Structural Investigation of the DedA Protein YqjA

Bree Streather

School of Biosciences
University of Kent

Thesis for MSc-R in Biochemistry

2021

Supervised by Dr Christopher Mulligan

Word Count: 12,424

University of Kent

Declaration

I declare that all the work carried out in this project and the wording of the report is entirely my own and no part has been copied from scientific journals, websites or other sources.

Signature: Bree Streather
Date: 15th November 2021

Abstract

As the threat of antimicrobial resistance (AMR) intensifies, resulting in a growing number of deaths worldwide every year, there is an increasing pressure to find solutions. One possibility is to reverse the resistance by determining its cause. DedA proteins are membrane transporters and have been shown to confer resistance to a variety of antibiotics in many different pathogenic bacteria. This project focussed on the structure of the DedA protein YqjA from *E. coli* which will aid in future drug design to block the functioning of this protein. The 3D structure was predicted via the EVcouplings algorithm which was then biochemically tested using Substituted Cysteine Accessibility Methods (SCAM). This uses single cysteine mutants of YqjA and, through treating the cells with various thiol-binding reagents, the location of the Cys residue in relation to the membrane can be determined. 8 out of 12 results from the SCAM analysis agreed with the predicted EVfold model showing it to be a reliable prediction of the 3D structure of YqjA. However, a rescue assay that was carried out to assess the functionality of the cysteine mutants needs to be repeated in order to have full confidence in the accuracy of the SCAM data.

Table of Contents

List of Figures

List of Tables

Abbreviations

Amp	Ampicillin
AMR	Antimicrobial resistance
Cam	Chloramphenicol
ChCl	Choline chloride
EtBr	Ethidium bromide
Kan	Kanamycin
LPS	Lipopolysaccharide
LB	Lysogeny broth
OD	Optical density
PMF	Proton motive force
SCAM	Substituted Cysteine Accessibility Methods
TMH	Transmembrane helix
WT	Wild type – *BW25113*
ΔΔ	*BW25113* Δ*ygjA*Δ*yghB*

1. Introduction

1.1. Antimicrobial Resistance (AMR)

One of the most rapidly increasing threats to public health is the development of resistance in pathogenic bacteria, viruses, fungi and parasites to current antimicrobials to such a level where they become an ineffective way of killing the pathogens. This can be inherent (i.e. naturally occurring via mutation, selection and evolution) but is also expedited by the improper use and overuse of antimicrobial agents (antibiotics, antivirals, antifungals and antiparasitics) in the treatment of humans, animals and plants (WHO, 2020). Once easily treatable infections are becoming progressively harder to deal with. This will disproportionately affect poorer people in developing countries as they have reduced access to clean water and sanitation, healthcare and information. It is and will continue to be more prevalent in hospitals and healthcare settings where patients are vulnerable and in close proximity. Recent predictions in the Review on Antimicrobial Resistance (Davies *et al.*, 2014) have estimated that the number of deaths directly attributed to AMR in 2050 will be 10 million which equates to one person every three seconds. This is an increase of more than ten-fold from the number recorded in 2014 (0.7 million) over less than 40 years. A comparison of the predicted number of deaths caused by AMR compared with cancer from this review is shown in Figure 1.

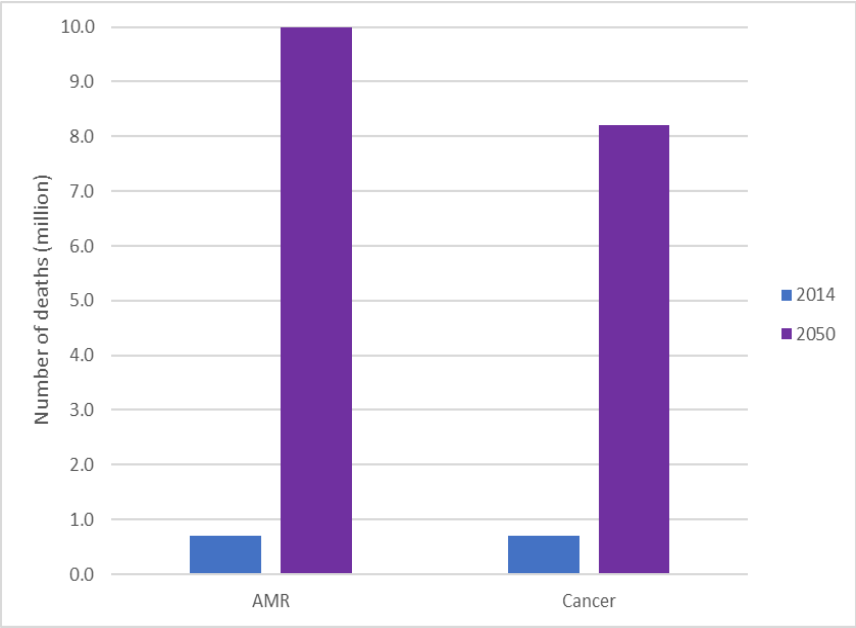


Figure 1 – Deaths attributable to AMR and cancer in 2014 vs those predicted for 2050 (Davies *et al.*, 2014).

Bacteria such as *Escherichia coli, Klebsiella pneumoniae* and *Staphylococcus aureus,* which cause many common and serious diseases from urinary tract and respiratory infections to sepsis, have shown high rates of resistance to the current antibiotics used against them according to the Global Antimicrobial Resistance and Use Surveillance System (GLASS) report (WHO, 2021). An example of this is the rise of methicillin-resistant *Staphylococcus aureus* (MRSA) infections in hospitals (Enright *et al.*, 2002). A number of countries have reported resistance to the drug of last resort colistin in the aforementioned bacterial species meaning

there is no longer an effective treatment for these infections. It is imperative that a solution to this ever-growing threat is found before more pathogens become untreatable leading to devasting consequences.

1.2. Structural Characterisation of Membrane Proteins

Membrane proteins are a vital class of proteins in all living cells. They have a variety of functions including transportation, cell recognition and signal transduction (Cournia *et al.*, 2015) and can be either peripheral (interacting with other proteins in the membrane or the lipids themselves) or integral (embedded in the membrane e.g. transmembrane proteins). Their significance is emphasised by the fact that they make up more than 60% of drug targets in humans (Yin and Flynn, 2016), for which it is fundamental to know the structure of the proteins. A smaller percentage of bacterial membrane proteins are targets for antibiotic drugs but these are still of great scientific importance (Bakheet and Doig, 2010).

The challenges faced when characterising the structure of membrane proteins corresponds with the challenges of purifying the proteins and reconstituting them into an appropriate membrane mimetic to prevent aggregation. A common technique is to extract the proteins using detergent but, more recently, attempts have been made to extract them in lipid nanodiscs using techniques such as styrene maleic acid copolymer lipid particles (SMALPs) so that they remain in an environment more similar to their native (Rawson *et al.*, 2016). The structure of the protein can be affected by different features of the membrane, like lipid composition, which are highly variable and can be influenced by the environment surrounding the cell, including the membrane proteins themselves (Lee, 2004). There are also issues with the overexpression required to collect a protein sample with a high enough concentration as this can be toxic to the cells.

Common successful methods of determining the 3D structure of soluble proteins, X-ray crystallography for example, are made much more difficult by the instability of membrane proteins once extracted from their native environment (Kang *et al.*, 2013). Although improvements in some techniques are being made, most structural data for membrane proteins is collected through mutagenesis and computational analysis (Almeida *et al.*, 2017).

1.3. DedA Family

The DedA/Tvp38 family of proteins is present in organisms across all three domains of life – bacteria, archaea and eukarya (Boughner and Doerrler, 2012). Despite thousands of members having been entered into the protein databases, they are not a well characterised family. Their exact roles are yet to be determined but they seem to perform various functions across the family including vesicle fusion in the late Golgi compartment for the yeast (*Saccharomyces cerevisiae*) homologue Tvp38 (Keller and Schneider, 2013) or autophagy for the human homologue TMEM41B (Morita *et al.*, 2019). From these functional studies, proteins in the DedA family have been classified as membrane proteins, more specifically membrane transporters (Kumar and Doerrler, 2014). Eukaryotic homologues, like Tvp38 or

TMEM41B, are located in the endoplasmic reticulum while bacterial homologues are found in the inner bacterial membrane. As well as a lack of functional data, there is also very little structural information known about these proteins, with only hydropathy and some mutational analysis having been carried out, and the oligomeric state is unknown. The most widely studied homologues are YqjA and YghB from *E. coli* (Keller, 2015) which have 61% amino acid identity to each other and are ~220 amino acids in length.

1.3.1. Predicted Functions of YqjA and YghB

When both of these proteins are knocked out (Δ*ygjA*Δ*yghB*, ΔΔ), the bacteria become sensitive to high temperatures (42°C) (Thompkins *et al.*, 2008), have cell division defects, an altered phospholipid composition in the membrane and a reduced proton motive force (PMF) (Keller, 2015). Δ*ygjA*Δ*yghB* is also unable to grow in reduced sodium conditions, below standard LB media concentrations. This implies that YqjA and YghB are involved in the growth of *E. coli* at increased temperatures, cell division processes, the integrity of the cell-surface membrane and possibly proton/ion transport. Interestingly, single knockouts Δ*yqjA* and Δ*yghB* individually do not show the temperature sensitivity phenotype but when both genes are knocked out, the bacteria are unable to grow above 30 °C. *yqjA* and *yghB* genes have been shown to be paralogous in some cases, indicating that one can compensate for the other (Sikdar and Doerrler, 2010).

The effect on the PMF from knocking out YqjA and YghB suggests that these proteins may function as proton pumps. YqjA alone is able to function as an osmosensor; that is, it detects an increase in osmolarity and, in response, its proton-dependent transport is enhanced (Kumar and Doerrler, 2015). Kumar and Doerrler also showed that a Δ*yqjA* mutant cannot grow at pH 8.5 - 9.5 and expression of a YqjA plasmid can restore the growth but only with a sufficient external Na^+ or K^+ concentration. Hence, it was concluded that YqjA is a Na^+/K^+-H^+ antiporter that is involved in the maintenance of cytoplasmic pH and the PMF. Furthermore, increasing external pH has been seen to cause an increase in the activity of YqjA and *vice versa* (Kumar *et al.*, 2017). YghB has not been observed functioning in this way.

Importantly, DedA proteins have been shown to confer resistance to a range of common and widely used antibiotics, such as EtBr, methyl viologen and tetracycline (Kumar and Doerrler, 2014), with this resistance having a possible link to the transport of protons across the membrane. The proteins may function as drug efflux pumps and exchange protons for the antibiotics. Of particular interest, is resistance to the antibiotic colistin which is used only as a last resort (Panta, 2019) i.e. if all other antibiotics have failed. Colistin acts by disrupting the cation bridges between lipopolysaccharide (LPS) molecules in both bacterial membranes, releasing the LPS and destabilising the bilayer, leading to cell lysis and death (Sabnis *et al.*, 2020). DedA proteins play a role in modifying the lipopolysaccharide lipid A in the outer membrane with positively charged aminoarabinose (Ara4N) to neutralise the charge of lipid A and prevent colistin from interacting with it. Sufficient PMF is essential for this process to occur, providing more evidence that DedA proteins are proton transporters.

In the ΔΔ strain, there are decreased levels of phosphatidylethanolamine (PE) and increased levels of both phosphatidylglycerol (PG) and cardiolipin (CL) (Thompkins *et al.*, 2008) (Liang *et al.*, 2010). PE is a non-bilayer lipid whereas PG and CL are bilayer forming lipids. Since the membrane is destablised by knocking out *yqjA* and *yghB,* the shift in membrane composition may be to counteract the destabilisation (Sikdar and Doerrler, 2010) suggesting that YqjA and YghB are not directly involved in phospholipid synthesis but are indirectly linked to the membrane composition (Keller and Schneider, 2013). Transcription of YqjA is regulated by the Cpx-signal transduction pathway (Yamamoto and Ishihama, 2006) (Price and Raivio, 2009) which plays a role in membrane integrity (Raivio *et al.*, 2013), among other things, and this stress response pathway is active in the ΔΔ (Sikdar *et al.*, 2013).

Moreover, YqjA and YghB have been linked to the twin arginine transport (Tat) pathway (Sikdar and Doerrler, 2010), possibly aiding export of certain periplasmic amidases that are involved in the septation stage of cell division. ΔΔ cells show a defect at this stage causing them to grow in chains with partially fused inner membranes. It has been hypothesised that this is a side effect of the irregular composition of membrane phospholipids because Tat substrates are able to form electrostatic interactions with PG and CL and hence are unable to cross the inner membrane into the periplasm.

With such a variety of different phenotypes, there is still a large amount of speculation as to the functions of YqjA and YghB and other DedA family members. On top of this, with all the different pathways implicated in their expression, there is still a lot more research to be done on their downstream effects.

1.3.2. Current Structural Data for YqjA and YghB

YqjA and YghB are both found in the inner membrane of *E. coli* and are around 220 amino acids in length. No successful attempts have been made to crystallise either of these proteins and the majority of the structural data collected up to now has been through mutational and computational analysis, consistent with that for the majority of membrane proteins. Through various predictive topology algorithms and hydropathy analysis, with some supporting biochemical analysis, it has been suggested that YqjA has been four and six transmembrane helices (Thompkins, 2008; Doerrler *et al*., 2013; Keller, 2015). Figure 2 shows the range of different 2D topology models of YqjA generated thus far including the current prediction (Scarsbrook *et al.*, 2021).

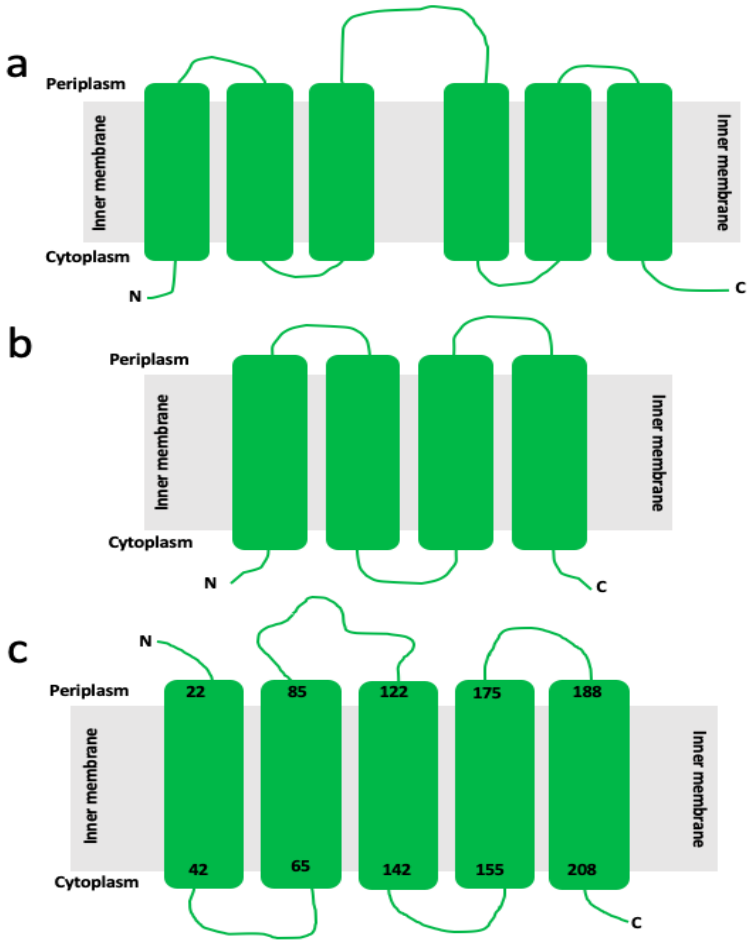


Figure 2 – Various 2D topological models of YqjA predicting between 4 and 6 transmembrane helices.
a) Thompkins, 2008 b) Doerrler *et al.* 2013 and c) Scarsbrook *et al.*, 2021

Two acidic residues, E39 and D51 (in both homologues), are essential for the proper functioning of the protein. If these are mutated to alanine, the temperature sensitivity, cell division defects and antimicrobial sensitivity phenotypes of the ΔΔ returns (Kumar and Doerrler, 2014). Intriguingly, while both of these residues are vital for alkaline tolerance, E39 seems to play more of a role at more extreme pH (above pH 9). Furthermore, mutations to mimic the protein's protonated state (E39Q and D51N) were unable to rescue the ΔΔ phenotypes showing that the charge of these residues is key for the protein to function. These acidic acid residues are conserved in other members of the DedA family too but with slight variation in location. For example, in *Borrelia burgdorferi,* the DedA protein BB0250 is essential for cell viability and knockouts have comparable phenotypes to ΔΔ although it has just 19% amino acid identity to YqjA. In this protein, the acidic acid residues are E39 and D40.

Other residues that have been of particular interest in YqjA are two native cysteine residues that may be important for oligomerisation. These are C83 in predicted TMH3 and C191 in predicted TMH6 (Keller, 2015). More specifically, C191 has been observed cross-linking two YqjA monomers in a C83A mutant although this does not necessarily provide evidence that YqjA is a dimer because a disulfide bond can only form between two Cys residues so more than one Cys would need to be present in each monomer to allow further oligomerisation. It has been hypothesised that a heterodimer between YqjA and YghB may

be possible but there is no evidence that this is true. YqjA can dimerise in the absence of YghB but other than this, there has not been much investigation into the oligomeric state of either YqjA or YghB. In fact, each protein is expressed by a different stress pathway – Cpx-signal transduction pathway (YqjA) and the quorum sensing molecule AI-2 (YghB) (Doerrler *et al.*, 2013) – and so perhaps they are not present in the cell at the time making them unlikely to interact *in vivo*.

Additionally, there are two arginine residues, R130 and R136, that are important for YqjA to function (Kumar *et al.*, 2016). These are both predicted to be in TMH3 (Scarsbrook *et al*., 2021) There is a charge dependence with R130 as the Arg residue can be substituted for Lys in this position (the same is not true for R136). Positively charged residues have been shown to play a role in proton transport (Sigal, 2005) which is in line with the hypothesis that DedA proteins, in particular YqjA and YghB, are proton transporters.

There is still much more to be understood about the structure of all DedA proteins including the most widely studied, YqjA and YghB.

1.4. Importance of Studying YqjA

Because there is evidence that YqjA is instrumental in the resistance to many different antibiotics seen in some strains of *E. coli* (Kumar and Doerrler, 2014), it is extremely important to understand as much as possible about how it functions and confers resistance so that this can be inhibited or counteracted. On top of this, structural information about the protein is crucial when designing drugs to target the protein and is intrinsically related to protein function e.g. structural changes upon substrate binding.

1.5. Project Aims

The aim of this project was to more accurately define the structure of the *E. coli* DedA protein YqjA. Computational modelling was undertaken using EVcouplings to predict a possible 3D structure of YqjA. Following this, substituted cysteine accessibility measurement (SCAM) analysis was carried out to determine where particular residues are in relation to the inner membrane of the bacteria. The data from SCAM was then used to evaluate the accuracy of the current 2D topological model of the protein (Scarsbrook *et al.*, 2021) and the 3D EV fold model. This is still preliminary structural analysis but is vital especially with the difficulties in obtaining crystal structures of membrane proteins.

2. Materials and Methods

2.1. Strains, Plasmids and Media

2.1.1. Strains

Strain of *E. coli*	Source	Genotype
BW25113 (wild type, WT)	Keio Collection	*lacI*$^{+}$*rrnB*$_{T14}$Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$*rph-1* Δ(*araB-D*)567 Δ(*rhaD-B*) 568 Δ*lacZ4787*(::*rrnB-3*) *hsdR514 rph-1*
BW25113 Δ*yqjA*:: Kan^{r}	Keio Collection	
BW25113 Δ*yghB*:: Kan^{r}	Keio Collection	
BW25113 Δ*yqjA*Δ*yghB*	This work	
NEB Turbo	NEB	*glnV44 thi-1* Δ(*lac-proAB*) *galE15 galK16 R*(*zgb-210*:: Tn*10*)Tet^{S} *endA1fhuA2* Δ(*mcrB-hsdSM*)5, (*r*$_{K}^{-}$ *M*$_{K}^{-}$) F'[*traD36 proAB*$^{+}$*lacI*^{q} *lacZ*Δ*M15*]
TOP10	Invitrogen	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80*lacZ*Δ*M15* Δ*lacX74 nupG recA1 araD139* Δ(*ara-leu*)7697 *galE15 galK16 rpsL*(*Str*^{R}) *endA1 λ*$^{-}$

Table 1 – List of *E. coli* strains used in this work, including their source and genotype.

2.1.2. Plasmids

Type	Plasmid details	Selective Resistance
Control	pBADhisGltph pBADhisYqjA	Amp^{r}
Cysteine mutants	pBADhisYqjACys-less pBADhisYqjAA46C pBADhisYqjAS52C pBADhisYqjAL53C pBADhisYqjAY68C pBADhisYqjAL75C pBADhisYqjAV99C pBADhisYqjAL117C pBADhisYqjAA143C	Amp^{r}

	pBADhisYqjAL148C	Amp^r
	pBADhisYqjAR152C	
	pBADhisYqjAL162C	
	pBADhisYqjAL175C	
	pBADhisYqjAT178C	
	pBADhisYqjAV180C	
	pBADhisYqjAY184C	
	pBADhisYqjAL188C	
	pBADhisYqjAL194C	
	pBADhisYqjAL195C	
	pBADhisYqjAC191S	
	pBADhisYqjAV200C	

Table 2 – List of plasmids used in this work and their associated selective antibiotic resistance.

2.1.3. Media and Solutions

Type of Media	Ingredients
Liquid LB	10 g/L NaCl, 10 g/L Tryptone, 5 g/L Yeast Extract in dH_2O, sterilised by autoclaving
LB agar	10 g/L NaCl, 10 g/L Tryptone, 5 g/L Yeast Extract, 15 g/L Agar in dH_2O, sterilised by autoclaving
LB agar with Ethidium Bromide (EtBr)	10 g/L NaCl, 10 g/L Tryptone, 5 g/L Yeast Extract in dH_2O, sterilised by autoclaving. 75 µg/ml EtBr added
LB agar choline chloride (ChCl) 100% molarity replacement	23.88 g/L ChCl, 10 g/L Tryptone, 5 g/L Yeast Extract in dH_2O, sterilised by autoclaving
LB agar of desired pH	10 g/L NaCl, 10 g/L Tryptone, 5 g/L Yeast Extract in dH_2O. 70 mM bis-tris propane (BTP) added, pH adjusted with HCl and sterilised by autoclaving
SOC	0.5% Yeast Extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM MgSO4, adjusted to pH 7.5 with NaOH and sterilised by autoclaving. 200 µl 1M sterile glucose (final concentration of 20 mM) added

Table 3 – Recipes for the media used for bacterial growth.

If an antibiotic was required for plasmid selection in liquid or solid media, it was at a final concentration of 100 µg/ml.

Solution	Ingredients
SDS-PAGE Gel (x2)	Resolving: 6.9 ml Bisacrylamide, 2.6 ml 1.875 M Tris pH 8.8, 130 µl 10% SDS, 3.77 ml dH_2O, 130 µl 10% APS and 13 µl Tetramethylethylenediamine (TEMED) Stacking: 1.69 ml Bisacrylamide, 1.625 ml 1 M Tris pH 6.8, 130 µl 10% SDS, 9.56 ml dH_2O, 130 µl 10% APS and 13 µl Tetramethylethylenediamine (TEMED)
4X SDS Sample Buffer	12 g glycerol, 3 ml dH_2O, 10 ml 10% SDS, 1 ml 1M Tris pH 7.2 (or 6.8), 0.06 g Bromophenol Blue
10X SDS Running Buffer	30 g Tris, 150 g glycine, 10 g SDS, made up to 1 L with dH_2O
1X Transfer Buffer	14.41 g glycine, 100 ml methanol, 25 ml 1M Tris pH 8.3, made up to 1 L with dH_2O
10X Tris-Buffered Saline Tween (TBST)	24.2 g Tris, 80 g NaCl, 10 ml Tween 20, adjusted to pH to 7.6 with HCl, made up to 1 L with dH_2O
Coomassie Brilliant Blue Dye	50% methanol, 50% acetic acid, 0.2% Coomassie Blue
Coomassie Destain	10% ethanol, 10% acetic acid, 80% dH_2O

Table 4 – Recipes for solutions used in this work.

2.2. Miniprep

To extract and purify the plasmid DNA from the bacteria, the QIAprep Spin Miniprep kit (Qiagen) was used. 5 ml sterile LB Amp (Table 3) was inoculated with *E. coli* cells containing the desired plasmid, either from a glycerol stock or a colony of transformed cells grown on LB Amp agar (Table 3), and allowed to grow overnight at 37 °C , 190 rpm. After this, the QIAprep Spin Miniprep Kit protocol was followed. Following the addition of buffer N3, the solution was centrifuged at 13,000 rpm for 8 minutes, instead of 10 as advised, and all other subsequent centrifugation steps lasted 60 seconds. Buffer PB was not used as all strains used were endA-. The DNA concentration and 260/280 values of the miniprep product were measured using a Thermo Scientific Nanodrop 2000.

2.3. Competent Cells

E. coli cells of the desired strain were streaked onto LB agar plates and grown overnight at 37 °C in order to obtain single colonies. A single colony was then used to inoculate 5 ml LB and incubated at 37 °C, 190 rpm overnight. 1 ml of the resulting culture was used to inoculate 100 ml sterile LB in a 500 ml flask which was grown at 37 °C, 190 rpm until the OD_{600} measured 0.4. The culture was split into two 50 ml sterile falcon tubes and placed on ice for 20 minutes.

Following this, the cells were harvested by centrifugation at 4 °C for 10 minutes at 3000 xg. The supernatant was discarded and the pellet resuspended in 3.75 ml of ice-cold $CaCl_2$ before a further 30 minutes on ice. Next, the cells were centrifuged as before and resuspended in 2 ml 0.1 M $CaCl_2$ + 15% glycerol. Aliquots of 50 or 100 µl were placed into 1.5 ml Eppendorf tubes, snap frozen and stored at -80 °C until required.

2.4. Transformation

2.4.1. Heat Shock

This method was used for chemically competent cells (see Section 2.3.). The competent cells were removed from the -80 °C freezer and allowed to thaw on ice. 100 ng of plasmid DNA was added to each 50 µl aliquot under aseptic conditions, mixed via pipetting and left on ice for 30 minutes. These were then heat-shocked using a 42 °C heat block for 90 seconds. 200 µl of SOC was added to each Eppendorf and the cells allowed to grow at 37 °C, 190 rpm for 1 hour. 10 µl of the transformant mixture was plated onto selective LB Amp plates and incubated at 37 °C overnight.

2.4.2. Freeze-thaw

An 5 ml LB overnight culture was set up from a glycerol stock of the desired *E. coli* strain to be transformed and grown at 37 °C, 190 rpm. The following day, 600 µl of this culture was used to inoculate 5 ml LB and grown for 1 hour at 37 °C, 190 rpm. 833 µl of this culture was used for each transformation and aliquoted into separate 1.5 ml Eppendorf tubes. Using a microfuge, the cells were harvested at 13,000 rpm for 1 minute. The cell pellet was resuspended in 50 µl ice-cold 1 M $CaCl_2$ and left on ice until required. 100 ng of plasmid DNA was added to each Eppendorf, mixed using a pipette and snap-frozen in dry-ice/ethanol for 90 seconds. The cells mixture was then thawed at 37 °C on a heat block for 2 minutes. 200 µl of SOC was added to each Eppendorf and the cells allowed to grow at 37 °C, 190 rpm for 1 hour. 10 µl of the transformant mixture was plated onto selective LB Amp plates and incubated at 37 °C overnight.

2.5. Production of double gene knockout Δ*yqjA*Δ*yghB* in *BW25113*

2.5.1. Preparation of Phage Lysates

5 ml LB cultures of *BW25113* Δ*yqjA* and Δ*yghB* from the Keio collection (Datsenko and Wanner, 2000) (Baba *et al.*, 2006) were grown overnight at 37 °C, 250 rpm. 400 µl of each culture was used to inoculate separate prewarmed flasks containing 25 ml LB with 0.2% glucose and 5 mM $CaCl_2$. These were incubated for 30 minutes at 37 °C, 250 rpm. 400 µl of premade phage lysate (provided by Dr Alex Moores from Kad Lab) was added to each culture and incubated for another 30 minutes as before. Every 30 minutes, the cultures were checked visually for cell lysis – indicated by a lack of turbidity and the appearance of

precipitates. 50 ml of this phage culture was transferred to a sterile falcon tube before adding 400 µl of chloroform and mixed by vortexing, allowing the phage to precipitate. The mixture was subjected to centrifugation at 4500 rpm for 10 minutes at 4 °C. 5 ml of the supernatant was extracted and 400 µl chloroform added before storing at 4 °C.

2.5.2. Curing of the Kan^r cassette

Day 1
BW25113 Δ*yqjA* and Δ*yghB* competent cells were made using the protocol described in section 2.3. A 50 µl aliquot of each of these cells was transformed with 2 µl of the plasmid pCP20 which encodes the arabinose-inducible yeast Flp recombinase along with chloramphenicol (Cam) and ampicillin resistance genes. The heat-shock method was used for this (see section 2.4.1.), with LB instead of SOC and incubated for 2.5 hours at 28 °C instead of 1 hour at 37 °C. Following incubation, the transformant mixture was centrifuged at 4000 rpm for 2 minutes. 100 µl of the supernatant was plated onto LB Amp agar and incubated at 28 °C overnight.

Day 2
One colony from the transformation plate was used to inoculate 5 ml LB which was incubated at 42 °C for 3 hours. The culture was then serially diluted 10-fold (6 series) with 100 µl of dilutions 10^{-3} to 10^{-6} being plated onto LB agar and incubated overnight at 42 °C.

Day 3
Selective LB kanamycin (Kan), LB Cam, LB Amp and LB agar plates were each split into a 6x6 grid. Single colonies were streaked onto each plate with the same colony being streaked onto the same grid square on each plate and then incubated overnight at 37 °C.

Day 4
Colonies that grow on the LB agar plate only were restreaked for single colonies on a fresh LB agar plate and incubated overnight at 37 °C. These colonies will no longer have the pCP20 plasmid (no Cam^r or Amp^r) and will have had the Kan^r cassette removed by the Flp recombinase.

Day 5
A single colony was used to inoculate 5 ml LB and grown overnight at 37 °C, 190 rpm in order to make a glycerol stock and stored at -80 °C.

2.5.3. P1vir Transduction

A 5ml LB overnight culture was set up of the recipient strain (*BW25113* Δ*yqjA* was used here but *BW25113* Δ*yghB* could also be used) and grown at 37 °C, 190 rpm. The next day, 200 µl of the culture was added to 200 µl phage buffer (10 mM $CaCl_2$, 20 mM $MgSO_4$), mixed and 100 µl distributed into 3 Eppendorf tubes. 10 µl of phage lysate was added to tube 1, 50 µl was added to tube 2 and both were mixed. Tube 3 was left as a negative control. Another tube was set up containing 50 µl of phage and lysate only. All samples were then incubated at 28 °C for 30 minutes before adding 1 ml of LB containing 0.1 M citrate to all four tubes. They were then incubated for 2-3 hours at 37 °C. 100 µl of each

mixture was plated onto LB Kan agar plates with tubes 1 and 2 being the only ones expected to result in colonies (these would be *BW25113* Δ*yqjA*Δ*yghB* colonies). Single colonies from these were grown overnight at 28 °C, 190 rpm in 5 ml LB Kan and used to make glycerol stocks to be stored at -80 °C.

2.5.4. Confirmation of Transduction

In order to confirm that both YqjA and YghB genes had both been deleted, the bacterial strain was tested for phenotypes associated with the double knockout Δ*yqjA*Δ*yghB* (ΔΔ): sensitivity to high temperatures (42 °C), alkaline sensitivity (pH 9.25), EtBr sensitivity and sensitivity to low sodium (ChCl replacement) (section 2.6). ΔΔ cells must always be grown at 30 °C.

2.6. Sensitivity Assay

From glycerol stocks, 5 ml LB was inoculated with ΔΔ cells and grown at 30 °C, 190 rpm overnight. An OD_{600} reading of the overnight was taken and normalised to 1. A 5-series, 10-fold dilution was carried out with the normalised culture and sterile LB. 5 µl of each dilution was spotted onto LB agar appropriate for each condition to be tested (see Table 3 and Section 2.5.4.). All plates were then incubated overnight at 30 °C (apart from the 42 °C condition).

2.7. Primer Design

All primers for site-directed mutagenesis were designed as per the guidelines of the QuickChange II Site-Directed Mutagenesis Kit instruction manual. The primers used in this project can be found in the appendix.

2.8. Site-Directed Mutagenesis via KOD PCR

This method was used to produce single cysteine mutants in a Cys-less (C93S/C181S) YqjA background.

Component	**Volume**
10x KOD buffer	2.5 µl
2 mM dNTPs	2.5 µl
25 mM $MgSO_4$	1 µl
5 µM Forward Primer	1.5 µl
5 µM Reverse Primer	1.5 µl
KOD Polymerase	0.5 µl
dH_2O	15 µl
Template DNA	0.5 µl
Total	**25 µl**

Table 5 – Contents of a 25 µl reaction mixture used for site-directed mutagenesis.

Time	Temperature
1 minute	95 °C
30 seconds 1 minute 6 minutes	95 °C 55 °C 68 °C x 16 cycles
5 minutes	72 °C
∞	12 °C

Table 6 – PCR cycle used for site-directed mutagenesis.

To run the mutagenesis cycle, a T1000 BioRad Thermocycler was used. The PCR product was then digested using 1 µl *DpnI* in 5 µl CutSmart buffer for 1 hour at 37 °C, 0 rpm. 2 µl of this was then used to transform 50 µl NEB Turbo $RbCl_2$ competent cells which were plated onto selective LB Amp plates and left to grow overnight at 37 °C. Single colonies were picked and grown in selective liquid LB overnight in order to make glycerol stocks (600 µl culture, 400 µl 50% glycerol) and to miniprep the plasmid.

2.9. Sequencing

Mutants produced for use in this study were sequenced by GATC sequencing provided by Eurofins. The data from this was analysed using the computer program A plasmid Editor (ApE) (Jorgensen, 2019). For this analysis, please refer to the appendix.

2.10. Rescue Assay

Plasmid DNA containing the single cysteine mutants of YqjA were transformed into Top10 *E. coli* cells (see section 2.4.). The control plasmid, Gltph, was transformed into both TOP10 and *BW25113* (WT) *E. coli* cells. 5 ml LB Amp was inoculated with one colony from each transformation and grown overnight at 37 °C, 190 rpm. Each culture was normalised to OD_{600} of 0.1 in fresh 5 ml LB Amp and grown to an OD_{600} of 0.4. Control LB and LB with a 100% molarity replacement of NaCl for ChCl (Table 3) containing Amp and 0.01% L-arabinose was inoculated to an OD_{600} of 0.05 with each culture and 200 µl was distributed into the wells of a 96-well plate. A Multiskan FC Microplate Photometer was used to take OD_{600} readings over 14 hours at 30 °C, medium rotation.

2.11. EVfold Modelling

The YqjA sequence was submitted to the EVcouplings server (Marks *et al.*, 2011) (Hopf *et al.*, 2012) with all settings left unchanged.

2.12. Substituted Cysteine Accessibility Methods (SCAM)

Plasmid DNA containing the single cysteine mutants of YqjA were transformed into Top10 *E. coli* cells (see section 2.4.). One colony was used to inoculate 5 ml LB Amp and grown overnight at 37 °C, 190 rpm. The resulting cultures were normalised to an OD_{600} of 0.2 in fresh 5 ml LB Amp and incubated at 30 °C, 190 rpm. After 90 minutes, the cells were induced with 0.1% L-arabinose and incubated under the same conditions for another hour. Next, the cells were harvested via centrifugation (5,000 rpm for 10 minutes) in order to concentrate them to an OD_{600} of 10 in 200 µl of 1 x PBS which was then distributed into four 50 µl aliquots. 5 µl of dH_2O was added to two of the aliquots, 5 µl 100 mM sodium (2-sulfonatoethyl) methanethiosulfonate (MTSES) was added to another aliquot of cells and 5 µl 100 mM N-Ethylmaleimide (NEM) was added to the other before incubating at room temperature in the dark for 1 hour. Cells were harvested by centrifugation at 13,000 rpm for 1 minute, washed with 100 µl 1 x PBS and centrifuged as above once more. The pellet was then resuspended in 22.5 µl lysis buffer (15 µl 1 M Tris-HCL pH 7.6, 300 µl 10% SDS, 0.376 g urea, made up to 1 ml with dH_2O) and 7.5 µl 25 mM methoxy-poly(ethylene glycol)-maleimide (mPEG5K) was added to three of the tubes for each mutant. 6 µl of DNase and 6 µl of protease inhibitors were added to all samples and then they were incubated for a further hour at room temperature in the dark. 17 µl of 4 x SDS sample buffer (see Table 4) was added to each sample and they were stored at -20 °C.

2.13. SDS-PAGE

SDS-PAGE gels were made as described in Table 4 and the Mini Gel Tank (BioRad) was used to run these gels. The sialic acid-binding periplasmic protein SiaP from Haemophilus influenzae (83.5 µM) was used as a positive control in all Western blot gels to identify successful transfer. 2 µl of Page Ruler Plus protein ladder, 10 µl of each sample and 1 µl of SiaP was loaded onto the gels and then set to 40 mA for 45 minutes for one gel or 60 mA for 1 hour for two gels.

2.14. Western Blot

The Mini Blot Module (BioRad) was used to carry out all Western blots. First, the polyvinylidene difluoride (PVDF) membrane was activated in 100% methanol before being equilibrated in transfer buffer along with the other components of the blot sandwich. The sandwich was prepared according to the manufacturer's instructions as follows: cathode core (-ve), sponge pad, filter paper, gel, membrane, filter paper, sponge pad, anode core (+ve). The tank was then filled with transfer buffer and set to run at 20 V (constant) for 1 hour. After this time, the gel was stained with Coomassie Brilliant Blue dye overnight and the membranes with Ponceau S stain for 5 minutes. The Ponceau was poured off and excess rinsed with dH_2O. 20 ml blocking solution of 5% skimmed milk powder in 1 x TBST was added to the membrane and placed on the shaker for 30 minutes. The blocking solution was discarded and the membranes washed in 1 x TBST for 15 minutes and then twice for 5 minutes. Primary hybridisation buffer was made with 1% skimmed milk powder in 1 x TBST with the tetraHis

primary antibody at a dilution of 1:5000. 5ml of this was added to the membrane which was sealed in plastic and shaken overnight at 4 °C.

The following day, the Coomassie stain was removed and the gel was soaked in Coomassie destain for a few hours. The primary hybridisation buffer was recycled and the membranes washed for 15 minutes in 1 x TBST followed by two 5 minute washes. 20 ml secondary hybridisation buffer (1:5000 anti-mouse secondary antibody (conjugated with horseradish peroxidase) in 1 x TBST with 1% skimmed milk powder) was added to the membrane and shaken at room temperature for 1 hour. It was then discarded and the membrane washed in 1 x TBST for 15 minutes and then 6 washes of 5 minutes each. The Supersignal West Pico PLUS Chemiluminescent kit (Thermo Scientific) was used to visualise the membrane, with a 1:1 ratio of each detection fluid (a total of 5 ml for each membrane), and then imaged using a Chemidoc BioRad on the ChemiHi Resolution setting with no filter. 10 images were taken across a 60 second exposure time, with the final image being used each time. A colourimetric image was also taken to capture the protein ladder and the two images merged. Images were also taken of the post-transfer gels using the Coomassie setting.

3. Results

3.1. Confirmation of ΔΔ Formation

The double gene knockout *BW25113* Δ*yqjA*Δ*yghB* (ΔΔ) was produced in this study for use in rescue assays. These assays determine whether the single cysteine mutants of YqjA, which were generated for substituted cysteine accessibility methods (SCAM) (see section 3.3.2.), are still functional and hence properly folded and inserted into the membrane. Each mutant is transformed into ΔΔ cells which are then grown under one of the stress conditions known to affect this strain (see below). If the mutant proteins are able to rescue the growth of the ΔΔ under these conditions, then they are functional. Their functionality is compared to WT YqjA which has been proven to be able to sufficiently rescue the growth of ΔΔ.

The last stage in the production of the ΔΔ is the introduction of Δ*yghB* into *BW25113* Δ*yqjA* via P1vir transduction. Following this, it is important to confirm that both YqjA and YghB genes have been knocked out of the genome. A sensitivity plate assay was used for this as *BW25113* ΔΔ has a distinctive set of phenotypes that can be tested. These are a sensitivity to high temperatures (42 °C), alkaline pH (pH 9.25), low sodium (replacement of NaCl with choline chloride (ChCl) in the media) and EtBr (75 µg/ml). The ΔΔ strain was remade for use in this study as anecdotal evidence has shown that it loses its distinctive phenotypes over time (C. Mulligan, personal communication), although it is unclear why. When this strain was first made, the sequence was confirmed via PCR therefore any subsequent productions of the ΔΔ are trusted to have the correct sequence as long as they display the appropriate phenotypes.

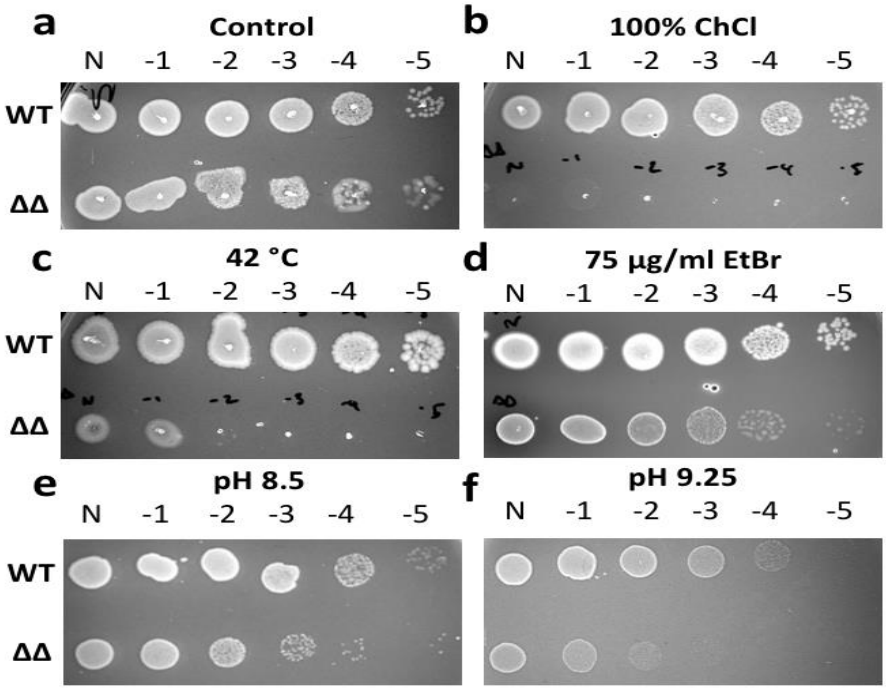
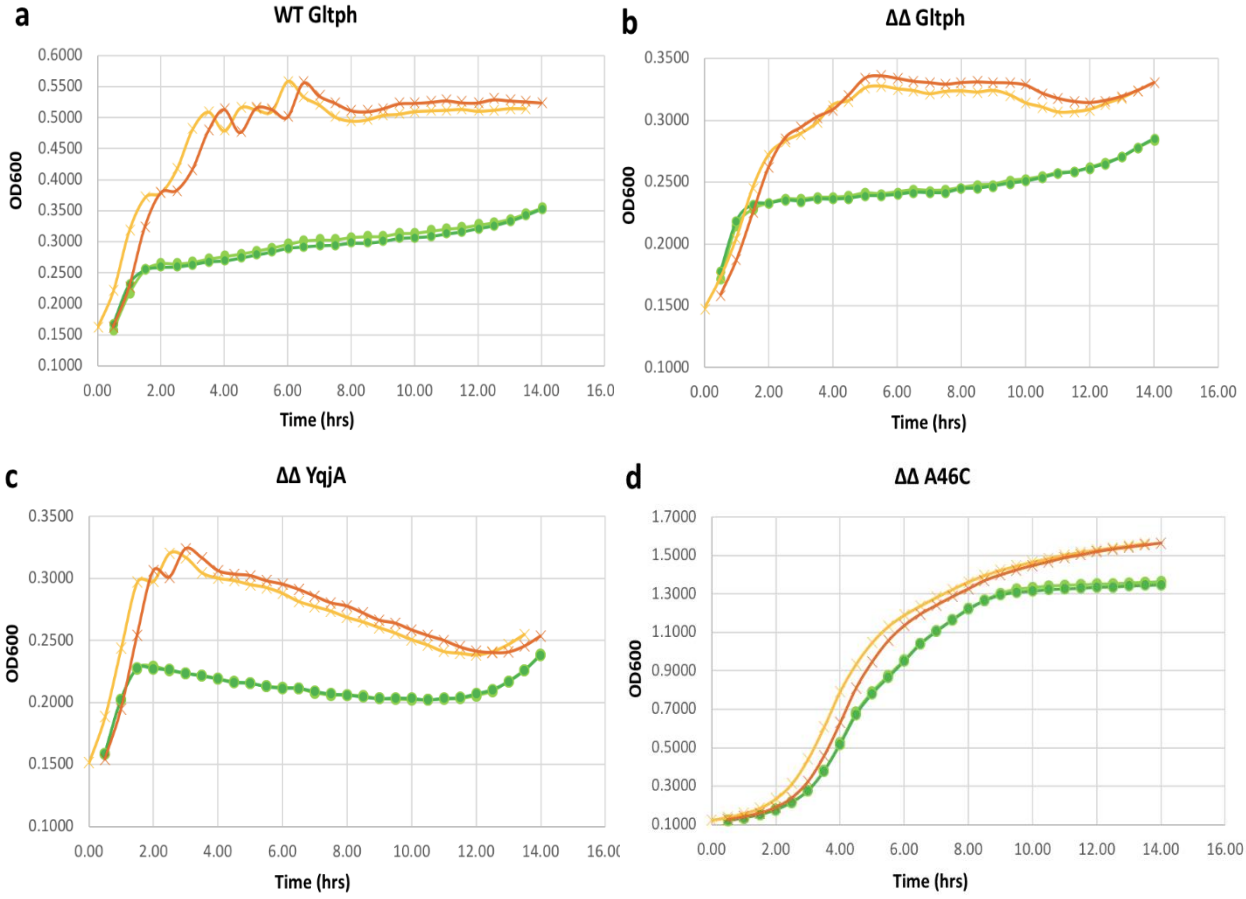


Figure 3 – Sensitivity plate assay of *BW25113* WT and *BW25113* Δ*yqjA*Δ*yghB* (ΔΔ) under various stress conditions known to affect growth of the ΔΔ.
The control consisted of standard LB agar and was grown at 30 °C (a). The stress conditions were: low sodium (100% molarity replacement of NaCl for ChCl in the media) (b), high temperature (42 °C) (c), 75 µg/ml EtBr (d) and high pH (9.25) (f). The annotations above each spot refer to the dilutions of culture. "N" stands for neat culture and then each spot after than is a 1 in 10 dilution of the previous.

BW25113 WT cells are expected to grow in all conditions and *BW25113* ΔΔ cells are only expected to grow in control conditions – LB agar made with NaCl (standard LB), 30 °C, pH 8.5 and no EtBr – pre-determined by past ΔΔ research (Thompkins, 2008; Kumar and Doerrler, 2015; Scarsbrook *et al.*, 2021). The ability of these strains to grow under the permissive conditions described above was evident from the standard LB plate grown at 30 °C (i.e. the control plate, see Fig. 3a) where both strains grew to the same level. There was growth for WT and ΔΔ in the neat culture spot (labelled N in Fig. 3) and all five dilutions with a few single colonies at the spot for the -5 dilution. The ΔΔ struggled to grow in the other conditions (the stress conditions) and the most severe condition was the low sodium/ChCl condition (Fig. 3b) in which the ΔΔ did not grow at all. The next most definitive phenotype was 42 °C (Fig. 4c) where the ΔΔ could only grow to -1 dilution. At pH 9.25 (Fig. 3f), the ΔΔ grew to the -2 dilution but overall growth was not as strong as for WT which grew to the -4 dilution and every spot had a much larger area of growth.

The EtBr sensitivity phenotype did not present as strongly as these other phenotypes (Fig. 3d). The ΔΔ was able to grow in LB containing 75 µg/ml EtBr across all dilutions, although not as strongly as the WT. In the -4 and -5 dilutions there were some single colonies and for the -5 dilution these were quite faint.

3.2. Functional Analysis of Single Cysteine Mutants of YqjA



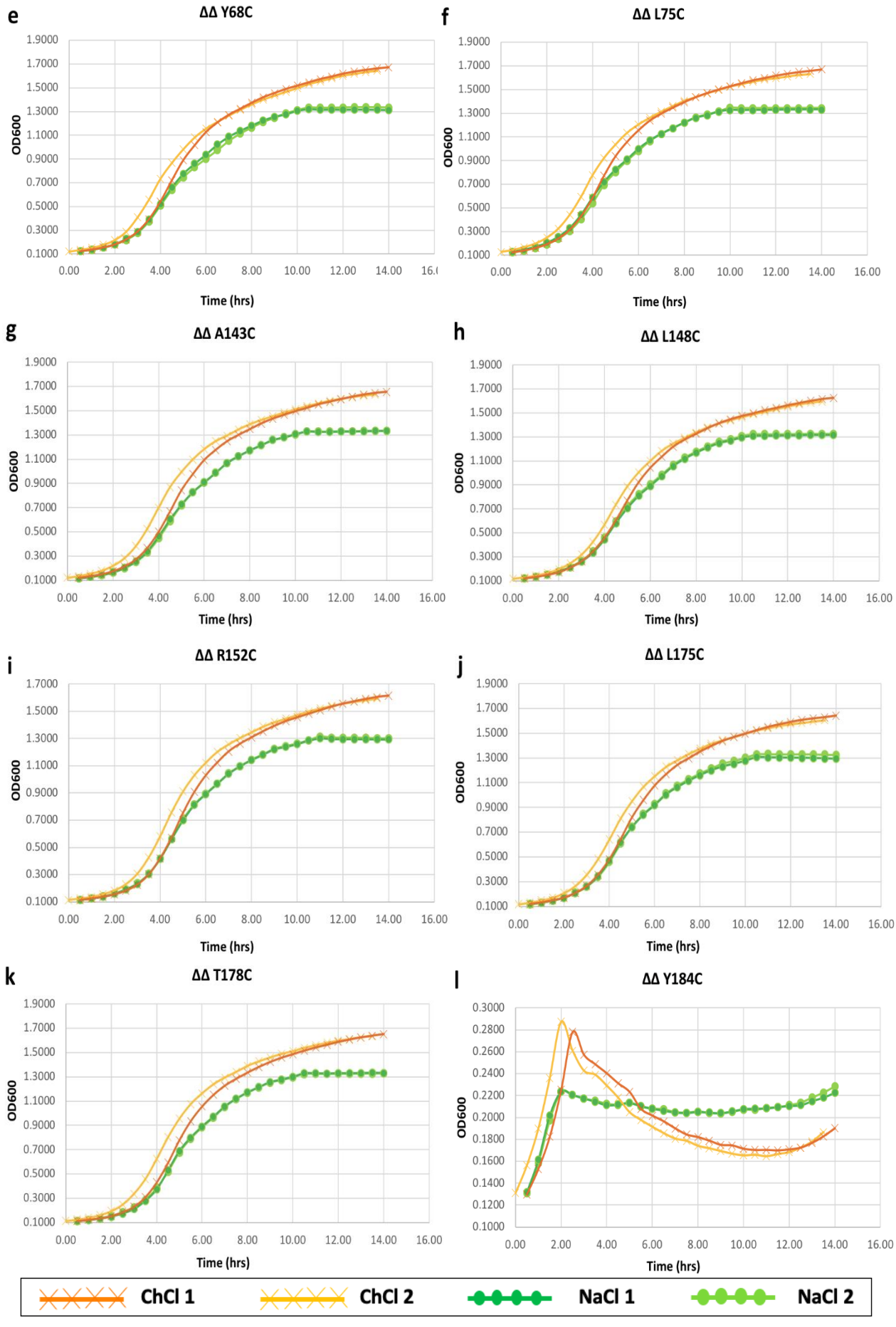

e
ΔΔ Y68C
f
ΔΔ L75C
g
ΔΔ A143C
h
ΔΔ L148C
i
ΔΔ R152C
j
ΔΔ L175C
k
ΔΔ T178C
l
ΔΔ Y184C
OD600
Time (hrs)
ChCl 1
ChCl 2
NaCl 1
NaCl 2

Figure 4 – Plate reader rescue assay for single cysteine mutants of YqjA in control liquid LB and low sodium (100% ChCl molarity replacement) liquid LB.
OD600 readings were taken every half an hour over a 14 hour period and there were duplicate technical repeats of each condition for each mutant. The NaCl condition (control) is shown in orange with the data points as crosses and the ChCl condition (stress) is shown in green with the data points as dots. The control strains were WT Gltph, ΔΔ Gltph and ΔΔ YqjA. This experiment was onlv carried out once.

All single cysteine mutants produced for use in structural analysis needed to be tested to ensure they are functional and this functionality then compared to that of the WT YqjA protein to see if the mutants are able to rescue the growth to a similar level. Functional proteins can be assumed to be properly folded and appropriately inserted into the membrane and it is important to ascertain whether the Cys mutations introduced into YqjA are affecting these processes. An incorrectly folded protein will have a different structure and possibly be inserted differently into the membrane, with the residues in different locations to that in the native protein. As a result, the biochemical data obtained using SCAM would be inaccurate because this data is used to determine the location of residues in WT YqjA relative to the membrane. In summary, a mutated YqjA that has a similar functionality to WT YqjA, in terms of rescue of growth of the ΔΔ under certain stress conditions, is likely folded and inserted into the membrane in the same way as the WT protein.

A rescue assay was used to determine whether the single Cys mutants of YqjA were able to recover the growth of ΔΔ cells under a condition known to be non-permissive to the growth of this strain and to what extent this rescue occurred i.e. to WT level or not. The phenotype tested here was the ability to grow in low sodium conditions using a complete molarity replacement of NaCl with ChCl in the liquid LB in which the cells grew. The samples were aliquoted into a plate reader at 30 °C, medium rotation and OD_{600} readings taken every half an hour for 14 hours.

Fig. 4 shows growth curves of nine of the Cys mutants used for other parts of this project (see section 3.3.2.), including controls. The three control cultures (WT Gltph, ΔΔ Gltph and ΔΔ YqjA) did not grow as expected. Their OD_{600} readings were much lower than that of the Cys mutants, throughout the whole experiment. For example, in the control NaCl condition, WT Gltph (Fig. 4a) reached an OD_{600} of 0.254 and ΔΔ Gltph (Fig. 4b) reached 0.230 after 1.5 hours but only 0.353 and 0.285, respectively, after 14 hours. The other control, ΔΔ YqjA (Fig. 4c), obtained an OD_{600} of 0.227 after 1.5 hours, but subsequently decreased to a low of 0.202 at 10.5 hours and then had a high of 0.238 after 14 hours. Conversely, ΔΔ YqjAA46C (Fig. 4d) produced a traditional growth curve with lag (0 - 3 hours), log (3 - 8.5 hours) and stationary phases and reached an OD_{600} of 1.356 after 14 hours. All other Cys mutants (Fig. 4e – 4k) showed a similar pattern and peaked at an OD_{600} of around 1.30 - 1.35 after 14 hours – with the exception of Y184C (Fig. 4l). This mutant had low overall absorbance readings, growing to just 0.234, dropping slightly and then peaking again at 0.225 after 14 hours. Its growth curve more closely resembled that of ΔΔ YqjA (Fig. 4c).

In the ChCl condition, WT Gltph (Fig. 4a) steadily increased to an OD_{600} of 0.557 at 6.5 hours before dropping slightly to 0.502 at 8.5 hours and, after 14 hours, the reading was 0.519. Thus, this strain was unable to grow again to the level it attained initially. It is unclear why this occurred as it should have been able to grow well in these conditions. ΔΔ Gltph (Fig. 4b)

grew in a similar pattern to an OD_{600} of 0.328 after 14 hours. This strain was not expected to grow in the ChCl condition, though, so the low OD_{600} readings are not surprising. What is surprising is the fact that the OD600 readings for all three controls (Fig. 5a, b and c) are higher for the ChCl condition than the NaCl control condition in which all strains and mutants should have been able to grow well. The growth curve of ΔΔ YqjA (Fig. 4c) in this condition was similar to its NaCl curve, reaching an OD_{600} of 0.322 after 3 hours, dropping to 0.239 at 12.5 hours and then growing to 0.254 at 14 hours. In contrast, and like the NaCl condition, ΔΔ YqjAA46C and all other mutants (Fig. 4d – 4k), apart from Y184C (Fig. 4l), displayed a standard growth curve that peaked after 14 hours with an OD_{600} between 1.55 and 1.65. This is slightly higher than for the NaCl condition. Y184C took 2.5 hours to reach an OD_{600} 0.283, dropped to 0.167 after 11.5 hours, then increased to just 0.188 by 14 hours. Therefore, it peaked higher than in the NaCl condition but did not display the traditional growth curve.

Overall, these growth curves (Fig. 4d – 4k) show that all the single Cys mutants of YqjA (besides Y184C, Fig. 4l) were able to rescue the phenotype of the ΔΔ and so are therefore functional. From this, it can be assumed that the mutants are properly folded and have been correctly inserted into the membrane. Unfortunately, because the irregular growth curve for WT YqjA (Fig. 4c) suggested that the WT protein itself was not functioning well or as expected in this assay, it is unknown whether the structure of the mutant proteins relative to the membrane is the same as the WT. Consequently, any SCAM data cannot be fully relied upon until this rescue assay is repeated and the functionality of the single Cys mutants appropriately compared to the WT protein. If similar results are obtained, further investigation will be required to determine the comparability of the Cys mutants to WT YqjA.

3.3. Structural Analysis of YqjA

3.3.1. EV Fold Modelling

Computational modelling is a vital tool in the prediction of protein structure, especially with membrane proteins or other proteins that are difficult to purify in their native state. Determining the structure of proteins using high resolution techniques such as electron microscopy (including cryo-EM) and X-ray crystallography is preferable but not without its complications (e.g. protein size and ability to crystallise) and is very expensive. There is little cost associated with computational protein modelling and often simply requires entering the amino acid sequence into a pre-established server which then runs a pre-designed algorithm.

The EVcouplings server performs a multiple sequence alignment with the protein sequence entered (specifically α-helical transmembrane proteins) following evolutionary constraints (Hopf, 2012). It relies on the hypothesis that key interactions between residues that aid the proper functioning or maintain the structure of a protein are highly likely to be conserved throughout evolution. There will be a limited number of possible mutations at these sites such as when there is a charge dependence. This alignment of thousands of evolutionarily related sequences enables *de novo* prediction of the 3D structure of the protein of interest. It does not use fragments, threading, or homologous 3D structures.

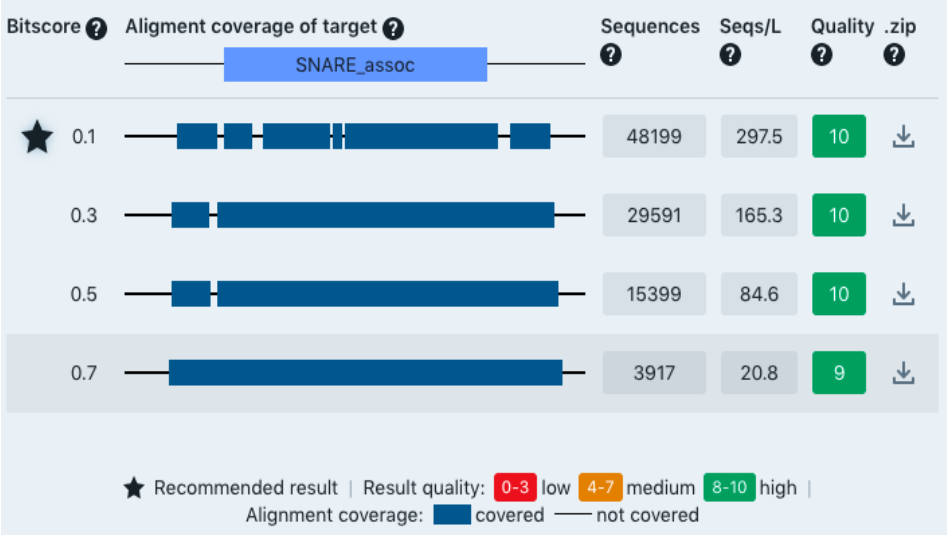


Figure 5 – Initial results for YqjA generated from the EVcouplings server. The four possible results sets, each with varying statistics such as bit score, alignment coverage and number of sequences aligned.

Four results were generated by the server with bit scores of 0.1, 0.3, 0.5 and 0.7. With increasing bit score, the number of sequence homologues to YqjA decreased but the quality of all the results was high (9 - 10). There were varying degrees of alignment coverage but the 0.7 bit score was the only result with full coverage (Fig. 5). This could be visualised when comparing the predicted 3D structures produced from each alignment result as, in the first three sets of predicted structures, the C-terminal helix was projecting away from the rest of the protein (Fig. 6a, b and c), whereas in the predicted structures for the 0.7 bit score, it was arranged in a more probable position – in line with the other helices (Fig. 6d). As well as the positioning of this helix, the rest of the protein was spread out over a large area in the first three model sets and there was more variation between each possible model. This was not the case for the 0.7 bit score which was more compact and each predicted model appeared to be in agreeance – only one possible structure had the C-terminal helix projecting out. Moreover, this result had 188 evolutionary couple positions while the others had 162, 179 and 182, respectively.

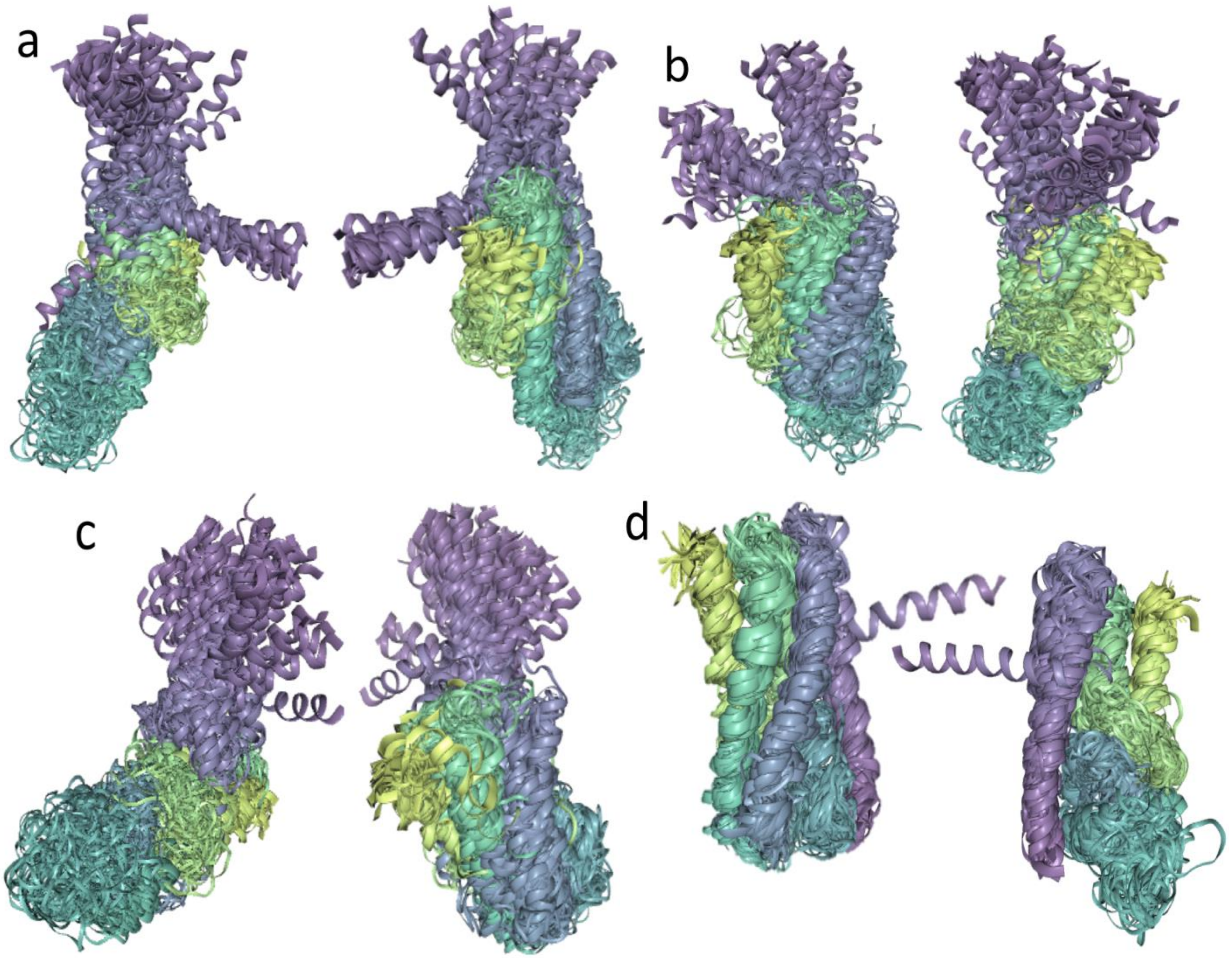


Figure 6 – Predicted models of YqjA for each result generated from the EVcouplings server. All possible models in each set are mapped on top of each other and the 4 sets of models represent the different bit score results (0.1, 0.3, 0.5 and 0.7 from 4a to 4d). Each model set is shown from either side of the model (~180° rotation) and individual predicted helices or partial helices are distinguished by different colours.

Following evaluation of the above information, the highest scoring model (named 124_9 and with a blind quality score of 0.69) in the 0.7 bit score results set was chosen to represent a possible 3D structure of YqjA (see Fig. 8). This is the most likely model out of those predicted for this bit score. In contrast with previous predictions based on hydropathy analysis (Scarsbrook *et al.*, 2021), which predicted five transmembrane helices (see Fig. 7), this model has just three transmembrane helices but also two re-entrant hairpin helices in its structure. From the N-terminus to C-terminus, the EV fold model adopts this structure (Fig. 8): re-entrant helix 1 (blue), TMH1, membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. The essential acidic residues (E39 and D51, red in Fig. 8) are located on the first of these partial transmembrane helices and the important basic residues (R130 and R136, pink in Fig. 8) are on the second.

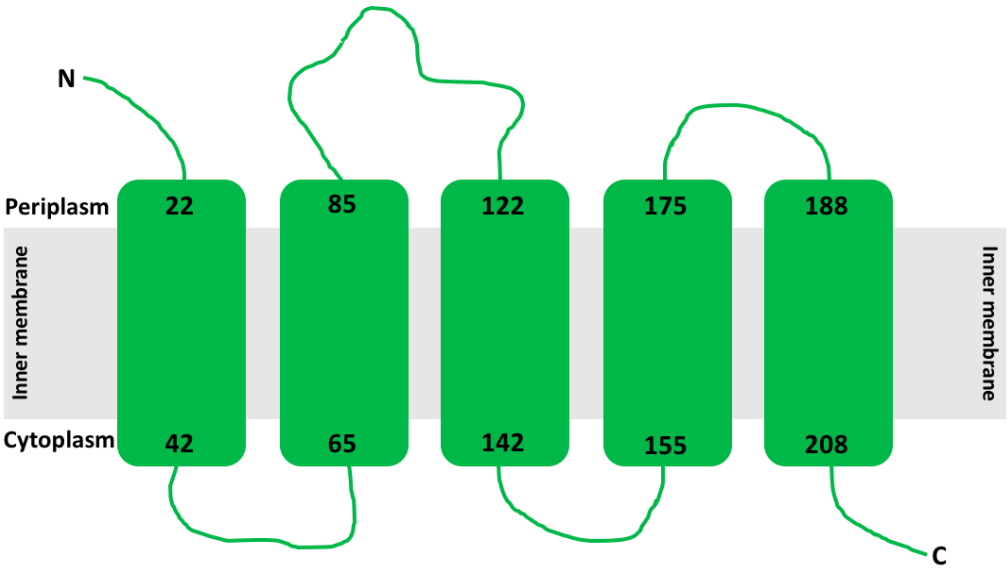


Figure 7 – 2D topology model of YqjA generated from previous computational data collected in our lab (Scarsbrook *et al.*, 2021).

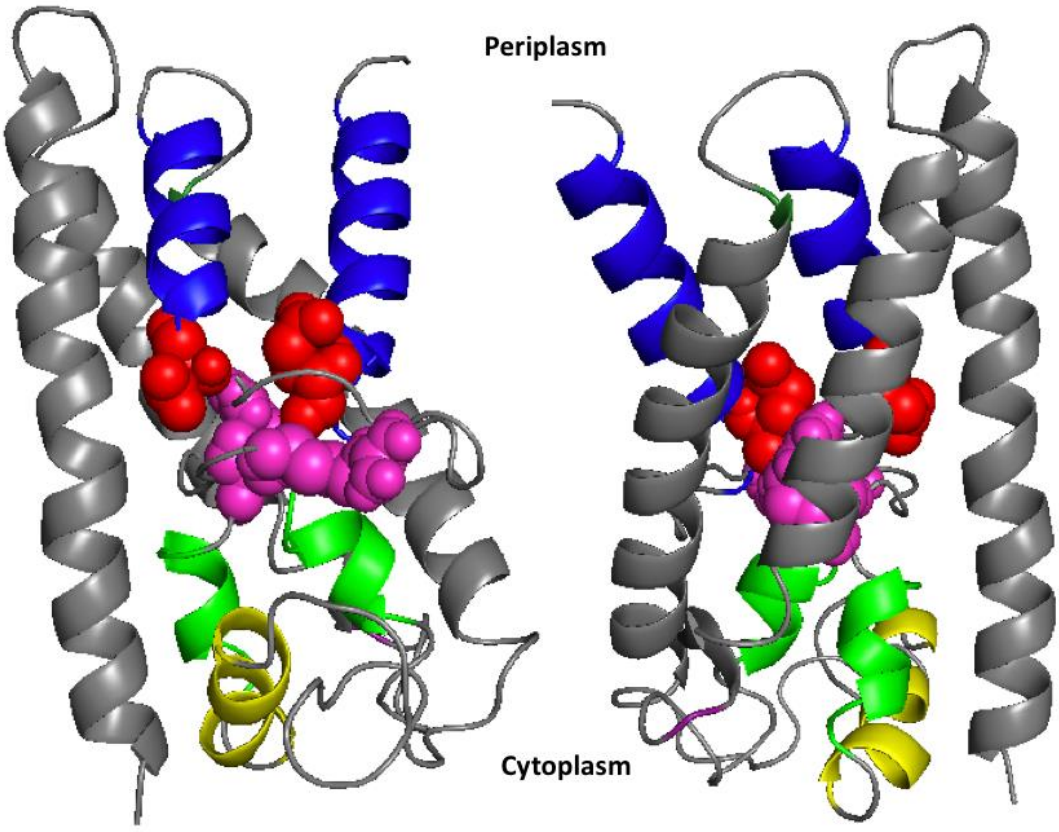


Figure 8 – Pymol structure of the EV fold model of YqjA showing the main structural features and important residues.

The structure consists of re-entrant helix 1 (blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. Essential residues E39 and D51 are in red and important residues R130 and R136 are in pink. The model is shown from two different angles (~180° rotation). This image was generated in PyMol.

Each partial helix that forms the two re-entrant helices features a helix-breaking residue, typically proline and glycine, at its boundary showing that this model fits some well-established protein patterns. The sidechain of proline is a ring making it too rigid and bulky and the sidechain of glycine is too flexible to form part of an organised structure so they disrupt (break) alpha helices (Imai and Mitaku, 2005). Residues P22/G26 and G41 (first half of re-entrant helix 1); P49/G50 and G64 (second half of re-entrant helix 1); G122 and G129 (first half of re-entrant helix 2) and G144/G147 (second half of re-entrant helix 2) of this model fit this pattern and are highlighted in the YqjA amino acid sequence in Figure 9.

```
        10         20         30         40         50
MELLTQLLQA LWAQDFETLA NPSMIGMLYF VLFVILFLEN GLLPAAFLPG
        60         70         80         90        100
DSLLVLVGVL IAKGAMGYPQ TILLLTVAAS LGCWVSYIQG RWLGNTRTVQ
       110        120        130        140        150
NWLSHLPAHY HQRAHHLFHK HGLSALLIGR FIAFVRTLLP TIAGLSGLNN
       160        170        180        190        200
ARFQFFNWMS GLLWVLILTT LGYMLGKTPV FLKYEDQLMS CLMLLPVVLL
       210        220
VFGLAGSLVV LWKKKYGNRG
```

Figure 9 – Amino acid sequence of YqjA with the main structural features identified by the EV fold model and helix-breaking residues highlighted.
The sequence is coloured as follows: re-entrant helix 1 (blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. Essential residues E39 and D51 are in red and important residues R130 and R136 are in pink. Helix breakers are highlighted green.

With the 2D topology model (Fig. 7; Scarsbrook *et al.*, 2021), it was found that it did not follow the positive inside rule (von Heijne, 1986) which states that there is a higher proportion of positively charged residues like arginine and lysine on the inside (cytoplasmic side) of the protein compared to the outside (periplasmic side). The same analysis was undertaken for the EV fold model produced in this study. K63 is on the outside in the EV fold model but was previously predicted to be inside. The opposite is true of R91, R97, R113 and K120 which were predicted to be outside based on the hydropathy model but inside on the EV fold model. R130 and R136 are located on the first re-entrant helix in the EV fold model which correlates with its location in TMH3 in the previous model. This re-entrant helix is positioned towards the cytoplasmic side of the membrane. On both models, R152 is on the inside. As with the hydropathy model, K177 and K183 are on the outside of the membrane, against the positive inside rule. The EVcouplings server did not align residues 210 - 220 and so K213, K214, K215 and R219 are not present on this model. They were located inside the membrane on the previous model and it is likely that they would also be inside on this model as the C-terminus is cytoplasmic. The locations of the positive residues compared with previously predicted transmembrane helices are shown in Figure 10. The same comparison alongside the EV fold model data can be seen in Figure 11 and Figure 12. There are six positive residues inside (cytoplasmic), two transmembrane, and six outside (periplasmic) in the 2D topology model. Conversely, in the EV fold model there are four positive residues inside (along with four not on this model but highly likely to be inside), two transmembrane and three outside. From the above data and visualised in Figure 12, it appears that, in the EV fold model of YqjA, this protein is in fact following the positive inside rule, contradicting previous modelling efforts.

```
        10          20          30          40          50
MELLTQLLQA LWAQDFETLA NPSMIGMLYF VLFVILFLEN GLLPAAFLPG
        60          70          80          90         100
DSLLVLVGVL IAKGAMGYPQ TILLLTVAAS LGCWVSYIQG RWLGNTRTVQ
       110         120         130         140         150
NWLSHLPAHY HQRAHHLFHK HGLSALLIGR FIAFVRTLLP TIAGLSGLNN
       160         170         180         190         200
ARFQFFNWMS GLLWVLILTT LGYMLGKTPV FLKYEDQLMS CLMLLPVVLL
       210         220
VFGLAGSLVV LWKKKYGNRG
```

Figure 10 – YqjA amino acid sequence with TMH predicted from TOPCONS and positive residues highlighted (Scarsbrook *et al.*, 2021).
Transmembrane helices are coloured red and the positive residues are highlighted yellow.

```
        10          20          30          40          50
MELLTQLLQA LWAQDFETLA NPSMIGMLYF VLFVILFLEN GLLPAAFLPG
        60          70          80          90         100
DSLLVLVGVL IAKGAMGYPQ TILLLTVAAS LGCWVSYIQG RWLGNTRTVQ
       110         120         130         140         150
NWLSHLPAHY HQRAHHLFHK HGLSALLIGR FIAFVRTLLP TIAGLSGLNN
       160         170         180         190         200
ARFQFFNWMS GLLWVLILTT LGYMLGKTPV FLKYEDQLMS CLMLLPVVLL
       210         220
VFGLAGSLVV LWKKKYGNRG
```

Figure 11 – Amino acid sequence of YqjA with main structural features identified by the EV fold model and positive residues highlighted.
The sequence is coloured as follows: re-entrant helix 1 (blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. Essential residues E39 and D51 are in red and important residues R130 and R136 are in pink. Positive residues are highlighted yellow.

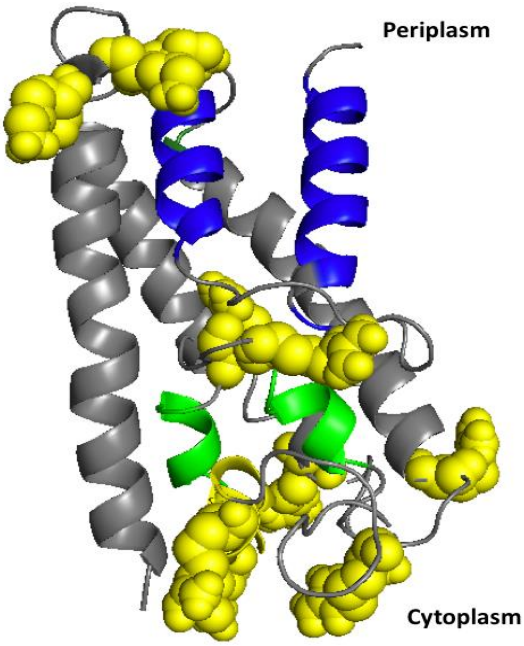


Figure 12 – Pymol structure of EV fold model of YqjA showing the main structural features and positive residues.
The structure consists of re-entrant helix 1 (blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. The positive residues – lysine (K) and arginine (R) – are in yellow. This image was generated in PyMol.

3.3.2. Substituted Cysteine Accessibility Methods (SCAM)

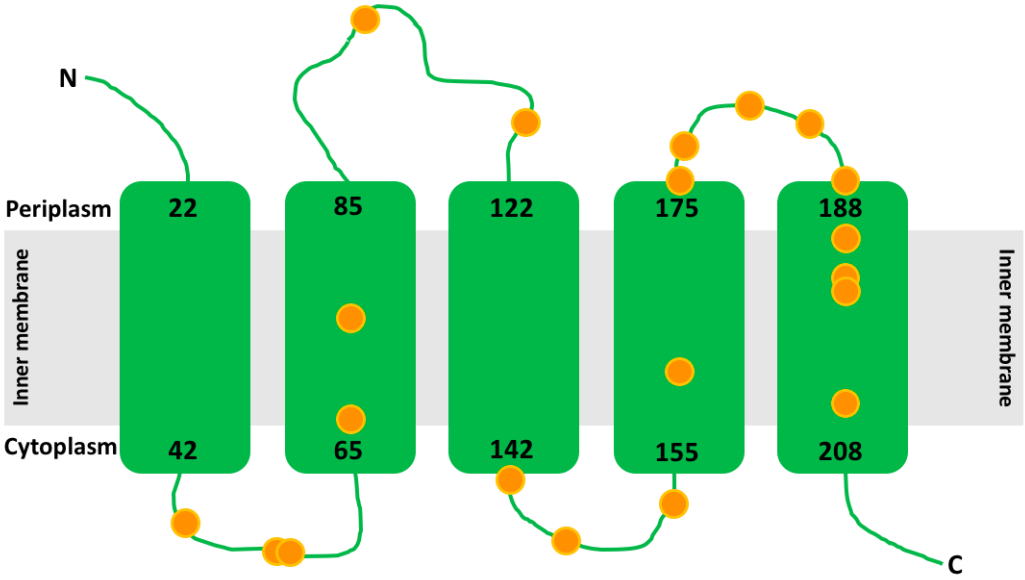


Figure 13 – 2D topology model of YqjA generated from previous computational data collected in our lab (Scarsbrook *et al., 2021*) with single cysteine mutants tested by SCAM analysis in the current project mapped on.

SCAM is used to determine the location of specific residues in relation to the inner bacterial membrane i.e. periplasmic, cytoplasmic or transmembrane. Single cysteine mutants of the YqjA protein were produced in a Cys-less background – the mutants tested in this project can be seen mapped onto the previous 2D topological map in Figure 13. This model was used to choose mutants to be tested and subsequently the EVfold model was used alongside.

The cells were treated with different thiol-blocking reagents which bind specifically to Cys residues. The reagents used in this study were (2-sulfonatoethyl)methanethiosulfonate (MTSES) and N-Ethylmaleimide (NEM). MTSES is permeable only to the outer bacterial membrane but NEM is permeable to both the outer and inner bacterial membranes. This means that NEM will bind to all Cys residues in the protein but MTSES can only bind Cys residues accessible from the periplasm. Following cell lysis, methoxy-poly(ethylene glycol)-maleimide (mPEG5K) was added which also binds to the thiol groups present in Cys residues. It will therefore bind to any unlabelled cysteines i.e. those that were inaccessible to the blocking reagents. It will increase the mass of the protein by 5 kDa which can be clearly observed using Western blotting. The change in mass when MTSES or NEM bind is insignificant. The Western blot data is used to show which of the thiol-blocking reagents were able to access and bind to the specific Cys residues subsequently preventing the binding of mPEG5K. Based on the known membrane permeability of MTSES and NEM, the whereabouts of each Cys residue in relation to the membrane can be inferred.

The four sample conditions prepared were: a negative control with no MTSES, NEM or mPEG5K; an mPEG5K positive control with neither MTSES or NEM; a sample with MTSES and mPEG5K and a sample with NEM and mPEG5K. A residue that is accessible from the periplasm will be protected by both MTSES and NEM and thus will be PEGylated in the

mPEG5K control alone. If the residue is accessible from the cytoplasm, it will be protected by NEM only and hence will be PEGylated in the mPEG5K control and the MTSES condition. Depending on their accessibility, some residues may only be partially protected by MTSES and NEM. These residues tend to be transmembrane.

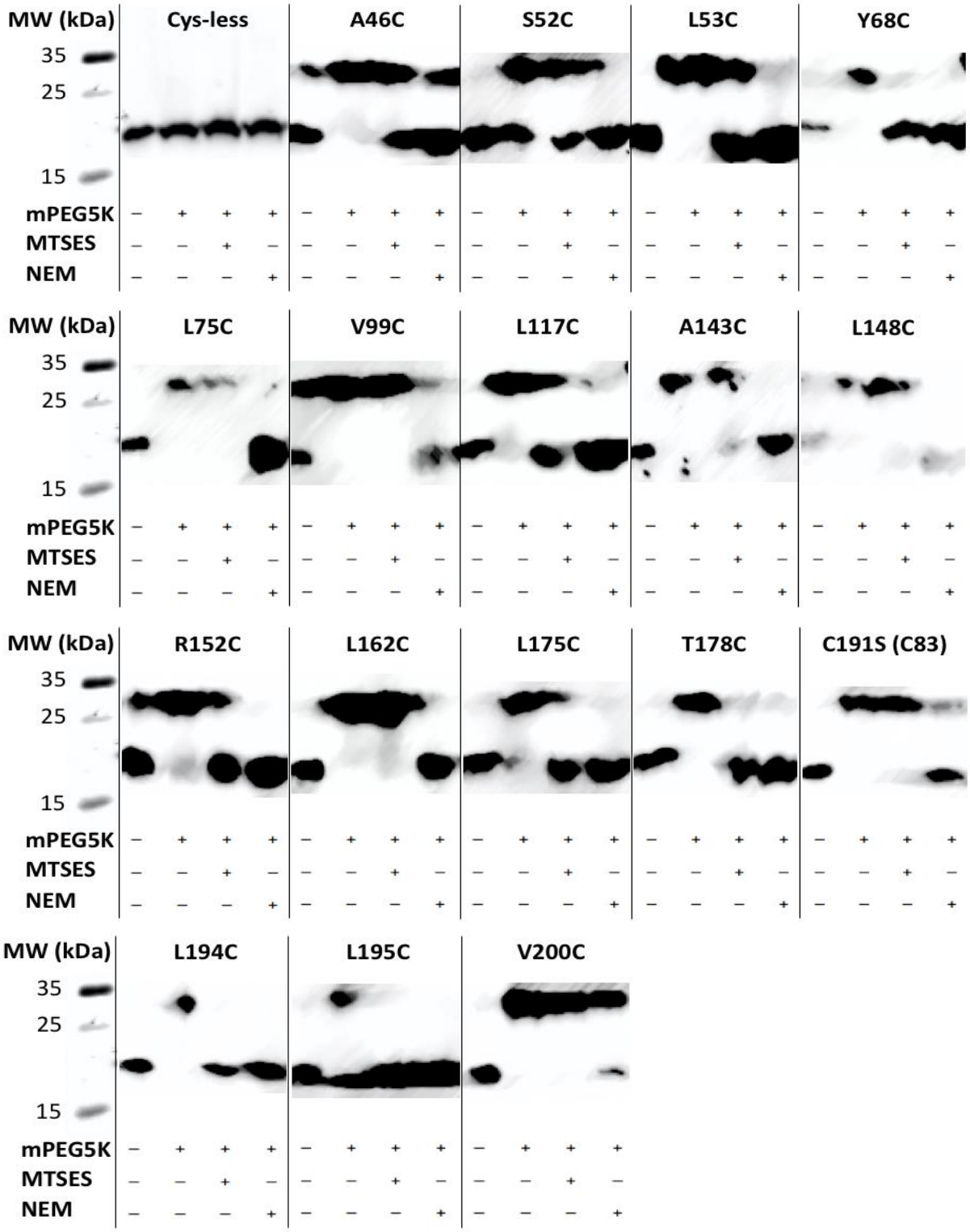


Figure 14 – Western blot data from substituted cysteine accessibility methods (SCAM) of single cysteine mutants of YqjA.
For each mutant there are four samples that were treated in different ways. The reagents added to each sample are indicated by a + or a –. At least three biological repeats were carried out for each mutant and this data is representative of all results obtained.

The Cys-less mutant acts as a control for this experiment as there are no Cys residues for the reagents to react with. There was no PEGylation across all conditions (see Fig. 14), as expected, showing that this cannot occur in the absence of Cys residues. Thus, the Cys-less mutant is an appropriate negative control with the unPEGylated bands being around 18 kDa in size. A46C seems to be fully PEGylated in the mPEG5K control condition and partially PEGylated in both MTSES and NEM conditions. On the full Western blot (not pictured), it appears that the PEGylation visible in the NEM condition of A46C could be overspill from the lane to its right. S52C and L53C are partially protected by MTSES and fully protected by NEM along with the standard PEGylation in the mPEG5K control. It is impossible to make a definitive conclusion from these results as to the location of A46, S52 and L53.

There was full protection from both MTSES and NEM for Y68C (Fig. 14) predicting it to be periplasmic. Both L75C and V99C showed PEGylation in the mPEG5K control and MTSES condition, as did A143C (albeit partial) and L148C. In contrast, L117C was protected by MTSES as well as NEM and so was only PEGylated in the mPEG5K control. It can be concluded from this data that L75, V99, A143 and L148 are cytoplasmic residues but the exact position of L117C in the protein is unclear.

With R152C (Fig. 14), there is full PEGylation in the mPEG5K control but only partial in the MTSES condition. It is possible that this is overspill from the mPEG5K control lane as there also appears to be a small amount of PEGylation in the negative control which would not be expected. This cannot be assumed, however, so the location of R152 remains inconclusive. MTSES was unable to protect L162C from PEGylation but almost fully protected L175C and fully protected T178C. All three of these mutants were fully protected from PEGylation by NEM and thus L162 was predicted to be in the cytoplasm and L175 and T178 in the periplasm. The C191S mutant showed complete PEGylation in the mPEG5K control and MTSES condition but only partial PEGylation in the NEM condition implying that it is in the transmembrane region of the protein. The cysteine residue present in this mutant is C83, the other native cysteine, it is this residue that the thiol-binding reagents are reacting with, producing the pattern of results seen in the Western blot. Therefore, these results represent the location of C83, not C191.

There was PEGylation in the mPEG5K control alone for L194C and L195C (Fig. 14), predicting them to be periplasmic, although for L195C this was possibly not full PEGylation – it is unclear why. The V200C mutant was PEGylated in all three conditions with a slight partial PEGylation in the NEM condition and so it is likely to be transmembrane.

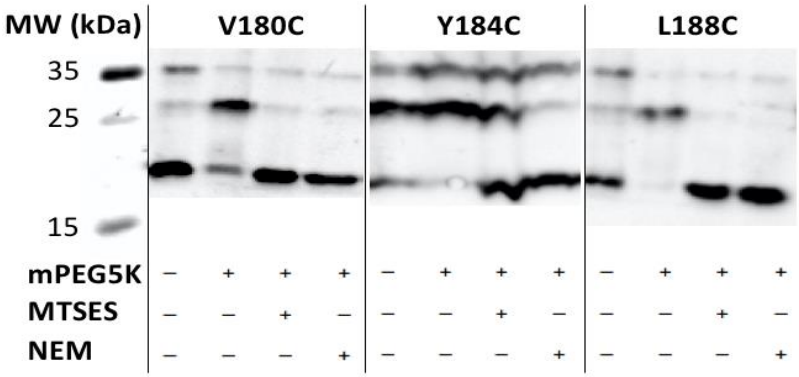


Figure 15 – Western blot data from substituted cysteine accessibility methods (SCAM) of single cysteine mutants V180C, Y184C and L188C of YqjA.
For each mutant there are four samples that were treated in different ways. The reagents added to each sample are indicated by a + or –. At least three biological repeats were carried out for each mutant and this data is representative of all results obtained.

Three of the mutants (V180C, Y184C and L188C) displayed an unusual pattern of results (see Fig. 15). For all four conditions, including the negative control, alongside the bands of unPEGylated size (~18 kDa) and PEGylated size (~26 kDa), there were other bands present at ~36 kDa. These higher MW bands are fainter than the other bands but are still well-defined. For V180C and L188C, the bands that are most distinct show PEGylation in the mPEG5K control and almost full protection in the MTSES and NEM conditions – the pattern of results associated with periplasmic residues. Y184C does not present a familiar pattern of bands and there is a considerable amount of PEGylation in the negative control compared with V180C and L188C. Overall, this means that the data for these residues is inconclusive.

The mutants S52C, L53C, V99C, L117C, L139C, L162C, V180C, C191S and V200C have been previously tested in our lab using SCAM (Scarsbrook *et al.*, 2021). Most of the data collected in this project agrees with the previous data and so confirms these findings. The exception is V180C which previously showed PEGylation in the mPEG5K control only, implying it is periplasmic. The data for V180C in Figure 15 is ambiguous but could suggest that V180 is periplasmic like has been shown before.

3.3.3. SCAM Data and the EV Fold Model

When mapped onto the EV fold model (see Fig. 16), 8 out of 20 residues show similar locations to that predicted by SCAM analysis (see Table 7). 8 of the Cys mutants produced inconclusive SCAM data and so it cannot be determined whether the EV fold model is accurate for these residues. Of the other 4 residues, L75C appeared to be periplasmic in the EV fold model but cytoplasmic based on the SCAM data and L162, L194 and L195 were all in a transmembrane region in the EV fold model but L162 was predicted to be in the cytoplasm and the other two in the periplasm using SCAM. One explanation for this is that, if a residue is accessible to MTSES, NEM and mPEG5K, it may not necessarily be in the cytoplasm or periplasm, just accessible from either of these sides of the membrane and not protected inside the membrane i.e. in a binding pocket. They appear to be the same distance into the membrane on TMH2 (L162) and THM3 (L194 and L195).

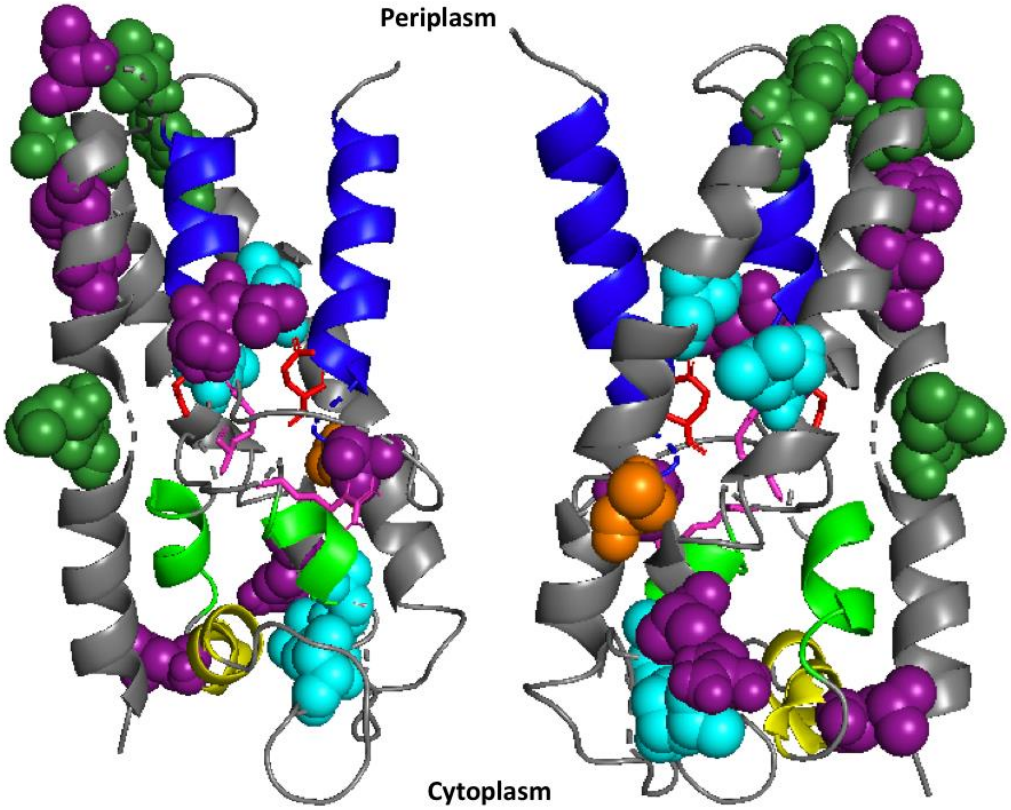


Figure 16 – SCAM data mapped onto EV fold model of YqjA.
Key structural features are identified: re-entrant helix 1 (dark blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (light green), TMH2 and TMH3. Residues are predicted to be either periplasmic, cytoplasmic or transmembrane by the SCAM analysis. Predicted periplasmic residues are coloured dark green, predicted cytoplasmic residues are coloured light blue, predicted transmembrane residues are coloured orange and inconclusive results are coloured purple. The SCAM residues are shown as spheres but important residues E39 and D51 (red) and R130 and R136 (pink) are shown as sticks. This image was generated in PyMol.

YqjA Mutant	Protected by MTSES?	Protected by NEM?	Location predicted by EV Fold	Location Predicted by SCAM
Cys-less	–	–	–	–
A46C	Partial	Partial	Transmembrane	Inconclusive
S52C	Partial	Yes	Transmembrane	Inconclusive
L53C	Partial	Yes	Transmembrane	Inconclusive
Y68C	Yes	Yes	Periplasm	Periplasm
L75C	No	Yes	Periplasm	Cytoplasm
V99C	No	Yes	Cytoplasm	Cytoplasm
L117C	Partial	Yes	Cytoplasm	Inconclusive
A143C	Partial	Yes	Cytoplasm	Cytoplasm
L148C	No	Yes	Cytoplasm	Cytoplasm
R152C	Partial	Yes	Cytoplasm	Inconclusive

L162C	No	Yes	Transmembrane	Cytoplasm
L175C	Partial	Yes	Periplasm	Periplasm
T178C	Yes	Yes	Periplasm	Periplasm
V180C	Partial	Partial	Periplasm	Inconclusive
Y184C	Partial	Partial	Periplasm	Inconclusive
L188C	Partial	Partial	Transmembrane	Inconclusive
C191S (C83)	No	Partial	Transmembrane	Transmembrane
L194C	Yes	Yes	Transmembrane	Periplasm
L195C	Yes	Yes	Transmembrane	Periplasm
V200C	No	Partial	Transmembrane	Transmembrane

Table 7 – Predicted locations of amino acid residues in YqjA based on an EV fold modelling and SCAM analysis.
This table outlines whether each single cysteine mutant was protected by the thiol-blocking agents MTSES and NEM (visualised by Western blotting) and the corresponding location predicted by SCAM analysis. It compares this predicted location with that predicted by the EV fold model produced by the EVcouplings server from evolutionarily constrained multiple sequence alignments. The colours correspond with those in the Pymol image in Fig. 16.

4. Discussion

A substantial amount of data, both biochemical and computational, was produced in this study regarding the structure of the *E. coli* DedA protein YqjA. Of those tested, the SCAM data predicted five periplasmic, five cytoplasmic and two transmembrane residues in YqjA. The other eight residues tested resulted in inconclusive data. The EVcouplings server produced a predicted model of YqjA with three transmembrane helices and two re-entrant helices. When mapped onto this model, the residues tested using SCAM were predicted as follows: six periplasmic, five cytoplasmic and nine transmembrane residues with eight out of twelve conclusive SCAM results agreeing with the EV fold model. Each of the Cys mutants was also tested using a rescue assay to determine whether they are functional and properly folded and hence their structure can be suitably compared to WT YqjA.

4.1. Functionality of Single Cysteine Mutants of YqjA

All of the growth curves for the single cysteine mutants (aside from Y184C) in Fig. 5 showed that the cells grew well in both the NaCl control and ChCl stress condition. This implies that all of the mutants of YqjA are functional and properly folded in the membrane allowing them to rescue this phenotype of ΔΔ. However, we cannot definitively make this conclusion as the growth of all the control strains was poor for both conditions and none of them were able to grow to a high enough OD_{600} after 14 hours. All of the mutants (including controls) seemed to grow slightly better in the ChCl, on average, which was not expected. The lack of sodium makes this a stress condition and so even though the cells can grow, they may not be able to grow as well as under optimal conditions i.e. with NaCl in the media. As a result of these unexpected data, the rescue assay needs to be repeated (it was only undertaken once in this project) and the mutants that have not yet been tested need to be included to confirm their functionality and see if the higher growth in low sodium conditions is repeatable. For now, the interpretation of the SCAM data can provisionally be assumed to be accurate for every mutant that was able to recover the growth of the ΔΔ cells in the ChCl condition, which is all but Y184C. If they are able to rescue the growth of the ΔΔ under this stress condition, as has been observed in the past with WT YqjA, then it is highly likely that these mutants are folded and inserted into the membrane in a way that is similar, if not identical, to the WT protein.

Those mutants for which SCAM analysis has been carried out previously (Scarsbrook *et al.*, 2021) (S52C, L53C, V99C, L117C, L139C, L162C, V180C, C191S and V200C), a rescue assay was carried out in that study using culture spotted onto solid media with various stress conditions (like with the confirmation of ΔΔ formation in this project – see section 3.1). These conditions were: high temperature (42 °C), varying levels of ChCl replacement and 75 µg/ml EtBr. All controls and mutants grew well in the control media (standard LB agar, grown at 30 °C) and ΔΔ Gltph did not grow well under any of the stress conditions, as expected (data not shown). All of the mutants were able to rescue the temperature and EtBr sensitivity phenotypes of the ΔΔ, showing that they are properly folded, inserted into the membrane and functional. WT YqjA and the mutants were unable to rescue the low sodium sensitivity phenotype. Despite this, YqjA has since been shown to be able to recover this phenotype (data not shown). There may be a difference in growth between solid and liquid media, caused by the

environmental differences between the two, because the other single cysteine mutants tested in this project were able to rescue the low sodium sensitivity phenotype of the ΔΔ in the plate reader assay. This is a phenomena observed in studies such as Fortuin *et al.*, 2020 which showed variations in the proteomes of the same strain of *E. coli* grown in liquid and on solid media.

4.2. Structure of YqjA

4.2.1. SCAM Analysis

The residues that were predicted to be periplasmic based on the SCAM analysis were Y68C, L175C, T178C, L194C and L195C and those predicted to be cytoplasmic were L75C, V99C, A143C, L148C and L162C (see Fig. 14 and Table 7).

A couple of the mutants, C191S and V200C, showed SCAM data associated with residues located in transmembrane regions of a protein. They were PEGylated in the mPEG5K control, MTSES and NEM conditions meaning that, for the most part, they are inaccessible from both sides of the membrane. There was a slight protection from NEM so at some point during the hour incubation with the thiol-blocking agents, NEM was able to bind to the Cys residue in a few of the mutant YqjA proteins in the sample.

A few of the mutants remain inconclusive based solely on the SCAM data collected so far. A46C showed partial PEGylation in both the MTSES and NEM conditions. One explanation for this is that it is located in a transmembrane region and so is only partially accessible to the reagents, as described above. This is plausible as A46 is between the two essential acidic residues (E39 and D51) that are predicted to be involved in proton transport. S52C and L53C showed the same pattern of PEGylation as each other with equal parts protection and PEGylation in the MTSES condition and full protection by NEM. This could mean that these residues are located in the membrane but closer to the cytoplasmic side as they are only partially accessible to MTSES. They are also very close in the primary structure to the acidic residue D51. It is possible that structural changes in the protein as it performs its function may change the accessibility of these three residues to the thiol-blocking agents. This theory will be further explored in section 4.2.2. Other mutants that showed a similar result to this were L117C and R152C but R152C appeared to have some PEGylation in the negative control and so needs to be repeated to produce more conclusive data. In fact, all of the mutants with ambiguous data should be repeated to see if more definitive data can be obtained.

Three of the mutants that produced inconclusive data showed a notably different pattern to the rest: V180C, Y184C and L188C (Fig. 15). As V180C and L188C show more well-defined bands in the pattern familiar with a residue located in the periplasm, they may be located towards the periplasmic side of the protein, agreeing with data previously collected (Scarsbrook *et al.*, 2021). The fact that there seems to be PEGylation in the negative control for Y184C means that prediction of the residue's location in the protein relative to the membrane is not possible. Also present in the data for these three mutants is another set of bands at ~36 kDa in all four conditions. This band is around double that of the

unPEGylated band i.e. the unmodified protein. SDS gels are denaturing and so proteins tend to be present in their monomeric forms but if monomers are associated via disulfide bonds then these are not broken by the SDS unless a reducing agent (e.g. β-mercaptoethanol) is added to the gel. An explanation for the ~36 kDa band could be that two monomers of the mutant YqjA proteins have formed dimers. These residues are located near to C191, which is predicted to be involved in oligomerisation (Keller, 2015). During protein folding, the Cys residues in two monomers of each of these mutants must have come into close enough contact to form a disulfide bond as C191 would in the native protein.

If the Y184C mutant is in fact forming a dimer, this could the unusual growth curve produced for both the NaCl and ChCl conditions of the rescue assay (Fig. 4). In order to test this hypothesis, a rescue assay would need to be carried out for V180C and L188C – the other mutants predicted to be forming dimers. As these dimers are forming between different residues to that in the native protein, it may be affecting the functionality of the protein. However, all other mutant YqjA proteins in the rescue assay are functioning as monomers: they were made in a Cys-less background (C83 and C191 have been mutated to serine) and so cannot form disulfide bonds.

4.2.2. Accuracy of the EV Fold model

This project used both computational predictions (EVcouplings server) and biochemical analysis (SCAM) to begin to define the 3D structure of YqjA. They can be used in conjunction with each other by simply highlighting the residues on the Pymol and making a judgement as to where they are in relation to the membrane (see Fig. 16 and Table 7). Through this, it can be inferred the accuracy of the predicted EV fold model. Unfortunately, these methods cannot be fully relied upon to determine the structure of YqjA as they are predictive and based on current knowledge of protein folding. Experimental techniques, such as X-ray crystallography, are required to complement the computational model and be more certain of the 3D structure of YqjA. Of the mutants that produced reliable data, 8 out of 12 had SCAM data that agreed with the EV fold model which implies that this prediction of the 3D structure of YqjA must have a high level of accuracy. Although the mutant V180C produced inconclusive results (like a few other mutants), in previous work (Scarsbrook *et al.*, 2021) it was believed to be periplasmic which agrees with the EV fold model. With more SCAM data across a wider number of Cys mutants, the accuracy of the EV fold model can be tested further.

In the past, it was thought that DedA proteins shared structural features with the Na^+ symporter from *Aquifex aeolicus,* LeuT (Khafizov, 2010). This protein has 12 transmembrane helices, 10 of which form two functional inverted repeats (pseudo 2-fold axis of symmetry in the plane of the membrane) with a re-entrant helix in each repeat that together form the ligand binding site (Yamashita *et al.*, 2005). Although there is little similarity in the primary amino acid sequences of DedA proteins and LeuT, there is similarity in their hydropathy profiles. Since this hypothesis was suggested, there has been little evidence that the DedA family share a common ancestor with LeuT. More structural studies have led to this hypothesis being disregarded.

The EVcouplings server results from this study predicted that YqjA has two re-entrant helices (see Fig. 8). This could have been where the assumption arose that YqjA adopts a LeuT fold arose due to the fact that the two re-entrant helices in LeuT form the ligand binding site for both leucine and two Na^+ ions. However, the EV fold model of YqjA is structurally very different from the LeuT fold and so it does not seem that the DedA family and LeuT are structurally related. In addition, E39 and D51, that are functionally essential to the protein and are thought to be involved in proton transport, are located on the first re-entrant helix in the EV fold model. R130 and R136, the other important (but not essential) residues, can be found on the second re-entrant helix, in close proximity to the acidic residues. Arginine has been shown to be involved in proton transport (Sigal, 2005) but this has not been observed in YqjA or any other DedA proteins, only in the *E. coli* multi-drug transporter MdfA.

Okawa *et al.* (2020) defined the DedA domain (found across the entire DedA superfamily), using trRoetta and EVfold, as re-entrant loop 1, TMH1, an extra-membrane region, re-entrant loop 2 and TMH2 (Fig. 17) which correlates with the results seen in this study for YqjA (with this protein having a further C-terminal TMH following the DedA domain). 3 out of 4 sets of possible models for YqjA (bit scores 0.1, 0.3 and 0.5) predicted the C-terminal helix to be directed away from the rest of the protein, further indicating that the DedA domain is present in YqjA and is made up of the features described above. The C-terminal helix is the only part of the model not in an appropriate position and is TMH3 i.e. the only part of YqjA not in the supposed DedA domain.

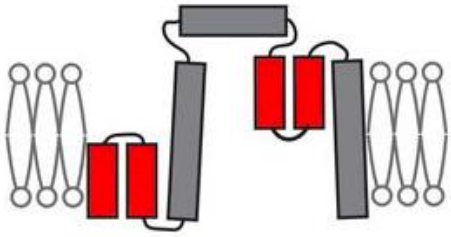


Figure 17 – Topology of the DedA domain as defined by Okawa *et al.* (2020).
This structural feature of DedA proteins consists of re-entrant loop 1, TMH1, an extra-membrane region, re-entrant loop 2 and TMH2. The re-entrant loops are in red.

These results for YqjA have been replicated, with some slight variation, using trRosetta (Mesdaghi *et al.* 2021) with the presence of the DedA domain as described above and the important residues – E39, D51, R130 and R136 – in similar locations on the two re-entrant helices (see Fig. 18). They predicted an extended DedA domain with a membrane-parallel amphipathic helix before each re-entrant helix (the topology of this can be seen in Fig. 19a). In our results, a possible membrane-parallel helix was only observed before the second re-entrant helix but this could simply be the extra-membrane region described by Okawa *et al.*, 2020.

Re-entrant helices have, so far, been found exclusively in transmembrane proteins that function as transporters (Yan and Luo, 2010). This, along with the predicted locations of

the acidic residues, provides substantial evidence that YqjA is a proton transporter. Mesdaghi, 2021 also came to this conclusion and drew parallels with the inverted repeat of re-entrant helices structure in Cl^- and H^+ transporter such as CLC transporter and UppP (Workman *et al.*, 2018) (Fig. 19). Only one of these transporters, 3orgA, had a membrane-parallel helix and so this must be not essential for function or it would be more highly conserved.

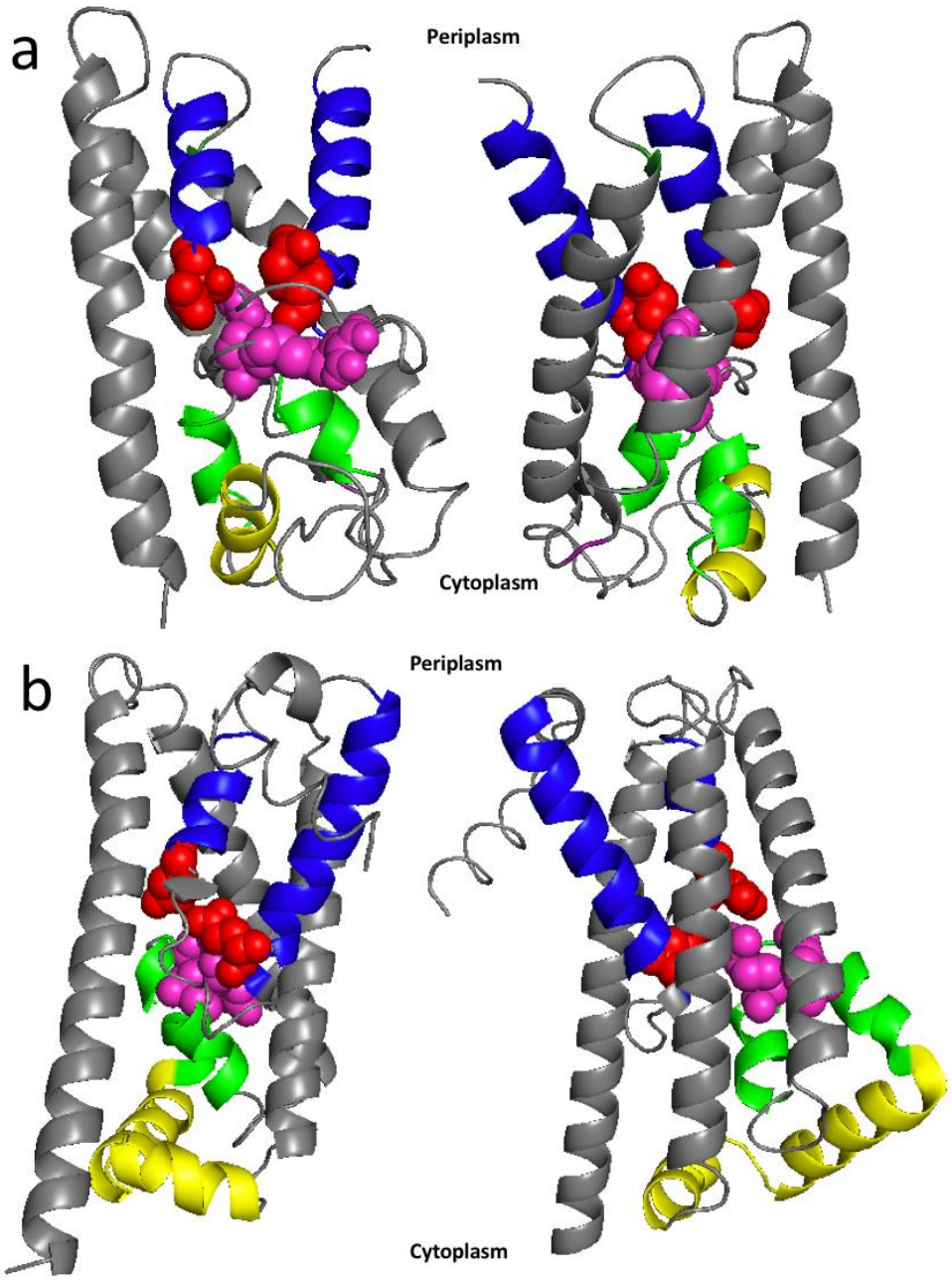


Figure 18 – Pymol structures of two predicted models of YqjA showing the main structural features and important residues: a) EV fold model from the current study and b) trRosetta model generated by Mesdaghi, 2021.
The structures both consist of re-entrant helix 1 (blue), transmembrane helix 1 (TMH1), membrane-parallel helix (yellow), re-entrant helix 2 (green), TMH2 and TMH3. Essential residues E39 and D51 are in red and important residues R130 and R136 are in pink. The models are shown from two different angles (~180° rotation). This imaga was generated in PyMol.

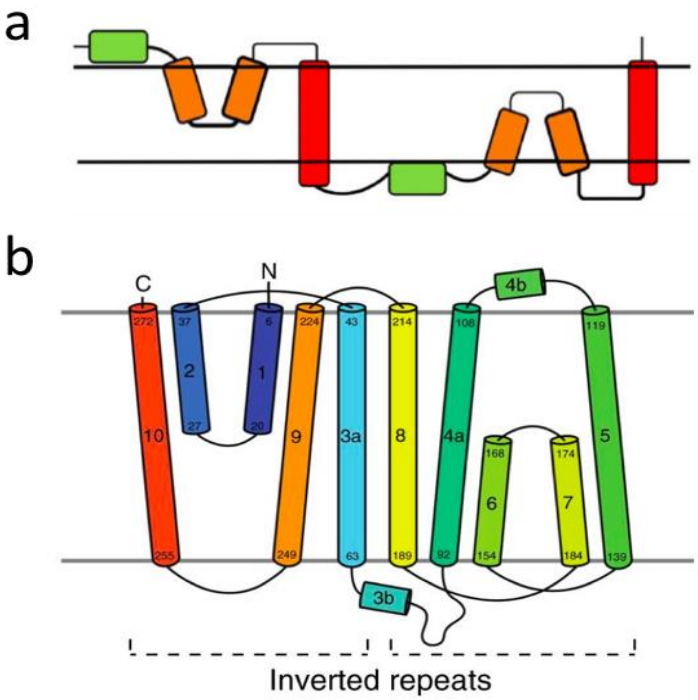


Figure 19 – Comparison of a) predicted topology of the extended DedA domain (Mesdaghi, 2021) and b) topology of *Ec*UppP showing structural similarities.
The inverted repeats in *Ec*UppP are comparable with the two re-entrant helices in the DedA domain and therefore it is possible that DedA proteins perform a similar function to *Ec*UppP – proton transport.

4.3. Future Work

Due to the unexpected results, particularly for the control strains, the rescue assay for the single Cys mutants already tested needs to be repeated with the mutants that are yet to be tested included as well. This will increase the confidence in the SCAM data and enhance its applicability to the structure of WT YqjA.

Now that a more accurate computational model of YqjA has been produced, further experimental investigation will be required to prove its accuracy with the ultimate aim being the production of a crystal or cryo-EM structure of the protein. On top of this 3D structural determination, the oligomeric state also needs to be ascertained. One way of doing this is through size exclusion chromatography multi-angle light scattering (SEC-MALS), with the protein isolated using detergent or another membrane mimetic.

The goal of determining the structure and function of YqjA is to reverse the antibiotic resistance which it has been shown to confer in *E. coli*. If it is understood how it functions then drugs can be developed to block it and prevent it from performing this function. These drugs will be designed with the structure of YqjA in mind so that they bind specifically and

perhaps irreversibly to the protein. Having a robust structural model is vital for the drug design process. An alternative method to refined drug design is to carry out a thorough screen to determine compounds that are able to bind to YqjA. Those that are able to bind could be further characterised to see if they act as inhibitors for YqjA and subsequently cross-screened for toxicity, metabolism etc.

As well as determining the structure of YqjA, the functional characterisation of the protein is also vital, especially as most of the data currently available relies solely on knockout studies of YqjA and YghB (Thompkins *et al*., 2008; Keller *et al*., 2015). The current hypothesis is that YqjA is a proton transporter and in order to prove this proton flux assays could be carried out. One way to undertake this would be to express the GFP variant pHlourin alongside YqjA which is sensitive to changes in pH and has excitation peaks at 395 nm and 475 nm (Mahon, 2011). As the extracellular environment acidifies, there is a decrease in the excitation at 395 nm and an increase in the excitation at 475 nm. This occurs in a ratiometric fashion meaning that the fluorescence of pHlourin can be used directly to determine changes in intracellular pH. If YqjA is able to transport protons, then there should be an influx caused by acidification of the extracellular environment and hence an increase in the electrochemical gradient.

More information is still needed about the individual roles of YqjA and YghB and their interaction with each other for most of the research carried out up to now has been focussed on the ΔΔ. It is important to understand how they both confer antibiotic resistance and if both are required because both may need to be blocked in order to reverse the resistance. If this is the case then similar structural investigation would need to be carried out for YghB in order to produce an accurate structural model.

4.4. Conclusion

In this study, it was found that YqjA consists of three transmembrane helices and two re-entrant helices, unlike the previous prediction of five TMH. These partial helices are arranged in an inverted conformation. The essential residues, E39 and D51, are located on re-entrant helix 1 and, as they are thought to be involved in proton transport in this protein, it can be hypothesised that the re-entrant helices come together in the membrane to form a pore to facilitate this transport.

5. Acknowledgements

I would like to thank my supervisor Dr Christopher Mulligan for providing me with all the guidance, support and training I needed throughout my project. I would also like to thank my fellow members of Mulligan Laboratory, Matthew Batson and Nicholas Massouh, for their help and support.

6. References

Almeida, J., Preto, A., Koukos, P., Bonvin, A. and Moreira, I., 2017. Membrane proteins structures: A review on computational modeling tools. Biochimica et Biophysica Acta (BBA) - Biomembranes, 1859(10), pp.2021-2039.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K., Tomita, M., Wanner, B. and Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology, 2(1).

Bakheet, T. and Doig, A., 2010. Properties and identification of antibiotic drug targets. BMC Bioinformatics, 11(1).

Boughner, L. and Doerrler, W., 2012. Multiple deletions reveal the essentiality of the DedA membrane protein family in *Escherichia coli*. Microbiology, 158(5), pp.1162-1171.

Cournia, Z., Allen, T., Andricioaei, I., Antonny, B., Baum, D., Brannigan, G., Buchete, N., Deckman, J., Delemotte, L., del Val, C., Friedman, R., Gkeka, P., Hege, H., Hénin, J., Kasimova, M., Kolocouris, A., Klein, M., Khalid, S., Lemieux, M., Lindow, N., Roy, M., Selent, J., Tarek, M., Tofoleanu, F., Vanni, S., Urban, S., Wales, D., Smith, J. and Bondar, A., 2015. Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. The Journal of Membrane Biology, 248(4), pp.611-640.

Datsenko, K. and Wanner, B., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences, 97(12), pp.6640-6645.

Davies S., Farrar J., Rex J., White L. J., Murry R., O'Neill J., in: The Review on Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. Ed.: Government H., 2014, pp. 1–14

Doerrler, W., Sikdar, R., Kumar, S. and Boughner, L., 2013. New Functions for the Ancient DedA Membrane Protein Family. Journal of Bacteriology, 195(1), pp.3-11.

Enright, M., Robinson, D., Randle, G., Feil, E., Grundmann, H. and Spratt, B., 2002. The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). Proceedings of the National Academy of Sciences, 99(11), pp.7687-7692.

Fortuin, S., Nel, A., Blackburn, J. and Soares, N., 2020. Comparison between the proteome of Escherichia coli single colony and during liquid culture. Journal of Proteomics, 228, p.103929.

Hopf, T., Colwell, L., Sheridan, R., Rost, B., Sander, C. and Marks, D., 2012. Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. Cell, 149(7), pp.1607-1621.

Imai, K. and Mitaku, S., 2005. Mechanisms of secondary structure breakers in soluble proteins. Biophysics, 1, pp.55-65.

Jorgensen.biology.utah.edu. (2019). ApE- A plasmid Editor. [online] Available at: http://jorgensen.biology.utah.edu/wayned/ape/

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

Keller, R., Schleppi, N., Weikum, J. and Schneider, D., 2015. Mutational analyses of YqjA, a Tvp38/DedA protein of *E. coli*. FEBS Letters, 589(7), pp.842-848.

Keller, R. and Schneider, D., 2013. Homologs of the yeast Tvp38 vesicle-associated protein are conserved in chloroplasts and cyanobacteria. Frontiers in Plant Science, 4, pp.467.

Khafizov, K., Staritzbichler, R., Stamm, M. and Forrest, L., 2010. A Study of the Evolution of Inverted-Topology Repeats from LeuT-Fold Transporters Using AlignMe. Biochemistry, 49(50), pp.10702-10713.

Kumar, S. and Doerrler, W., 2014. Members of the Conserved DedA Family Are Likely Membrane Transporters and Are Required for Drug Resistance in Escherichia coli. Antimicrobial Agents and Chemotherapy, 58(2), pp.923-930.

Kumar, S. and Doerrler, W., 2015. *Escherichia coli* YqjA, a Member of the Conserved DedA/Tvp38 Membrane Protein Family, Is a Putative Osmosensing Transporter Required for Growth at Alkaline pH. Journal of Bacteriology, 197(14), pp.2292-2300.

Kumar, S., Bradley, C., Mukashyaka, P. and Doerrler, W., 2016. Identification of essential arginine residues of *Escherichia coli* DedA/Tvp38 family membrane proteins YqjA and YghB. FEMS Microbiology Letters, 363(13), p.fnw133.

Kumar, S., Tiwari, V. and Doerrler, W., 2017. Cpx-dependent expression of YqjA requires cations at elevated pH. FEMS Microbiology Letters, 364(12).

Lee, A., 2004. How lipids affect the activities of integral membrane proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes, 1666(1-2), pp.62-87.

Liang, F., Xu, Q., Sikdar, R., Xiao, Y., Cox, J. and Doerrler, W., 2010. BB0250 of Borrelia burgdorferi Is a Conserved and Essential Inner Membrane Protein Required for Cell Division. Journal of Bacteriology, 192(23), pp.6105-6115.

Mahon, M., 2011. pHluorin2: An enhanced, ratiometric, pH-sensitive green florescent protein. Advances in Bioscience and Biotechnology, 02(03), pp.132-137.

Marks, D., Colwell, L., Sheridan, R., Hopf, T., Pagnani, A., Zecchina, R. and Sander, C., 2011. Protein 3D Structure Computed from Evolutionary Sequence Variation. PLoS ONE, 6(12), p.e28766. Available at: https://v2.evcouplings.org/

Mesdaghi, S., Murphy, D., Sánchez Rodríguez, F., Burgos-Mármol, J. and Rigden, D., 2021. *In silico* prediction of structure and function for a large family of transmembrane proteins that includes human Tmem41b. F1000Research, 9, p.1395.

Morita, K., Hama, Y. and Mizushima, N., 2019. TMEM41B functions with VMP1 in autophagosome formation. Autophagy, 15(5), pp.922-923.

Okawa, F., Hama, Y., Zhang, S., Morishita, H., Yamamoto, H., Levine, T. and Mizushima, N., 2021. Evolution and insights into the structure and function of the DedA superfamily containing TMEM41B and VMP1. Journal of Cell Science,.

Panta, P., Kumar, S., Stafford, C., Billiot, C., Douglass, M., Herrera, C., Trent, M. and Doerrler, W., 2019. A DedA Family Membrane Protein Is Required for *Burkholderia thailandensis* Colistin Resistance. Frontiers in Microbiology, 10.

Price, N. and Raivio, T., 2009. Characterization of the Cpx Regulon in Escherichia coli Strain MC4100. Journal of Bacteriology, 191(6), pp.1798-1815.

PyMol. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Raivio, T., Leblanc, S. and Price, N., 2013. The Escherichia coli Cpx Envelope Stress Response Regulates Genes of Diverse Function That Impact Antibiotic Resistance and Membrane Integrity. Journal of Bacteriology, 195(12), pp.2755-2767.

Rawson, S., Davies, S., Lippiat, J. and Muench, S., 2016. The changing landscape of membrane protein structural biology through developments in electron microscopy. Molecular Membrane Biology, 33(1-2), pp.12-22.

Sabnis, A., Hagart, K., Klöckner, A., Becce, M., Evans, L., Furniss, R., Mavridou, D., Murphy, R., Stevens, M., Davies, J., Larrouy-Maumus, G., Clarke, T. and Edwards, A., 2021. Colistin kills bacteria by targeting lipopolysaccharide in the cytoplasmic membrane. eLife, 10.

Scarsbrook, H., Urban, R., Streather, B., Moores, A. and Mulligan, C., 2021. Topological analysis of a bacterial DedA protein associated with alkaline tolerance and antimicrobial resistance. Microbiology, 167(12).

Sigal, N., Vardy, E., Molshanski-Mor, S., Eitan, A., Pilpel, Y., Schuldiner, S. and Bibi, E., 2005. 3D Model of the Escherichia coli Multidrug Transporter MdfA Reveals an Essential Membrane-Embedded Positive Charge. Biochemistry, 44(45), pp.14870-14880.

Sikdar, R. and Doerrler, W., 2010. Inefficient Tat-Dependent Export of Periplasmic Amidases in an Escherichia coli Strain with Mutations in Two DedA Family Genes. Journal of Bacteriology, 192(3), pp.807-818.

Sikdar, R., Simmons, A. and Doerrler, W., 2013. Multiple Envelope Stress Response Pathways Are Activated in an Escherichia coli Strain with Mutations in Two Members of the DedA Membrane Protein Family. Journal of Bacteriology, 195(1), pp.12-24.

Thompkins, K., Chattopadhyay, B., Xiao, Y., Henk, M. and Doerrler, W., 2008. Temperature Sensitivity and Cell Division Defects in an *Escherichia coli* Strain with Mutations in *yghB* and *yqjA*, Encoding Related and Conserved Inner Membrane Proteins. Journal of Bacteriology, 190(13), pp.4489-4500.

von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. The EMBO Journal, 5(11), pp.3021-3027.

WHO., 2020. Antimicrobial Resistance. [online] Available at: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

WHO., 2021. Global Antimicrobial Resistance and Use Surveillance System (GLASS) report., ISBN: 978-92-4-002733-6., Licence: CC BY-NC-SA 3.0 IGO.

Workman, S., Worrall, L. and Strynadka, N., 2018. Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. Nature Communications, 9(1).

Yamamoto, K. and Ishihama, A., 2006. Characterization of Copper-Inducible Promoters Regulated by CpxA/CpxR in Escherichia coli. Bioscience, Biotechnology, and Biochemistry, 70(7), pp.1688-1695.

Yamashita, A., Singh, S., Kawate, T., Jin, Y. and Gouaux, E., 2005. Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. Nature, 437(7056), pp.215-223.

Yan, C. and Luo, J., 2010. An Analysis of Re-entrant Loops. The Protein Journal, 29(5), pp.350-354.

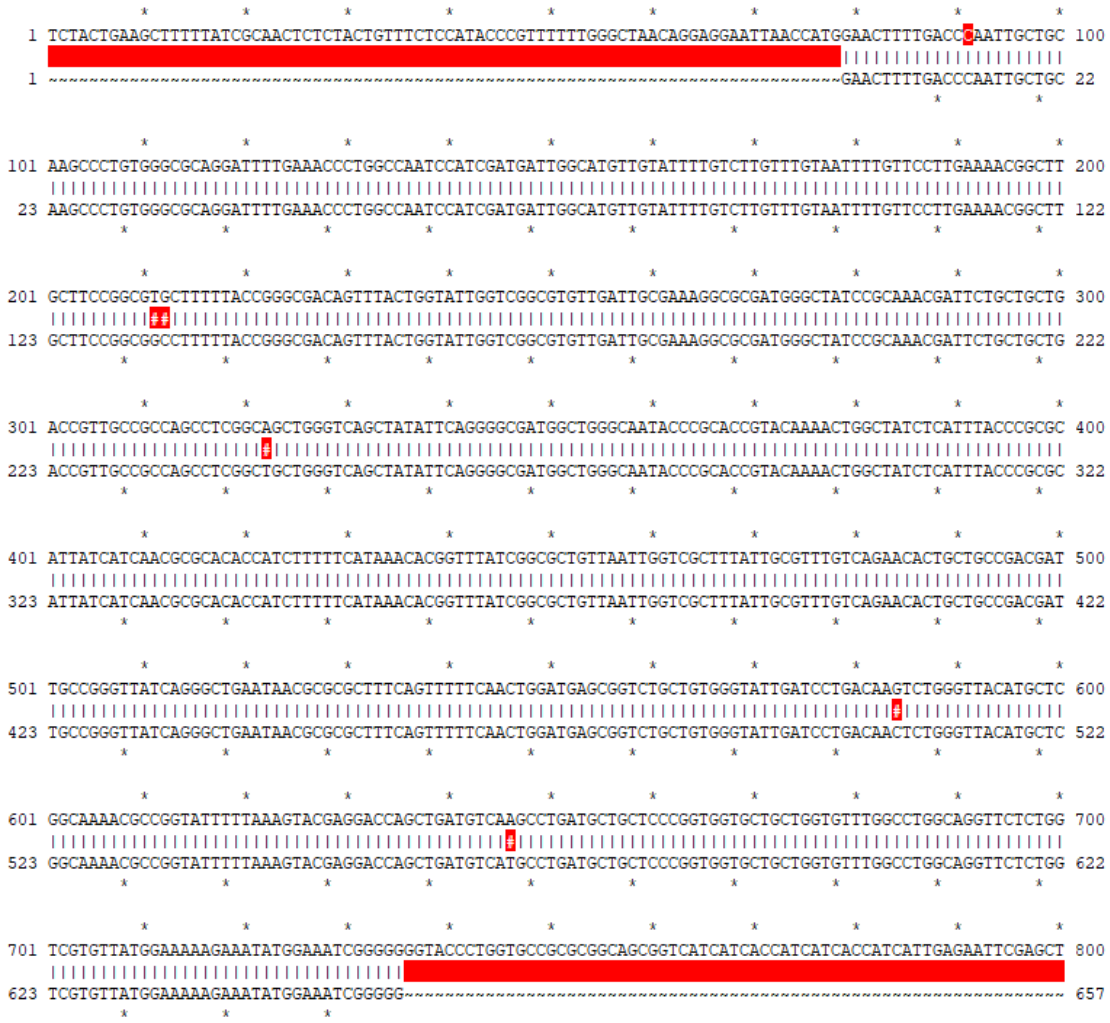
Yin, H. and Flynn, A., 2016. Drugging Membrane Protein Interactions. Annual Review of Biomedical Engineering, 18(1), pp.51-76.

7. Appendix

Cysteine Mutant	Forward Primer	Reverse Primer
YqjAA46C	GCTTGCTTCCGGCG**TGC**TTTTTACCGGGCGAC	GTCGCCCGGTAAAAA**GCA**CGCCGGAAGCAAGC
YqjAY68C	GAAAGGCGCGATGGGC**TGT**CCGCAAACGATTCTG	CAGAATCGTTTGCGG**ACA**GCCCATCGCGCCTTTC
YqjAL75C	CAAACGATTCTGCTG**TGT**ACCGTTGCCGCCAGCC	GGCTGGCGGCAACGGT**ACA**CAGCAGAATCGTTTG
YqjAA143C	CTGCTGCCGACGATT**TGC**GGGTTATCAGGGCTG	CAGCCCTGATAACCC**GCA**AATCGTCGGCAGCAG
YqjAL148C	GCCGGGTTATCAGGG**TGC**AATAACGCGCGCTTTC	GAAAGCGCGCGTTATT**GCA**CCCTGATAACCCGGC
YqjAR152C	CAGGGCTGAATAACGCG**TGC**TTTCAGTTTTTCAACTGG	CCAGTTGAAAAACTGAAA**GCA**CGCGTTATTCAGCCCTG
YqjAL175C	CAACTCTGGGTTACATG**TGC**GGCAAAACGCCGG	CCGGCGTTTTGCC**GCA**CATGTAACCCAGAGTTG
YqjAT178C	GGGTTACATGCTCGGCAAA**TGC**CCGGTATTTTTTAAAGTACGAG	CTCGTACTTTAAAAATACCGG**GCA**TTTGCCGAGCATGTAACCC
YqjAY184C	CAAAACGCCGGTATTTTTAAAG**TGC**GAGGACCAGCTGATGTC	GACATCAGCTGGTCCTC**GCA**CTTTAAAAATACCGGCGTTTTG
YqjAL188C	GTACGAGGACCAG**TGC**ATGTCAAGCCTGATGCTGC	GCAGCATCAGGCTTGACAT**GCA**CTGGTCCTCGTAC
YqjAL194C	GTCAAGCCTGATG**TGC**CTCCCGGTGGTGCTG	CAGCACCACCGGGAG**GCA**CATCAGGCTTGAC
YqjAL195C	CAAGCCTGATGCTG**TGC**CCGGTGGTGCTGCTG	CAGCAGCACCACCGG**GCA**CAGCATCAGGCTTG

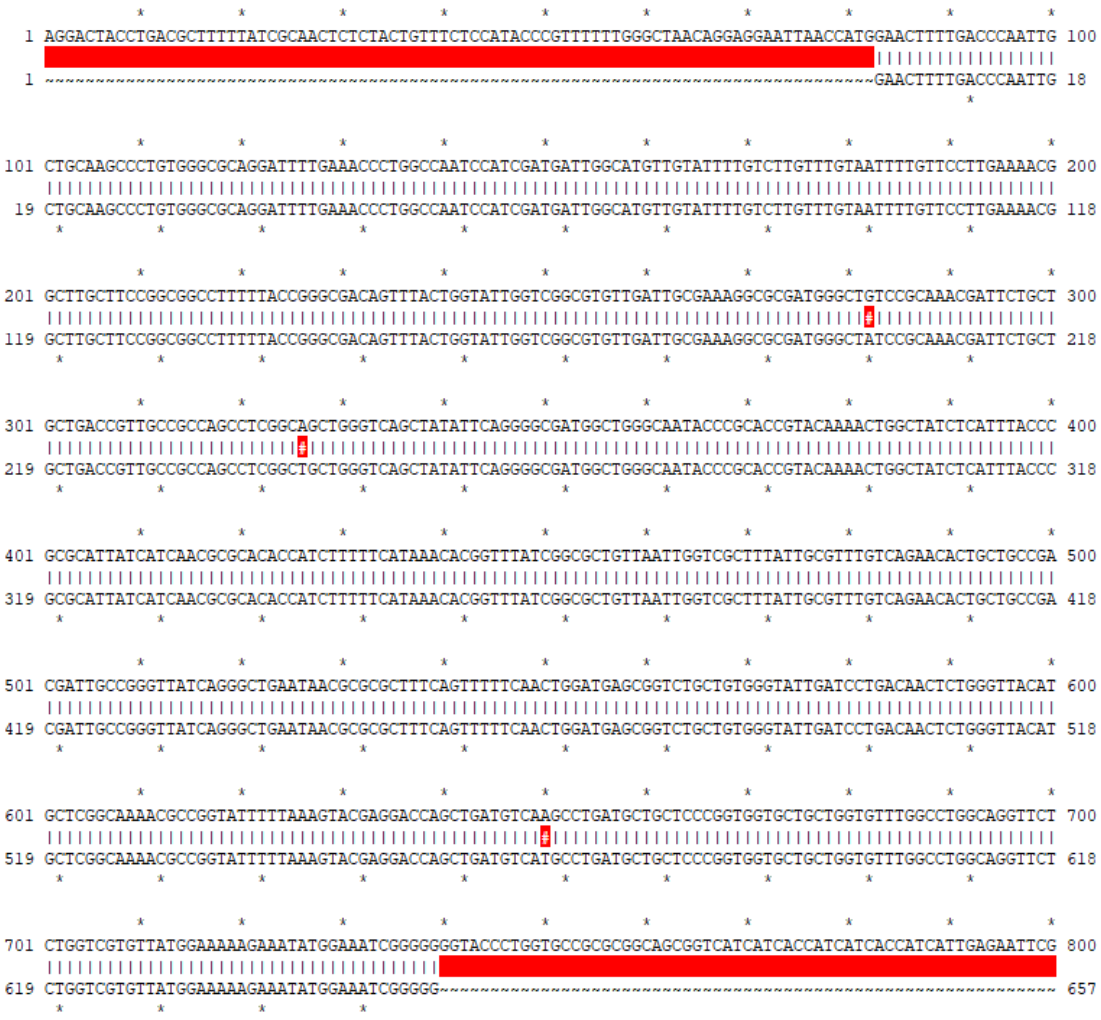
Appendix 1 – Forward and reverse primers used for site-directed mutagenesis to produce single cysteine mutants of YqjA

Nucleotides that correspond to the change in the DNA sequence required for the mutagenesis are in bold.



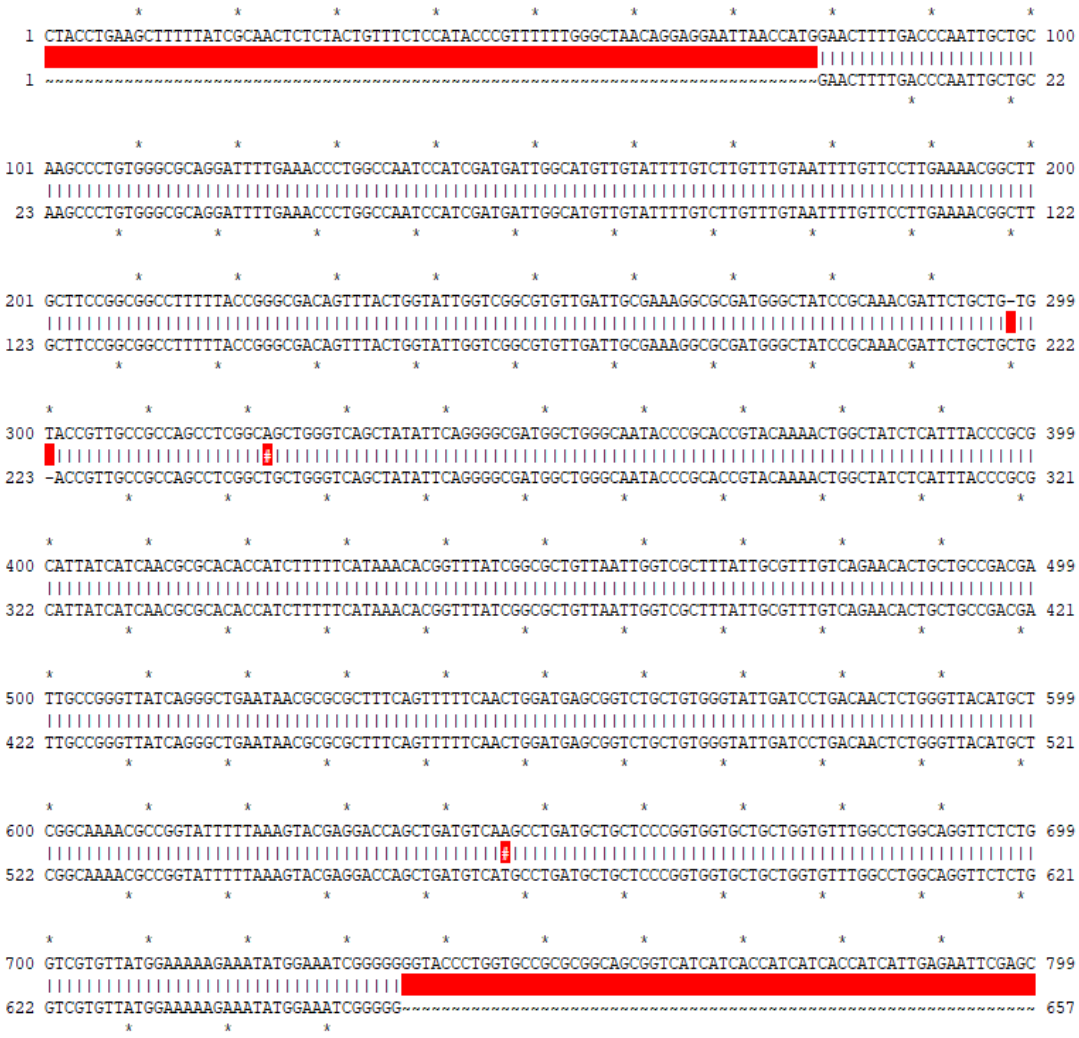
Appendix 2 – Sequencing analysis of YqjAA46C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAA46C result having been produced in a Cys-less background. GCC to TGC at nucleotide 133 for A46C mutation.



Appendix 3 – Sequencing analysis of YqjAY68C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAY68C result having been produced in a Cys-less background. TAT to TGT at nucleotide 199 for Y68C mutation.

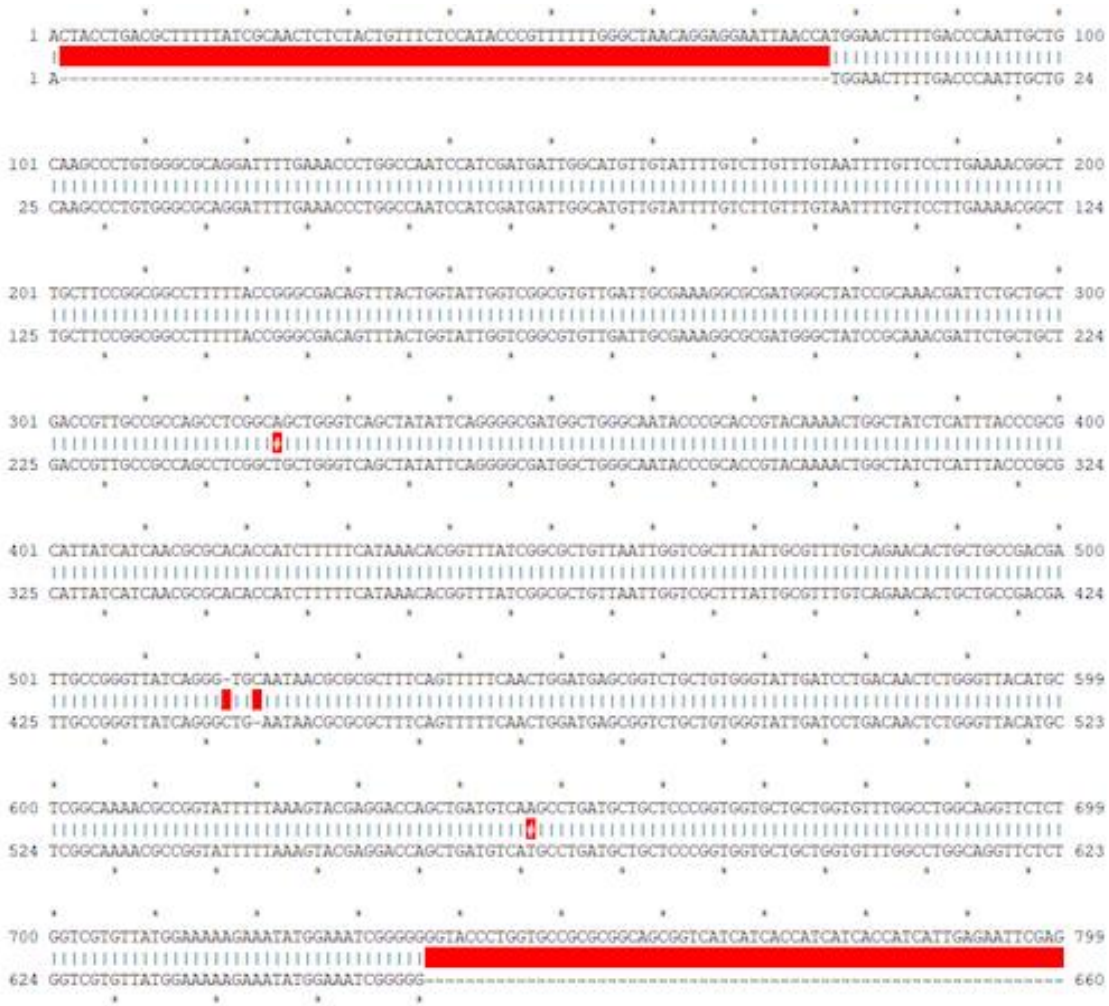


Appendix 4 – Sequencing analysis of YqjAL75C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL75C result having been produced in a Cys-less background. CTG to TGT at nucleotide 220 for L75C mutation.

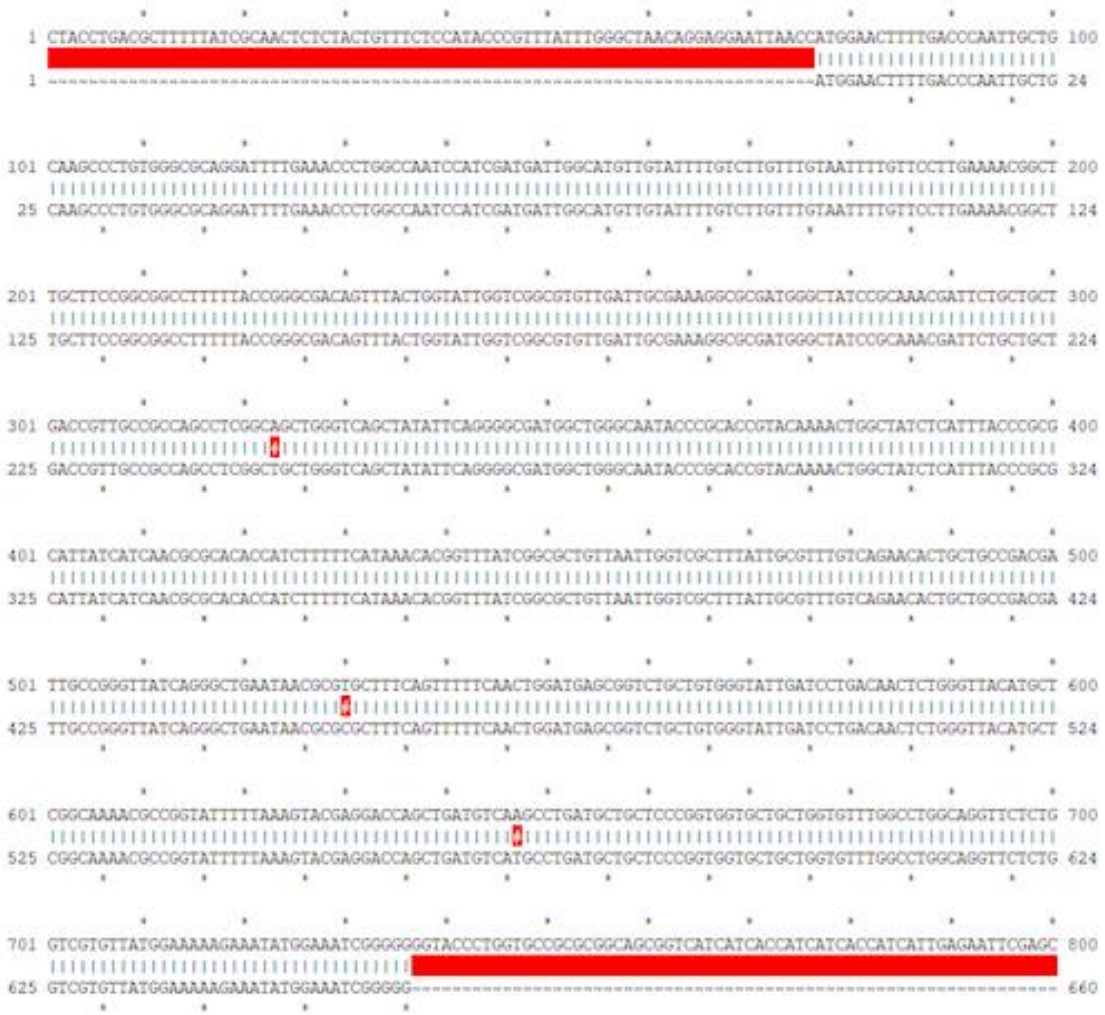
Appendix 5 – Sequencing analysis of YqjAA143C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAA143C result having been produced in a Cys-less background. GCC to TGC at nucleotide 427 for A143C mutation.



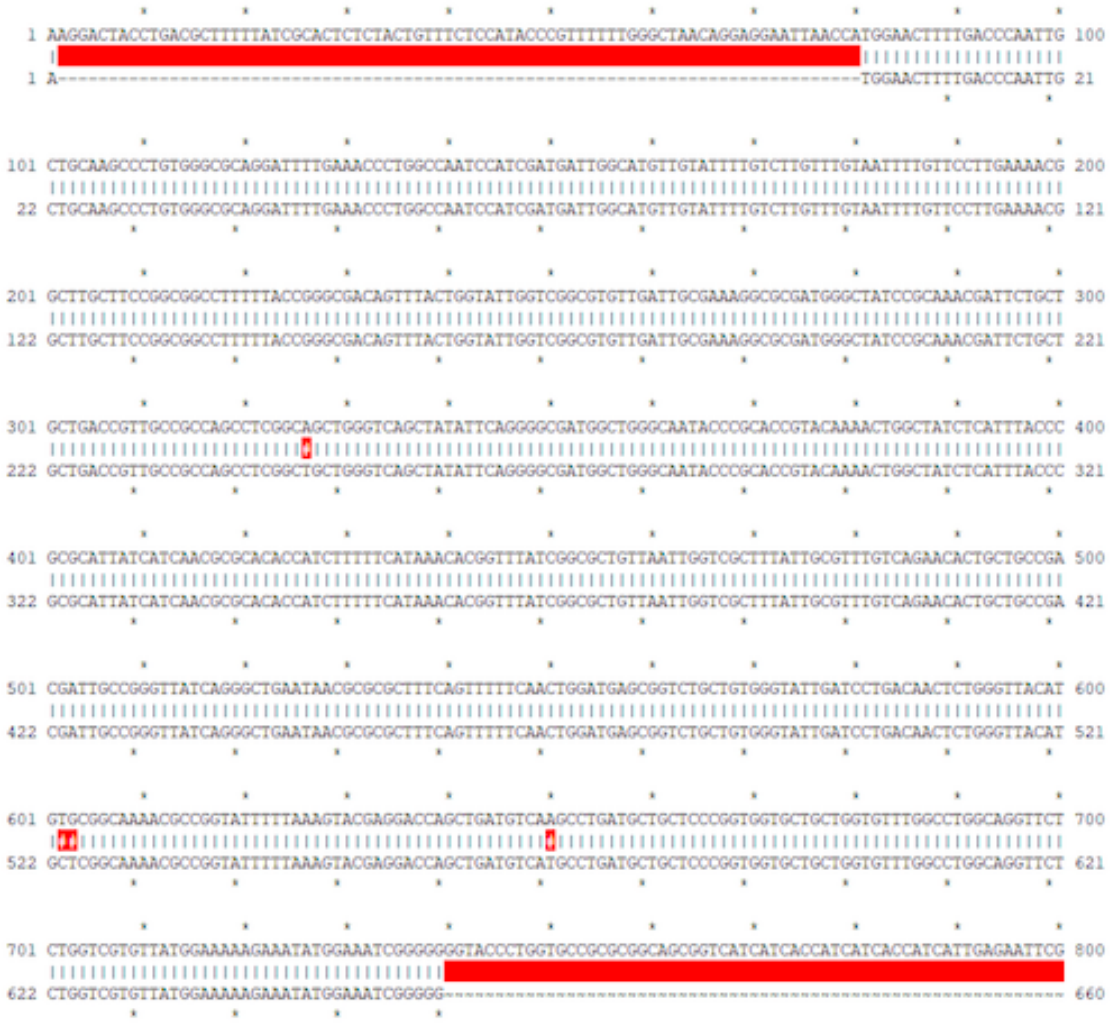
Appendix 6 – Sequencing analysis of YqjAL148C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL148C result having been produced in a Cys-less background. CTG to TGC at nucleotide 442 for L148C mutation.



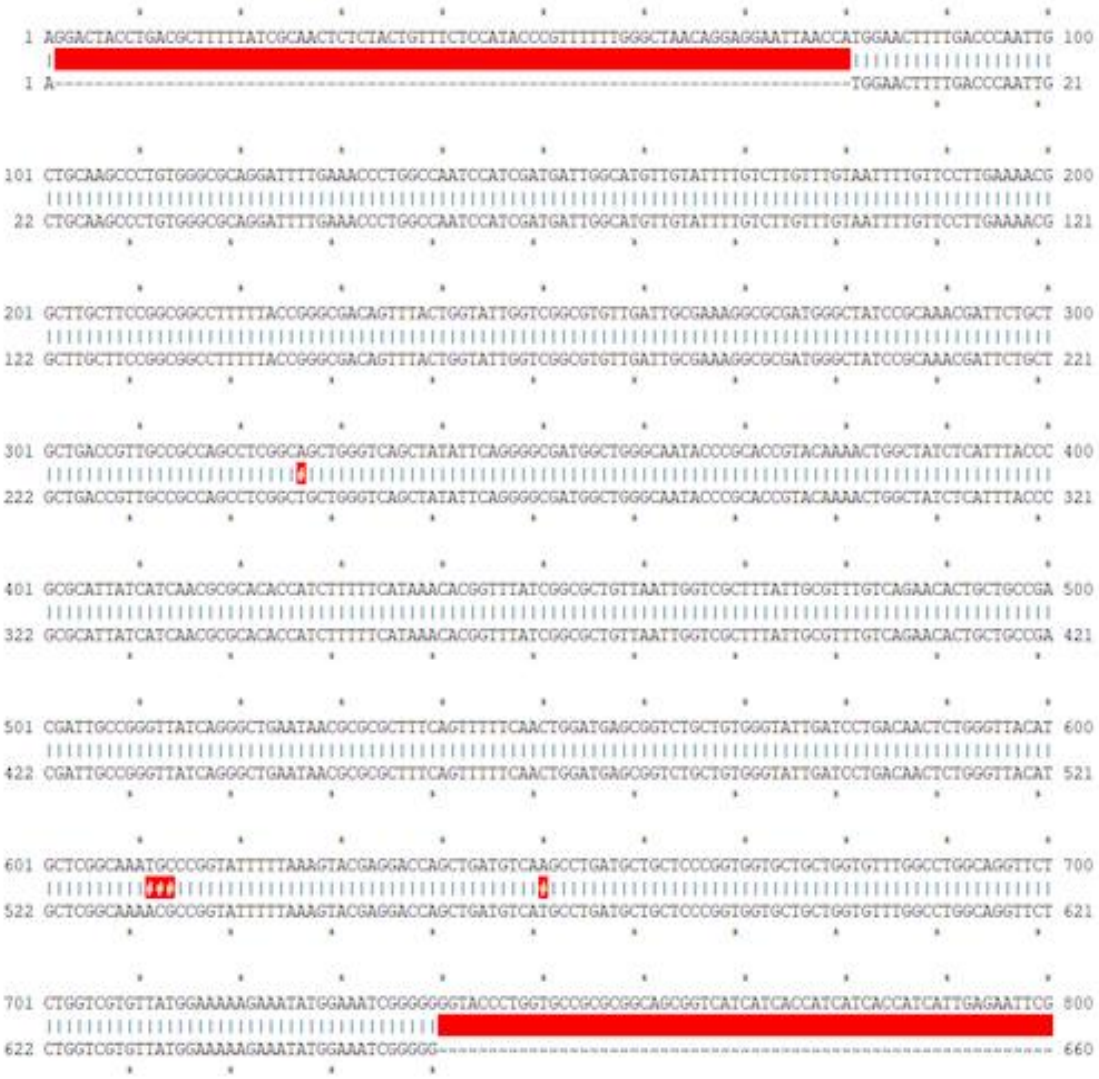
Appendix 7 – Sequencing analysis of YqjAR152C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAR152C result having been produced in a Cys-less background. CGC to TGC at nucleotide 454 for R152C mutation.



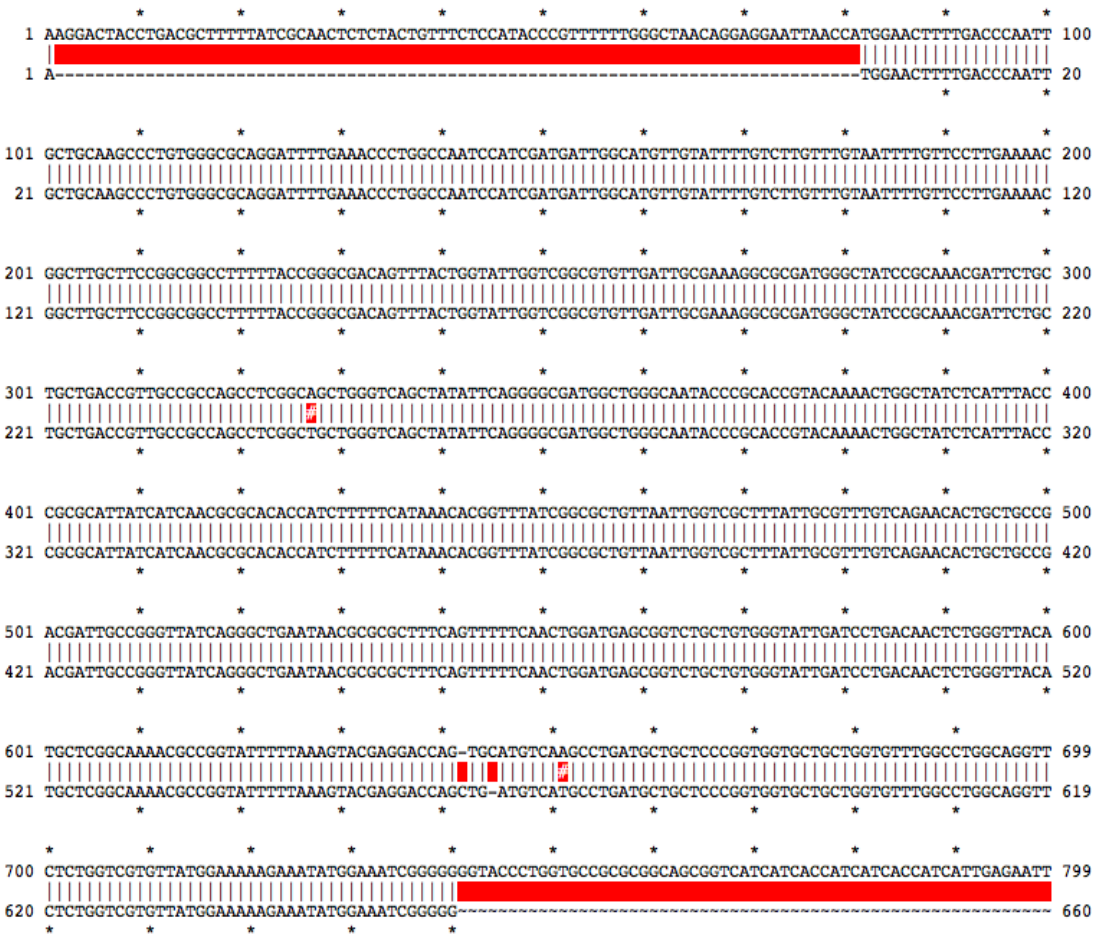
Appendix 8 – Sequencing analysis of YqjAL175C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL175C result having been produced in a Cys-less background. CTC to TGC at nucleotide 523 for L175C mutation.



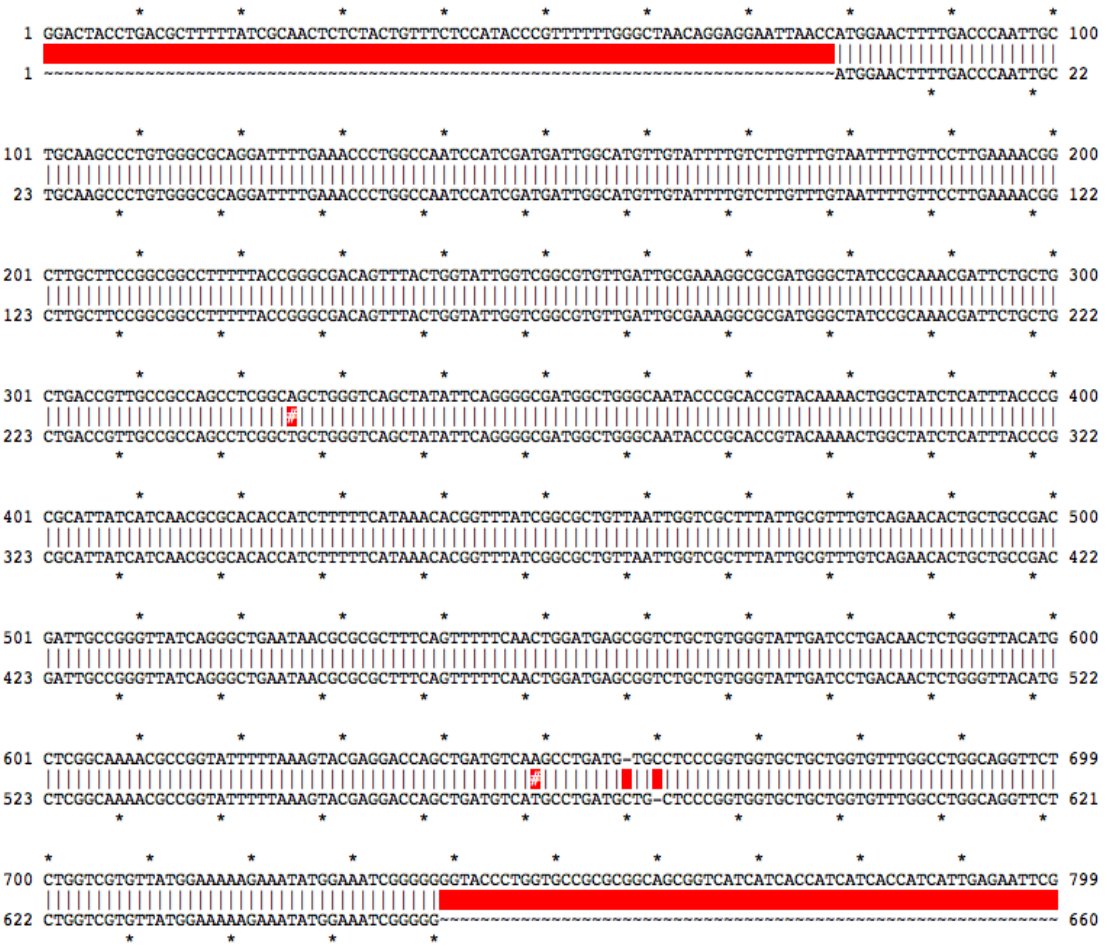
Appendix 9 – Sequencing analysis of YqjAT178C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAT178C result having been produced in a Cys-less background. ACG to TGC at nucleotide 532 for T178C mutation.



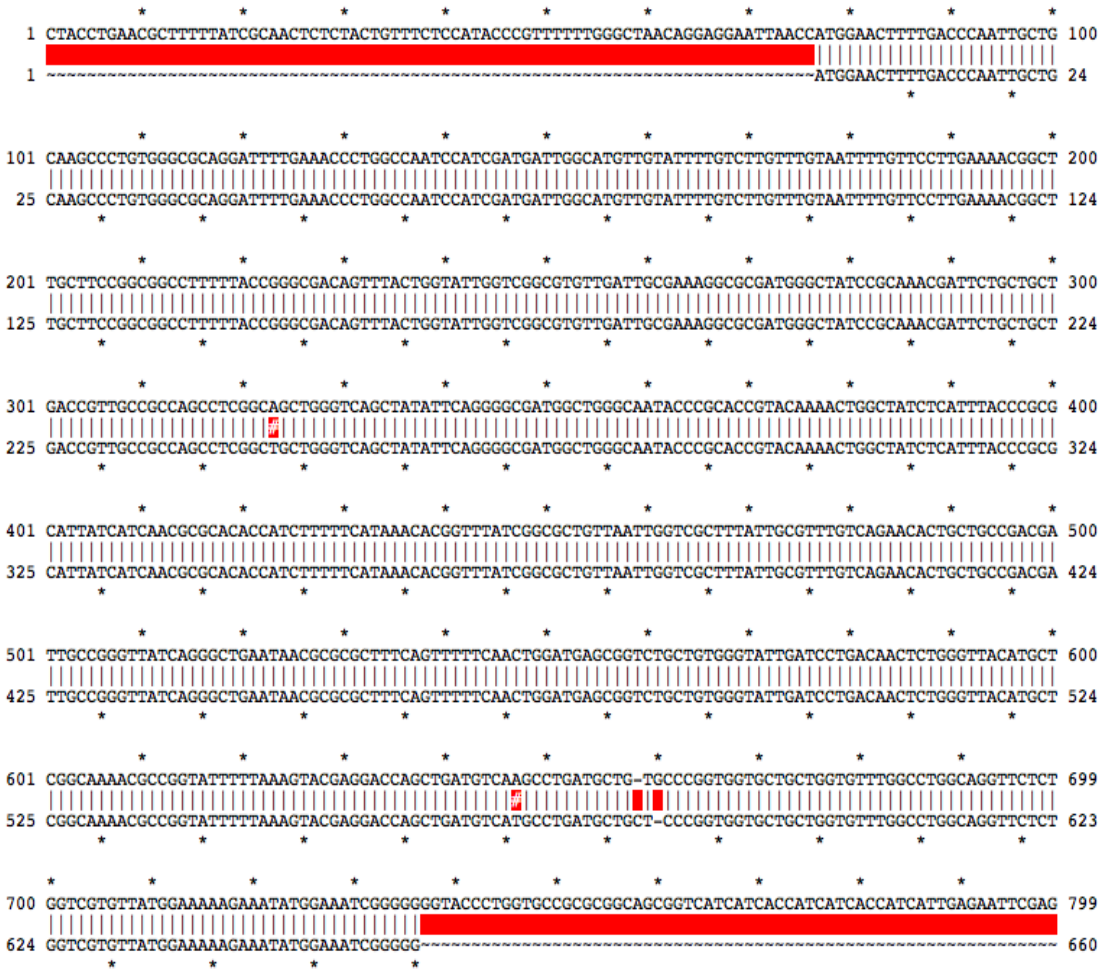
Appendix 10 – Sequencing analysis of YqjAL188C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL188C result having been produced in a Cys-less background. CTG to TGC at nucleotide 562 for L188C mutation.



Appendix 11 – Sequencing analysis of YqjAL194C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL194C result having been produced in a Cys-less background. CTG to TGC at nucleotide 580 for L194C mutation.



Appendix 12 – Sequencing analysis of YqjAL195C

Wild-type *yqjA* (bottom row) aligned with the sequencing result confirming that the native cysteines (C83 and C191) are serine residues in YqjAL195C result having been produced in a Cys-less background. CTC to TGC at nucleotide 583 for L195C mutation.