2014-08-21

The Evaluation of Conserved Amino Acids that Influence Polysubstrate Specificity in the Multidrug Resistance Transporter, EmrE

Saleh, Marwah

UNIVERSITY OF CALGARY

The Evaluation of Conserved Amino Acids that Influence Polysubstrate Specificity in the Multidrug Resistance Transporter, EmrE

by

Marwah Saleh

A THESIS SUBMITTED

TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

AUGUST 2014

Abstract

The study of the molecular mechanism of substrate recognition/binding by multidrug resistance transporters has taken different approaches. I have utilized a PCR-driven site-directed mutagenesis analysis to alter conserved amino acid residues within the *Escherichia coli* small multidrug transporter, EmrE. The EmrE variants generated were further assessed for their resistance ability to 19 structurally different quaternary cationic compounds (QCC) using a high-throughput microtitre plate assay. The underlying hypothesis is that the mutation of conserved amino acids will alter the resistance profile of EmrE to structurally different QCC and identify the specificity of these residues to specific characteristic(s) of QCC. Based on 1,254 resistance profiles, the plasticity of EmrE binding pocket can be explained by the presence of conserved amino acid residues with different substrate preferences. The significance of studying EmrE is that it provides an excellent model for understanding the polysubstrate specificity on a molecular basis relatable to other major transporters.

Acknowledgment

I would like to show my gratitude to the following:

My supervisor, Dr. Raymond J. Turner

My committee members: Dr. Elke Lohmeier-Vogel and Dr. Douglas Storey

All the Turner lab members with special thanks to Dr. Denice Bay for her continuous guidance throughout the process

Finally to my family and loyal friends

“Doubt is a pain too lonely to know that faith is his twin brother” Khalil Gibran

If it weren’t for my mother (**Sundus**), father (**Abdul-Salam**), and my two sisters (**Ayat & Tabark**) who had faith in me and supported me every step of the way, it would have taken longer to successfully complete this work.

I’d like to offer my deepest gratitude to **Dr. Raymond Turner** and **Dr. Denice Bay**, who have guided me every time there was a moment of doubt.

Finally, for those moments of pain and joy, I have to thank my encouraging and caring **friends**!

Thank you,

Table of Content

List of Figures

Chapter One: Introduction

Chapter Two: General Methods

Chapter Three: Altering Key Amino Acid Residues within EmrE

Chapter Five: Advanced Data Analysis

Chapter Six: Discussion

List of Tables

Chapter One: Introduction

Chapter Two: General Methods

Chapter Three: Altering Key Amino Acid Residues within EmrE

Chapter Four: Resistance Profiles

Chapter Five: Advanced Data Analysis

Chapter 6: Discussion

Abbreviations List

ABC- ATP-binding cassette

ACR- acriflavine

ANOVA- Analysis of variance

BZ- benzalkonium

CB & CC - Cetylpyridinium (bromide & chloride)

CET- Cetrimide

Cryo-EM- Cryo-electron microscopy

CT.C or Ct.CL- Cetalkonium (chloride)

CV- Crystal violet

DC- Dequalinium

DCCD- dicyclohexylcarbodiimide

DMSO- Dimethylsulfoxide

EB or Et.Br- Ethidium (bromide)

EmrE- *E.coli*-Ethidium multidrug resistance transporter

EPR- Electron spin resonance

Hclust: Hierarchical agglomerative clustering

HE- Hexamethylenetetramine

LB- Luria broth/ LB agar plates

MATE- Multidrug and toxic compound extrusion

MC- Myristalkonium

MDR- Multidrug resistance/resistant

MFS- Major facilitator superfamily

MIC- Minimum inhibitory concentration

MTP- Methyltriphenylphosphonium

MV- Methyl viologen

NEM- N-ethylmalemide

NMR- Nuclear magnetic resonance

OD= Optical density

Omp- Outer membrane protein

PCR- Polymerase chain reaction

PISEMA- Polarization inversion spin exchange at magic angle

PMF- Proton motive force

PRO- Proflavine

PSMR- Paired SMR proteins

PY- Pyronin

QCC- Quaternary cationic compounds

RH- Rhodamine

RND- Resistance nodulation cell division

SCAM- Scanning cysteine accessibility method

Sd- Mean synonymous nucleotide substitution

SDM- Site directed mutagenesis

SMP- Small multidrug pumps

SMR- Small multidrug resistance

STAC- Stearyltrimethylammonium

SUG- Suppressor of *groEL* mutation proteins

TMS- Transmembrane segment(s)

TPA- Tetraphenylarsonium

TPP- Tetraphenylphosphonium

Chapter 1: Introduction

1.1 Multidrug Resistance

In recent years, there has been an apparent elevation in the production of antiseptics, disinfectants and antibiotics (Levy, 2002; Mellon et al., 2001; Barza and Gorbach, 2002). This is mainly because of the mounting concerns over the potential for microbial contamination and infection risks in food, cosmetics and other general consumer products. Therefore, the phenomena of the indiscriminate use of these antimicrobial agents in hospitals, other health care settings and by the general public resulted in the selection of pathogenic bacteria that are resistant to several antimicrobial agents (Levy, 2002). These species are identified as multidrug resistant (MDR) bacteria. Therefore, the question becomes what are the mechanisms that are giving these bacteria the power to resist such toxic compounds?

There are several resistance mechanisms utilized by bacteria in order to extrude antimicrobial agents, which are reviewed by Nikaido (2009). First resistance mechanism involves the mutation of the proteins that are targeted by the drugs making it less susceptible to them. For example, erythromycin is known to inhibit protein synthesis by its effect on ribosome's function (Welsblum, 1995). Bacteria carrying the *erm* gene, which codes for N-methyltransferase, results in the methylation of the adenine at position 2058 of the 50S rRNA (Welsblum, 1995). This modification of the rRNA inhibits the interaction of macrolides (erythromycin and others) with the ribosome, thereby conferring resistance to these agents. The second mechanism involves an enzymatic inactivation of the antimicrobial agent. For example, the β-lactams (penicillin, cephalosporin, and carbapenem

such as imipenem) are inactivated by enzymatic hydrolysis by β-lactamases, which usually takes place in the periplasm (Bush et al., 1995). The third mechanism involves the acquisition of genes from foreign sources to allow bacteria to become resistant, because they code for protein variants that are tolerant. For example, *Streptococcus pneumoniae* acquires penicillin resistant proteins from other organisms (Spratt, 1994). This is mainly because of *S. pneumoniae's* ability for natural transformation and import foreign DNA. The fourth mechanism involves non-specific exclusion of antimicrobial agents, which is a mechanism utilized by bacteria with porin mutations. The mutations within the coding sequences of porins have been reported to reduce the diffusion uptake rates of β-lactams (Achouak et al., 2001). The fifth and last mechanism involves the bacteria that possess efflux pumps with the ability to transport multiple antimicrobial agents, also known as multidrug resistance (MDR) pumps.

MDR transporters are membrane translocases with the ability to extrude a variety of structurally unrelated compounds from the cell (Lewis, 1994). Since the early 1990's, there has been an immense interest in studying bacterial multidrug resistance transporters since they defy the general notion of enzymes substrate-specificity. The subsequent sections of this chapter will shine a light on the different MDR families found in bacteria leading to the multidrug resistance protein, EmrE, that is the focus of this study.

1.2 Multidrug Resistance Efflux Pumps

Microorganisms have developed various mechanisms to resist the toxic effects of antimicrobial agents. Some of these mechanisms like reduced membrane permeability are not sufficient in preventing the toxicity of these agents once they are inside the cell.

Therefore, active efflux of these compounds is essential to keep the cellular concentration of the toxin low enough for the cell to maintain biochemical and physiological function. There are multidrug transporters that can extrude a wide variety of structurally different agents, called multidrug resistance (MDR) transporters (Lewis 1994; Nikaido, 1996; Paulsen et al., 1996B; Van-Veen et al., 1998). MDR transporters are divided into two major groups based on bioenergetics and structural criteria. Secondary multidrug transporters utilize the electrochemical proton gradient to drive the extrusion of substrates and the ATP-binding cassette (ABC) group that uses the free energy of ATP hydrolysis to efflux the compounds out.

The ABC transporters constitute one of the largest superfamilies of proteins (Dassa et al., 1999). They are composed of four domains where two are highly hydrophobic, which consist of 6 transmembrane α-helices, and two hydrophilic nucleotide-binding domains (Hyde et al., 1990). In total there are 80 ABC transporters in *Escherichia coli* and 48 in humans (Saier, 2000). The ABC efflux transporters in Gram-negative organisms are composed of the inner membrane ABC transporter and an outer membrane channel (i.e. TolC channel), where both are tightly coupled and an absolute requirement for substrate extrusion (Figure 1.1). The homologues of Gram-negative ABC transporters in humans are the P-glycoproteins.

The secondary multidrug transporters are divided into four distinct families based on their size and their primary/secondary structures: the major facilitator superfamily (MFS) (Marger and Saier, 1993), the resistance nodulation cell division (RND) family (Saier et al., 1994), the metabolite and toxic compound extrusion (MATE) family (Brown et al., 1999), and the small multidrug resistance (SMR) family (Paulsen et al., 1996C).

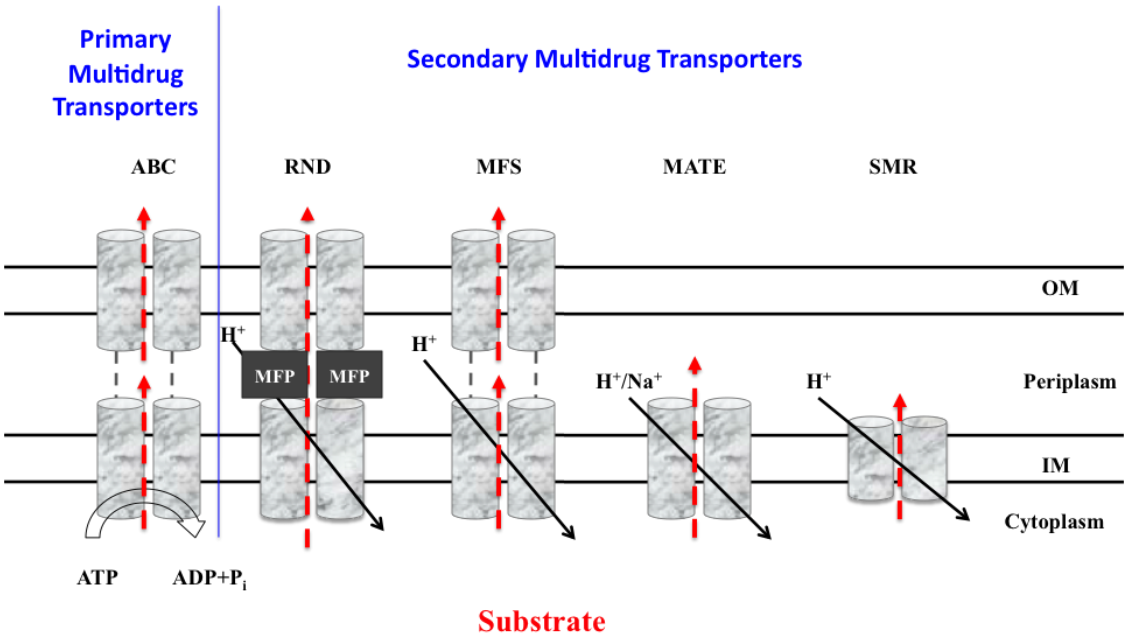


Figure 1. 1: The multidrug resistance transporter families of Gram-negative bacteria.

OM=outer membrane and IM=inner membrane. MDR are divided into two major groups based on bioenergetics and structural criteria. First group is the ATP-binding cassette (ABC) group that uses the free energy of ATP hydrolysis to efflux the compounds out and the second group is the secondary multidrug transporters utilize the electrochemical proton gradient to drive the extrusion of substrates. The secondary MDR are divided into four distinct families based on their size and their primary/secondary structures: the major facilitator superfamily (MFS), the resistance nodulation cell division (RND) family, the multidrug and toxic compound extrusion (MATE) family, and the small multidrug resistance (SMR) family.

The MFS family is the largest group of secondary active transporters (Saier et al., 1998). Some examples of its members include Bmr and Blt of *B. subtilis*, NorA and QacA of *S. aureus*, and MdfA of *E. coli*. Usually in Gram-negative bacteria, due to the inner and outer membrane, the MFS efflux system is composed of three components (Figure 1.1) with inner membrane MFS component, periplasmic protein, and an outer membrane channel (i.e. EmrAB-TolC). This family of transporters is known to be involved in symport or antiport activities of several substrates, such as sugars, phosphate esters, oligosaccharides, and antibiotics (Marger and Saier, 1993). For example, MdfA in *E. coli* confers resistance to neutral and positively charged substrates. Also, the substitution of the Glu26 in that protein leads to reduced resistance to positively charged drugs indicating the importance of the charged residue embedded in the transmembrane segment (Edgar and Bibi, 1999). Another important pathogenic bacterium that expresses a multidrug resistance pump of the MFS family is *Vibrio cholerae*. VceAB pump in *V. cholerae* is found to be involved in conferring resistance to toxic agents/antibiotics, such as deoxycholate, nalidixic acid and chloramphenicol (Colmer et al., 1998).

The RND transporters family is a proton/substrate antiport system that is particularly predominant in Gram-negative bacteria. Examples of RND transporters include the AcrB and MexB, which form a tripartite complex (Figure 1.1) with an inner membrane component, a membrane fusion protein (MFP) that is a periplasmic adaptor (AcrA and MexA), and finally an outer membrane channel (TolC and OprM) (Nikaido, 2009; Seeger et al., 2006; Murakami, 2008). This family of transporters consists of 12 transmembrane segments (TMS). The AcrA/B-TolC system in *E. coli* is known to confer resistance to basic dyes, detergents and antibiotics (Nakamura, 1968).

The MATE family of transporters are involved in a wide variety of biological functions. In bacteria, MATE transporters function as proton/Na+ antiport and confer resistance to xenobiotics and multiple cationic toxic agents (Becker et al., 2009; Tsuda et al., 2009). However, MATE substrate-resistance profiles are narrower than the RND family (Li and Nikaido, 2009).

The SMR family contains more than 250 annotated members and is divided into three subclasses: the small multidrug pumps (SMP), the suppressor of *groEL* mutation proteins (SUG), and paired SMR proteins (PSMR) (Bay et al., 2008). SMR proteins are encoded on the chromosomes, plasmids or associated with integrons (Bay and Turner, 2009). As the smallest efflux transporter known, SMR typically have 105-125 amino acid residues (Putman et al., 2000). These proteins tend to function as oligomer complexes, in part considered because of their small sizes (Bay et al., 2008). Also, SMR proteins are known to confer resistance to variety of lipophilic cations and clinically relevant antibacterials (i.e. aminoglycosides) (Bay et al., 2008; Li et al., 2003). Although SMR members consist of an inner membrane component, which exports substrate to the periplasmic space (Figure 1.1), these proteins can give rise to significant resistance. It is hypothesized that substrates can be taken up by the outer membrane channel of other pumps (i.e. TolC of AcrA/B-TolC pump) (Schuldiner, 2009). Further details on SMR family of transporters will follow in the subsequent sections of this chapter mainly with the focus on the most studied member of the family, EmrE.

1.3 Small Multidrug Resistance Proteins

The small multidrug resistance (SMR) transporters have been found on a variety of plasmids and transposable elements (Bay et al., 2008; Bay & Turner, 2009). They are also known to confer resistance to wide range of lipophilic cationic drugs (Paulsen et al., 1993; Grinius et al., 1992) and antibiotics, such as β-lactams (Sidhu et al., 2001), cephalosporins (Doi et al., 2002), dihydrofolate inhibitors (Burnside et al., 1996), and aminoglycosides (Li et al., 2003). Also, it has been suggested that antibiotic and SMR resistance genes tend to be tightly linked (Sidhu et al., 2001), which makes the spread of SMR homologues a critical clinical and industrial issue. Therefore, the study of the transport mechanisms of the proteins related to the SMR family becomes essential.

SMR transporters appear to be an ideal system to study the steps of the transport and drug efflux due to their small size (~12kDa) that range from 100-140 amino acids in length (Paulsen et al., 1996C). Due to their small size, SMR protein multimerization is suggested to be a requirement for substrate transport (Yerushalmi et al., 1996; Masaoka et al., 2000; Zhang et al., 2007; Elbaz et al., 2004). Also, SMR family consists of over 250 annotated proteins, with 52% of completely sequenced bacteria having SMR homologues and 31% of Archaea have SMR homologues (Bay et al., 2008).

The substrate efflux of the SMR proteins is energy dependent and driven by the proton motive force (PMF). Also, particular members of the SMR family (i.e. Sau-Smr & Eco-EmrE) are known to be resistant to antibiotics, such as erythromycin and tetracycline (Heir et al., 1999; Fuentes et al., 2005). The earliest characterized member of the SMR family is QacC, which is a staphylococcal multidrug efflux protein that is known to confer

resistance to a variety of lipophilic cationic compounds (Lyon and Skurray, 1987; Sasatsu et al., 1989; Grinius et al., 1992).

Early amino acid sequence alignment of 7 members of the SMR family revealed conserved motifs with signature sequences that distinguish the SMR family from the other MDR transporters (Grinius et al., 1992; Bairoch, 1992; Paulsen et al., 1993). Thus, these signature sequences are used to identify new members of the family as they become sequenced.

The SMR family is divided into 3 subclasses, which include the SMP, SUG, and the PSMR (Bay et al., 2008). The small multidrug pumps (SMP) are grouped according to their functional and structural similarity (Chung and Saier, 2001). They are also found in Gram-negative (i.e. EmrE from *E. coli*), Gram-positive bacteria (Smr from *S. aureus*), and Archaea (Hsmr from *H. salinarum*) to confer multidrug resistance to quaternary cationic compounds (QCC) (Saier, 2000). The second subclass is referred to as suppressor of *groEL* mutation protein (SUG). The GroEL chaperone is part of the chaperone complex of GroEL/GroES, which helps in protein folding in bacteria (Radford, 2006). The SUG transporters also confer resistance to QCC. However, members of the SMP subclass have unique resistance profiles to a wider range of compounds (reviewed by Bay et al., 2008). The paired SMR (PSMR) subclass requires 2 copies of each SMR homologue in order to be functional in conferring drug resistance, hence the name 'paired' (Chung and Saier, 2001). Some examples of PSMR include YdgF-YdgE and EbrA-EbrB.

The difference in the functionality between SUG and SMP may reside in the different conserved residues found in the TMS of the protein. In TMS4, SUG have more

positively charged residues than SMP proteins. Also, SUG and SMP differ in the presence or absence of aromatic residues and Ala or Gly at positions 14-15 in TMS1 (Bay et al., 2008). Moreover, a highly conserved Glu in loop #1 is found in SMP while a His is found in SUG (Paulsen et al., 1996C). In summary, there are unique conserved motifs that are specific to each subclass of the SMR family, which may explain the differences in functionality.

The transport mechanism of SMR proteins remains to be a compelling question with different possible answers. However, one of the proposed drug transport mechanisms of SMR by Paulsen et al. (1996C) suggested that the substrate interacts initially with the protonated acidic residue (i.e. Glu) in TMS1 to allow the exchange between the substrate and H^+. Then, the transporter undergoes a conformational change that allows for the translocation of the substrate, which is mediated by competitive binding of the proton(s) to replace the substrate. Since then, most mechanisms proposed are based on some aspect of this theme. The reports of the proton stoichiometry of H^+ to SMR vary from 1:1 (Rotem and Schuldiner, 2004), 1:2 (Tate et al., 2003), and 2:3 (Soskine et al., 2004). Finally, the complex binds the protons releasing the substrate to the periplasmic space and returning to its native conformation (Figure 1.2). More specific aspects of proposed transport mechanisms have been mainly explored with the protein EmrE of the SMR group, which will be described in the following section.

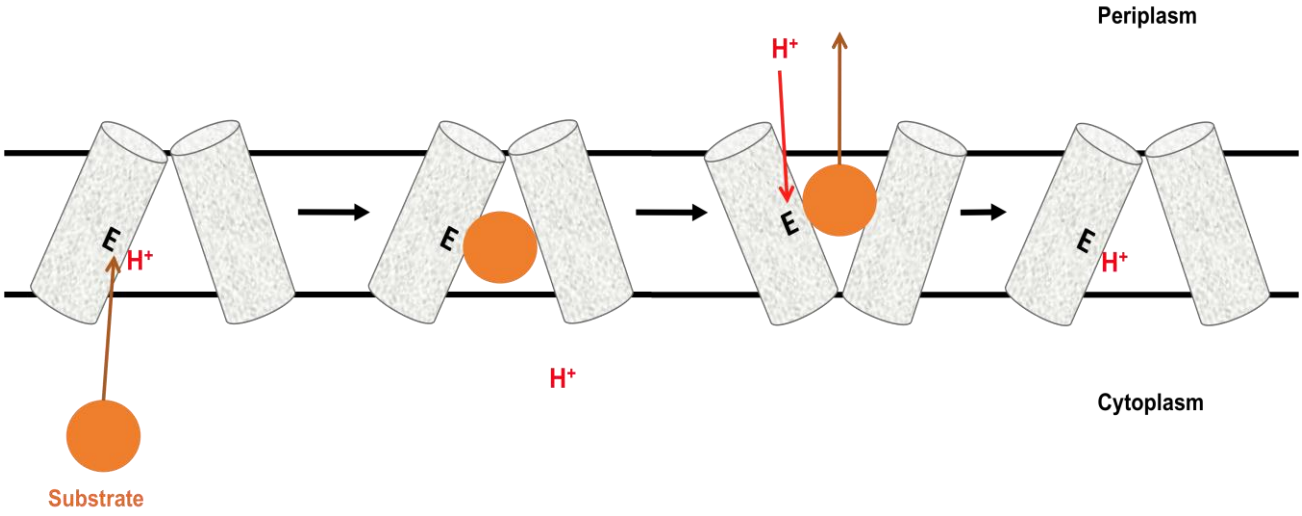


Figure 1. 2: The general transport mechanism proposed for small multidrug resistance (SMR) proteins.

The oligomerization of the SMR protein has been suggested to be a requirement for the drug resistance function (Yerushalmi et al., 1996; Masaoka et al., 2000; Winstone et al., 2005; Zhang et al., 2007; Elbaz et al., 2004), hence the protein above is cartooned as a dimer. The above stages and sequence of substrate translocation were originally suggested by Paulsen et al., 1996C and details explained in Section 1.3. Overall, the mechanism follows the alternating access model. This model goes through steps were the binding site has alternating access to the inside and outside of the cell.

1.4 *E. coli* - Ethidium Multidrug Resistance (EmrE) Transporter

Ethidium multidrug transporter, EmrE, is an *E. coli* member of the SMR family of transporters. EmrE has been identified to confer resistance to a variety of QCC, particularly ethidium (Purewal, 1991) and methyl viologen (Morimyo et al., 1992). The predicted amino acid sequence of the protein suggests that it is highly hydrophobic (12 kDa) protein with four transmembrane segments (TMS) and only one charged residue (Glu) that is embedded in TMS1 (Yerushalmi et al., 1995) (Figure 1.3). The Glu14 residue is conserved in more than 100 homologous proteins and involved in the coupling mechanism of proton/substrate exchange.

EmrE provides an excellent experimental paradigm to study the multidrug resistance mechanism because of several advantages. First, its small size offers an excellent model for studying membrane protein folding and ligand binding as there are less amino acids to explore by mutagenesis studies that focuses on unique replacements of specific residues. Thus, potentially making it easier for identifying the residues that are possibly involved in the substrate-binding pocket. Second, EmrE is soluble in organic solvent, which makes purification procedures quick and efficient without the need of an affinity epitope tag. This allows for the study of the secondary structure of the protein using X-ray structural analysis (Chen et al., 2007) and nuclear magnetic resonance (NMR) (Schwaiger et al., 1998). Moreover, the solubilization and the purification of the protein do not abolish the transport activity of EmrE since reconstitution displays transport activity (Schuldiner et al., 2001) and ligand binding (Sikora and Turner, 2005). Therefore, EmrE multidrug resistance activity can be usefully explored using *in vivo* and *in vitro* assays.

In the subsequent sections of this chapter, the transport mechanisms proposed for EmrE, the dynamic structure and oligomerization, and the plasticity of the protein will be described as illustrated in the previous literature. Moreover, the studies carried out to explore the amino acids involved in EmrE binding-site will be outlined. By the end of this chapter, the reader will hopefully develop an understanding of what has been explored in the multidrug resistance functionality of EmrE and what remains to be studied.

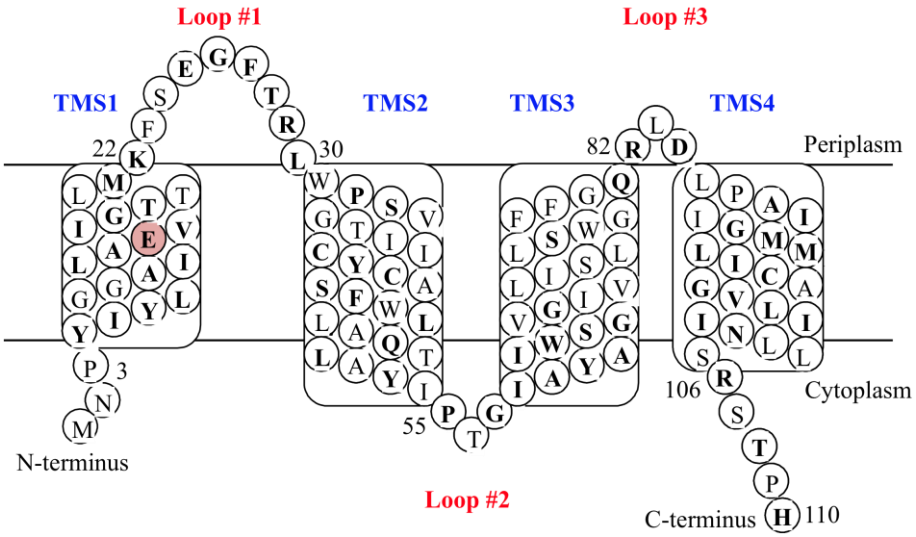


Figure 1. 3: EmrE secondary structure.

Loops and transmembrane segments (TMS) are organized according to residue accessibility experiments preformed by Mordoch et al., 1999. The bold highlighted residue is the Glu14 that is involved in substrate/proton coupling process of EmrE transport mechanism (Bay et al., 2008- obtained with permission).

1.4.1 Biochemical and Structural Studies of EmrE

The experiments that confirmed the presence of highly hydrophobic TMS involved the use of scanning cysteine accessibility method (SCAM) (Mordoch et al., 1999). SCAM involves the generation of mutants with unique Cys residues that are positioned at desired

locations in the protein. Then, sulfhydryl reagents (i.e. N-ethylmalemide), which react covalently with thiol groups, were added to assess the accessibility of certain residues at various positions in the protein. The results indicated that none of the residues speculated to be in the TMS were accessible to N-ethylmalemide. Therefore, this confirmed the presence of 4 tightly packed helices that are embedded in the membrane (Figure 1.3).

The significance of Glu14 residue in the functionality of EmrE was examined using chemical modifications with carbodiimides (Yerushalmi and Schuldiner, 2000). A carbodiimide (i.e. dicyclohexylcarbodiimide (DCCD)), would react with the carboxyl group on residues like Asp and Glu resulting in the inhibition of substrate uptake by wild-type EmrE. However, when Glu25 and Asp84 (the other charged residues found in EmrE loops) were substituted with Cys, the inhibition of EmrE uptake activity was observed indicating that Glu14 was the only carboxylic acid residue where DCCD mechanism of action took place. Moreover, substrate-binding activity was inhibited for wild-type EmrE by DCCD in a dose-dependent manner (Yerushalmi et al., 2001). Also, the inactivation of substrate binding was inhibited upon the addition of substrates, such as ethidium, acriflavine, and benzalkonium, during the incubation with DCCD occured in a dose-dependent manner. Therefore, the binding of these substrates to Glu14 prevented the action of DCCD, which further confirmed the presence of Glu14 in the binding-site.

The oligomerization of EmrE was first demonstrated using the mixing approach between wild-type and mutant-EmrE monomers (Yerushalmi et al., 1996). EmrE variants, which replaced Glu14, Tyr60, and Trp63 with Cys or Phe, were co-expressed with a wild-type protein and showed a significant reduction in resistance to methyl viologen, ethidium, and acriflavine. There was an inhibition of methyl viologen-uptake activity when EmrE

variants were purified and co-reconstituted in proteoliposomes with the wild-type protein in a dose dependent manner. The inhibition in the uptake activity was attributed to the formation of mixed oligomers. Thus, the transport mechanism of EmrE may require the formation of homodimers and the participation of Glu14, Tyr60 and Trp63 from each protomer in the transport activity.

In previous research, EmrE has been found to exist in multiple functional forms depending on the experimental conditions. The different forms vary from monomers (Winston et al., 2005; Klammt et al., 2004; Soskine et al., 2002), dimers (Tate et al., 2001 & 2003; Ubarretxena-Belandia et al., 2003 & 2004; Rotem et al., 2001; Elbaz et al., 2004; Ilag et al., 2004; Dutta et al., 2014), to trimers (Yerushalmi et al., 1996; Torres and Arkin, 2000; Butler et al., 2004), and tetramers (or dimers of dimers) (Tate et al., 2003; Pornillos et al., 2005; Elbaz et al., 2004). This led to varying opinions on the transport mechanism of EmrE depending on the structural model that is proposed for the protein.

The transport mechanism of EmrE is similar to the SMR generic mechanism in that it is also driven by a proton electrochemical gradient (Yerushalmi et al., 1995). The stoichiometry of proton to substrate is likely to be $2H^+$ to every 1 substrate (Grinius and Goldberg, 1994; Paulsen et al., 1996C; Yerushalmi et al., 1995) yet this is still debated.

The transport mechanism suggested by Yerushalmi and Schuldiner (2000) and Muth and Schuldiner (2000) proposed an EmrE trimer model. This model suggested that the substrate (TPP+) interacts with the trimeric cluster of the Glu14 residues, where two charges are neutralized by protons and one permanent negative charge stabilizes the interaction with the cationic substrate. The electrostatic interaction of the substrate with the

Glu14 cluster leads to the deprotonation of the two neutralized residues. Subsequent conformational changes to the protein-substrate complex opens the pocket to face the periplasmic side of the membrane. The release of the substrate depends on the binding of 2 protons from the periplasmic space to the two Glu14 of the cluster. This is supported by the experiment to which decreasing the pH helps in accelerating substrate release. Following the protonation and substrate release, the binding site relaxes back to initial conformation (open towards the cytoplasm).

On the other hand, using proton-substrate binding assay of detergent solubilized EmrE proposed a transport mechanism for a monomeric EmrE (Soskine et al., 2004). First, the deprotonation of Glu14 leads to the binding of the substrate. Then, the substrate is translocated to the periplasm, where a new proton binds Glu14 displacing the substrate. Therefore, substrate transport is based on the competition between proton and substrate for Glu14 binding, which means that protonation/deprotonation can happen on either side of EmrE but protons are more likely to be released to the cytoplasm.

Other researchers have suggested that EmrE functions as a dimer using Cryo-electron microscopy (Tate et al., 2001 & 2003; Ubarretxena-Belandia et al., 2003). Based on this model, the substrate binds upon the loss of 2 protons from the Glu14 of each monomer. Subsequent conformational changes that result in tilting TMS1-3 of each monomer for 20° exposes the substrate to the periplasm. Then, the protonation of the 2 Glu14 displaces the substrate and causes a conformational change that puts the dimer in its original state. This suggested mechanism has also been confirmed by Koteiche et al. (2003), which demonstrated a V-shaped chamber of EmrE near the N-terminus with Glu14 residing within the hydrophobic pocket.

The last model suggested by Winston et al. (2005) tried to accommodate both biochemical and structural differences of the models proposed above. This model suggested the flexibility of EmrE with respect to multimerization in that oligomerization process may be dependent on the nature of the substrate. First, the substrate binds monomeric EmrE and is released upon Glu14 protonation as suggested earlier by Soskine et al. (2004). However, if the substrate requires more EmrE protomers to accommodate for substrate's structural differences (i.e. extra charges), then oligomerization of EmrE becomes necessary. This speculation was further confirmed by Bay et al. (2010) and Bay and Turner (2012), where EmrE multimerization occurred in a protein concentration-dependent manner and was enhanced by chemically diverse QCC. Therefore, the transport ability of the EmrE monomer and the demonstration of EmrE dimer/trimer structures (Tate et al., 2001; Ubarretxena-Belandia et al., 2003; Yerushalmi and Schuldiner, 2000) further strengthens the Winston et al. (2005) model and suggests a functional EmrE transport model for all forms of the protein since multimerization is substrate-dependent (Bay and Turner 2012).

Additional structural models of EmrE utilized X-ray crystallography to analyze the oligomerization of EmrE. Ma and Chang (2004) put together a tetrameric model for an unbound EmrE that was retracted later on due to incorrect structure and topology analysis. That X-ray structure had conflicting packing angles between each pair of the α-helices with the Cryo-EM structure (Fleishman et al., 2006). Moreover, the other retracted dimer X-ray structure of EmrE that was bound to TPP+ also had major differences from the Cryo-EM structure, in that there was difference in the tilt angles and the positions of α-helices (Pornillos et al., 2005). Upon reanalyzing their data, Chen et al (2007) were able to produce an anti-parallel dimer model of EmrE. The TPP+ bound X-ray structure of EmrE illustrated

two monomers with the most conserved helices (TMS1and TMS3) being in the centre and flanked by less conserved helices (TMS2 and TMS4). This structure also agreed with the models produced from the Cryo-EM analysis (Fleishman et al., 2006), where the substrate-binding chamber is composed of six helices (TMS1, 2 and 3 from each monomer) and the two TMS4 helices participate in the dimerization interaction. Furthermore, the orientations of the TMS are similar in both Cryo-EM and X-ray structures, where Glu14 from each monomer point towards the binding chamber and in contact with TPP+, confirming its involvement in substrate binding (Muth and Schuldiner, 2000). Also, residues in TMS1, 2, and 3 that have been identified to be important for TPP+ binding in EmrE (Muth and Schuldiner, 2000; Elbaz et al., 2005; Gutman et al., 2003; Mordoch et al., 1999) were mapped on the walls of the substrate-binding chamber as identified by X-ray structure (Chen et al., 2007).

Tryptophan scanning analysis further supported the X-ray model of EmrE anti-parallel dimer (Lloris-Garcera et al., 2013). The tryptophan scanning studies involved the replacement of residues at random locations in the protein with Trp. This would disrupt the packing of EmrE and abolishes its function since Trp has a bulky side chain. However, the lipid-exposed residues tend to be impervious to Trp replacements producing functional EmrE proteins. Therefore, Trp scanning can produce structural models of EmrE α-helices based on lipid-exposure data obtained from coarse-grained molecular dynamics simulations. From the 60 Trp mutants tested (Lloris-Garcera et al., 2013), all lipid-exposures agreed with the previous anti-parallel homodimer model of EmrE obtained by X-ray crystallography (Chen et al., 2007). Recent research with EmrE also confirmed the stability of the dimer model of EmrE and that EmrE dimers do not dissociate once they are

formed (Dutta et al., 2014). Therefore, recent emerging data are promoting the anti-parallel dimer model of EmrE.

Considering that EmrE substrate-transport mechanism only moves the substrate to the periplasmic space yet the expression of the protein leads to significant multidrug resistance, there must be an aiding outer membrane protein that finishes the job. A recent study by Beketskaia et al. (2014) suggested that an outer membrane protein, OmpW, participates in the efflux of EmrE substrates across the outer membrane. This study involved assessing the growth of *E. coli* transformed with plasmid carrying *emrE* under alkaline conditions. It is believed that under alkaline conditions, EmrE transports osmoprotectant molecules (i.e. betaine) resulting in loss-of-growth phenotype (Bay and Turner, 2012). Therefore, several *E. coli* strains (expressing EmrE) with one outer membrane protein deletion were used to assess the rescue-of-growth phenotype. The deletion of *ompW* displayed loss-of-growth phenotype; hence OmpW is the outer membrane protein that completes EmrE substrate-extrusion function.

In summary, the transport mechanism of EmrE possibly involves the oligomerization of the protein to accommodate for certain substrate molecular structures. Furthermore, the involvement of OmpW in leading the toxic substrates across the outer membrane space suggests more proteins are involved in the complete transport pathway out of the cell for the substrates.

1.4.2 The Plasticity of EmrE Binding-Pocket

The transport mechanism of EmrE has been explored by several groups. However, the structural basis and the conformational changes mediating the transport have yet to be

elucidated. Also, the plasticity of the binding pocket that can accommodate such a diverse range of substrates and whether the conformational changes are influenced by substrate identity have been recently investigated. Korkhov and Tate (2008) have suggested that the remodeling of the regions surrounding the substrate-binding site as a result of the binding of planar substrates (ethidium, dequalinium, and propidium) is different from the changes observed with Cryo-EM when the more spherical substrate, TPP+, is bound. Therefore, they concluded that EmrE remodeling is dependent on the geometry of the substrate bound and the remodeling process may involve subtle changes to the transmembrane domains.

Furthermore, Amadi et al. (2010) identified those subtle changes to the transmembrane segments that are induced as a result of substrate binding. The conformational changes of EmrE described by the transition from inward-open to out-ward open (Figure 1.4) involves repacking of TMS1, tilting of TMS2, and changes in the backbone configuration of TMS3 and the adjacent loop connecting to TMS4. Electron spin resonance (EPR) spectroscopy was used to identify the changes/movements of TMS upon the binding of TPP+. The EPR data suggested that the changes of the TMS involved a slight kink in TMS1 (near Glu14 residue), which leads to tighter packing of the N-terminal to the Glu14 increasing the steric restrictions at residues 8 and 10. Hence, TMS1 has limited helix rotation. Moreover, TMS2 is tilted and TMS3 backbone is rearranged as a result of residue accessibility changes when TPP+ is bound. Finally, loop #3 (Figure 1.3) becomes more exposed to the lipid side of the membrane, which suggests a rearrangement of TMS4 to accommodate for that. The ordering of TMS3 is further confirmed by Gayen et al. (2013). The PISEMA (polarization inversion spin exchange at magic angle) spectroscopy data indicated structural rearrangements of TMS3 involving helix bending

around Gly67 (Gly65-Val66-Gly67-Ile68-Va69 region), which agrees with the distinguishable kink (See Figure 1.5) observed in the Cryo-EM images generated by Fleishman et al. (2006).

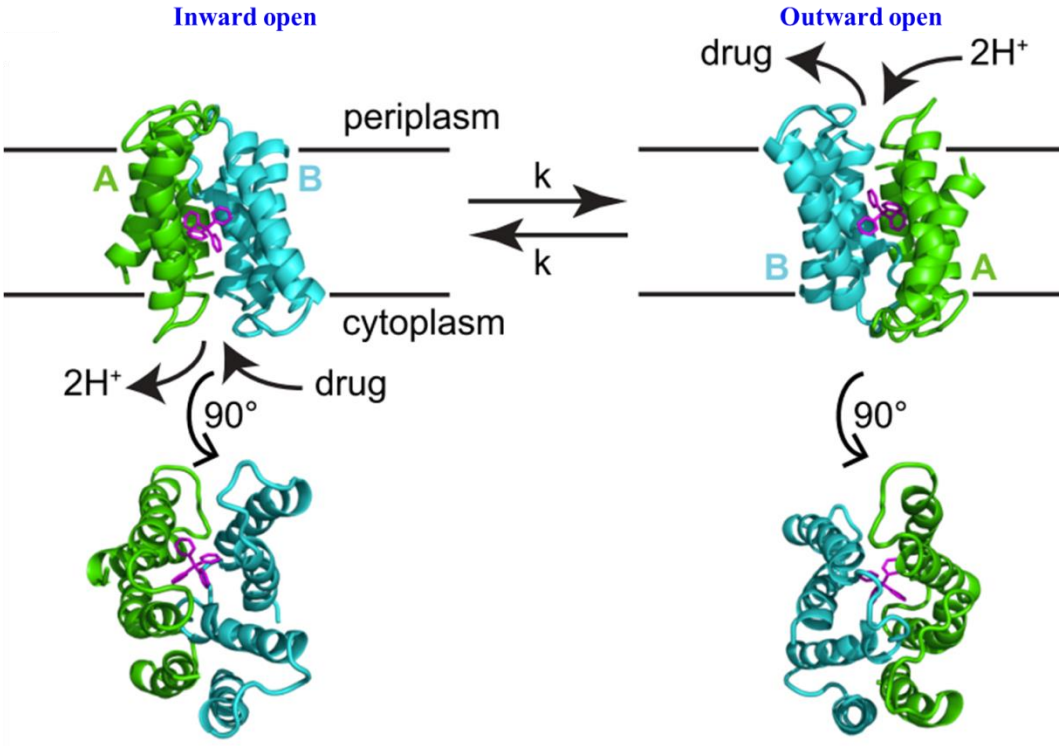


Figure 1. 4: EmrE conformational changes upon substrate binding.

The EmrE dimer exchanges between an inward-open and outward open confirmations as a result of substrate (TPP+ in purple) binding/release and proton release/binding. A 90° rotation permits a view of the open and closed faces of EmrE (PDB 3B5D, PyMOL). The figure was obtained from Morrison and Henzler-Wildman (2014) with permission.

In addition, recent research by Morrison and Henzler-Wildman (2014) identified another aspect of EmrE conformational changes that are influenced by substrate identity. Using a variety of NMR techniques along with binding/efflux assays, they identified that the rate of conversion between inward and outward-open state is dependent on substrate characteristics (particularly substrate hydrophobicity that influenced substrate binding affinities). Therefore, taking all these experiments into consideration, the plasticity of

EmrE binding region is influenced by the identity of the substrate and the conditions surrounding the experiment, which leads to particular rearrangements of the transmembrane domains to accommodate for particular substrates.

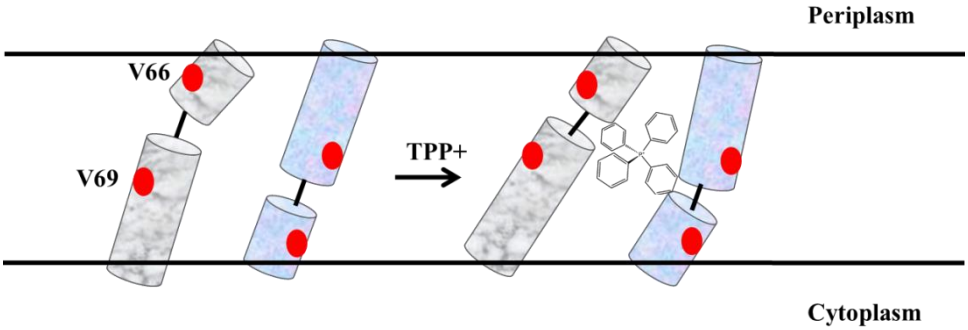


Figure 1. 5: The bending motion of EmrE transmembrane segment three.

The kink of TMS3 around residues Val66-Val69 changes between the two monomers upon the binding of the substrate (TPP+). Adapted from Fleishman et al. (2006) and Gayen et al. (2013). The red dots indicate the residues Val66 and Val69 that span the Val66-Gly67-Ile68-Va69 region, which is responsible for the distinguishable kink in TMS3.

Furthermore, the plasticity of EmrE binding-pocket may be explained by the idea of a large cavity of the protein with discrete niches for the different substrates. These discrete niches are the different amino acid residues with preferences in interacting with different substrates. Earlier studies by the Schuldiner lab explored the role of several residues involved in giving resistance to a finite group of substrates. The following section will outline those residues and their proposed roles and the following chapters will outline my contributions towards understanding the poly-substrate specificity of EmrE.

1.4.3 Resistance profiles of mutated amino acid residues in previous literature work

The group led by Simon Schuldiner has explored the role of several amino acids in the EmrE by altering them through site-directed mutagenesis. The table below describes those residues and the possible functions they play in the protein based on *in vivo* and *in vitro* analysis completed by his group. One of the highly conserved and only charged membrane-embedded residue is Glu14. When EmrE loses the acidic residue at that position (E14C), there is an apparent loss in transport activity of TPP+. When Glu14 is changed to another acidic residue that lacks the methylene group (E14D), there is an ~88% decrease in the transport of TPP+ (Muth and Schuldiner, 2000). Also, E14C/D abolishes resistance to methyl viologen, acriflavine, and ethidium in liquid media *in vivo* assay. The results above suggest that this glutamic acid with an apparently high pKa of 7.5 is important for a functional EmrE and efficient multidrug transport (Muth and Schuldiner, 2000). Other charged residues found in the loops (such as Asp84 and Glu25) of EmrE were altered and have been found to not participate in the transport of the positively charged substrate (only tested with TPP+). Also, the fact that altering these residues does not affect the binding/transport activity indicates that these residues may not participate in protein

insertion or stability (Yerushalmi et al., 2001). Moreover, the replacement of other charged residues (arginine 29, 82, 106, lysine 22, and histidine 110) with cysteine or another charged amino acid produces an active EmrE that confers resistance to methyl viologen, acriflavine, and ethidium (Yerushalmi and Schuldiner, 2000). Thus, Glu14 is the only charged residue that is essential for the activity of EmrE.

Taking the above into consideration, Schuldiner group further explored the role of certain conserved residues surrounding the Glu14. There was no apparent measurable resistance against methyl viologen, ethidium and acriflavine upon Leu7, Ala10, Ile11, Gly17, and Thr18 replacement with Cys, which are found on the same face as Glu14 with the helix modeled as a 3.6 residues per turn pitch. Moreover, L7C, A10C, T18C, I11C, and G17C variants have no to very low binding to TPP+ indicating that these residues are possibly involved in substrate binding. Furthermore, when Ala10 has been mutated to Gly, Val, Cys, Leu, and Pro, the pKa of Glu14 has been changed and most of the substitutions led to a nonfunctional protein. This suggests that a close interaction between Ala10 and Glu14 is possibly happening, which mediates substrate transport (Gutman et al., 2003). The pKa of Glu14 in wild-type EmrE and in EmrE variants were measured using a pH–dependent TPP+ binding assay, where the effect of pH changes on the number of TPP+ bound was graphed in a sigmoidal function (Muth and Schuldiner, 2000). Thus, the inflection point of the graph estimates the pKa of the binding site.

Further *in vitro* studies that explored TPP+ binding upon the incubation of EmrE with NEM, a thiol alkylating agent, shows that substrate binding is not affected when Tyr4, Ile5 or Tyr6 are replaced with Cys and exposed to NEM suggesting that they are not involved in the binding area of TPP+. However, I11C, A10C, E14C and L7C all

experienced a significant decrease in TPP+ binding suggesting their involvement in the binding pocket.

The heterodimer studies that were performed by Schuldiner group, in which a mutant EmrE is mixed with a wild-type EmrE showed some interesting results (Sharoni et al., 2005). For example, G17C+wildtype-EmrE heterodimer produced a non-functional protein suggesting that Gly17 may be involved in dimer conformational changes and not only substrate binding. Also, the heterodimers of W63C+wildtype-EmrE demonstrated an inhibition of the binding activity thereby identifying that Trp63 might be essential for substrate recognition and interaction (Sharoni et al., 2005).

Further investigation of the role of tryptophan residues in EmrE by Schuldiner group (Elbaz et al., 2005) suggested the involvement of Trp63 in the binding pocket while Trp45 and 76 might be important for folding or insertion. The Trp63 replacement with Cys conferred no resistance to acriflavine, ethidium, and methyl viologen and was unable to bind TPP+ in the binding assay indicating its importance for ligand binding. While the cysteine substitutions to Trp31, Trp45, Trp76 did not affect the resistance activity of EmrE to those substrates. However, double mutations of W45C-W76C lost the capability to grow on all three substrates. Therefore, the replacements of more than one Trp residue can have a deleterious effect on the function of EmrE. Moreover, the double mutant W45C-W76C when tested for methyl viologen uptake *in vitro* displayed levels of uptake similar to those of the wild-type despite the loss of activity seen *in vivo* assays (Elbaz et al., 2005). Hence, the exact function of tryptophans is still not known and it is possible that the residues Trp45 and Trp76 play a role in protein folding and/or insertion into the membrane.

The tyrosine residues in EmrE were also targeted for mutagenesis studies by the Schuldiner group to identify their significance in EmrE. When Tyr6 and Tyr53 were altered to Cys, Phe, and/or Lys, all mutants conferred resistance to acriflavine, methyl viologen, and ethidium (Rolem et al., 2006). However, when Tyr40 was substituted to Phe, Leu, Met, Ser, Thr, and Val, it didn't confer resistance to acriflavine and ethidium. Also, mutant Tyr40 displayed low TPP+ binding and low methyl viologen uptake indicating the importance of that residue in substrate binding. Moreover, the hetero-oligomer generated *in vitro* between a wild-type EmrE and Y40C displayed normal levels of TPP+ binding, which confirms the importance of this residue in substrate binding. Also, the mutation of Tyr4 to Cys reduced the resistance of the mutant significantly to all three substrates but the mutation to Phe or Trp displayed a resistance profile similar to that of the wild-type. Furthermore, Tyr4 when replaced with Cys can bind TPP+ but cannot transport TPP+ against its concentration gradient in the presence of a proton electrochemical gradient, yet can carry the substrate with its concentration gradient in the absence of proton electrochemical force. Hence, Y4C might be a residue that is required for proper coupling between substrate transport and proton gradient (Rolem et al., 2006). The alteration of Tyr60 to Cys, Phe, Ser, and Thr abolished resistance activity to all 3 substrates, resulted in low TPP+ binding, and low methyl viologen uptake. Therefore, Tyr60 was speculated to be also involved in substrate binding/recognition. The cysteine substitutions of Tyr40, Phe44 and Leu93 resulted in non-detectable expression/accumulation levels of the protein and hence they concluded that these residues were possibly involved in membrane insertion, folding, or the stability of the protein.

The residue replacements in TMS2 and TMS3 generated modified specificity to at

least one of the 3 substrates tested. For example, all amino acid replacements of TMS3 resulted in a decreased resistance to acriflavine, while replacements of 8 residues in TMS2 displayed decreased resistance to methyl viologen. These 8 residues in TMS2 cluster into two faces of the helix, where some residues tend to cluster on the same face as Tyr40 and Phe44, which have been found to be possibly involved in protein folding and the others are found on a different face suggesting that different faces of TMS might participate in different functions (Mordoch et al., 1999). Table 1.1, summarizes the 59 residues that have been previously explored (For amino acids abbreviations and structures, see Appendix D)

Table 1. 1: Previously explored amino acid residues in the protein, EmrE.

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
[2]E14	D, C	Involved in substrate binding and proton release	In vitro-binding/release assay	TPP+	Muth & Schuldiner, 2000
			In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
E25	C, D	No effect	In vitro-binding/release assay	TPP+	Yerushalmi et al., 2001
			In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
D84	C	No effect	In vitro-binding/release assay	TPP+	Yerushalmi et al., 2001
K22	R	No effect	In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
R29	C	No effect	In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
R82	C, K	No effect	In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
R106	C, K	No effect	In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000

Amino acid	**Mutated to**	**[1]Proposed role**	***In vivo/in vitro* study**	**Compounds tested**	**Reference**
H110	Δ, C	No effect	In vivo-qualitative resistance assay	ethidium, acriflavine, methyl viologen	Yerushalmi & Schuldiner, 2000
Y6	C, F, L	No effect	In vivo-qualitative resistance assay In vitro- TPP +binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
L12	C	Maybe involved in protein folding	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
A13	C	Maybe involved in protein folding	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]V15	C	No effect	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
I16	C	No effect	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]M21	C	No effect	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]L7	C	Possibly involved in substrate binding	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]A10	C, L, V, P	Possibly involved in substrate binding	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]I11	C, G	May participate in protein folding/ stability	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]G17	C	Possibly involved in conformational changes	In vivo-qualitative resistance assay In vitro-binding assay, hetero-oligomer studies	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
[2]T18	C, S, A, G	May participate in protein folding/ stability	In vivo-qualitative resistance assay In vitro-binding assay	ethidium, acriflavine, methyl viologen TPP+	Gutman et al., 2003
[2]Y40	C, F, L, M, S, T, V	Involved in substrate binding/ recognition	In vitro-binding assay+ hetero-oligomer studies	TPP+	Sharoni et al., 2005
			In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999; Rolem et al., 2006
[2]W63	C	Involved in substrate binding	In vitro-binding assay+ hetero-oligomer studies	TPP+	Sharoni et al., 2005
T28	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L30	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]W31	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]P32	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference:
S33	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
V34	C	Reduces resistance activity to acriflavine and methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
G35	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
T36	C	Reduces resistance activity to acriflavine and methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L37	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
I38	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
A42	C	Reduces resistance activity to methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
S43	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]F44	C	Reduces resistance activity to Acriflavine+ methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]W45	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L46	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]L47	C	Reduces resistance activity to acriflavine and methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
A48	C	Reduces resistance activity to Methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
Q49	C	Reduces resistance activity to methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
T50	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L51	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
A52	C	Reduces resistance activity to Methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
Y53	C	No effect	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
I54	C	Reduces resistance activity to Methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]**P55**	C	Probably involved in protein folding	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
T56	C	Probably involved in protein folding	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]**G57**	C	Probably involved in protein folding	In vivo- qualitative resistance assay/ In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
[2]S72	C	Reduces resistance activity to acriflavine when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L73	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
L74	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
S75	C	Reduces resistance activity to acriflavine when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]L93	C	Reduces resistance activity to Acriflavine+ methyl viologen when mutated	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]I94	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
A96	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999

Amino acid	Mutated to	[1]Proposed role	*In vivo/in vitro* study	Compounds tested	Reference
[2]**G97**	C	Unknown	In vivo- qualitative resistance assay In vitro- transport assay	ethidium, acriflavine, methyl viologen	Mordoch et al., 1999
[2]**W63**	C, Y, F	Involved in substrate binding	In vivo- qualitative resistance assay In vitro- transport/TPP+ binding assays	Acriflavine, methyl viologen, ethidium	Elbaz et al., 2005
[2]**W76**	C	Possibly involved in folding/insertion	In vivo- qualitative resistance assay In vitro- transport/binding assays	Acriflavine, methyl viologen, ethidium	Elbaz et al., 2005
[2]**Y60**	C, F, S, T	Involved in substrate binding/ recognition	In vivo- qualitative resistance assay In vitro- methyl viologen uptake assay, TPP+ binding assay	Acriflavine, methyl viologen, ethidium	Rolem et al., 2006
Y4	C, F, W	Involved in proper coupling between substrate transport and proton gradient	In vivo- qualitative resistance assay In vitro- methyl viologen uptake assay, TPP+ binding assay	Acriflavine, methyl viologen, ethidium	Rolem et al., 2006

[1] Each amino acid residue has been explored either with *in vivo* or *in vitro* assays to help in understanding the role it plays in the protein EmrE by altering each residue with site-directed mutagenesis to another amino acid of interest.

[2] The **blue amino acid** residues represent the amino acid residues that were also studied in this thesis.

1.5 Quaternary Cationic Compounds

Quaternary ammonium compounds (QAC) or the more generic term, quaternary cationic compounds (QCC) are permanently charged compounds with a positive atom bound to four alkyl and/or aryl groups. QCC are widely used biocides that have antimicrobial effects against a broad range of microorganisms. These compounds can be found in water treatment products, antifungal treatments, pharmaceuticals and everyday consumer products (Hegstad et al., 2010). QCC are used in preserving agents found in hair products, for their ability to provide conditioning, and found in skin, nails and lip products for their antimicrobial activity. They are also added to mouthwashes to prevent dental biofilm formation and gingivitis (Gilbert and Moore, 2005). An example of a QCC is benzalkonium, which is universally used in the majority of eye drops, nose drops (as a decongestant), facial moisturizers, facial cleansers, acne treatments, sun protection products and hand sanitizers. QCC antiseptics are also used to counter microbial influenced corrosion. Benzalkonium is commonly used in the oil field industry for such purposes.

Bacterial resistance towards QCC has been a phenomenon that is spreading among several microorganisms (Levy and Marshal, 2004). Multiple mechanisms (that were discussed earlier in Section 1.1) facilitate the resistance activity. Some of these mechanisms include the modification of the membrane composition, expression of stress response or repair systems, or the expression of efflux pumps genes. Resistance to such highly prevalent compounds is problematic especially when there is increasing evidence of coresistance between QCC and clinically relevant antibiotics (Carson et al., 2008). Previous research has illustrated a correlation between high minimum inhibitory concentrations (MIC) for benzalkonium and antibiotic resistance phenomena in staphylococcal species,

which raises the concern that extensive usage of QCC may exert a selective pressure resulting in reduced susceptibility for antibiotics (Carson et al., 2008). Such co-selection may be promoted through the acquisition of genetic units carrying antimicrobial resistance that are found on plasmids or transposons. The widespread of these plasmids and transposons lead to the increase in the number of resistant microbes or pathogenic bacteria. Thus, controlled usage of antimicrobial agents and the extensive research exploring multidrug resistance mechanisms are promising measures to limit the expansion of bacterial 'superbugs'.

QCC have several modes of action against microbial cells depending on the structure and the chemical nature of the compound. However, the main mode of action is believed to be the interaction of QCC with cell membranes (Fredell, 1994). The damage caused to the cellular membrane leads to the release of cellular content and cellular death. Other QCC may be involved in the denaturation of structural proteins and enzymes or altering the DNA. The mode of action for each compound used in this study will be outlined below.

I have studied the resistance profiles of *E. coli* for 19 QCC; tetraphenylarsonium (TPA+), tetraphenylphosphonium (TPP+), methyltriphenylphosphonium (MTP+), acriflavine (ACR), proflavine (PRO), pyronin (PY), rhodamine (RH), ethidium (EB), crystal violet (CV), hexamethylenetetramine (HE), myristalkonium (MC), cetrimide (CET), cetylpyridinium (CC and CB), cetalkonium (Ct.CL), benzalkonium (BZ), stearyltrimethylammonium (STAC), methyl viologen (MV), and dequalinium (DC) (For structures of these compounds see Figure 2.7).

Tetraphenylarsonium chloride, tetraphenylphosphonium chloride, and methyltriphenylphosphonium bromide are lipophilic salts made from inorganic and organometallic anions. They are usually used as phase-transfer catalysts (Starks, 1971). The phase-transfer catalyst helps in solubilizing salts into the organic phase hence can function as detergents. Detergents are known to disrupt the secondary and tertiary structures of proteins. Therefore, TPP+, MTP+, and TPA+ may possibly be involved in denaturing structural proteins or solubilizing bacterial membranes, which explains their antibacterial activity.

Acriflavine is an antiseptic with various applications in the medical field. Acriflavine has demonstrated anti-cancer activity by preventing blood vessels growth in tumor cells (Lee et al., 2009). This is achieved by inhibiting Hypoxia-inducible factors (HIF), which are transcription factors that promote the formation of blood vessels under hypoxia (Benizri et al., 2008). Moreover, the bactericidal activity of acriflavine involves cell membrane damaging, which affects cell permeability and results in the leakage of cellular components, thereby cellular death (Browning and Gulbransen, 1921).

Hexamethylenetetramine is a cyclic compound that is involved in the synthesis of variety of products, such as plastic, rubber additives, and food preservative (UK Food Standards Agency, 2011). The main medical use of HE involves the treatment of urinary tract infection (Greenwood an Slack, 1981). This is mainly because HE decomposes to formaldehyde and ammonia under the acidic pH of urine. Formaldehyde is known to denature proteins and cause the alkylation of nucleic acids (Russell, 1983). Therefore, the bactericidal activity of HE may involve protein denaturation and modification of nucleic acids.

Rhodamine is a fluorescent dye that possesses antibacterial activity. The bactericidal mechanism of RH remains unexplored. However, RH has been found to be an inhibitor of mitochondrial oxidative phosphorylation (Gear, 1974). The speculated mechanism of the inhibition involves the disruption of the lipid-protein interaction of adenine nucleotide translocase since RH is lipid-soluble. Therefore, we can speculate that the bactericidal activity of RH involves the disruption of the lipid-protein interaction of integral proteins that are vital for bacterial growth.

Proflavine, ethidium, crystal violet, and pyronin show antibacterial activity due to their ability to intercalate with nucleic acids. Proflavine is an acriflavine derivative with different mode of action when it comes to its bactericidal activity. It acts by intercalating between double stranded DNA, thereby disrupting DNA synthesis or leading to mutations and preventing bacterial growth (Bradbury and Linnell, 1942). Ethidium and crystal violet are also intercalating dyes that cause deformation of the DNA, thereby interrupting transcription and replication processes (Waring, 1964; Wakelin et al., 1981). Finally, pyronin is another intercalating cationic dye that preferentially interacts with RNA, hence interrupting important cellular process such as protein translation (Darzynkiewicz et al., 1986).

Methyl viologen or paraquat is an electron acceptor in redox reactions. It is used in herbicides for its quick-acting and non-selective killing ability (Bus et al., 1984). Methyl viologen can act as an antibacterial agent for its ability to interfere with electron transfer of reactions that are vital for life. Usually MV accepts electrons and transfers them to molecular oxygen forming reactive oxygen species resulting in cellular oxidative damage (Bus et al., 1984).

Dequalinium is an antimicrobial agent that was first described by Babbs et al. (1956). The proposed mechanism for this compound involves damaging the cytoplasmic membrane of the cell and it may penetrate the membrane making it to the cytoplasm (Cox, 1965). Subsequently, research by Hugo and Frier (1969) suggested that dequalinium was involved in several cellular inhibitory processes that included attacking enzymes involved in aerobic metabolism and precipitating cytoplasmic material, specifically bacterial DNA. Hence, dequalinium is considered to be a highly toxic compound to the majority of microbes.

The acyl-chained group of compounds, such as myristalkonium, cetrimide, cetylpyridinium, cetalkonium, benzalkonium, and stearyltrimethylammonium, are commonly used as general antiseptics or surfactant. MC is an antimicrobial preservative with no known mechanism of activity. Cetrimide is an antiseptic that is frequently used in cleaning products and in some pharmaceutical products used to treat skin injuries. Cetylpyridinium is another antiseptic agent that is highly used in mouthwash, toothpaste and nasal sprays for their antibacterial activity. Also, cetalkonium is used for oral treatments, such as mouth ulcers and denture sores. Benzalkonium is used in many industrial products, such as skin antiseptics, algaecides, preservatives and surfactants. Stearyltrimethylammonium is also used as a surfactant with unexplored bactericidal activity. The specifics of each one of these compounds' bactericidal mechanisms have not been explored. However, due to the lengthy acyl tails (12-16 carbons), these compounds are expected to disrupt the membrane lipid bilayer. They form mixed micellar aggregates that solubilize the membrane leading to the progressive leakage of cytoplasmic constituents, hence their designation as surfactants (Gilbert and Moore, 2005). They damage the

membrane by binding to the head group of the acidic phospholipids through their positively charged atom. Then, the hydrophobic tail interconnects with the hydrophobic membrane core. Such interactions will increase surface pressure in the exposed leaflet of the membrane and decrease membrane fluidity, causing the loss of membrane's osmoregulatory and physiological functions.

The selections of QCC compounds discussed above are widely utilized as antimicrobial agents in daily-used products. Therefore, the indiscriminate use of these compounds constitutes many challenges to the medical and the industrial fields, especially when their effects are inhibited by multidrug resistance mechanisms. As a result, it becomes essential to study the efflux pumps involved in the extrusion of QCC to help in the development of more powerful antimicrobial agents.

1.6 Hypothesis and Research Goals

The adaptability of multidrug resistance efflux systems presents a great challenge to both molecular biologists and to pharmaceutical industry. A review by Putman et al. (2000) indicated that there has been progressive research in the field of studying MDR mechanisms identifying the amino acid residues involved in substrate binding: the past decade has seen this trend continue. Some mutants of particular ABC and other secondary multidrug transporters identified the importance of aromatic and negatively charged residues in substrate binding, which leads to the belief of the potential of multiple substrate binding sites in some of the MDR proteins. However, due to the complications associated with purification and crystallizing MDR proteins that can bind a variety of substrates *in vitro*, unique approaches to understand the binding sites of these MDR are needed. Also, upon further understanding of the molecular mechanisms underlying the polysubstrate-specific nature of these transporters, it becomes possible to design new reagents/inhibitors that are able to elude such systems. For example, in order to understand the poly-substrate specificity of EmrE, one can ask whether there is a central anchoring point in the binding site that could be recognizing only very basic characteristics of all EmrE substrates or multiple amino acid residues are participating in recognizing different substrates and thus mediating the exporting activity? A review by Bay et al (2008) discussing the previous studies that mutated single amino acid residue in EmrE via site-directed mutagenesis have demonstrated different substrate resistance profiles of these EmrE variants, where some variants increased bacteria's susceptibility to ethidium and methyl viologen while others did not affect resistance.

I hypothesize that the mutation of conserved amino acids in EmrE will alter the plasticity of the ligand-binding site, which will help in identifying the specificity of these residues to a subgroup of QCC or all of the QCC tested. The research goals are thus defined as:

1. To generate different EmrE variants through targeting one conserved amino acid at a time by using site-directed mutagenesis with a two primer-polymerase chain reaction (PCR) protocol.
2. To assess the QCC resistance of the different EmrE variants in order to generate overlaps of substrate profiles with the residues that are important for EmrE functionality/specificity.

For goal (1), thirty-three highly conserved amino acid residues were altered to generate different EmrE variants that are discussed in Chapter 3. The site-directed mutagenesis protocol utilized to produce those variants is outlined in Section 2.4. The EmrE variants' plasmids were transformed in *Escherichia coli ΔacrB* and their growth phenotype was assessed on solid media containing 19 different QCC, explained in Section 2.5. For goal (2), the resistance profiles of the 33 EmrE variants will help us in deciphering the specificity of the different conserved amino acid residues to the 19 QCC, which differ in their physiochemical characteristics (See Chapter 4). Moreover, multivariate analysis of the data generated and helical wheel mapping of the different conserved residues will be explained in Chapter 5.

1.7 Contribution of Co-authors

I hereby declare that this thesis incorporates material that is a result of joint research undertaken by undergraduate project students, Simon Taylor and Leeanna El- Houjeiri under the supervision of Dr. Raymond J. Turner and Dr. Denice Bay. The collaboration involved the designing of 42 primers for PCR based site-directed mutagenesis that was altering one codon at a time in *emrE* sequence. They were 42 primers that were designed by Dr. Raymond J. Turner and Dr. Denice Bay while Simon Taylor and Leeanna El-Houjeiri have carried out the PCR-based site-directed mutagenesis of six each for an initial 12 variants, further mutagenesis led them to generate the 21 single codon mutants of EmrE that I refer to as EmrE variants. The variants were generated during the Fall of 2011 and the Winter of 2012 semesters and Spring 2012 and they were; L7A, A10C, E14D, G17A, W31A, Y40A, W45A, A59C, Y60A, I62C, W63A, S64C, V66C, V69C, S72C, W76A, L93C, I94C, G97C, L103C, and S105C. I transformed the 21 EmrE variants in an expression strain of *Escherichia coli* and used them for Quaternary Cationic Compounds (QCC) resistance assay to obtain part of the results presented in this thesis.

Moreover, I would like to declare that in Section 2.5.1, Table 2.5 was a product of the spectrophotometer-based QCC resistance assay that was in part developed by the undergraduate student, Simon Taylor. The assay will be described in details in Chapter 2, Section 2.5.1.

I am aware of the University of Calgary Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained permission from the co-authors to include the above materials in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work.

Chapter 2: General Methods

2.1 Plasmid

The expression plasmid that has been used for this study is pMS119EH with the infusion of the *emrE* gene sequence. As described in the figures below, the empty vector was utilized as a negative control for the quaternary cationic compounds (QCC) resistance analysis (Section 2.5) while the vector containing *emrE* was utilized to generate the different EmrE variants as well as serving as a comparator positive control of EmrE function (Section 2.4).

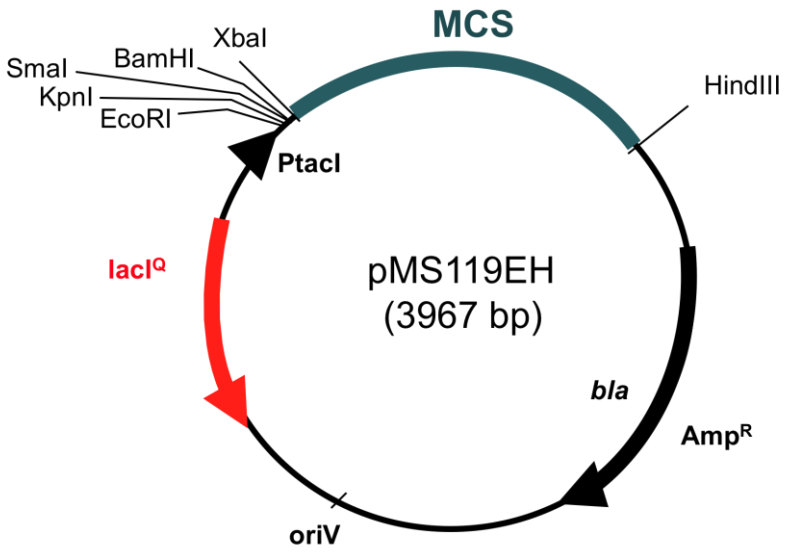


Figure 2. 1: The empty vector, pMS119EH.

The multiple cloning site (MCS), where the amplified *E. coli emrE* gene was cloned by exploiting the restriction sites (XbaI and HindIII) that are spanning the MCS region. The ampicillin resistance marker on the plasmid will be exploited for screening purposes upon the transformation of the plasmid. For further details on the plasmid, refer to Furste *et al*. (1986).

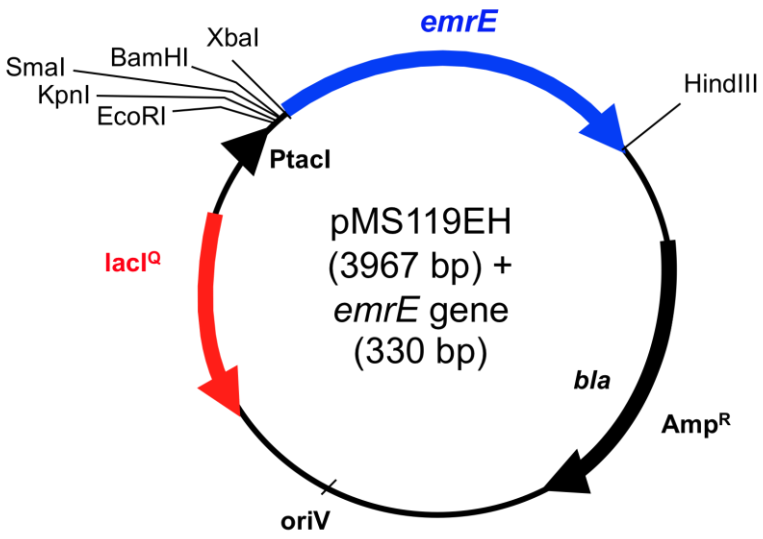


Figure 2. 2: The pEmr11 vector.

PCR amplified *E. coli emrE* gene from pASP505 cloned into XbaI/ HindIII sites of the multiple cloning site (MCS) of pMS119EH vector.

2.2 Strains

Two strains of *Escherichia coli* were used in this study, one of which is DH5α and the other is JW0451. The table below describes in detail the genotypes of the strains and their significance in the study.

Table 2. 1: *E. coli* strains used in this study.

Strain	Genotype	Significance in the study	Reference
DH5α	F– Δ*lac*Z, ΔM15, Δ(*lac*ZYA-*arg*F), U169 *rec*A1 *end*A1 *hsd*R17 (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1	-The *endA1* mutation allows for lower endonuclease degradation, which ensures higher plasmid transfer rates. -The *recA1* mutation reduces homologous recombination for a more stable insert.	Taylor et al., 1993

Strain	Genotype	Significance in the study	Reference
JW0451	F– *Δ(araD-araB)567, ΔlacZ4787*(::rrnB-3), *ΔacrB747::kan, λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514*	The *acrB* deletion disrupts the function of the AcrAB-TolC drug efflux pump in order to prevent masking the phenotype observed with EmrE.	Baba et al., 2006

2.3 Media

Luria Broth (LB) was used for diluting culture and for preparing overnight culture. It was made from 10g of tryptone powder, 5g of yeast extract powder, and 5g of sodium chloride mixed in 1L of distilled, deionized water (ddH_2O).

Also, Luria (LB) agar plates were prepared by mixing 1L of prepared LB (same as above) with 12.5g of agar (Invitrogen). After autoclaving the LB agar media, the media is allowed to cool and then 1mL of ampicillin at 100mg/mL was added. Then, LB media is poured in sterile circular Petri dishes for later use.

For long-term storage of cell culture, 2mL of cell culture at log phase was mixed with 1mL of sterile LB+24% (v/v) dimethylsulfoxide (DMSO) in glass cryovials and stored at -80°C. The preparation of LB+24% (volume-by-volume) DMSO involved dissolving 2.5g of Tryptone powder, 1.25g of yeast extract powder, 1.25g of sodium chloride, and 60mL of 100% DMSO in a total volume of 190mL of ddH_2O.

2.4 Mutagenesis Protocol

A two primer PCR-based site-directed mutagenesis (PCR-SDM) has been used to generate the different variants of the protein EmrE. This allows one to target a conserved

amino acid in the *emrE* sequence and change the DNA codon to the amino acid codon of interest. The selection of the amino acid to be altered is described in details in Section 3.1. The mutagenesis protocol involved primer design, PCR-SDM (Figure 2.4), plasmid amplification/isolation, and the confirmation of the newly generated *emrE* variants' DNA sequences. The wild-type sequence of *emrE* gene is depicted in Figure 2.3.

```
ATGAACCCTTATATTTATCTTGGTGGTGCAATACTTGCAGAGGTCATTGGTACAACCTTA
 M  N  P  Y  I  Y  L  G  G  A  I  L  A  E  V  I  G  T  T  L
ATGAAGTTTTCAGAAGGTTTTACACGGTTATGGCCATCTGTTGGTACAATTATTTGTTAT
 M  K  F  S  E  G  F  T  R  L  W  P  S  V  G  T  I  I  C  Y
TGTGCATCATTCTGGTTATTAGCTCAGACGCTGGCTTATATTCCTACAGGGATTGCTTAT
 C  A  S  F  W  L  L  A  Q  T  L  A  Y  I  P  T  G  I  A  Y
GCTATCTGGTCAGGAGTCGGTATTGTCCTGATTAGCTTACTGTCATGGGGATTTTTCGGC
 A  I  W  S  G  V  G  I  V  L  I  S  L  L  S  W  G  F  F  G
CAACGGCTGGACCTGCCAGCCATTATAGGCATGATGTTGATTTGTGCCGGTGTGTTGATT
 Q  R  L  D  L  P  A  I  I  G  M  M  L  I  C  A  G  V  L  I
ATTAATTTATTGTCACGAAGCACACCACATTAA
 I  N  L  L  S  R  S  T  P  H  -
```

Figure 2. 3: The DNA and the amino acid sequence of EmrE.

The letters in **blue** indicate the codon sequence of *emrE* (5'→3') while the letters in red are the amino acid (1 letter abbreviation) sequence. The protein consists of 110 amino acids and the – indicates the stop codon.

2.4.1 Primer Design

Using the template sequence of *emrE*, 84 primers were generated (Table 2.2) in order to produce a total of 42 EmrE variants. However, there were only 33 successful EmrE variants that were generated.

In order to ensure a high success rate for the mutations generated in the *emrE* sequence targeting one codon, several factors were taken into consideration. First, since most of the changes of the conserved amino acids were to cysteine or alanine, we used

codons for these two amino acids that were most prevalent in the native *emrE* sequence (i.e. for Alanine, the codon was GAG while for cysteine, the codon was TGT). With hydrophobic membrane proteins, scanning with cysteine has the advantage of the cysteine side chain being an intermediate bulk, amenable to highly specific modification, and relatively hydrophobic (Dunten et al., 1993; Frillingos et al., 1998). Also, considering the amino acid distributions in integral membrane protein structures, cysteine seems to have no preference as to where it would be localized in the protein (Ulmschneider and Sansom, 2001). Hence, the introduction of cysteine residues in any location of the transmembrane protein should not have a detrimental effect on protein structure. Alanine is also used because its non-bulky, chemically inert, and has a methyl functional group that mimics the secondary structure preferences of many other amino acids (Morrison and Weiss, 2001). Second, the primers generated were between 34-43 nucleotides long and the targeted codon was embedded in the center of the primer. Thus, the melting temperature of the primers was in the range of 67-78°C. Also, for almost all of the primers, each primer ended with G or C nucleotides in order to enhance the annealing of the primer.

The primer sequences for both forward and reverse oligonucleotides were submitted to Integrated DNA Technologies. Upon the arrival of the primers, the oligonucleotides were resuspended in sterile (DNase/RNase free) ddH_2O to a concentration of 0.1mM. Then, two dilutions (10μM and 1μM) of each primer were prepared for later use in PCR-SDM and stored at -10°C.

Table 2. 2: The primer pairs used to generate single codon replacement of conserved amino acids in *emrE*.

[1]EmrE Variant	[2, 3] Primers
L7A	Forward: 5′GAACCCTTATATTTAT**GCA**GGTGGTGCAATACTTGC3′ Reverse: 5′GCAAGTATTGCACCACC**TGC**ATAAATATAAGGGTTC3′
A10C	Forward: 5′CCTTATATTTATCTTGGTGGT**TGT**ATACTTGCAGAGGTC3′ Reverse: 5′GACCTCTGCAAGTAT**ACA**ACCACCAAGATAAATATAAGG3′
I11C	Forward: 5′CTTGGTGGTGCA**TGT**CTTGCAGAGGTCATTGGTACAACC3′ Reverse: 5′GGTTGTACCAATGACCTCTGCAAG**ACA**TGCACCACCAAG3′
E14D	Forward: 5′GTGGTGCAATACTTGCA**GAC**GTCATTGGTACAACC3′ Reverse: 5′GGTTGTACCAATGAC**GTC**TGCAAGTATTGCACCAC3′
V15C	Forward: 5′GTGGTGCAATACTTGCAGAG**TGT**ATTGGTACAACCTTAATG3′ Reverse: 5′CATTAAGGTTGTACCAAT**ACA**CTCTGCAAGTATTGCACCAC3′
G17A	Forward: 5′CTTGCAGAGGTCATT**GCG**ACAACCTTAATGAAGTTTC3′ Reverse: 5′GAAACTTCATTAAGGTTG**TCG**CAATGACCTCTGCAAG3′
T18C	Forward: 5′CTTGCAGAGGTCATTGGT**TGT**ACCTTAATGAAGTTTTCAG3′ Reverse: 5′ CTGAAAACTTCATTAAGGT**ACA**ACCAATGACCTCTGCAAG3′
L20C	Forward: 5′CAGAGGTCATTGGTACAACC**TGT**ATGAAGTTTTCAGAAGG3′ Reverse: 5′CCTTCTGAAAACTTCAT**ACA**GGTTGTACCAATGACCTCTG3′
M21C	Forward:5′CATTGGTACAACCTTA**TGT**AAGTTTTCAGAAGGTTTTACACGG3′ Reverse:5′CCGTGTAAAACCTTCTGAAAACTT**ACA**TAAGGTTGTACCAATG3′
G26C	Forward: 5′CCTTAATGAAGTTTTCAGAA**TGT**TTTACACGGTTATGGCC3′ Reverse: 5′ GGCCATAACCGTGTAAA**ACA**TTCTGAAAACTTCATTAAGG3′
F27C	Forward: 5′GAAGTTTTCAGAAGGT**TGT**ACACGGTTATGGCCATCTG3′ Reverse: 5′CAGATGGCCATAACCGTGT**ACA**ACCTTCTGAAAACTTC3′
W31A	Forward: 5′GAAGGTTTTACACGGTTA**GCA**CCATCTGTTGGTAC3′ Reverse: 5′GTACCAACAGATGG**TGC**TAACCGTGTAAAACCTTC3′
P32C	Forward: 5′GGTTTTACACGGTTATGG**TGT**TCTGTTGGTACAATTATTTG3′ Reverse: 5′CAAATAATTGTACCAACAGA**ACA**CCATAACCGTGTAAAACC3′
Y40A	Forward: 5′CTGTTGGTACAATTATTTGT**GCA**TGTGCATCATTCTGG3′ Reverse: 5′CCAGAATGATGCACA**TGC**ACAAATAATTGTACCAACAG3′
F44C	Forward: 5′GTTATTGTGCATCA**TGT**TGGTTATTAGCTCAGACGCTGG3′ Reverse: 5′CCAGCGTCTGAGCTAATAACCA**ACA**TGATGCACAATAAC3′

EmrE Variant	Primers
W45A	Forward: 5′GTTATTGTGCATCATTC**GCA**TTATTAGCTCAGACGCTG3′ Reverse: 5′CAGCGTCTGAGCTAATAA**TGC**GAATGATGCACAATAAC3′
L47C	Forward: 5′GCATCATTCTGGTTA**TGT**GCTCAGACGCTGGCTTATATTCC3′ Reverse: 5′ GGAATATAAGCCAGCGTCTGAGC**ACA**TAACCAGAATGATGC3′
P55C	Forward: 5′GCTCAGACGCTGGCTTATATT**TGT**ACAGGGATTGCTTATGC3′ Reverse: 5′ GCATAAGCAATCCCTGT**ACA**AATATAAGCCAGCGTCTGAGC3′
G57C	Forward: 5′CTGGCTTATATTCCTACA**TGT**ATTGCTTATGCTATCTGG3′ Reverse: 5′CCAGATAGCATAAGCAAT**ACA**TGTAGGAATATAAGCCAG3′
A59C	Forward: 5'CTTATATTCCTACAGGGATT**TGT**TATGCTATCTGGTC3' Reverse: 5'GACCAGATAGCATA**ACA**AATCCCTGTAGGAATATAAG3'
Y60A	Forward: 5′CCTACAGGGATTGCT**GCA**GCTATCTGGTCAGGAGTC3′ Reverse: 5′GACTCCTGACCAGATAGC**TGC**AGCAATCCCTGTAGG3′
I62C	Forward: 5'CAGGGATTGCTTATGCT**TGT**TGGTCAGGAGTCGG3' Reverse: 5'CCGACTCCTGACCA**ACA**AGCATAAGCAATCCCTG3'
W63A	Forward: 5′GGATTGCTTATGCTATC**GCA**TCAGGAGTCGGTATTG3′ Reverse: 5′CAATACCGACTCCTGA**TGC**GATAGCATAAGCAATCC3′
S64C	Forward: 5'GCTTATGCTATCTGG**TGT**GGAGTCGGTATTGTCCTG3' Reverse: 5'CAGGACAATACCGACTCC**ACA**CCAGATAGCATAAGC3'
V66C	Forward: 5'GCTTATGCTATCTGGTCAGGA**TGT**GGTATTGTCCTGATTAG3' Reverse: 5'CTAATCAGGACAATACC**ACA**TCCTGACCAGATAGCATAAGC3'
I68C	Forward: 5′ CTATCTGGTCAGGAGTCGGT**TGT**GTCCTGATTAGCTTAC 3′ Reverse: 5′GTAAGCTAATCAGGAC**ACA**ACCGACTCCTGACCAGATAG3′
V69C	Forward: 5'GGTCAGGAGTCGGTATT**TGT**CTGATTAGCTTACTG3' Reverse: 5'CAGTAAGCTAATCAG**ACA**AATACCGACTCCTGACC3'
S72C	Forward:5'GGTCAGGAGTCGGTATTGTCCTGATT**TGT**TTACTGTCATGGG3' Reverse:5'CCCATGACAGTAA**ACA**AATCAGGACAATACCGACTCCTGACC3'
W76A	Forward: 5′CTGATTAGCTTACTGTCA**GCA**GGATTTTTCGGCCAAC3′ Reverse: 5′GTTGGCCGAAAAATCC**TGC**TGACAGTAAGCTAATCAG3′
Q81C	Forward:5′CGGC**TGT**CGGCTGGACCTGCCAGCCATTATAGGCATGATGTTG3′ Reverse:5′CAACATCATGCCTATAATGGCTGGCAGGTCCAGCCG**ACA**GCCG3′
D84C	Forward:5′CAACGGCTG**TGT**CTGCCAGCCATTATAGGCATGATGTTGATTTG3′ Reverse:5′CAAATCAACATCATGCCTATAATGGCTGGCAG**ACA**CAGCCGTTG3′
G90C	Forward: 5′CTGGACCTGCCAGCCATTATA**TGT**ATGATGTTGATTTGTG3′ Reverse: 5′ CACAAATCAACATCAT**ACA**TATAATGGCTGGCAGGTCCAG3′
L93C	Forward: 5′CATTATAGGCATGATG**TGT**ATTTGTGCCGGTGTGTTG3′ Reverse: 5′CAACACACCGGAACAAAT**ACA**CATCATGCCTATAATG3′
I94C	Forward: 5'GCCATTATAGGCATGATGTTG**TGT**TGTGCCGGTGTGTTG3' Reverse: 5'CAACACACCGGCACA**ACA**CAACATCATGCCTATAATGGC3'

EmrE Variant	Primers
G97C	Forward: 5'GATGTTGATTTGTGCC**TGT**GTGTTGATTATTAATTTATTG3' Reverse: 5'CAATAAATTAATAATCAACAC**ACA**GGCACAAATCAACATC3'
V98C	Forward: 5′GCCGGT**TGT**TTGATTATTAATTTATTGTCACGAAGCACAC3′ Reverse: 5′GTGTGCTTCGTGACAATAAATTAATAATCAA**ACA**ACCGGC3′
N102C	Forward: 5′GGTGTGTTGATTATT**TGT**TTATTGTCACGAAGCACACCAC3′ Reverse: 5′GTGGTGTGCTTCGTGACAATAA**ACA**AATAATCAACACACC3′
L103C	Forward: 5'GTGTTGATTATTAAT**TGT**TTGTCACGAAGCACACC3' Reverse: 5'GGTGTGCTTCGTGACAA**ACA**ATTAATAATCAACAC3'
S105C	Forward: 5'GTTGATTATTAATTTATTG**TGT**CGAAGCACACCAC3' Reverse: 5'GTGGTGTGCTTCG**AC**ACAATAAATTAATAATCAAC3'

[1] Each variant is shown as a one-letter abbreviation for the amino acid that has been altered and the second letter is for the new amino acid that has been generated.

[2] The codon that has been changed to alanine, cysteine or aspartic acid has been **bolded.**

[3] The reverse primer is in blue while the forward primer is in black. All primers were designed for the plasmid pEMR11 (Figure 2.2).

2.4.2 PCR-SDM

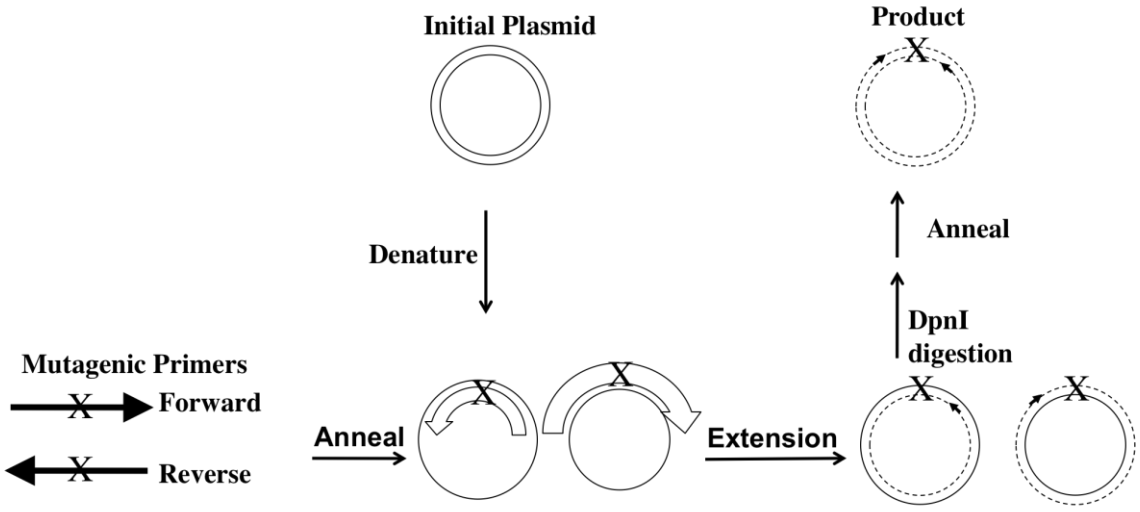


Figure 2. 4: The basic PCR-site directed mutagenesis reaction sequence.
Adapted from: Zheng *et al*., 2004.

For PCR-SDM reactions, the KAPA HiFi PCR kit from KAPABiosystems was used. The PCR reaction was set up in small thin-wall PCR tubes on ice, where each tube contained the reaction buffer, plasmid, dNTPs, nuclease free water and the appropriate primers (Table 2.3). Finally, 1uL of the KAPA Hifi Taq polymerase was added to each reaction tube.

Table 2. 3: PCR-site directed mutagenesis master reaction.

All of the reactions utilized the KAPA HiFi PCR kits components, which came with its PCR buffer, dNTP mix, magnesium chloride, and Taq polymerase.

Component	**Volume**
Fidelity 5X KAPA HiFi buffer	10uL
pEMR11 vector	0.5uL
10mM KAPA dNTPs mix	2uL
10mM $MgCl_2$	2uL
10µM Forward primer	5uL
10µM Reverse primer	5uL
1 U/µl KAPA HiFi DNA Polymerase	1uL
Sterile ddH2O (DNase/RNase Free)	24.5uL

The reaction mixtures were then placed in the thermocycler and ran at the program indicated in Table 2.4. After running the PCR products on 0.1% agrose gel, it becomes possible to assess the success of the PCR reaction when the negative control (unamplified plasmid) is run on the gel along with PCR products. When the PCR products show a thicker band than the negative control at the expected size of the plasmid (~4500bp), it means that the template plasmid has been amplified. Although all primers were designed to be used at the same temperature, some PCR reactions were successful at an annealing temperature of 60°C while others were successful at 58°C. This was because some of the EmrE variants did not have a discernable band on the agrose gel at an annealing temperature of 58°C and an accumulation of primer dimers. Thus, the annealing temperature was increased to 60°C, where a notable decrease in primer dimerization was observed and an accumulation of the band at ~4500bp occurred. The variants that were generated using an annealing temperature of 60°C were I11C, V15C, T18C, M21C, G26C, F27C, L7A, A10C, E14D, G17A, W31A, Y40A, W45A, A59C, Y60A, I62C, W63A, S64C, V66C, V69C, S72C, W76A, L93C, I94C, G97C, L103C and S105C, and the ones generated using an annealing temperature of 58°C were P32C, F44C, L47C, P55C, G57C and G90C.

Table 2. 4: Thermocycler program.

Step	Temperature (°C)	Duration (min:sec)	Cycles
Initial Denaturation	95	3:00	1
Denaturation	98	0:20	16
Annealing	58-60	0:20	
Extension	72	3:00	
Final extension	72	4:30	1

Following the PCR-SDM reaction, 1uL of DpnI restriction enzyme is added to each tube in order to digest the methylated (template) pEmrE11, leaving the newly synthesized plasmid, which is unmethylated.

2.4.3 Plasmid Amplification/Isolation

Plasmid amplification occurred through the transformation of DpnI digested PCR products in *E. coli* DH5α. The heat-shock transformation method was utilized by adding 10uL of digested PCR-SDM products to 100uL of competent DH5α and keeping them on ice for 15 minutes. The mixture of DNA and the cells were then heat-shocked for 90 seconds at 42°C, following which the tubes were immediately transferred back on ice. Then, 1mL of sterile LB was added to each tube and incubated at 37°C for 1 hour in an incubator that shook the tubes at 180rpm. After an hour the tubes were spun down at 10,000rpm for 1 minute and the supernatant was removed and the pellet was resuspended with fresh 100uL of LB media. Then, the 100uL of culture for each PCR-SDM reaction was spread on plates containing 100ug/mL ampicillin and were incubated at 37°C for 16 hours. The ampicillin helped in selecting for transformed DH5α cells.

After 16 hours, discrete colonies formed on each of the plates. Three colonies from each plate were selected for amplification, plasmid isolation, and ultimately sequencing. Colonies were picked up using an autoclaved sterile wooden dowel and then placed in a test

tube with 5mL of LB and 5uL of 100mg/mL ampicillin. This was repeated so that there were three tubes for each EmrE variant (plate). The tubes were then placed in a shaking incubator at 37°C for 16 hours at 180rpm.

Following the 16 hours-incubation, the *E. coli* strains transformed with the EmrE variants had reached confluence. Using the standard plasmid miniprep isolation procedure (E.Z.N.A Plasmid Mini Kit II, from OMEGA Bio-Tek), 3-5mL of the cell mixture was used and repeated for each EmrE variant/test tube. The isolated plasmids were eluted in 40μL of sterile ddH_2O (DNase/RNase free).

The isolated plasmids were then sent to Eurofins Genomics Sequencing Service to obtain DNA sequences of the EmrE variants. The output sequences were then aligned with pEMR11 sequence using CLUSTAL W alignment program for confirmation of the variants generated.

The confirmed EmrE purified plasmids transformed into the cell strain *Escherichia coli* JW0451 *ΔacrB* using the previously described heat-shock transformation method. In addition, two control plasmids were transformed in this strain, which include the pEMR11 and the empty vector pMS119EH (Figure 2.1/2.2). Each one of the transformed cultures was then mixed with LB+24% DMSO solution and stored at -80°C for later usage in the QCC resistance assay.

2.5 QCC Resistance Assay

2.5.1 Spectrophotometer-based Resistance Assay

QCC were prepared with initial concentrations ranging from 10-20mg/mL and

stored at -20°C. Then, a serial dilution of factor ½ is prepared for each compound based on the concentration ranges seen in Table 2.5. In this assay, the EmrE variants along with the two controls were amplified for 16 hours before the assay in 5mL of LB and shaken at 37°C. Following that, a small aliquot of each test tube was used to determine the optical density (OD) at 550nm for each culture. Then, each culture was adjusted to within +/-0.01 units of each other OD values by adding LB to the requisite tubes. Then, normalized cultures were further diluted to make 1/10 and 1/100 cell dilution solutions for each EmrE variant/control. A 96-well plate was used for this assay where 80µL of LB was placed in every column except #12, which is not used. Next 20µL of the QCC stock solution of choice was added to each well in column 1. After which 20µL from column 1 was taken and added to column 2 and then mixed. Following mixing, 20µL was taken from each well in column 2 and added to column 3. This procedure was repeated until column 10, no drug was added to column 11. Column 12 was for waste.

Once the drug has been added the next step was to add the appropriate cell solutions to the rows of the plate. The first row has no cells added; instead 80µL of LB was placed in each well of this row. Next, the 1/10-cell solution of a particular culture was added to rows B, C, and D respectively. In rows E, F, and G, the same procedure was carried out, but with the 1/100-cell solution. Row H was kept empty; a pictorial description of the 96-well plate can be seen in Figure 2.5.

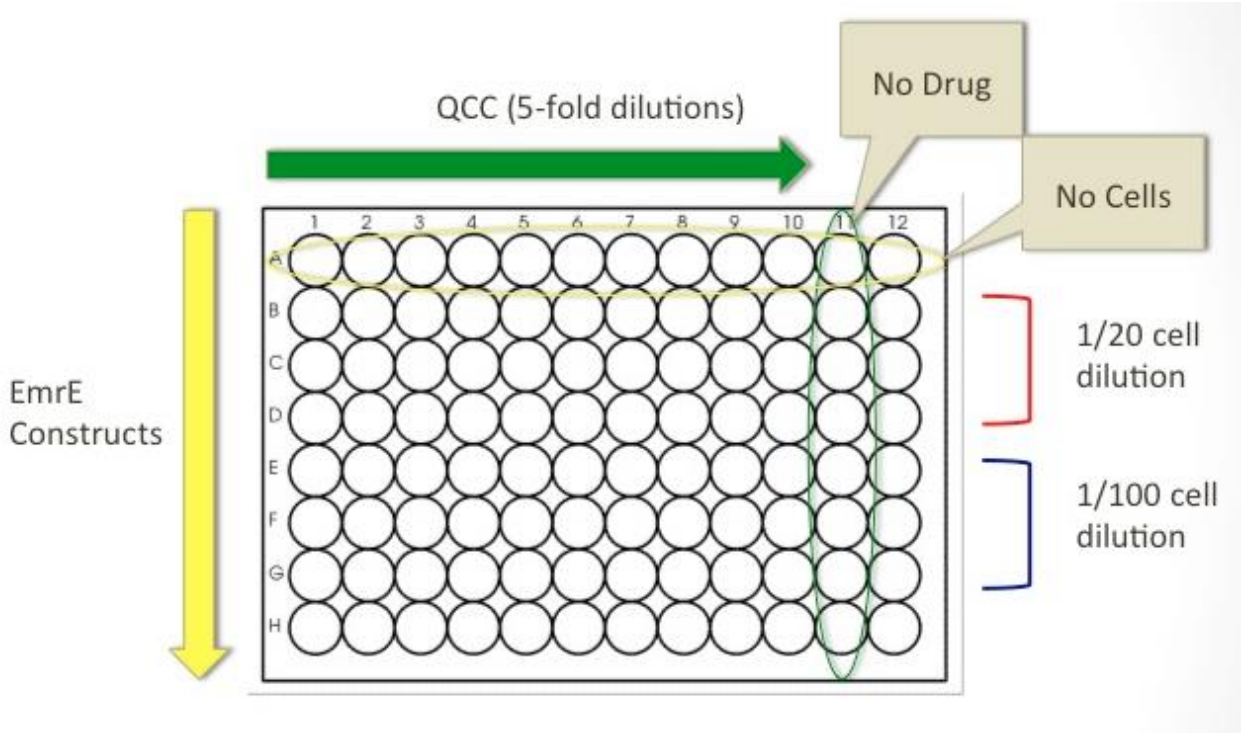


Figure 2. 5: Microtitre plate Spectrophotometer-based resistance assay set up.

These plates were then incubated at 37°C for 16 hours on a shaker at 120rpm. After 16 hours, the plates were removed from the incubator and placed in a plate reader; wherein the OD values at 550nm of each well in each plate were read. For each EmrE variant, an X/Y plot was generated with the change in OD values against the QCC concentrations. Along with each variant, the plot of the wild-type EmrE (pEMR11) was plotted and a sigmoidal line of best fit was added to each data set. Therefore, the minimum inhibitory concentration (MIC) was determined by finding the concentration at the inflection point of the sigmoidal curve, which is the point on the curve where the curvature changes sign from positive curvature to negative curvature. Several trials of the assay helped in determining the concentration ranges of each QCC needed to observe an effect on culture growth seen in Table 2.5.

Table 2. 5: Concentration ranges used for QCC resistance assay.

Drug Name:	**[1] Concentration ranges (ug/mL)**									
Tube #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**[2] 10**
Methyl Viologen	512	256	128	64	32	16	8	4	2	0
Dequalinium chloride	500	250	125	62.5	31.3	15.6	8	4	2	0
Acriflavine	128	64	32	16	8	4	2	1	0.5	0
Proflavine	128	64	32	16	8	4	2	1	0.5	0
Crystal Violet	64	32	16	8	4	2	1	0.5	0.25	0
Rhodamine 6G	64	32	16	8	4	2	1	0.5	0.25	0
Pyronin Y	64	32	16	8	4	2	1	0.5	0.25	0
Ethidium bromide	256	128	64	32	16	8	4	2	1	0
Hexamethylenetetramine	4000	2000	1000	500	250	125	62.5	31.3	15.6	0
Benzalkonium chloride	64	32	16	8	4	2	1	0.5	0.25	0
Cetrimide (CTAB-cetrimonium bromide)	128	64	32	16	8	4	2	1	0.5	0
*Stearyltrimethylammonium chloride	2000	1000	500	250	125	62.5	31.3	15.6	7.8	0
Cetalkonium chloride (Banjela) (16 Chain)	128	64	32	16	8	4	2	1	0.5	0
Myristalkonium chloride (14C chain)	64	32	16	8	4	2	1	0.5	0.25	0
Cetylpyridinium bromide	512	256	128	64	32	16	8	4	2	0
Cetylpyridinium chloride	512	256	128	64	32	16	8	4	2	0
Tetraphenylphosphonium chloride	32	16	8	4	2	1	0.5	0.25	0.125	0
Methyltriphenyl phosphonium bromide	240	120	60	30	15.5	7.75	3.9	2	1	0
Tetraphenylarsonium chloride	64	32	16	8	4	2	1	0.5	0.25	0

[1] The concentration ranges were obtained by the spectrophotometer-based resistance assay.

[2] Tube10 was the negative control where no QCC was added.

2.5.2 Plate-based Resistance Assay

The above spectrophotometer-QCC based method faced a number of complications, particularly the ability of certain QCC compounds to absorb light at the same wavelength as the culture (550nm). Therefore, as a start the concentration ranges from Table 2.5 initially prepared for the broth bath assay were used to generate the different concentrations of QCC on the solid LB agar.

The other advantage of using a plate-based resistance assay rather than an spectrophotometer-based assay is the fact that the plate assay will allow the determination of the MIC when zero colony growth is observed on a plate, which is a more definite observation/value than the inflection point of a sigmoidal graph generated from optical density. Moreover, other phenotypic changes of the EmrE variants can be noted from colony morphology (i.e. the ability to absorb the dye from the color of the colony).

The media for the plate-based assay was prepared from LB agar (prepared as in Section 2.3), which was mixed with one QCC at a time for the resistance assay preparation. Serial dilutions with a dilution factor of ½ were prepared for each QCC according to Table 2.5. The dilutions of the QCC were prepared in 1.6mL microcentrifuge tubes and the content of the tubes were then poured into circular Petri dishes with 20mL of LB agar, mixed, and allowed to solidify at 4°C (Figure 2.6).

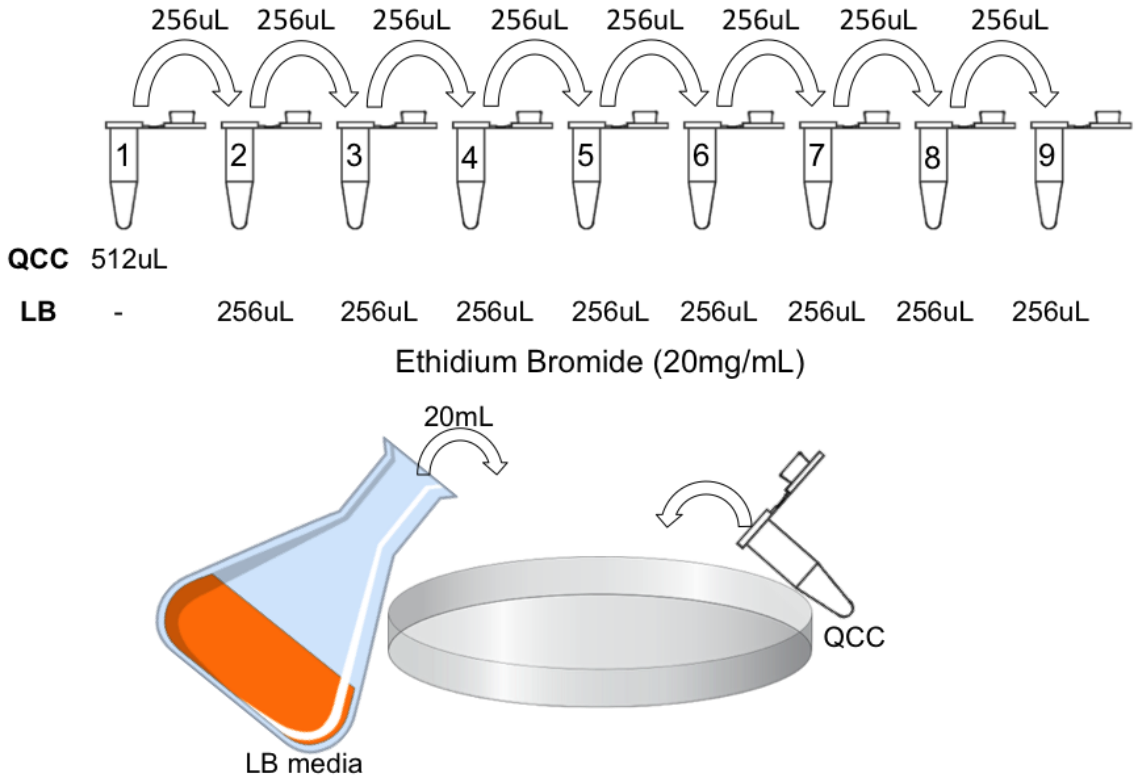


Figure 2. 6: Preparation of QCC plates.

Using Ethidium bromide as an example, the highest concentration of ethidium bromide in the MIC ranges Table 2. E is 256ug/mL. Thus, with 20mL of LB agar in the Petri dish, 256uL of the 20mg/mL Ethidium bromide is needed to prepare the highest concentration plate ($C_1V_1=C_2V_2$, where C is the concentration and V is the volume so in this case 20mg/mL x 0.512mL=0.256mg/mL x 20mL). With 9 tubes of ½ dilutions of Ethidium bromide, 9 plates for each QCC were prepared leaving the 10^{th} plate with only 20mL of agar (growth control). The remaining 18 QCC plates were also prepared in a similar fashion to the figure above.

There were 19 quaternary cationic compounds used in this study (Table 2.5). These QCC compounds were divided into four groups based on the characteristics of their chemical structures (Figure 2.7). For example, sphere-forming compounds form spherical shaped structures when observed by X-ray crystallography, as a result of Sp4 orbitals and the bonding of the bulky R groups. Also, the poly-aromatic group consists of QCC

containing multiple aromatic rings without lengthy acyl chains. The acyl-chained QCC include the group of compounds with long saturated acyl tails ranging from 12-16 carbons. Finally, the poly-charged group includes the compounds that contain more than one charged atom.

The procedure I worked for the plate-based QCC assay, which I will refer to as QCC resistance assay, was described in Figure 2.6. After the preparation of the LB agar containing a particular QCC, the transformed *E. coli* JW0451Δ*AcrB* cells containing the EmrE variants were grown for 16 hours at 37°C shaker with ampicillin at a concentration of 100ug/mL. Then, the mixture of cells was normalized to an optical density value of 1.6 using the spectrophotometer at a wavelength of 550nm. At this optical density value, the cells were at mid-log phase. Then, the cells were further diluted to a dilution factor of 1/10 and 1/100 for each EmrE variant.

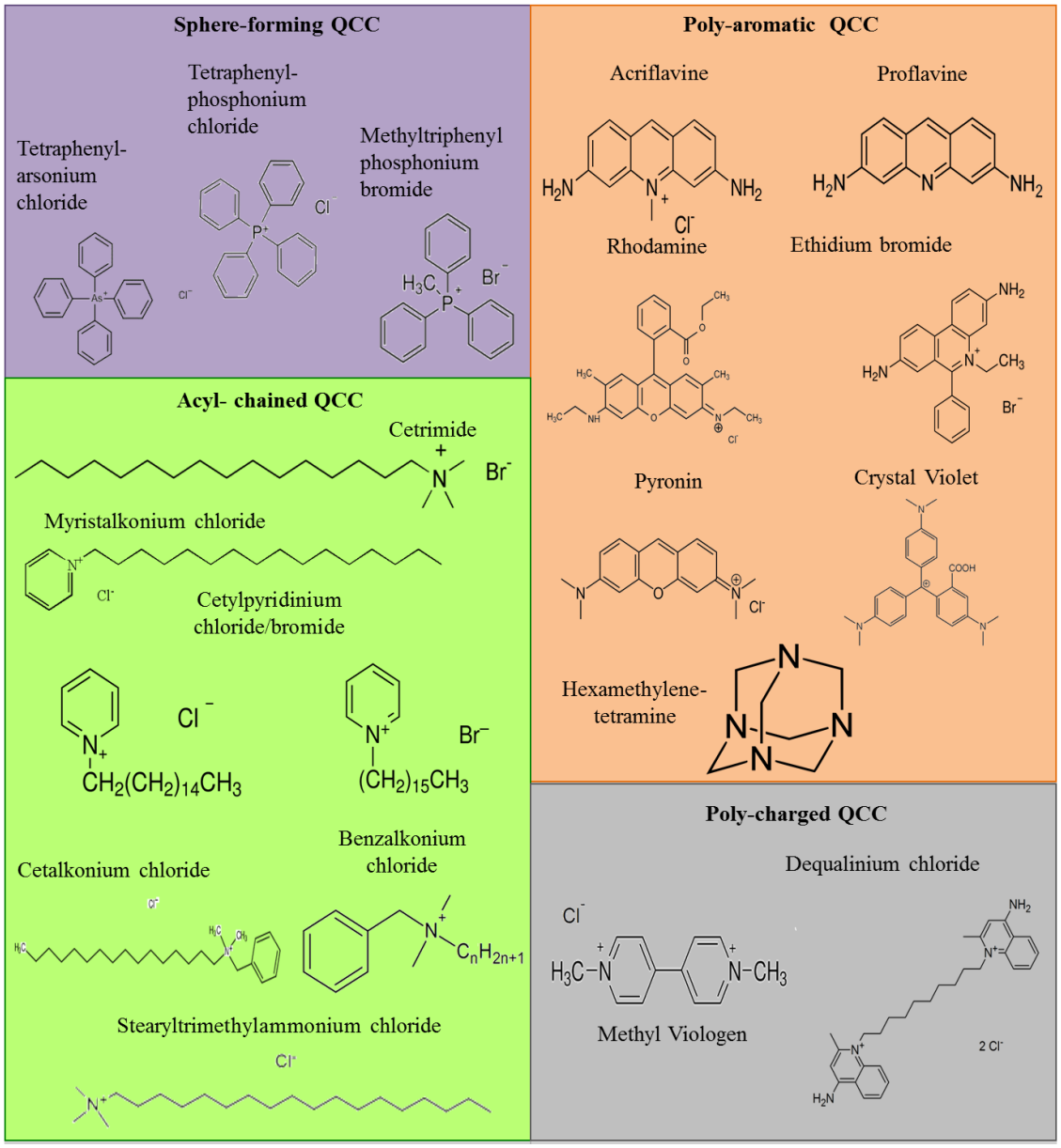


Figure 2. 7: The quaternary cationic compounds of this study.

The list includes 19 QCC with varied structural features and they have been categorized based on general structural characteristic into four groups.

Using a 96-well plate, 48 wells were filled with a total of 10 EmrE variants and the two controls for each dilution. The two controls for the QCC assay include one *E. coli* JW0451 *ΔacrB* transformed with an empty pMS119EH vector and one *E. coli* JW0451 *ΔacrB* transformed with pMS119EH vector containing the wild-type *emrE* gene. Considering that each one of the 12 samples was diluted to 1/10 and 1/100 culture, there were a total of 24 different samples. In addition, for internal experimental replicates, each sample was placed in 2 different wells, therefore, completing the 48 wells (Figure 2.8 illustrates an example of one assay). Using the sterilized 48-pins Microplate replicator, the replicator was inserted into the 48 wells of the 96-well plate and then gently spotted on an LB plate containing a specific QCC concentration; usually spotting started from the plate containing the lowest concentration of QCC to the highest and the sterilization of the 48-pins happened each time the QCC changed.

There were two sets of plates for each QCC concentration making each EmrE variant of a particular dilution have a total of 4 replicates (4 replicates for 1/10 & 4 for 1/100).

L103C	**pMS119EH**	**S64C**	**WT**	**V66C**	**L103C**
A59C	**L93C**	**S64C**	**S105C**	**V66C**	**S72C**
A59C	**L93C**	**WT**	**S105C**	**pMS119EH**	**I62C**
A59C	**L93C**	**pMS119EH**	**S105C**	**V66C**	**V69C**
A59C	**L93C**	**G97C**	**S105C**	**V66C**	**V69C**
WT	**S64C**	**G97C**	**S72C**	**I62C**	**V69C**
pMS119EH	**S64C**	**G97C**	**S72C**	**I62C**	**V69C**
L103C	**WT**	**G97C**	**S72C**	**I62C**	**L103C**

Figure 2. 8: The 48-wells set-up of the 96-well plate.

Each EmrE variant is shown as one amino acid abbreviation; the first letter indicating the conserved amino acid in the wild-type EmrE while the second letter is the new amino acid. The red indicates the 1/100 dilution of EmrE variant culture while the **black** is the **1/10** dilution of the same culture. The variants are placed in a randomized fashion and the organization changes from one assay to the other. The above illustrates only one example of one of the assay set-ups with those particular EmrE variants. pMS119EH is the empty vector that is transformed in *E. coli ΔacrB* while WT is the wild-type EmrE transformed in *E. coli ΔacrB*.

Once all the plates have been spotted, they were incubated for 16 hours at 37°C incubator to be read for the next day.

The MIC were recorded for each replicate of each sample as the concentration at which the colony stops growing. Then, the MIC of the negative control and the positive control (wild-type EmrE) can be compared to the MIC of the EmrE variants in order to identify the importance of certain residues in conferring resistance to the specific QCC used

in this study. In addition to the MIC, the colony morphology of each EmrE variant/controls were recorded through taking photographs of the plates and recording the concentrations of QCC at which the microlawns started to disperse, fade or shrink into smaller colonies (Figure 2.9).

In the following chapters of this thesis, the results of the plate-based QCC resistance assays will be organized/analyzed and presented in several ways in order to identify the patterns seen in the data.

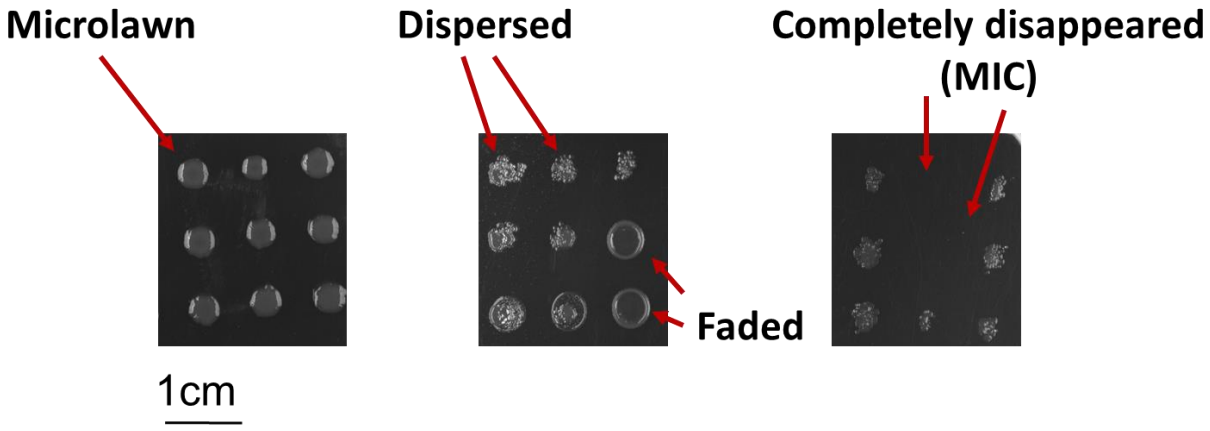


Figure 2. 9: Colony morphology changes upon introducing QCC to the media.

Microlawn represents a cellular culture of normally healthy EmrE variants without the presence of QCC. As the concentration of QCC increases some microlawns begin to disperse into smaller colonies or fade and become transparent. With further increase of QCC concentrations, the microlawn completely disappears and this is the point where the minimum inhibitory concentration (MIC) is recorded.

2.6 Hierarchical Agglomerative Clustering Method

The hierarchical agglomerative clustering plots were prepared using the scaled-MIC data for the resistance profiles of the 33 EmrE variants and the two controls. The scaled-MIC tables can be viewed in Appendix C, where the MIC values were scaled from 1 to 4 and the wild-type MIC was given the value of 3 (i.e. anything with lower MIC than the wild-type were given 1-2 and anything with higher MIC than the wild-type were given 4).

For all the hierarchical agglomerative clustering plots, we have used R Statistics software[1] (version 3.1.0) with R Studio[2] (version 0.98.507) Integrated Development Environment (IDE) interface. The clustered heat-maps were generated using the function heatmap.2(x)[3] associated with the library package 'gplots', and 'RColorBrewer'. Clustering was performed using the function dist(x,..), where an Euclidean distance method was applied, which is a method that measures the distance between two pointes as defined by the square root of the sum of the squares of the differences between the corresponding coordinates of the points. The hclust(x)[4] function was used to generate a heat-map for every one of the following linkage metrics: Ward, Average, Complete, Centroid, and Single. These different methods will be discussed in Section 5.1.1 and the heat-maps generated will be presented in Appendix C and Section 5.1.2.

Links for the software used:

[1] http://cran.r-project.org/bin/windows/base/

[2] https://www.rstudio.com/ide/download/

[3] http://hosho.ees.hokudai.ac.jp/~kubo/Rdoc/library/gplots/html/heatmap.2.html

[4] http://stat.ethz.ch/R-manual/R-patched/library/stats/html/hclust.html

Chapter 3: Altering Key Amino Acid Residues within EmrE

3.1 EmrE Conserved Amino Acid Residues

EmrE, as described earlier, belongs to the SMP subclass of the SMR family and serves as an excellent paradigm for all SMR members (Bay et al., 2008). Previous bioinformatics analysis that explored the evolutionary relatedness of SMR protein sequences shined a light on the degree of amino acid conservation and the selection pressures that were exerted upon those proteins (Bay and Turner, 2009).

The examination of protein sequence alignments of the two subclasses of the SMR family illustrated that SMP and SUG groups have unique conserved amino acid motifs that reflect their functional differences (Bay et al., 2008). The study also investigated the selective pressure that was exerted on the two different subclasses that led to the functional differences by determining the rate of synonymous to non-synonymous nucleotide substitution. On the other hand, the extent of amino acid conservation was also explored at each position within the loops and the transmembrane segments (TMS) by generating the overall amino acid consensus from the protein alignment of each SMR subclass member. The mean synonymous nucleotide substitution (Sd) values were obtained by taking the average of 20 SMP members' Sd values at each codon position in order to determine the selective pressure exerted on that particular residue of the protein. In addition, the percent consensus of each amino acid is calculated for each amino acid residue in order to indicate the frequency of a particular residue found at that particular position in the sequence alignment of the 685 SMR members. The Figure below describes a summary of both selective pressure analysis (Mean Sd values) and amino acid conservation (% consensus) for the SMP subclass.

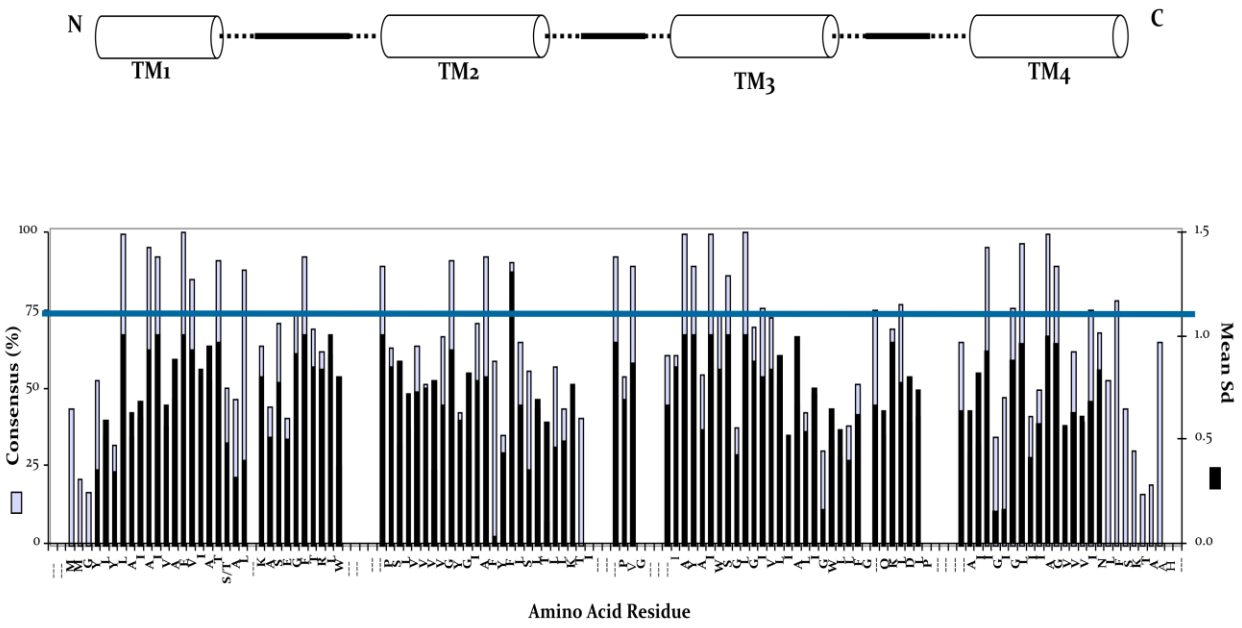


Figure 3. 1: Summary of SMP amino acid residue consensus and mean synonymous substitutions (Mean Sd) observed at each residue position within the overall sequence alignment of the 685 SMR members.

The amino acids with the highest frequency at each position within the alignment are indicated on the x-axis and dashes indicate position lacking amino acid alignment. The degree of conservation of the listed amino acid is based on its percent consensus (grey bars, left hand y-axis). Mean Sd values (black bars) on the right-hand y-axis represent the average of pair-wise comparisons of Sd values for the 20 SMP members. The blue line represents the 75% consensus cut-off that was chosen in this experiment and any residue with 75% consensus or higher will be altered by site-directed mutagenesis (Bay and Turner, 2009, obtained with permission).

When compared to the other subclasses of SMR, the SMP subclass demonstrates a unique amino acid consensus at all four TMS and loops. Moreover, the highly conserved amino acids tend to appear in a periodic fashion resembling the turns of an alpha-helix (Bay and Turner, 2009). Therefore, it is predicted that certain face(s) of TMS will have more conserved residues than other faces.

In order to explore the importance of some of these conserved residues to the functionality of EmrE in conferring resistance to QCC, conserved residues were altered using site-directed mutagenesis. Any residue with 75% of conservation (% consensus) or higher were targeted in the mutagenesis analysis. The 75% cut-off was chosen arbitrary in order to explore if these amino acids play any significant role in the folding and function of EmrE. Thus, if no significant trends are observed with residues of lower percent conservation in the 75-100% range, then the focus will be on the residues with higher % consensus (i.e. 90% consensus residues). However, if the majority of the residues with 75% or higher consensus played an important role in the function of EmrE, then future studies can further investigate the amino acids with lower percent consensus (i.e. 60% or 50% conservation).

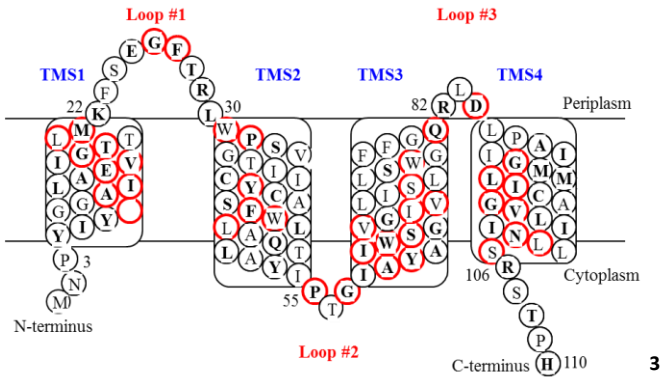
In the following section, the amino acids that were chosen for mutagenesis studies and the successful EmrE variants that were generated will be discussed.

3.2 EmrE variants generated

Table 3.1 below shows the amino acids that will be investigated and altered using site-directed mutagenesis to either cysteine or alanine. There were 6 residues out of the 39 conserved residues that were not mutated due to complications in the procedure of altering

these amino acids to cysteine. The QCC resistance profiles of the remaining 33 EmrE variants generated will be discussed in details in the following chapters with various ways of data presentation.

Table 3. 1: Amino acid residues targeted for mutagenesis.



[2] EmrE variants	% Consensus
1. L7	98
2. A10	94
3. I11	90
4. E14	98
5. V15	85
6. G17	70
7. T18	88
8. L20	87
9. M21	87
10. G26	75
11. F27	90
12. [1] W31	**60**
13. P32	88
14. Y40	94
15. F44	94
16. [1] W45	**70**
17. L47	90
18. P55	94
19. G57	90
20. A59	98
21. Y60	88
22. I62	98
23. W63	75
24. S64	87
25. V66	98
26. I68	75
27. V69	75
28. S72	75
29. [1] W76	**50**
30. Q81	75

[2] EmrE variants	% Consensus
31. D84	78
32. G90	92
33. L93	76
34. I94	90
35. G97	97
36. V98	88
37. N102	75
38. L103	75
39. S105	80

[1] The tryptophan (W) residues **(bolded)** don't obey the rule of ≥75% consensus.

[2] All black colored residues were successfully mutated while the red residues were not.

[3] The diagram above the table illustrates the location of the residues that will be explored.

Chapter 4: Resistance Profiles

4.1 Introduction

A transporter as small as EmrE with an ability to transport a broad range of lipophilic toxins sparks the question of how does such a transporter bind to and transport this vast array of ligands?

EmrE has only one membrane-embedded charged residue, Glu14, which is a highly conserved residue in more than 100 homologous proteins (Ninio *et al.,* 2001). The acidic side chain of Glu14 (the carboxyl group) has been postulated to be involved in the recognition of both substrates and the coupling ion, H^+. Also, previous studies have shown that the replacement of Glu14 with Cys impairs the ability of the EmrE protein to confer resistance against ethidium bromide because the protein is incapable of driving the proton motive force to transport the molecule (Schuldiner *et al*., 1997). Furthermore, some of the residues that are found on the same face of the transmembrane helix as Glu14 may also be involved in substrate recognition and possibly the stabilization of the negative charge in the membrane (Gutman *et al*., 2003).

Moreover, the unique pattern of conserved amino acid residues seen in EmrE transmembrane helices and the helical periodicity of conservation in these helices (Bay and Turner, 2009) suggest the importance of these conserved residues to the function of EmrE (Figure 3.1)**.** These residues may possibly be involved in protein folding, dimerization, or substrate binding/transport.

Some of the previously studied amino acid residues using SDM analysis in EmrE

overlap with the residues we have targeted in this study (Section 1.4.3). However, site-directed mutagenesis in this study focused on altering the conserved amino acids with ≥75% consensus (Bay et al., 2009). Also, the previously studied EmrE variants were tested against 3-4 substrates (ethidium, methyl viologen, TTP+, and/or acriflavine) **(**Mordoch et al., 1999; Muth and Schuldiner, 2000; Yerushalmi and Schuldiner, 2000; Gutman et al., 2003; Elbaz et al., 2005; Sharoni et al., 2005; Rolem et al., 2006**).** Nonetheless, I have used a larger number and more diverse groups of compounds to develop my conclusions and have used different amino acid replacements in some cases.

The loss of resistance phenotype when conserved residues in EmrE are mutated can be explained by many possible factors. First, a mutation of a residue that is critical for protein folding will lead to improper folding of the protein leading to proteolytic degradation of the protein. The specific side-chains of these residues may play a critical role in maintaining stability to the 3-dimentional structure of the protein through hydrophobic, covalent, hydrogen bonding, or van der Waals interactions. Also, some residues may participate in the multimerization of EmrE, which is critical to the functionality of the protein. EmrE has been previously determined to be more functionally favorable in an oligomeric state and hence interruption of the oligomerization process can cause full inactivation of the protein and loss of resistance (Yerushalmi et al., 1996). Moreover, loss of resistance can be as a result of improper protein insertion into the membrane, improper ligand binding or ligand movement. Also, some residues may participate in mediating the proton exchange and the conformational changes of the protein upon the binding of the ligand. Therefore, changes to these amino acids may influence any of the above processes that are critical for the transport mechanism of the ligand and hence

causing loss of resistance. However, if mutations affect the resistance toward QCC differently, then many of these factors can be ignored and we can attribute the changes in resistance to changes in binding and transport.

I have studied 19 quaternary cationic compounds (QCC) (refer to Figure 2.7) and compared QCC resistance profiles of 33 mutants of EmrE with single conserved residue replacement. This led to 1,330 different resistance profiles that were analyzed taking into account the replicates of each assay and the resistance profiles of the controls. In the following sections of this chapter, the results of the QCC resistance assays will be presented.

4.2 Minimum Inhibitory Concentrations of EmrE Variants

Thirty-three EmrE variants were tested against 19 different QCC. Of the 19 different QCC chosen, hexamethylenetetramine did not show difference in inhibition for any of the 33 variants or the two controls even at the highest concentration used in the assay (4mg/mL). Hexamethylenetetramine is known to have a cage-like structure and as such may make it difficult to penetrate cellular membrane or interact with key cellular components interfering with cell growth**.** Previous *in vitro* genotoxicity studies with *S. typhimurium* have shown that with hexamethylenetetramine concentrations increased up to 5mg/mL, the substrate still was not toxics to the bacteria. Consequently, they attributed the lack of the drug's effect on the bacteria to inadequate penetration of the bacterial membrane (Crebelli et al., 1984). Also, there was no difference between the MIC of the two controls for stearyltrimethylammonium suggesting that the overexpression of wild-type EmrE protein by the transformed plasmid did not affect the resistance of *E. coli ΔacrB* to that

drug. Thus, the resistance profiles with wild-type EmrE and without (negative control) have similar sensitivity to stearyltrimethylammonium, which made any deviations in the resistance of the EmrE variants from the wild-type EmrE insignificant. As a result, hexamethylenetetramine and stearyltrimethylammonium were excluded from the future QCC resistance profiles analysis.

Based on the premise that conserved amino acids at ≥75% consensus would influence the QCC resistance activity seen in the wild-type EmrE, 33 conserved amino acids were mutated to either alanine or cysteine residues. The MIC of each EmrE variant seemed to cover a large range of numbers suggesting that no single residue alteration can lead to 'all or nothing response' except for I11C, E14D, V15C, and P32C (Table 4.1). The other 29 residues were either important for conferring resistance to one particular QCC group or more than one group of QCC. Therefore, it can be speculated that these residues are involved in substrate binding/transport. This is because if these residues were involved in protein folding or stabilization then their mutation would result in the loss of the protein, EmrE, thereby showing a reduced host-tolerance to all the compounds used.

As presented in Table 4.1, I11C, V15C, and P32C variants did not reduce host resistance to the majority of QCC and sometimes had MIC values that were higher than the wild-type control. This indicates that these residues are not likely involved in substrate recognition, protein dimerization, or protein folding/insertion. Thus, they don't seem to play an important role in the activity of EmrE despite how highly conserved they are. Moreover, the replacements of these residues resulted in enhanced host resistance to some of the compounds (i.e. Acriflavine and Proflavine). What is surprising is that the extreme change of the biochemistry of the side chain from Pro to Cys still resulted in a functional

protein. This is can be explained by the fact that this residue is found in loop-1 of the protein and the residues in the loops have shown very minimal involvement in the activity of the protein when mutated based on Simon Schuldiner's research group studies (University of Jerusalem). Also, because amino acids in loops are not constrained by space and environment as the amino acids found in TMS, and do not have an effect on the arrangement of secondary structures in the core region of the protein, substitutions of residues in the loops may not show an affect on the protein activity.

Moreover, Ile11 and Val15 substitutions with Cys did not affect the activity of the protein possibly because of the simplicity of the side-chains of these amino acids. With the ethyl or methyl being changed to a thiol group it seems to not affect the ability of the protein to interact with the majority of the substrates and in some cases it was a favored interaction with certain substrates (acriflavine and proflavine). Therefore, despite the conservation of these two residues, they don't seem to play an important role in the multidrug resistance activity of EmrE.

Glu14 residue, on the other hand, is the amino acid that when mutated, reduces the MIC of the cells expressing E14D to the majority of the QCC (except for rhodamine and myristalkonium) when compared to the wild-type. This is attributed to the fact that Glu14 is considered to be key in the deprotonation activity upon the binding of the substrate to the residue. It is also the only charged residue embedded into the transmembrane helices making it critical for interacting with cationic substrates. It has also been apparent that the substitution of Glu-14 to Asp decreases the pKa of the residue thereby making it more acidic (Muth and Schuldiner, 2000). The more acidic the residue at this position, the better it is at deprotonation and the worse it is at binding a cationic substrate. Thus, Asp

substitution of Glu14 affects the binding pocket activity significantly resulting in lower resistance of the variant to the majority of the cationic substrates except for rhodamine and myristalkonium.

The MIC values for each QCC were scaled to values from 0-1 (See Appendix A) and each value was designated to a color to generate a heat-map for each compound that can be used for further analysis on the involvement of each residue in the poly-substrate specificity of EmrE. The following section will analyze the trends seen in the scaled data and draw some conclusions on the patterns observed.

Table 4. 1: Minimum inhibitory concentration (MIC) of EmrE variants in µg/mL.

[1, 2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
1. L7A	32	128	64	32	4	32	**32**	**120**	**16**	16	250	8	16	500	8	8	64	16
2. A10C	16	64	32	32	2	32	**32**	**60**	**32**	8	64	16	16	250	8	8	32	4
3. I11C	128	128	32	32	32	32	**32**	**120**	**32**	500	500	16	16	500	8	8	128	8
4. E14D	8	32	16	32	4	16	**16**	**60**	**16**	8	64	8	16	250	4	4	64	4
5. V15C	128	>128	32	8	32	32	**32**	**120**	**32**	500	500	16	16	500	8	8	128	16
6. G17A	128	64	64	32	16	32	**32**	**120**	**16**	8	250	16	16	250	16	8	128	16
7. T18C	64	64	4	64	4	16	**32**	**120**	**32**	500	125	16	16	250	16	8	64	16
8. M21C	64	128	4	8	4	16	**32**	**120**	**32**	32	500	16	16	500	8	8	128	8
9. G26C	64	128	8	8	4	16	**32**	**120**	**32**	16	500	16	16	500	8	4	128	8
10. F27C	64	128	4	8	4	32	**16**	**120**	**32**	64	500	16	16	500	8	4	128	8
11. W31A	64	64	64	32	16	64	**16**	**120**	**16**	16	500	16	16	250	8	8	64	8
12. P32C	128	>128	32	32	32	32	**32**	**120**	**32**	500	500	16	16	500	8	8	128	16
13. Y40A	128	64	16	32	8	64	**32**	**120**	**8**	8	500	16	32	250	8	8	128	16
14. F44C	64	128	4	8	2	32	**32**	**60**	**32**	16	250	8	16	500	8	4	128	4
15. W45A	64	128	32	32	16	64	**32**	**120**	**16**	8	500	8	16	250	8	8	64	8
16. L47C	16	64	8	16	1	16	**32**	**120**	**32**	16	32	8	16	500	8	4	128	4
17. P55C	64	128	8	4	2	32	**16**	**60**	**16**	500	500	16	16	500	4	4	128	4

[1,2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
18. G57C	64	128	4	16	2	32	**32**	**120**	**32**	500	125	16	16	500	8	8	64	8
19. A59C	32	128	16	32	4	64	**32**	**60**	**16**	16	500	8	16	500	8	8	64	8
20. Y60A	32	64	32	32	8	32	**32**	**120**	**16**	16	250	8	16	250	16	8	64	8
21. I62C	64	128	64	32	4	128	**32**	**120**	**8**	16	250	16	16	500	8	8	64	8
22. W63A	16	32	32	16	2	16	**16**	**>120**	**8**	500	64	16	16	500	16	16	128	16
23. S64C	64	128	32	32	8	64	**16**	**60**	**16**	16	128	8	16	500	8	8	64	8
24. V66C	128	128	32	32	16	64	**32**	**120**	**16**	16	250	16	16	250	8	8	128	8
25. V69C	128	128	32	32	16	64	**16**	**60**	**32**	8	500	8	16	500	8	8	128	8
26. S72C	64	128	32	32	16	64	**32**	**120**	**32**	16	500	8	16	250	8	8	64	16
27. W76A	64	64	16	16	8	32	**16**	**60**	**8**	16	500	4	16	250	8	8	128	16
28. G90C	64	64	4	16	2	32	**32**	**120**	**32**	64	250	8	16	500	8	8	128	8
29. L93C	256	128	32	32	16	64	**32**	**120**	**16**	16	500	8	16	250	8	8	128	8
30. I94C	32	64	32	16	2	16	**32**	**>120**	**8**	500	250	8	16	500	8	8	64	16
31. G97C	16	64	32	32	2	NA	**32**	**60**	**32**	8	250	8	8	500	8	8	64	4
32. L103C	64	128	32	32	8	64	**32**	**120**	**16**	16	500	8	32	500	8	8	64	8
33. S105C	64	128	32	32	4	128	**32**	**60**	**16**	16	500	8	16	500	8	8	64	4

[1,2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
[3] pMS119EH	16	16	8	4	2	16	**8**	**60**	**16**	16	250	8	8	500	4	4	64	4
[3] pEmrE(WT)	64	128	32	32	8	32	**32**	**120**	**32**	500	500	16	16	500	8	8	128	16

[1] The columns showing the 33 different plasmids with *emrE* single codon mutants transformed in *E. coli* Δ*acrB,* called EmrE variants.

[2] The culture of each EmrE variant had an optical density value that was standardized to 1.5 and then diluted to 1/10. Comparable MIC values were observed with 1/100 culture (see Appendix B).

[3] pMS119EH empty vector transformed in *E. coli* Δ*acrB,* representing the negative control while the positive control is *E. coli* Δ*acrB* expressing wild-type EmrE (pEMR11).

[4] The QCC abbreviations are: ACR= Acriflavine, PR= Proflavine, CV= Crystal violet, RH= Rhodamine, PY= Pyronin, EB=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.C= Cetalkonium chloride, STAC= Stearyltrimethylammonium chloride, and CET= Cetrimide. Also, the grouping of the QCC followed this coloring scheme: Poly-aromatic QCC, Sphere-forming QCC, Poly-charged QCC, Acyl-chained QCC.

4.3 QCC Resistance Profiles of EmrE Variants

I used the native form of the EmrE protein for my studies, which contained 3 cysteine residues (Cys39, Cys41, & Cys95). The introduction of another cysteine residue as a result of mutating a conserved amino acid to a cysteine did not result in intra-protein cysteine cross-linking due to the lack of a reducing environment *in vivo*. Using the native EmrE, I have generated 27 single Cys or Ala replacements throughout the four transmembrane segments (TMS) of the protein and 6 replacements in the 3 loops of the protein. Eight residues were altered in TMS1, four residues in TMS2, nine residues in TMS3, and six residues in TMS4. Out of the 33 residues, 30 residues have 75% conservation or higher within the overall sequence alignment of the 685 SMR members. The other 3 residues are the tryptophans that are believed to play an important role in substrate recognition but have 50-70% conservation (Table 3.1). Therefore, they are expected to play an important role in the multidrug resistance activity of EmrE. The majority of the replacements reduced the protein activity as judged from the ability of the EmrE variant to confer resistance against the toxic QCC tested.

The activity of each EmrE variant has been tested *in vivo*, where the resistance displayed by each variant was assessed by studying the ability of the *E. coli* cells expressing them to grow on solid media containing one QCC at various concentrations. The cells that were carrying the empty vector without *emrE* insert were used as a background negative control while the cells expressing the wild-type EmrE represented the comparator positive control. The cells expressing the EmrE variants and the two controls grew to a similar degree in the control plates containing only ampicillin and without any QCC. The

resistance profiles of the different variants to the different substrates are being used to obtain qualitative idea of the activity of these mutants when compared to the resistance profile of the wild-type. Therefore, any amino acid alterations that result in *E. coli* cells with lower MIC than the wild-type will be referred to as an EmrE variant with reduced host resistance. Despite the possibility that the levels of EmrE expression may differ among the various mutants, several conclusions on the involvement of these residues in the activity of the protein can be made. It is recognized that a given mutation could affect expression, membrane insertion, folding, or protomer assembly. If this were the case for any of these residues, we would expect the mutation to affect all substrates equally. Our data below shows that this was not the case (Table 4.2). Thus, we are able to interpret the results of each mutation and its effect on substrate recognition.

In order to more easily interpret the large data set, the values were weighted to generate a 'heat-map' (Table 4.2). By evaluating the trends within the heat-map for only the poly-aromatic QCC class, it is apparent that there is diversity in host resistance among the different variants illustrating different degrees of residue interactions with that particular QCC class. For example, when the 33 amino acids were substituted, they were found to be divided in groups in all of the colors in the MIC-scaled spectrum (0-1) for acriflavine, crystal violet, rhodamine, ethidium and pyronin demonstrating a range of resistance to those compounds. Also, several EmrE variants with extremely low MIC (red-orange) are found under these compounds showing the detrimental affect of these compounds on the mutated EmrE.

Table 4. 2: Heat-map summarizing QCC minimum inhibitory concentration results for *E. coli ΔacrB* strains expressing different plasmids with *emrE* single amino acid variants. [1, 2]

[3] ACR	PR	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	CC	CB	Ct. C	CET
[4] WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
[4] -VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE	-VE
E14D	E14D	T18C	P55C	A10C	E14D	E14D	A10C	Y40A	L7A	L47C	W76C	G97C	E14D	E14D	A10C	A10C
A10C	W63A	M21C	V15C	F44C	T18C	F27C	E14D	I62C	A10C	A10C	L7A	L7A	P55C	G26C	L7A	E14D
L47C	A10C	F27C	M21C	L47C	M21C	W31A	F44C	W63A	E14D	E14D	E14D	A10C	L7A	F27C	E14D	F44C
W63A	G17A	F44C	G26C	P55C	G26C	P55C	P55C	W76A	G17A	W63A	F44C	I11C	A10C	F44C	T18C	L47C
G97C	T18C	G57C	F27C	G57C	L47C	W63A	A59C	I94C	G26C	T18C	W45A	E14D	I11C	L47C	W31A	P55C
L7A	W31A	G90C	F44C	W63A	W63A	S64C	S64C	L7A	W31A	G57C	L47C	V15C	V15C	P55C	W45A	G97C
A59C	Y40A	G26C	L47C	G90C	I94C	W76A	V69C	E14D	F44C	S64C	A59C	G17A	M21C	L7A	G57C	S105C
Y60A	L47C	L47C	G57C	I94C	L7A	V69C	W76C	G17A	L47C	L7A	Y60A	T18C	G26C	A10C	A59C	I11C
I94C	Y60A	P55C	W63C	G97C	A10C	L7A	G97C	W31A	A59C	G17A	S64C	M21C	F27C	I11C	Y60A	M21C
T18C	W76A	E14D	W76A	L7A	I11C	A10C	S105C	W45A	Y60A	F44C	V69C	G26C	W31A	V15C	I62C	G26C
M21C	G90C	Y40A	G90C	E14D	V15C	I11C	L7A	P55C	I62C	Y60A	S72C	F27C	P32C	G17A	S64C	F27C
G26C	I94C	A59C	I94C	T18C	G17A	V15C	I11C	A59C	S64C	I62C	G90C	W31A	Y40A	T18C	S72C	W31A
F27C	G97C	W76A	L7A	M21C	F27C	G17A	V15C	Y60A	V66C	V66C	L93C	P32C	F44C	M21C	I94C	W45A
W31A	L7A	A10C	A10C	G26C	P32C	T18C	G17A	S64C	V69C	L93C	I94C	F44C	W45A	W31A	G97C	G57C
F44C	I11C	V15C	I11C	F27C	F44C	M21C	T18C	V66C	W45A	I94C	G97C	W45A	L47C	P32C	L103C	A59C
W45A	V15C	P32C	E14D	A59C	P55C	G26C	M21C	L93C	Y40A	G97C	L103C	L47C	G57C	Y40A	S105C	Y60A
P55C	M21C	W45A	G17A	I62C	G57C	Y40A	G26C	L103C	S72C	I11C	S105C	P55C	A59C	W45A	I11C	I62C
G57C	G26C	Y60A	T18C	S105C	Y60A	P32C	F27C	S105C	W76A	V15C	A10C	G57C	I62C	G57C	V15C	S64C
I62C	F27C	W63A	W31A	Y40A	W76A	F44C	W31A	A10C	L93C	M21C	I11C	A59C	S64C	A59C	G17A	V66C
S64C	P32C	S64C	P32C	Y60A	G90C	W45A	P32C	I11C	G97C	G26C	V15C	Y60A	V66C	Y60A	M21C	V69C
S72C	F44C	V66C	Y40A	S64C	W31A	L47C	Y40A	V15C	L103C	F27C	G17C	I62C	V69C	I62C	G26C	G90
W76A	W45C	V69C	W45A	W76A	Y40A	G57C	W45A	T18C	S105C	W31A	T18C	W63A	S72C	W63A	F27C	L93C
G90C	P55C	S72C	A59C	L103C	W45A	A59C	L47C	M21C	M21C	P32C	M21C	S64C	W76C	S64C	P32C	L103C
L103C	G57C	L93C	Y60A	G17A	A59C	Y60A	G57C	G26C	F27C	Y40A	G26C	V66C	G90C	V66C	Y40A	L7A
S105C	A59C	I94C	I62C	W31A	I62C	I62C	Y60A	F27C	G90C	W45A	F27C	V69C	L93C	V69C	F44C	V15C
I11C	I62C	G97C	S64C	W45A	S64C	V66C	I62C	P32C	I11C	P55C	W31A	S72C	I94C	S72C	L47C	G17A
V15C	S64C	L103C	V66C	V66C	V66C	S72C	W63C	F44C	V15C	A59C	P32C	W76C	G97C	W76C	P55C	T18C
G17A	V66C	S105C	V69C	V69C	V69C	G90C	V66C	L47C	T18C	V69C	Y40A	G90C	L103C	G90C	W63C	P32C
P32C	V69C	I11C	S72C	S72C	S72C	L93C	S72C	G57C	P32C	S72C	P55C	L93C	S105C	L93C	V66C	Y40A
Y40A	S72C	L7A	L93C	L93C	L93C	I94C	G90C	V69C	P55C	W76A	G57C	I94C	G17A	I94C	V69C	W63A
V66C	L93C	G17A	G97C	I11C	L103C	G97C	L93C	S72C	G57C	L93C	I62C	S105C	T18C	G97C	W76A	S72C
V69C	L103C	W31A	L103C	V15C	S105C	L103C	I94C	G90C	W63A	L103C	W63A	Y40A	Y60A	L103C	G90C	W76A
L93C	S105C	I62C	S105C	P32C	N/A	S105C	L103C	G97C	I94C	S105C	V66C	L103C	W63A	S105C	L93C	I94C

[1] MIC values were scaled (0-1) to reflect the lowest (red) and highest MIC (green) values obtained for *emrE* mutant transformed in *E. coli ΔacrB* strain and grown on plates containing one of the 17 QCC tested (see Appendix A).

[2] The color scheme represent the following scale, 0 0.1 0.2, 0.3, 0.4, 0.5, and 1.

[3] The QCC abbreviations are: ACR= Acriflavine, PR= Proflavine, CV= Crystal violet, RH= Rhodamine, PY= Pyronin, EB=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.C= Cetalkonium chloride, and CET= Cetrimide. Also, the grouping of the QCC followed this coloring scheme: Poly-aromatic QCC, Sphere-forming QCC, Poly-charged QCC, Acyl-chained QCC.

[4] Controls are *E. coli ΔacrB* with plasmid containing the wild-type EmrE (WT) and *E. coli ΔacrB* with the empty plasmid pMS119EH (-VE).

On the other hand, for a compound like proflavine, EmrE variants were divided into 3 groups; 2 EmrE variants with the low MIC indicating reduction in QCC resistance, second group with intermediate MIC indicating QCC resistance levels that are close to wild-type, and the third group with the highest MIC indicating wild-type level of QCC resistance (Table 4.2). Although proflavine is a derivative of acriflavine, it showed less of an extreme effect on EmrE variants. There were fewer EmrE variants that were sensitive to extremely low concentrations of proflavine, which was seen as no variants in the 0-0.2 range of standardized MIC. Thus, the data suggests that acriflavine may have a more potent effect on the cells than proflavine. Previous research has shown acriflavine binds more efficiently than proflavine to cellular constituents making it more toxic than proflavine (Silver et al., 1968). Acriflavine may cause denaturation of proteins and cell membrane damage (McDonnell and Russell, 1999). Therefore, treatment with acriflavine produced stronger inhibition activity of *E. coli* cells expressing EmrE variants which is seen in larger numbers of amino acids in the red-orange zone (very low MIC). With proflavine, the inhibition of resistance activity (MIC-scaled value of 0, red box) is not observed with any of the EmrE variants and the next MIC-scaled value of 0.1 is only observed with 2 EmrE variants (E14D and W63A). These residues have been identified to participate in the binding pocket, where Glu14 is the only charged transmembrane residue that is involved in the substrate efflux and proton influx. Thus, it is expected that alterations to these residues will reduce the resistance significantly in response to the majority of the substrates used.

This is also confirmed when looking at the resistance profile of W63A and E14D in which they cluster in the low MIC ranges (red-orange) for the majority of the poly-aromatic QCC, except E14D showed similar resistance levels to that of wild-type under rhodamine

treatment and W63A showed similar resistance levels to that of the wild-type under crystal violet treatment. These results may suggest that the presence of an aspartic acid at position 14 may not interfere with the interaction with a bulkier multi-aromatic structure like rhodamine. Also, the presence of an alanine in position 63 seems not to affect the resistance to crystal violet and neither any of the other tryptophan substitutions affects the resistance to that compound, which possibly indicates a minimal role of tryptophan interaction with crystal violet.

A considerable number of EmrE variants (M21C, G26C, F27C, A10C, T18C, F44C, L47C, G57C, G90C, I94C, A59C, W63A, P55C, G97C, E14D, and W76A) showed a substantial loss of resistance activity to the majority of the poly-aromatic QCC. The previous exploration of A10C involvement in substrate binding activity, its critical role for the proton/substrate coupling activity, its conservation, and its close proximity to Glu-14 can explain the importance of this residue for poly-aromatic QCC resistance (Gutman et al., 2003). The very short side chain of alanine at that particular location seems to be very critical for the recognition of compounds with multiple aromatic rings. However, the complexity of crystal violet and rhodamine seems not to be very challenging substrates to A10C variant. Thus, it may suggest that an alanine is important for conferring resistance to poly-aromatic compounds with less complex structures and other residues might be important for the more complex poly-aromatic compounds. This may suggest that Ala10 is an amino acid that is involved in the binding pocket and confers a level of substrate preferences even within the poly-aromatic QCC group.

The substitutions of Met21, Gly26, and Phe27 with Cys are found to reduce host resistance to only poly-aromatic QCC. Met21 is found in the first transmembrane helix and

very close to the hydrophilic loop-1. Thus, the results suggest that the shortening of the amino acid side chain from Met to Cys affected this residue's interaction with compounds containing multiple aromatic rings. In addition, the other two residues (Phe27 & Gly26) are found on loop-1, which might suggest that the part of TMS1 that is close to loop-1 (Met21) and loop-1 residues that are close to TMS1 are important for interacting with poly-aromatic QCC. Hence, the resistance to these compounds is inhibited upon the alteration of such residues.

Residues Phe44, Leu47, Gly90, Ile94, Ala59 and Gly97 do not have established roles in EmrE in the previous literature. However, it is worth noting that F44C, L47C, G90C, and I94C have reduced host resistance to 10 compounds out of the 19 QCC tested. The residues, Ala59 and Gly97, seem to be involved in conferring resistance to more than one group of QCC but with different sensitivities to the compounds (having varied MIC). This may suggest that these residues are involved in interacting with the different QCC to different degrees and hence they are important for conferring resistance to particular compounds. The 'broad' specificity of these residues to the different compounds may suggest the involvement of these amino acids in the binding pocket of EmrE.

Previous studies have suggested that Thr18, Pro55, Trp76, and Gly57 are residues that may play an important role in EmrE folding and stability. From the data presented here, these residues tend to participate in conferring resistance to about 6-8 compounds from the 19 compounds studied. If they possibly participate in protein folding, alterations to these residues should have resulted in deleterious effects on the growth of *E. coli* expressing them under the majority of the QCC if not all of them. Therefore, these residues may possibly participate in substrate selectivity.

Considering the sphere-forming QCC, the MIC values for the EmrE variants fall into two main categories, those with high MIC (green) and those with very low MIC (red). Hence, loss in the diversity of host resistance is observed in this group. This may suggest that the conserved residues tested are involved in 'all or nothing' response to this particular group of compounds. The specific amino acid substitutions of the conserved residues resulted in either the inhibition of resistance to sphere-forming compounds or resistance that is similar to the wild-type.

The residues that seem to be important for conferring resistance to sphere-forming compounds are Trp31, Ser105, Ser64, Val69, Ala59, Trp63, Glu14, Pro55, and Trp76 (variants displaying lower MIC than the wild-type). As described previously, the Glu14 is an essential residue for substrate and proton binding/transport. Hence, the alteration of Glu14 to Asp reduces host resistance to this group of QCC as well. Interestingly, 3 out of the 4 altered tryptophan residues (Trp63, 76, and 31) seem to be important for conferring resistance to this group of QCC. The tryptophan residue at position 63 has previously been identified as a critical residue involved in interacting directly with the substrate (Elbaz et al., 2005). Moreover, the alteration of more than one tryptophan residue in EmrE resulted in lower tetraphenylphosphonium binding affinity and loss of transport activity, which suggests the importance of tryptophan residues for a functional EmrE protein (Elbaz et al., 2005). Furthermore, tryptophans have been found to maintain interactions between a protein and its substrate through hydrophobic stacking interactions (Vazques-Ibar et al., 2003). Therefore, the aromatic residues at these positions may have hydrophobic stacking interactions with aromatic rings found sphere-forming QCC and poly-aromatic QCC, which explains the importance of these residues for conferring resistance against these

compounds. The rest of the residues (Ser105, Ser64, Val69, Ala59, and Pro55) have not been identified with precise roles in EmrE in the previous literature. However, the data here suggests that they are important for conferring resistance to sphere-forming compounds and hence involved in EmrE substrate selectivity.

Under poly-charged QCC group, which is the group with two compounds that contain two cations each, the mutations of conserved residues resulted in extreme reduction in host resistance indicated by the remarkable decrease in MIC when compared to wild-type. More than 50% of the conserved residues studied (21 residues) have been found to participate in conferring resistance against poly-charged QCC (Leu7, Gly17, Tyr40, Ile62, Val66, Ser72, Leu93, Leu103, Ala10, Thr18, Phe44, Leu47, Gly57, Gly90, Ile94, Ser64, Val69, Ala59, Trp63, Glu14, Gly97, and Tyr60). This can be explained by the extra charge on the compound, which may not only require Glu14 to interact with the substrate but a network of residues are needed to stabilize the charge on the compound; hence mediating the transport of the compound. In recent work by Morrison and Henzler-Wildman (2014), they have discussed the importance of substrate properties on influencing EmrE transport activity. Substrate properties such as charge and geometry have been found to affect the energy needed for the transition state of the protein from in-ward open to out-ward open (See intro Figure 1.4). Therefore, substrates like methyl viologen and dequalinium chloride with an extra charge exert a burden on the protein transition state for these EmrE variants suggesting that some of the 21 residues may play an important role in the transition state of EmrE. Out of these 21 residues, there are 8 residues that are specifically involved in conferring resistance for only poly-charged QCC (Leu7, Gly17, Tyr40, Ile62, Val66, Ser72, Leu93, and Leu103). These residues can be involved in substrate recognition due to their

specific preference to poly-charged substrates and thus mediate the transition state of EmrE upon the binding of these residues. Given that the above residues are mostly hydrophobic, they might be involved in stabilizing the second ring in a substrate like methyl viologen. However, the Ser72 could be the residue involved to help with the extra charge, particularly if the charge is delocalized through the ring. Given EmrE is the only MDR in *E. coli* that can transport poly-charged QCC, our data begins to extract some unique features.

When examining the resistance profiles for acyl-chained QCC, it is apparent that fewer conserved residues are important for conferring resistance to that particular group of substrates. The acyl-chained compounds diverge remarkably in terms of structure from the remaining QCC groups, mainly due to the lengthy tails of carbons associated with their structure. These substrates have properties similar to surfactants/detergents. The results here demonstrate the residues that are important for conferring resistance to the majority of the compounds under acyl-chained QCC group are Trp45, Tyr60, Glu14, and Gly97. The majority of these residues also play a role in conferring resistance to other groups of QCC except for Trp45.

The mutation of Trp45 affected host resistance specifically to 50% of the acyl-chained compounds suggesting that this particular residue is important for the recognition of substrates with long acyl chains. Previous *in vivo* and *in vitro* studies of Trp45 suggested that this residue does not have a known function and possibly not involved in substrate binding (Elbaz et al., 2005 & Sharoni et al., 2005). However, most of these studies were done using acriflavine, methyl viologen, ethidium and TPP+ and did not explore acyl-chained compounds. Thus, from the qualitative results presented in this study, I can

speculate that Trp45 is involved in substrate recognition and possibly binding of acyl-chained compounds.

Although the remaining 3 residues (Y60, G97 and E14) are also involved in conferring resistance to other groups of QCC, an important thing to note is that the four residues (W45 included) are distributed in all four transmembrane helices of the protein. This suggests that at least one amino acid from each TMS is required to mediate the binding/transport of acyl-chained substrates. Also, Glu14 and Gly97 residues tend to be required for conferring resistance to the majority of the compounds used in this study; hence there is no surprise in the involvement of these two residues in conferring resistance to acyl-chained QCC as well.

In this study, we have found that Tyr60's fine-tuning activity is critical for conferring resistance to acyl-chained QCC and poly-charged QCC. Tyr60 does not participate in substrate interaction with compounds like the sphere-forming and poly-aromatic. Previous studies have shown that the aromatic ring with the hydroxyl group on Tyr residue at position 60 is critical for EmrE activity because the replacement of Tyr with Ser or Thr does not rescue the resistance phenotype to acriflavine, methyl viologen, and ethidium (Rolem et al., 2006). A possible explanation is the pKa of the phenolic hydroxyl group is ~10.5 whereas for Ser/Thr is 13. This may be a key difference in the functionality but also the aromatic ring may interact with the specific features of the acyl-chained or poly-charged substrate providing fine-tuning to the resistance activity of EmrE.

In summary, the generation of single residue variants of EmrE has identified the preferences of key conserved residues to specific QCC groups. Also, the heat-map helped

in identifying the residues involved in explaining the poly-substrate specific nature of EmrE by providing a spectrum from most resistant variant (high MIC-green) to least resistant variant (low MIC-red). This spectrum helped to clarify different degrees of residues involvement in interacting with specific substrates. Thus, speculating that EmrE binding pocket may involve a variety of amino acids, where each has its own responsibility of recognizing and interacting with specific substrates with specific characteristics (i.e. structural features). Also, these residues bind their substrates to different degrees explaining why some residues result in a more detrimental affect on resistance (i.e. residues in red boxes) while other may have a less of significant affect on conferring resistance to particular compounds (i.e. residues in yellow or light green boxes). This can be explained by the fact that certain residues are of a particular importance to that specific substrates while the other residues with less of an involvement in host resistance may be part of residues network that are as a group critical for substrate recognition.

In the following chapters, the conclusions that can be possibly drawn from the above analysis will be summarized diagrammatically using multivariate analysis and helical-wheel mapping.

Chapter 5: Advanced Data Analysis

The study of the resistance profiles of the 33 EmrE variants, in addition to the 2 controls, that were tested against 19 QCC produced 665 MIC values. The data was analyzed first by scaling the MIC values to values from 0-1 or from 1-4 in order to produce the Table 4.2 and the Figure 5.1 in this chapter. However, the data needed to be mined for additional relevant trends. For example, multivariate analysis of the scaled MIC values using R statistics software was used to generate heat-maps, which cluster EmrE variants according to their QCC resistance and clusters QCC according to their toxicity. Moreover, the different conserved amino acid residues that were altered in this study were mapped on helical wheels to identify particular faces of helices that may be involved in the binding site.

5.1.1 Introduction to Multivariate Analysis

Cluster analysis is a multivariate analysis that attempts to form groups or clusters from complex data sets, which are similar to each other, but differ among the clusters. The R statistics software offers an amazing variety of functions for cluster analysis. We have used the hierarchical agglomerative clustering (hclust) method, which is described as a "bottom up" approach to which each observation starts as its own cluster, the two most similar observations are fused to form the first cluster, and more similar observations are merged as one moves up the hierarchy (Zepeda-Mendoza and Resendis-Antonio, 2013) Also, the hclust can calculate a cluster analysis from either a similarity or dissimilarity matrix, but we have used a similarity matrix approach. The similarity is measured from the ratio of the number of characteristics shared by two observations compared to their total

characteristics. So observations that have everything in common will be identical and have a similarity value of 1.0, while objects with nothing in common will have a similarity value of 0.0.

The agglomerative cluster has several algorithms that differ in the calculation of similarity between observations. The data presented in Section 5.1.2 explores one of the clustering algorithms, which is the ‘complete’ linkage method. However, Appendix C shows the different heat-maps obtained by the other algorithms that are commonly used in R software for comparison.

In the ‘complete’ linkage, the similarity of an observation to a cluster is calculated as the minimum similarity of the observation to any member of the cluster. Thus, the distance between two clusters is the maximum distance between an observation in one cluster and an observation in the other cluster. The other linkage methods include ‘Ward’, ‘single’, ‘centroid’, and ‘average’. In the ‘single’ linkage, the similarity of an observation to a cluster is determined by the maximum similarity of the observation to any of the observations in the cluster. The name "single linkage" arises because the observation only needs to be similar to a single member of the cluster to join. In the ‘average’ linkage, the similarity of an observation to a cluster is defined by the mean similarity of the plot to all the members of the cluster. In contrast to single linkage, an observation needs to be relatively similar to all the members of the cluster to join, rather than just one. In the ‘centroid’ linkage, the distance between two clusters is defined as the (squared) Euclidean distance between their centroids or means. The centroid method is more robust to outliers than the other hierarchical linkage methods but may not perform as well as Ward, complete, or average linkage (Milligan and Sokol 1980). In the Ward linkage method, the distance

between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables. At each step the pair of clusters with the minimum between-cluster distances are fused.

5.1.2 Trends and Observations Collected From Multivariate Analysis

Using ‘complete’ linkage clustering, a heat-map illustrating the grouping of amino acid residues with their resistance profiles to QCC is presented in Figure 5.1. Complete-linkage clusters each observation as its own cluster at the beginning. In the following steps, the distance between clusters equals the distance between those two observations (one in each cluster) that are farthest away from each other. The shortest of these links that remains at any step causes the fusion of the two clusters.

A result of the ‘complete’ clustering gives a categorization of the QCC based on the behavior of EmrE variants towards these compounds as measured from their MIC-scaled values (Section 5.1.2.1). Also, EmrE variants are clustered according to the intensity of their sensitivity to the toxicity of the QCC (Section 5.1.2.2).

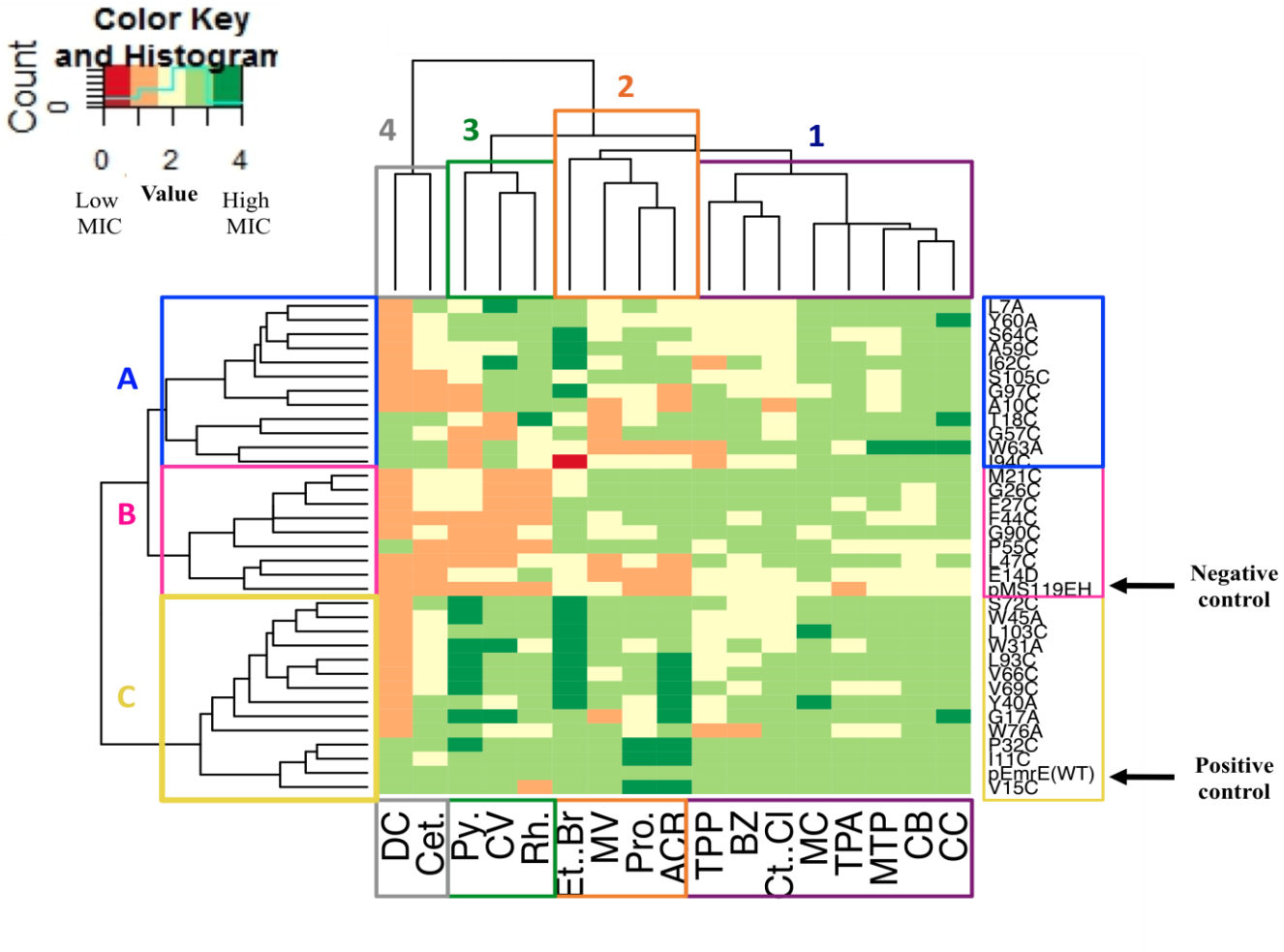


Figure 5. 1: The 'complete' linkage clustering of EmrE variants' resistance profiles to 17 QCC.

The heat map was generated using hierarchical agglomerative clustering plots, which used R Statistics software (See Sections 2.6 and 5.1.1 for details). The figure above presents the 'complete' linkage clustering method, where the similarity of an observation to a cluster is calculated as the minimum similarity of the observation to any member of the cluster. The bottom of the figure displays the different 17 QCC used in this study (ACR= Acriflavine, Pro= Proflavine, CV= Crystal violet, Rh.= Rhodamine, Py.= Pyronin, Et.Br=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.Cl= Cetalkonium chloride, and Cet.= Cetrimide). The right side of the diagram lists all the EmrE variants in addition to the two controls (*E. coli acrB* with wild-type EmrE protein and *E. coli acrB* with empty plasmid pMS119EH). The colored legend represents the values of scaled-MIC found in Appendix C. The rectangular grouping of the EmrE variants and the compounds is based on the dendrogram clustering.

5.1.2.1 ‘Complete’ Clustering of QCC

The clustering of QCC in the ‘complete’ linkage analysis was based on their toxicity to the EmrE variants and the controls of this study. However, identifying the physical and structural properties of these compounds may help us in pinpointing the relationship between these properties and the resistance phenotype that are observed with the variants. Markedly, the grouping of the compounds under the ‘complete’ linkage clustering is different from the initial grouping of QCC that was solely based on the overall chemical structure of the compounds seen in Section 1.5. Therefore, in order to understand the grouping of QCC presented with the ‘complete’ linkage clustering, additional properties of the compound will be explored.

Looking at Table 5.1, group #1 can be distinguished from the rest of the groups for having the least number of aromatic rings and the lack of amine groups in their structure. This group tends to have no particular trend in terms of their water solubility nor the length of acyl-chain. Moreover, according to the heat-map coloring scheme, group #1 seemed to have the least toxic effect on *E. coli*. This is also comparable to the results observed in Table 4.2 since the sphere-forming compounds (TPP, MTP, and TPA) and the acyl-chained compounds (MC, BZ, CC, CB, and Ct.C) showed the least diversity in terms of MIC among the different EmrE variant. Under these compounds only few EmrE variants were sensitive to the low concentrations of these compounds and the majority of the variants seemed to have MIC similar to that of the wild-type control. Thus, group #1 is the least toxic group of the QCC studied.

Table 5. 1: The 'complete' linkage clustering of QCC and their structural and physical properties.

QCC	Number of Aromatic Rings	Number of Amine groups	Water Solubility (mg/mL)	Acyl chain length
[1] **Group #1**				
Cetylpyridinium chloride	1	0	50	16
Cetylpyridinium bromide	1	0	50	16
Methyltriphenyl-phosphonium bromide	3	0	400	0
Tetraphenylarsonium chloride	4	0	50	0
Myristalkonium chloride	1	0	20	14
Cetalkonium chloride	1	0	20	16
Benzalkonium chloride	1	0	100	12
Tetraphenylphosphonium chloride	4	0	5	0
[1] **Group #2**				
Acriflavine	**3**	**2**	**330**	**0**
Proflavine	3	3	500	0
Methyl viologen	2	2	700	0
Ethidium bromide	4	2	10	0
[1] **Group #3**				
Rhodamine	4	1	20	0
Crystal violet	3	2	16	0
Pyronin	3	1	30	0
[1] **Group #4**				
Cetrimide	0	0	560	16
Dequalinium chloride	4	2	10	10

[1] The grouping of QCC and the color scheme represents the clustering of the dendrogram in Figure 5. 1.

When considering group #2 of the QCC ‘complete’ clustering, this group has higher number of aromatic rings, larger number of amine groups and relatively high water solubility values, which distinguish it from the remaining groups (Table 5.1). Also, this group of compounds (ACR, PRO, MV, and Et.Br) tends to have an intermediate to low level of toxicity on the majority of EmrE variants. Most of the compounds in that group belong to the poly-aromatic QCC except for MV, which belongs to the poly-charged QCC. In Table 4.2, EmrE variants had a very diverse MIC range and extreme loss of resistance for these groups of compounds. However, when you look at each substrate within the poly-charged and poly-aromatic compounds, you realize that about 15-40% of the variants were more sensitive to lower concentrations of MV, Et.Br, PRO, and ACR than the wild-type EmrE. Hence, this confirms that this group of compounds has an intermediate toxicity effect when compared to the remaining 3 groups (Figure 5.1). Also, with ‘complete’ linkage, the QCC are more likely confined according to their toxicity not necessarily according to their overall chemical structure.

Group #3 of the QCC includes CV, RH, and PY and they differ from the previous group mainly due to their lower water solubility values and having less amine groups in their structure. This group of compounds belongs to the poly-aromatic QCC original grouping and seems to belong to the higher toxicity class of compounds as reflected in the larger number of variants with very low MIC values. This is also comparable to the heat-maps generated in Figure 4.2, as more than 40% of the EmrE variants had extremely low scaled-MIC values. Therefore, the high toxicity of this group of compounds, their lower solubility, and lower number of amine groups sets it apart for the rest of the groups in the dendrogram.

Finally, group #4 seems to be the group of outliers since there are no distinguishable characteristics that sets this group from the rest. However, since we know the clustering is mainly based on the resistance profiles of EmrE variants to these compounds, it is apparent that the extreme toxicity of these compounds that set them apart from the rest of the groups. There are no chemical features shared between these two compounds but possibly they share the mechanism of action against bacterial cells since both have long acyl chains. Both DC and CET are found to direct their action towards the cytoplasm membrane causing the damage of the membrane and the release of cellular components (D'Auria et al., 1989). Therefore, the remarkable toxicity of these two compounds resulted in reducing the resistance of the majority of the EmrE variants.

The 'complete' clustering of QCC that helped in identifying different toxicity levels of the compounds differed from our original grouping of the compounds, which was solely based on their structures. The subsequent section evaluates a different grouping of EmrE variants.

5.1.2.2 'Complete' Clustering of EmrE Variants

The clustering of EmrE variants categorizes them into 3 main groups labeled A, B, and C in Figure 5.1. Starting with group C, EmrE variants, such as S72C, W45A, L103C, W31A, L93C, V66C, V69C, Y40A, G17A, W76A, P32C, I11C, and V15C congregated with the wild-type EmrE. This indicates that the resistance profiles of the above 13 variants are mostly similar to that of the wild-type. The above 13 variants, also, did not have extremely low MIC values, where most of these variants are seen in the light orange to light green color range in our earlier analysis (Table 4.2). Thus, the substitution of the Ser72, Trp45, Lue103, Trp31, Lue93, Val66, Val69, Tyr40, Gly17, Trp76, Pro32, Ile11, and

Val15 did not result in an extreme loss of resistance phenotype and those mutants were comparable to the wild-type phenotype. Moreover, the dendrogram indicates that S72C, W45A, L103C, W31A, L93C, V66C, V69C, Y40A, G17A, and W76A variants are sub-grouped from V15C, I11C, P32C and wild-type EmrE. This suggests that the resistance profile of V15C, P32C, and I11C is almost identical to that of the wild-type, which has been explained in Section 4.3 that Val15, Pro32, and Ile11, although highly conserved, they do not appear to participate in substrate recognition (at least for the substrates chosen in this study). The other 10 variants (S72C, W45A, L103C, W31A, L93C, V66C, V69C, Y40A, G17A, and W76A) may be involved in substrate recognition or indirectly facilitate in the substrate transport mechanism with a network of other residues. Furthermore, group C variants do not have common characteristics in terms of their loop/TMS location or their polarity (Table 5.2). Therefore, from the clustering of these variants (Figure 5.1), we cannot extract the trends or common characteristics that have grouped these variants together other than their resistance phenotype. However, it is apparent that group C variants tend to cluster in the group with the lower end of the 75% consensus cut-off (conservation). The above 13 amino acid residues (Ser72, Trp45, Lue103, Trp31, Lue93, Val66, Val69, Tyr40, Gly17, Trp76, Pro32, Ile11, and Val15) have percent consensus between 50-90%, where the majority of the residues (~60%) have percent consensus lower than 80%. Hence, these residues may be considered moderately conserved when compared to the remaining 20 conserved residues thereby altering them with site-directed mutagenesis did not result in an extreme loss of resistance phenotype.

Table 5. 2: The 'complete' linkage clustering of the conserved amino acid residues in EmrE and their properties.

Amino acid	TMS/Loop (L)	Polarity Characteristics	% Consensus
[1] Group A			
L7	TMS1	Non-polar	98
Y60	TMS3	Polar	88
S64	TMS3	Polar	87
A59	TMS3	Non-polar	98
I62	TMS3	Non-polar	98
S105	TMS4	Polar	80
G97	TMS4	Polar	97
A10	TMS1	Non-polar	94
T18	TMS1	Polar	88
G57	L2	Polar	90
W63	TMS3	Polar	75
I94	TMS4	Non-polar	90
[1] Group B			
M21	TMS1	Non-polar	87
G26	L1	Polar	75
F27	L1	Non-polar	90
F44	TMS2	Non-polar	94
G90	TMS4	Polar	92
P55	L2	Non-polar	94
L47	TMS2	Non-polar	90
E14	TMS1	Acidic	98
pMS119EH	0	0	0

Amino acid	TMS/Loop (L)	Polarity Characteristics	% Consensus
[1] Group C			
S72	TMS3	Polar	75
W45	TMS2	Non-polar	70
L103	TMS4	Non-polar	75
W31	L1	Non-polar	60
L93	TMS4	Non-polar	76
V66	TMS3	Non-polar	98
V69	TMS3	Non-polar	75
Y40	TMS2	Polar	94
G17	TMS1	Polar	70
W76	TMS3	Non-polar	50
P32	L1	Non-polar	88
I11	TMS1	Non-polar	90
Wild-type	0	0	0
V15	TMS1	Non-polar	85

[1] The grouping of EmrE amino acid residues and the color scheme represents the clustering of the dendrogram in Figure 5. 1.

Group B of the EmrE variants included M21C, G26C, F27C, F44C, G90C, P55C, L47C, and E14D. The resistance profiles of these variants were most similar to that of the negative control, where *E. coli ΔacrB* was transformed with an empty pMS119EH plasmid. This indicates that the mutations of Met21, Gly26, Phe27, Phe44, Gly90, Pro55, Leu47, and Glu14 to another amino acid caused remarkable loss of resistance to the majority of QCC. This agrees with the earlier heat-map results (Table 4.2), where the above variants mostly occupied the red and dark orange cluster with the lowest MIC values for the majority of QCC. Thus, Met21, Gly26, Phe27, Phe44, Gly90, Pro55, Leu47, and Glu14 residues are critical for the resistance function of EmrE suggesting their involvement in substrate recognition or transport. Moreover, there are no distinguishable common features between the above 8 residues in terms of their loop/TMS location or polarity. However, the majority of the residues cluster in the group of residues with considerably high percent consensus, which is an indication of their conservation. The 8 residues (Met21, Gly26, Phe27, Phe44, Gly90, Pro55, Leu47, and Glu14) have percent consensus that ranges from 75-98%, where 6 out of the 8 residues have ≥90% consensus. Therefore, these residues are highly conserved and hence can explain their critical role for proper functionality of the protein, EmrE, and their involvement in the substrate recognition or the transport mechanism.

Group A variants include L7A, Y60A, S64C, A59C, I62C, S105C, G97C, A10C, T18C, G57C, W63A, and I94C. This group clusters closer to group B and the furthest away from group C in the dendrogram (Figure 5.1). This means that group A variants are resistant to more QCC and may have slightly higher MIC than group B. However, they are resistant to fewer QCC and have lower MIC than group C (Figure 5.1 and Figure 4.2). Also, the

amino acid residues (Leu7, Tyr60, Ser64, Ala59, Ile62, Ser105, Gly97, Ala10, Thr18, Gly57, Trp63, and Ile94) do not cluster because of their TMS/loop location or polarity characteristics (Table 5.2). However, these residues tend to have remarkably high percent consensus that ranges from 75% to 98% similar to that of group B. Also, ~60% of the 12 residues listed above have percent consensus of 90 or higher. Therefore, group A and group B variants are clustered according to their resistance profiles that illustrate lower MIC values than group C and are also distinguished form group C due to their remarkably high percent consensus (conservation). Consequently, this also confirms our earlier hypothesis that amino acids, which are highly conserved, play a critical role in the resistance functionality of EmrE.

The hierarchical agglomerative clustering analysis led to conclusions about the amino acids that are similar to the ones presented in section 4.2 and 4.3. However, the clustering of the compounds differed from how the compounds were grouped initially, which was solely based on the chemical structure of the compounds. The 'complete' linkage clustering predicted a different grouping of the compounds based on the compounds toxicity to EmrE variants. From the clustering seen in Table 5.1, we can predict that the compounds were also grouped by the *E. coli* sensitivity according to certain structural and physical properties, such as the number of amine groups, the number of aromatic rings, length of acyl-chains and solubility in water. Moreover, EmrE variants clustering predicted the presence of 3 groups, with one group having similar resistance phenotype to that of the wild-type protein (group C), another group with resistance phenotype that is similar to the negative control (in another word, the resistance phenotype was abolished in group B), and the final group with an intermediate resistance phenotype

(group A). In the following section, the significance of the conserved amino acid residues and their role in the resistance mechanism of EmrE will be explored in terms of these residues possible position and locality in the protein helices.

5.2 Helical Wheel Analysis

The helical wheel diagram is a projection down the axis of an alpha helix. In Figure 5.2 the helical wheels are used to illustrate the localization of each one of the conserved amino acids that when mutated, they alter the resistance phenotype of the protein EmrE. The helical wheels also include a vector arrow, which is predicted to point towards the lipid-exposed surface of the helix. This prediction is based on the hydrophathy moments of the residues within the helix and the arrow predicts the directionality of the residues facing the lipid side. Thus, the residues facing to the inside of the protein are expected to be on the opposite direction of which the arrow is pointing to. Consequently, the arrows point away from the majority of colored residues indicating that these colored residues in all TMS are not exposed to the lipid and are likely facing the inner core of the protein involved in substrate recognition or protein-protein interaction (oligomerization). The vectors in Figure 5.2 are in agreement with the calculations of Edwards and Turner (1998) that also calculated the amino acid sequence variability, and hydropathy on a set of six homologous proteins of the SMR family.

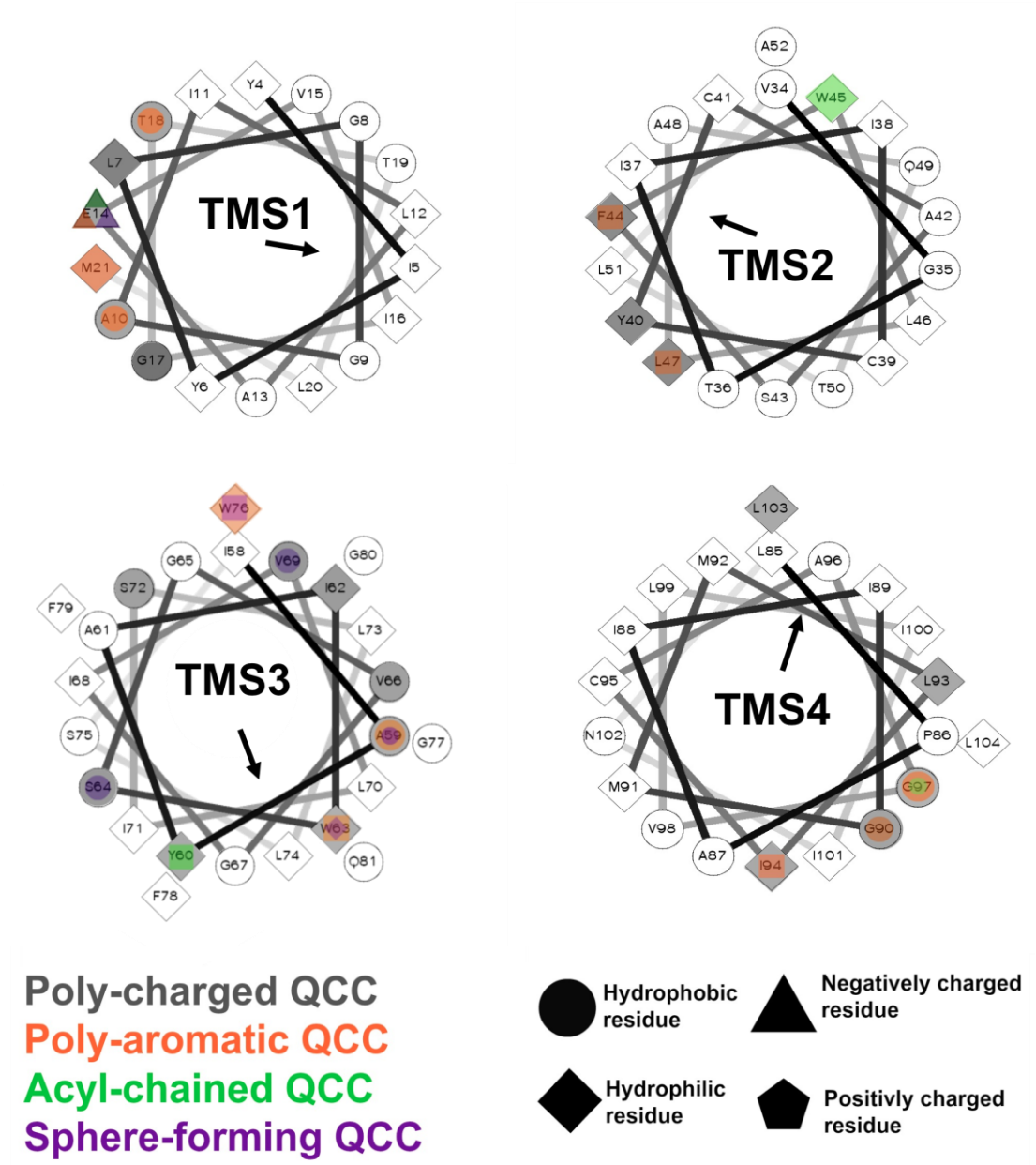


Figure 5. 2: Helical wheel analysis of the 33 EmrE conserved amino acid residues. The helical wheels projections of the 4 transmembrane strands (TMS) were generated using Wheel.PI at 3.6 residues per turn pitch. The amino acids are indicated with one letter abbreviation and the geometric shape surrounding each residue indicates the property of that residue according to the four classifications of QCC used in this study (see inset legend). The coloring scheme of the residues represents how the mutation of these residues affects the resistance to group(s) of QCC. The arrows a predicted vector pointing towards the lipid-exposed surface of each helix.

In TMS1, there were 9 residues with 75% conservation or higher (Figure 3.1). Out of the 9, 8 residues were successfully mutated generating 8 EmrE variants. As seen in Figure 5.2, most of the residues that displayed importance for QCC resistance (colored residues) were grouped on one face of the TMS. Also, all of these residues were on the same face as the Glu14 residue and mostly were important for poly-charged and/or poly-aromatic QCC. Glu 14 is required for both substrate and proton exchange mechanism (Muth and Schuldiner, 2000). It is also postulated that Glu14 is deprotonated upon ligand binding and protonated when the ligand is released on the other face of the membrane. This suggests that this face of the TMS participates in the assembly of the binding pocket and also displays preference to 2 groups of the QCC. Our data supports previously studied amino residues that were on the same face as Glu14, which also were mutated using site-directed mutagenesis (Gutman et al., 2003). When they replaced Leu7, Ala10, Ile11, Gly17, and Thr18 replacement with Cys, they noticed that L7C, A10C, T18C, I11C, and G17C variants had no binding or very low binding to TPP+ indicating that these residues are possibly involved in substrate binding. Therefore, with our resistance phenotype analysis against 19 QCC, we have found that Leu7, Ala10, Thr18, Gly17, Met21and Glu14 are critically important for substrate recognition and we have also determined substrate preferences of these residues. For example, Met21 recognizes poly-aromatic QCC while Leu7 and Gly17 prefer poly-charged QCC. On the other hand, Thr18 and Ala10 prefer both poly-aromatic and poly-charged QCC. Hence, we speculate with certainty that residues on the same face as Glu14 participate in substrate recognition and possibly mediate their transport.

In TMS2, where the least number of conserved amino acid residues were found, 4 residues showed conservation of 75% or higher. All of these 4 residues were tested and one face containing 3 residues of the TMS displayed specificity towards poly-charged/poly-aromatic QCC. However, on a completely different face, one residue displayed selectivity towards a specific group of compounds, which was acyl-chained QCC. The acyl-chained QCC are complex structures that span in the lengths of 12-16 carbons. Therefore, it is justifiable that a different amino acid on a different face is required for interacting with such complex compounds that diverge significantly from the rest of the QCC groups. This could also suggest that in order for EmrE to export such compounds, TMS2 is required to undergo a reconformation to allow such an activity. Previous analysis with planar compounds (i.e. ethidium) has shown small rearrangements of the transmembrane domain and the binding site region of the EmrE dimers are remodeled upon the take up of such compounds (Korkhov and Tate, 2008). Therefore, further analysis using Cryo-EM can be worthwhile to confirm if remodeling of TMS occurs upon the interaction with acyl-chained QCC.

TMS3 contains the highest number of conserved amino acids with 75% conservation or higher. Out of the 11 conserved residues, only 9 were successfully mutated generating the 9 EmrE variants. As seen in Figure 5.2, the residues seemed to be scattered on different faces of the TMS and mostly displayed resistance to poly-charged QCC. A particular face of the TMS was important for conferring resistance to 3 groups of the QCC where a tryptophan residue and an alanine are found. Based on previous biochemical and mutagenesis studies, Trp63 has been implicated in substrate binding and transport (Chen et al., 2007). Hence, it suggests that this particular face possibly participates in the assembly

of the binding pocket and possibly faces the same side as the Glu14 residue. Moreover, EmrE functional transporter has been considered to form a trimer in addition to dimers (Yerushalmi et al., 1996), I speculate that TMS3 may be localized in the centre of a dimer or trimer structure where each face of TMS3 may be participating in the binding pocket of every oligomer. Furthermore, the organization of TMS presented by X-ray structure (Chen et al., 2007) illustrated that TMS3 and TMS1 are the central helices surrounded by TMS 4 and TMS1. Consequently, the different faces of TMS3 can be involved in the substrate-binding pocket or protein-protein interactions with the helices mediating the transport of the different substrates.

In TMS4, there were 7 residues with 75% conservation or more. Out of the 7 residues, 5 EmrE variants were successfully generated. It is also notable that 3 of the residues (Ile94, Gly90, and Gly97) were grouped in one face of the TMS and displayed sensitivity to poly-charged and poly-aromatic QCC when mutated. Also, all of the 7 conserved residues are also found in the same face as Ile94, Gly90, and Gly97 except for Leu103 and Leu93. Therefore, the conservation and the importance of these 3 residues for conferring resistance to more than one QCC group suggest that these residues may participate in the assembly of the substrate-binding pocket possibly facing the same side as the Glu14 residue. Moreover, the specificity of Leu93 and Leu103 to only one group of QCC (poly-charged) suggests the preference of a different face to a particular group of drugs as seen in TMS2 (face-specific interaction). Hence, remodeling of the TMS maybe happening to accommodate for the export of such compounds.

The helical wheel analysis further strengthened our hypothesis that conserved amino acids play an important role in the poly-substrate specificity of EmrE. Also, some of

these residues cluster around particular faces of the TMS may suggest helical reconformation in order to accommodate for QCC with particular features. Recent NMR study also supports the idea of conformational plasticity in ligand binding (Cho et al., 2014). The recent studies illustrated that the unbound EmrE oscillates between the inward-open and outward-open structural conformations at an exchange rate (k_{ex}) of $\sim 300\ s^{-1}$, which suggests that EmrE is able to overcome the energy barrier for the conformational changes enabling it to achieve the broad multidrug recognition and resistance (Cho et al., 2014 & Morrison and Henzler-Wildman, 2014). Moreover, studies completed by Bay and Turner (2012) showed that the increase of particular substrates (TPP+, methyl viologen, ethidium and cetylpyridinium) concentration enhanced the multimerization of the protein. Also, the reduction in multimerization of the protein upon the mutation of Glu14 (active site residue) suggested that substrate binding influences multimerization. Therefore, the fact that all faces of TMS3 are involved in different substrate recognition can further confirm that EmrE functions as a oligomer, where particular TMS is found in the centre of the complex and the residues of that TMS recognize particular substrates, hence mediating the multimerization. In the following chapters, EmrE poly-substrate specificity will be further illustrated in a Venn diagram.

Chapter 6: Discussion

6.1 Poly-Substrate Specificity of Multidrug Resistance Transporters Binding-Pocket

Multidrug resistance presents an imminent challenge to healthcare and scientists. The most recurring reason for this phenomenon is the active transporters that extrude out of the cell a broad range of chemically and structurally distinct cytotoxic substrates. The question then becomes how is this multi-specificity for substrates achieved? First, allow me to define multi/poly-specificity for substrates. Transporters with flexible binding pockets become accessible to many dissimilar molecules. However, those dissimilar molecules must have particular chemical or structural feature(s) that allow for their recognition by the protein thereby the poly-specificity.

The main challenge with understanding the molecular mechanism of multidrug transporters is that these proteins contradict the main dogma of enzymes 'lock and key' model (Fischer, 1894). The model explains an effective binding of substrates that have complementary geometry to the specific binding/active site of the enzyme. However, Neyfakh (2002) indicates that this specificity for enzyme substrates is essential since the majority of their substrates are hydrophilic molecules. Therefore, these molecules will have strong and competitive interactions with water, where the only way to overcome this competition is by possessing a complementary spatial match for the substrate. On the other hand, MDR transporters bind hydrophobic substrates that do not have to compete with water molecules for binding (Neyfakh, 2002). Thus, a hydrophobic binding site is a sufficient requirement for binding such substrates making 'lock and key' system unnecessary.

Further research on understanding MDR structures and poly-specific transport mechanisms attempted to explain the plasticity of the binding site of these transporters. Two opinions emerged on drug-binding specificity of MDR transporters. The first argues that MDR transporters have multiple separate binding sites to accommodate for the different substrates (Tamai and Safa, 1991; Putman et al., 1999; Mitchell et al., 1999; Schumacher et al., 2004). The second opinion implicates that MDR transporters have a single large and flexible binding pocket that can accommodate a wide range of substrates (Paulsen et al, 1996B; Klyachko et al., 1997; Schumacher et al., 2001; Loo et al., 2001A/B; Adler and Bibi, 2004). The majority of the data on several MDR transporters suggest that the later opinion is the most plausible. However, I will briefly discuss the experimental work that led to each one of the opinions presented.

Research by Putman et al. (1999) suggested that LmrP, an MDR transporter, has multiple drug binding sites. LmrP of *Lactococcus lactis* is a transporter that belongs to the MFS family of transporters and is known to mediate the extrusion of a dye called Hoechst, which is a mutagenic and carcinogenic molecule that interferes with DNA. The kinetics analysis demonstrated the inhibition of Hoechst transport by LmrP as a result of competitive binding to the binding cavity by quinine and verapamil drugs. Also, there was a noncompetitive inhibition of Hoechst transport by nicardipin and vinblastin and uncompetitive inhibition by TPP+. Therefore, the three types of Hoechst transport inhibition indicated the presence of multiple binding sites for these drugs in the LmrP transporter (Putman et al., 1999). Also, another MFS family member, QacA from *Staphylococcus aureus*, which confers resistance to a wide range of QCC may have distinct binding sites for monovalent and divalent QCC (Mitchell et al., 1999). The kinetics

analysis indicated that the export of ethidium by QacA was competitively inhibited by monovalent cations (i.e. benzalkonium) while noncompetitively inhibited by divalent cations (i.e. propamidine). Therefore, different binding sites may exist in QacA with different substrate preferences. Additional kinetic analysis on P-glycoprotein, which is a member of the ABC family of transporters, indicated noncompetitive inhibition on azidopine (a photoactive dihydropyridine calcium channel blocker) binding by cyclosporin A and Vinblastine (Tamai and Safa, 1991). Hence, they suggested that P-glycoprotein also has multiple binding sites.

In addition to competitive and noncompetitive inhibition experiments, other observations of MDR transporters binding two substrates simultaneously reinforced the thought of having several binding sites in one transporter. QacR, for example, which is a *S. aureus* multidrug-binding transcription repressor, can bind both ethidium and proflavine simultaneously (Schumacher et al., 2004). The binding affinity of QacR to proflavine was similar in the unbound and in the ethidium-bound QacR. However, ethidium affinity was significantly decreased once proflavine was bound to QacR. Also QacR-substrate bound structures revealed a shift in the position of ethidium binding when the protein was pre-bound to proflavine as compared to the protein bound to ethidium alone. This suggested that the binding of proflavine is mediated by the interaction of the substrate with the charged residue Glu57 and Glu58 of the protein while ethidium binding prefers the interaction to aromatic and hydrophobic residues. Therefore, QacR is believed to have different binding sites for different substrates.

Despite the competitive, noncompetitive analysis, and the models of two substrates-bound transporters that speculate the presence of separate multiple binding sites in one

transporter, all of the above experimental data is also compatible with the second opinion of one large and flexible binding pocket of MDR transporters. That can be attributed to the fact that one large and flexible binding pocket will fit more than one substrate at the same time, which can explain the noncompetitive, uncompetitive and the two-substrates bound phenomena. Also, the flexibility of the binding pocket provided by the diversity of amino acid residues and their preferences to interacting with substrates of particular features can explain the competitive and noncompetitive inhibition of the binding phenomena. Therefore, the recent research and my current data lean towards explaining the poly-specificity of MDR transporters by having a single large and highly flexible binding pocket, where within it several niches of amino acid residues that interact with substrates of particular characteristic(s). This notion tries to accommodate both opinions while stressing the concept of single binding pocket and emphasizes the important role of residues lining the binding cavity thereby the substrate-transport mechanism does not solely rely on the highly conserved charged residue of these transporters (i.e. Asp or Glu).

There are several ways in which the substrate establishes itself in the binding pocket of the transporter. One of the major forces that drive the binding is van der Waal, where a hydrophobic interaction occurs between hydrophobic and aromatic residues of the binding pocket, and the hydrophobic substrates (Higgins, 2007). Such interactions are present in large flexible binding chamber of several MDR transporters, including P-glycoprotein, MsbA, MdfA, LmrP, Bmr, QacA, QacB, BmrR, QacR and EmrE.

The large chamber of P-glycoprotein is formed by two transmembrane domains, where each is composed of six α-helices (Toyoshima et al., 2000). The binding-pocket (with 60-70Å external diameter) can transport a diversity of substrates, including large

polypeptides and a number of drugs used in cancer chemotherapy, immunosuppression, hypertension, allergy, infection, and inflammation (Rosenberg et al., 2001; Kim, 2002). Further research also indicated the presence of a common binding pocket for P-glycoprotein substrates, where within it different substrates have overlapping or distinct regions of interactions (Loo et al., 2003A). The rhodamine binding region overlaps with that of colchicine and calcein, which is in close proximity to the region that interacts verampil and Hoechst. Loo et al. (2003B) further confirms the presence of a large binding pocket that can accommodate both verapmil and Tris-2-maleimidoethylamine simultaneously suggesting that they occupy different regions in the common binding pocket. The large binding pocket is particularly emphasized because of the shape of P-glycoprotein binding chamber that is suggested to look like a 'funnel' and the evidence for substrate-induced fit mechanism for the drug binding (Loo et al., 2003B). Therefore, P-glycoprotein is believed to have a large binding pocket, where there could be linked-distinct regions of substrates-residues interactions or overlapping regions.

The MsbA, which is a P-glycoprotein homologue in *E. coli*, also shares the feature of having a large binding-pocket to accommodate the different substrates it transports. MsbA binding pocket has all of its α-helices tilted by 30-40° from the membrane, forming a cone shaped structure (~25Å) that allows the transport of variety of amino acids, sugars, toxins, lipids and other medically related drugs (Chang and Roth, 2001).

Moreover, MdfA, which is an *Escherichia coli* MFS transporter, is shown to have a large hydrophobic pocket with embedded charged residues that mediate the transport of positively charged drugs and proton transfer (Edgar and Bibi, 1999). In addition to the charged residues, mutational analysis demonstrated that the residues lining the central

binding cavity of the transporter participate in substrate recognition and translocation (Higgins, 2007). These residues include hydrophobic, aromatic, and some polar amino acids. These observations indicated that the different residues could be more or less important for the binding of different substrates. Additional support of the plasticity of MdfA binding pocket indicated that the loop between TMS10 and TMS11 of the protein collaborates with the membrane-embedded residues to mediate the binding of particular substrates (Adler and Bibi, 2004). Therefore, the involvement of protein loops in addition to the several amino acid residues in the binding-pocket of MdfA indicate that the transporter possesses a large and flexible binding pocket with regions of residues having different substrate preferences.

The *B. subtilis* multidrug transporter that belongs to the MFS family, Bmr, confers resistance to several toxins, such as ethidium, rhodamine, acridine dyes, doxorubicin, chloamphenicol, and fluoroquinolone antibiotics (Neyfakh et al., 1991). The mutation studies of Val268, Phe134, and Phe306 in Bmr identified how critical these residues to the binding of reserpine substrate (Klyachko et al., 1997). Interestingly, these residues are found in different regions of the Bmr transporter, where Phe143 found on TMS5, Val268 found on TMS9, and Phe306 on TMS10, which suggested that substrate-binding pocket of Bmr is large enough to involve several interacting transmembrane segments. Moreover, these residues displayed different substrate preferences, where the mutations of Phe143 affected the transport of reserpine, norfloxacin, acriflavine, and ethidium, but not rhodamine or TPP+. On the other hand, Val286 is only important for reserpine transport while Phe306 does not affect the transport of ethidium. Therefore, different substrates interact with different regions of the protein.

To further confirm that the binding pocket of MDR transporters has different amino acid niches that bind particular substrates, it has been demonstrated that the difference in substrate specificity between QacA and QacB results from differences in only a few amino acid residues (Paulsen et al., 1996A). QacA and QacB are *S. aureus* multidrug transporters that belong to the MFS family. Nucleotide sequencing showed that *qacA* gene differed from *qacB* gene by seven nucleotides substitutions thereby giving QacA the ability to confer resistance against divalent cations while QacB confers lower level of resistance to the same substrates. This is believed to be as result of the presence of acidic amino acid residue at position 322 in QacA that is not found in QacB (Paulsen et al, 1996A). Therefore, the differences in substrate-specificity can result from subtle changes in the amino acid residues of the binding pocket, which leads to the notion that diverse amino acid composition within the binding pocket can explain the plasticity of most MDR transporters.

The plasticity of a large binding pocket is not a phenomenon that is only observed among the MDR transporters. Instead, we have gained a wealth of knowledge on drug-binding sites from multidrug-binding transcription factors, such as QacR and BmrR. QacR is a *Staphylococcus aureus* multidrug-binding protein that represses the transcription of QacA. The binding pocket of QacR contains several glutamates and aromatic residues, where the acidic residues are found to be involved in drug binding (Glu57, 78, 90, and 120) and the aromatic residues have electrostatic interactions with the aromatic substrates (Schumacher and Brennan, 2002). Structural analysis of QacR showed that substrates, like pentamidine and hexamidine that differ by a single methylene carbon, bind different determinants in the protein. Also, pentamidine binding does not necessarily require the

presence of a complementary charged residue (i.e. Glu) (Murray et al., 2004). In addition, substrate bound-QacR structures revealed a number of linked binding regions within one extended and multifaceted binding pocket (Schumacher et al., 2001). Thus, the ability of QacR to bind to very similar aromatic substrates (pentamidine and hexamidine) in different ways and such a multifaceted pocket underscores the plasticity of its large multidrug-binding pocket.

BmrR, bacterial multidrug resistance regulator from *Bacillus subtilis*, can bind several structurally unrelated lipophilic cationic substrates and activate the transcription of the multidrug transporter gene, *bmr* (Godsey et al., 2002). There are many similarities between BmrR and QacR binding pockets in that both have the positively charged substrate interacting with their acidic residues and the hydrophobic interactions between aromatic and hydrophobic residues, and the substrates to strengthen those electrostatic binding (Neyfakh, 2002). Hence, the plasticity of transcriptional factors' binding pockets further supports the notion of a common large binding pocket for the different substrates.

Consequently, these data reinforce the idea of the plastic multi-drug binding pocket with a number of shared 'sub-sites' or regions that are linked to accommodate the wide range of structural and chemical dissimilarities of their substrates. This can also be seen in the data obtained from my EmrE mutagenesis study, where the subtle alterations of conserved amino acid residues changes the resistance profiles of the protein in a substrate-dependent manner. Some amino acid mutations resulted in complete loss of the resistance activity to particular compounds when compared to wild-type (i.e. residues colored in red in Table 4.2) while others slightly reduced the resistance to particular compounds (i.e. residues in colored in yellow-light green in Table 4.2). Also, no single residue, except for

the only membrane-embedded charged Glu14, resulted in the loss of resistance to the entire set of compounds tested. Moreover, the conserved residues that have been found to be involved in substrate recognition and possibly binding were hydrophobic, aromatic or polar. Therefore, the structural and chemical diversity of the 19 substrates tested in addition to the involvement of multiple amino acids in substrate recognition strongly agrees with the earlier binding-pocket hypothesis. Figure 6.1 illustrates a cartoon representation of EmrE binding-pocket, which shows the different residues' substrate preferences. EmrE, an anti-parallel dimer, helps in the formation of a large and flexible binding pocket that allows the binding of substrates with different features, such as poly-aromatic, acyl-chained, poly-charged, or sphere-forming QCC. Based on the helical organization suggested by X-ray crystallization and Cryo-EM structural studies (Chen et al., 2007; Fleishman et al., 2006), there is an overlap in substrate recognition between the different conserved residues since there is no particular TMS that is specifically involved in interacting with one group of QCC. Hence, these observations confirm that the plasticity of the binding-pocket results from a large binding pocket with several interacting transmembrane segments (all eight TMS) rather than several unlinked binding sites.

To summarize, the negatively charged residue found in the binding-chamber of the majority of MDR may play a key role in proton-mediated transport of positively charged substrates but is not the only residue that mediates substrate recognition and binding. Instead, several amino acid residues that are possibly lining the binding-pocket play critical roles in the binding of a wide range of substrates. Therefore, these amino acid niches are linked within a single large binding pocket where two substrates can bind simultaneously or structurally distinct substrates can bind.

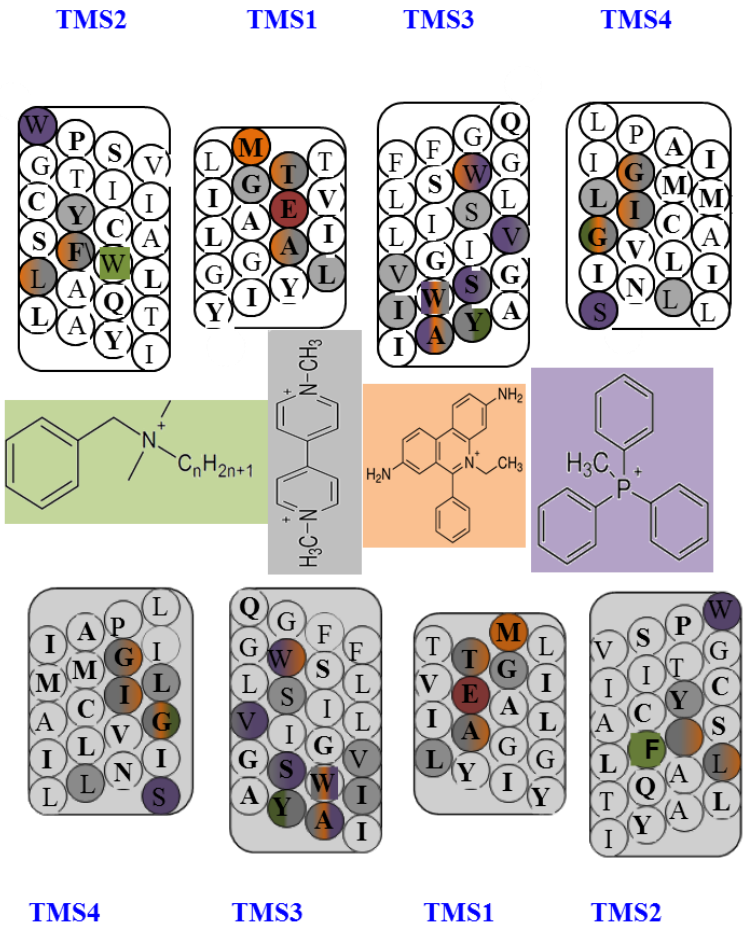


Figure 6. 1: The large binding-pocket of EmrE.

The cartoon suggests a single large binding pocket for EmrE where the poly-substrate specificity can be explained by the involvement of different amino acid residues in recognizing or binding different substrates. The mutations of the conserved amino acid residues affected the resistance to the four different groups of QCC, Poly-aromatic QCC, Sphere-forming QCC, Poly-charged QCC, Acyl-chained QCC. The residues are colored according to the group of QCC that they are involved in recognizing. This was determined based on the MIC of EmrE-mutants to each one of he 19 compounds, where each one of the conserved amino acid was mutated to a different residue. Therefore, loss of resistance is seen as decrease in MIC when compared to the wild-type EmrE. Only the protein transmembrane segments (TMS) are shown here and organized in anti-parallel model according to Cryo-EM and X-ray structural analysis (Fleishman et al., 2006; Chen et al., 2007). The Glu14 is shaded in red. The overlap in the substrates' recognition by the different residues suggests that there is a single binding pocket where different residues are involved in mediating the binding to the wide range of substrates.

6.2 Conclusions: Conserved Amino Acids Can Explain The Poly-specificity of EmrE

Conferring resistance to a particular toxic quaternary cationic compound reflects the balance between multiple factors, including the levels of EmrE expression and the antagonism between passive inward leaks of the substrate and its elimination by the transporter. Although the level of EmrE expression is unknown for each variant, the lack of resistance to a particular substrate is sufficient qualitative data to indicate changes in substrate recognition and transport. This is especially true when none of the EmrE variants tested resulted in abolishing the resistance of the cells expressing them towards all the compounds tested. Thus, we can speculate that these residues are most likely involved in substrate recognition/transport rather than protein expression, folding, insertion, stability or oligomerization.

6.2.1 Conclusions from Initial Data Analysis

Based on the above assumption and the pictorial representation of the Venn diagram (Figure 6.2), I hypothesize that EmrE has a large binding pocket with multiple residues involved in giving EmrE the poly-substrate specific nature. Previous studies with other multidrug resistance transporters (MDR) have produced high-resolution structures of proteins that interact with multiple drugs confirming the hypothesis that different substrates may interact with different residues in the large binding pockets of these MDR (Neyfack, 2002; Godsey et al., 2002; Murray et al., 2004). Also, from Figure 6.2, we can see that this large binding pocket of EmrE may have two components, where there is a common binding site of the various substrates and another unconventional binding site that may require conformational changes or rotation of certain helices to accommodate for specific substrates. For example, we can say that a residue that is involved in conferring resistance

to more than one group of QCC is part of the common binding site. These residues include Glu14, Gly97, Ala59, Trp63, Pro55, Trp76, Tyr60, Ala10, Thr18, Phe44, Leu47, Gly57, Gly90, and Ile94. Residues like Glu14, Gly97, Ala59, Ser64, Val69 and Trp63 tend to be important for a wide range of QCC (3 groups or more) and these residues are found on TMS1, TMS3, and TMS4. As discussed earlier, TMS1-3 are believed to be involved in substrate binding/transport while TMS4 is more involved in the dimerization of the protein (Chen et al., 2007). Thus, these residues are critical for having a robust QCC resistance activity for EmrE and are believed to be important residues in the common binding site since they have the ability to interact with compounds of different chemical characteristics.

The residue Glu14, for instance, participates in the substrate binding/transport upon its deprotonation. The change of Glu14 to Asp, which is another acidic residue that lacks the methylene group, would be positioned differently in the binding pocket when compared to the wild-type. Hence, altering the common binding site, which results in reduced resistance against the majority of substrates tested. Moreover, E14D variant has a lower pKa at this residue position when compared to wild-type (pKa=7.5 for Glu14 and pKa=5 for Asp14), which affects the coupling mechanism of binding and release of the substrate with the release and binding of the protons (Gutman et al., 2003). This is because at lower pKa the Asp residue is a stronger acid than Glu making it better at deprotonation but needs a stronger acidic environment for it to bind the substrate. Thus, Asp14 lacks the coupling between substrate/proton transport, which results in accumulation of substrates.

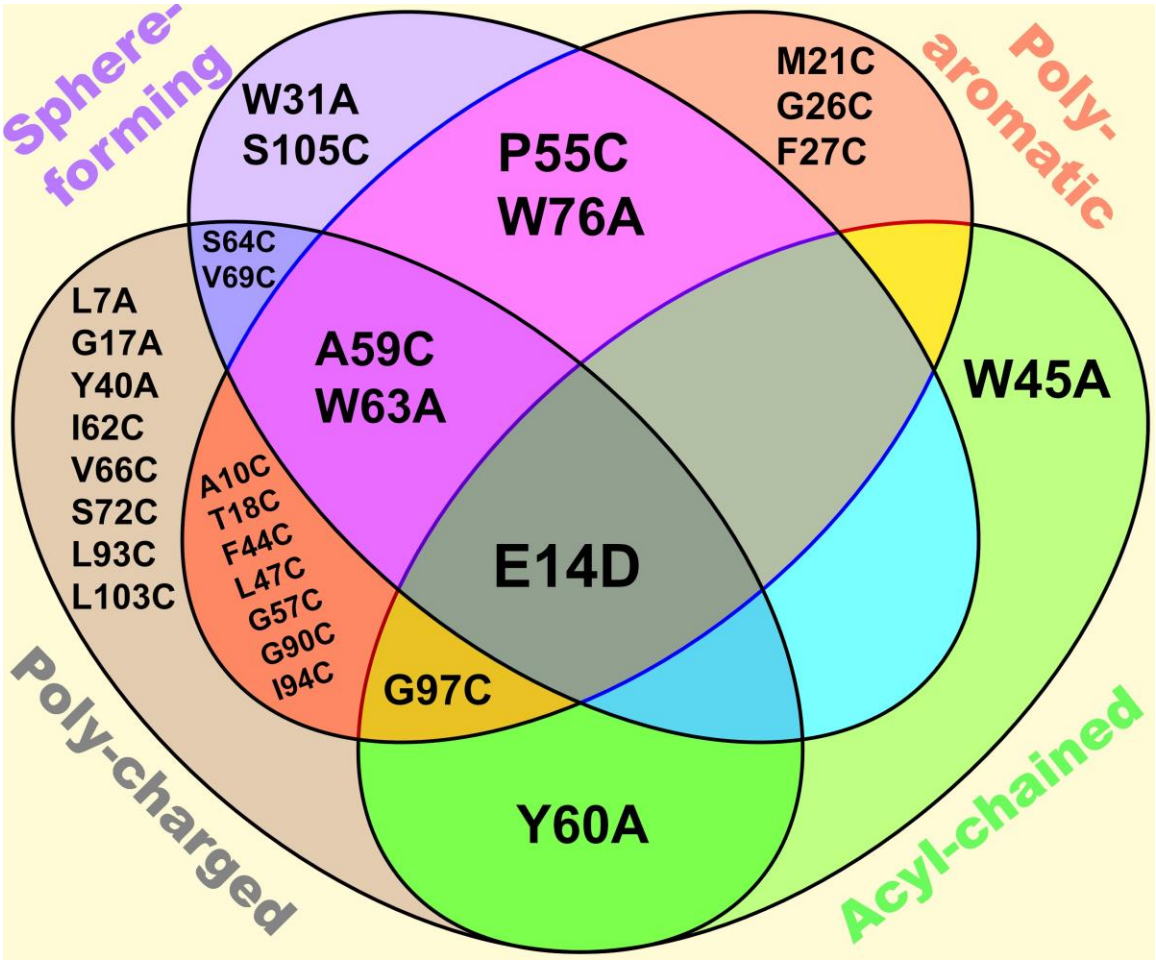


Figure 6. 2: Venn diagram illustrating that multiple amino acid residues are involved in explaining the poly-specific nature of EmrE.

Diagram depicts the sensitivities of EmrE variants to the toxicity of the 4 groups of QCC. Each EmrE variant is generated via site-directed mutagenesis altering single conserved amino acid in EmrE and each was tested against a variety of QCC. Each EmrE variant is written where the first letter represents the amino acid targeted for mutation and the second letter is the new amino acid. The selection of EmrE variant was based on which variant shows a $\geq$ 2-fold decrease in MIC when compared to wild-type EmrE and this decrease in MIC must be observed in $\geq$ 50% of the compounds under particular QCC group. The four groups of QCC have been categorized based on their structural similarities (See Section 2.5.2).

Residues like Ala59 and Gly97, which are responsible for conferring resistance to 3 out of the 4 groups of QCC do not have a known function that are discussed in previous literature. Knowing that Gly97 is in TMS4, it might be involved in the dimerization of the protein. However, the importance of this residue for the poly-substrate specific activity of EmrE strongly suggests that this residue may participate in the common binding site of the large binding pocket and hence is involved in interacting with the majority of the substrates tested. In addition, Ala59 is another residue that might be involved in the common binding site due to the same reasons explained above. Since Ala59 and Gly97 are found on different TMS, it can explain the differences between substrate preferences of these two residues. For example, Gly97 is important for conferring resistance to acyl-chained, poly-aromatic, and poly-charged QCC while Ala59 is important for conferring resistance to poly-aromatic, poly-charged and sphere-forming QCC. Thus, the presence of both residues is an absolute requirement for EmrE to confer resistance to all four groups of QCC. Both of these residues have small to no side-chains and alanine differs from glycine by an extra methyl group. Thus, the substitution of these residues to cysteine (introduction of the thiol group) has altered the side chain significantly in the common binding site hence altering the specificity of the binding site to the substrates. We can conclude that conserved Ala59 and Gly97 are critical for conferring resistance to QCC because of their participation in the common binding site of the binding pocket and their ability to recognize substrates of different structural features.

The last residue that is involved in conferring resistance to 3 out of the 4 QCC groups (poly-aromatic, poly-charged, and sphere-forming QCC) is Trp63. As described earlier, Trp63 is the tryptophan with the highest percent consensus (conservation) when

compared to the other tryptophans in EmrE (Table 3.1). It is also believed to be involved in substrate binding/transport and the stabilization of Glu14 (Elbaz et al., 2005). This tryptophan residue may play an important role in maintaining protein interaction with substrates through hydrophobic stacking interactions (Vazquez-Ibar et al., 2003). Thus, these characteristics of Trp63 can explain its importance for EmrE-substrate interactions with the majority of QCC tested. Moreover, the involvement of the other aromatic residues (i.e. Trp76 and Tyr60) in conferring resistance to two groups of QCC may confirm the involvement of aromatic residues in the large binding pocket of EmrE. Previous studies with other MDR such as QacR and BmrR suggested that aromatic residues like tyrosine and tryptophan could provide an environment that may explain the unusually high pKa of central carboxyl residues (i.e. glutamate) and allow for interactions with substrates (Schumacher et al., 2001; Zheleznova et al., 1999). The substitution of aromatic residues like in Trp63, Tyr60 and Trp76 with an alanine results in lowering the proteins resistance activity to the majority of QCC tested and seems to have an unfavorable effect on the binding pocket of EmrE, which suggests that the aromatic rings at these positions are critical for protein activity. Therefore, we can speculate that Trp63, Tyr60, and Trp76 are critical residues for the common binding site of the large binding pocket because they are involved in interacting with substrates that possess different chemical characteristics. Again, the presence of all the above 3 aromatic residues on TMS3 create a niche within the binding pocket that allows the protein to recognize compounds in all 4 groups of QCC.

Other residues that are involved in conferring resistance to 2 QCC groups, such as Pro55, Ser64, Val69, Ala10, Thr18, Phe44, Leu47, Gly57, Gly90, and Ile94 may participate in the common binding site as well. From these residues, only few have known functions

that were suggested in previous literature. For example, Ala10 is believed to play a central role in interacting with the carbonyl group on Glu14 and may contribute to the increase of pKa at that position (Gutman et al., 2003). Moreover, the Schuldiner group (2002) suggests that Ala10 residue plays a key role in the coupling activity between proton and substrate efflux of Glu14. Therefore, the close proximity of Ala10 to Glu14 and the conservation of that residue can justify it is importance in conferring resistance to cationic compounds with different characteristics (poly-charged and poly-aromatic). Furthermore, previous studies with A10C variant showed decrease host resistance to ethidium, acriflavine, and methyl viologen, which confirms the importance of this residue to poly-charged and poly-aromatic QCC. My studies with larger group of compounds suggests the existence of A10C in the common binding site were the determinants on that residue can interact with different features on the substrates.

The other residues include Thr18, Phe44, Leu47, Gly57, Gly90, Ile94, Ser64, Val69 and Pro55 do not have known functions. However, I speculate that they are involved in the common binding site, which is part of the larger binding pocket and are critical residues for conferring resistance to at least 2 groups of QCC. Thus, they either participate in substrate recognition or in the transport mechanism. Some of these residues with smaller or less complex side-chains (i.e. Gly and Val) may play an important role in giving flexibility to the TMS they are located in. The small side-chains give the helix more freedom for conformational changes or tilting which has been found to be happening in other MDR to mediate substrate transport (O'Neil and DeGrado, 1990). The other more complex amino acid side-chains are important for stabilizing the common binding site allowing for specific interactions between the amino acid and the substrate of its own preference hence

explaining the poly-substrate specificity of EmrE.

Moreover, there is a group of amino acids that are involved in conferring resistance to only one group of QCC with particular chemical characteristics (Met21, Gly26, Phe27, Trp31, Ser105, Trp45, Leu7, Gly17, Tyr40, Ile62, Val66, Ser72, Leu93, & Leu103). These amino acids can participate in the unconventional binding site found within the larger binding pocket. The unconventional binding site happens as a result of helix tilting, conformational changes or other structural dynamics that result in exposing those amino acids to the binding pocket and thus allowing for the recognition/transport of the substrates. Some of these residues may be found on the same face of the other residues (see Section 5.2) and participate in interacting with one particular group of compounds (Met21, Leu7, Glly17, Ser72, Ile62, Val66, Ser105 and Tyr40) while other residues are present on a different face of the helix to which helix rotations may be required for the transport of the substrate (Leu93, Leu103, and Trp45). Also, residues, which are found on the loops, are involved in interacting with substrates of particular structural features include Trp31, Phe27, and Gly26. Thus, the specificity of these residues towards interacting with a group of compounds with particular feature suggests that loop #1 (Figure 1.3) helps in recognizing particular substrates and mediating substrate transport within the unconventional binding site.

Other residues like Ile11, Val15, and Pro32 were not included in the Venn diagram because cells expressing EmrE mutants of these residues had a resistant phenotype that was similar to the cells expressing the wild-type EmrE. Therefore, these residues don't play an important role in the functionality of EmrE that involves the extrusion of QCC thereby mutating these three residues did not affect the resistance phenotype of the cells expressing

them.

In summary, I speculated that EmrE has a large binding pocket that consists of a common binding site with residues that are important for interacting with substrates of different chemical characteristics and a unconventional binding site that contains residues, which only interact with compounds of a particular structural feature. In the subsequent section, the above speculation will further be supported by conclusions drawn from the advanced analysis of the data.

6.2.2 Conclusions from Advanced Data Analysis

When the data were reanalyzed using R-Statistical software, a different clustering of QCC has emerged based on the toxicity of these compounds to the different EmrE variants. As a result, a new Venn diagram is produced for the four groups of QCC (Figure 6.3).

As a general observation, there are fewer amino acid residues that are involved in recognizing only one group of QCC. Trp31, Tyr40, Trp45, Val66, Val69, Ser72, Leu93, Gly97, and Leu103 are the group of residues that are involved in conferring resistance to only group #4 of QCC (CET and DC). The residues that overlap in their substrate preferences between the two Venn diagrams are Tyr40, Ile62, Val66, Ser72, Leu93 and Leu103. These 6 residues are found to be important for conferring resistance to poly-charged QCC (Figure 6.2), specifically the compound, dequalinium (Figure 6.3). Also, Trp45 is found to have preference to recognizing acyl-chained compounds (Figure 6.2), particularly cetrimide as the new clustering analysis indicate. Gly97 is important for conferring resistance to group #4 of QCC which agrees with the initial clustering of QCC since some of the groups this residue is found to be important for recognizing are acyl-

chained and poly-charged QCC. On the other hand, residues like Val69 and Trp 31 showed different grouping from the initial data analysis. For example, Trp31 is found to be important in conferring resistance to sphere-forming compounds (i.e. TPP, MTP and TPA) while with the new clustering it is showing a preference to group #4 of QCC, which include a compound from the acyl-chained family and one from the poly-charged family. This can be attributed to the differences of the clustering methods of QCC used in both analysis and also the way the Venn diagram is generated. The Venn diagram is generated based on showing a 2-fold or more decrease in the MIC values of each EmrE variant when compared to the wild-type to more than 50% of the compounds in that group; hence it is a generalized organization.

The amino acid residues that showed preferences to more than one group of compounds are Ser64, Leu7, Ala10, Gly17, Tyr60, Leu47, Gly90, Met21, Gly26, Phe27, Phe44, Gly7, Pro55, Ala59, Trp76, Thr18, Trp63, Ile94, and Glu14. Referring to the speculation that was made earlier about EmrE's common binding site and the unconventional binding site, these residues are expected to participate in the common binding site since they recognize multiple groups of compounds. Out of 19 residues, 5 residues (Leu7, Gly17, Met21, Gly26, and Phe27) were expected to participate in the unconventional binding site since they had a preference to only one group of compounds based on the initial clustering (Figure 6.2). However, with the R-statistical clustering, these residues are found to participate in recognizing two groups of compounds (Figure 6.3). Residues like Leu7 and Gly17 have preference towards recognizing poly-charged substrates, which explains their preference to group #4 (particularly dequalinium). The second group of compounds these residues have preference to include group #2 (EB, PRO,

ACR, and MV). This can be explained from Table 4.1, where the mutations of these two residues displayed lower MIC than the wild-type to these four compounds. For Met21, Gly26, and Phe27, where they initially have been shown to only display substrate preference towards poly-aromatic QCC, the R-statistical clustering shows that they recognize group #3 and 4 substrates. The MIC values of the three EmrE variants (M21C, G26C, and F27C) were ≥ 2-fold lower than the wild-type EmrE for all of the compounds under group #3 and 4 (PY, CV, RH, DC, and CET).

The remaining 14 residues (Ser64, Ala10, Tyr60, Leu47, Gly90, Phe44, Gly7, Pro55, Ala59, Trp76, Thr18, Trp63, Ile94, and Glu14) have shown to be able to recognize compounds with different characteristics in both data analyses. From the above residues, Glu14, A10, Trp63 have been shown to participate in substrate binding and facilitate the substrate transport activity of EmrE (Yerushalmi and Schuldiner, 2000; Gutman et al., 2003; Elbaz et al., 2005; Sharoni et al., 2005). The remaining residues do not have a definite known function in the protein. Therefore, based on our results, Ser64, Tyr60, Leu47, Gly90, Phe44, Gly7, Pro55, Ala59, Trp76, Thr18, and Ile94 residues are expected to participate in the common binding site of the large binding pocket of EmrE. Each one of these residues is important for interacting with different characteristics of the substrates allowing EmrE to recognize all the compounds of this study.

It is important to note that the majority of the residues' substrate preferences as a result of clustering QCC according to their toxicity by R-Statistical program agree with the preferences seen when compounds were clustered based on structural differences (28 residues out of the 33). Therefore, the structural characteristics of the QCC influence the toxicity of these compounds on EmrE variants and substrate recognition. From both

analyses we can draw the conclusion that compounds with multiple charged atoms and compounds loaded with several aromatic rings are the most toxic compounds to cells expressing the EmrE variants. Also both analyses confirm the notion of a large binding pocket for EmrE with different niches, where different residues are responsible for recognizing compounds with particular feature(s).

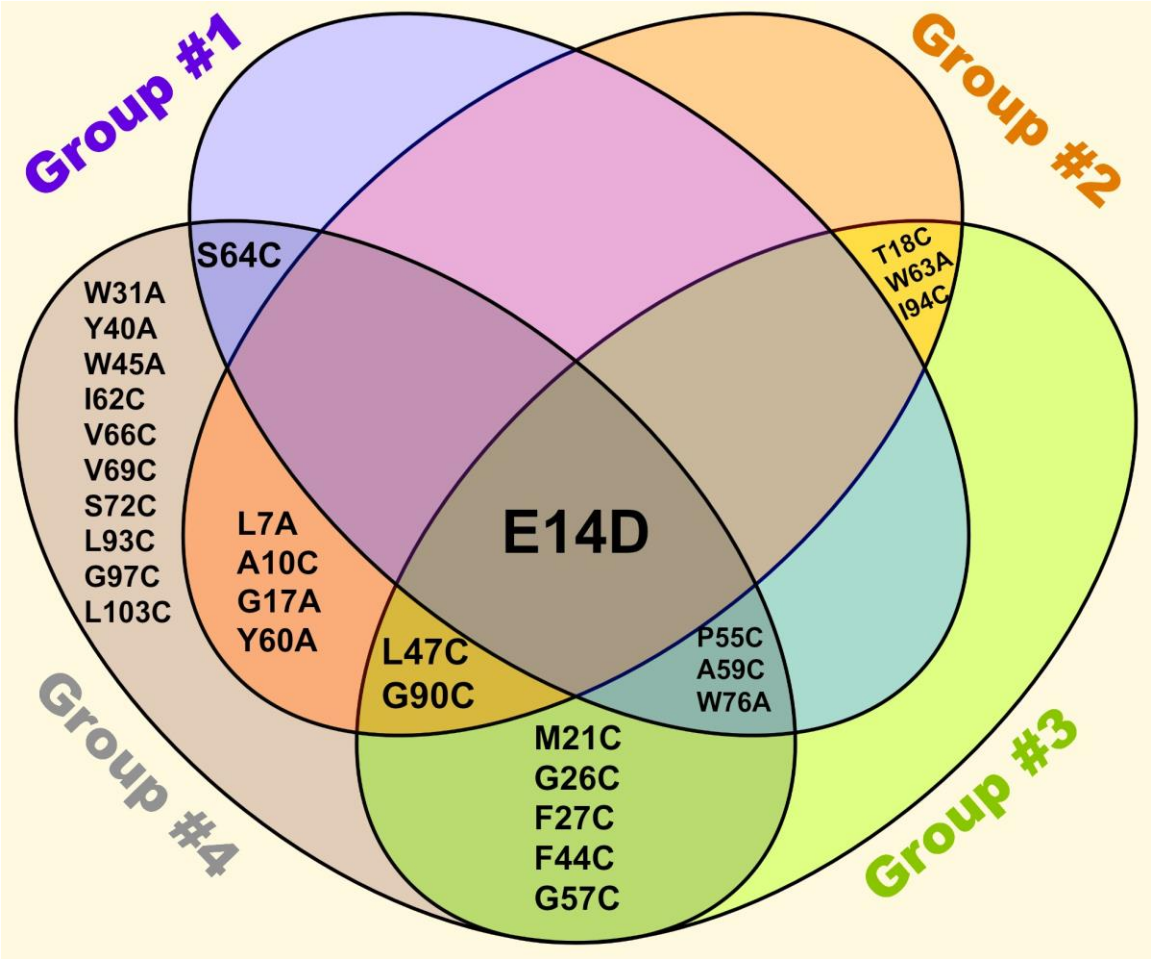


Figure 6. 3: Venn diagram illustrating that multiple amino acid residues are involved in explaining the poly-specific nature of EmrE.

Diagram depicts the sensitivities of EmrE variants to the toxicity of the 4 groups of QCC. Each EmrE variant is generated via site-directed mutagenesis altering single conserved amino acid in EmrE and each was tested against a variety of QCC. Each EmrE variant is written where the first letter represents the amino acid targeted for mutation and the second letter is the new amino acid. The selection of EmrE variant was based on which variant shows a ≥ 2-fold decrease in MIC when compared to wild-type EmrE and this decrease in

MIC must be observed in ≥ 50% of the compounds under particular QCC group. The four groups of QCC have been categorized based on ‘complete’ clustering analysis (See Section 5.1.2.1).

6.3 Future Initiatives

The conclusions of this thesis did not finish the story of the small multidrug transporter, EmrE. Instead, they brought several important questions that can be answered through the future work of motivated scientists. In this study, highly conserved amino acid residues were altered using site directed mutagenesis in order to assess the multidrug resistance activity of EmrE against 19 diverse cationic compounds. Out of the 33 residues tested, 30 were found to participate in substrate recognition and/or transport. Substrate recognition involves the requirement of particular amino acid side-chains that border and shape the binding pocket allowing the substrate to reside in it. However, substrate transport involves amino acid residues that directly bind the substrate and release it upon protein conformational changes. Therefore, in order to confirm if residues are involved in substrate recognition or substrate binding/release, certain *in vitro* transport assays can be performed (Gutman et al., 2003). Also, substrate-uptake kinetics can be evaluated using fluorescence spectroscopy-based transport assays (Son et al., 2003) or the radioactivity of C^{14}-labeled substrate (Gutman et al., 2003). Furthermore, select EmrE variants can be targeted for expression, purification and applied to *in vitro* protein ligand binding experiments using fluorescence, isothermal titration calorimetery, and native polyacrylamide gel electrophoresis methods (Bay et al., 2010; Sikora and Turner, 2005; Cilley and Williamson, 1997).

Furthermore, we can use heterooligomer experiments to confirm the participation of certain amino acid residues in substrate binding/transport rather than oligomerization (Sharoni et al., 2005). These studies involve expressing one wild-type protomer with another EmrE protomer that has an alteration in one of its conserved amino acids and

evaluating the binding affinity and transport activity of the dimer. If the multidrug resistance activity of the heterodimer is similar to the wild-type homodimer, then these substrates are not involved in the dimerization of the protein but involved in substrate transport. Also, if the binding affinity of the substrate is different from that of the wild-type yet can still bind the substrate, then that further confirms the involvement of that residue in the binding activity of EmrE.

Considering the data presented in Figure 5.2, there were residues that particularly recognized 1 group of substrates and found on a different face of the helix from the majority of the residues involved in substrate recognition. These residues include, Leu103 and Trp45 and their corresponding TMS may require helical rotaion or tilting in order to mediate subtrate transport. The chemical shifts of amino acid residues and helical tilting angle can be studied using NMR spectroscopy analysis using the 19 different substrates of this study. The NMR spectroscopy data of Morrison and Henzler-Wildman (2014) demonstrated that conformational change rates differ between planar and tetrahedral substrates. Their conclusion tested wild-type EmrE conformational changes as a result of interacting with TPP, ethidium, propidium, and dequalinium. Therefore, we can express and purify the different EmrE variants (as well as the wild-type protein) and test their conformational and helical adjustements as a result of interacting with the 19 different QCC of this study to confirm if particular residues induce helical rotaion/tilting as a result of interacting with substrates of specific features.

Another interesting observation indicates the significance of residues found in Loop #1(Figure 1.3), such as Gly26, Phe27, and Trp31, for substrate transport with particular features (Figure 6.2 & 6.3). Since loop #1 faces the periplasmic space, this leads to the

question of whether these residues are important for mediating the interaction between EmrE and a periplasmic protein or an outermembrane protein. Thus, mutations of such residues inhibit further efflux of the substrate to the outside of the cell. Since recent research has demonstrated the participation of OmpW in completing the efflux of EmrE substrates to the outside of the cell (Beketskaia et al., 2014), using the same assays it will be worthwhile to test the effect of mutating Gly26, Phe27, and Trp31 residues on rescuing the growth-phenotype of *E. coli*. If the mutation of any of the 3 residues resulted in the rescue of the growth-phenotype under alkaline conditions as a result of accumulation of osmoprotectant molecules that EmrE usually transports, these residues (loop #1) possibly participate in interacting with the outermembrane, OmpW. If no rescue of the growth-phenotype observed, then either the residues do not participate in intercting with OmpW to mediate the completion of the efflux process or the involvment of other periplasmic or outermembrane proteins in the process. Also, the specificity of the residues in loop #1 to particular substrates (Figure 6.2 & 6.3) may suggest the involvement of different periplasmic proteins or outermembrane proteins for the transfer of the different substrates if these residues were found to be important for EmrE-periplasmic/outermembrane protein interactions.

Studies that can futher confirm the presence of amino acid niches that participate in transporting specific substrates may involve double-mutant analysis. The double-mutant analysis would focus on altering two conserved amino acids at the same time, particularily the ones that participate in conferring resistance to only one group of QCC (Figure 6.1), and assess the resistance profiles of the variant generated agaist the 19 QCC. Preliminary data of a double-mutant EmrE generated by a project student (Joanna, 2014) showed

promising results. Gly17 is found to be important for recognizing poly-charged QCC while Ser105 is important for recognizing sphere-forming QCC. Therefore, upon the mutation of both residues, producing G17A/S105C EmrE variant, it was expected that the new variant would lose its resistance ability to both sphere-forming and poly-charged QCC. The double-mutant, however, showed a loss of resistance to poly-aromatic and poly-charged QCC when compared to wild-type EmrE (Table 6.1). If we compare the resistance profile of the double-mutant to the single-mutants, we notice that the loss of resistance of a single-mutant is reflected as a loss of resistance in the double mutant for 5 out of the 8 compounds tested (MV, DC, EB, CC, and BZ). This means that when Gly17, for example, was mutated, there was a loss of resistance to EB, MV, and DC which is also seen in the doubl-mutant. Therefore, Gly17 is critical for the recognition and transport of EB, MV, and DC substrates. Moreover, Ser105 is critical for conferring resistance particularly to DC and BZ since the double-mutant also experiences a loss of resistance to these substrates. On the other hand, they were certain subtrates' resistance profiles that gave unexpected results that require further investigation, such as CV, TPP and MTP. The double-mutant conferred resistance to MTP and TPP despite how the single mutants were sensitive to these compounds. Since this assay had only one replication, more replications are required to confirm the above observation. Regardless, the double-mutant suggest that the presence of an Ala at position 17 and a Cys at position 105 cancels out the effect of each mutant on it is own and restores resistance to sphere-forming QCC (TPP and MTP). Also, the double-mutant was senstive to CV, while the single mutants conferred resistance to CV. This suggests that at least one of the two residues (Gly17 or Ser105) are required for confering resistance to CV hence a double-mutation resultes in abolishing the resistance to CV. Therefore, the preliminary results presented in Table 6.1 can offer some insight on how double mutaions of conserved

amino acid residues affect the resistance profiles of EmrE and can further confirm the involvement of particular residues in creating niches within the binding pocket that are responsible for interacting with substrates of specific features.

Finally using the approach used in this study in addition to changing the amino acid substitutions from cysteine and alanine to more structurally and chemically divergent residues (i.e. Trp, Asp, or Phe) can help in producing useful qualitiative data on poly-substrate specificity of multidrug resitance transporters. These studies can provide an excellent glimpse on the organization of the binding pocket of these transporters and explain their plasticitiy. This knowledge can be a useful tool for the development of muldtidrug transporters' inhibitors.

Table 6. 1: Double-mutant resistance profile.

EmrE Variants	[1,2] QCC							
	Poly-aromatic		Acyl-chained		Sphere-forming		Poly-charged	
	[5] EB	CV	CC	BZ	TPP	MTP	DC	MV
G17A/S105C	32	4	8	8	32	120	62.5	250
G17A	32	8	16	16	16	120	8	250
S105C	256	8	8	8	16	62.5	16	500
[3] WT	64	8	8	16	32	120	500	500
[4] pMS119EH	32	2	4	8	16	60	32	250

[1] The minimum inhibitory concentrations (MIC) for each QCC in µg/mL

[2] The red colored boxes respresent ≥2-fold decrease in MIC while the green boxes represent MIC values that are similar to wild-type (WT) or higher

[3] WT is the wild-type EmrE-containing plasmid that has been transformed in *E. coli ΔacrB*

[4] pMS119EH is the empty vector that is transfromed in *E. coli ΔacrB*

[5] For QCC abbreviations, see Abbreviations list (page x)

Chapter 7: Bibliography

Achouak, W., Heulin, T., & Pagès, J. M. (2001). Multiple facets of bacterial porins. *FEMS Microbiology Letters*, *199*(1), 1–7.

Adler, J., & Bibi, E. (2004). Determinants of substrate recognition by the *Escherichia coli* multidrug transporter MdfA identified on both sides of the membrane. *The Journal of Biological Chemistry*, *279*(10), 8957–65.

Amadi, S. T., Koteiche, H. A., Mishra, S., & McHaourab, H. S. (2010). Structure, dynamics, and substrate-induced conformational changes of the multidrug transporter EmrE in liposomes. *The Journal of Biological Chemistry*, *285*(34), 26710–26718.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006-8.

Babbs, M., Collier, H., Austin, W., Potter, M., & EP, T. (1956). Salts of decamethylene-bis-4-aminoquinaldinium (dequadin); a new antimicrobial agent. *Journal of Pharmacy and Pharmacology*, *8*(2), 110–9.

Bairoch, A. (1992). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic acids Research*, *20 Suppl*(March 1992), 2013–8.

Barza, M., & Gorbach, S. (2002). The Need to Improve Antimicrobial Use in Agriculture. *Clinical Infectious Diseases*, *24*, 71–75.

Bay, D. C., Budiman, R. A., Nieh, M.P., & Turner, R. J. (2010). Multimeric forms of the small multidrug resistance protein EmrE in anionic detergent. *Biochemical and Biophysical Research Communications*, *1798*(3), 526–535.

Bay, D. C., Rommens, K. L., & Turner, R. J. (2008). Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochemical and Biophysical Research Communications*, *1778*(9), 1814–1838.

Bay, D. C., & Turner, R. J. (2009). Diversity and evolution of the small multidrug resistance protein family. *BMC Evolutionary Biology*, *9*, 140–67.

Bay, D. C., & Turner, R. J. (2012). Membrane composition influences the topology bias of bacterial integral membrane proteins. *Biochemical and Biophysical Research Communication*, *1828*(2), 260–70.

Becker, M. L., Visser, L. E., Schaik, R. H. N. Van, & Hofman, A. (2009). Genetic Variation in the Multidrug and Toxin Extrusion 1 Transporter Protein Influences the Glucose-Lowering Effect of Metformin in Patients With Diabetes: A Preliminary Study. *Diabetes*, *58*, 745–749.

Beketskaia, M. S., Bay, D. C., & Turner, R. J. (2014). Outer membrane protein OmpW participates with small multidrug resistance protein member EmrE in quaternary cationic compound efflux. *Journal of Bacteriology*, *196*(10), 1908–14.

Benizri, E., Ginouvès, a, & Berra, E. (2008). The magic of the hypoxia-signaling cascade. *Cellular and Molecular Life Sciences*, *65*(7-8), 1133–49.

Bradbury, F. R., & Linnell, W. H. (1942). Chemotherapeutic studies in the acridine series. Part VIII. The chloroaminoacridines. *Journal of Pharmacy and Pharmacology*, *15*, 31–40.

Brown, M. H., Paulsen, I. T., & Skurray, R. A. (1999). The multidrug efflux protein NormM is a prototype of a new family of transporters. *Molecular Microbiology*, *31*(1), 393–395.

Browning, C. H., & Gulbransen, R. (1921). The antiseptic potency of acriflavine, with considerations on the variability of results in testing antiseptics. *The British Journal of Experimental Pathology*, 2 (2), 95–103.

Bus, J. S., & Gibson, J. E. (1984). Paraquat: model for oxidant-initiated toxicity. *Environmental Health Perspectives*, *55*, 37–46.

Bush, K., Jacoby, G. A., & Medeiros, A. A. (1995). A Functional Classification Scheme for beta-Lactamases and Its Correlation with Molecular Structure. *Antimicrobial Agents and Chemotherapy*, *39*(6), 1211–1233.

Butler, M. I., Gray, J., Goodwin, T. J. D., & Poulter, R. T. M. (2006). The distribution and evolutionary history of the PRP8 intein. *BMC Evolutionary Biology*, *6*, 42.

Carson, R. T., Larson, E., Levy, S. B., Marshall, B. M., & Aiello, A. E. (2008). Use of antibacterial consumer products containing quaternary ammonium compounds and drug resistance in the community. *The Journal of Antimicrobial Chemotherapy*, *62*(5), 1160–1162.

Chang, G., & Roth, C. B. (2001). Structure of MsbA from *E. coli*: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science*, *293*(5536), 1793–800.

Chen, Y. J., Pornillos, O., Lieu, S., Ma, C., Chen, A. P., & Chang, G. (2007). X-ray structure of EmrE supports dual topology model. *Proceedings of The National Academy of Sciences of the United States of America*, *104*(48), 18999–19004.

Cho, M. K., Gayen, A., Banigan, J. R., Leninger, M., & Traaseth, N. J. (2014). Intrinsic Conformational Plasticity of Native EmrE Provides a Pathway for Multidrug Resistance. *Journal of the American Chemical Society*, *136*, 8072–80.

Chung, Y. J., & Saier, M. (2001). SMR-type multidrug resistance pumps. *Current Opinion in Drug Discovery and Development*, *4*(2), 237–45.

Cilley, C. D., & Williamson, J. R. (1997). Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA*, *3*(1), 57–67.

Colmer, J., Fralick, J., & Hamood, N. (1998). Isolation and characterization of a putative multidrug resistance pump from *Vibrio cholerae*. *Molecular Microbiology*, *27*(1), 63–72.

Cox, W. A. (1965). Site of Action of Certain Antibacterial Heterocyclic Quaternary Ammonium Compounds. *Applied and Environmental Microbiology*, *13*(6), 956–86.

Crebelli, R., Falcone, E., Aquilina, G., Carere, A., Paoletti, A., & Fabri, G. (1984). Mutagenicity studies in a tyre plant: *in vitro* activity of urine concentrates and rubber chemicals. *IARC Scientific Publication*, *59*, 289–95.

D'Auria, M. V., Riccio, R., Minale, E. U., Tanaka, J., & Higa, T. (1989). Isolation and structure elucidation of seven new polyhydroxylated sulfated sterols from the ophiuroid, *Ophiolepis superba*. *Journal of Organic Chemistry*, *54*(1), 234–239.

Darzynkiewicz, Z., Kapuscinski, J., Carter, S. P., Schmid, F. A., & Melamed, M. (1986). Cytostatic and Cytotoxic Properties of Pyronin Y: Relation to Mitochondrial Localization of the Dye and Its Interaction with RNA. *Cancer Research*, *46*, 5760–5766.

Dassa, E., Hofnung, M., Paulsen, I. T., & Saier, M. H. (1999). Structure, function, and evolution of bacterial ATP- binding cassette systems. *Molecular Microbiology*, *32*(4), 881–891.

Doi, Y., Shibata, N., Shibayama, K., Kurokawa, H., Yokoyama, K., Yagi, T., & Kamachi, K. (2002). Characterization of a Novel Plasmid-Mediated Cephalosporinase (CMY-9) and Its Genetic Environment in an *Escherichia coli* Clinical Isolate. *Antimicrobial Agents and Chemotherapy*, *46*(8), 2427–36.

Dunten, R. L., Sahin-Tóth, M., & Kaback, H. R. (1993). Cysteine scanning mutagenesis of putative helix XI in the lactose permease of *Escherichia coli*. *Biochemistry*, *32*(47), 12644–12650.

Dutta, S., Morrison, E. a, & Henzler-Wildman, K. (2014). EmrE dimerization depends on membrane environment. *Biochemical and Biophysical Research Communications*, *1838*(7), 1817–22.

Edgar, R., & Bibi, E. (1999). A single membrane-embedded negative charge is critical for recognizing positively charged drugs by the *Escherichia coli* multidrug resistance protein MdfA. *The EMBO Journal*, *18*(4), 822–32.

Edwards, R., & Turner, R. J. (1998). Alpha-periodicity analysis of small multidrug resistance (SMR) efflux transporters. *Biochemistry and Cell Biology*, *76*(5), 791–797.

Elbaz, Y., Steiner-Mordoch, S., Danieli, T., & Schuldiner, S. (2004). *In vitro* synthesis of fully functional EmrE, a multidrug transporter, and study of its oligomeric state. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(6), 1519–1524.

Elbaz, Y., Tayer, N., Steinfels, E., Steiner-mordoch, S., & Schuldiner, S. (2005). Substrate-Induced Tryptophan Fluorescence Changes in EmrE , the Smallest. *Biochemistry*, *44*, 7369–7377.

Fischer, E. (1894). Influence of configuration on the action of enzymes. *The Journal of the American Chemical Society in Berlin*, *27*(3), 2985–93.

Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G., & Ben-Tal, N. (2006). Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *Journal of Molecular Biology*, *364*(1), 54–67.

Fredell, D. L. (1994). Biological properties and applications of cationic surfactants. *Cationic Surfactants* (J. Cross a., Vol. 24, pp. 31–60). New York: Marcel Dekker, Inc.

Frillingos, S., Sahin-Tóth, M., Wu, J., & Kaback, H. R. (1998). Cys-scanning mutagenesis: a novel approach to structure function relationships in polytopic membrane proteins. *Federation of American Societies for Experimental Biology*, *12*(13), 1281–1299.

Fuentes, D. E., Navarro, C., Tantaleán, J. C., Araya, M., Saavedra, C. P., Pérez, J. M., Calderón, I., Youderian, P., Mora, G., & Vasques, C. (2005). The product of the *qacC* gene of *Staphylococcus epidermidis* mediates resistance to beta-lactam antibiotics in Gram-positive and Gram-negative bacteria. *Research in Microbiology*, *156*(4), 472–7.

Fürste, J. P., Pansegrau, W., Frank, R., Blöcker, H., Scholz, P., Bagdasarian, M., & Lanka, E. (1986). Molecular cloning of the plasmid RP4 primase region in a multi-host-range tacP expression vector. *Gene*, *48*(1), 119–31.

Gayen, A., Banigan, J. R., & Traaseth, N. J. (2013). Ligand-induced conformational changes of the multidrug resistance transporter EmrE probed by oriented solid-state NMR spectroscopy. *Angewandte Chemie (International ed. in English)*, *52*(39), 10321–4.

Gear, A. (1974). Rhodamine 6G: A potent inhibitor of mitochondrial oxidative phosphorylation. *The Journal of Biological Chemistry*, *249*, 3628–3637.

Gilbert, P., & Moore, L. E. (2005). Cationic antiseptics: diversity of action under a common epithet. *Journal of Applied Microbiology*, *99*(4), 703–15.

Godsey, M. H., Zheleznova Heldwein, E. E., & Brennan, R. G. (2002). Structural biology of bacterial multidrug resistance gene regulators. *The Journal of Biological Chemistry*, *277*(43), 40169–72.

Greenwood, D., & Slack, R. C. (1981). The antibacterial activity of hexamine (methenamine), hexamine hippurate and hexamine mandelate. *Infection*, *9*(5), 223–7.

Grinius, L., Dreguniene, G., Goldberg, E. B., Liao, C. H., & Projan, S. J. (1992). A staphylococcal multidrug resistance gene product is a member of a new protein family. *Plasmid*, *27*(2), 119–29.

Grinius, L. L., & Goldberg, E. B. (1994). Bacterial multidrug resistance is due to a single membrane protein which functions as a drug pump. *The Journal of Biological Chemistry*, *269*(47), 29998–30004.

Gutman, N., Steiner-Mordoch, S., & Schuldiner, S. (2003). An amino acid cluster around the essential Glu-14 is part of the substrate- and proton-binding domain of EmrE, a multidrug transporter from *Escherichia coli*. *The Journal of Biological Chemistry*, *278*(18), 16082–16087.

Hegstad, K., Langsrud, S., Lunestad, B. T., Scheie, A. A., Sunde, M., & Yazdankhah, S. P. (2010). Does the wide use of quaternary ammonium compounds enhance the selection and spread of antimicrobial resistance and thus threaten our health? *Microbial Drug Resistance*, *16*(2), 91–104.

Heir, E., Sundheim, G., & Holck, a L. (1999). Identification and characterization of quaternary ammonium compound resistant *staphylococc*i from the food industry. *International Journal of Food Microbiology*, *48*(3), 211–9.

Higgins, C. F. (2007). Multiple molecular mechanisms for multidrug resistance transporters. *Nature*, *446*(7137), 749–57.

Hugo, W. B., & Frier, M. (1969). Mode of action of the antibacterial compound dequalinium acetate. *Applied Microbiology*, *17*(1), 118–27.

Hyde, S., Emsley, P., Hartshorn, M. J., Mimmack, M. M., Gileadi, U., Pearce, S. R., Gallagher, M. P., Gill, D., Hubbard, R., Higgins, C. (1990). Structural model of ATP-binding proteing associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature*, *346*, 362–66.

Ilag, L. L., Ubarretxena-Belandia, I., Tate, C. G., & Robinson, C. V. (2004). Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. *Journal of the American Chemical Society*, *126*(44), 14362–3.

Kim, R. B. (2002). Drugs as P-glycoprotein substrates, inhibitors, and inducers. *Drug Metabolism Reviews*, *34*(1-2), 47–54.

Klammt, C., Lohr, F., Schafer, B., Haase, W., Dotsch, V., Ruterjans, H., Glaubitz, C., & Frank, B. (2004). High level cell-free expression and specific labeling of integral membrane proteins. *European Journal of Biochemistry*, *271*(3), 568–580.

Klyachko, K. A., Schuldiner, S., & Neyfakh, A. A. (1997). Mutations affecting substrate specificity of the *Bacillus subtilis* multidrug transporter Bmr. *Journal of Bacteriology*, *179*(7), 2189–95.

Korkhov, V. M., & Tate, C. G. (2008). Electron crystallography reveals plasticity within the drug binding site of the small multidrug transporter EmrE. *Journal of Molecular Biology*, *377*(4), 1094–103.

Koteiche, H. a, Reeves, M. D., & McHaourab, H. S. (2003). Structure of the substrate binding pocket of the multidrug transporter EmrE: site-directed spin labeling of transmembrane segment 1. *Biochemistry*, *42*(20), 6099–105.

Lee, K., Zhang, H., Qian, D. Z., Rey, S., Liu, J. O., & Semenza, G. L. (2009). Acriflavine inhibits HIF-1 dimerization, tumor growth, and vascularization. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(42), 17910–15.

Levy, S B. (2000). Antibiotic and antiseptic resistance: impact on public health. *The Pediatric Infectious Disease Journal*, *19*(10), 120–2.

Levy, Stuart B, & Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nature medicine*, *10*(12), 122–9.

Lewis, K. (1994). Multidrug resistance pumps in bacteria: variations on a theme. *Trends in Biochemical Sciences*, *19*(March), 119–123.

Li, X., Poole, K., & Nikaido, H. (2003). Contributions of MexAB-OprM and an EmrE Homolog to Intrinsic Resistance of *Pseudomonas aeruginosa* to Aminoglycosides and Dyes. *Antimicrobial Agents and Chemotherapy*, *47*(1), 27–35.

Li, X.Z., & Nikaido, H. (2009). Efflux mediated drug resistance in Bacteria: an update. *Drugs*, *69*(12), 1555–1623.

Lloris-Garcerá, P., Slusky, J. S. G., Seppälä, S., Prieb, M., Schäfer, L. V, & Von Heijne, G. (2013). *In vivo* Trp scanning of the small multidrug resistance protein EmrE confirms 3D structure models'. *Journal of Molecular Biology*, *425*(22), 4642–51.

Loo, T. W., Bartlett, M. C., & Clarke, D. M. (2003A). Methanethiosulfonate derivatives of rhodamine and verapamil activate human P-glycoprotein at different sites. *The Journal of Biological Chemistry*, *278*(50), 50136–41.

Loo, T. W., Bartlett, M. C., & Clarke, D. M. (2003B). Simultaneous binding of two different drugs in the binding pocket of the human multidrug resistance P-glycoprotein. *The Journal of Biological Chemistry*, *278*(41), 39706–10.

Lyon, B. R., & Skurray, R. O. N. (1987). Antimicrobial Resistance of *Staphylococcus aureus*: Genetic Basis. *Microbiology and Molecular Biology Reviews*, *51*(1), 88–136.

Ma, C., & Chang, G. (2004). Structure of the multidrug resistance efflux transporter EmrE from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 2852–2857.

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

Masaoka, Y., Ueno, Y., Morita, Y., & Kuroda, T. (2000). A Two-Component Multidrug Efflux Pump , EbrAB , in *Bacillus subtilis*. *Journal of Bacteriology*, *182*(8), 2307–12.

McDonnell, G., & Russell, D. (1999). Antiseptics and disinfectants: activity, action, and resistance. *Clinical Microbiology Reviews*, *12*(1), 147–79.

Mellon, M, Benbrook, C, & Benbrook, K. (2001, January). Estimates of antimicrobial abuse in livestock. *Union of Concerned Scientists*, (January), 1–14.

Milligan, G. W., & Sokol, L. M. (1980). A Two-Stage Clustering Algorithm with Robust Recovery Characteristics. *Educational and Psychological Measurement*, *40*, 755–59.

Mitchell, B., Paulsen, I. T., Brown, M. H., & Skurray, R. (1999). Bioenergetics of the Staphylococcal Multidrug Export Protein QacA: Identification of distinct binding sites for monovalent and divalent cations. *Journal of Biological Chemistry*, *274*(6), 3541–3548.

Mordoch, S. S., Granot, D., Schuldiner, S., Chem, J. B., & Lebendiker, M. (1999). Scanning Cysteine Accessibility of EmrE , an H + -coupled Multidrug Transporter from *Escherichia coli* , Reveals a Hydrophobic Pathway for Solutes. *The Journal of Biological Chemistry*, *274*, 19480–86.

Morimyo, M., Hongo, E., Hama-Inaba, H., & Machida, I. (1992). Cloning and characterization of the *mvrC* gene of Escherichia coli K-12 which confers resistance against methyl viologen toxicity. *Nucleic Acids Research*, *20*(12), 3159–65.

Morrison, E. & Henzler-Wildman, K. (2014). Transported Substrate Determines Exchange Rate in the Multidrug Resistance Transporter EmrE. *The Journal of Biological Chemistry*, *289*(10), 6825–36.

Morrison, K. L., & Weiss, G. (2001). Combinatorial alanine-scanning. *Current Opinion in Chemical Biology*, *5*(3), 302–307.

Murakami, S. (2008). Multidrug efflux transporter, AcrB-the pumping mechanism. *Current Opinion in Structural Biology*, *18*(4), 459–65.

Murray, D. S., Schumacher, M. & Brennan, R. G. (2004). Crystal structures of QacR-diamidine complexes reveal additional multidrug-binding modes and a novel mechanism of drug charge neutralization. *The Journal of Biological Chemistry*, *279*(14), 14365–71.

Muth, T. R., & Schuldiner, S. (2000). A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. *The European Molecular Biology Organization Journal*, *19*(2), 234–40.

Nakamura, H. (1968). Genetic determination of resistance to acriflavine, phenethyl alcohol, and sodium dodecyl sulfate in *Escherichia coli*. *Journal of Bacteriology*, *96*(4), 987–96.

Neyfakh, A. A., Bidnenko, V. E., & Chen, L. B. (1991). Efflux-mediated multidrug resistance in *Bacillus subtilis*: similarities and dissimilarities with the mammalian system. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(11), 4781–5.

Neyfakh, A. (2002). Mystery of multidrug transporters: the answer can be simple. *Molecular Microbiology*, *44*(5), 1123–30.

Nikaido, H. (1996). Multidrug Efflux Pumps of Gram-Negative Bacteria. *Journal of Bacteriology*, *178*(20), 5853–5859.

Nikaido, H. (2009). Multidrug resistance in bacteria. *Annual Review of Biochemistry*, *78*, 119–146.

Ninio, S., Rotem, D., & Schuldiner, S. (2001). Functional analysis of novel multidrug transporters from human pathogens. *The Journal of Biological Chemistry*, *276*(51), 48250–6.

O'Neil, K. T., & DeGrado, W. F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, *250*(4981), 646–51.

Paulsen, I. A. N. T., Littlejohn, T. I. M. G., Radstrom, P., Sundstrom, L., Skold, O. L. A., Swedberg, G., & Skurrayl, R. A. (1993). The 3 ' Conserved Segment of Integrons Contains a Gene Associated with Multidrug Resistance to Antiseptics and Disinfectants. *Antimicrobial Agents and Chemotherapy*, *37*(4), 761–770.

Paulsen, I. T., Brown, M. H., Littlejohn, T. G., Mitchell, B. a, & Skurray, R. a. (1996A). Multidrug resistance proteins QacA and QacB from *Staphylococcus aureus*: membrane topology and identification of residues involved in substrate specificity. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(8), 3630–5.

Paulsen, I. T., Brown, M. H., & Skurray, R. a. (1996B). Proton-dependent multidrug efflux systems. *Microbiological reviews*, *60*(4), 575–608.

Paulsen, I. T., Skurray, R., Tam, R., Saier, M. H., Turner, R. J., Weiner, J. H., Goldberg, E. B., Grinius, L. (1996C). The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. *Molecular microbiology*, *19*(6), 1167–1175.

Pornillos, O., Chen, Y. J., Chen, A. P., & Chang, G. (2005). X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science (New York, N.Y.)*, *310*(5756), 1950–3.

Purewal, S. (1991). Nucleotide sequence of the ethidium efflux gene from *Escherichia coli*. *Federation of European Microbiological Societies Microbiology Letters*, *66*(2), 229–231.

Putman, M., Koole, L. A., Veen, H. W. Van, & Konings, W. N. (1999). The Secondary Multidrug Transporter LmrP Contains Multiple Drug Interaction Sites. *Biochemistry*, *38*, 13900–13905.

Putman, M., Veen, H. W. Van, Konings, W. N., & Hendrik, W. (2000). Molecular Properties of Bacterial Multidrug Transporters Molecular Properties of Bacterial Multidrug Transporters. *Microbiology and Molecular Biology Reviews*, *64*(4), 672–95.

Rosenberg, M. F., Velarde, G., Ford, R. C., Martin, C., Berridge, G., Kerr, I. D., Callaghan, R., Schmidlin, A., Wooding, C., Linton, K., & Higgins, C. (2001). Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *The European Molecular Biology Organization Journal*, *20*(20), 5615–25.

Rotem, D, Sal-man, N., & Schuldiner, S. (2001). *In vitro* monomer swapping in EmrE, a multidrug transporter from *Escherichia coli*, reveals that the oligomer is the functional unit. *The Journal of Biological Chemistry*, *276*(51), 48243–9.

Rotem, D., Steiner-Mordoch, S., & Schuldiner, S. (2006). Identification of tyrosine residues critical for the function of an ion-coupled multidrug transporter. *The Journal of Biological Chemistry*, *281*(27), 18715–22.

Russell, A. D. (1983). *Principles of antimicrobial activity. In Disinfection, sterilization and preservation* (3rd Ed., pp. 717–745). Philadelphia: Lea & Febiger.

Saier, M. H. (2000). A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiology and Molecular Biology Reviews*, *64*(2), 345–401.

Saier, M. H., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R., & Nikaido, H. (1998). Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *Federation of American Societies for Experimental Biology*, *12*(3), 265–74.

Saier, M. H., Tarn, R., Reizer, A., & Reizer, J. (1994). Two novel families of bacterial membrane proteins concerned with nodulation , cell division and transport. *Molecular Microbiology*, *11*(5), 841–848.

Sasatsu, M., Shima, K., Shibata, Y., & Kono, M. (1989). Nucleotide sequence of a gene that encodes resistance to ethidium bromide from a transferable plasmid in *Staphylococcus aureus*. *Nucleic Acids Research*, *17*(23), 101-103.

Schuldiner, S, Granot, D., Steiner, S., Ninio, S., Rotem, D., Soskin, M., & Yerushalmi, H. (2001). Precious things come in little packages. *Journal of Molecular Microbiology and Biotechnology*, *3*(2), 155–62.

Schuldiner, S, Lebendiker, M., & Yerushalmi, H. (1997). EmrE, the smallest ion-coupled transporter, provides a unique paradigm for structure-function studies. *The Journal of Experimental Biology*, *200*(2), 335–341.

Schuldiner, Shimon. (2009). EmrE, a model for studying evolution and mechanism of ion-coupled transporters. *Biochemical and Biophysical Research Communication*, *1794*(5), 748–62.

Schumacher, M, Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. & Brennan, R. G. (2001). Structural mechanisms of QacR induction and multidrug recognition. *Science*, *294*(5549), 2158–63.

Schumacher, M., Miller, M. C., & Brennan, R. G. (2004). Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *The European Molecular Biology Organization Journal*, *23*(15), 2923–30.

Schwaiger, M., Lebendiker, M., Yerushalmi, H., Coles, M., Gröger, Schwarz, C., Schuldiner, S., & Kesller, H. (1998). NMR investigation of the multidrug transporter EmrE, an integral membrane protein. *European Journal of Biochemistry*, *254*(3), 610–9.

Seeger, M. A, Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., & Pos, K. M. (2006). Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. *Science*, *313*(5791), 1295–8.

Sharoni, M., Steiner-Mordoch, S., & Schuldiner, S. (2005). Exploring the binding domain of EmrE, the smallest multidrug transporter. *The Journal of Biological Chemistry*, *280*(38), 32849–32855.

Sidhu, M. S., Heir, E., Sørum, H., & Holck, A. (2001). Genetic linkage between resistance to quaternary ammonium compounds and beta-lactam antibiotics in food-related *Staphylococcus* spp. *Microbial Drug Resistance*, *7*(4), 363–71.

Sikora, C. W., & Turner, R. J. (2005). Investigation of ligand binding to the multidrug resistance protein EmrE by isothermal titration calorimetry. *Biophysical Journal*, *88*(1), 475–482.

Silver, S., Levine, & E., Philip, (1968). Acridine Binding by *Escherichia coli*: pH Dependency and Strain. *Journal of Bacteriology*, *95*(2), 333–341.

Son, M. S., Del Castilho, C., Duncalf, K. a, Carney, D., Weiner, J. H., & Turner, R. J. (2003). Mutagenesis of SugE, a small multidrug resistance protein. *Biochemical and Biophysical Research Communications*, *312*(4), 914–921.

Soskine, M., Adam, Y., & Schuldiner, S. (2004). Direct evidence for substrate-induced proton release in detergent-solubilized EmrE, a multidrug transporter. *The Journal of Biological Chemistry*, *279*(11), 9951–5.

Soskine, M., Steiner-Mordoch, S., & Schuldiner, S. (2002). Crosslinking of membrane-embedded cysteines reveals contact points in the EmrE oligomer. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(19), 12043–12048.

Spratt, B. G. (1994). Resistance to antibiotics mediated by target alterations. *Science*, *264*(5157), 388–93.

Starks, C. M. (1971). Phase-transfer catalysis: Heterogeneous reactions involving anion transfer by quaternary ammonium and phosphonium salts. *Journal of American Chemical Society*, *93*(1), 195–9.

Tamai, I., & Safa, R. (1991). Azidopine noncompetitively interacts with vinblastine and cyclosporin A binding to P-glycoprotein in multidrug resistant cells. *The Journal of Biological Chemistry*, *266*(25), 16796–800.

Tate, C G, Kunji, E. R., Lebendiker, M., & Schuldiner, S. (2001). The projection structure of EmrE, a proton-linked multidrug transporter from *Escherichia coli*, at 7°A resolution. *The European Molecular Biology Organization Journal*, *20*(1-2), 77–81.

Tate, Christopher G, Ubarretxena-Belandia, I., & Baldwin, J. M. (2003). Conformational Changes in the Multidrug Transporter EmrE Associated with Substrate Binding. *Journal of Molecular Biology*, *332*(1), 229–242.

Taylor, R. G., Walker, D. C., & Mclnnes, R. R. (1993). *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research*, *21*(7), 1677–1678.

Torres, J. & Arkin, I. T. (2000). Recursive use of evolutionary conservation data in molecular modeling of membrane proteins: A model of the multidrug H+ antiporter *emrE*. *European Journal of Biochemistry*, *267*(12), 3422–31.

Toyoshima, C., Nakasako, M., Nomura, H., & Ogawa, H. (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. *Nature*, *405*(6787), 647–55.

Tsuda, M., Terada, T., Ueba, M., Sato, T., Masuda, S., Katsura, T., & Inui, K. (2009). Involvement of Human Multidrug and Toxin Extrusion 1 in the Drug Interaction between Cimetidine and Metformin in Renal Epithelial Cells. *The Journal of Pharmacology and Experimental Therapeutics*, *329*(1), 185–191.

Ubarretxena-Belandia, I, & Tate, C. G. (2004). New insights into the structure and oligomeric state of the bacterial multidrug transporter EmrE: an unusual asymmetric homo-dimer. *Federation of European Biochemical Societies Letters*, *564*(3), 234–8.

Ubarretxena-Belandia, Iban, Baldwin, J. M., Schuldiner, S., & Tate, C. G. (2003). Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *The European Molecular Biology Organization Journal*, *22*(23), 6175–81.

UK Food Standards Agency. (2013). Food Additives Legislation Guidance to Compliance November 2013. Retrieved June 14, 2014, from: http://www.food.gov.uk/science/additives/enumberlist#.U6-0xhbQdbw

Ulmschneider, M. B., & Sansom, M. S. (2001). Amino acid distributions in integral membrane protein structures. *Biochemical and Biophysical Research Communication*, *1512*(1), 1–14.

Van-Veen, H. W., Callaghan, R., Soceneantu, L., Sardini, A., Konings, W., & Higgins, C. (1998). A bacterial antibiotic-resistance gene that complements the human multidrug resistance gene P-glycoprotein gene. *Nature*, *391*, 291–295.

Vázquez-Ibar, J. L., Guan, L., Svrakic, M., & Kaback, H. R. (2003). Exploiting luminescence spectroscopy to elucidate the interaction between sugar and a tryptophan residue in the lactose permease of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(22), 12706–11.

Wakelin, L., Adams, A., Hunter, C., & Waring, M. (1981). Interaction of crystal violet with nucleic acids. *Biochemistry*, *20*(20), 5779–87.

Waring, M. (1964). Complex formation with DNA and inhibition of *Escherichia coli* RNA polymerase by ethidium bromide. *Biochemical and Biophysical Research Communications*, *87*, 358–61.

Weisblum, B. (1995). Insights into erythromycin action from studies of its activity as inducer of resistance. *Antimicrobial Agents and Chemotherapy*, *39*(4), 797–807.

Winstone, T. L., Jidenko, M., Le Maire, M., Ebel, C., Duncalf, K. & Turner, R. J. (2005). Organic solvent extracted EmrE solubilized in dodecyl maltoside is monomeric and binds drug ligand. *Biochemical and Biophysical Research Communications*, *327*(2), 437–45.

Yerushalmi, H., Lebendiker, M., & Schuldiner, S. (1996). Negative Dominance Studies Demonstrate the Oligomeric Structure of EmrE, a Multidrug Antiporter from *Escherichia coli*. *The Journal of Biological Chemistry*, *271*(49), 31044–31048.

Yerushalmi, H., Mordoch, S. S., & Schuldiner, S. (2001). A single carboxyl mutant of the multidrug transporter EmrE is fully functional. *The Journal of Biological Chemistry*, *276*(16), 12744–8.

Yerushalmi, H., & Schuldiner, S. (2000). An Essential Glutamyl Residue in EmrE, a Multidrug Antiporter from *Escherichia coli*. *The Journal of Biological Chemistry*, *275*(8), 5264–5269.

Yerushalmi, H. & Mario, & Lebendiker, S. S. (1995). EmrE, an *Escherichia coli* 12-kDa Multidrug Tranporter, Exchanges Toxic Cations and H+ and is Soluble in Organic Solvents. *The Journal of Biological Chemistry*, *270*(12), 6856–6863.

Zepeda-Mendoza, M. L., & Resendis-Antonio, O. (2013). Hierarchical Agglomerative Clustering. *Encyclopedia of Systems Biology*. Springer New York.

Zhang, Z., Ma, C., Pornillos, O., Xiu, X., Chang, G., & Saier, M. H. (2007). Functional characterization of the heterooligomeric EbrAB multidrug efflux transporter of *Bacillus subtilis*. *Biochemistry*, *46*(17), 5218–25.

Zheleznova, E. E., Markham, P. N., Neyfakh, A. A., & Brennan, R. G. (1999). Structural basis of multidrug recognition by BmrR, a transcription activator of a multidrug transporter. *Cell*, *96*(3), 353–62.

Zheng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, *32*(14), e115.

Appendix A: Scaled Data

MIC values were scaled according to the following formula:

$$\textbf{Scaled value} = \frac{(MIC - MIC_{min})}{MIC_{max}}$$

The MIC_{min} is the lowest MIC value, which is found under one particular QCC in Table 4.1, is subtracted from the MIC of each variant and divided by MIC_{max}, which is the highest MIC value found under that same particular QCC in Table 4.1, giving the product as a scaled value from 0-1. Therefore, the coloring scheme of the heat-map presented in Table 4.2 is based on the scaled values of MIC from 0-1.

Appendix B: Table of Minimum Inhibitory Concentrations of EmrE variants in µg/mL for 1/100 cultures.

[1,2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
1. L7A	16	32	8	8	2	32	16	128	16	8	256	8	16	256	8	8	64	8
2. A10C	16	32	4	8	2	16	8	64	16	8	8	8	16	256	4	4	4	1
3. I11C	64	64	4	8	4	32	16	128	16	16	512	16	16	512	8	8	128	8
4. E14D	8	16	2	8	2	16	8	64	8	8	4	8	16	256	4	4	64	4
5. V15C	128	128	4	8	4	64	16	128	16	16	512	16	8	512	8	8	128	16
6. G17A	128	32	8	8	8	32	16	128	8	8	256	8	16	125	4	8	128	16
7. T18C	64	64	2	16	0.5	32	16	128	16	16	64	16	16	256	8	8	64	16
8. M21C	64	64	2	4	4	32	16	128	16	16	512	16	8	512	4	8	64	4
9. G26C	64	128	2	8	1	32	16	128	16	16	512	16	16	512	8	8	128	8
10. F27C	64	64	4	8	1	32	16	128	16	16	512	16	16	512	8	8	64	16
11. W31A	16	64	8	8	8	32	8	64	16	8	256	8	16	256	8	8	64	8
12. P32C	128	128	4	16	4	64	16	128	16	16	125	16	16	512	8	8	64	8
13. Y40A	64	32	4	8	8	32	16	128	8	8	256	8	16	256	8	8	65	16
14. F44C	32	128	4	4	1	32	16	64	16	16	512	8	8	512	4	4	64	4
15. W45A	32	64	4	8	8	32	8	128	16	8	512	4	16	125	4	4	64	4
16. L47C	16	64	2	8	0.5	32	32	128	8	16	256	8	16	512	8	8	64	8

[1,2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
17. P55C	32	64	4	4	2	32	16	128	16	16	512	16	16	512	4	8	128	4
18. G57C	64	128	4	8	1	32	16	128	16	16	32	8	8	512	8	8	64	4
19. A59C	32	64	8	16	4	32	8	64	8	8	256	8	16	256	8	8	64	4
20. Y60A	16	32	8	8	4	32	16	0	16	8	256	8	16	256	8	4	64	4
21. I62C	32	64	8	8	4	64	16	128	8	16	256	8	16	256	8	8	64	8
22. W63A	16	16	8	8	2	8	4	0	4	512	64	8	8	256	8	8	64	8
23. S64C	32	64	4	8	8	32	8	64	16	8	128	8	16	256	8	8	64	8
24. V66C	64	64	4	8	16	64	16	128	8	256	8	8	16	256	8	8	9	128
25. V69C	128	64	4	8	8	64	8	64	32	8	256	8	16	256	8	8	64	8
26. S72C	64	64	4	8	4	64	8	64	32	8	512	8	8	256	8	8	64	8
27. W76A	64	64	2	4	4	32	4	64	4	8	256	4	16	256	8	8	128	16
28. G90C	64	64	2	16	1	32	16	128	8	16	256	16	16	512	8	8	128	8
29. L93C	128	64	8	4	8	64	8	128	8	8	512	8	16	256	8	8	128	8
30. I94C	16	32	8	8	1	16	8	0	8	256	128	8	16	256	8	4	64	16
31. G97C	16	32	8	16	1	32	16	64	32	8	128	4	16	256	8	8	64	4
32. L103C	32	64	8	8	4	32	16	64	8	16	256	8	16	256	8	8	64	8
33. S105C	32	64	4	8	2	32	16	64	16	8	256	8	16	256	4	4	37.5	4

[1,2] EmrE Variants	[4]ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
[3] pMS119EH	16	8	2	4	2	16	4	64	16	8	128	8	8	256	4	4	64	4
[3] pEmrE(WT)	64	32	8	16	8	32	8	128	32	256	256	16	16	256	8	8	128	8

[1] The columns showing the 33 different plasmids with *emrE* single codon mutants transformed in *E. coli* Δ*acrB,* called EmrE variants.

[2] The culture of each EmrE variant had an optical density value that was standardized to 1.5 and then diluted to 1/100. Comparable MIC values were observed with 1/10 culture (see Table 4.1).

[3] pMS119EH empty vector transformed in *E. coli* Δ*acrB,* representing the negative control while the positive control is *E. coli* Δ*acrB* expressing wild-type EmrE (pEMR11).

[4] The QCC abbreviations are: ACR= Acriflavine, PR= Proflavine, CV= Crystal violet, RH= Rhodamine, PY= Pyronin, EB=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.C= Cetalkonium chloride, STAC= Stearyltrimethylammonium chloride, and CET= Cetrimide. Also, the grouping of the QCC followed this coloring scheme: Poly-aromatic QCC, Sphere-forming QCC, Poly-charged QCC, Acyl-chained QCC.

Appendix C: Advanced Analysis

The advanced analysis was obtained using R-statistical software. The heat-maps that were generated by the program used the scaled data in Table C. Five types of clustering was used on the scaled data to extract additional trends. The ‘Complete’ clustering is presented and discussed in Chapter 5. The figures below represent the other four types of clustering.

Table C: Scaled Minimum Inhibitory Concentration values for EmrE variants

[1,2] EmrE Variants	ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
L7A	2	3	4	3	2	3	3	3	2	1	2	2	3	3	3	3	2	3
A10C	1	2	3	3	1	3	3	2	3	1	1	3	3	2	3	3	1	1
I11C	4	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
E14D	1	1	2	3	2	2	2	2	2	1	1	2	3	2	2	2	2	1
V15C	4	4	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3
G17A	4	2	4	3	4	3	3	3	2	1	1	3	3	2	4	3	3	3
T18C	3	2	1	4	2	2	3	3	3	3	1	3	3	2	4	3	2	3
M21C	3	3	1	1	2	2	3	3	3	1	3	3	3	3	3	3	3	2
G26C	3	3	1	1	2	2	3	3	3	1	3	3	3	3	3	2	3	2
F27C	3	3	1	1	2	3	2	3	3	1	3	3	3	2	3	2	3	2
W31A	3	2	4	2	4	4	2	3	2	1	3	3	3	2	3	3	2	2
P32C	4	4	3	3	4	3	3	3	3	3	3	3	3	3	3	3	3	3
Y40A	4	2	2	3	3	4	3	3	2	1	3	3	4	2	3	3	3	3
F44C	3	3	1	1	1	3	3	2	3	1	2	2	3	3	3	2	3	1
W45A	3	3	3	3	4	4	3	3	2	1	3	2	3	2	3	3	2	2
L47C	1	2	1	2	1	2	3	3	3	1	1	2	3	3	3	2	3	1
P55C	3	3	1	1	1	3	2	2	2	3	3	3	3	3	2	2	3	1

[1, 2] EmrE Variants	ACR	PRO	CV	RH	PY	EB	TPA	MTP	TPP	DC	MV	BZ	MC	STAC	CC	CB	Ct.C	CET
G57C	3	3	1	2	1	3	3	3	3	3	1	3	3	3	3	3	2	2
A59C	2	3	2	3	2	4	3	2	2	1	3	2	3	3	3	3	2	2
Y60A	2	2	3	3	3	3	3	3	2	1	2	2	3	2	4	3	2	2
I62C	3	3	4	3	2	4	3	3	1	1	2	3	3	3	3	3	2	2
W63A	1	1	3	2	1	2	2	4	1	3	1	3	3	3	4	4	3	3
S64C	3	3	3	3	3	4	2	2	2	1	2	2	3	3	3	3	2	2
V66C	4	3	3	3	4	4	3	3	2	1	2	3	3	2	3	3	3	2
V69C	4	3	3	3	4	4	2	2	3	1	3	2	3	3	3	3	3	2
S72C	3	3	3	3	4	4	3	3	3	1	3	2	3	2	3	3	2	3
W76A	3	2	2	2	3	3	2	2	1	1	3	1	3	2	3	3	3	3
G90C	3	2	1	2	1	3	3	3	3	1	2	3	2	3	3	3	3	2
L93C	4	3	3	3	4	4	3	3	2	1	3	2	3	2	3	3	3	2
I94C	2	2	3	2	1	2	3	3	1	3	2	2	3	3	3	3	2	3
G97C	1	2	3	3	1	0	3	2	3	1	2	2	2	3	3	3	2	1
L103C	3	3	3	3	3	4	3	3	2	1	3	2	4	3	3	3	2	2
S105C	3	3	3	3	2	4	3	2	2	1	3	2	3	3	3	3	2	1
pMS119EH	1	1	1	1	1	2	1	2	2	1	2	2	2	3	2	2	2	1
pEmrE(WT)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

[1] The columns showing the 33 different plasmids with *emrE* single codon mutants transformed in *E. coli* Δ*acrB,* called EmrE variants. The culture of each EmrE variant had an optical density value that was standardized to 1.5 and then diluted to 1/100.

[2] The scaled values represent the following scheme, where wild-type MIC values were given number 3, number 2 is for 2-fold decrease in MIC, number 1 is more than 2-fold decrease in MIC, and finally number 4 is given for 2-fold or more increase in MIC.

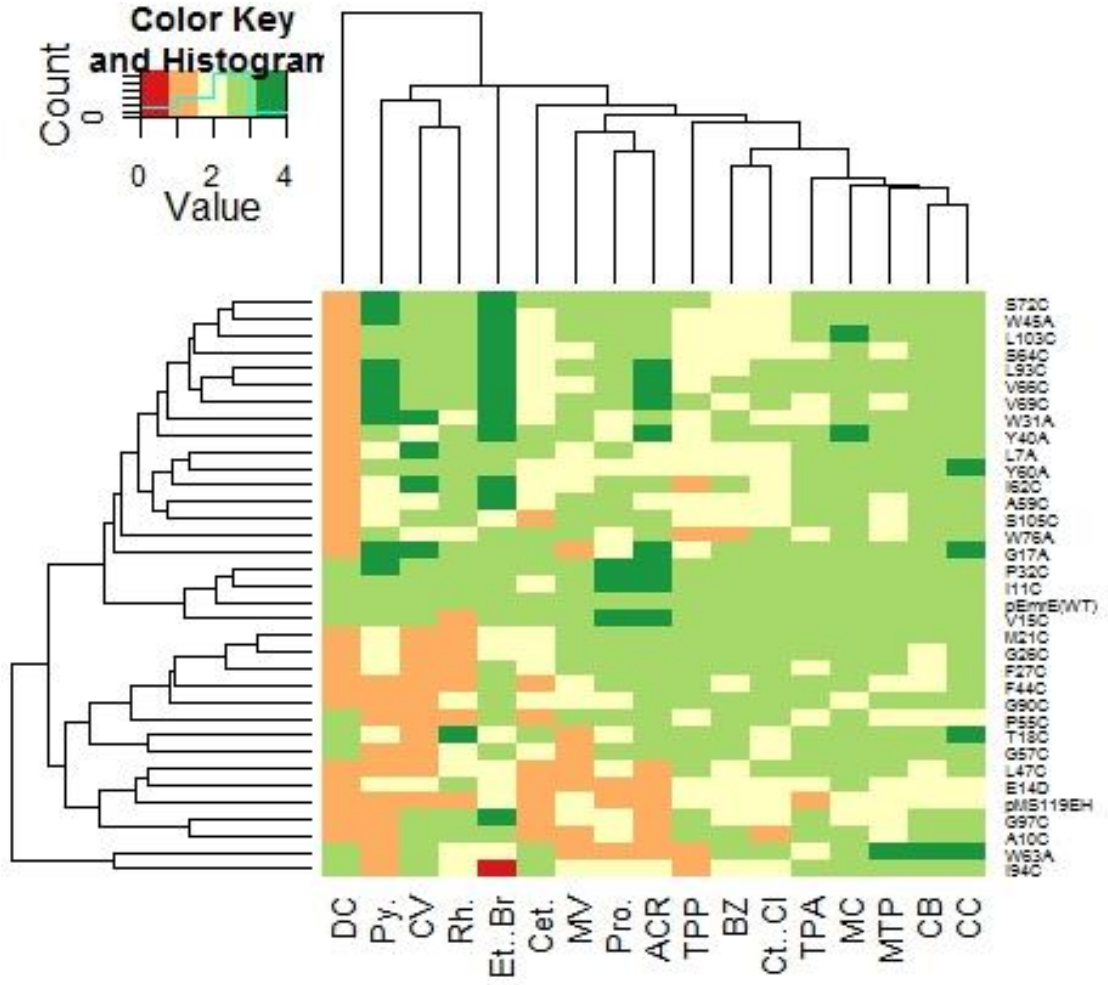


Figure C.1: The 'average' linkage clustering of EmrE variants' resistance profiles to 17 QCC. The heat map was generated using hierarchical agglomerative clustering plots, which used R Statistics software (See Sections 2.F and 5.A.i for details). In the 'average' linkage, the similarity of an observation to a cluster is defined by the mean similarity of the plot to all the members of the cluster. The bottom of the figure displays the different 17 QCC used in this study (ACR= Acriflavine, Pro= Proflavine, CV= Crystal violet, Rh.= Rhodamine, Py.= Pyronin, Et.Br=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.Cl= Cetalkonium chloride, and Cet.= Cetrimide). The right side of the diagram lists all the EmrE variants in addition to the two controls (*E. coli acrB* with wild-type EmrE protein and *E. coli acrB* with empty plasmid pMS119EH). The colored legend represents the values of scaled-MIC found in Table C.

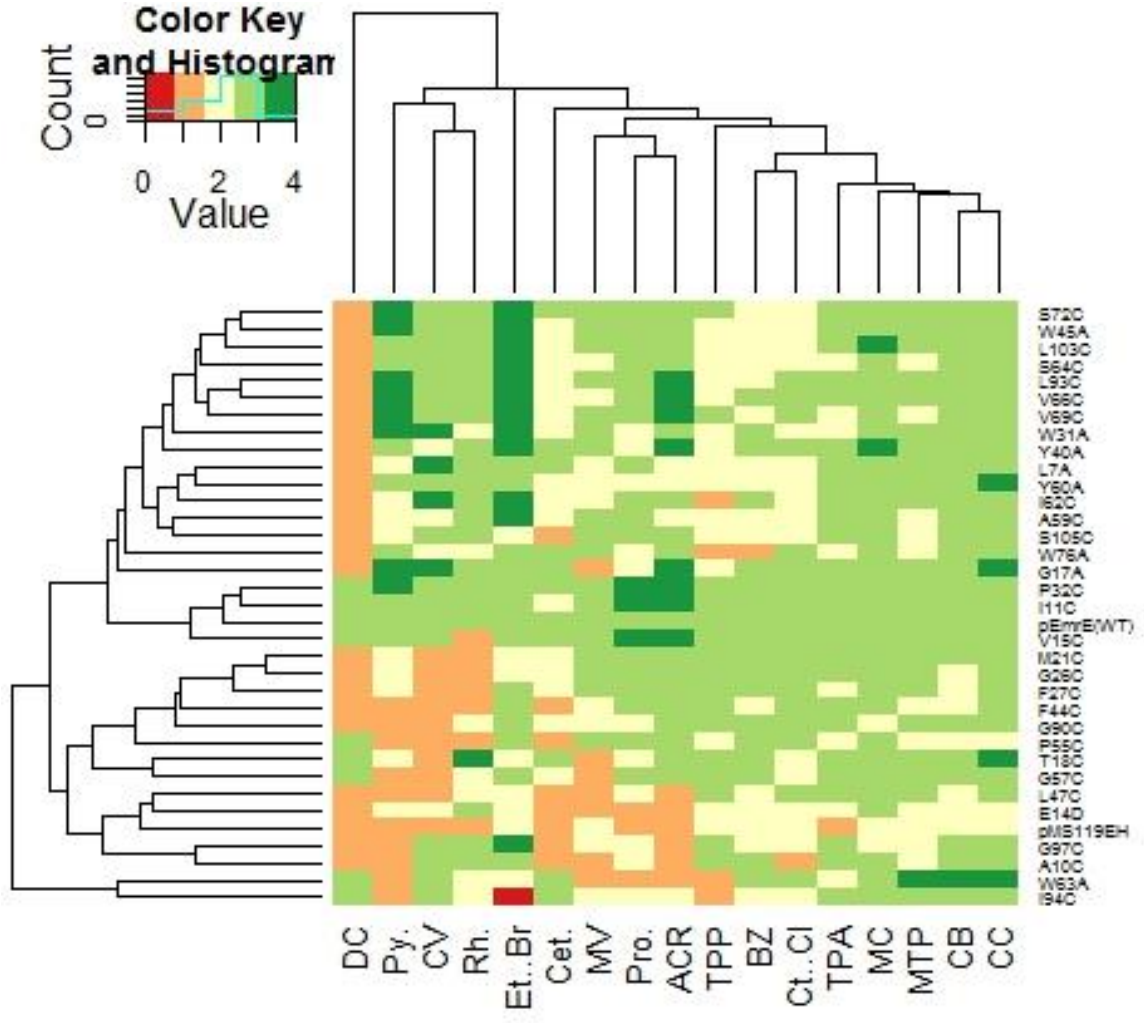


Figure C.2: The 'centroid' linkage clustering of EmrE variants' resistance profiles to 17 QCC. The heat map was generated using hierarchical agglomerative clustering plots, which used R Statistics software (See Sections 2.F and 5.A.i for details). In the 'centroid' linkage, the distance between two clusters is defined as the (squared) Euclidean distance between their centroids or mean. The bottom of the figure displays the different 17 QCC used in this study (ACR= Acriflavine, Pro= Proflavine, CV= Crystal violet, Rh.= Rhodamine, Py.= Pyronin, Et.Br=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.Cl= Cetalkonium chloride, and Cet.= Cetrimide). The right side of the diagram lists all the EmrE variants in addition to the two controls (*E. coli acrB* with wild-type EmrE protein and *E. coli acrB* with empty plasmid pMS119EH). The colored legend represents the values of scaled-MIC found in Table C.

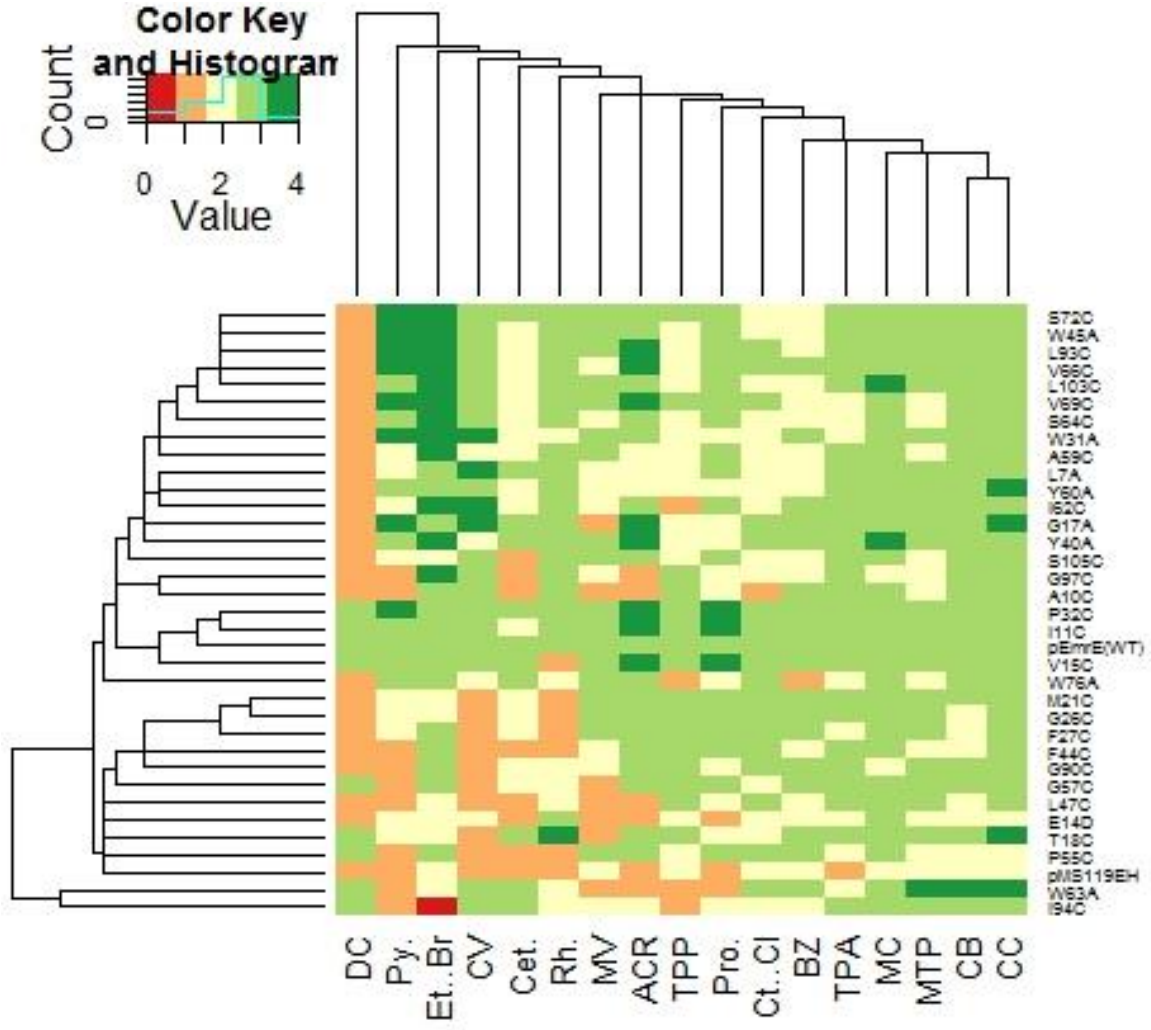


Figure C.3: The 'single' linkage clustering of EmrE variants' resistance profiles to 17 QCC. The heat map was generated using hierarchical agglomerative clustering plots, which used R Statistics software (See Sections 2.F and 5.A.i for details). The name "single linkage" arises because the observation only needs to be similar to a single member of the cluster to join. The bottom of the figure displays the different 17 QCC used in this study (ACR= Acriflavine, Pro= Proflavine, CV= Crystal violet, Rh.= Rhodamine, Py.= Pyronin, Et.Br=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.Cl= Cetalkonium chloride, and Cet.= Cetrimide). The right side of the diagram lists all the EmrE variants in addition to the two controls (*E. coli acrB* with wild-type EmrE protein and *E. coli acrB* with empty plasmid pMS119EH). The colored legend represents the values of scaled-MIC found in Table C.

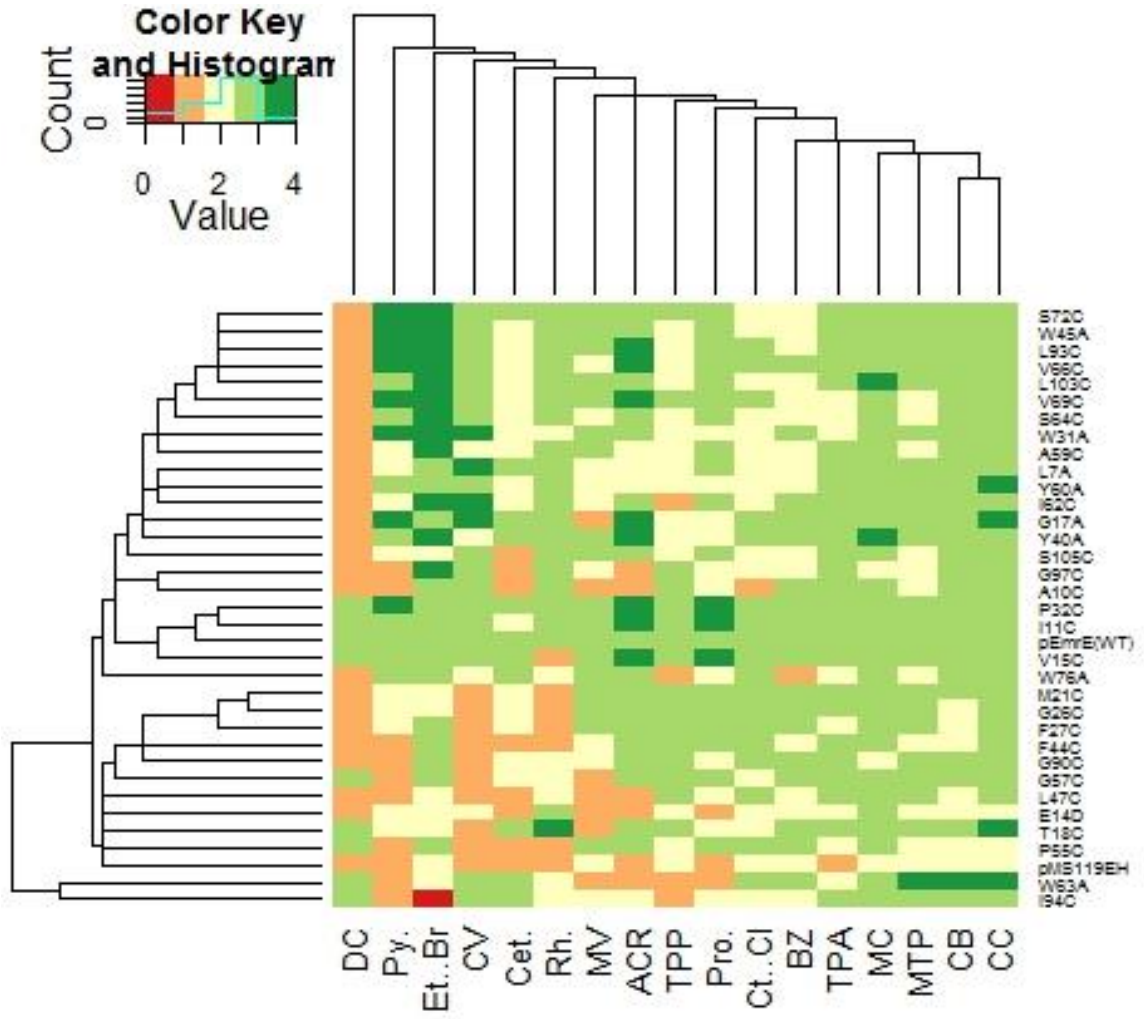
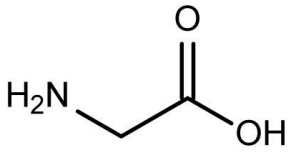
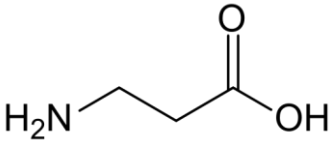
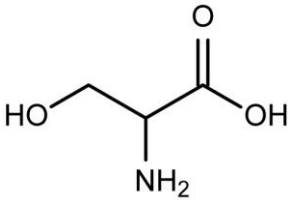
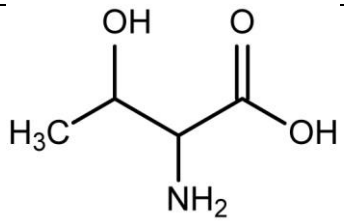
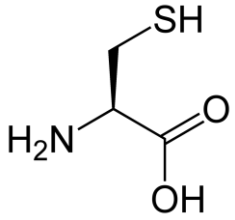
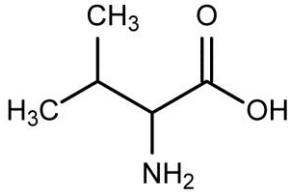
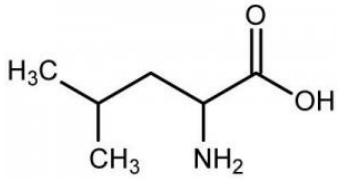
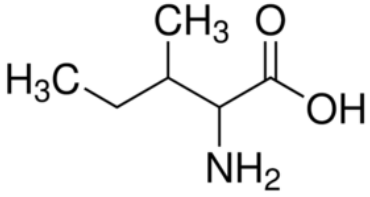
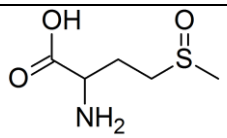
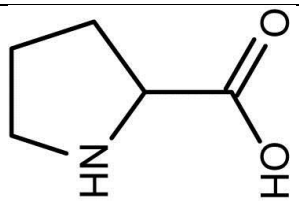
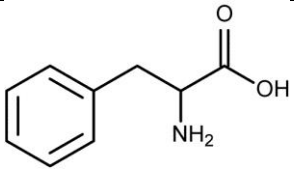
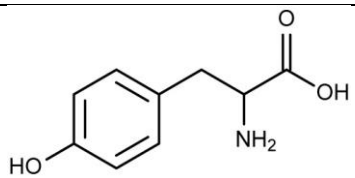
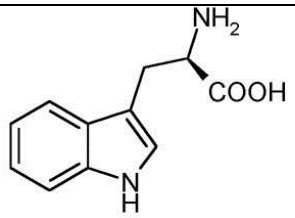
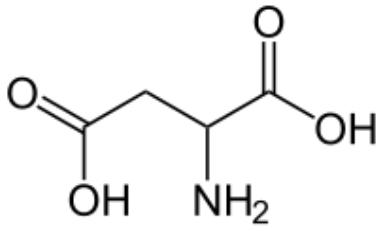
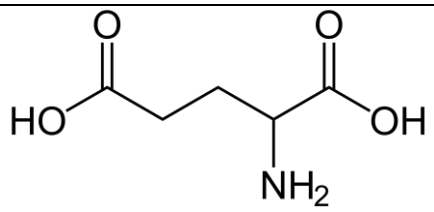
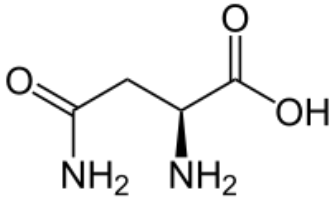
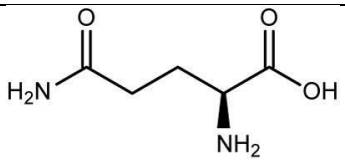
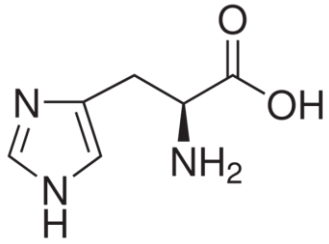
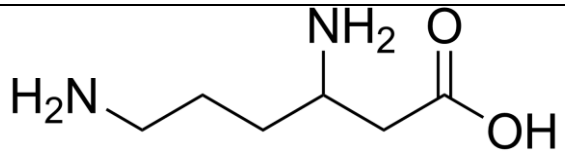
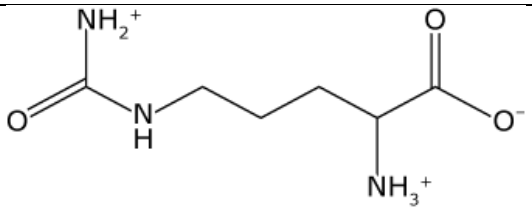


Figure C.4: The 'Ward' linkage clustering of EmrE variants' resistance profiles to 17 QCC. The heat map was generated using hierarchical agglomerative clustering plots, which used R Statistics software (See Sections 2.F and 5.A.i for details). In Ward linkage method, the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all the variables. At each step the pair of clusters with the minimum between-cluster distances are fused. The bottom of the figure displays the different 17 QCC used in this study (ACR= Acriflavine, Pro= Proflavine, CV= Crystal violet, Rh.= Rhodamine, Py.= Pyronin, Et.Br=ethidium bromide, TPA= Tetraphenylarsonium chloride, MTP= Methyltriphenyl phosphonium bromide, TPP= Tetraphenylphosphonium chloride, DC= Dequalinium chloride, MV= Methyl viologen, BZ= Benzalkonium chloride, MC= Myristalkonium chloride, CC= Cetylpyridinium chloride, CB= Cetylpyridinium bromide, Ct.Cl= Cetalkonium chloride, and Cet.= Cetrimide). The right side of the diagram lists all the EmrE variants in addition to the two controls (*E. coli acrB* with wild-type EmrE protein and *E. coli acrB* with empty plasmid pMS119EH). The colored legend represents the values of scaled-MIC found in Table C.

Appendix D: Amino Acid Structures and Abbreviations

Amino Acid	Abbreviation	Structure
Glycine	Gly, G	
Alanine	Ala, A	
Serine	Ser, S	
Threonine	Thr, T	
Cysteine	Cys, C	
Valine	Val, V	

Amino Acid	Abbreviation	Structure
Leucine	Leu, L	
Isoleucine	Ile, I	
Methionine	Met, M	
Proline	Pro, P	
Phenylalanine	Phe, F	
Tyrosine	Tyr, Y	
Tryptophan	Trp, W	

Amino Acid	Abbreviation	Structure
Aspartic Acid	Asp, D	
Glutamic Acid	Glu, E	
Asparagine	Asn, N	
Glutamine	Gln, Q	
Histidine	His, H	
Lysine	Lys, K	
Arginine	Arg, R	

Permission Letter for Figure 1. 3

Permission Letter for Figure 1. 4

11200 Rockville Pike
Suite 302
Rockville, Maryland 20852

August 19, 2011

American Society for Biochemistry and Molecular Biology

To whom it may concern,

It is the policy of the American Society for Biochemistry and Molecular Biology to allow reuse of any material published in its journals (the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research) in a thesis or dissertation at no cost and with no explicit permission needed. Please see our copyright permissions page on the journal site for more information.

Best wishes,

Sarah Crespi

American Society for Biochemistry and Molecular Biology
11200 Rockville Pike, Rockville, MD
Suite 302
240-283-6616
JBC | MCP | JLR

Tel: 240-283-6600 • Fax: 240-881-2080 • E-mail: asbmb@asbmb.org

Permission Letter for Figure 3. 1

creative commons

LEGAL CODE

Attribution 2.0

CREATIVE COMMONS CORPORATION IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS LICENSE DOES NOT CREATE AN ATTORNEY-CLIENT RELATIONSHIP. CREATIVE COMMONS PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. CREATIVE COMMONS MAKES NO WARRANTIES REGARDING THE INFORMATION PROVIDED, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM ITS USE.

License

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

1. Definitions

a. **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this License.

b. **"Derivative Work"** means a work based upon the Work or upon the Work and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work will not be considered a Derivative Work for the purpose of this License. For the avoidance of doubt, where the Work is a musical composition or sound recording, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered a Derivative Work for the purpose of this License.

c. **"Licensor"** means the individual or entity that offers the Work under the terms of this License.

d. **"Original Author"** means the individual or entity who created the Work.

e. **"Work"** means the copyrightable work of authorship offered under the terms of this License.

f. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License

despite a previous violation.

2. Fair Use Rights. Nothing in this license is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3. License Grant. Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

a. to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;

b. to create and reproduce Derivative Works;

c. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works;

d. to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission Derivative Works.

e. For the avoidance of doubt, where the work is a musical composition:

   i. **Performance Royalties Under Blanket Licenses**. Licensor waives the exclusive right to collect, whether individually or via a performance rights society (e.g. ASCAP, BMI, SESAC), royalties for the public performance or public digital performance (e.g. webcast) of the Work.

   ii. **Mechanical Rights and Statutory Royalties**. Licensor waives the exclusive right to collect, whether individually or via a music rights agency or designated agent (e.g. Harry Fox Agency), royalties for any phonorecord You create from the Work ("cover version") and distribute, subject to the compulsory license created by 17 USC Section 115 of the US Copyright Act (or the equivalent in other jurisdictions).

f. **Webcasting Rights and Statutory Royalties**. For the avoidance of doubt, where the Work is a sound recording, Licensor waives the exclusive right to collect, whether individually or via a performance-rights society (e.g. SoundExchange), royalties for the public digital performance (e.g. webcast) of the Work, subject to the compulsory license created by 17 USC Section 114 of the US Copyright Act (or the equivalent in other jurisdictions).

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4. Restrictions. The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a. You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this License, and You must include a copy of, or the Uniform Resource Identifier for, this License with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this License or the recipients' exercise of the rights granted hereunder.

You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this License Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this License. If You create a Collective Work, upon notice from any Licensor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested. If You create a Derivative Work, upon notice from any Licensor You must, to the extent practicable, remove from the Derivative Work any reference to such Licensor or the Original Author, as requested.

b. If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Derivative Works or Collective Works, You must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied; to the extent reasonably practicable, the Uniform Resource Identifier, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and in the case of a Derivative Work, a credit identifying the use of the Work in the Derivative Work (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Derivative Work or Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5. Representations, Warranties and Disclaimer

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

6. Limitation on Liability. EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

7. Termination

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Derivative Works or Collective Works from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.

b. Subject to the above terms and conditions, the license granted here is perpetual

(for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

8. Miscellaneous

a. Each time You distribute or publicly digitally perform the Work or a Collective Work, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.

b. Each time You distribute or publicly digitally perform a Derivative Work, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.

c. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

d. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

e. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.

Creative Commons is not a party to this License, and makes no warranty whatsoever in connection with the Work. Creative Commons will not be liable to You or any party on any legal theory for any damages whatsoever, including without limitation any general, special, incidental or consequential damages arising in connection to this license. Notwithstanding the foregoing two (2) sentences, if Creative Commons has expressly identified itself as the Licensor hereunder, it shall have all rights and obligations of Licensor.

Except for the limited purpose of indicating to the public that the Work is licensed under the CCPL, neither party will use the trademark "Creative Commons" or any related trademark or logo of Creative Commons without the prior written consent of Creative Commons. Any permitted use will be in compliance with Creative Commons' then-current trademark usage guidelines, as may be published on its website or otherwise made available upon request from time to time.

Creative Commons may be contacted at http://creativecommons.org/.

« Back to Commons Deed