Examination of a putative Cu^+ importer in *Pseudomonas aeruginosa*

A Major Qualifying Project

Submitted to the Faculty of

Worcester Polytechnic Institute

in partial fulfillment of the requirements for the

Degree in Bachelor of Science

in

Biochemistry

Bioinformatics and Computational Biology

By

Virginia Núñez Mir

04/22/2018

Professor José M. Argüello

Professor Scarlet Shell

WPI Project advisors

Abstract

The opportunistic bacteria *Pseudomonas aeruginosa* is among the greatest causes of infection in hospitals. Due to its inherent resistance against many antibiotics, it is important to find new and different ways to attack the bacteria. One promising method is to target the bacteria's copper homeostasis network, which is used to carefully balance their needs for copper with the toxicity of the metal. However, the relationships between the proteins in this network are still not well understood. The focus of this project is to gain understanding on the putative copper importer PA5030 and its relationship to other proteins in the copper homeostasis network. First, recombination techniques were employed to successfully develop a double knockout mutant of the PA5030 gene and a different copper importer, PA1848 (ccoA). Second, in-silico exploration of the structure of the protein, accompanied by the development of a homolog filtering script, revealed two potential copper binding sites, M157-M173-H355 and C65-M185-M262, and 925 homologs that suggest that PA5030 is a potential copper/arabinose antiporter.

Acknowledgements

I would like to thank Professor José Argüello for guiding me through this Project and my undergraduate career, as well as for giving me the opportunity to work at his lab and grow as a professional. I also take this opportunity to thank Professor Scarlet Shell and Professor Liz Ryder for their invaluable advice these four years. Finally, special thanks to all members of the Argüello lab for making my working experience truly rewarding, to Drs. Lorena Novoa-Aponte and David Ramirez for all their useful tips, and to Dr. Julia Quintana for helping me develop into the scientist I am today.

Table of Contents

Abstract 2
Acknowledgements 3
Introduction 5
Methods 9
Results 15
Discussion 25
References 28
Supplementary Material 36

Table of Figures

Figure 1: DNA constructs 9
Table 1: Antibiotic Concentrations 11
Figure 2: Plasmid constructs 12
Figure 3: Mutant confirmation 15
Figure 4: Antibiotic sensitivity 16
Figure 5: Building plasmid construct 17
Figure 6: Screening 18
Figure 7: Final mutant confirmation 18
Table 2: BLAST results 20
Table 3: Conserved residues 21
Figure 8: Phylogenetic analysis 23
Figure 9: Copper binding sites 24
Supplementary Table 1: Primers 36
Supplementary Table 2: PCR Programs 36
Supplementary Figure 1: Identity Alignment 37
Supplementary Figure 2: C/M/H Alignment 38
Supplementary Figure 3: PA5030 models 39

Introduction

Pseudomonas aeruginosa is a Gram-negative opportunistic pathogen. It can grow in a variety of environments, such as soil, water, soap, and a myriad of hosts ranging from plants to humans (14). Its inherent resistance against many antibiotics, its ability to grow on surfaces in the form of complex communities called biofilms, which have been associated with increased resistance to antibiotics and difficulty to eradicate persistent infections, and its great ability to adapt has helped it become a nearly ubiquitous bacteria (31,40). *P. aeruginosa* is one of the major causes of infection in hospitals, encompassing more than 10% of all nosocomial infections, and in cystic fibrosis patients, where they can form biofilms in the lungs (21,47). Therefore it is important to find new ways to target the bacteria during infections. Furthermore, its status as a model organism adds an extra benefit to studying this bacteria, as the information gained can be extrapolated to understand other gram-negative bacteria, which in total account for more than 50% of all nosocomial infections (7,47).

Copper is a transition metal that bacteria need for a variety of processes such as catalyzing electron transfer reactions (33). However, in high concentrations, copper becomes toxic to the cell by increasing oxidative-stress through disruption of iron-sulfur clusters (33). Therefore, bacteria must carefully balance the amount of copper in the cell to ensure they can satisfy their needs without being killed by the metal. To reach that balance, bacteria employ a complex network of proteins that import and export copper in and out of the cell, as well as proteins the chaperone the metal from one location to another. By disrupting this network, we can destroy

the bacteria's ability to survive, or provide a method of killing bacteria who are resistant to many current antibiotics. Knowledge of the workings of copper homeostasis is also important in understanding human diseases such as Wilson's disease, which is caused by build-up of copper in the liver (1).

Many proteins important to copper homeostasis have been described. Movement of Cu^+ from one protein to another in the cytoplasm and periplasm is facilitated by chaperones, such as CopZ1, CopZ2, and CusF. These have been shown to bind Cu^+ with high affinity, eliminating the presence of free Cu^+ in the cell (32,34). The balance between these proteins is mediated by several regulators, such as CopR and CueR (34,42). CopA1 and CopA2 are two P-type ATPases responsible for efflux of Cu^+ from the cytoplasm into the periplasm in *P. aeruginosa* (11). They work in tandem with proteins involved in export of Cu^+ from the periplasm into the extracellular environment, such as PcoB (23,34). In opposition to the efflux channels, several Cu^+ importers have been identified. OprC is an outer-membrane protein that moves Cu^+ from the extracellular surroundings into the periplasm (46). It has been suggested that OprC works in conjunction with PA5030, a protein suspected of importing Cu^+ from the periplasm into the cytoplasm (34). Another membrane protein that has been attributed the function of importing Cu^+ from the periplasm into the cytoplasm is CcoA (9). However, the relationships between the proteins in this network, for example the relationship between the putative importer PA5030 and *P. aeruginosa* CcoA (PA1848), are still not well understood.

Expression of PA5030 is highly down-regulated during Cu^+ stress events (34). Furthermore, mutant strains of *P. aeruginosa* with a knockout of the PA5030 gene have an increase of ability to survive Cu^+ stress (34). These two factors, coupled with the abundance of methionines and histidines in its sequence, suggest that PA5030 might be a copper importer (34).

CcoA has been shown to be necessary for Cu^+ uptake in *Rhodobacter capsulatus* (9). The protein contains two conserved Cu^+ binding domains following a MXXXM and HXXXM motif (20). The protein has also been shown to be important for a cbb3-type cytochrome c oxidase, an enzyme involved in electron transfer, activity, potentially serving as a provider of Cu^+ for heme copper formation (9).

Both proteins have been predicted to have 10-12 transmembrane domains and belong to the Major Facilitator Superfamily (MFS). Alignment of the two proteins (39) reveals that they have 15% identity and 38% similarity at 85% coverage. Interestingly, the same experiments that saw repression of PA5030 under Cu^+ stress, saw no change in expression of *P. aeruginosa* CcoA (34). However, the relationship between PA5030 and CcoA has yet to be tested.

The focus of this project was to gain understanding of the PA5030 protein and its relationship to other proteins in the copper homeostasis network by accomplishing two different goals corresponding to a biochemical and a bioinformatic approach. One way to explore the interactions between PA5030 and other proteins in the copper homeostasis network, is by comparing mutant strains where PA5030 and another protein is knocked-out. However, despite *P. aeruginosa*'s multiple readily available mutants, and the existence of a comprehensive mutant

library (15,19), of double knockout mutant strains are not available. Therefore, the first goal was to create a double mutant strain by knocking out two genes encoding putative copper importers, PA5030 and CcoA, using recombination techniques (16). This mutant will then be useful for future experiments that wish to explore the relationship between these two proteins.

The second focus of the project was on the structure of the protein, utilizing a bioinformatic approaches. We aimed to identify potential copper-binding sites in PA5030 by finding homologs of the PA5030 protein in other bacteria species and developing a model of the protein. These homologs helped determine conserved residues in the protein, particularly methionines, cysteines, and histidines since these are potential copper-binding residues. A python script was also written to aid with the process of finding homologs. The model structures using 3WDO as a template then furthered this analysis by allowing visualization of the spatial proximity of the identified residues.

Methods

Extraction of genomic DNA. Cells from an overnight culture of *Pseudomonas aeruginosa* PAO1 were collected by centrifugation and washed with Sarkosyl 0.1% in 10 mM Tris, 1 mM EDTA pH 8.0 (TE buffer). The pellet was re-suspended in TE buffer and treated with Proteinase K, RNase, and 5% SDS. Following incubation at 37 °C for 1 h, cold 5M NaCl aqueous solution was added, and incubated on ice for 30 min. The supernatant was combined with an equal volume of phenol/chloroform/isoamyl alcohol (25:25:1). The nucleic acids were precipitated from the aqueous phase by adding 2.5 volumes of 100% ethanol and incubating at -20 °C for 1 h. The pellet was re-suspended in 100 µL of TE buffer once all the ethanol had evaporated.

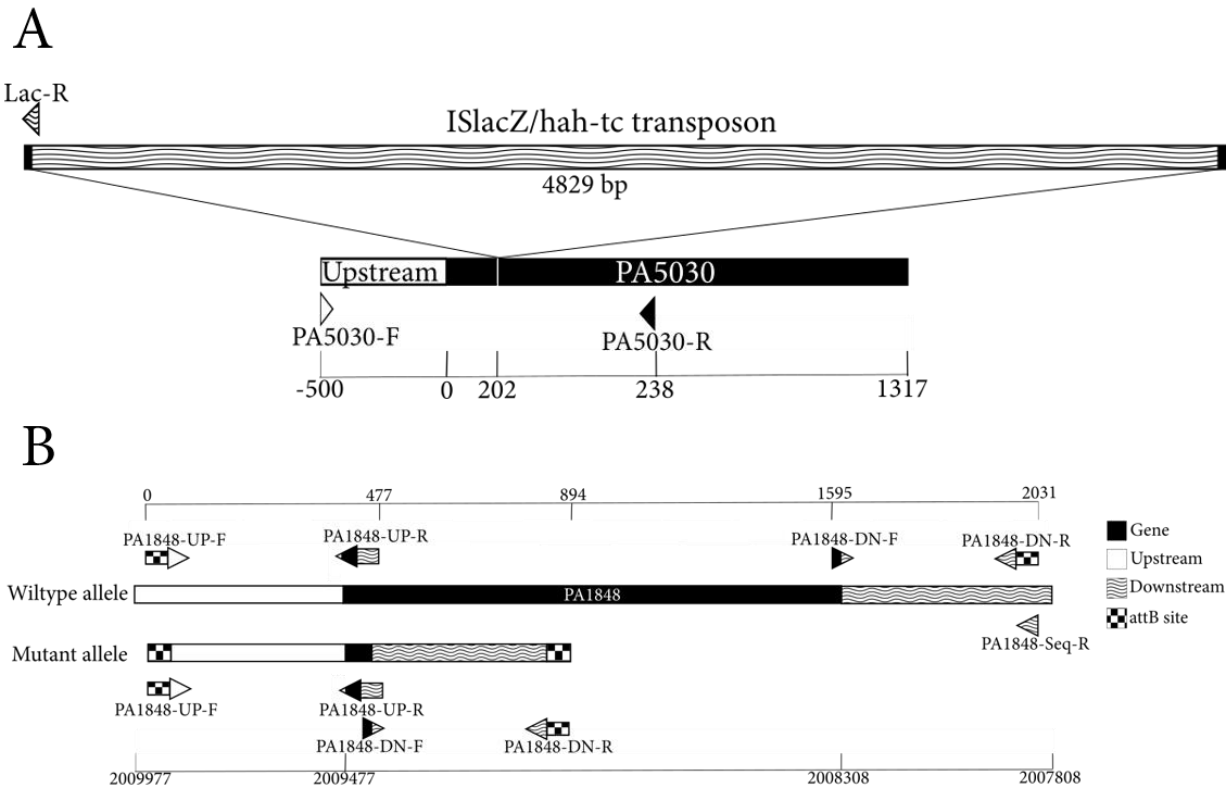


Figure 1: DNA constructs. A) Location of LacZ transposon insertion in the ΔPA5030 strain. B) Construction of mutant deletion allele. The numbers at the top represent length, and the numbers at the bottom represent location in genome. The primers used for manipulation of the gene are represented as triangles.

Confirmation of ΔPA5030 mutation. The ΔPA5030 strain was obtained from the University of Washington mutant library (15,19). Presence of the transposon insertion in the mutant was confirmed by polymerase chain reaction (PCR) using a 5' primer annealing before the location of the insertion, and a 3' primer sitting within the transposon (Fig. 1A). PCR using primers that amplify the gene of interest confirmed that the full gene can be amplified in PAO1 but not in either of the mutant strains. Following PCR, all reactions were analyzed by agarose gel electrophoresis.

PA5030/PA1848 mutations. Mutation of PA1848 in the ΔPA5030 background was done following Hmelo et al. (2015). Two sets of primers were designed to amplify the upstream and downstream regions of the target gene with the addition of attB sites necessary for Gateway cloning. Splicing by overlap extension (SOE) PCR was used to construct an allele consisting on the upstream and downstream region and only 5% of the target gene (Fig. 1B). A Gateway cloning recombination reaction was used to introduce the deletion allele into the pDONRpEX18-Gm plasmid. Successful recombinants were screened in DH5-α competent cells. Biparental conjugation was performed with S17-1 colonies carrying the plasmid with the deletion allele and the mutant ΔPA5030 using puddle mating. Merodiploids were screened using citrate as a nutritional counter-selection and the desired mutants were isolated from counter selection no-salt LB + 15 % sucrose (NSLB) plates. Final confirmation of the double mutants was performed by screening colonies from NSLB plates with three separate PCR reactions: one to confirm that the original mutation was still present, a second reaction confirmed the presence of the mutant allele,

and a final reaction was run to confirm that the mutant allele was introduced in the correct location. For this reaction, the 5' primer used to amplify the downstream area of the gene was used with a 3' primer located further downstream than the deletion allele (Fig. 1B). Following PCR, all reactions were analyzed by agarose gel electrophoresis. Confirmation of normal behavior in the mutants was performed by plating on LB plates with no antibiotic, LB + Gm50 plates, and Pseudomonas isolation agar plates.

Determining antibiotic sensitivity of the ΔPA5030 strain. A vector designed for Gateway cloning was selected to confer resistance to an antibiotic to which the ΔPA5030 strain were sensitive. Antibiotic sensitivity was determined by growing each mutant strain in LB media with increasing concentrations of each antibiotic being tested: ampicillin, gentamycin, and chloramphenicol (Table 1). Tubes with no antibiotic, inoculated with each of the strains, were used as controls. Growth was observed after incubating overnight at 37 °C. The vector used for cloning, pDONRpEX18-Gm, was obtained from Dr. Joe J Harrison (University of Calgary) (16).

Antibiotic	Concentrations Tested (μg/ mL)				
ampicillin	10	50	100	150	200
gentamycin	5	8	10	30	50
chloramphenicol	5	10	25	50	100

Table 1: Antibiotic Concentrations. Antibiotic sensitivity of the ΔPA5030 strain was tested against ampicillin, gentamycin, and chloramphenicol. The strain was grown overnight in LB media with increasing concentrations each antibiotic. Wild-type *P. aeruginosa* and cultures with no antibiotic were used as control.

Confirming plasmid construct by digestion. Proper construction of the allele-carrying vector was confirmed by restriction enzyme digestion. A digestion with BamHI cuts once in the vector carrying the deletion allele, and twice in the empty vector, as an extra site lies within a section of the plasmid that is excised during the BP-clonase reaction (Fig. 2). Conversely, a second digestion with NheI cuts once in the empty vector, and twice in the vector carrying the deletion allele (Fig. 2). Reactions with each enzyme but no DNA template, and reactions with each DNA template but no enzyme were used as controls. All reactions were analyzed by agarose gel electrophoresis.

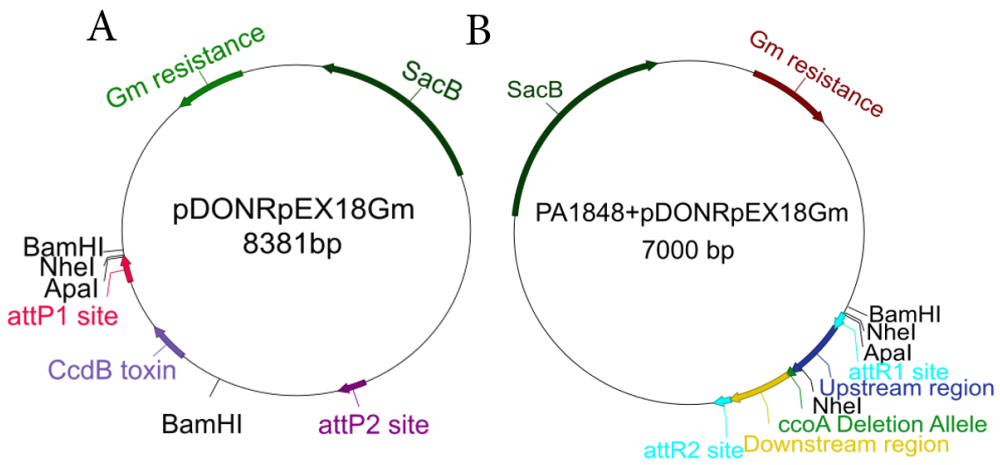


Figure 2: Plasmid constructs. A) Wild-type empty pDONRpEX18Gm vector. B) pDONRpEX18Gm vector after the introduction of the mutant allele by BP clonase reaction.

Bioinformatic analysis

Finding homologs of PA5030.

Filtering manually. The top 5000 sequences aligning to PA5030 were obtained by running the sequence of the protein (44) through a BLASTp in the NCBI database (2), excluding the Pseudomonas family. The 5000 sequences were filtered by selecting every sequence that belonged

to a species that had not been previously selected, eliminating duplicates. Sequences that did not belong to a defined species (ie. an uncultured bacteria) were ignored. Only sequences with a percent identity between 60% and 100% were selected. The aligned sequences were downloaded and sent through the Clustal Omega alignment algorithm (12,30,38). The alignment was visualized in JalView (43). Methionines, cysteines, and histidines conserved 50% or higher were tabulated.

Filtering with the "Filter BLAST Searches" script. The top 5000 sequences aligning to PA5030 were obtained by running the sequence of the protein (44) through a BLASTp in the NCBI database (2), excluding the Pseudomonas family. The downloaded file was renamed "FileToFilter", and moved to the folder containing the "Filter Blast Searches" python script. The file produced by running the script, always named "filtered", was renamed to prevent further uses of the script to alter the contents of the file. The file was submitted to the Clustal Omega alignment algorithm (12,30,38), and the alignment file obtained was visualized in JalView (43). Methionines, cysteines, and histidines conserved 50% or higher were tabulated. Note: If only a range of percent identities was desired, entries outside of the range were manually deleted in the file downloaded from the NCBI BLAST.

Modeling of the PA5030 protein

The full sequence was obtained from the Pseudomonas database (44) and ran through three servers which utilize three different modelling algorithms: I-Tasser (36,45,48), Swiss-Model (3-6,13), and EV-coupling (16,17,27,28).

Phylogenetic analysis

A multiple sequence alignment of PA5030 and its identified homologs was sent through the GBlocks cleaning algorithm, which eliminates poor alignments to improve the accuracy of phylogenetic analysis (8,40). An un-rooted phylogenetic tree was created from the resulting alignment using PhyML (23), and visualized using iTOL (24-26).

Results

Successful development of a ΔPA1848/ΔPA5030 double mutant strain. In order to gain understanding of the interactions of proteins in the *P. aeruginosa* copper homeostasis network, here we knock out the PA1848 gene in the ΔPA5030 mutant background to create a double mutant strain that can be used for future experiments exploring their relationship. This was achieved by carrying a deletion mutation using the recombination techniques outlined in Hmelo et al. (2015). Two other mutations, PA4221 in the ΔPA5030 background and in the PA4221 in ΔPA3790 strain, were attempted with unsuccessful results.

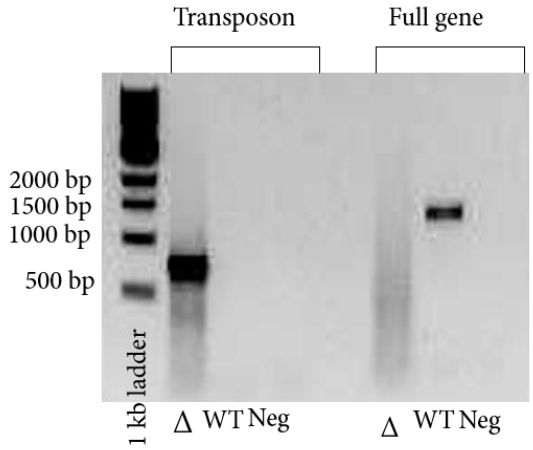


Figure 3: Mutant confirmation. PCR was used to amplify a region of the transposon insertion (left), and to amplify the full gene (right). The transposon could only be amplified in the mutant and the gene could only be amplified in the wild-type.

Polymerase chain reaction (PCR) using primers sitting upstream of the insertion and within the transposon was performed to confirm that the ΔPA5030 strain had the desired mutation (Fig. 1A). The presence of the transposon insertion was confirmed by the appearance of a 746 bp band for the mutant strains while no band appeared for the wild-type strain (PAO1) as expected (Fig. 3). Furthermore, PCR showed that the mutant gene could not be amplified in full in either mutant strain (Fig. 3).

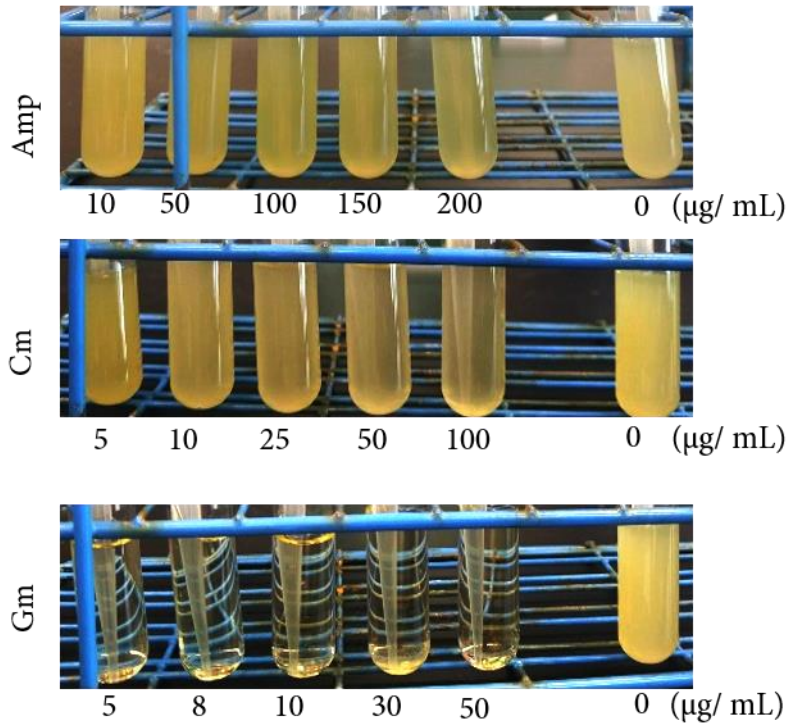


Figure 4: Antibiotic sensitivity. The ΔPA5030 strain was grown overnight in LB with incresing concentrations of antibiotic. It was able to grow in both ampicilling and chloramphenicol, but unable to grow in gentamycin.

In order to select a vector carrying the appropriate antibiotic resistance for the ΔPA5030 strain, antibiotic sensitivity against ampicillin, gentamycin, and chloramphenicol was tested. The strain was able to grow in ampicillin and chloramphenicol even at the highest concentration tested, and unable to grow in any concentration of gentamycin (Fig. 4).

Amplification of the upstream and downstream regions of the target gene, followed by splicing by overlap extension (SOE) PCR, yielded products of 477 bp (upstream), 436 bp (downstream) and 894 bp (SOE-PCR product) (Fig. 5A). Following introduction of the deletion allele into the vector pDONRpEX18Gm and transformation into E. *coli* DH5-α, a successful colony was found and selected to purify the allele-carrying vector (Fig. 5B). Two restriction enzyme digestions were carried to further confirm the vector construct was correctly built. Digestion of the purified plasmid with NheI showed that the allele was properly inserted into the plasmid (Fig. 5C).

Digestion of the purified plasmid with BamHI showed that the CcdB toxin cassette, which serves to kill cells transformed with the empty plasmid, was properly removed from the plasmid during the Gateway reaction (Fig. 5C). The confirmed purified plasmid was transformed into E. *coli* S-17.1, and a positive colony was selected to perform biparental conjugation.

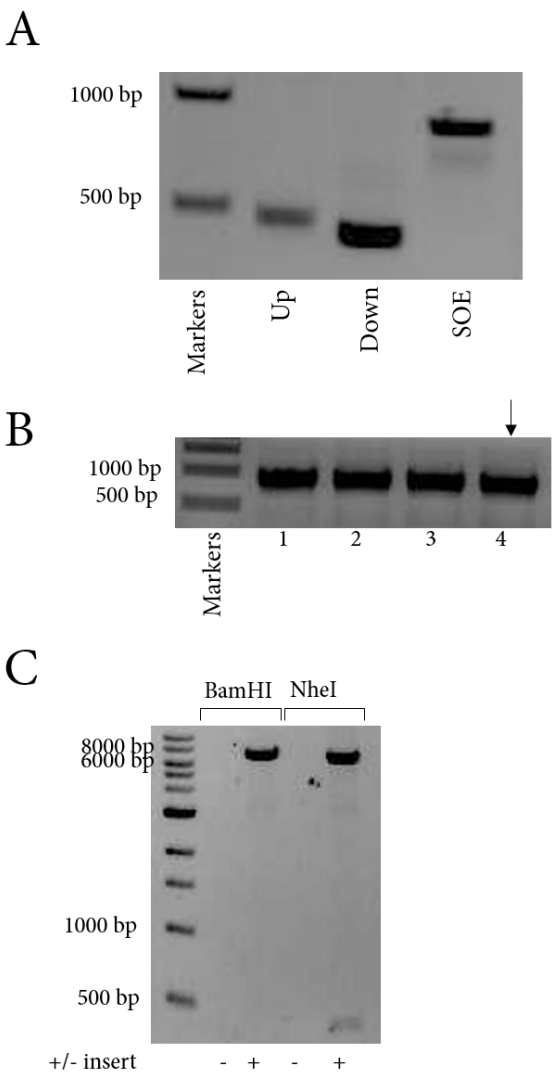


Figure 5: Building plasmid construct. A) Upstream and downstream region of PA5030 were amplified, and then used as templates for a second round of PCR (SOE).

Colonies isolated from counter selection no-salt LB + 15 % sucrose (NSLB) plates were screened by PCR and plated on LB plates with no antibiotic, LB + Gm50 plates, and Pseudomonas isolation agar (PIA) plates. For PA1848/ΔPA5030, 4 out of the 12 colonies screened carried the mutant allele, while the rest carried the wild-type allele (Fig. 6A). All four of those colonies exhibited normal mutant behavior, growing in the LB with no antibiotic and PIA plates, but not in the LB + Gm50 plates (Fig. 6B).

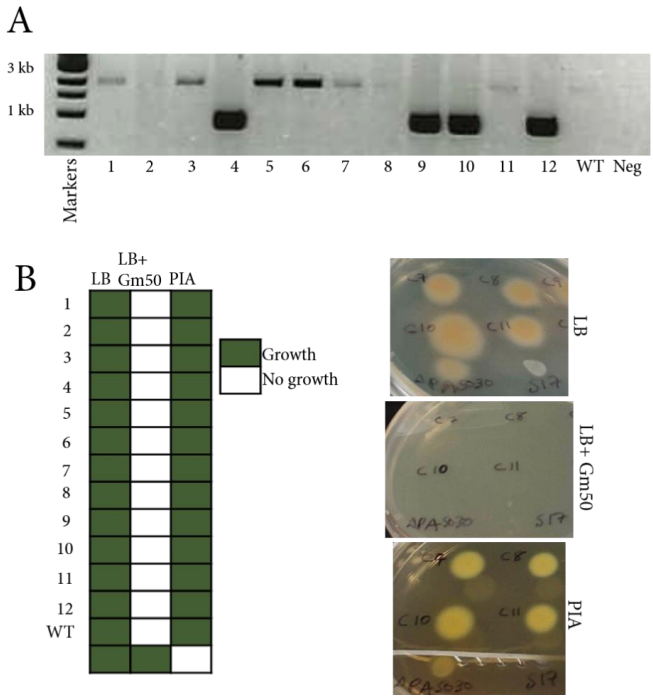


Figure 6: Screening. A) PCR amplifying the PA1848 gene. 12 colonies were checked, of whoch four carried the mutant allele. B) Colonies were grown in LB agar plates without antibiotics, and with Gm50, and in *Pseudomonas* Isolation Agar (PIA). All colonies grow in LB plates, but the mutants, which have lost the gentamycin cassette, cannot grow in Gm50. PIA was used to confirm the colonies are *P. aeruginosa.*

One mutant colony from the PA1848/ΔPA5030 NSLB plate was further confirmed by PCR. In one PCR reaction, the protocol to check for the ΔPA5030 transposon was repeated, to ensure the original mutation was still present. In a second PCR reaction, recombination of the mutant allele in the proper location of the genome was verified by amplifying with the 5' primer used for the downstream region and a 3' primer sitting further downstream of the allele (Fig. 1A). Just as the wild-type strain, the mutant colony yielded a band at 746 bp for the first reaction, and a band at 426 bp for the second PCR reaction (Fig. 7). An overnight culture from one colony was stored in 16% glycerol at -80°C and archived for future uses.

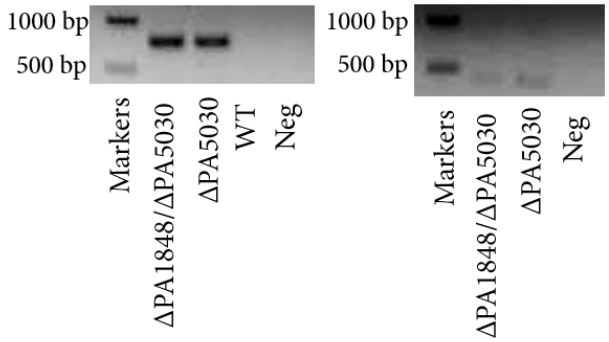


Figure 7: Final mutant confirmation. PCR was used to amplify a region of the transposon insertion. (left), and to amplify the downstream region of the gene including the area outside the deletion allele to confirm correct localization of the mutation (right).

Bioinformatics

Homologs of PA5030 show conserved residues related to copper binding. Performing a BLASTp with the complete protein sequence of PA5030, selecting to view the top 5000 results, and excluding the Pseudomonas family, yields 5000 hits with a percent identity to PA5030 between 35% and 100%. These 500 hits, however, include many sequences corresponding to the same species and hits not corresponding to a defined species (ie. "uncultured bacteria"). At this point, two different methods were used to filter these results. Filtering manually involved going through every hit and only selecting the appropriate ones that had not been previously seen. Due to time constraints, the manual filtering only consisted of those results with percent identities 60% or higher. Of the 1772 hits with at least 60% identity, 198 results were identified. In order to reduce the amount of time taken to filter, a Python script was written to automate the process. The script tries to mimic the filtering decisions made by hand, and therefore only includes unique results and excludes results deemed low quality, such as those with no defined species. Running the script with the same 1772 hits with at least 60% identity yields 203 unique and significant results. A comparison of the 198 results obtained manually and the 203 results obtained by the script shows that the manually filtered file still contained 11 duplicate or non-significant results. Furthermore, the script was able to identify an additional 16 unique and significant hits that were missed by the manual filtering. When the full 5000 results were filtered with the script, 892 species were identified. An additional BLASTp was performed using only the sequences of the transmembrane domains

and changing all other residues to "X". The top 5000 hits from this search were run through the script, and an additional 33 species were identified. These results are summarized in Table 2.

(BLAST full seq)(60-100) 1772 -> (filtered) 198
(BLAST full seq)(60-100) 1772 -> (filtered) 203
(BLAST full seq)(35-100) 5000 -> (filtered) 892
(BLAST full seq)(35-100) 5000 + (BLAST TM) 5000 -> (filtered) 925

Table 2: BLAST results. Number of sequences obtained with the different filtering methods.

Unsurprisingly, all the filtered sequences belong to proteins in the Major Facilitator Superfamily (MFS). An identical protein, SAJ30679.1 ynfM in *Enterobacter cloacae,* was found. While most of the filtered proteins were only labeled as MFS transporters, 2 proteins were identified as potential multidrug resistance (MDR) transporters, 14 proteins were labeled as permeases, and of those, 3 were identified as potential arabinose efflux transporters. Using these proteins as a search query revealed an additional 4 potential multidrug resistance transporters similar to PA5030, an extra 59 potential permeases, and another 8 potential arabinose efflux permeases.

Furthermore, the filtered sequences were aligned, visualized, and methionines, cysteines, and histidines conserved 50% or higher were tabulated. When aligning the sequences with 60% identity or higher, 9 conserved methionines, 1 conserved cysteine, and 2 conserved histidines were identified (Table 3). The same residues were identified regardless of whether the BLAST search results were manually filtered or if the script was used. Similarly, the percentage of conservation

shown for the residues did not vary by more than 4% between the two filtering techniques. When aligning the sequences filtered from the full 5000 hits, 6 conserved methionines, 1 conserved cysteine, and 1 conserved histidine were identified (Table 3). The additional 33 species, identified by aligning the transmembrane domains, slightly lower the percentage of conservation, however which residues are identified as highly conserved does not change, and the decrease in conservation was at most 1.5% (Table 3).

TM	% conserved (manual 60%)	% conserved (AUTO >60%)	% conserved (AUTO >35%)	% conserved (AUTO>60%+TM)	% conserved (AUTO>35%+TM)
1	C65: 91.4%	C65: 90.1%	C65: 58.3%	C65: 58.6%	C65: 57.6%
2	M94: 60.6%	M94: 56.7%	-	-	-
3	M115: 98.5% M132: 94.4%	M115: 96.6% M132: 93.6%	M115: 98.2%	M115: 98.2%	M115: 97.9%
4	M141: 90.9% M157: 100.0%	M141: 88.7% M157: 100.0%	M157: 99.3%	M157: 99.0%	M157: 98.7%
5	M173: 98.5% M185: 100.0%	M173: 98.5% M185: 100.0%	M173: 91.1% M185: 83.0%	M173: 89.9% M185: 82.3%	M173: 89.7 M185: 81.8%
6	-	-	-	-	-
7	M262: 100.0%	M262: 99.5%	M262: 88.9%	M262: 90.7	M262: 88.6%
8	-	-	-	-	-
9	M325: 98.0%	M325: 98.0%	M325: 67.3%	M325: 66.1%	M325: 66.9%
10	H355: 98.5	H355: 98.5%	H355: 99.7	H355: 95.5%	H355: 99.1
11	H397: 73.2	H397: 72.9%	-	-	-
12	-	-	-	-	-

Table 3: Conserved residues. Cysteine, methionine, and histidine residues conserved 50% or higher in the different transmembrane domains (TM) of the homologs obtained through manual filtering (manual) and using the filtering script (AUTO), using BLAST sequences with either greater than 60% identity or greater than 35% identity.

Homology of these proteins was further confirmed using phylogenetic analysis. A sugar efflux transporter (SetA), an iron-regulated transporter (PDB: 5AYN), two known MDR transporters (MdfA and EmrD), two copper importers (CcoA and OprC), a different arabinose transporter (AraJ), a non-MDR sugar transporter (LacY), and the protein used as a template to

make the PA5030 model (yajR) were added to the analysis. The developed phylogenetic tree showed high clustering of the 925 sequences with PA5030 when compared to the non-homologous sequences with similar functions (Fig. 8). PA5030 and its homologs cluster closer to any of the sugar transporters than to any of the heavy-metal transporters or the known MDR proteins.

In-silico modeling of PA5030 reveals two potential copper binding sites. 17 models obtained by running the full PA5030 sequence through three different modelling servers, were compared. The two best models, one from I-Tasser and one from Swiss-Model, were selected based on knowledge of the characteristics of the protein. In both models, residues M157-M173-H355 are located close enough to coordinate copper binding (Fig. 9). A second site, C65-M185-M262, was identified (Fig. 9). While the residues in this second site are located far apart in the outward-open conformation of the protein shown in the model, it is possible that these residues are brought together by the conformational change induced by substrate binding.

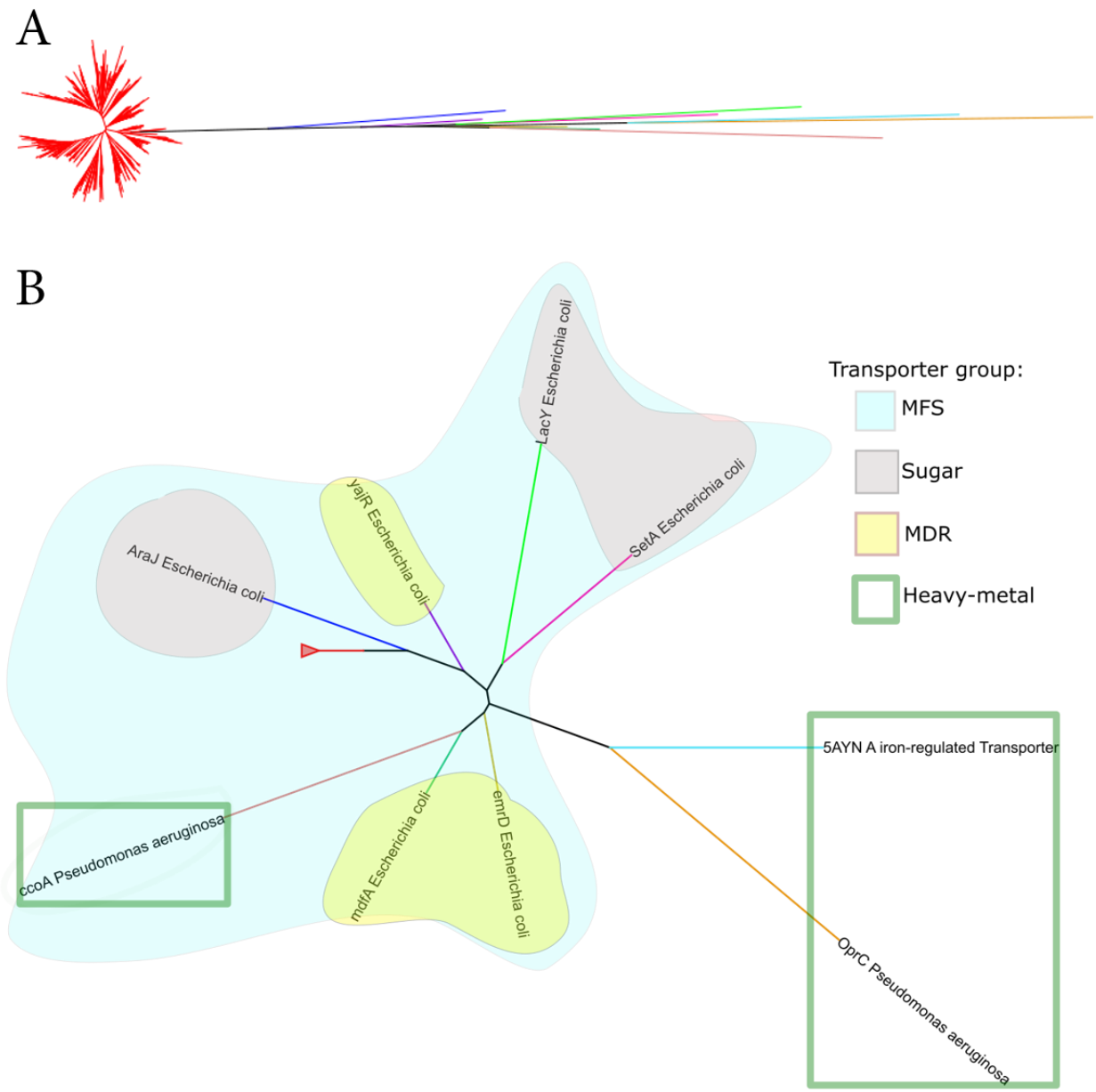


Figure 8: Phylogenetic analysis. A) Phylogenetic tree of PA5030 and its homologs (red), and other proteins with similar functions. B) The clade containing PA5030 and its homologs was condensed (red triangle) to allow for better viewing of the distances to the other proteins. Here, the proteins are also divided by categories corresponding to the family of transporters they belong to.

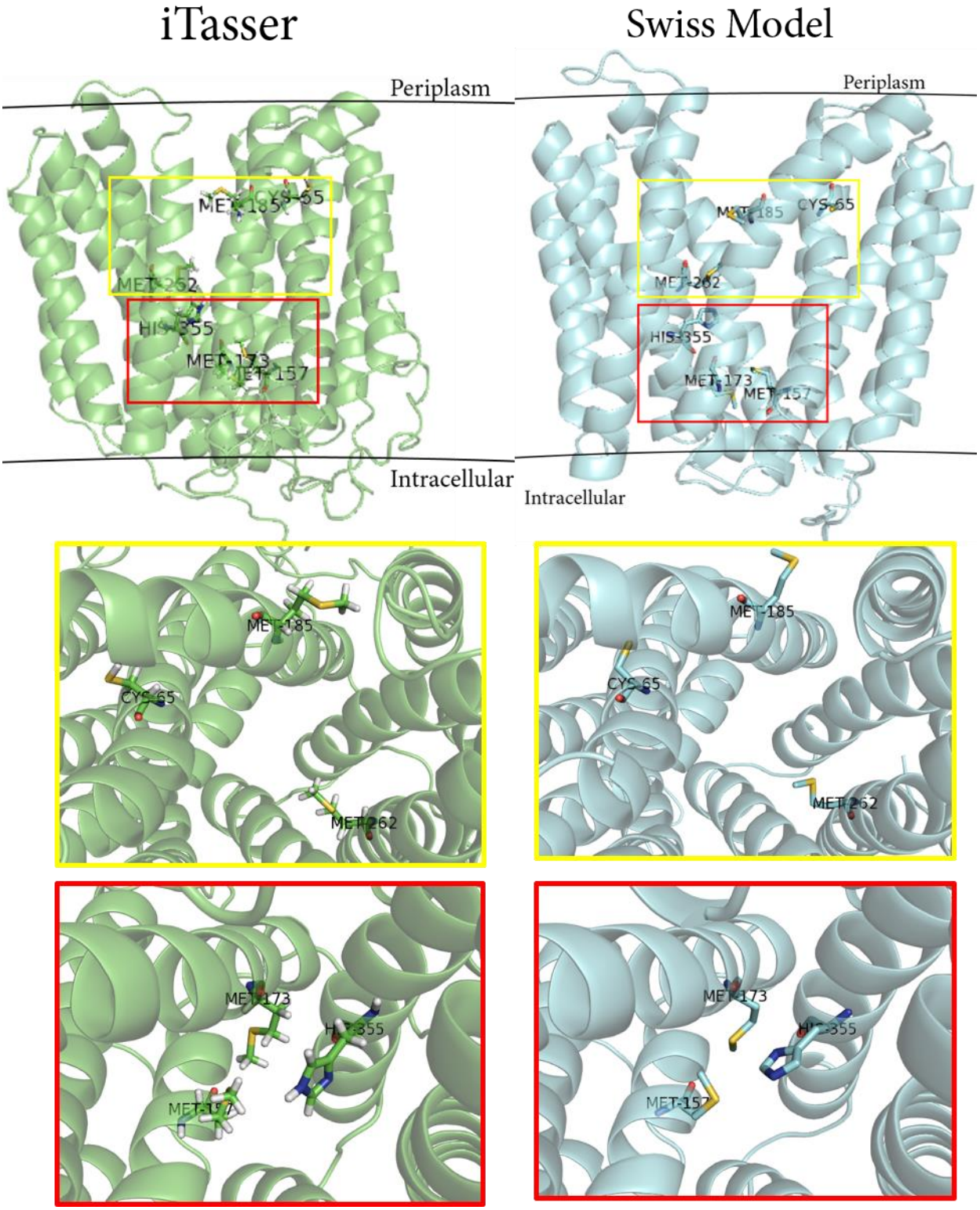


Figure 9: Copper binding sites. (Top) The predicted outward-open Cu^+ binding site is shown in a red box, and the predicted inward-open Cu^+ binding site is shown in a yellow box. B) Close-up of the residues in the outward-open Cu^+ binding site.

Discussion

PA5030 is a Cu^+ transport protein.

One goal of this project was to identify possible homologs of the PA5030 protein. Of particular interest were ynfM in *Enterobacter cloacae,* a protein identical to PA5030, and ynfM in *Acinetobacter baumannii*, a protein with a single P->A substitution. There are no publications on *E. cloacae* ynfM or *A. baumannii* ynfM, however, the protein can be compared to *Escherichia coli* ynfM, which when aligned to PA5030 with AlignMe, an algorithm optimized for membrane proteins (39), shows 58% identity and 74% similarity at 95% coverage. Notably, *E. coli* ynfM has also been suspected to be an arabinose efflux transporter (22). While this information seems to contradict the previous findings that suggest that PA5030 is a copper importer (34), no concrete evidence was found to support that they apply to PA5030. Therefore, with the addition of the results presented in this paper, PA5030 being a Cu^+ transport protein remains as the most plausible conclusion.

PA5030 transports Cu^+ across the membrane through binding at two sites.

The presence of two analogous binding sites at either end of the protein suggests that the protein might obtain copper at one side of the membrane, and undergo a conformational change that changes the shape of the protein from outward-open to inward-open, enabling the copper to be transferred to the second binding site and out into the other side of the membrane. This hypothesis is consistent with both the "clamp-and-switch" model of transport and the "rocker-switch" mechanism of MFS transporters (10,35). However, no information can be derived on the exact

mechanism of this transfer, or whether certain chaperones need to participate in the exchange. Furthermore, the accuracy of the model itself is limited by the shortage of crystal structures available to serve as templates. Considering the best template was selected for this project based on current knowledge of PA5030, perhaps the increased knowledge on PA5030's function could lead to the use of a different template, and a more accurate model. For example, the models derived by using the available crystal structures of arabinose transport proteins as templates could be explored. Additionally, further research into the binding affinity of known Cu^+ chaperones to PA5030 could yield information on the transfer mechanism of this protein.

Conclusion.

In conclusion, future research into this protein should focus on determining its functional role and confirming its location in the inner membrane. Based on the findings in this paper, it would be interesting to check the proteins binding affinity to Cu^+, arabinose, and drugs coordinated by heavy metals. It would also be interesting to repeat the bioinformatics portion of this paper, but with a focus on finding an arabinose binding site in PA5030. The ΔPA1848/ΔPA5030 double mutant can be used in future research to determine the strain's ability to uptake Cu^+. Past research into the ΔPA1848 and ΔPA5030 have shown that Cu^+ stress represses expression of PA5030 but not PA1848 (ccoA) (34). Therefore, it would be interesting to see if tolerance to Cu^+ is affected in the ΔPA1848/ΔPA5030 strain's, when compared to the single mutants. This research could give clues to the different regulatory pathways in the copper homeostasis network. Finally, if PA5030 is

indeed a MDR protein, research into its repression under Cu^+ stress could lead to the development of methods to prevent the efflux of antibiotics and lead to better treatment of infections.

References

1. Ala, A., Walker, A. P., Ashkan, K., Dooley, J. S., & Schilsky, M. L. Wilson's disease. *The Lancet, 369*(9559), 397-408. doi:10.1016/S0140-6736(07)60196-2

2. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410. doi:https://doi.org/10.1016/S0022-2836(05)80360-2

3. Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics, 27*(3), 343-350. doi:10.1093/bioinformatics/btq662

4. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports, 7*(1), 10480. doi:10.1038/s41598-017-09654-8

5. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research, 42*(Web Server issue), W252-W258. doi:10.1093/nar/gku340

6. Bienert, S., Waterhouse, A., de Beer, Tjaart A P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research, 45*(Database issue), D313-D319. doi:10.1093/nar/gkw1132

7. Breathnach, A. S. Nosocomial infections. *Medicine, 37*(10), 557-561. doi:10.1016/j.mpmed.2009.07.008

8. Castresana, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution, 17*(4), 540-552. doi:10.1093/oxfordjournals.molbev.a026334

9. Ekici, S., Yang, H., Koch, H.-G., & Daldal, F. (2012). Novel Transporter Required for Biogenesis of cbb(3)-Type Cytochrome c Oxidase in Rhodobacter capsulatus. *mBio, 3*(1), e00293-00211. doi:10.1128/mBio.00293-11

10. Forrest, L. R., Krämer, R., & Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1807*(2), 167-188. doi:https://doi.org/10.1016/j.bbabio.2010.10.014

11. González-Guerrero, M., Raimunda, D., Cheng, X., & Argüello, J. M. (2010). Distinct functional roles of homologous Cu+ efflux ATPases in Pseudomonas aeruginosa. *Molecular Microbiology, 78*(5), 1246-1258. doi:doi:10.1111/j.1365-2958.2010.07402.x

12. Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research, 38*(suppl_2), W695-W699. doi:10.1093/nar/gkq313

13. Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *ELECTROPHORESIS, 30*(S1), S162-S173. doi:doi:10.1002/elps.200900140

14. Hardalo, C., & Edberg, S. C. (1997). Pseudomonas aeruginosa: Assessment of Risk from Drinking Water. *Critical Reviews in Microbiology, 23*(1), 47-75. doi:10.3109/10408419709115130

15. Held, K., Ramage, E., Jacobs, M., Gallagher, L., & Manoil, C. (2012). Sequence-Verified Two-Allele Transposon Mutant Library for Pseudomonas aeruginosa PAO1. *Journal of Bacteriology, 194*(23), 6387-6389. doi:10.1128/JB.01479-12

16. Hmelo, L. R., Borlee, B. R., Almblad, H., Love, M. E., Randall, T. E., Tseng, B. S., . . . Harrison, J. J. (2015). Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. *Nature Protocols, 10*, 1820. doi:10.1038/nprot.2015.115

17. Hopf, T. A., Colwell, L. J., Sheridan, R., Rost, B., Sander, C., & Marks, D. S. (2012). 3D structures of membrane proteins from genomic sequencing. *Cell, 149*(7), 1607-1621. doi:10.1016/j.cell.2012.04.012

18. Hopf, T. A., Schärfe, C. P. I., Rodrigues, J. P. G. L. M., Green, A. G., Kohlbacher, O., Sander, C., . . . Marks, D. S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife, 3*, e03430. doi:10.7554/eLife.03430

19. Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., . . . Manoil, C. (2003). Comprehensive transposon mutant library of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences, 100*(24), 14339-14344. doi:10.1073/pnas.2036282100

20. Khalfaoui-Hassani, B., Verissimo, A. F., Koch, H.-G., & Daldal, F. (2016). Uncovering the Transmembrane Metal Binding Site of the Novel Bacterial Major Facilitator Superfamily-Type Copper Importer CcoA. *mBio, 7*(1). doi:10.1128/mBio.01981-15

21. Khan, H. A., Ahmad, A., & Mehboob, R. (2015). Nosocomial infections and their control strategies. *Asian Pacific Journal of Tropical Biomedicine, 5*(7), 509-514. doi:https://doi.org/10.1016/j.apjtb.2015.05.001

22. Koita, K., & Rao, C. V. (2012). Identification and Analysis of the Putative Pentose Sugar Efflux Transporters in Escherichia coli. *PLOS ONE, 7*(8), e43700. doi:10.1371/journal.pone.0043700

23. Lee, S. M., Grass, G., Rensing, C., Barrett, S. R., Yates, C. J. D., Stoyanov, J. V., & Brown, N. L. (2002). The Pco proteins are involved in periplasmic copper handling in Escherichia coli. *Biochemical and Biophysical Research Communications, 295*(3), 616-620. doi:https://doi.org/10.1016/S0006-291X(02)00726-X

24. Lefort, V., Longueville, J.-E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution, 34*(9), 2422-2424. doi:10.1093/molbev/msx149

25. Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics, 23*(1), 127-128. doi:10.1093/bioinformatics/btl529

26. Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research, 39*(Web Server issue), W475-W478. doi:10.1093/nar/gkr201

27. Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research, 44*(Web Server issue), W242-W245. doi:10.1093/nar/gkw290

28. Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., & Sander, C. (2011). Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLOS ONE, 6*(12), e28766. doi:10.1371/journal.pone.0028766

29. Marks, D. S., Hopf, T. A., & Sander, C. (2012). Protein structure prediction from sequence variation. *Nature Biotechnology, 30*, 1072. doi:10.1038/nbt.2419

30. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., . . . Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research, 41*(W1), W597-W600. doi:10.1093/nar/gkt376

31. O'Toole, G. A., & Kolter, R. (1998). Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development. *Molecular Microbiology, 30*(2), 295-304. doi:doi:10.1046/j.1365-2958.1998.01062.x

32. Padilla-Benavides, T., George Thompson, A. M., McEvoy, M. M., & Argüello, J. M. (2014). Mechanism of ATPase-mediated Cu+ Export and Delivery to Periplasmic

Chaperones: THE INTERACTION OF ESCHERICHIA COLI CopA AND CusF. *Journal of Biological Chemistry, 289*(30), 20492-20501. doi:10.1074/jbc.M114.577668

33. Porcheron, G., Garenaux, A., Proulx, J., Sabri, M., & Dozois, C. (2013). Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. *Frontiers in Cellular and Infection Microbiology, 3*(90). doi:10.3389/fcimb.2013.00090

34. Quintana, J., Novoa-Aponte, L., & Argüello, J. M. (2017). Copper homeostasis networks in the bacterium Pseudomonas aeruginosa. *Journal of Biological Chemistry, 292*(38), 15691-15704. doi:10.1074/jbc.M117.804492

35. Quistgaard, E. M., Löw, C., Guettou, F., & Nordlund, P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nature Reviews Molecular Cell Biology, 17*, 123. doi:10.1038/nrm.2015.25

36. Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols, 5*(4), 725-738. doi:10.1038/nprot.2010.5

37. Saier, M. H., & Paulsen, I. T. (2001). Phylogeny of multidrug transporters. *Seminars in Cell and Developmental Biology, 12*(3), 205-213. doi:10.1006/scdb.2000.0246

38. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*(1). doi:10.1038/msb.2011.75

39. Stamm, M., Staritzbichler, R., Khafizov, K., & Forrest, L. R. (2013). Alignment of Helical Membrane Protein Sequences Using AlignMe. *PLOS ONE, 8*(3), e57731. doi:10.1371/journal.pone.0057731

40. Stewart, P. S., & William Costerton, J. Antibiotic resistance of bacteria in biofilms. *The Lancet, 358*(9276), 135-138. doi:10.1016/S0140-6736(01)05321-1

41. Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology, 56*(4), 564-577. doi:10.1080/10635150701472164

42. Thaden, J. T., Lory, S., & Gardner, T. S. (2010). Quorum-Sensing Regulation of a Copper Toxicity System in Pseudomonas aeruginosa. *Journal of Bacteriology, 192*(10), 2557-2568. doi:10.1128/JB.01528-09

43. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*(9), 1189-1191. doi:10.1093/bioinformatics/btp033

44. Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., & Brinkman, Fiona S L. (2016). Enhanced annotations and features for comparing thousands of

Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Research, 44*(Database issue), D646-D653. doi:10.1093/nar/gkv1227

45. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. *Nature Methods, 12*, 7. doi:10.1038/nmeth.3213

46. Yoneyama, H., & Nakae, T. (1996). Protein C (OprC) of the outer membrane of Pseudomonas aeruginosa is a copper-regulated channel protein. *Microbiology, 142*(8), 2137-2144. doi:doi:10.1099/13500872-142-8-2137

47. Zaragoza, R., Ramírez, P., & López-Pueyo, M. J. (2014). Infección nosocomial en las unidades de cuidados intensivos. *Enfermedades Infecciosas y Microbiología Clínica, 32,* 320-327.

48. Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*(1), 40. doi:10.1186/1471-2105-9-40

Supplementary Material

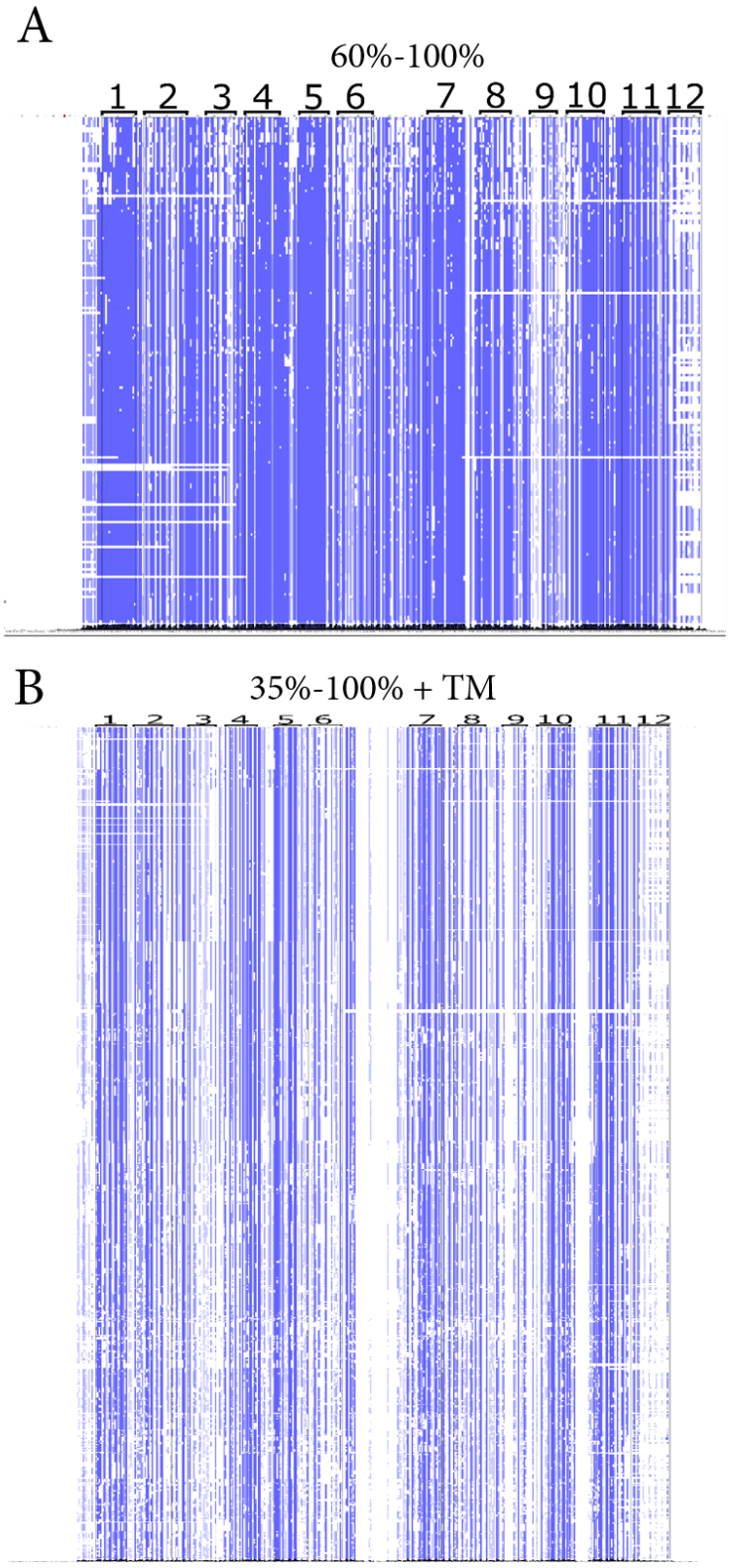
Supplementary Table 1: Primers

Experiment	Primer Name	Primer Sequence	Direction
Check PA5030 mutation	PA5030-F	5' GGGTATGGTCGAGGTGAAG 3'	5'
	PA5030-R	5' CAGGAAGCCGATCAGGAA 3'	3'
	Lac-R	5' GGTTCCGTCCAGGACGCTACTTGTG 3'	3'
PA1848 mutation	PA1848-UP-F	5' GGGGACAAGTTTGTACAAAAAAGCAGGCTACCCCAGGCG TCGATCGCCA 3'	5'
	PA1848-UP-R	5' CGTTGCCGGCGTTTCAGCGGAGCAGGGTCGGGTTGCG 3'	3'
	PA1848-DN-F	5' CGCTGAAACGCCGGCAACG 3'	5'
	PA1848-DN-R	5' GGGGACCACTTTGTACAAGAAAGCTGGGTAGACTACGCCA CCGACCGC 3'	3'
	PA1848-Seq-R	5' TCCTCGAAGATGCGGTGGTG 3'	3'

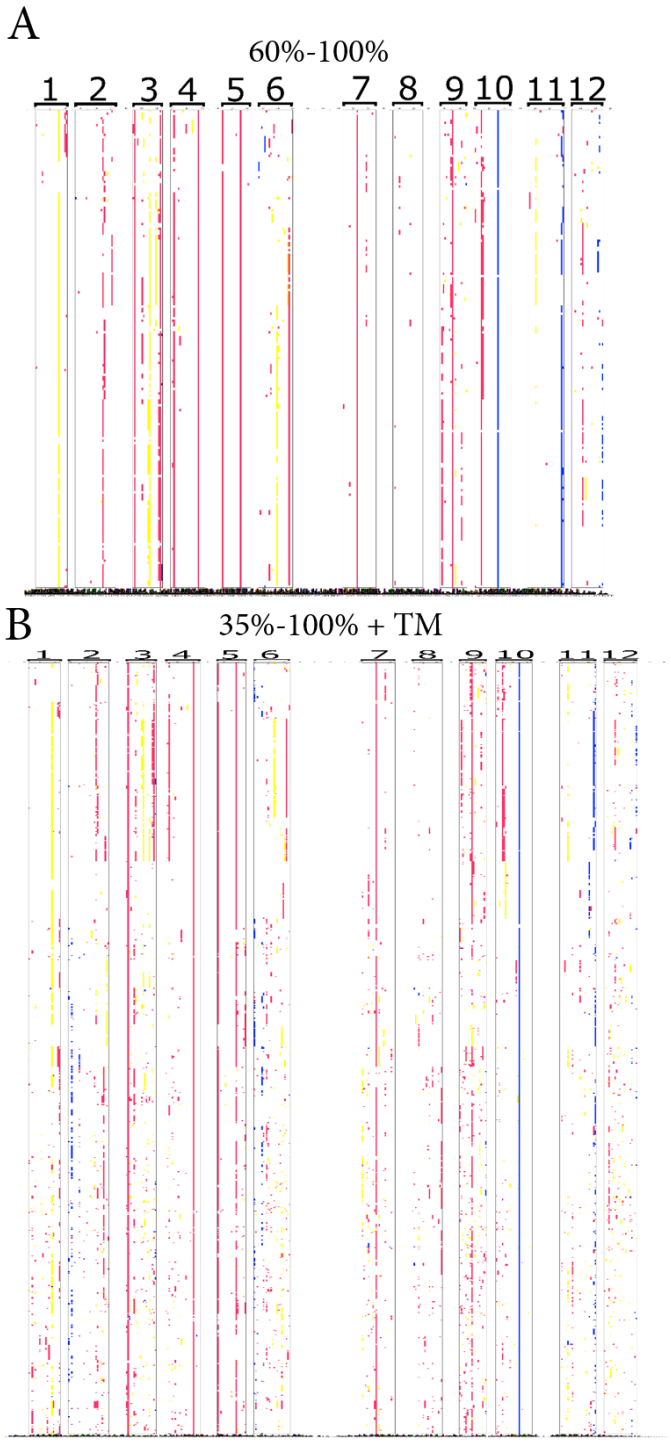
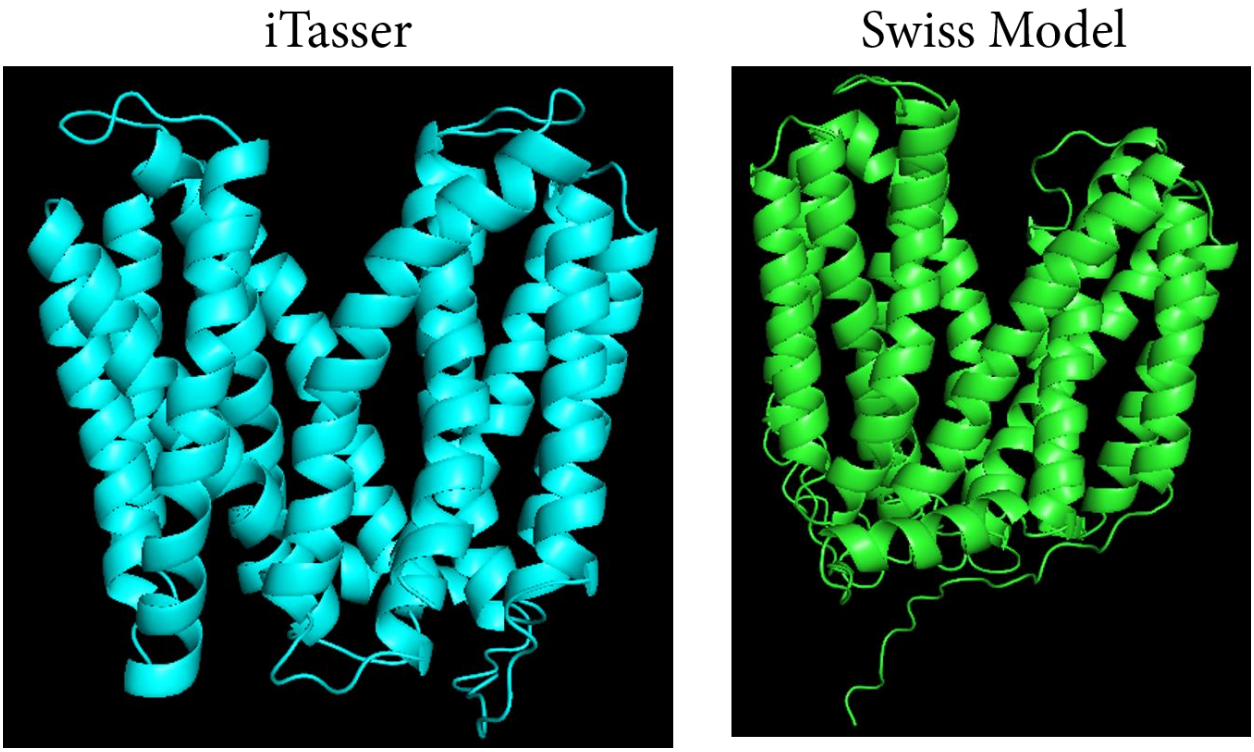
Supplementary Table 2: PCR Programs

		Initial denat.	Denat.	Anneal		Ext.		Final ext.	
				Primers used					
Check mutation	PA5030	94°C/ 5 min	94°C/ 30 sec	PA5030-F + Lac-R	57 °C/ 30 sec	68°C/ 2 min	Repeat Steps 2-4 for an additional 24 cycles	68°C/ 5 min	Hold at 10°C
				PA5030-F + PA5030-R	52 °C/ 30 sec				
Cloning Round 1	PA1848	98°C/ 30 sec	98°C/ 10 sec	PA1848-UP-F + PA1848-UP-R	72°C/ 30 sec	72°C/ 45 sec	Repeat Steps 2-4 for an additional 34 cycles	72°C/ 2 min	Hold at 10°C
				PA1848-DN-F + PA1848-DN-R	68°C/ 30 sec				
				PA1848-UP-F + PA1848-DN-R	72°C/ 30 sec	72°C/ 75 sec			
Screening colonies	PA1848	94°C/ 5 min	94°C/ 30 sec	PA1848-UP-F + PA1848-DN-R	68°C/ 30 sec	68°C/ 2 min	Repeat Steps 2-4 for an additional 24 cycles	68°C/ 5 min	Hold at 10°C
Final mutant confirmation	PA1848	94°C/ 5 min	94°C/ 30 sec	PA1848-DN-F + PA1848-Seq-R	57°C/ 30 sec	68°C/ 45 sec	Repeat Steps 2-4 for an additional 24 cycles	68°C/ 5 min	Hold at 10°C

Supplementary Figure 1: Identity Alignment. Conservation of all residues in the alignment by identity for sequences with >60% identity (A) and >35% identity (B) to PA5030.



Supplementary Figure 2: C/M/H Alignment. All cysteines, methionines, and histidines in the transmembrane domains in the alignment for sequences with >60% identity (A) and >35% identity (B) to PA5030.



Supplementary Figure 3: PA5030 models. Full view of the PA5030 models.