully cloned into the pTTQ18 vector. However, the proportion of extracted clones which contained the *xylE-gusB* chimera was low: only clone 23A was characterised by restriction mapping to contain the chimera (Figure 3.19). This high proportion of background pTTQ18 and

pMJB33 (Figure 3.18; Figure 3.19; Figure 3.20; Figure 3.21; Figure 3.22) indicates that the process was not efficient.

The presence of pMJB33 indicates that *Dpn*I treatment was not 100% effective. *Dpn*I cleaves DNA which has been methylated by DNA adenine methylase (Dam methylase) (Geier and Modrich 1979). Within *E. coli*, Dam methylation assists in regulating DNA replication and repair, and gene expression (Russell and Zinder 1987; Barras and Marinus 1989; Murphy et al. 2013). Cleavage of the Dam methylated sites allows for the PCR template DNA to be degraded after its use. This is especially important for transformation because *E. coli* will only take up one plasmid of the same origin of replication (Novick 1987; Maniatis et al. 1982). pMJB33 shares the same replication origin as the chimeric *xylE-gusB* containing recombinant DNA, and so it acts as competition during transformation. Therefore improved efficiency of *Dpn*I treatment would improve the success of the cloning process.

Clone characterisation by restriction digestion revealed that a high proportion of clones contained background vector pTTQ18 (Figure 3.18; Figure 3.19; Figure 3.20; Figure 3.21; Figure 3.22). Several steps were taken to reduce the proportion of pTTQ18 background transformants. These included rSAP treatment, spin-column purification and *Sph*I restriction digestion. rSAP treatment was employed to dephosphorylate the 5'- and 3'- ends of a linearised pTTQ18 vector to prevent religation (Olsen et al. 1991; Nilsen et al. 2001). Spin-column chromatography was employed to remove the multiple cloning site (MCS) which was excised during pTTQ18 digestion with *Eco*RI and *Hind*III (Maniatis et al. 1982). Since the sticky ends generated by *Eco*RI and *Hind*III are not compatible with each other, the removal of the MCS prevents religation of the vector. However, colonies were present on the transformation control plates which contained the linearised pTTQ18 and DNA ligase (C1) (Figure 3.12; Figure 3.13). This indicates that rSAP treatment and spin-column chromatography were not 100% effective. In addition, digestion of ligated samples with restriction endonuclease *Sph*I was employed to reduce pTTQ18 background transformants.

*Sph*I recognises a single restriction site within the MCS of pTTQ18, which is not present in the *xylE-gusB* chimeras. Therefore background pTTQ18 vectors (which had religated with the MCS, or had not been fully digested by *Eco*RI and *Hind*III), should have been digested and prevented from establishing during transformation. Due to the high proportion of extracted clones containing the pTTQ18 vector, the success of restriction digestion with *Sph*I was low, which may have resulted from not adding enough *Sph*I. Further, only one colony was present on a control plate which contained only the pTTQ18 vector (C2) (Figure 3.9), suggesting that digestion of pTTQ18 with *Eco*RI and *Hind*III was fairly efficient. Overall, increased efficiency of the aforementioned steps would reduce the proportion of background transformants, and facilitate the extraction of the chimeric *xylE-gusB* clones.

Some of the colonies extracted did not contain plasmid DNA (Figure 3.18; Figure 3.20; Figure 3.21; Figure 3.22). These colonies may be satellite colonies. pTTQ18, pMJB33 and the chimeric *xylE-gusB* recombinant DNA possess the ampicillin resistance gene, *bla*. *E. coli* harbouring plasmids containing *bla* secrete the β-lactamase enzyme, which depletes surrounding antibiotics. This may have permitted cells which do not possess resistance to cultivate (Korpimaki et al. 2003; Peubez et al. 2010). Shortening the overnight incubation period after transformation should reduce the likelihood of ampicillin depletion and therefore satellite colonies.

Sequence analysis revealed few ambiguities within the extracted DNA sequences (Appendix 10). This suggests that the protocols employed for DNA purification and extraction were adequate.

4.2 Troubleshooting within the project

Many first-stage PCR troubleshoots were performed before *xylE* and *gusB* amplicons were successfully generated (Appendix 11). Prior to performing

first-stage PCR it was recognised that difficulties may occur regarding primer binding specificities. Therefore, Touchdown PCR (TD-PCR) was employed.

TD-PCR starts at an annealing temperature above the predicted primer annealing temperatures, which is gradually lowered over the course of successive cycles (Roux 1995; Korbie and Mattick 2008). This transition favours primer-template annealing of high complementarity and creates a twofold advantage per cycle for these amplicons (Korbie and Mattick 2008). The desired amplicons should outcompete non-specific products which are created as the temperature is decreased. However, despite multiple troubleshoots TD-PCR was only able to generate a maximum six *xylE* or *gusB* amplicons. Of these, only four amplicons contributed matching pairs for a designed *xylE-gusB* fusion product. Multiple PCR variables were altered in an attempt to achieve successful amplification. These included: altering temperature steps between successive cycles; altering the total number of successive cycles; and alterations in the duration of both primer annealing and *Taq* polymerisation. At this point, it was suspected that the primer annealing temperatures calculated by the manufacturer were too high, and thus conventional PCR was performed at a lower temperature. However, the results from these reactions were equally as poor; the proportion of successful *xylE* and *gusB* amplicons was extremely low, and the number of non-specific amplification products was high.

After careful deliberation over the variables employed throughout these troubleshoots, a mistake was noticed for the concentration of template DNA used. When performing PCR reactions a final concentration of 5 ng/µl of template DNA within a 50 µl reaction is required for amplification (Roux 1995). However, in the previous reactions a total of 5 ng of template DNA was added into a 50 µl reaction, creating a final template DNA concentration of 0.1 ng/µl. This mistake was rectified, and successful amplicons were generated for all *xylE* and *gusB* primer pairs (Figure 3.2).

4.3 Future work

For future work, clone 23A should be sequenced to confirm that it contains the *xylE-gusB* chimera and assessed for PCR generated mutations. Functionality assays using other chromogenic glucuronides should be employed to ensure that the loss in functionality is not substrate specific.

In addition to this, other *xylE-gusB* chimera containing transformants need to be identified and characterised for function. This may be difficult due to the high amount of background transformants, therefore many colonies will need to be extracted to maximise the chances of extracting a *xylE-gusB* mutant clone. Enzymatic assays using glucuronide molecules which can differentiate between the wild-type GusB and genomic GusB should then be employed. Further, the *E. coli* cell count should be standardised prior to assay, to ensure that any differences visualised are comparable between samples. Enzymatic assays of chromogenic substrates, such as para-Nitro-phenyl-β-D-glucuronide, should also be employed to measure changes in the rate of glucuronide transport.

Finally, primers were designed during this project which incorporated *gfp* between the separate domains of *xylE* and *gusB* (Appendix 9). The GFP protein would have acted as a selection marker to facilitate colony extraction. Due to time constraints, these primers were not utilised within this project. Future work should utilise these primers to maximise the likelihood of extracting a mutant *xylE-gusB* chimera. In addition, the incorporation of the GFP between the N- and C- domains of the chimeric XylE-GusB protein would provide flexibility between the two domains, which may have been a limiting factor for transport functionality.

5. Conclusion

This research project was designed to locate the H^+ translocation domain of GusB. Six out of seven of the objectives of this project were achieved. Primers were successfully designed, and *xylE-gusB* chimeras were constructed and cloned into the vector pTTQ18. The functionality of the clones was assessed, and their identities were characterised through restriction mapping and DNA sequencing. The research achievements so far cannot completely confirm if the H^+ translocation site is located in the first 6-TM helices. However, so far one clone (clone 23A) can be further examined for its precise molecular mechanisms. For future work, many of the other created XylE-GusB chimeras should be examined. This research highlighted some of the challenges in molecular cloning and studying molecular mechanisms. Throughout this research, troubleshooting was routinely conducted which shows that these objectives can be achieved in a fully optimal laboratory set up. The strategy was proved to be correct, however, the efficiency of removing background transformants needs to be further refined. If clone 23A is indeed a XylE-GusB chimera, this research can direct further characterisation of the H^+ recognition domain. Future work which needs to be conducted and the significance of this work is highlighted.

6. References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. and Iwata, S., 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. *Science*, 301 (5633), 610-615.

Abramson, J., Kaback, H. R. and Iwata, S., 2004. Structural comparison of lactose permease and the glycerol-3-phosphate antiporter: members of the major facilitator superfamily. *Current Opinions in Structural Biology*, 14 (4), 413-419.

Alkharfy, K. M., Jan, B. L., Afzal, S., Al-Jenoobi, F. I., Al-Mohizea, A. M., Al-Muhsen, S., Halwani, R., Parvez, M. H. and Al-Dosari, M. S., 2017. Prevalence of UDP-glucuronosyltransferase polymorphisms (UGT1A6*2, 1A7*12, 1A8*3, 1A9*3, 2B7*2, and 2B15*2) in a Saudi population. *Saudi Pharmaceutical Journal*, 25 (2), 224-230.

Arias, I. M., Lawrence, M. G., Seifter, S. and Furman, M., 1964. Prolonged Neonatal Unconjugated Hyperbilirubinemia Associated with Breast Feeding and a Steroid, Pregnane3 (Alpha), 20 (Beta) -Diol, in Maternal Milk That Inhibits Glucuronide Formation In Vitro. *Journal of Clinical Investigation*, 43 (11).

Arthur, J. C. and Jobin, C., 2011. The struggle within: microbial influences on colorectal cancer. *Inflammatory Bowel Diseases*, 17 (1), 396-409.

Axelrod, J., Inscoe, J., Tomkins, G. M., 1958. Enzymatic synthesis of N-glucosyluronic acid conjugates. *Journal of Biological Chemistry*, 232, 835–851.

Bachmann, B. J., 1983. Linkage map of *Escherichia coli* K-12, edition 7. *Microbiological Reviews*, 47 (2), 180-230.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., Semenkovich, C. F. and Gordon, J. I., 2004. The gut microbiota as an

environmental factor that regulates fat storage. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (44), 15718–15723.

Barras, F. and Marinus, M. G., 1989. The great GATC: DNA methylation in *E. coli*. *Trends in Genetics*, 5, 139-143.

Basu, C., Kausch, A. P. and Chandlee, J. M., 2004. Use of beta-glucuronidase reporter gene for gene expression analysis in turfgrasses. *Biochemical and Biophysical Research Communications*, 320 (1), 7-10.

Beckwith, J. R., 1967. Regulation of the Lac Operon. *Science*, 156 (3775), 597-604.

Bélanger, A., Hum, D. W., Beaulieu, M., Lévesque, E., Guillemette, C., Tchernof, A., Bélanger, G., Turgeon, D. and Dubois, S., 1998. Characterization and regulation of UDP-glucuronosyltransferases in steroid target tissues. *The Journal of Steroid Biochemistry and Molecular Biology*, 65, 301-310.

Bidart, G. N., Rodríguez-Díaz, J., Pérez-Martínez, G. and Yebra, M. J., 2018. The lactose operon from *Lactobacillus casei* is involved in the transport and metabolism of the human milk oligosaccharide core-2 N-acetyllactosamine. *Scientific Reports*, 8 (1), 7152.

Billing, B. H., Cole, P. G. and Lathe, G. H., 1957. The excretion of bilirubin as a diglucuronide giving the direct van den Bergh reaction. *Biochemical Journal*, 65 (4), 774-784.

Birch, J., Axford, D., Foadi, J., Meyer, A., Eckhardt, A., Thielmann, Y. and Moraes. I., 2018. The fine art of integral membrane protein crystallisation. *Methods*, In press. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29778646 [Accessedd 11 June 2018].

Black, M. and Billing, B. H., 1969. Hepatic Bilirubin UDP-Glucuronyl Transferase Activity in Liver Disease and Gilbert's Syndrome. *The New England Journal of Medicine*, 280, 1266-1271

Black, M., Billing, B. H. and Heirwegh, K. P. M., 1970. Determination of bilirubin UDP-glucuronyl transferase activity in needle-biopsy specimens of human liver. *Clinica Chimica Acta*, 29, 27-35.

Blanco, C., Ritzenthaler, P. and Mata-Gilsinger, M., 1982. Cloning and endonuclease restriction analysis of *uidA* and *uidR* genes in *Escherichia coli* K-12: determination of transcription direction for the *uidA* gene. *Journal of Bacteriology*, 149 (2), 587-594.

Blanco, C., Mata-Gilsinger, M. and Ritzenthaler, P., 1985. The use of gene fusions to study the expression of *uidR*, a negative regulatory gene of *Escherichia coli* K-12. *Gene*, 36 (2), 159-167.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y., 1997. The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277, 1453-1474.

Bossuyt, X. and Blanckaert, N., 1997. Carrier-mediated transport of uridine diphosphoglucuronic acid across the endoplasmic reticulum membrane is a prerequisite for UDP-glucuronosyltransferase activity in rat liver. *Biochemical Journal*, 323, 645-648.

Bryksin, A. V. and Matsumura, I., 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques*, 48 (6), 463-465.

Burchell, B. and Coughtrie, M. W. H., 1989. UDP-Glucuronosyltransferases. *Pharmacology and Therapeutics*, 43, 261-289.

Carbone, J. V. and Grodsky, G. M., 1957. Constitutional Nonhemolytic Hyperbilirubinemia in the Rat Defect of Bilirubin Conjugation. *Experimental Biology and Medicine*, 94 (3), 461-463.

Carpenter, E. P., Beis, K., Cameron, A. D. and Iwata, S., 2008. Overcoming the challenges of membrane protein crystallography. *Current Opinions in Structural Biology*, 18 (5), 581-586.

Chen, D. E., Podell, S., Sauer, J. D., Swanson, M. S. and Saier, M. H. Jr., 2008. The phagosomal nutrient transporter (Pht) family. *Microbiology*, 154 (1), 42-53.

Cline, J., Braman, J. C. and Hogrefe, H. H., 1996. PCR Fidelity of *Pfu* DNA Polymerase and Other Thermostable DNA Polymerases. *Nucleic Acid Research*, 24 (18), 3546-3551.

Conte, S. S. and Walker, E. L., 2012. Genetic and biochemical approaches for studying the yellow stripe-like transporter family in plants. *Current Topics in Membranes*, 69, 295-322.

Davies, J. and Jacob, F., 1968. Genetic Mapping of the Regulator and Operator Genes of the Lac Operon. *Journal of Molecular Biology*, 36 (3), 413-417.

Davis, E. O., Jones-Mortimer, M. C. and Henderson, P. J. F., 1984. Location of a Structural Gene for Xylose-H+ Symport at 91 Min on the Linkage Map of *Escherichia coli* K12. *The Journal of Biological Chemistry*, 259 (3), 1520-1525.

Davis, E. O., Henderson, P. J., 1987. The cloning and DNA sequence of the gene *xylE* for xylose-proton symport in *Escherichia coli* K12. *Journal of Biological Chemistry*, 262 (29), 13928-13932.

Deng, D., Sun, P., Yan, C., Ke, M., Jiang, X., Xiong, L., Ren, W., Hirata, K., Yamamoto, M., Fan, S. and Yan, N., 2015. Molecular basis of ligand recognition and transport by glucose transporters. *Nature*, 526 (7573), 391-396.

Deng, D. and Yan, N., 2016. GLUT, SGLT, and SWEET: Structural and mechanistic investigations of the glucose transporters. *Protein Science*, 25 (3), 546-558.

Deng, D. and Yan, N., 2018. Crystallization and Structural Determination of the Human Glucose Transporters GLUT1 and GLUT3. *Methods in Molecular Biology*, 1713, 15-29.

Doyle, M. L., Katzman, P. A. and Doisy, E. A., 1955. Production and properties of bacterial beta-glucuronidase. *Journal of Biological Chemistry*, 217 (2), 921-930.

Dutton, G. J. and Illing, H. P. A., 1972. Mechanism of Biosynthesis of Thio-p-D-glucuronides and Thio-p-D-glucosides. *Biochemical Journal*, 129, 539-550.

Dutton, G. J. and Storey, I. D. E., 1953. The isolation of a compound of uridine diphosphate and glucuronic acid from liver. *Journal of Biochemistry*, 53, 37–38.

Dutton, G. J. and Storey, I. D. E., 1954. Uridine compounds in glucuronic acid metabolism. 1. The formation of glucuronides in liver suspensions. *Biochemical Journal*, 57 (2), 275-283.

Egger, S., Chaikuad, A., Kavanagh, K. L., Oppermann, U. and Nidetzky, B., 2010. UDP-glucose dehydrogenase: structure and function of a potential drug target. *Biochemical Society Transactions*, 38 (5), 1378-1385.

Ethayathulla, A. S., Yousef, M. S., Amin, A., Leblanc, G., Kaback, H. R. and Guan, L., 2014. Structure-based mechanism for Na(+)/melibiose symport by MelB. *Nature Communication*, 5 (3009).

European Molecular Biology Laboratory., 2018. *Cloning - PCR Strategy - Primer Design* [online], Heidelberg: European Molecular Biology Laboratory. Available from: https://www.embl.de/pepcore/pepcore_services/cloning/pcr_strategy/primer_design/ [Accessed 25 May 2018].

Fretzayas, A., Moustaki, M., Liapi, O. and Karpathios, T., 2012. Gilbert syndrome. *European Journal of Pediatrics*, 171 (1), 11-15.

Geier, G. E. and Modrich, P.,1979. Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of *Dpn*I endonuclease. *Journal of Biological Chemistry*, 254 (4), 1408-1413.

Goffeau, A., Park, J., Paulsen, I. T., Jonniaux, J. L., Dinh, T., Mordant, P. and Saier, M. H. Jr., 1997. Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames with the major facilitator superfamily. *Yeast*, 13 (1), 43-54.

Grisshammer, R. and Tate, C. G., 1995. Overexpression of integral membrane proteins for structural studies. *Quarterly Reviews of Biophysics*, 28 (3), 315-422.

Guillemette, C., 2003. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *The Pharmacogenomics Journal*, 3, 136-158.

Harwood J, P. and Peterkofsky, A., 1975. Glucose-sensitive adenylate cyclase in toluene-treated cells of *Escherichia coli* B. *The Journal of Biological Chemistry*, 250 (12), 4656-4662.

Henderson, P. J. and Maiden, M. C., 1990. Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 326 (1236), 391-410.

Hobert, O., 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic *C. elegans*. *Biotechniques*, 32 (4), 728-730.

Hoffmann, D., Adams, J. D., Brunnemann, K. D. and Hecht, S. S., 1979. Assessment of Tobacco-specific N-Nitrosamines in Tobacco Products. *Cancer Research*, 39 (7), 2502-2509.

Hogema, B. M., Arents, J. C., Bader, R. and Postma, P. W., 1999. Autoregulation of lactose uptake through the LacY permease by enzyme IIA^{Glc} of the PTS in *Escherichia coli* K-12. *Molecular Microbiology*, 31 (6), 1825-1833.

Huang, M-M., Arnheim, N. and Goodman, M. F., 1992. Extension of base mispairs by *Taq* DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Research*, 20 (17), 4567-4573.

Huang, Y., Lemieux, M. J., Song, J., Auer, M. and Wang, D. N., 2003. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science*, 301 (5633), 616-620.

Huang, J., Pang, J., Watanabe, T., Ng, H. K. and Ohgaki, H., 2009. Whole genome amplification for array comparative genomic hybridization using DNA extracted from formalin-fixed, paraffin-embedded histological sections. *Journal of Molecular Diagnostics*, 11 (2), 109-116.

Humblot, C., Murkovic, M., Rigottier-Gois, L., Bensaada, M., Bouclet, A., Andrieux, C., Anba, J. and Rabot, S., 2007. Beta-glucuronidase in human intestinal microbiota is necessary for the colonic genotoxicity of the food-borne carcinogen 2-amino-3-methylimidazo[4,5-f]quinoline in rats. *Carcinogenesis*, 28 (11), 2419-2425.

Jacob, F., Perrin, D., Sanchez, C. and Monod, J., 1960. The Operon: A Group of Genes Whose Expression is Coordinated by an Operator. *Comptes Rendus*

Hebdomadaires des Séances de l'Académie des Sciences, 250 (6), 1727–1729.

Jardetzky, O., 1966. Simple allosteric model for membrane pumps. *Nature*, 211 (5052), 969-970.

Jefferson, R. A., Burgess, S. M. and Hirsh, D., 1986. Beta-Glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proceedings of the National Academy of Sciences of the United States of America*, 83 (22), 8447-8451.

Jefferson, R. A., Kavanagh, T. A. and Bevan, M. W., 1987. GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal*, 6 (13), 3901-3907.

Jefferson, R. A., 1989. The GUS reporter gene system. *Nature*, 342 (6251), 837-838.

Jude, A. R., Little, J. M., Bull, A. W., Podgorski, I. and Radominska-Pandya, A., 2001. 13-Hydroxy- and 13-Oxooctadecadienoic acids: Novel Substrates for Human UDP-Glucuronosyltransferases. *Drug Metabolism and Disposition*, 29 (5), 652-655.

Kalckar, H. M., Maxwell, E. S. and Strominger, J. L., 1956. Some properties of uridine diphosphoglucose dehydrogenase. *Archives of Biochemistry and Biophysics*, 65, 2–10.

Kallionpää, R. A., Järvinen, E. and Finel, M., 2015. Glucuronidation of estrone and 16α-hydroxyestrone by human UGT enzymes: The key roles of UGT1A10 and UGT2B7. *The Journal of Steroid Biochemistry and Molecular Biology*, 154, 104-111.

Kang, H. J., Lee, C. and Drew, D., 2013. Breaking the barriers in membrane protein crystallography. *The International Journal of Biochemistry and Cell Biology*, 45 (3), 636-644.

Kim, D. H. and Jin, Y. H., 2001. Intestinal bacterial beta-glucuronidase activity of patients with colon cancer. *Archives of Pharmacal Research*, 24 (6), 564-567.

Kolkhof, P., 1992. Specificities of three tight-binding Lac repressors. *Nucleic Acids Research*, 20 (19), 5035-5039.

Kolb, A., Busby, S., Buc, H., Garges, S. and Adhya, S., 1993. Transcriptional Regulation by cAMP and Its Receptor Protein. *Annual Review of Biochemistry*, 62, 749-797.

König, J., Nies, A. T., Cui, Y., Leier, I. and Keppler, D., 1999. Conjugate export pumps of the multidrug resistance protein (MRP) family: localization, substrate specificity, and MRP2-mediated drug resistance. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1461 (2), 377–394.

Korbie, D. J. and Mattick, J. S., 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, 3 (9), 1452-1456.

Korpimaki, T., Kurittu, J., Karp, M., 2003. Surprisingly fast disappearance of beta-lactam selection pressure in cultivation as detected with novel biosensing approaches. *Journal of Microbiological Methods*, 53 (1), 37-42.

Kumar, H., Kasho, V., Smirnova, I., Finer-Moore, J. S., Kaback, H. R. and Stroud, R. M., 2014. Structure of sugar-bound LacY. *Proceedings of the National Academy of Sciences of the United States of America*. 111 (5), 1784-1788.

Kunkel, T. A., 1992. DNA Replication Fidelity. *The Journal of Biological Chemistry*, 267 (26), 18251-18254.

Kuroda, T. and Tsuchiya, T., 2009. Multidrug efflux transporters in the MATE family. *Biochimica et Biophysica Acta*, 1794 (5), 763-768.

Kwok, C. T., Ward, R. L., Hawkins, N. J. and Hitchins, M. P., 2010. Detection of allelic imbalance in *MLH1* expression by pyrosequencing serves as a tool for

the identification of germline defects in Lynch syndrome. *Familial Cancer*, 9 (3), 345-356.

Lam, V. M., Daruwalla, K. R., Henderson, P. J., Jones-Mortimer, M. C., 1980. Proton-linked D-xylose transport in *Escherichia coli*. *Journal of Bacteriology*, 143 (1), 396-402.

Lautala, P., Ethell, B. T., Taskinen, J. and Burchell, B., 2000. The specificity of glucuronidation of entacapone and tolcapone by recombinant human UDP-glucuronosyltransferases. *Drug Metabolism and Disposition*, 28 (11), 1385-1389.

Lawson, C. L., Swigon, D., Murakami, K. S., Darst, S. A., Berman, H. M. and Ebright, R. H., 2004. Catabolite activator protein: DNA binding and transcription activation. *Current Opinion in Structural Biology*, 14, 10-20.

Leonard, S. R., Lacher, D. W. and Lampel, K. A., 2015. Acquisition of the *lac* operon by *Salmonella enterica*. *BMC Microbiology*, 15 (173).

Lester, R. and Klein, P. D., 1966. Bile pigment excretion: a comparison of the biliary excretion of bilirubin and bilirubin derivatives. *The Journal of Clinical Investigation*, 45 (12), 1839-1846.

Levvy, G. A., 1956. Glucuronide Metabolism, with Special Reference to the Steroid Hormones. *Vitamins and Hormones*, 14, 267-303.

Liang, W-J., 1992. The Glucuronide Transport System of *Escherichia coli*. Thesis (PhD). University of Cambridge.

Liang, W-J., Wilson, K. J., Xie, H., Knol, J., Suzuki, S., Rutherford, N. G., Henderson, P. J. F., Jefferson, R. A., 2005. The gusBC Genes of *Escherichia coli* Encode a Glucuronide Transport System. *Journal of Bacteriology*, 187, (7), 2377-2385.

Lin, M., Liu, J., Zhou, H., Wu, M., Lv, D., Huang, Y., Zheng, Y., Shentu, J. and Wu, L., 2017. Effects of UDP-glucuronosyltransferase (UGT) polymorphisms on the pharmacokinetics of febuxostat in healthy Chinese volunteers. *Drug Metabolism and Pharmacokinetics*, 32 (1), 77-84.

Lorca, G. L., Barabote, R. D., Zlotopolski, V., Tran, C., Winnen, B., Hvorup, R. N., Stonestrom, A. J., Nguyen, E., Huang, L. W., Kim, D. S. and Saier, M. H. Jr., 2007. Transport capabilities of eleven gram-positive bacteria: comparative genomic analyses. *Biochimica et Biophysica Acta*, 1768 (6), 1342-1366.

Ishii, N., 2010. Investigation on stability of transporter protein, glucuronide transporter from *Escherichia coli*. *The Journal of Membrane Biology*, 235 (2), 63-72.

Ishii, N., 2013. Two-dimensional crystalline array formation of glucuronide transporter from *Escherichia coli* by the use of polystyrene beads for detergent removal. *The Journal of Membrane Biology*, 246 (3), 199-207.

Loe, D. W., Almquist, K. C., Cole, S. P. C. and Deeley, R. G., 1996. ATP-dependent 17 beta-estradiol 17-(beta-D-glucuronide) transport by multidrug resistance protein (MRP). Inhibition by cholestatic steroids. *The Journal of Biological Chemistry*, 271, 9683-9689.

Loomis, W. F. and Magasanik, B., 1965. Genetic control of catabolite repression of the *lac* operon in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 20 (2), 230-234.

Loomis, W. F. and Magasanik, B., 1967. The catabolite repression gene of the *Lac* operon in *Escherichia coli*. *Journal of Molecular Biology*, 23 (3), 487-494.

Mackenzie, P. I. and Owens, I. S., 1984. Cleavage of nascent UDP glucuronosyltransferase from rat liver by dog pancreatic microsomes. *Biochemical and Biophysical Research Communications*, 122 (3), 1441-1449.

Madej, M. G., 2015. Comparative Sequence-Function Analysis of the Major Facilitator Superfamily: The "Mix-and-Match" Method. *Methods in Enzymology*, 557, 521-549.

Magdalou, J., Fournel-Gigleux, S., Pritchard, M. and Siest, G., 1993. Peroxisome proliferators as inducers and substrates of UDP-glucuronosyltransferases. *Biology of the Cell*, 77, 13-16.

Makman, R. S. and Sutherland, E. W., 1965. Adenosine 3', 5'-Phosphate in *Escherichia coli*. *The Journal of Biological Chemistry*, 240 (3), 1309-1314.

Mandelstam, J., 1960. The intracellular turnover of protein and nucleic acids and its role in biochemical differentiation. *Bacteriology Reviews*, 24 (3), 289-308.

Mandelstam, J., 1961. Induction and Repression of p-Galactosidase in Non-growing *Escherichia coli*. *Journal of Biochemistry,* 79, 489.

Mandelstam, J., 1962. The Repression of Constitutive p-Galactosidase in *Escherichia coli* by Glucose and other Carbon Sources. *Journal of Biochemistry*, 82, 489.

Mandrand-Berthelot, M. A., Condemine, G. and Hugouvieux-Cotte-Pattat, N., 2004. Catabolism of Hexuronides, Hexuronates, Aldonates, and Aldarates. *EcoSal Plus*, 1 (1).

Maniatis, T., Fritsch, E. F. and Sambrook, J., 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory.

Mao, Q. and Unadkat, J. D., 2015. Role of the breast cancer resistance protein (BCRP/ABCG2) in drug transport--an update. *American Association of Pharmaceutical Scientists Journal*, 17 (1), 65-82.

Marbach, A. and Bettenbrock, K., 2012. *lac* operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *Journal of Biotechnology*, 157 (1), 82-88.

Marger, M. D. and Saier, M. H. Jr., 1993. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends in Biochemical Sciences*, 18 (1), 13-20.

Meech, R. and Mackenzie, P. I., 1997. Structure and Function of Uridine Diphosphate Glucuronosyltransferases. *Clinical and Experimental Pharmacology and Physiology*, 24 (12), 907-915.

Miners, J. O. and Mackenzie, P. I., 1991. Drug glucuronidation in humans. *Pharmacology and Therapeutics*, 51 (3), 347-369.

Monod, J., Changeux, J-P. and Jacob, F., 1963. Allosteric proteins and cellular control systems. *Journal of Molecular Biology*, 6 (4), 306-329.

Monod, J., Wyman, J. and Changeux, J-P., 1965. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12, 88-188.

Mullis, K. B., 1990. The Unusual Origins of the Polymerase Chain Reaction. *Scientific American*, 262 (4), 36-43.

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A. and van Sinderen, D., 2013. Bacteriophage Orphan DNA Methyltransferases: Insights from Their Bacterial Origin, Function, and Occurrence. *Applied and Environmental Microbiology*, 79 (24), 7547-7555.

Nakada, D. and Magasanik, B., 1964. The roles of inducer and catabolite repressor in the synthesis of β-galactosidase by *Escherichia coli*. *Journal of Molecular Biology*, 8, 105-127.

National Library of Medicine (US), National Center for Biotechnology Information, 2017. *Nucleotide: Accession No. NC_000913.3, Escherichia coli str. K-12 substr. MG1655, complete genome*. Bethesda. Available from: https://www.ncbi.nlm.nih.gov/nuccore/NC_000913.3?report=graph# [Accessed 4 November 2017].

Nelson, E. and Morioka, T., 1963. Kinetics of the Metabolism of Acetaminophen by Humans. *Journal of Pharmaceutical Sciences*, 52 (9), 864-868.

Nelson, S. O., Wright, J. K. and Postma, P. W., 1983. The mechanism of inducer exclusion. Direct interaction between purified III^{Glc} of the phosphoenolpyruvate:sugar phosphotransferase system and the lactose carrier of *Escherichia coli*. *The EMBO Journal*, 2 (5), 715-720.

Nigam, S. K., Bush, K. T., Martovetsky, G., Ahn, S-Y., Liu, H. C., Richard, E., Bhatnagar, V. and Wu, W., 2015. The Organic Anion Transporter (OAT) Family: A Systems Biology Perspective. *Physiological Reviews*, 95 (1), 83-123.

Nilsen, I. W., Øverbø, K. and Olsen, R. L., 2001. Thermolabile alkaline phosphatase from Northern shrimp (*Pandalus borealis*): protein and cDNA sequence analyses. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry and Molecular Biology*, 129 (4), 853-861.

Novel, G., Didier-Fichet, M. L. and Stoeber, F., 1974. Inducibility of beta-glucuronidase in wild-type and hexuronate-negative mutants of *Escherichia coli* K-12. *Journal of Bacteriology*, 120 (1). 89-95.

Novel, M. and Novel, G., 1976a. Regulation of beta-glucuronidase synthesis in *Escherichia coli* K-12: constitutive mutants specifically derepressed for *uidA* expression. *Journal of Bacteriology*, 127 (1), 406-417.

Novel, M. and Novel, G., 1976b. Regulation of beta-glucuronidase synthesis in *Escherichia coli* K-12: pleiotropic constitutive mutations affecting *uxu* and *uidA* expression. *Journal of Bacteriology*, 127 (1) 418-432.

Novick, R. P., 1987. Plasmid Incompatibility. *Microbiological Reviews*, 51 (4), 381-395.

Olsen, R. L., Øverbø, K. and Myrnes, B., 1991. Alkaline phophatase from the hepatopancreas of shrimp (*Pandalus borealis*): A dimeric enzyme with catalytically active subunits. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*. 99 (4), 755-761.

Pacifici, G. M., Giuliani, L. and Calcaprina, R., 1986. Glucuronidation of 1-Naphthol in Nuclear and Microsomal Fractions of the Human Intestine. *Pharmacology*, 33, 103–109.

Pao, S. S., Paulsen, I. T. and Saier, M. H. Jr., 1998. Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews*, 62 (1), 1-34.

Parquet, M., Pessah, M., Sacquet, E., Salvat, C. and Raizman, A., 1988. Effective glucuronidation of 6a-hydroxylated bile acids by human hepatic and renal microsomes. *European Journal of Biochemistry*, 171, 329-334.

Pastan, I. and Perlman, R., 1970. Cyclic Adenosine Monophosphate in Bacteria. *Science*, 169 (3943), 339-344.

Paulsen, I. T. and Skurray, R. A., 1994. The POT family of transport proteins. *Trends Biochemical Sciences*, 19 (10), 404.

Peng, J., Wen, F., Lister, R. L. and Hodges, T. K., 1995. Inheritance of *gusA* and *neo* genes in transgenic rice. *Plant Molecular Biology*, 27 (1), 91-104.

Peters, W. H., Kock, L., Nagengast, F. M. and Kremers, P. G., 1991. Biotransformation enzymes in human intestine: critical low levels in the colon? *Gut*, 32 (4), 408-412.

Peubez, I., Chaudet, N., Mignon, C., Hild, G., Husson, S., Courtois, V., 2010. Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microbial Cell Factories*, 9 (65).

Poolman, B., Knol, J., Van Der Does, C., Henderson, P. J. F., Liang, W-J., Leblanc, G., Pourcher, T. and Mus–Veteau, I., 1996. Cation and sugar selectivity determinants in a novel family of transport proteins. *Molecular Microbiology*, 19 (5), 911-922.

Popovych, N., Tzeng, S-R., Tonelli, M., Ebright, R. H. and Kalodimos, C. G., 2009. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (17), 6927-6932.

Putaala, H., Sainio, K., Sariola, H. and Tryggvason, K., 2000. Primary structure of mouse and rat nephrin cDNA and structure and expression of the mouse gene. *Journal of the American Society of Nephrology*, 11 (6), 991-1001.

Quistgaard, E. M., Löw, C., Moberg, P., Trésaugues, L. and Nordlund, P., 2013. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nature Structural and Molecular Biology*, 20 (6), 766-768.

Quistgaard, E. M., Löw, C., Guettou, F. and Nordlund, P., 2016. Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews. Molecular and Cell Biology*, 17 (2), 123-132.

Radestock, S. and Forrest, L. R., 2011. The alternating-access mechanism of MFS transporters arises from inverted-topology repeats. *Journal of Molecular Biology*, 407 (5), 698-715.

Radominska-Pandya, A., Czernik, P. J., Little, J. M., Battaglia, E. and Mackenzie, P. I., 1999. Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metabolism Reviews*, 31 (4), 817–899.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. and Medzhitov, R., 2004. Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis. *Cell*, 118 (2), 229-241.

Reddy, V. S., Shlykov, M. A., Castillo, R., Sun, E. I. and Saier, M. H. Jr., 2012. The major facilitator superfamily (MFS) revisited. *The FEBS Journal*, 279 (11), 2022-2035.

Ren, Q., Murphy, S. E., Zheng, Z. and Lazarus, P., 2000. *O*-Glucuronidation of the Lung Carcinogen 4-(methylnitrosamino)-1- (3-Pyridyl)-1-Butanol (nnal) by Human Udp-Glucuronosyltransferases 2b7 and 1a9. *Drug metabolism and Disposition*, 28 (11), 1352-1360.

Roberts, M. S., Magnusson, B. M., Burczynski, F. J. and Weiss, M., 2002. Enterohepatic circulation: physiological, pharmacokinetic and clinical implications. *Clinical Pharmacokinetics*, 41 (10), 751-790.

Roth, M., Obaidat, A. and Hagenbuch, B., 2012. OATPs, OATs and OCTs: the organic anion and cation transporters of the *SLCO* and *SLC22A* gene superfamilies. *The British Journal of Pharmacology*, 165 (5), 1260-1287.

Roux, K. H., 1995. Optimization and Troubleshooting in PCR. *Genome Research*, 4, S185-S194.

Rowland, A., Miners, J. O. and Mackenzie, P. I., 2013. The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. *The International Journal of Biochemistry and Cell Biology*, 45 (6), 1121-1132.

Russell, D. W. and Zinder, N. D., 1987. Hemimethylation prevents DNA replication in *E. coli*. *Cell*, 50 (7), 1071-1079.

Saier, M. H. Jr., Beatty, J. T., Goffeau, A., Harley, K. T., Heijne, W. H., Huang, S. C., Jack, D. L., Jähn, P. S., Lew, K., Liu, J., Pao, S. S., Paulsen, I. T., Tseng, T. T. and Virk, P. S., 1999. The major facilitator superfamily. *Journal of Molecular Microbiology and Biotechnology*, 1 (2), 257-279.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H. A. and Arnhein, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 230 (4732), 471-477.

Schmid, R., 1956. Direct-Reacting Bilirubin, Bilirubin Glucuronide, in Serum, Bile, and Urine. *Science*, 124 (3211), 76-77.

Shi, Y., 2013. Common folds and transport mechanisms of secondary active transporters. *Annual Review of Biophysics*, 42, 51-72.

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

Stappenbeck, T. S., Hooper, L. V. and Gordon, J. I., 2002. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (24), 15451-15455.

Storey, I. D. E. and Dutton,G. J., 1955. Uridine compounds in glucuronic acid metabolism. 2. The isolation and structure of `uridine-diphosphate-glucuronic acid'. *Biochemical Journal*, 59 (2), 279-288.

Strominger, J. L., Kalckar, H. M., Axelrod, J. and Maxwell E. S., 1954. Enzymatic oxidation of uridine diphosphate glucose to uridine diphosphate glucuronic acid. *Journal of the American Chemistry Society*, 76, 6411–6412.

Sun, L., Zeng, X., Yan, C., Sun, X., Gong, X., Rao, Y. and Yan, N., 2012. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, 490 (7420), 361-366.

Sun, H., Ma, Z., Lu, D. and Wu, B., 2015. Regio- and isoform-specific glucuronidation of psoralidin: evaluation of 3-o-glucuronidation as a functional marker for UGT1A9. *Journal of Pharmaceutical Sciences*, 104 (7), 2369-2377.

Takada, H., Hirooka, T., Hiramatsu, Y. and Yamamoto, M., 1982. Effect of beta-glucuronidase inhibitor on azoxymethane-induced colonic carcinogenesis in rats. *Cancer Research*, 42 (1), 331-334.

Titgemeyer, F., Mason, R. E. and Saier, M. H., 1994. Regulation of the Raffinose Permease of *Escherichia coli* by the Glucose-Specific Enzyme IIA of the Phosphoenolpyruvate:Sugar Phosphotransferase System. *Journal of Bacteriology*, 76 (2), 543-546.

Tukey, R. H. and Strassburg, C. P., 2000. Human UDP-Glucuronosyltransferases: Metabolism, Expression, and Disease. *Annual Review of Pharmacology and Toxicology*, 40, 581-616.

Wallace, B. D., Wang, H., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L. A., Mani, S. and Redinbo, M. R., 2010. Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. *Science,* 330 (6005), 831-835.

Wennberg, R. P., Ahlfors, C. E., Bhutani, V. K., Johnson, L. H. and Shapiro, S. M., 2006. Toward understanding kernicterus: a challenge to improve the management of jaundiced newborns. *Pediatrics*, 117 (2). 474-485.

Wilson, K. J., Jefferson, R. A. and Hughes, S. G., 1992. The *Escherichia Coli gus* Operon: Induction and Expression of the *gus* Operon in *E. coli* and the occurrence and use of GUS in other bacteria. In: Gallagher, S .R., ed. *Gus Protocols: Using the GUS gene as a reporter of gene expression.* 1st ed. San Diego: Academic Press, 7-22.

Winkler, H. H., and Wilson, T. H., 1967. Inhibition of β-galactoside transport by substrates of the glucose transport system in *Escherichia coli*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 135 (5), 1030-1051.

Wisedchaisri, G., Park, M. S., Iadanza, M. G., Zheng, H. and Gonen, T., 2014. Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nature Communications*, 5 (4521).

Wu, B., Jiang, W., Yin, T., Gao, S. and Hu, M., 2012. A new strategy to rapidly evaluate kinetics of glucuronide efflux by breast cancer resistance protein (BCRP/ABCG2). *Pharmaceutical Research*, 29 (11), 3199-3208.

Yan, N., 2013. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends in Biochemical Sciences*, 38, 151-159.

Yan, N., 2015. Structural Biology of the Major Facilitator Superfamily Transporters. *Annual Review of Biophysics*, 44, 257-283.

Zhao, P. and Sun, M. X., 2015. The Maternal-to-Zygotic Transition in Higher Plants: Available Approaches, Critical Limitations, and Technical Requirements. *Current Topics in Developmental Biology*, 113, 373-398.

Zheng, Z., Fang, J. L. and Lazarus, P., 2002. Glucuronidation: an important mechanism for detoxification of benzo[a]pyrene metabolites in aerodigestive

tract tissues. *Drug Metabolism and Disposition: The Biological Fate of Chemicals*, 30 (4), 397-403.

Zubay, G., Schwartz, D. and Beckwith, J., 1970. Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proceedings of the National Academy of Sciences of the United States of America*, 66, 104-110.