The Industrial Microbiome as a Resource for Biotechnology: Identification and Isolation of Butyl Methacrylate Tolerant *Pseudomonas* Species for Enhanced Sustainable Bio-Production of Methacrylate Esters

Thomas Haworth Hender

Under the supervision of

Dr Nicholas Tucker, Dr Stephen Cartman

And Dr Leighton Pritchard

A thesis submitted in partial fulfilment for the degree of Doctor of Philosophy

Strathclyde Institute of Pharmacy and Biomedical Sciences

2024

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed:

Date:

1 Table of Contents

Table 1 Table of Abbreviations

Abbreviation	Full Term	Abbreviation	Full Term
ABPS	Adaptive Poisson-Boltzmann Solver	MLST	Multi-locus sequence typing
ANI	Average nucleotide identity	MMA	Methyl Methacrylate
ANOVA	Analysis of Variance	MOA	mdtA-oqxB7-arpC
BGC	Biosynthetic Gene Cluster	PCR	Polymerase chain reaction
BLAST	Basic Local Alignment Search Tool	PET	Polyethylene terephthalate
BMA	Butyl Methacrylate	PMMA	Polymethylmethacrylate
CFU	Colony Forming Unit	RND	Resistance nodulation division
DNA	Deoxyribonucleic acid	T6SS	Type VI secretion system
HSD	Honestly significant difference	USML	Unsupervised machine learning
IPCA	Incremental Principal Component Analysis	VFG	Virulence Factor Genes
MA	Methacrylic acid	WGS	Whole Genome Sequencing
MLA	Machine learning algorithms	XLogP3	Octanol-water partition coefficient (logP) prediction

2 Abstract

The increased demand for sustainable acrylic plastic has led to interest in developing a fermentation process to produce the monomers used to make this versatile compound. This would provide an alternative to the current finite and environmentally damaging petrochemical processes. Butyl Methacrylate (BMA) is an appealing precursor to these monomers. The current bottleneck for the fermentation of BMA is toxicity; its hydrophobic properties, which make it ideal for extraction from a bioreactor, are toxic to microorganisms interfering with cell membranes. *Pseudomonas* species, known for thriving in industrially contaminated sites, provide a potential solution to this bottleneck. By targeting industrial environments heavily contaminated with hydrophobic solvents, we isolated *Pseudomonas* species with tolerance to BMA. The most tolerant isolate, *P. putida*, held a plasmid, which significantly contributed to conferring tolerance to BMA. The primary tolerance mechanism was an RND (resistance-nodulation-division) efflux pump, which when transferred to *P. putida* KT2440, significantly increased its tolerance to BMA. Features of this RND efflux pump were used to identify other efflux pumps which may also confer tolerance to hydrophobic compounds. To offer additional tolerance and intracellular protection from BMA, we also used unsupervised machine learning to sort compounds produced by

biosynthetic gene clusters (BGC) and BMA. The natural compound FR-900098 an antimalarial candidate with a similar electrostatic profile to BMA was identified, and the corresponding resistance gene *frbF* associated with the BGC was cloned into a high-copy plasmid with a constitutive promotor. When transformed into *P. putida* KT2440, this resulted in a significant increase in BMA tolerance. In summary, the newly isolated *P. putida* strain (RCS09A) may be advantaged as an BMA production strain, due to its intrinsic tolerance to this hydrophobic molecule and a candidate for engineering for production. Alternatively, the identified RND efflux pump and the resistance protein may be used to 'upgrade' the tolerance of other potential BMA production strains. Improved BMA tolerance could lead to increased product titres and lower the costs of the sustainable fermentation process.

3 Introduction

3.1 Petrochemicals as a Feedstock for the Plastic Industry are Unsustainable.

Petrochemicals are used widely across the globe to satisfy the huge industrial demand for their extensive, valuable applications. Currently 12% of oil usage is devoted to petrochemical feedstocks for manufacturing including for plastics and fertilisers (*The Future of Petrochemicals – Analysis - IEA*, n.d.). It is well documented that refining crude oil and using its products releases greenhouse gases into the environment, contributing to climate change and, more specifically, global warming (Perera, 2018). With the peak in fossil fuel usage predicted to occur between 2027 and 2040, the damage to the climate is only set to increase along with the increase in fossil fuel usage (Zou et al., 2016). Innovative solutions in sustainable chemical production are needed. Industrial Biotechnology can provide alternatives to petrochemical feedstocks and processes.

3.2 The Production of Acrylic Plastics

Acrylic plastics are essential for modern society and are used in medical implants, construction, aviation, robotics, submersibles and scientific research. They are formed by the chemical polymerisation of methacrylate esters such as methyl methacrylate (MMA - Figure 1b) and n-butyl-methacrylate (BMA - Figure 1c) to form polymethylmethacrylate (PMMA - Figure 1). The raw materials for the manufacture of MMA and BMA are currently derived from petrochemical feedstocks. One of the world's largest producers of these acrylic precursors is Mitsubishi Chemical Group (MCHEM). The demand for acrylic plastics has translated into increased production of methacrylate esters and, therefore, the requirement for petrochemical feedstocks. Currently, these precursor molecules are generated through highly energy-intensive petrochemical processes.

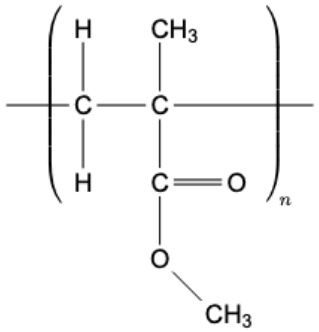


Figure 1: The chemical structure of Polymethylmethacrylate

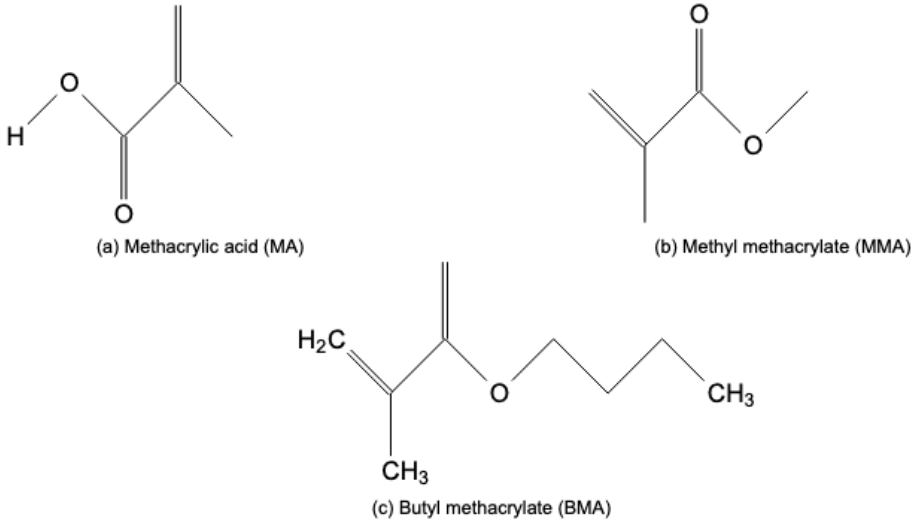


Figure 1 - Precursors of Polymethylmethacrylate. Figure 1a Methacrylic acid, which is converted into Methyl methacrylate, is shown in Figure 1b. Figure 1c Butyl methacrylate is a promising methacrylate ester for bioproduction due to its hydrophobic qualities, which allow for easy separation of the aqueous phase.

3.3 Traditional Routes of Methyl Methacrylate Production by Mitsubishi Chemical Group

Methyl Methacrylate (MMA) was first produced in the 1930s using a process that reacts acetone with hydrogen cyanide to produce acetone cyanohydrin. Sulphuric acid converts the cyanohydrin to methacrylamide sulphate. Then, a methanol and water mixture is used to treat the methacrylamide sulphate whilst being heated. This produces the MMA and an ammonium bisulphate by-product. Mitsubishi Gas Chemical updated the process so that it does not produce ammonium bisulphate waste but instead, a recyclable hydrocyanic acid that can be used again as a raw material.

Mitsubishi Chemical Group (formally Lucite International) developed Alpha (C2) process which takes ethylene, methanol and carbon monoxide to produce MMA (Figure 2). This process does not require toxic hydrogen cyanide and produces no toxic byproducts, in contrast to the acetone (C3) process mentioned above. The Alpha process also reduces the production costs by 40%. Despite the clear advantages of the Alpha C2 process, a petrochemical input derived from refined crude oil is still in use. The carbon-footprint of the petrochemical chemical process is, therefore, still high, and an alternative process must be developed to meet sustainability requirements. .

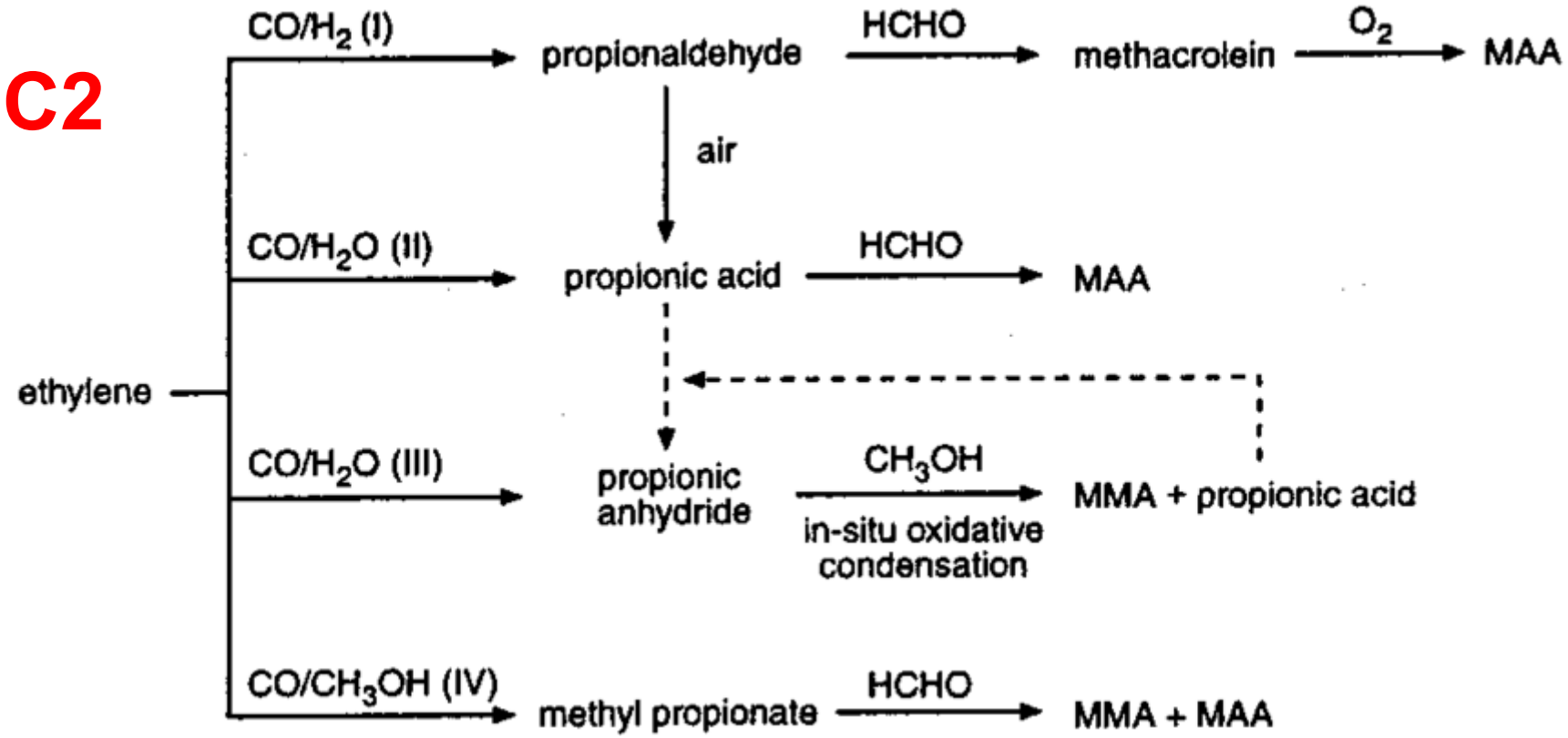


Figure 2 - The Mitsubishi Chemicals Alpha C2 Process for the production of MMA is shown. As shown in the pathway, hydrogen cyanide is not a required feedstock, significantly reducing the volume of toxic waste produced during MMA synthesis (Spivey et al., 1997).

3.4 Synthesis of Novel Molecules and Platform Chemicals for the Plastics Industry Using Microorganisms

Microbial life has been a valuable platform for synthesising many compounds, from proteins such as insulin using bacteria to simpler molecules like antimicrobials and anticancer compounds produced in various organisms (Riggs, 1981; Pham et al., 2019). With advances in the study of novel biosynthetic pathways, these can be tailored to engineer other valuable molecules such as caffeine. This process uses a plant-derived gene encoding a caffeine synthase and a novel biosynthetic pathway to convert glucose into guanine, a precursor of caffeine (M. Li et al., 2017).

The biosynthesis of MMA and other platform chemicals provides an alternative to the petrochemical route and presents an opportunity to utilise sustainably sourced or even recycled feedstocks. The production of Methacrylic Acid (MA) and MMA from biologically derived precursor molecules may be achieved in the laboratory from isobutyric acid, hydroxy isobutyric acid (HIBA), citramalic/citraconic acids, itaconic acid and mesaconic acid, as shown in Figure 3. Routes via the intermediates itaconic acid and citramalic acid (highlighted in blue in Table 2 and Figure 3) have the highest titres to date, but still fall short in terms of their titres to

MA required for an economically viable process at commercial scale (Lebeau et al., 2020).

Table 2 – Current Biological Production routes of precusors to MMA. Highlighting the two most mature biological synthesis routes (Lebeau et al., 2020)

Maturity	Route	Best demonstrated performance	Bioprocess challenges	Chemistry challenges
1	Itaconic acid/decarboxylation	220g/L (*U. maydis*) (Tehrani et al., 2019) 0.45 g/L-h 51% bioprocess yield 48% conversion to MA (Johnson et al., 2009, 2012; Pirmoradi and Kastner, 2017)	Fermentation rates & yields	Yield, catalyst costs
2	Citramalic acid/decarboxylation & dehydration	80g/L (*E. coli*) (Webb et al., 2018) 1.85 g/L-h 58% bioprocess yield 81% conversion to MA (Johnson et al., 2009, 2012; Pirmoradi and Kastner, 2017)	Fermentation rates & yields	Yield, catalyst development
3	Isobutyric acid/dehydrogenation	90g/L (*E. coli*) (Xiong et al., 2015) 0.625 g/L-h 80% bioprocess yield 40–60% conversion to MA (Pirmoradi and Kastner, 2017)	Fermentation rates & yields	Catalyst development
4	2-HIBA/dehydration	6.4g/L (*C. necator H16*) (Hoefel et al., 2010) 0.09 g/L-h 6.3% bioprocess yield 71.5% conversion to MA (Pirmoradi and Kastner, 2017)	Enzymology	Yield, catalyst development
5	Mesaconic acid/decarboxylation	23g/L (*E. coli*) (Wang et al., 2018) 0.36 g/L-h 64% bioprocess yield 52% conversion to MA (Pirmoradi and Kastner, 2017)	Enzymology	Yield, catalyst development
6	Methacrylic acid production	0.0146g/L (*E. coli*) (Eastham et al., 2018) 0.0007 g/L-h 0.62% bioprocess yield	Rates, yields, engineering resistance	NA

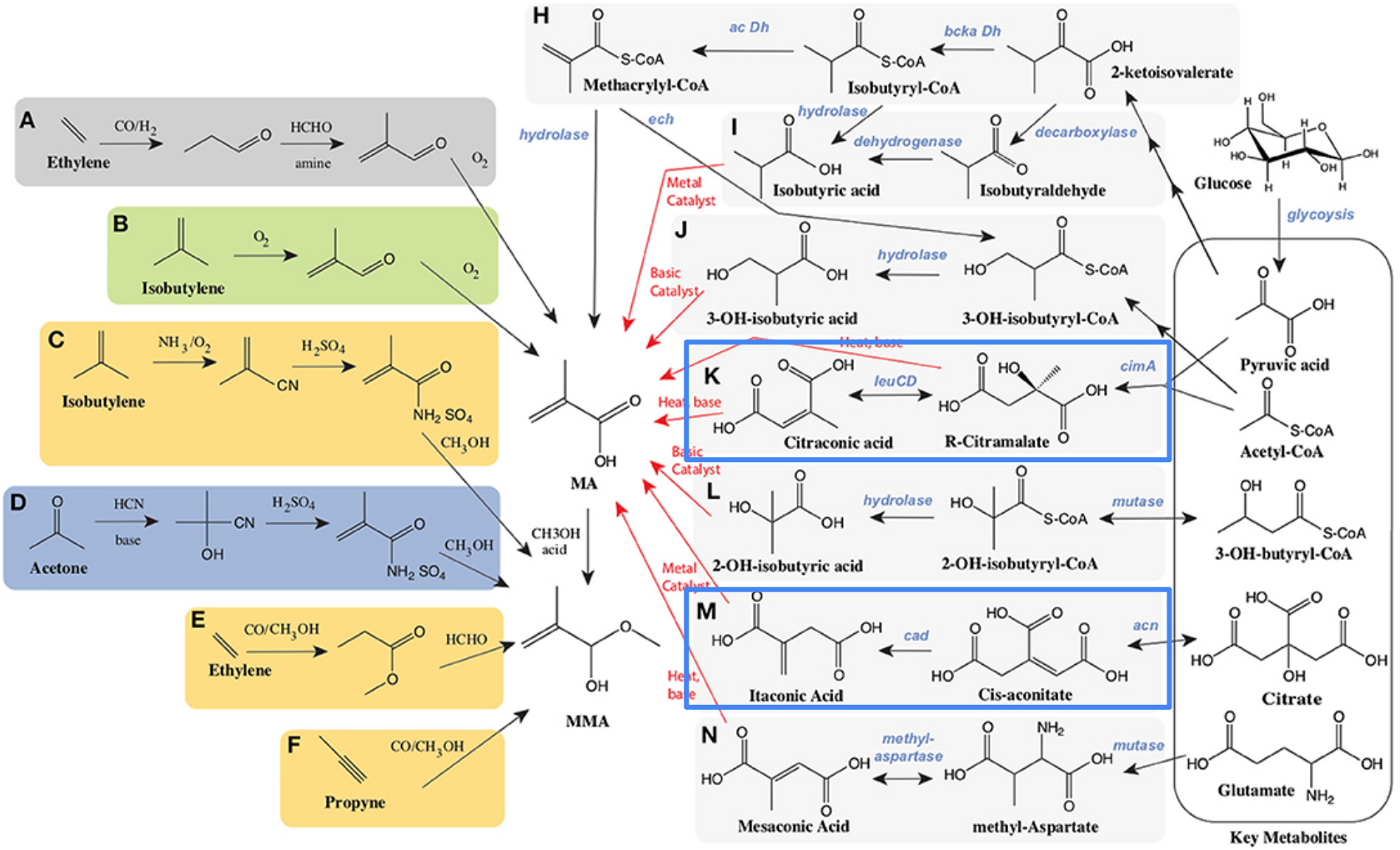


Figure 3 - Biological (H-N) and Petrochemical (A-F) routes to MMA precursor molecules highlighting the two most mature routes (K and M). M is syntheisised using *E. coli* and is curreltly limited by yield at 220g/L and route K produced using *U. maydis* also limited by yeild at 80g/L (Lebeau et al., 2020).

Itaconic acid has been produced in yeasts such as *Aspergiullus terreus* since the 1940s but has production issues due to the propensity for *A. terreus* to form mycelium and stop production. *Ustilago maydis* is a natural producer of Itaconic acid (in which production is not linked with a certain morphology) and is a promising alternative *to A. terreus*. Deleting an itaconate oxidase (Cyp3) and overexpressing the Ria1 regulator produced titres of 220g/L (Okabe et al., 2009). Compared to a titre of 80g/L from *A. terreus,* this makes *U. maydis* a far more promising candidate for further development. Currently, the cost of biologically produced itaconic acid is around $2 per kilogram, and to make it a viable replacement for petrochemical-based products, it would need to be $ 0.5 per kilogram (Lebeau et al., 2020).

Citramalic acid is another promising candidate as a biological intermediate for subsequent chemical conversion to MA. The conversion of citramalic acid to MA is relatively simple and can be done using hot pressurised water (Lebeau et al., 2020). Citramalate can be produced by *E. coil* by expressing a variant of the *cimA3.7* gene for a citramalate synthase from *Methanococcus jannaschii* (see Figure 4 A). Citramalate is subsequently converted to MA/MMA via a chemo-catalytic route (see Figure 4 B). The best titre of citramalate reported to date is 82±1.5g L^{-1} with a conversion efficiency of 0.48g citramalate g^{-1} glucose (Webb et al., 2018).

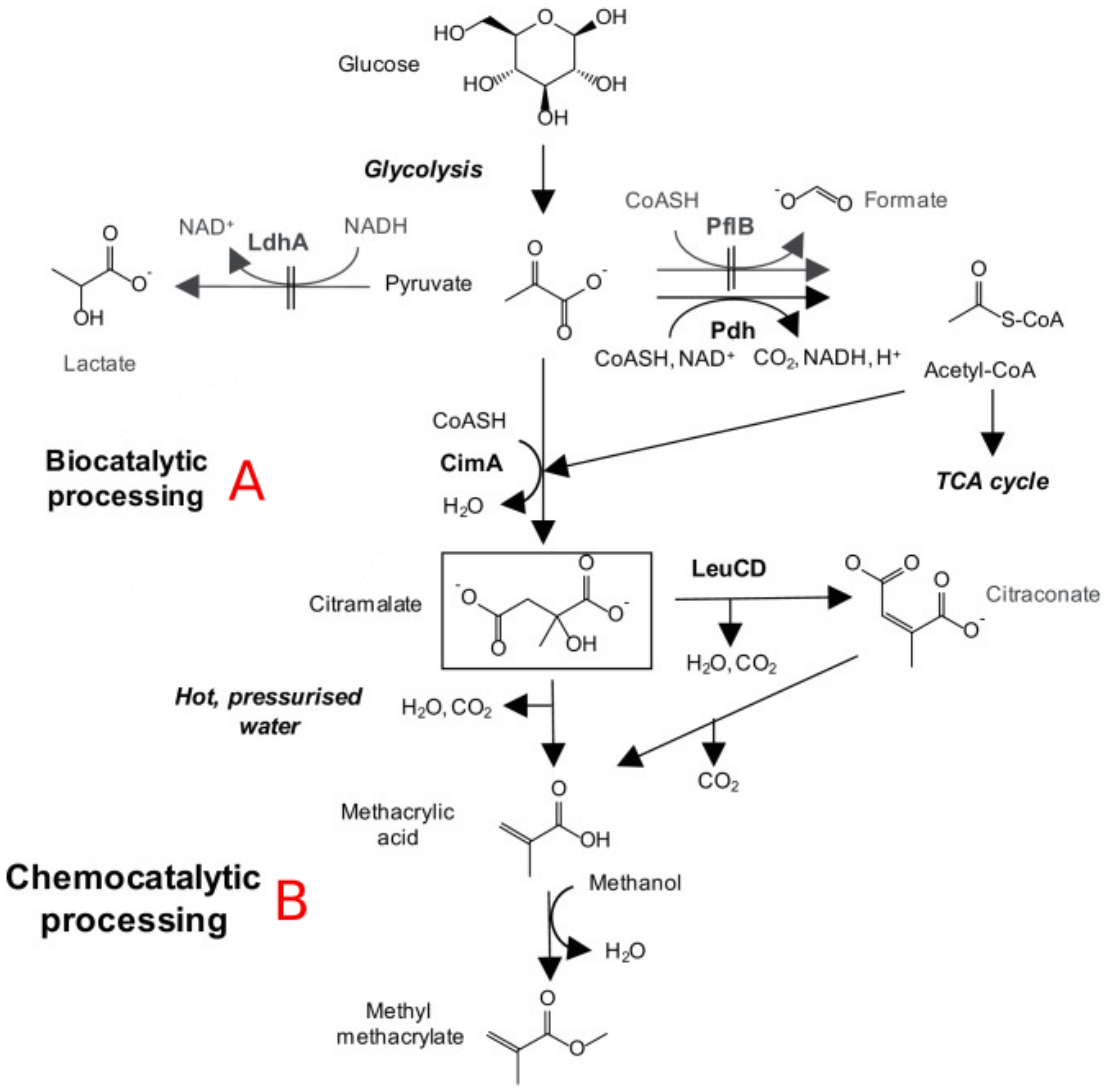


Figure 4 - A) Bioproduction of citramalate from Glucose. Using a citramalate synthase from *Methanococcus jannaschii*. B) Catalytic production of methylmethacrylate from citramalate (Webb et al., 2018).

A common issue with these routes is that the biologically produced intermediate/precursor molecule is generally hydrophilic, making extraction from the bulk fermentation broth challenging. In contrast, BMA is an attractive fermentation product, as it is hydrophobic and forms a second liquid-phase in aqueous systems. However, this property, which makes BMA attractive for separation from the fermentation broth, also make it highly toxic to cells. Hydrophobic molecules can cause a multitude of problems for living cells, especially membrane disruption (Rajagopal, 1996). The Mitsubishi Chemical Group have engineered microorganisms such as *E. coli* and *P. putida* to produce BMA and other MMA precursors, but they are currently limited by the toxicity bottleneck described or refinement processes. However, an organism isolated from an environment with intrinsic tolerance to hydrophobic solvents may solve this biotechnological issue.

3.5 Targeting the Industrial Microbiome for BMA-tolerant *Pseudomonas* species

Pseudomonas species are a highly versatile and adaptive genus. Medical isolates of *P. aeruginosa* are known for their multi-drug resistance phenotypes (Aloush et al., 2006). *P. aeruginosa* has also been shown to grow in jet fuel (Striebich et al., 2014), and *Pseudomonas* species are often the contaminating organisms of industrially produced organic molecules (*The Future of Petrochemicals – Analysis - IEA*, n.d.). *P. aeruginosa* has been isolated from home cleaning products, laundry detergents,

cosmetics, metal working fluids and timber care products (Weiser et al., 2019a). Industrial isolates of *P. aeruginosa* have been shown to have larger genomes than their clinical and environmental counterparts this could be provide adaptability to toxic environments (Weiser et al., 2019b). Furthermore, *P. putida* strains have demonstrated tolerance of the plastic precursor Styrene and tolerance of up to 90% v/v toluene (Hosseini et al., 2017b; Verhoef et al., 2009b).

Sampling environments with extreme physiochemical conditions to isolate organisms with desired molecular machinery has been successful for the isolation of the thermophile *Thermus aquaticus*, which held a DNA polymerase suited to high temperatures, allowing for the development of PCR (Chien et al., 1976a). Similarly, an organism containing a Polyethylene teraphtalate-ase (PETase) enzyme suited to the degradation of PET plastics was isolated from a PET recycling facility in Japan (Yoshida et al., 2021).

Knowing that *Pseudomonas* species are ubiquitous in the environment and have the ability to tolerate extreme industrial contamination, we might expect to be able to isolate organisms with the molecular machinery conferring tolerance to butyl methacrylate in industrial environments enriched with BMA and its related compounds. It has been previously demonstrated that the primary mechanism for conferring tolerance to BMA is through RND efflux pumps (Personal Communication, Tucker Group).

3.6 RND Efflux Pumps

A major mechanism that *Pseudomonas* species and other microorganism use to tolerate environments which are saturated in chemical contaminants such as antimicrobials and solvents are resistance-nodulation-division (RND) efflux pumps. These are transmembrane pores actively extract toxic compounds and metabolites into the external environment (Colclough et al., 2020). This type of tolerance is advantageous in the biological synthesis of toxic compounds as the tolerance mechanism removing it from the cell also provides the compound into the culture for downstream processing.

3.6.1 Overall Structure of RND Efflux Pumps

The RND efflux pump typically is made up of three components each made of subunits of one protein. The system components are:

- Inner membrane transporter unit (Figure 5 – **BLUE - MexB**): This component is nested in the inner membrane and is responsible for recognising and binding substrates to export and uses a proton gradient from the periplasm to drive the active export into the upper pore components.
- Periplasmic adaptor unit (Figure 5 – **RED - MexA**): This is the bridging unit of the inner and outer membrane units.
- Outermembrane unit (Figure 5 – **YELLOW - OrpM**): The exit channel nested into the Outermembrane.

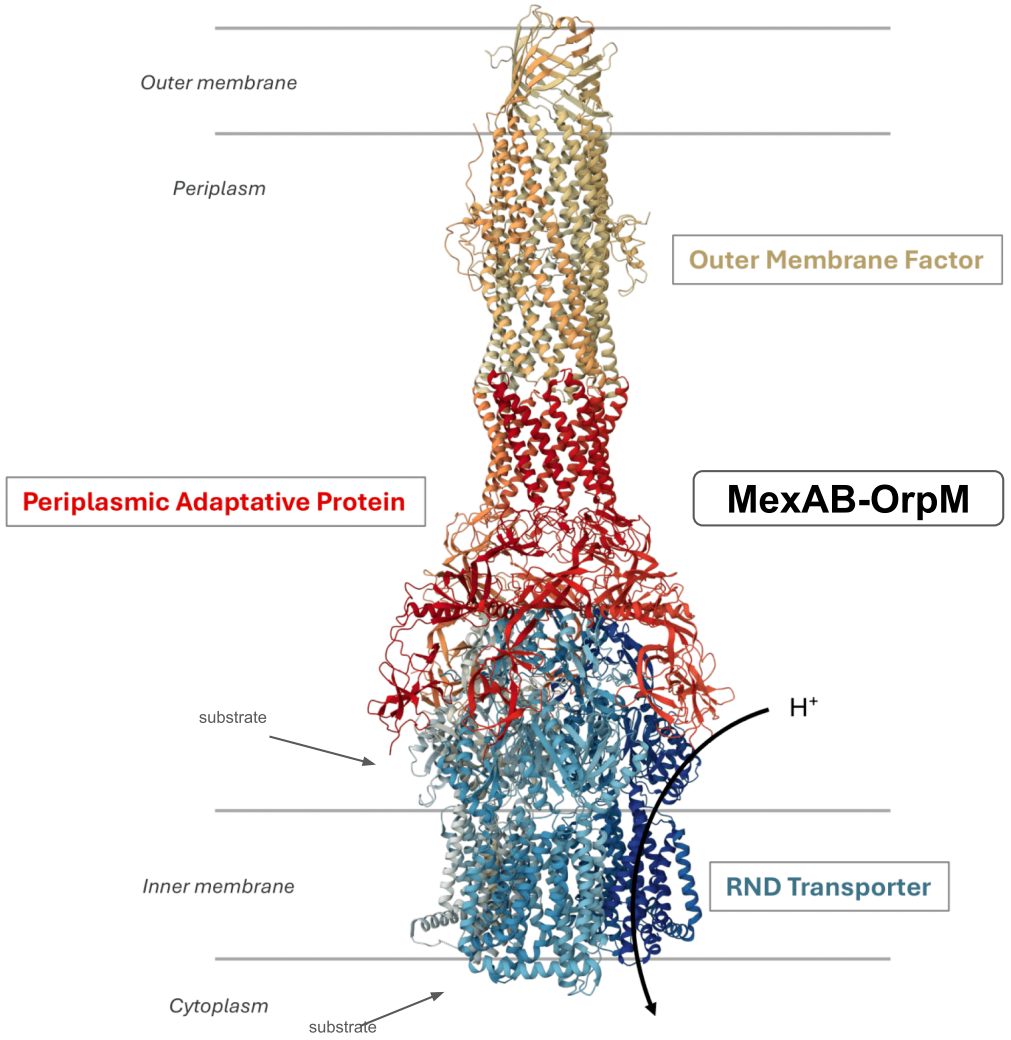


Figure 5 – The structure of the MexAB-OrpM RND efflux pump. The RND transport unit, the active component powered by a proton gradient in the periplasm versus the cytoplasm (**BLUE**), periplasmic adaptative protein unit (**RED**) and outer membrane unit (**Yellow**). The substrate can be taken from the cytoplasm and the periplasm (marked with arrows) *(Novelli & Bolla, 2024)*

3.6.2 Regulation of RND Efflux Pumps

RND efflux pumps are regulated by locally via a regulator local to the RND efflux operon and/or by a global regulator. The local regulator is typically a repressor which binds to the promotor region of the RND efflux operon. It will undergo a conformational change when an undesirable substrate such as an antibiotic is present in the cytoplasm or periplasm and binds to the repressor protein which disengages from the RND promotor allowing for transcription of the RND genes and the eventual translation of the RND pump and export of the substrate. When the levels of the compound are low enough not to bind the repressor protein its conformation will return to its RND promotor binding configuration and block transcription of the pump. In the case of MexAB-OrpM its transcription is regulated by the local repressor MexR (Srikumar et al., 2000).

Local regulators of other RND efflux pumps have also been shown to block the transcription when two pumps may have a cross over of substrates. The system better suited to the efflux of the contaminating compound. The SrpR regulator binds the *ttgABC* promotor with a higher affinity than its own local regulator TtgR (Yao et al., 2021).

Global regulators regulate a variety of genes involved in antimicrobial resistance and compound tolerance the CpxR is such a regulator it has

been shown to influence cell membrane repair and cell survival. Increase resistance to beta-lactamase and in deficient MexR strains of *P. aeruginosa* the over expression of *cpxR* has been shown to also increase expression of *mexAB-OrpM* (Tian et al., 2016).

In the context of biotechnology and the synthesis of platform chemicals to increase tolerance the regulatory system can be taken advantage of to increase tolerance. Creating strains with defective/deleted local promotors that are unable block transcription of the RND pump can be used to increase tolerance and export the product into the environment. Increasing transcription of the global regulators such as *cpxR* could also provide protection for the cell by maintaining the membrane and increasing production of RND efflux pumps and other systems within the network that contribute to tolerance.

3.6.3 RND Efflux and Antimicrobial Resistance and Solvent Tolerance

Due to their broad activity of various compound classes the RND efflux system contribute to antimicrobial resistance and in particular multidrug resistance (MDR). *P. aeruginosa* has been known to have twelve or more different RND efflux pumps four of which have been noted to contribute to MDR (Martinus Pos et al., 2015). This repertoire of RND efflux pumps makes *P. aeruginosa* particularly hard to treat as it cannot simply be inhibited like a beta-lactamase. Pumps are known to have cross over of substrates and

can use an alternative in the case of one being defective which makes them a difficult drug development target (Duque et al., 2001; Yao et al., 2021).

In a medical setting RND efflux pump can provide an obstacle for treating MDR infections. Regarding industrial biotechnology they may provide a useful tool due to their broad chemical compatibility as not only are they effective at effluxing antimicrobials, but they also are capable of transporting industrial solvents and platform chemicals (Duque et al., 2001; Weber et al., 1993).

3.7 Efflux is a Major Component in Tolerance and Ideal for Biotechnology Applications

The RND efflux pump was shown to confer BMA tolerance in the laboratory reference strains *P. putida* KT2440, which uses its TtgABC efflux pump, and *P. aeruginosa* PA01, which uses its MexAB-orpM efflux pump. These versatile efflux pumps provide significantly more tolerance to BMA than other efflux pumps and tolerance mechanisms characterised in *E. coli* (Personal Communication—Tucker Group).

The efflux pumps that provide the reference strains *P. putida* KT2440 and *P. Aeruginosa* PA01 have not evolved in an environment that would select adaptations suited to hydrophobic solvents such as BMA. However, the industrial environment may contain organisms with RND efflux pumps and

such adaptations. The strain *P. putida* DOT-T1E has been shown to increase transcription levels of the ttgDEF efflux pump and decrease the levels of ttgABC when challenged with the hydrophobic compound toluene, demonstrating that the efflux pump TtgABC held by *P. putida* KT2440 (current strain in development) may not be best suited for all hydrophobic compounds when an alternative is available (Ramos et al., 1998). An alternative efflux pump which has evolved within an industrial microbiome may be advantageous in terms of conferring tolerance to an organism producing BMA through fermentation. These adaptations could increase tolerance by increasing the amount of BMA expelled from the cell in the environment; thus, an increase in tolerance would also increase titres accessibility for downstream processing.

3.8 Objectives of the Study

The main objectives of this study:

To address the need to produce the precursors for acrylic plastic via biotechnology rather than petrochemical routes. The toxicity bottleneck must be resolved. The following objectives were achieved:

- Isolate organisms with increased tolerance to solvents, particularly BMA, compared to *P. putida* KT2440 (Chapter 5).
- Identify non-pathogenic organisms with potential as production hosts for a BMA fermentation process (Chapter 5.3).
- Identify and characterise the systems these environmental isolates use to confer their solvent tolerance (i.e., RND efflux pumps) (Chapter 6).
- Transfer systems identified into *P. putida* KT2440 and demonstrate enhanced tolerance to butyl methacrylate as a result (Chapter 6.5.1).
- Use the features of efflux pumps known to confer solvent tolerance to identify other pumps which may also be evolved to solvent tolerance from online sequence data (Chapter 6.6)
- Address the issue of intracellular tolerance versus tolerance conferred through export. Using machine learning identified resistance genes (Chapter 7).

4 Materials and Methods

4.1 Media Composition

Medium	Composition	Quantity	Instructions
R-2A	Yeast extract	0.5g/L	Dissolve in water, autoclave
	Proteose peptone	0.5g/L	
	Casein hydrolysate	0.5g/L	
	Glucose	0.5g/L	
	Soluble starch	0.5g/L	
	Sodium pyruvate	0.3g/L	
	Dipotassium hydrogenphospat e	0.3g/L	Add 15g/L Agar for solid media
	Magnesium sulphate	0.024g/L	
LB	Tryptone	10g/L	Dissolve in water, autoclave
	Yeast extract	5g/L	
	NaCl	5g/L	Add 15g/L Agar for solid media
M9 Minimal Media	Sterile H2O	750ml	
	5x M9 Salts	200ml	Add to water (If bought) To make: To 400mL of ddH2O add: a. 64g (Na2HPO4•7H2O) b. 15g (KH2PO4) c. 2.5g (NaCl) d. 5.0g (NH4Cl) Make to 1 L with ddH20
	1 M MgSO4	2ml	Add to the solution
	20% Glucose	20ml	Add to the solution
Charcoal Transport Medium	Sodium Chloride	3g/L	Dissolve in water, Autoclave. Add to 10mls to 15ml falcon tubes and insert Swab once sample collected
	Potassium Chloride	0.3g/L	
	Calcium Chloride	0.1g/L	
	Magnesium Chloride	0.1g/L	
	Monopotassium Phosphate	0.2g/L	
	Disodium Phosphate	1.15g/L	
	Sodium Thioglycolate	1g/L	
	Charcoal	10g/L	
	Agar	4g/L	

4.2 Cultivation of Isolates

Isolates were cultivated by inoculating 10mL of LB in a universal tube and incubating for 12 hours at 30°C 210rpm.

All cultures containing BMA were processed using falcon tubes or glass conical flasks as BMA degrades other types of plastics.

4.3 Cryostock Preparation

5mL cultures of Isolated organisms were pelleted by centrifugation. Supernatant was removed from samples and pellets were resuspended using 500µl of 60% (v/v) glycerol and transferred to a sterile 1.5mL centrifuge tube. Samples were stored at -80°C.

4.4 Nucleic Acid Isolation

4.4.1 PowerSoil Bead Basher Extraction

Nucleic acid was isolated from environmental samples using the DNeasy PowerSoil Pro Kit (Cat #47014 Qiagen), following the Quick-Start Protocol (*DNeasy PowerSoil Pro Kits*, n.d.).

4.4.2 High Molecular Weight DNA Isolation

High molecular weight DNA was isolated from samples for use in long-read sequencing using the Promega Wizard® HMW DNA Extraction Kit (Cat #A2920 Promega). The Wizard® HMW DNA Extraction Kit Technical Manual was used (Corporation, n.d.).

4.5 Isolation of BMA Tolerant Organisms

Environmental samples from varied sources and volumes (swabs, collected liquid, soil) were cultivated using R-2A media. 50mL cultures were incubated for 48 hours at 30°C 210rpm. Cultures were treated with 25% (v/v) BMA from inoculation.

After cultivation, 20µl of 10^{-1}, $10^{-2,}$ 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} dilutions were generated. Dilutions were prepared using 1.5mL centrifuge tubes and Ringer's solution as diluent. Dilutions were dropped in 20µl volumes onto a square LB agar plate in the series described and incubated in a static incubator for 24 hours at 30°C.

After incubation, colonies present in the highest dilution were deemed the most tolerant as that culture had the most abundant cells treated with BMA. Colonies from these samples were re-streaked on fresh LB agar plates and subjected to further incubation in a static incubator for 24 hours at 30°C.

4.6 Butyl Methacrylate Tolerance Assay

Isolates were prepared using the methods described in section 4.2. The optical density of cultures was measured at the wavelength 600nm. The OD of isolates was standardised to 0.05OD by dilution using LB to generate a final volume of 15 mL. 5 mL of BMA was added to each sample to replicate an ideal bio-fermentation environment . All assays were conducted in triplicate to conduct

statistical analysis for each experimental group. Samples were cultivated overnight as per the described method in 4.2.

After cultivation, 20µl of 10^{-1}, $10^{-2,}$ 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6} dilutions were generated. Dilutions were prepared using 1.5mL centrifuge tubes and Ringer's solution as diluent. Dilutions were dropped onto a square LB agar plate in the series described and incubated in a static incubator for 24 hours at 30°C.

After incubation, the number of colonies present in the most dilute sample were counted and these values were used to determine the number of colony-forming units (CFU) within each original sample using Equation 1.

$$\mathrm{CFU/mL} = \mathrm{VN} \times \mathrm{D}$$

Equation 1 - **N** = Number of colonies counted on the agar plate, **D** = Dilution factor, **V=** Volume plated

4.7 16S rRNA Gene Colony PCR and Sanger Sequencing

4.7.1 DNA Isolation

Genomic DNA was isolated from colonies by adding them to 25 µL of filter-sterilised H_2O in a 1.5mL centrifuge tube using a sterile pipette tip. The colonies were boiled at 100°C using a heat block for 15 minutes before pelleting by centrifugation.

4.7.2 PCR Amplification of 16S rRNA Gene

Segments of the 16S rRNA gene were amplified by PCR using the following primers:

- Forward Primer: 27F (5'-AGA GTT TGA TCC TGG CTC AG-3')
- Reverse Primer 1: 1492R - (5´- TAC GGC TAC CTT GTT ACG ACT T - 3´)
- Reverse Primer 2: 1522R (5´- AAG GAG GTG ATC CAG CCG CA - 3´)
- Reverse Primer 3: 1525R (5´- AAG GAG GTG ATC CAG CC - 3´)

  (Frank et al., 2008)

4.7.3 PCR Reaction Setup Using Q5 High-Fidelity Polymerase – M0491S

Each 50µL PCR reaction contains:

- 10µL of Q5 Reaction Buffer (B9028AVIAL - New England Biolabs)
- 2.5µL of Forward Primer (10µM working concentration)
- 2.5µL of Reverse Primer (10µM working concentration)
- 1µL of dNTPs 10mM (N0447S - New England Biolabs)
- 0.5µL of Q5 Polymerase (M0491SVIAL - New England Biolabs)
- 5µL of the Colony Lysis Material (Template DNA)
- 28.5µL of Nuclease Free Water

4.7.4 Thermocycler Conditions – API Applied Biosystems PCR System 2700

1. Initial Denaturation Step at 98°C for 30 seconds
2. 35 Cycles of:
    - Denature 98°C - 10 Seconds
    - Anneal 65°C - 20 Seconds

- Extension 72°C – 45 Seconds

3. Final Extension 72°C for 2 Minutes
4. Hold 12°C ∞

4.7.5 Agarose Gel Electrophoresis

After PCR, samples were analysed using agarose gel electrophoresis.

5µL of each PCR reaction was loaded into a well within a 1% (w/v) agarose 1 X TAE gel with added 0.1% (w/v) ethidium bromide. Gels were placed in 1 XT AE buffer and resolved at 85V for 60 minutes.

4.7.6 Purification and Sequencing

PCR Reactions with a single band were purified using a PCR cleanup kit, and reactions with more than one band were extracted and cleaned up using a Gel purification kit (Wizard® SV Gel and PCR Clean-Up System – #A9281). Fragments were sequenced using the Eurofins Sanger Sequencing Service. Two Tubes were prepared per PCR reaction, each containing the 16S forward and reverse primer. The returned sequence were analysed using the NCBI BLAST tool to approximate the genus of isolates.

4.8 Metagenomic Analysis of Bacterial Solvent Enriched Cultures

Environmental Samples if swabs were added to ringers' solution and vortexed for 30 seconds then the ringer's solution was added. Liquid or soil samples

were homogenised to allow for an equal distribution of populations between experimental groups.

Samples were cultivated using R-2A media with various solvents such as BMA, Styrene and ethyl benzene for 48 hours, and DNA was isolated using the bead bashing technique described in Section 4.4.1. DNA Libraries were prepared using the Illumina workflow described in Section 4.10. DNA Libraries were sequenced using the run analysis as described in Section 4.15.5.

4.9 Oxford Nanopore Sequencing Kits and Protocols

4.9.1 Equipment and Reagents:

All MinION Sequencing runs were completed using a MinION Mk 1C(MIN-101C) following the manufactures protocol for the below kits:

DNA was prepared using a Ligation Sequencing Kit (SQK-LSK109)

DNA was barcoded using Native Barcoding Expansion 1-12 (EXP-NBD104)

Flow Cells primed with Flow Cell Priming Kit (EXP-FLP002)

Flow Cell was SPOT ON FLOW CELL MK 1 R9 Version (FLO-MIN106D)

4.10 Illumina Sequencing Kits and Protocols

Illumina Sequencing for metagenomics and WGS was performed using an Illumina MiniSeq. Using the MiniSeq High Output Reagent Kit 150-cycles (FC-420-1002) following the MiniSeq System Guide(Illumina, 2020) .

4.11 Preparation and Transformation of Electrocompetent *Pseudomonas* Cells

4.11.1 Media and Buffers

- Sucrose 300mM (102.7g/L Sucrose Autoclaved)
- LB
- SOC (15544034 Thermofisher Scientific)

4.11.2 Preparation

- 20 mL overnight cultures of the desired *Pseudomonas* strains were prepared in LB.
- After cultivation, centrifuge (Bench Top Centrifuge – Thermofisher Scientific) at 4700 rpm for 10 minutes.
- The supernatant was discarded, and the pellet was resuspended in 10 mL sucrose 300 (mM).
- Samples were centrifuged at 4700 rpm for 10 minutes.
- The supernatant was discarded, and the pellet was resuspended in 1 mL sucrose 300mM.
- Transferred to a 1.5mL centrifuge tube.
- Centrifuged for 2 minutes at maximum speed.
- The supernatant was discarded, and the pellet was resuspended in 500 µL sucrose 300 (mM).
- Samples were distributed into 100 µL 1.5 mL tubes

4.11.3 Transformation

- 1 μL of Plasmid at 100ng/μl was mixed with 100μl aliquots of electrocompetent cells then transferred into prechilled electro-cuvette (Bio-Rad Gene Pulser/MicroPulser Cuvette 0.2 mm #1652082)
- Samples were pulsed using programme Ec2 (Bio-Rad MicroPulser #165-2100) and 900μl SOC was added directly to the cuvette.
- The sample was collected with a syringe and large gauge needle and transferred into a 1.5 mL tube before incubation at 30°C at 650 rpm for 2 hours (Eppendorf Thermomixer R #5355).
- Sampled were plated by spreading 200 μL and 20μL on agar plates with the desired antibiotic.

4.12 Plasmid DNA purification

Purification of plasmids was carried out using a Promega Mini-Prep Kit (Wizard® Plus SV Minipreps DNA – #A1330) using the Quick Protocol (Corporation, 1999).

4.13 Plasmid Construction

4.13.1 PCR amplification of Efflux Operon and pUC186LT-mini-Tn7t-Km

PCR Primers with Efflux Insert and Vector Backbone Overhangs were used:

pUC18R6LT-Mini Back Amplification Forward:

(5’-GACTTGTTCAAACCCCTGGAAGCTTGGGCCCGGTACCT-3’)

pUC18R6LT-Mini back Amplification Reverse

(5’-CTTCCTGAAAAGCCAAGGCGGGGGATCCACTAGTGAGCT-3’)

RCS09A Efflux Operon Forward:

(5’- AGGTACCGGGCCCAAGCTTCCAGGGGTTTGAACAAGTC-3’)

RCS09A Efflux Operon Reverse:

(5’-AGCTCACTAGTGGATCCCCCGCCTTGGCTTTTCAGGAAG-3’)

(Choi & Schweizer, 2006a)

PCR Reaction:

Each 50µl PCR reaction contained:

- 10µL of Q5 Reaction Buffer (B9028AVIAL - New England Biolabs)
- 2.5µL of Forward Primer (10µM)
- 2.5µL of Reverse Primer (10µM)
- 1µL of dNTPs (N0447S - New England Biolabs)
- 0.5µL of Q5 Polymerase (M0491SVIAL - New England Biolabs)
- 5µL of the Colony Lysis Material (Template DNA)
- 28.5µL of Nuclease Free Water

Thermocycler Conditions for The Two Separate Amplifications:

1. Initial Denaturation Step at 98°C for 30 seconds
2. 35 Cycles of:
    - Denature 98°C - 10 Seconds
    - Anneal 67°C - Efflux Operon, 64°C - Vector - 20 Seconds

- Extension 72°C – Efflux Operon 4 minutes, Vector 3 minutes

3. Final Extension 72°C for 2 Minutes
4. Hold 12°C ∞

4.13.2 Gibson Assembly

Gibson assembly was carried out using the NEBuilder Protocol calculator tool

Efflux Fragment is 7236 bp at 60ng/µl

Vector is 4891bp at 60ng/µl

Reaction:

- 2.5 µl of Vector PCR product (0.05 pmoles)
- 7.4µl of Efflux Insert (0.100 pmoles)
- Deionised H20 0.1µl
- NEBuilder HiFi DNA Assembly Master Mix 10µl (E2621S)

50°C for 60 minutes in a Thermocycler

4.13.3 Pir1 Transformation

Pir1 Transformation was carried out by following the protocol:

- Defrost one Vial of One Shot™ PIR1 Chemically Competent *E. coli* (C101010) on Ice

- Add 2µl of Assembly Mix to Pir1 vial mix by tapping and add to ice for 30 minutes
- Heat shock at 42°C for 30 Seconds
- Incubate at 37°C degrees for one hour in either a shaking incubator with a rack or a Thermomixer on 650rpm
- Spread 20µl and 200µl on a plate with Kanamycin 50µg/mL

4.13.4 Colony PCR for Construct and mdtA Efflux Component

Select ten colonies from the transformation plate and add 25µl of Sterile H_2O.

Construct PCR Primers:

Forward (Vector) Primer: (5'- AACTGGGTGTAGCGTCGTAA-3')

Reverse (Insert) Primer: (5'- AACGCCAACCTCGACCAG-3")

PCR 50µl Reaction Mix:

Using OneTaq® Quick-Load® 2x Master Mix (M0486S)

- 25µl OneTaq Master Mix
- 1µl Forward Primer
- 1µl Reverse Primer
- 5µl colony lysis
- 18µl H_2O

Thermocycler Conditions for The Two Separate Amplifications:

5. Initial Denaturation Step at 94°C for 30 seconds
6. 30 Cycles of:
    - Denature 94°C - 10 Seconds
    - Anneal 55°C – mdtA, 56°C - Construct - 20 Seconds
    - Extension 68°C – 20 seconds (both)
7. Final Extension 68°C for 5 Minutes
8. Hold 12°C ∞

4.13.5 *P. putida* KT2440 Transformation

Using the protocol in section 4.11, prepare KT2440 for transformation, and using overnights of successfully transformed Pir1, prepare the plasmid material section according to section 4.12. Continue section 4.11 using the Pir1 prepared plasmid as the transformant.

4.14 Statistical Analysis

Statistical Analysis was performed either using the Python package bioinfokit version 2.13 using a 2-sample T-Test for two isolates. A one-way ANOVA when comparing more than two isolates

4.15 Bioinformatic Tools

4.15.1 Genome Assembly

For long-read data sequenced on the ONT Mk1c, Flye Assembler (Version 2.94) was used(Kolmogorov et al., 2019).

Example Command:

flye --nano-corr input.fasta --our-dir input_flye_assm/

4.15.2 Mapping Reads for Depth and Genome Polishing

BWA MEM Aligner:

To align reads to genome assembly, BWA MEM (version 0.7.17) was used(H. Li & Durbin, 2009).

Example Command:

bwa mem reference.fasta reads.fastq > reference_reads.sam

SamTools Sort:

To sort the sam file to the bam file, samtools (version 1.18) was used(Danecek et al., 2021).

Example Command:

samtools sort reference_reads.sam > reference_reads.bam

samtools index reference_reads.bam

4.15.3 Genome Polishing

To polish long-read assemblies and correct long-read sequence data with Illumina reads sequenced on the Illumina Miniseq. Pilon (version 1.24) was used (Walker et al., 2014).

Example Command:

pilon --genome reference.fasta --frags reference_reads.bam --output reference_polished –outdir reference_pilon

4.15.4 Bacterial Identification Using autoMLST and pyANI

Preliminary identification using autoMLST webtool if the estimated ANI of related genomes is less than 95% pyANI (Version 0.2.7) was used to

confirm actual ANI falls below 95% (Pritchard et al., 2015). PyANI was used on a larger dataset of related organisms to further confirm identity.

4.15.5 Metagenomic Analysis

Metagenomic analysis was performed using Kraken2 (version 2.1.3) and the Maxikraken database (Wood et al., 2019), which requires 140 GB of RAM.

Example Command:

Kraken2 --db Maxikraken_1903_140gb/ --output sample.txt --report sample_report.txt sample_reads.fastq

Visualisations were created with Recentrifuge (Martí, 2019).

4.15.6 BLAST Screening

Local BLAST screening for tolerance conferring systems was done locally using Blastn (version 2.15.0) and Bandage (version 0.8.1). The assembly .fasta was loaded as the map, and the building BLAST function was used to load a custom BLAST database of known tolerance conferring systems. This allowed for a visual representation of the matches and easy extraction from the assembly to search using the NCBI BLAST database online (Camacho et al., 2009).

4.15.7 IPCA and K-Mers, DBSCAN Clustering

Incremental Principal Component Analysis and K-Mers, DBSCAN clustering were performed using Sci Kit Learn see manual for usage (Pedregosa et al., 2011).

4.15.8 AutoDock Blind Docking

AutoDock was used to predict compound binding to proteins of interest. Set to Blind docking by encapsulating the entire protein in the Gridbox (Morris et al., 2009).

4.15.9 Prediction of Protein Structure using Alphafold2

Alphafold2 was used to set the default parameters and determine the structure of proteins from the amino acid sequence (Jumper et al., 2021).

4.15.10 Fpocket Protein Pocket Analysis

The analysis of protein pockets was performed using Fpocket set to the default settings (Le Guilloux et al., 2009a).

5 Results 1 - Sampling, Isolation and Identification and Tolerance Profiling of Organisms:

5.1 Introduction

The unique biochemical properties found in extreme environments have provided the basis for transformational innovations in biotechnology. Exploration of these microbiomes has led to the discovery of tools essential for modern-day biological sciences and sustainability goals. For example, the isolation of the bacterium *Thermus aquaticus* from the extreme hot spring environment allowed for the development of PCR with its thermostable Taq-polymerase (Chien et al., 1976b). In addition, a PET plastic-rich environment allowed for the isolation of *Ideonella sakaiensis*, producing PETase, which has been shown to degrade PET and could develop bioremediation processes to combat the world's plastic crisis (Yoshida et al., 2016). Both examples demonstrate that targeting environments with unique biochemical properties can yield molecular tools with desired properties.

This study focused on tolerance to BMA, a hydrophobic methacrylate ester, and precursor to acrylic-based products, including durable plastic materials and surface coatings. The hydrophobic nature of BMA makes it an ideal candidate for bioreactor production due to the product's phase

separation from the culture, making the extraction process more straightforward. Current strains in development for the production of BMA are limited by toxicity (Mitsubishi Chemical, Personal Communication). Targeting a solvent/methacrylate-rich environment could provide the molecular mechanisms and tools to relieve this toxicity bottleneck and alternative production strains.

In sampling the environment for microorganisms with intrinsically desirable traits, areas that contain high amounts of human or agricultural contamination should be avoided due to the high antimicrobial resistance gene and pathogen content (Begmatov et al., 2024). These organisms may provide systems that confer tolerance but are unsuitable as production strains.

In addition to searching for systems that confer tolerance to BMA, other hydrophobic toxic industrially relevant compounds used in the production of plastics will also be investigated. Ethyl benzene is used in the production of Styrene (Zhu et al., 2021). Styrene which has been shown to cause membrane and DNA damage but does not influence expression efflux transporters in *E. coli* at the transcriptional level (Machas et al., 2021).

Sampling sites for the isolation of solvent-tolerant organisms were chosen based on environmental enrichment of the area by various toxic compounds, specifically methacrylate esters. It was hypothesised that

microorganisms within these environments could have evolved tolerance mechanisms to methacrylate esters, which could be beneficial in overcoming the toxicity bottleneck previously described.

The Mitsubishi Chemical Group BMA production facility selected as a sampling site for this study was initially unavailable for sampling due to travel restrictions associated with the COVID-19 pandemic. Because of this, environmental samples were obtained from interim alternative locations such as sites with heavy metal contamination, sites previously used for industrial activity, and environmental sites near industrial areas. As mechanisms that confer heavy metal, solvent tolerance can be broadly acting across chemical classes. These samples were tested for growth with 50% (v/v) and 25% (v/v) BMA, but no growth was visually observed.

Like antimicrobial resistance, solvent tolerance can typically be facilitated by broadly acting molecular mechanisms. Mechanisms such as efflux pumps which are transmembrane channel proteins that remove compounds from the cell into the environment. These fall into six categories: ATP-Binding Cassette (ABC) Transporters, Major Facilitator Superfamily (MFS), Small Multidrug Resistance (SMR) Family, Multidrug and Toxic Compound Extrusion (MATE) Family, Proteobacterial Antimicrobial Compound Efflux (PACE) and Resistance-Nodulation-Division (RND) Family (Blair et al., 2014). The RND efflux pump found in Gram-negative bacteria removes compounds from various chemical

families. A proton gradient facilitates effluxing from the cell cytosol and periplasm into the environment. This broad-ranging activity in pathogenic organisms contributes to multidrug antimicrobial resistance and is involved in hydrophobic solvent tolerance (Ramos et al., 1998). The MexAB-OrpM system has been shown to transport 54 substrates (Carrara et al., 2022a). These systems are typically regulated by a repressor, freed from the efflux repressor/regulator promotor via conformational changes to the repressor protein when it binds to a substrate (Yao et al., 2021). Tolerance conferred via efflux is ideal for biotechnological applications, as synthesised toxic products are removed from the cell into the environment where they are accessible for downstream processing. It was hypothesised that a location in which a wide variety of solvents, plastics, heavy metals, and adhesive compounds are used could harbour a highly tolerant organism conferring their tolerance through broadly acting RND efflux pumps. An art studio was the obvious choice, as it not only contains desired contaminants and is not maintained in a sterile fashion with low levels of human contamination. The Royal Conservatories of Glasgow, a theatre and performing arts institute set design department, was chosen as a sampling site.

5.2 Royal Conservatoire Samples

The Royal Conservatories of Glasgow's set design department was sampled. Bulky liquid samples from the solvent waste and drain u-bend

were collected in Falcon tubes. Swabs were taken from sites that appeared to be chemically contaminated and preserved in a charcoal transport medium. In total, 10 sites were sampled, as detailed in Table 3. Bulky samples were aliquoted into 1.5mL centrifuge tubes and stored at -80°C. The swabs were stored at 4°C until they could be processed in the lab. Each sample was aseptically transferred into R-2A media with BMA 25% (v/v) for two days at 30°C. Sample Royal Conservatoire Site 09 (RCS09), I.e., the sink U-bend sample, was the only sample observed visually to support growth at the end of the experiment (Table 3). Therefore, population analysis of sample RCS09 treated with Styrene, Ethyl Benzene and BMA was carried out using metagenomics.

Table 3 – Royal Conservatoire of Glasgow Set Design Department Sample and R-2A Growth – Sampling information including sampling location and sample type. Growth is a record of visually observed culture growth where 'clear' indicates no microbial growth and 'visually confirmed' indicates presumed microbial growth, based on the culture becoming 'cloudy' following incubation for 24 hours at 30°C treated with BMA 25% (v/v). The visual conformation approach was used as this allowed for the processing of a large number of samples and the shorter culture time to allow for organism well suited for a quicker culture time under BMA treatment.

Sample ID - Royal Conservatoire	Type	Growth in R-2A BMA 48 hours
RCS01 - Work area	Swab - Charcoal Transport Media	Clear
RCS02 - Paint brush cleaning	Swab - Charcoal Transport Media	Clear
RCS03 - Work Area	Swab - Charcoal Transport Media	Clear
RCS04 - Work Area	Swab - Charcoal Transport Media	Clear
RCS05 - Wastewater	Swab - Charcoal Transport Media	Clear
RCS06 - Solvent waste	Swab - Charcoal Transport Media	Clear
RCS07 - Tap	Swab - Charcoal Transport Media	Clear
RCS08 - Sink	Swab - Charcoal Transport Media	Clear
RCS09 - Drain u bend	50ml Collection in Falcon Tube	Visually confirmed
RCS10 - Solvent waste	50ml Collection in Falcon Tube	Clear

5.2.1 Culture-Based Solvent Metagenomics was conducted on Sample RCS09

To understand how the RCS09 sample population is altered when treated with solvents R-2A medium was inoculated with sample RCS09 and incubated in the presence of styrene (5mM, 10mM), ethylbenzene (5mM, 10mM), and BMA (25% v/v: and 50% v/v) for 48 hours at 30°C 220rpm. Due to the limitation of the Illumina MiniSeq sequencing technology, it was decided that instead of biological replicates, it would be more insightful to study a variety of industrially relevant solvents. R-2A Media was chosen as it has previously been shown to facilitate the growth of organisms typically outcompeted in rich culture media by organisms well suited to these conditions (Reasoner & Geldreich, 1985a). Therefore, the enriched culture better represents the original sample population while promoting organisms suited to growing at standard culture conditions applicable to ideal biotechnological constraints. DNA was extracted using a DNeasy PowerSoil bead-bashing kit (Qiagen). Illumina Libraries were prepared using the high-output kit (Illumina). DNA Libraries were analysed using an Agilent bioanalyzer and were within 'normal' parameters for sequencing (i.e., Library size ranging from 200-1000bp – average length 303bp – Concentration 1029 pg/μL 5615pmol/L). Then, libraries were sequenced using the Illumina Miniseq.

Metagenomic reads were allocated to reference genomes within the Maxikraken 140gb database using Kraken2. This approach has limitations when analysing environmental samples as it uses a set database to align metagenomics reads against. If organisms within the population are outside this database, they will be assigned to the closest match within the database, which is mainly comprised of well-studied microorganisms, and so can lead to a bias (Smith et al., 2022). This limitation is highlighted by the confidence scores assigned to each classification. The confidence score indicates the alignment quality, the number of K-Mer matches from each query sequence and the confidence of the classification.

Microbial population percentages were generated from these data, showing the population distribution for sample RCS09 treated with each solvent concentration (Figures 7 and 8). A dynamic interactive data representation was generated using Recentrifuge. This allowed for easy visualisation of shifts in subpopulations and display of confidence scores for each taxonomic group assignment.

5.2.2 RCS09 Treated with Butyl Methacrylate Showed the Majority of Reads assigned to *Pseudomonas* species

RCS09 sample treated with BMA at all the tested concentrations tested showed the largest shifts in population, with *Pseudomonads* occupying 95% of the reads at BMA 25% (v/v) and 93% at BMA 50% (v/v) compared to

19% for the no solvent control. *Pseudomonas putida* (confidence score 90.1 – BMA 25% (v/v) and 102.9 – BMA 50% (v/v)) was observed to be the dominant species within the sample, shifting from just 4% of the control sample to 28% for BMA 25% v/v and 52% of the reads for BMA 50% v/ (Figure 7). The remaining population represented in each BMA-treated sample contained *Pseudomonas fluorescens*, *Pseudomonas syringae* and other unclassified *Pseudomonas* species. RCS09 treated with BMA at 25% v/v and 50% v/v concentrations had minimal reads allocated to the pathogenic species *Enterococcus faecium*: 25% v/v BMA treatment showed 3% population representation with a confidence score of 39.5 (Figure 6). 50% v/v BMA treatment showed a 3% population representation of *E. coli* (confidence score of 38.9). These low confidence scores for both pathogenic classifications suggest that they are not accurate.

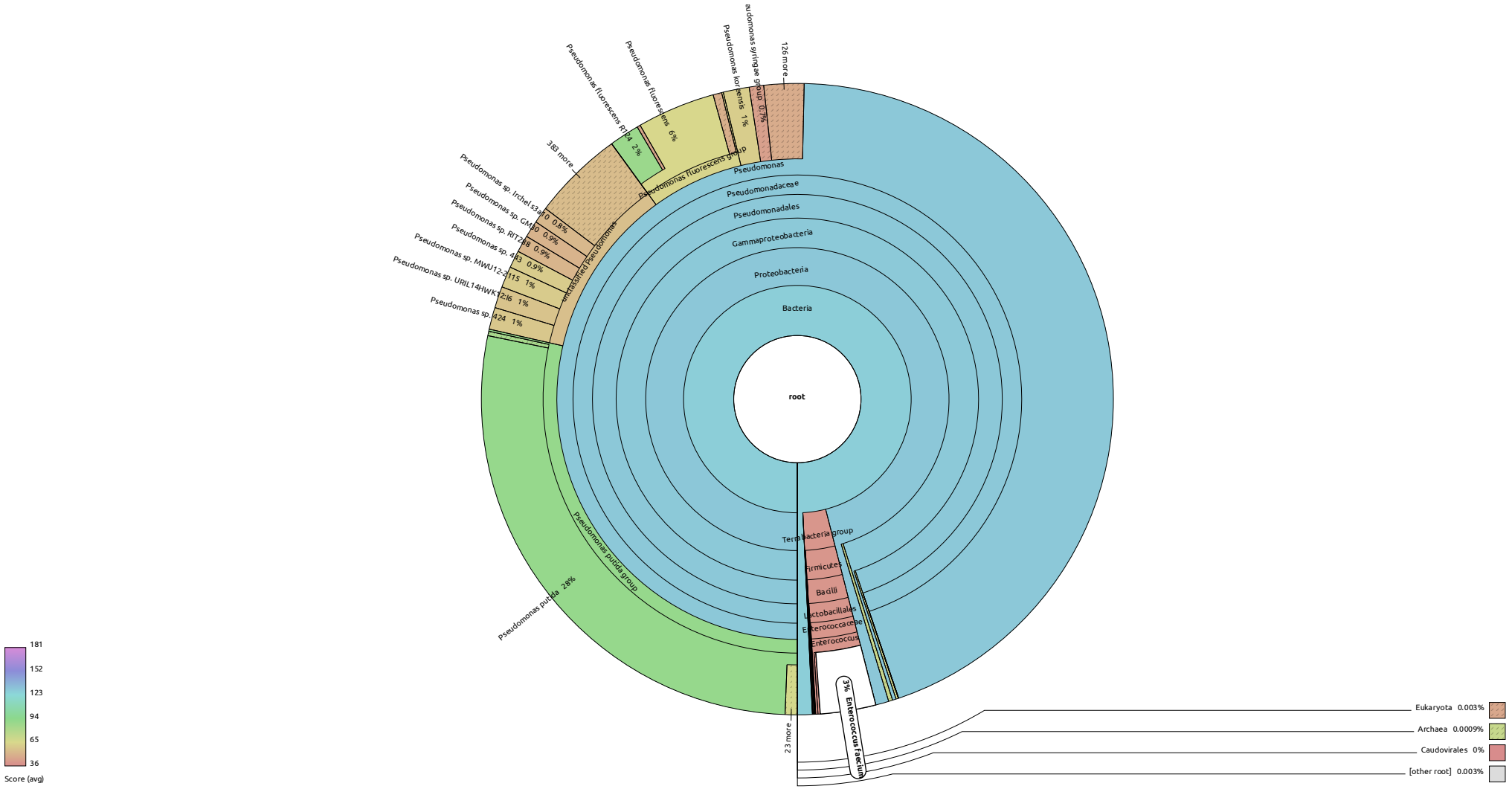


Figure 6 BMA 25% v/v (1.56 M) Treatment of RCS09 Metagenomic read allocations visualised using recentrifuge. *Pseudomonas putida* allocated 28% (C-Score 90.1 Green) of Reads at the species level. At the Genus level, *Pseudomonas* is allocated 95% (C-Score 128.8 Blue) of the reads. Enterococcus faecium was allocated 3%, with a low confidence score of 39.8 (Red), which is not likely to be an accurate allocation.

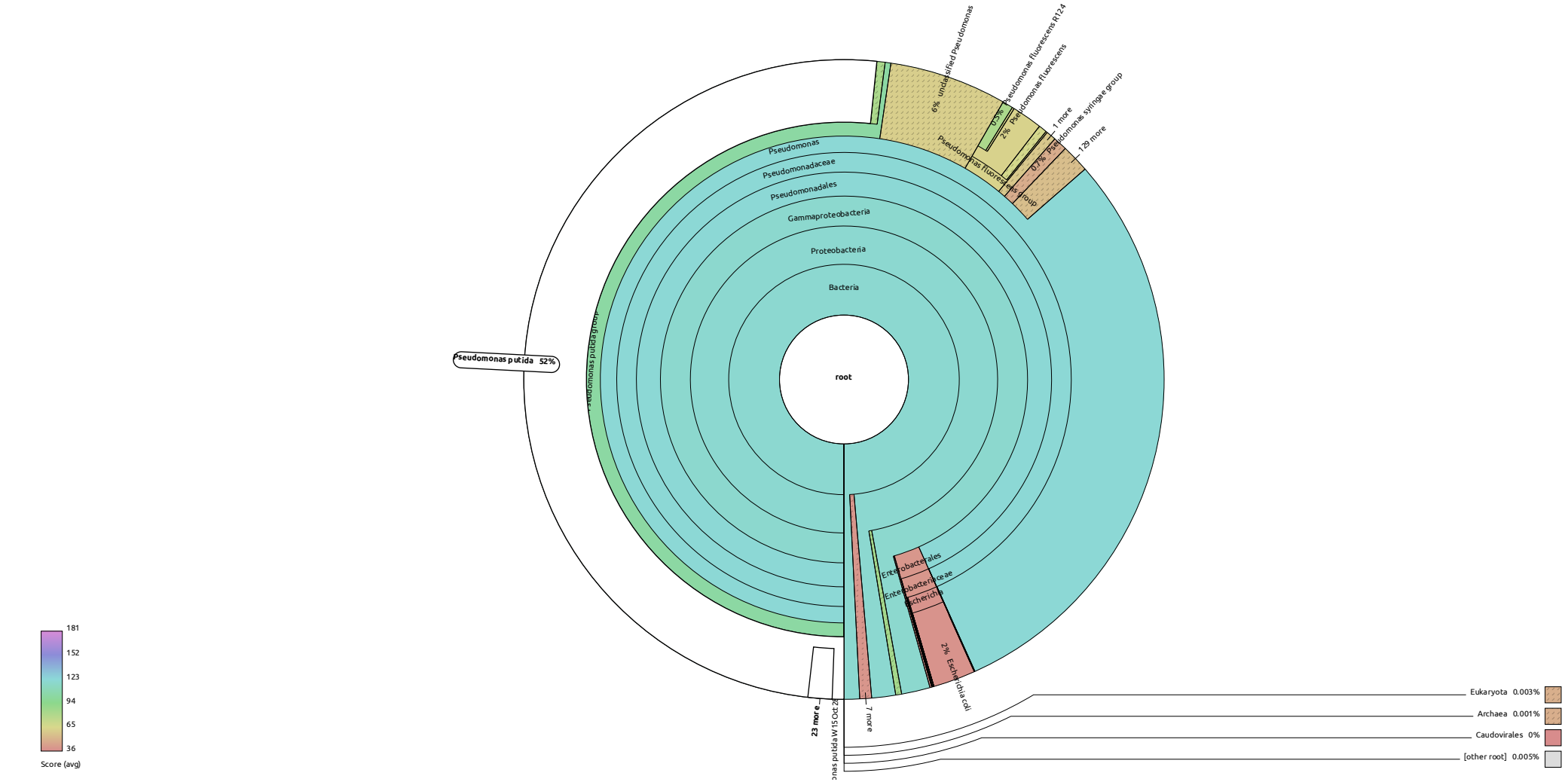


Figure 7 BMA 50% v/v (3.12 M) Concentration Treatment of RCS09 Metagenomic read allocations visualised using recentrifuge. *Pseudomonas putida* allocated 52% of the reads (C-Score 102.9) . *Pseudomonas* genus has 93% of allocated reads (C-Score 121.6). E.coli has 2% of the reads but has a low confidence score of 38.9, suggesting that this is not an accurate allocation. An increase in reads being allocated to *P. putida* and also an increase in the C-Score of *P. putida* when compared to the BMA 25% v/v (1.56 M) treatment grou*P.*

5.2.3 RCS09 Treated with Styrene Showed a Population Shift

Unlike the BMA-treated samples, the RCS09 cultures treated with Styrene population did not show a major shift in population towards one species. The metagenomic data of RCS09 treated with Styrene showed increased reads allocated to Gammaproteobacteria from 42% of the control group to 67% (confidence score of 110.7) at 5 mM and 61% (confidence score of 119) at 10 mM. For the sample treated with 5 mM styrene (Figure 8), at the genus-level *Pseudomonas* occupied 33% (confidence score of 117.8) of the population, followed by *Acinetobacter* at 28% (confidence score of 103.9). At the species level, the algorithm allocated 7% of the reads in the RCS09 sample treated with 5mM Styrene to *P. putida* (confidence score of 81.3) and *A. johnsonnii* at 5% (confidence score of 74). For the 10 mM styrene treatment (Figure 9), the metagenomic reads allocated to *P. putida* are reduced to 0.8% (confidence score of 106.8) and reads assigned to *Acinetobacter* species represented 33% (confidence score of 122) of the population at 10mM Styrene Treatment. The 10 mM Styrene treatment group showed reads that were mapped back to pathogenic species, including *Acinetobacter baumannii*, *Stenotrophomonas maltophilia*, and *Enterobacter cloacae* complex, which were not at notable levels in the 5mM treatment group. *Enterobacter* species represented 16% (confidence score of 71.5) of the reads in the 10 mM treatment group. This project focuses on acrylic plastic, with BMA being the candidate for biosynthesis

and not the synthesis of styrene. Therefore, this is not an issue, but as this sample yielded reads allocated to pathogenic styrene-tolerant organisms, it would not be pursued as a production strain candidate. The majority of reads in the 10 mM treatment group were allocated to the *Acinetobacter* genera, but as there were no species that were allocated more than other species within the treatment group, this suggests that the species may be outside of the Maxikraken database.

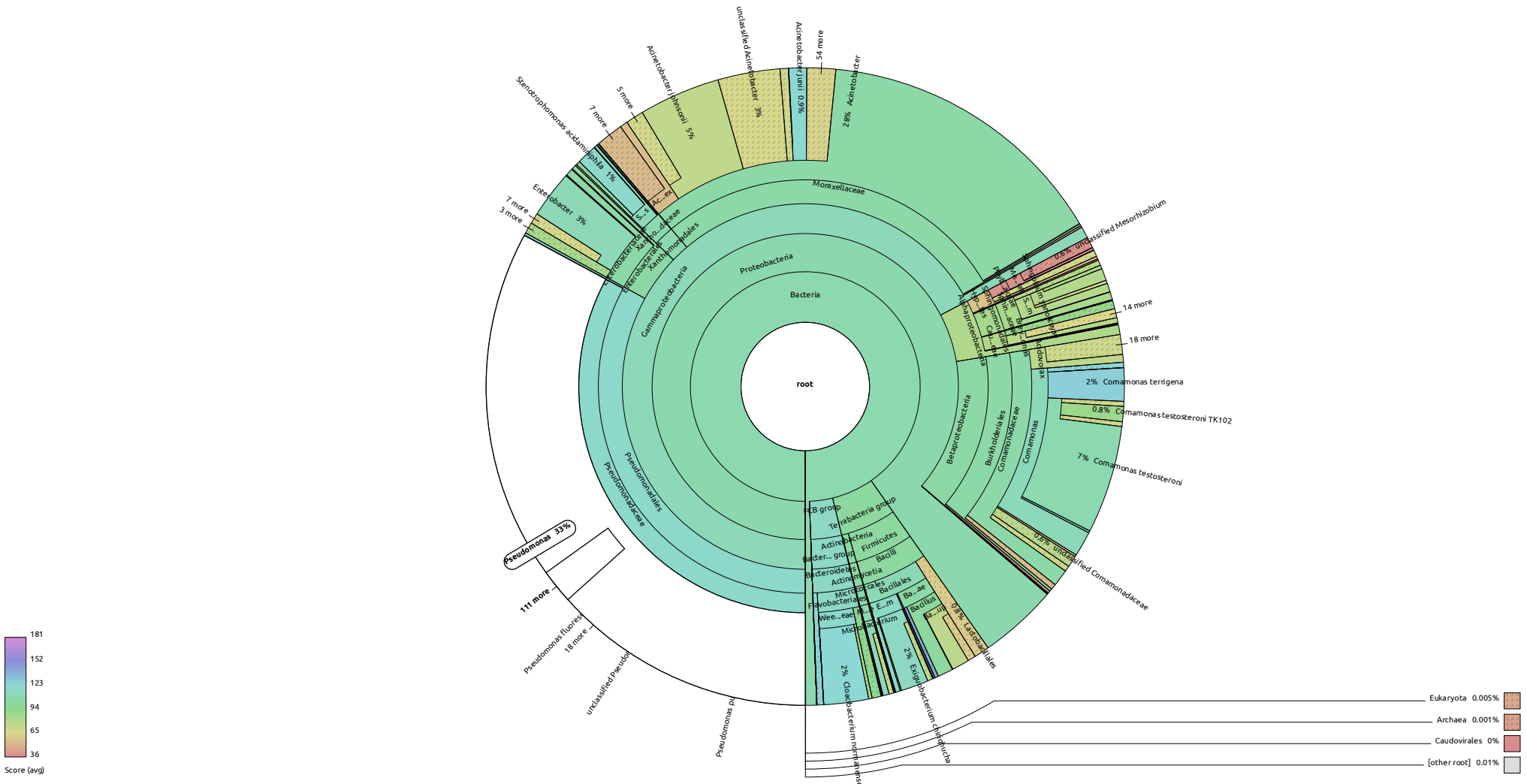


Figure 8- Styrene (5mM) Treatment of RCS09 Metagenomic read. Allocations visualised using recentrifuge. *Pseudomonas* species has 33% of the reads allocated (C-score117.8). *P. putida* has 7% of the allocated reads (C-score 81.3). *Acinetobacter* genus has 28% of the allocated reads (C-Score 103.9) But has no species with higher read allocation that the species *P. putida*.

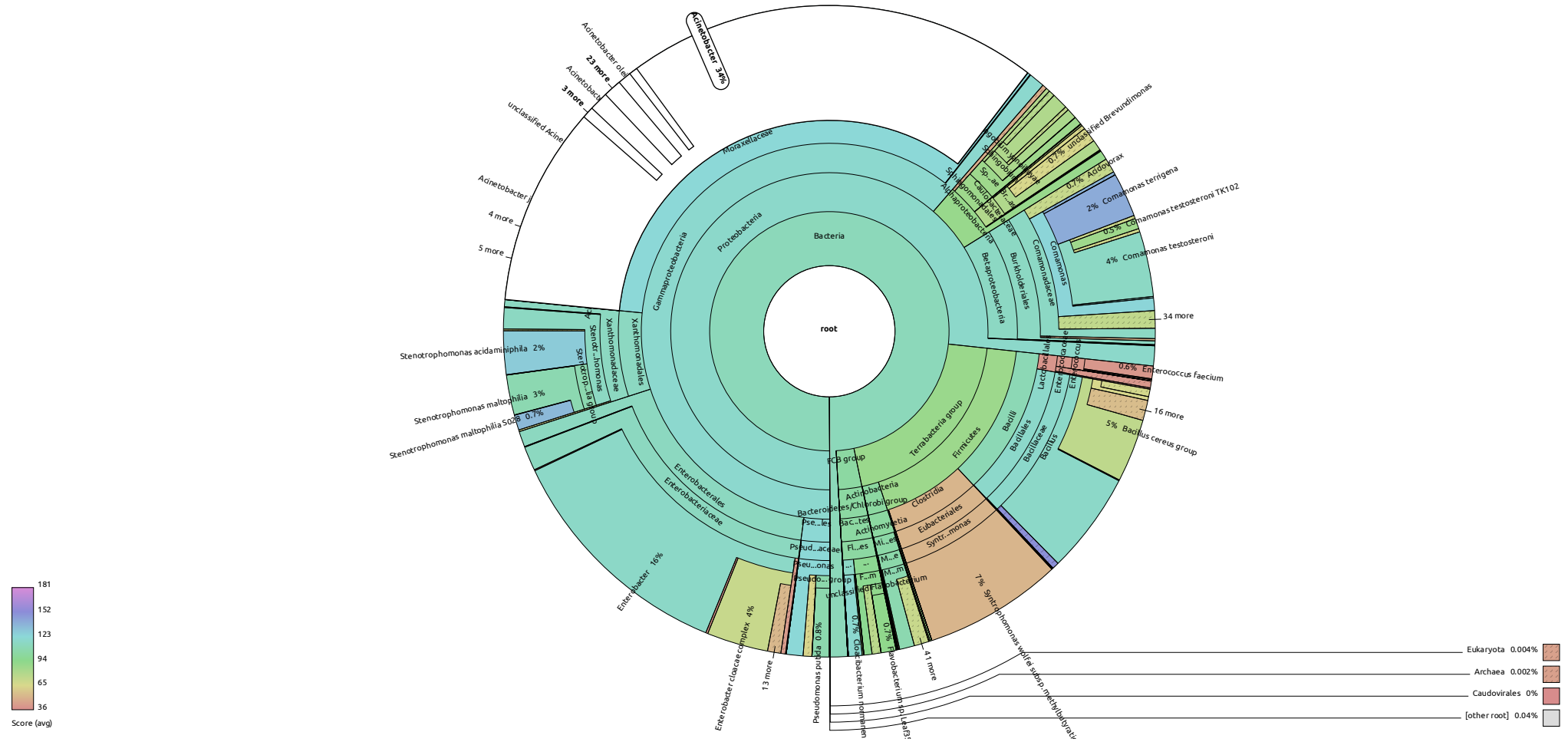


Figure 9- Styrene (10mM) Treatment of RCS09 Metagenomic reads allocations visualised using recentrifuge. The genus *Acinetobacter* has 34% of the read allocation (C-Score 122), the highest of all the genera on the plot. There is no distinct *Acinetobacter* species that is higher than other allocated reads allocated to a species. Suggesting that either it is a complex mixed population of *Acinetobacter*, or the reads belong to *Acinetobacter* species outside of the Kraken2 Maxikraken Database.

5.2.4 RCS09 R-2A Population with Ethyl Benzene

Like the Styrene treatment group, the data for the Ethyl Benzene treatment group were allocated to a much more diverse sample population than BMA. *Acinetobacter* reads accounted for 37% (confidence score of 120) at 10 mM (Figure 11) and 21% (confidence score of 109.3) at 5 mM (Figure 10). Both concentrations held no reads of notable pathogenic species with high confidence scores. *Comamonas terrigena* holds the highest amount of reads of a species (6%) with a high confidence score of 129.2. As no species within *Acinetobacter* held a higher read allocation *than C. terrigena*, even though its genus read allocation was the highest of the treatment group, it suggests that the dominant number of reads are from within the genus *Acinetobacter* but of species outside of the Maxikraken database.

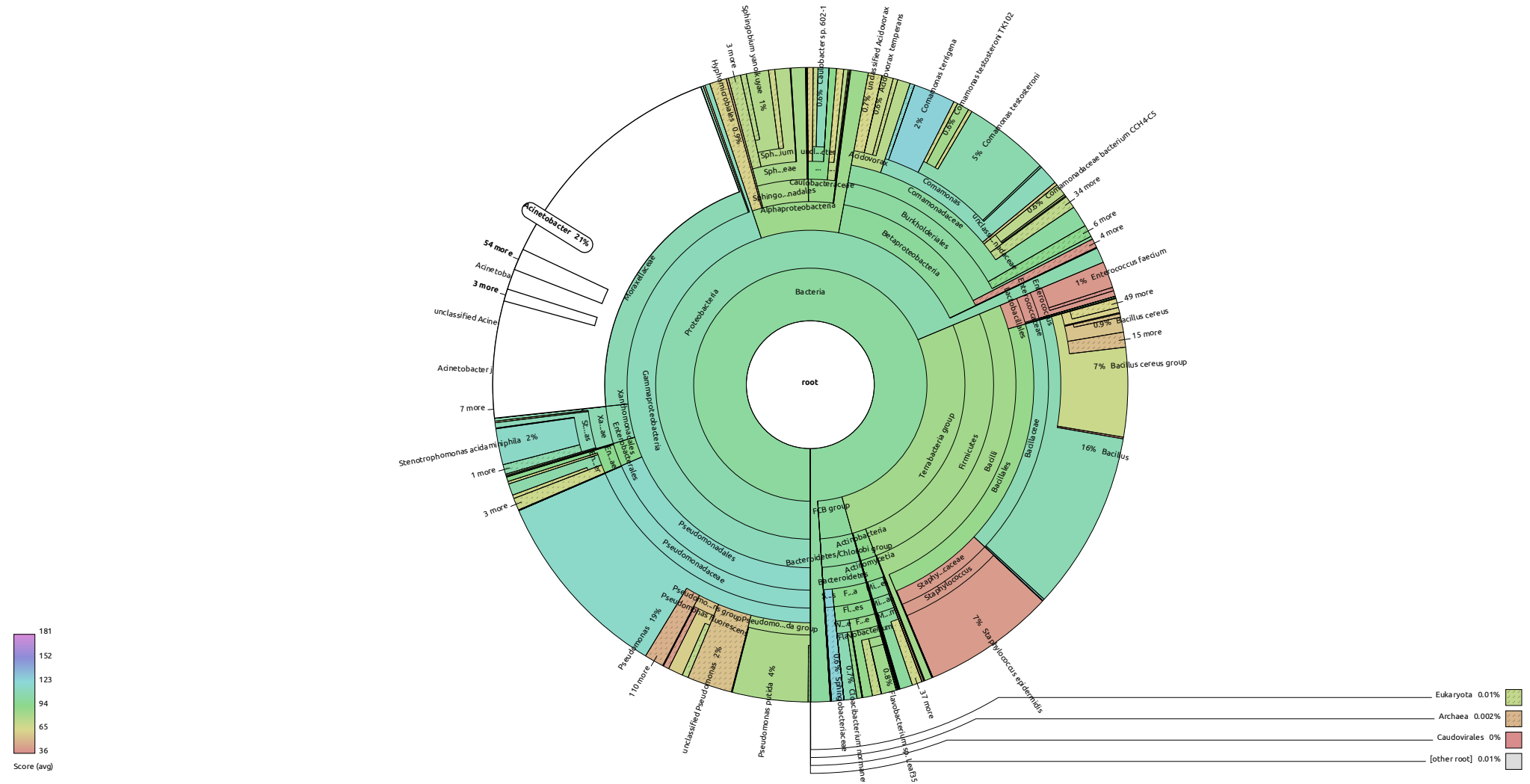


Figure 10 – Ethyl Benzene (5 mM) Treatment of RCS09 Metagenomic reads allocations visualised using recentrifuge. *Acinetobacter* genus has 21% of the reads allocated (C-Score 109.3), and *Pseudomonas* has 19% of the reads allocated (C-Score 117.1). *P. putida* has 4% of the reads (C-Score 81.2). *Staphylococcus epidermidis* has 7% of the reads allocated but with a confidence score of 41.9 indicating this is not a accurate allocation.

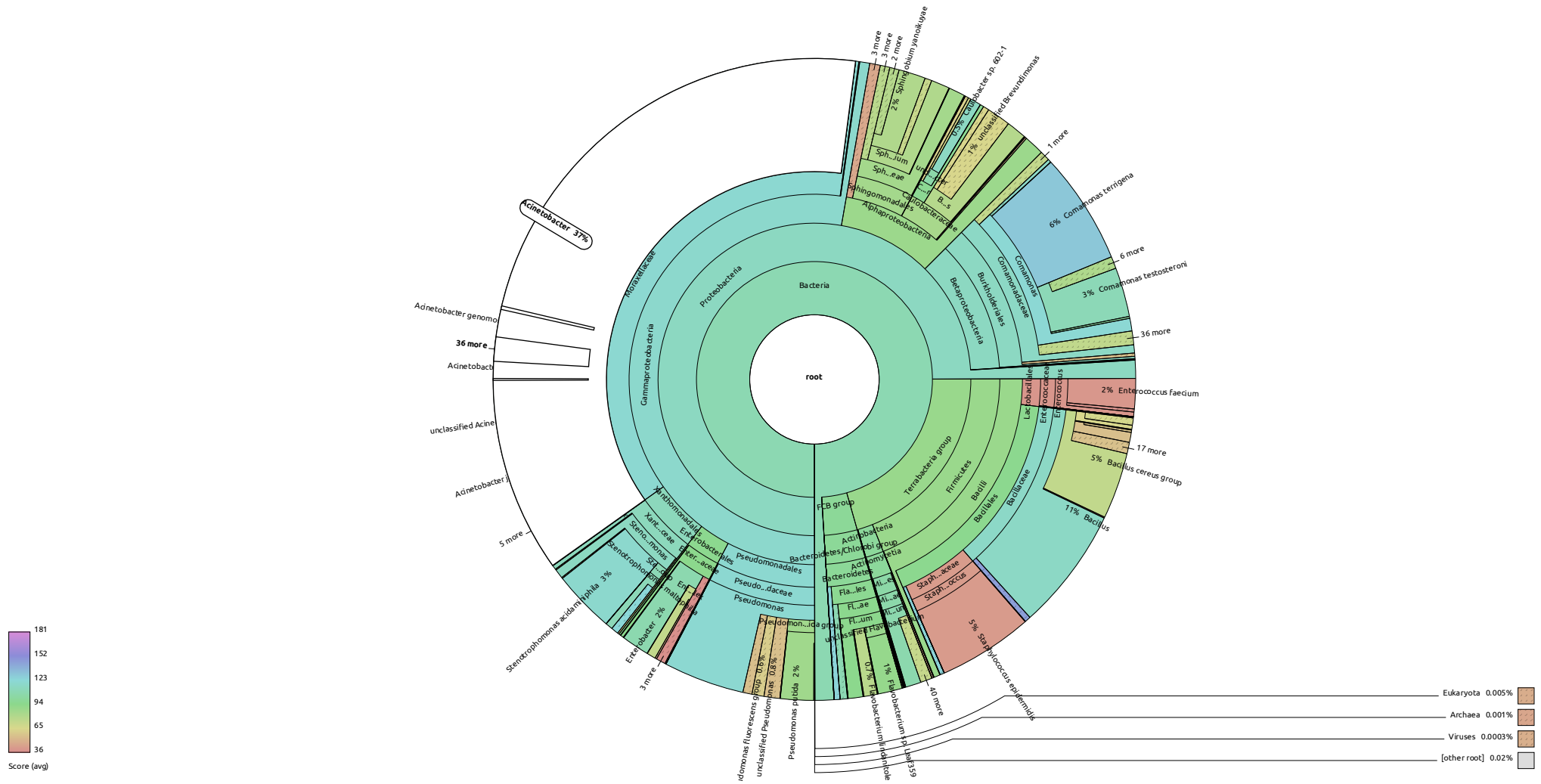


Figure 11 Ethyl Benzene 10mM Treatment of RCS09 Metagenomic reads allocations visualised using recentrifuge

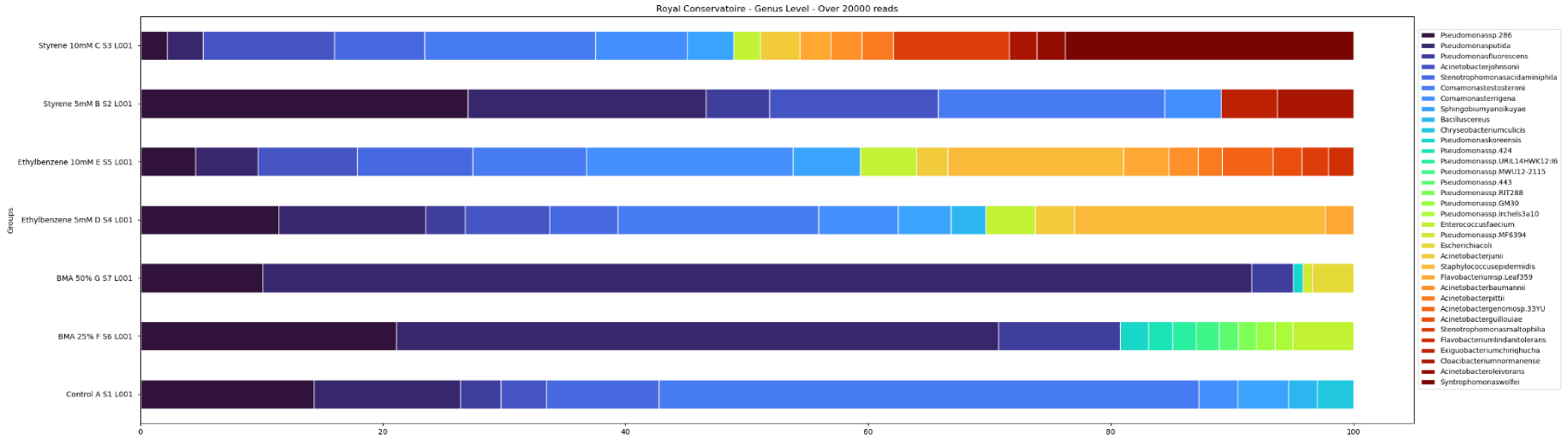


Figure 12 – Royal Conservatoire Solvent Culture Metagenomics experiment. Showing only organisms that exceed 20,000 allocated reads to reduce noise in the data. For BMA 25% v/v (1.56 M) and 50% v/v (3.12 M) *P. putida* is the dominant species along with *Pseudomonas* species 286 and *Pseudomonas fluorescens*. The same *P. putida* is represented in higher level of reads in the lower concentration of Styrene and Ethylbenzene but does not hold this at the higher concentrations.

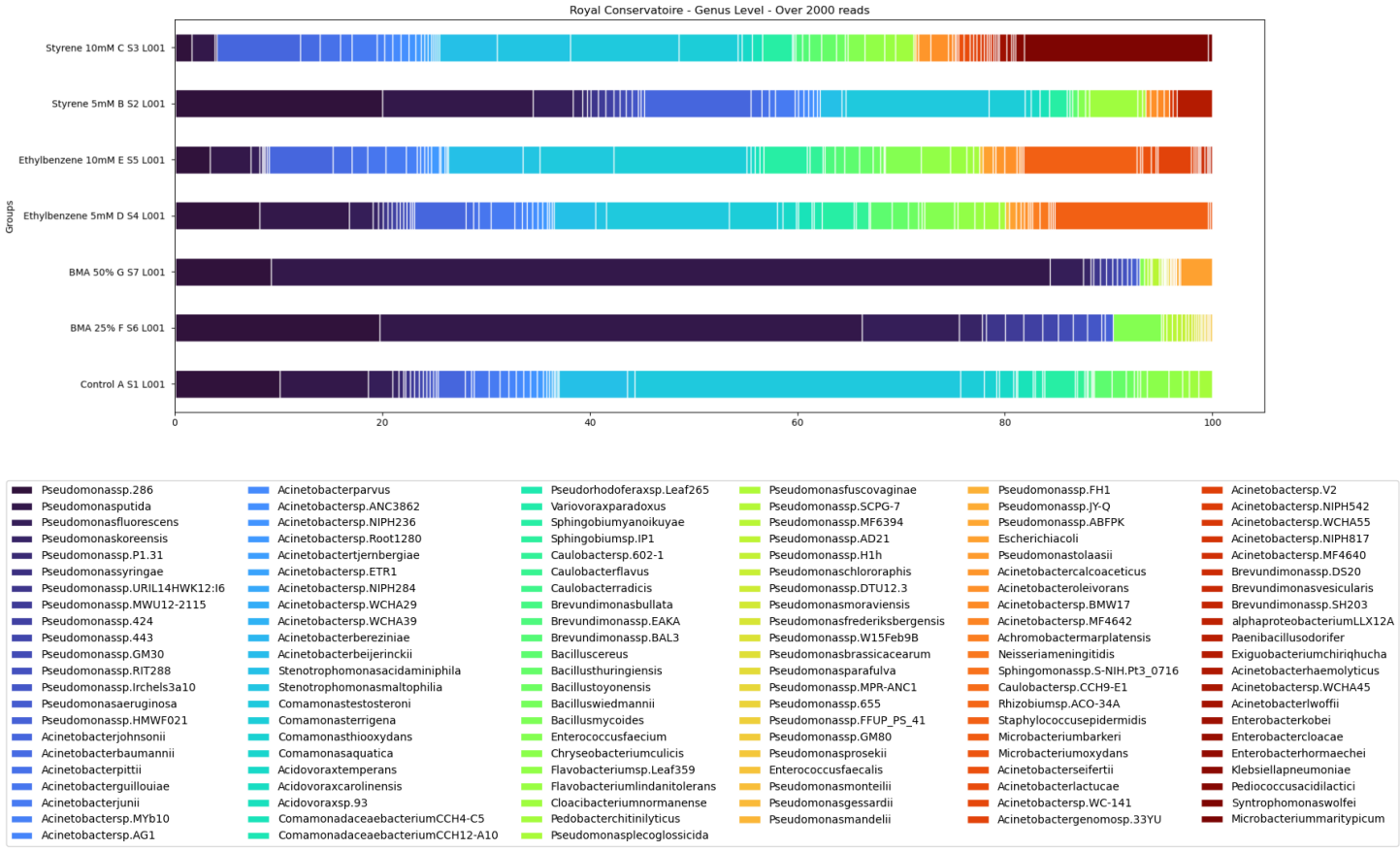


Figure 13 – RCS09 Culture-Based Solvent Metagenomic study. Only organisms that have over 2000 reads allocated to them are displayed. This figure is included to illustrate the diversity of the sample, whereas Figure 12 illustrates the dominant species.

5.2.5 RCS09 treated with BMA 50% v/v Metagenomic Assembly

Metagenomic reads from RCS09 sample treated with BMA 50% v/v were assembled using SPAdes. They were then visualised using Bandage and screened using BLAST for systems previously noted to confer solvent tolerance, such as RND efflux pumps (X. Z. Li et al., 1998a). A notable contig was highlighted with a match to the *mexB* subunit of the *mexAB-orpM* efflux system of *P. aeruginosa*. When annotated, the contig node 63 (~50kb) contained many features such as transfer gene *traD*, toxin-antitoxin system *vapBC* and a complete RND Efflux System with a regulator, suggesting it was a plasmid.

Using the NCBI BLAST database, the top match returned a plasmid from a *Pseudomonas aeruginosa* isolate RW109 (Weiser et al., 2019b). This isolate also contained a larger mega-plasmid, which held an RND efflux pump with 99.95% identity to the assembled plasmid, suggesting that they are the same system. The plasmid/organism was associated with extreme stress-resilient phenotypes, including solvent tolerance. When the metagenomic reads are mapped back to node 63, the average read depth is 233.13. Comparing this, the reads that mapped back to a *P. putida* reference genome have a read depth of 99.1, suggesting that this plasmid has a higher copy number, as shown in Figure 14 and Figure 15.

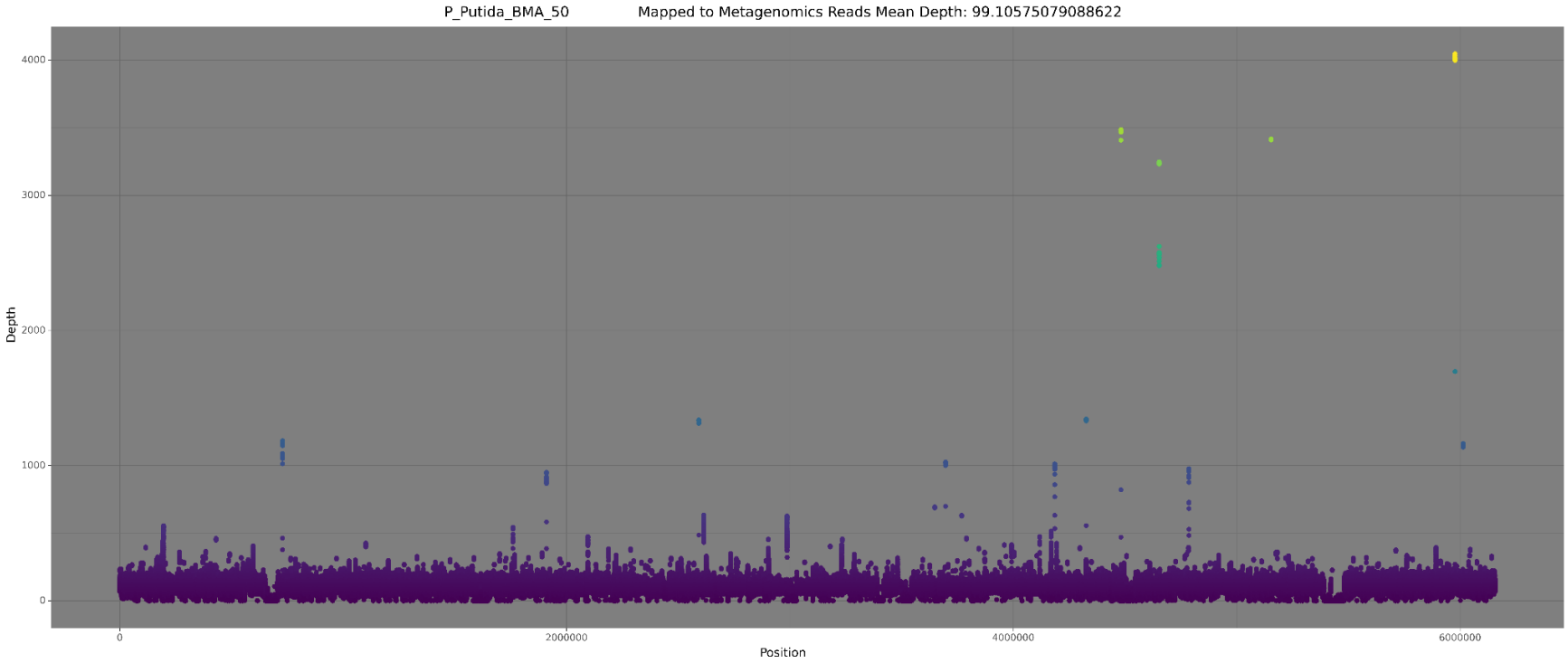


Figure 14 – RCS09 BMA 50% v/v Metagenomic Reads Mapped back to a *P. putida* reference genome using BMA-MEM aligner. The average read depth is 99.1

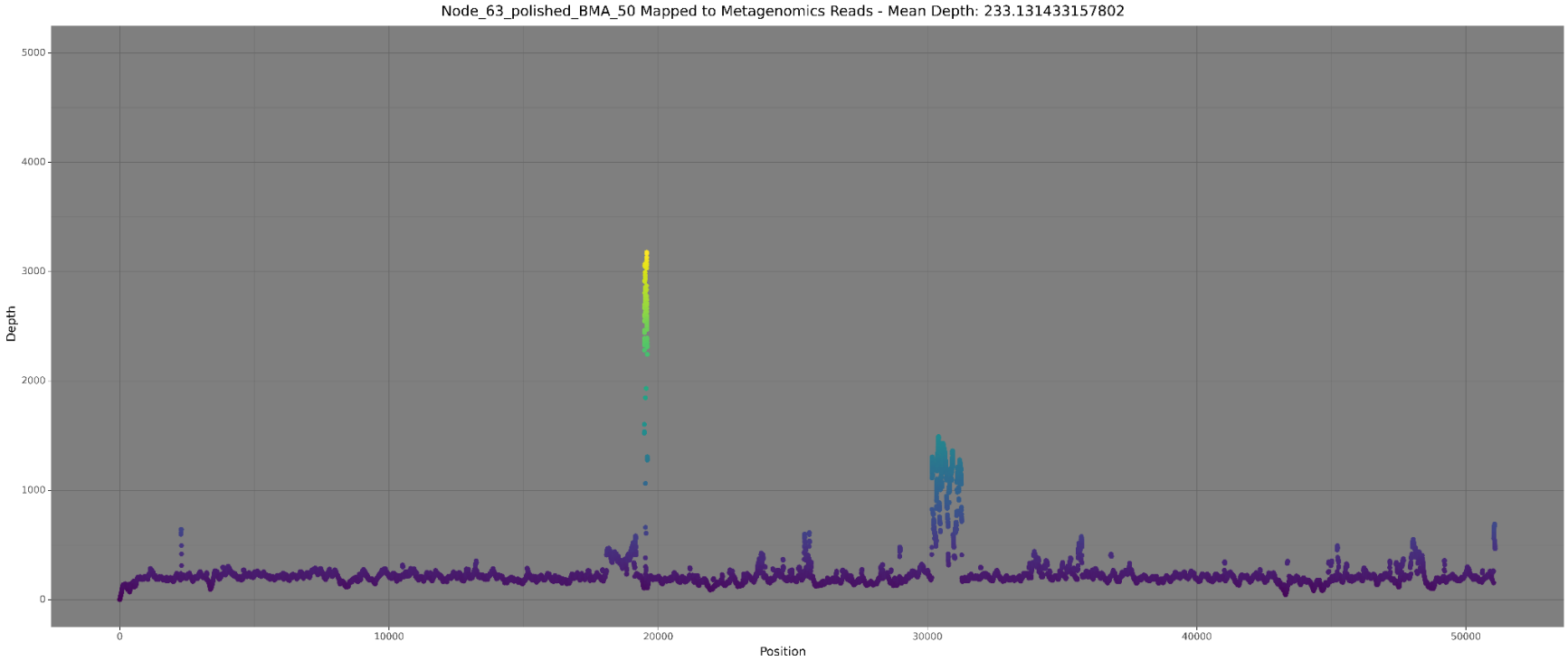


Figure 15 - RCS09 BMA 50% v/v Metagenomic Reads mapped back to the assembly contig/node 63, which contains the RND Efflux System. The average read depth is 233.13, which suggests that the potential plasmid is at a higher copy number than the chromosome, which has an average read depth of 99.1. There are several sections in which there are higher levels of reads, and these align with transposases, which are ubiquitous in genetic content.

5.3 Isolate - RCS09A – *Pseudomonas putida*

Serial dilutions were made from the mixed culture used for the metagenomic study, and the organism present in the highest dilution was isolated. This organism was approximately identified as *Pseudomonas putida* using 16S rRNA gene PCR and Sanger sequencing. To assess its tolerance to BMA compared to *P. putida* KT2440, the current industry-tolerant organism, both organisms were challenged in a BMA tolerance assay. There was a significant increase (p-value 0.004) of $1.65x10^6$ CFU of RCS09A, an increase of 2352% compared to the CFU of KT2440 (Figure 16).

WGS revealed that this organism contains the plasmid identified in the metagenomic experiment. The WGS assembly reported the size of the plasmid to be approximately 15kb larger at approximately ~65kb, but the content of particular interest; namely, the RND efflux pump, was consistent, with an identified match of 99.8%. The Flye assembly report stated the coverage of the chromosome contig to have a read depth of 34, compared to and the plasmid at 118, suggesting further that multiple copies of this plasmid exist within each cell.

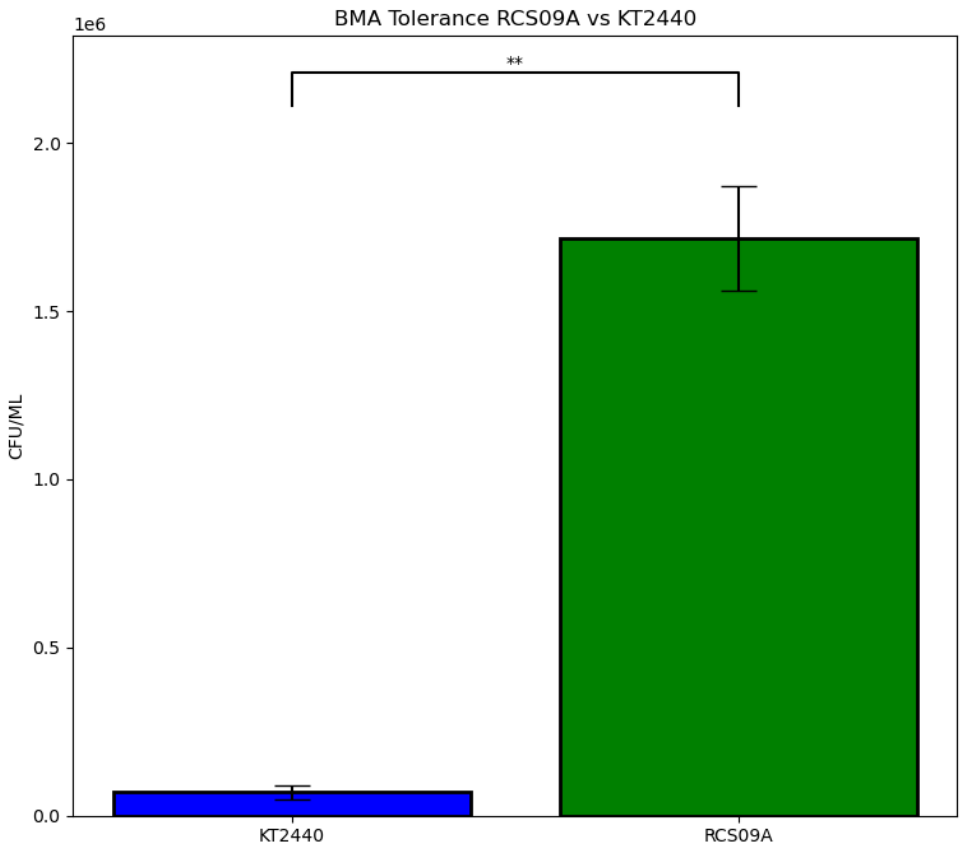


Figure 16 – BMA Tolerance Assay Results of KT2440 and RCS09A grown in 25% v/v BMA and LB media. RCS09A has significantly (P-Value 0.004 Two Sample T-Test) more CFU measured at 24 hours. Three replicates per group at 30°C.

5.3.1 Failed identification of RCS09A using autoMLST

WGS of RCS09A used a combination of nanopore long-reads, assembled using Flye and polished using Illumina short-reads with Pilon. autoMLST did not confirm the assembly/isolate to be *P. putida,* returning an estimated average nucleotide identity to *P. putida* GB-1 of 91.9%. Furthermore, the closest ANI to *Pseudomonas* SP S13.1.2 was 94.7% which is below the threshold of 95-96% to confirm the isolate as this species (Table 4) (Richter & Rosselló-Móra, 2009).

Table 4 – autoMLST results for RCS09A – No estimated ANI is about 95%

Query organism	Reference assembly ID	Reference name	Mash distance	estimated ANI	P-value	Genus	Order	Type strain
QS--Node26_Illumina_pol.fa	GCF_000292285	Pseudomonas sp. S13.1.2	0.0527	94.7%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_001320655	Pseudomonas sp. NBRC 111134	0.0571	94.3%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_001753975	Pseudomonas sp. NBRC 111141	0.0573	94.3%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_000019125	Pseudomonas putida GB-1	0.0811	91.9%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_001320565	Pseudomonas sp. NBRC 111132	0.0825	91.8%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_001320885	Pseudomonas sp. NBRC 111140	0.0830	91.7%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_000514335	Pseudomonas sp. LAIL14HWK12:I12	0.0839	91.6%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_001753915	Pseudomonas sp. NBRC 111122	0.0841	91.6%	0.0000	Pseudomonas	Pseudomonadales	False
QS--Node26_Illumina_pol.fa	GCF_002112825	Pseudomonas sp. B5(2017)	0.0843	91.6%	0.0000	Pseudomonas	Pseudomonadales	False

5.3.2 RCS09A Closest Genome Match *P. putida* using pyANI

The estimated average nucleotide identity of the organisms identified by autoMLST estimations was confirmed using pyANI, shown in Figure 17. To further understand the isolates' identity, pyANI was used to determine the average nucleotide identity of a more extensive database of 221 *Pseudomonas putida* genomes downloaded from the NCBI database. Thirty genomes from this analysis passed the 95% threshold shown in Figure 18 and Figure 19 confirming RCS09A to be *P. putida*. To further characterise the strain identity down to its closest neighbour, ANI match the thirty genomes that passed the 95% ANI threshold were investigated further.

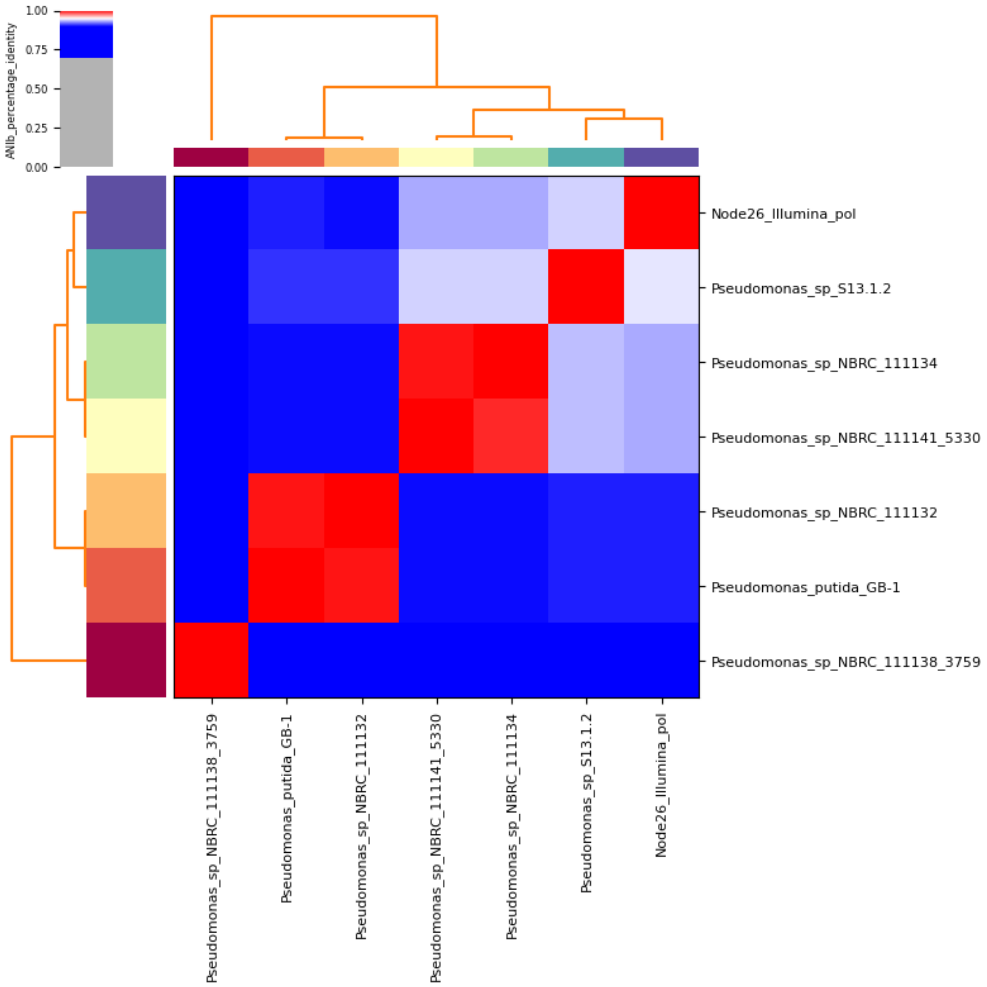


Figure 17 - pyANI results of top hits from autoMLST and Node26 the chromosome assembles of RCS09A. No Chromosome has an ANI of over 95%, which would be indicated by a pink-red colour.

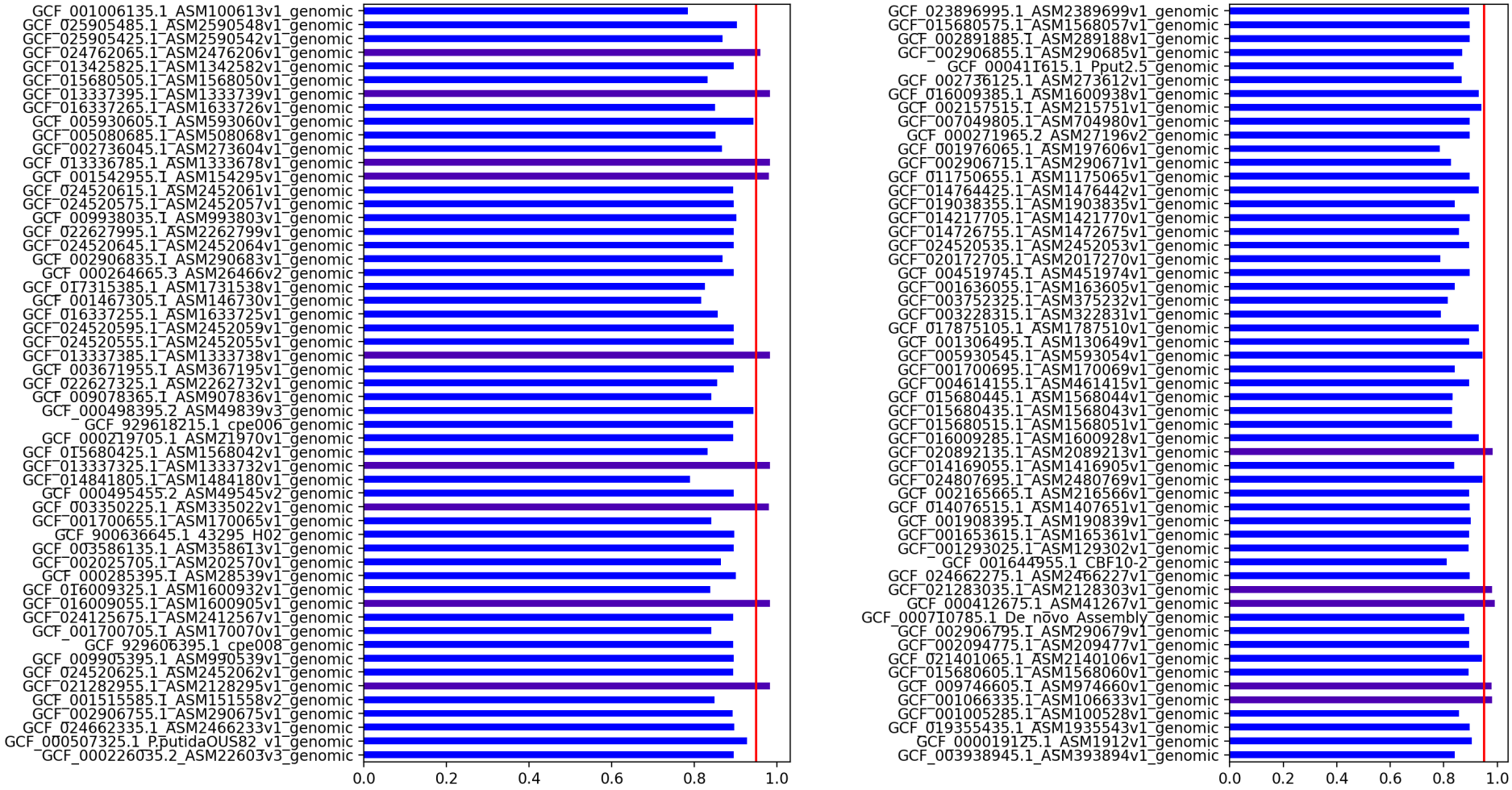


Figure 18 -Average Nucleotide Identity of Royal Conservatoire Isolate RCS09A chromosome contig Node 26 compared to 221 *P. putida* genomes. All bars crossing the red line indicate an ANI of over 95% confirming the RCS09A chromosome contig Node 26 as a, *P. putida* chromosome. This plot shows 14 chromosomes with ANI greater than 95%

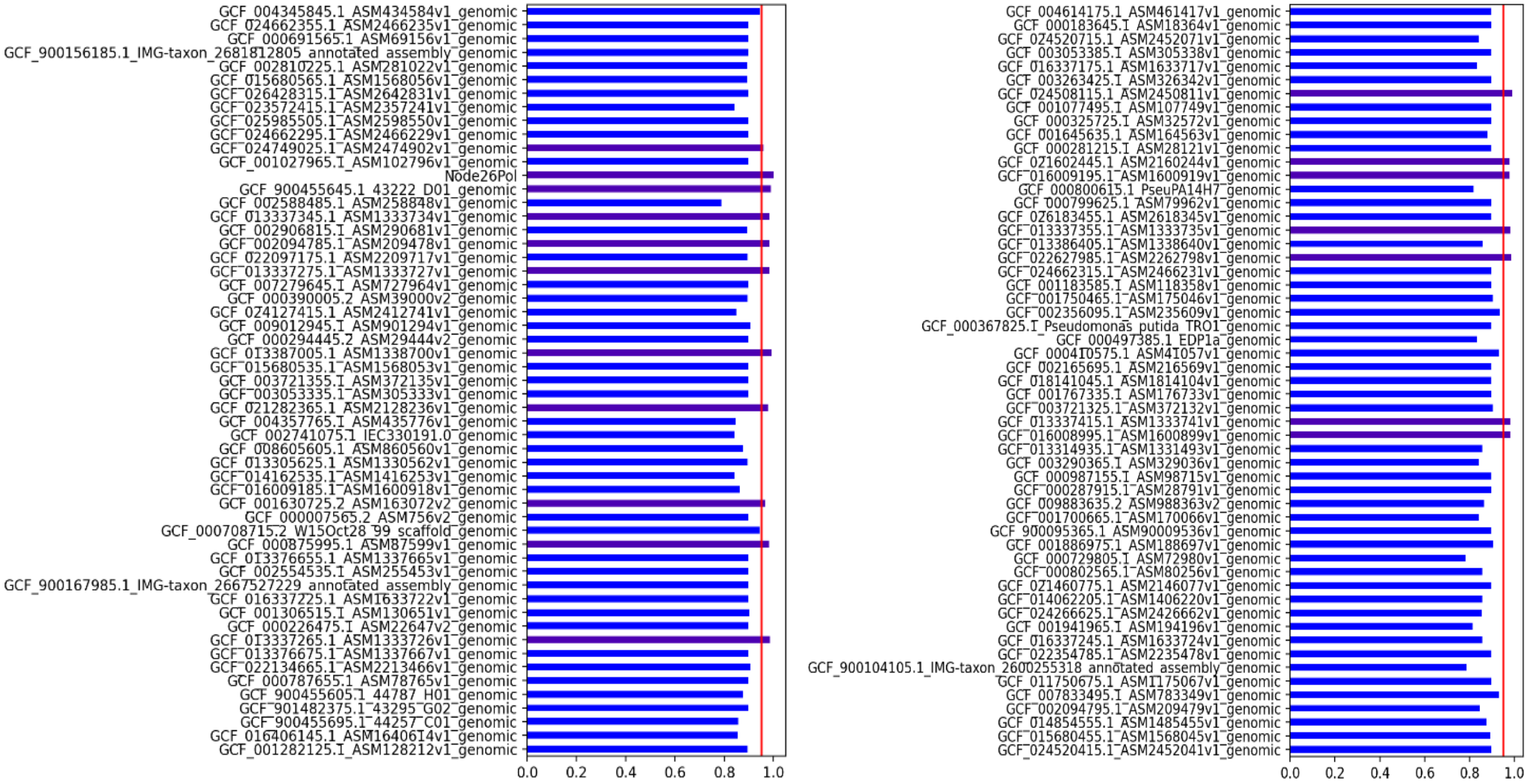


Figure 19 - Average Nucleotide Identity of Royal Conservatoire Isolate RCS09A chromosome contig Node 26 compared to 221 *P. putida* genomes. All bars crossing the red line indicate an ANI of over 95%, confirming the RCS09A chromosome contig Node 26 as a *P. putida* chromosome. This plot shows 16 chromosomes with over 95% ANI of RCS09A

5.3.3 RCS09A Closest Genome Match is *P. putida* DZ-C20 using ANI and Genome Length

To further ascertain the closest genome, we used average nucleotide identity and genome length to find the genome with a similar size and ANI. Figure 20 shows genome GCF_002094785.1, which is closest in genome length and ANI. The genome belongs to a *P. putida* DZ-C20 strain isolated in Beijing from a chicken swab, thus confirming its identity as *P. putida* (Valadbeigi et al., 2023).

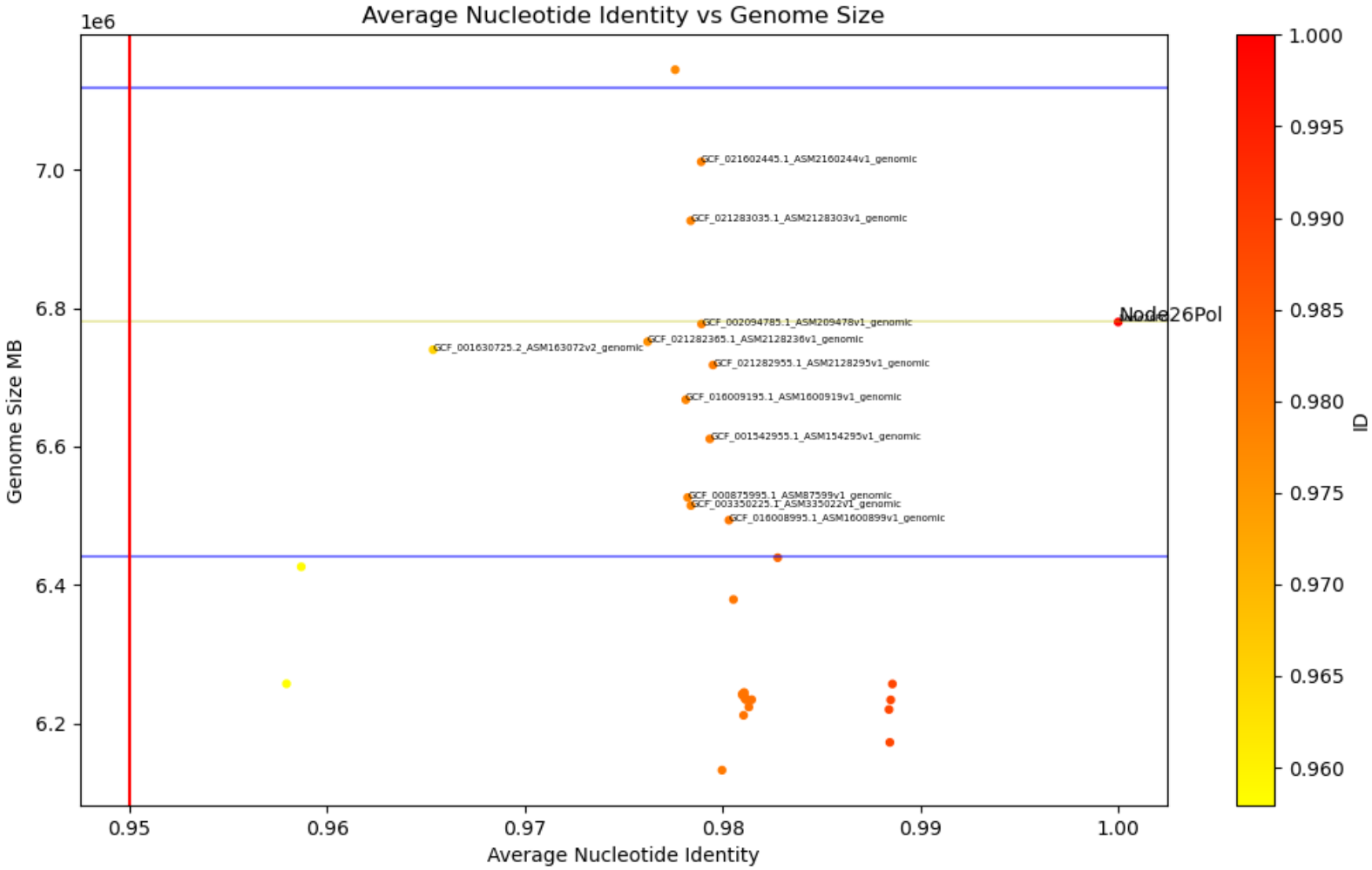


Figure 20 —Average Nucleotide Identity plotted against Genome size. RCS09A chromosome contig Node 26 shows 100% ANI on the right. *P. putida* GCF_002094785.1 shares 98% ANI and an almost identical genome size. This has been identified as *P. putida* DZ-C20 isolated from a chicken in Beijing. The yellow line indicates the length of the RCS09A chromosome contig Node26. The blue lines are a 10% allowance of the length of Node 26. The red line indicates the 95% threshold to be considered the same species.

5.4 Isolation of RCS09A(2) and RCS09A(3)

During the final months of this study, it was required to re-isolate RCS09A from the original RCS09 sample challenged with BMA. This was due to RCS09A failing various PCR-based quality control checks, which were necessary to resolve a contamination of the -80 glycerol stock. In the end, RCS09A(2) and RCS09A(3) were re-isolated from the original RCS09 environmental sample as well as RCS09A. However, these also failed the PCR quality checks, confirming them to be different isolates from the initial RCS09A isolate. Despite this, they were still tested to ascertain if they were more tolerant to BMA than RCS09A. BMA tolerance assays were conducted using KT2440 as a benchmark.

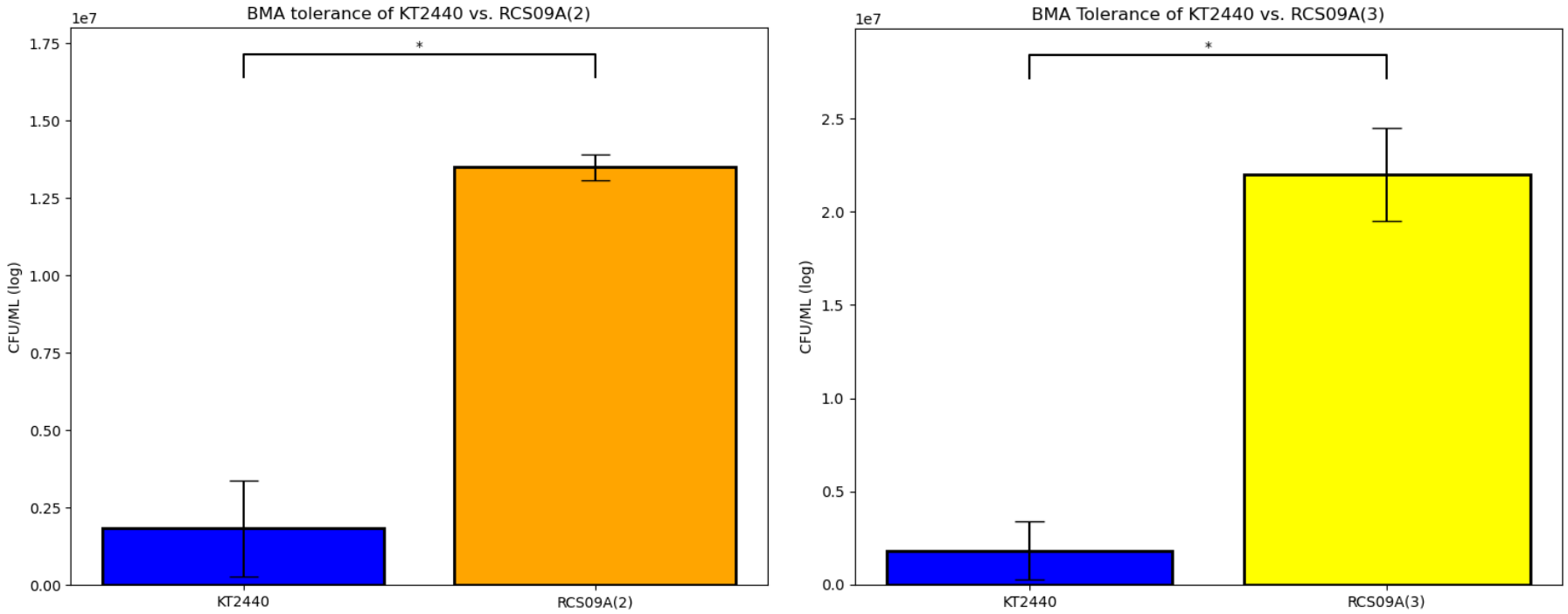


Figure 21 – BMA tolerance assay of RCS09A(2) (**LEFT**) and RCS09A(3) (**RIGHT**) grown in LB for 24 hours treated with BMA 25% v/v. Both isolates are shown to be significantly more tolerant to BMA than KT2440 (P-Value = <0.05).

5.4.1 BMA Tolerance of RCS09A(2) and RCS09A(3)

RCS09A(2) and RCS09(3) were significantly more tolerant to BMA than *P. putida* KT2440 (Figure 21). To further understand their tolerance compared to that of the original RCS09 tolerant isolate RCS09A, the increase in CFU were presented as a percentage increase compared to KT2440.

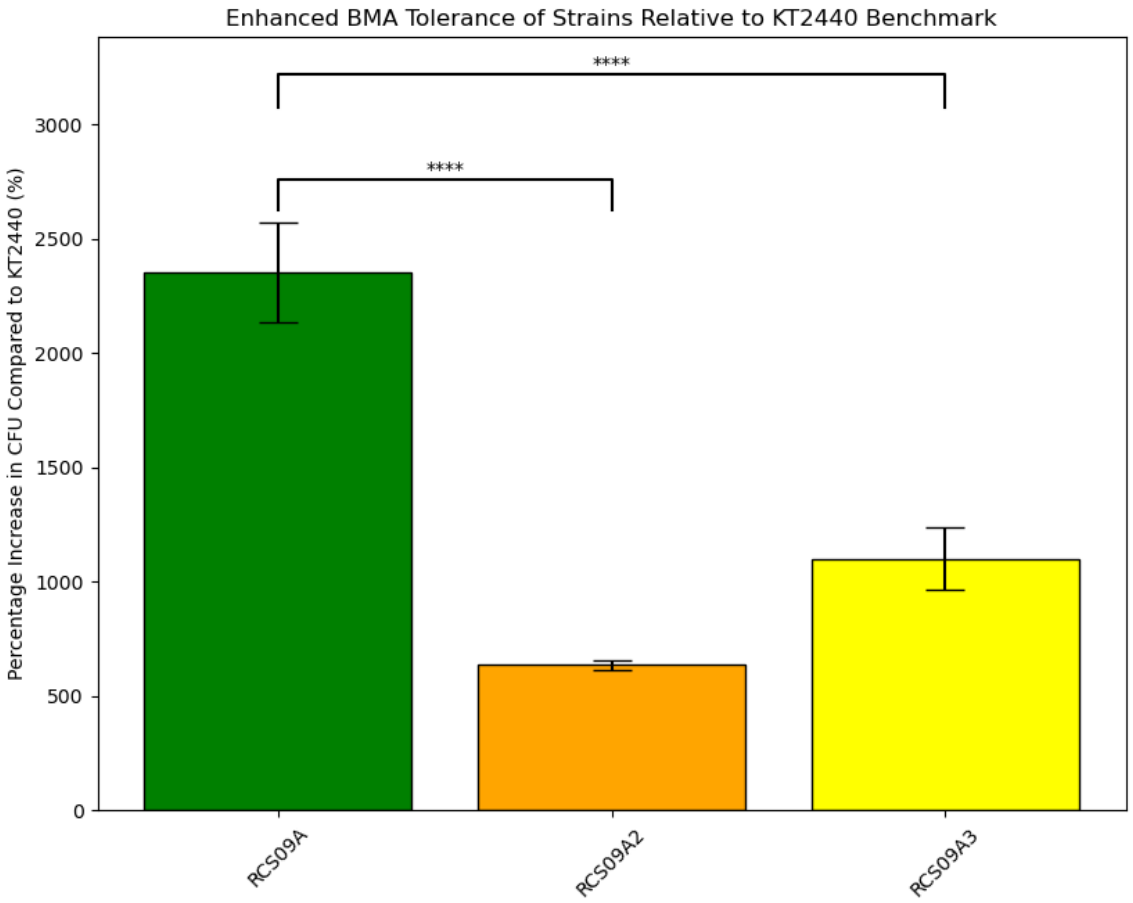


Figure 22 – BMA Tolerance represented as a percentage increase of CFU when compared to KT2440. A one-way ANOVA detected variance mean percentage increase and confirmed using Tukey HSD.

RCS09A percentage increase relative to KT2440 was significantly higher than that of RCS09A(2) and RCS09A(3) (P-Value = <0.0001).

5.4.2 Comparison of BMA tolerance relative to KT2440 of RCS09A, RCS09A(2) and RCS09A(3)

When comparing the BMA tolerance relative to their percentage increase in CFU against the KT2440 benchmark, RCS09A was shown to be significantly more tolerant than RCS09A(2) and RCS09A(3) (Figure 22). As these isolates were isolated during the final month of the study, Eurofins bacterial whole genome sequencing service was used to charcterise RCS09A(2) and (3). The service only produced reads for RCS09A(2), but not enough to complete an assembly. The largest contig was analysed using autoMLST and was shown to have the highest estimated ANI with *Pseudomonas* sp. S13.1.2 (GCF_000292285) with an estimated ANI of 91%. Due to the incomplete assemblies for each of these isolates, and not being as BMA tolerant as RCS09A and the time constraints on the project, they underwent no further investigation but were included as tolerant organisms.

5.5 Mitsubishi Chemical Production Facility Teesside Isolate

When travel restrictions were lifted and travel to the Mitsubishi Chemical Production facility was permitted, nine sites were sampled. Several samples were taken from each location for a diverse range of experiments. Swabs were directly added to Ringer's solution and R-2A for 16S rRNA gen PCR applications, another set as before for potential full metagenomics, a more prominent sample location permitting, and a swab preserved in charcoal transport media (Table 5).

Table 5 – Mitsubishi Sample Key

Site Location	16s Metagenomic Sample Swab in A-Media or B-Ringers	Metagenomics Sample Swab in A-Media or B-Ringers	Large Sample	Charcoal Transport Media Sample Swab
Red Barrel (A)	MCA-01-A/B	MCB-01-A/B	MCA-03	
Gully 1 (B)	MCB-01-A/B	MCB-02-A/B	MCB-03	MCB-04
(C)	MCC-01-A/B	MCC-02-A/B	MCC-03	MCC-04
Sump Drainage (D)	MCD-01-A/B	MCD-02-A/B	MCD-03	MCD-04
Gully 2 (E)	MCE-01-A/B	MCE-02-A/B		MCE-04
TE-IBC (F)				MCF-04
Gully 3 (G)	MCG-01-A/B	MCG-02-A/B	MCG-03	MCG-04
Gully 4 (H)				MCH-04
Guly (I)				MCI-04

5.5.1 Mitsubishi Chemical Site 16S Metagenomics

Full metagenomics is expensive, and processing all the sampling sites in the Mitsubishi Chemical Production facility would be time-consuming. Therefore, an initial 16S rRNA gene Metagenomic Study was used to pick sites that indicated species of interest highlighted in the previous experiment with samples from the Royal Conservatoire. Picking the most promising sites also allows for biological replicates and statistical analysis to be carried out in a full-metagenomic study. Our Illumina sequencer can process 12 barcodes and output 2 million reads per barcode, which allowed for four samples with three replicates each.

The samples collected at the production facility were not like the environmental samples which had been taken from elsewhere. They were heavily contaminated with chemicals and viscous. Several different sampling methods were used. Swabs in damp areas were taken and preserved in Charcoal Transport Media. Scoops of sludge were collected and stored in falcon tubes. Liquids were pipetted, added directly to both R-2A and LB media, and stored on ice for transport.

Once back in the lab, each sample was subjected to direct 16S PCR, but these failed, possibly due to chemical contamination from the site. This left the use of culture-based methods to grow bacteria from swab and liquid sample innocula in R-2A Media to achieve the best representation of the

original population (Reasoner & Geldreich, 1985b). DNA was then extracted using a PowerSoil kit, as this has also been demonstrated to give the best representation of the bacterial population (Pearman et al., 2020). 16S amplicons were prepared and sequenced on an Oxford Nanopore Sequencer. After approximately 30,000 reads were obtained for each site, the sequencing run was ended, and the flow cell was washed for further use.

Sites G, C, and D containing *Pseudomonas* and *Acinetobacter* were identified, as shown in Figure 23. Samples at these sites were selected for the next solvent culture metagenomics experiment. Unfortunately, these samples were not viable when challenged with BMA, whereas *Pseudomonas* species had previously been shown to grow in BMA in the RCS09 metagenomic experiment (Figure 7). Samples from sites G, C, and D were streaked onto *Pseudomonas* Isolation Agar, and each isolate was individually challenged with BMA. Only one isolate, taken from the site G plate, grew in media supplemented with BMA 25% v/v.

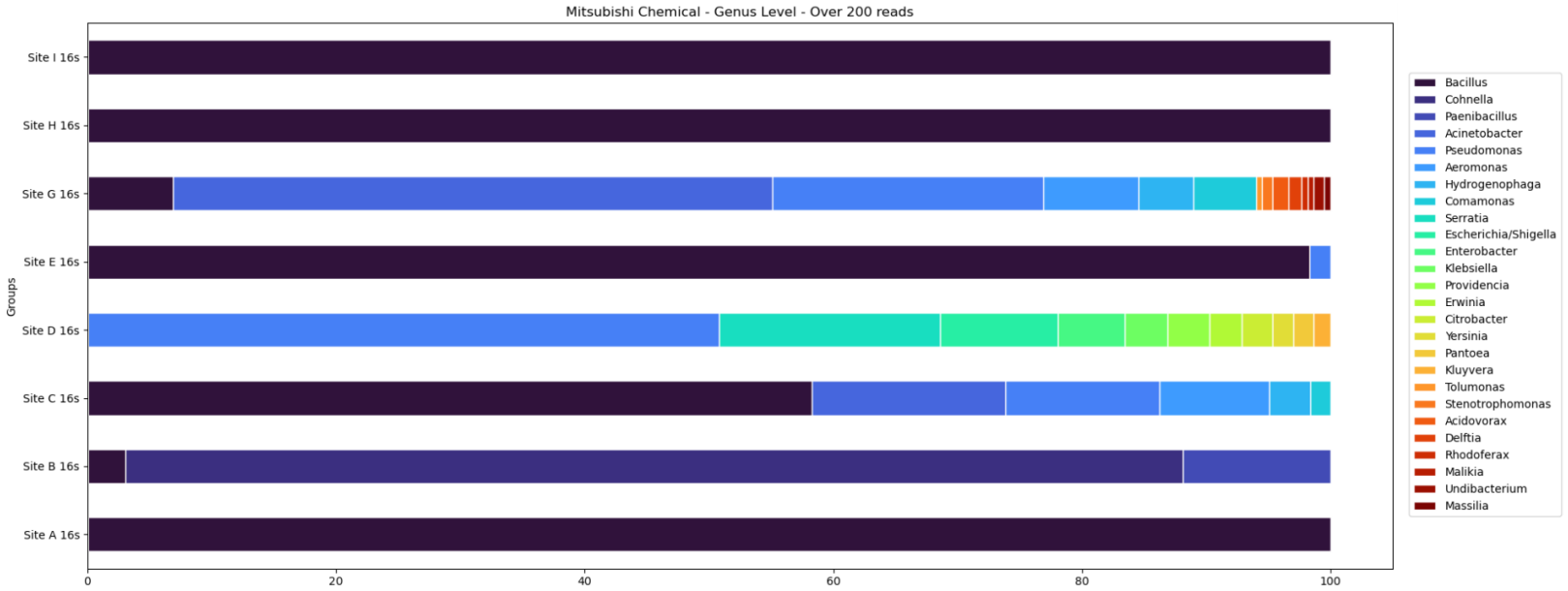


Figure 23 – Mitsubishi Chemical Production Facility 16s Amplicon Metagenomics of cultures grown from swabs and liquid sample innocula from sites A-I grown in R-2A. Amplicons were sequenced using a MinION MK1C to 30,000 reads per sample. Reads were processed and allocated to genera using Kraken2 and the SILVA 16S database. Sites G, D and C had reads assigned to the genera *Pseudomonas* and *Acinetobacter*. Both these genera demonstrated tolerance to platform chemicals used in the plastics industry based on the previous study conducted on the royal conservatoire sample.

5.5.2 Isolate -MC-G-03 – *Pseudomonas* veronii

The BMA-tolerant Isolate MC-G-03, isolated from site G and grown from sampling type 3 a swab, was approximately identified using 16S PCR and Sanger sequencing as *Pseudomonas veronii*. WGS of MC-G-03 and autoMLST confirmed its identity as *P. veronii*, which is used commonly used as a bioremediation organism to resolve environmental chemical spills containing simple aromatic organic molecules (Nam et al., 2003a)(Onaca et al., 2007a). To assess its tolerance to BMA when compared to KT2440, both organisms were subjected to a BMA tolerance assay.

The results of the tolerance assay showed that MC-G-03 was significantly more tolerant to BMA than KT2440 (Figure *24*). This experiment measured tolerance as CFU compared to KT2440 when treated with BMA. KT2440 was used as a baseline measurement of tolerance, as this is Mitsubishi Chemical's current production strain in development (Mitsubishi Chemical, personal communication). A 618% increase in CFU was measured when treated with BMA compared to the baseline strain.

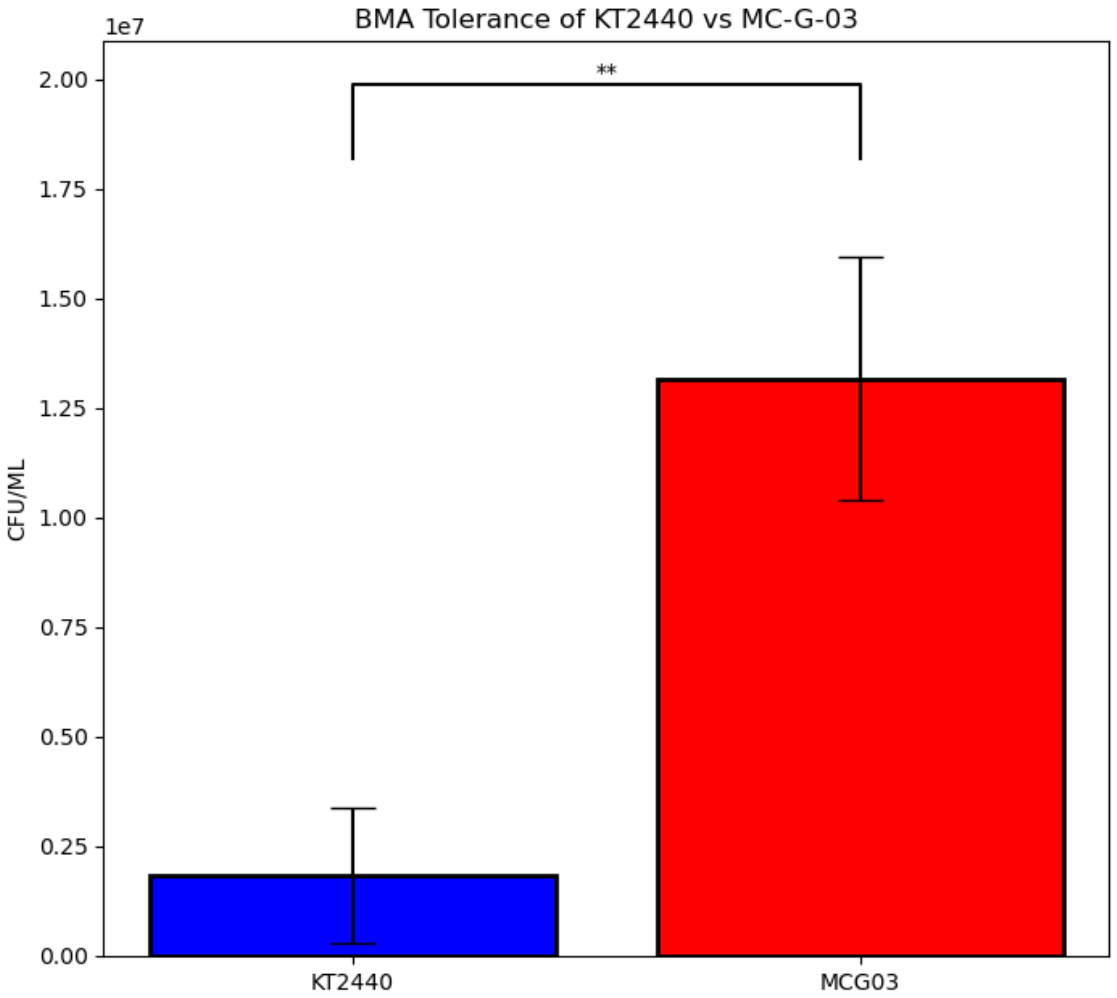


Figure 24 - BMA Tolerance Assay of KT2440 and MC-G-03 Mitsubishi Chemical Isolate. 11333333 increase CFU in MCG03. 618% significantly (P-Value 0.002 ** Two Sample T-Test) higher than KT2440.

5.5.3 BMA-tolerant isolates were Analysed using KT2440 as a Baseline Strain for Tolerance

To compare each isolate's tolerance, their growth compared to KT2440 was used as a baseline for the current tolerance standard. The CFU of each isolate was quantified and expressed as a percentage increase relative to KT2440. This normalised the comparison between various tolerance assays and allowed for direct comparison between isolates,

A one-way ANOVA and Tukey HSD was conducted on the percentage increases to determine if there were statistically significant differences in the CFU increase and tolerance between the isolates. RCS09A showed a significant increase compared to all other isolated organisms shown in Figure 25. There was no other significance between other isolates.

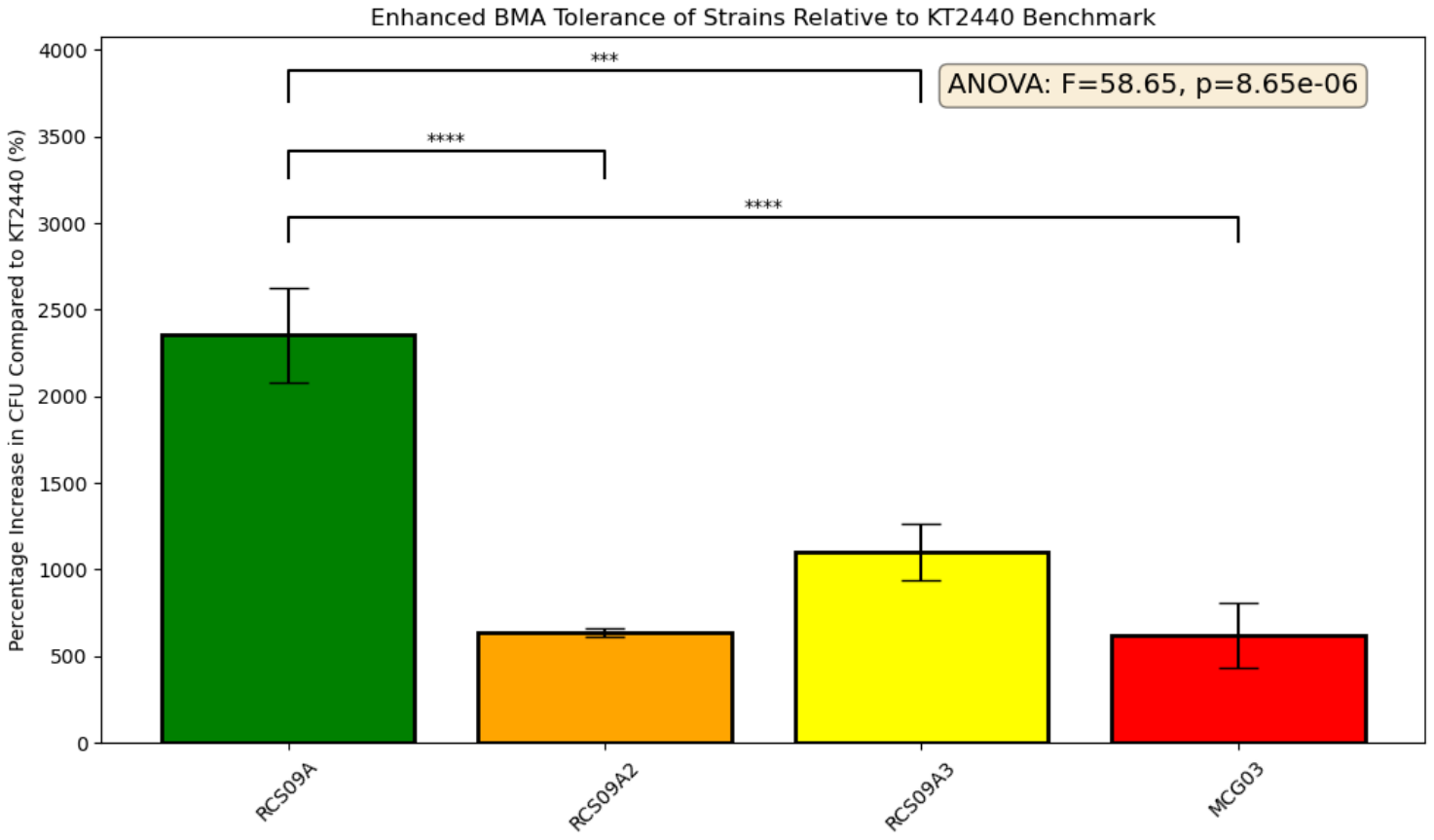


Figure 25 – BMA tolerance using KT2440 as a benchmark. Percentage increase of CFU count when compared to KT2440 to allow for direct comparison of strains between experiments RCS09A was significantly more tolerant when compared to KT2440 than RCS09A(2) (P-Value = <0.0001),(3) (P-Value = <0.001) and MC-G-03 (P-Value = <0.0001) using a one-way ANOVA and a Tukey HSD.

5.6 Comparative Genomics

5.6.1 Introduction

To understand BMA tolerance, we can use comparative genomics of the reference organism *P. putida* KT2440 compared with our the Royal Conservatoire isolate RCS09A (*P. putida*) and the Mitsubishi Chemical isolate MC-G-03 (*P.* veronii). Each environmental isolate shows an increase in tolerance to BMA when compared with KT2440. The work described in this section aimed to investigate what genetic mechanisms (if any) the two isolates share. The plasmid carried by RCS09A was not included, as this was studied separately in the work discussed in the next chapter. Roary was used to analyse the pangenome of the BMA tolerant isolates and *P. putida* KT2440.

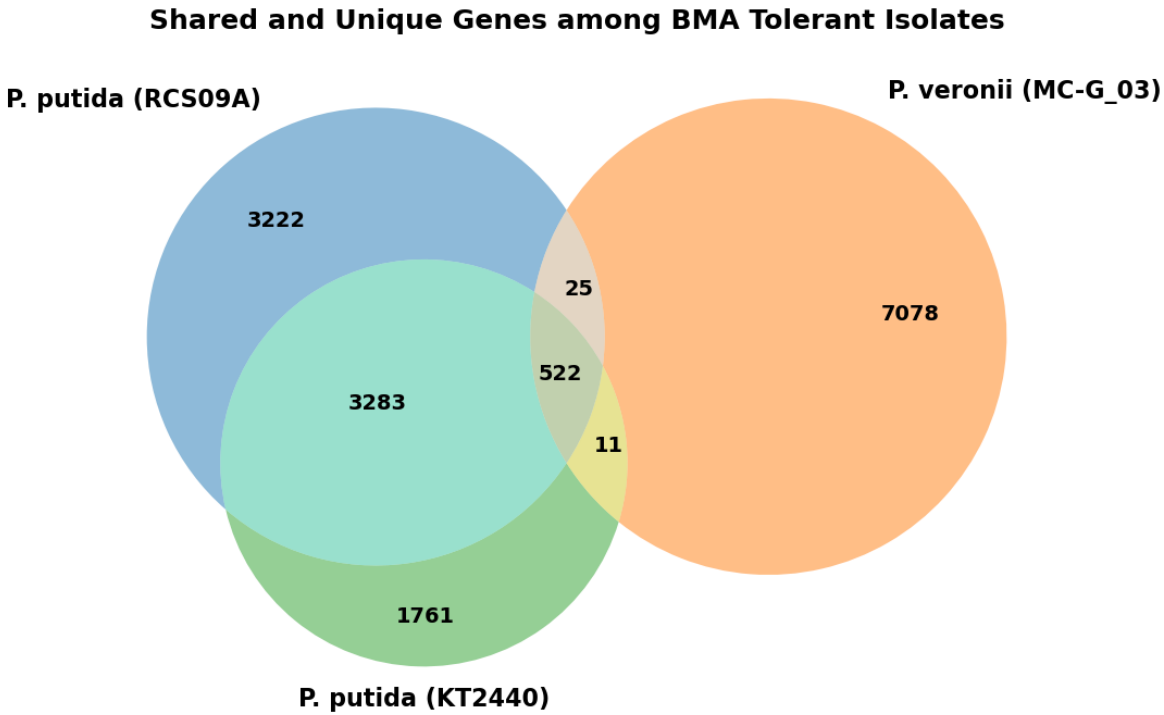


Figure 26- Pan chromosomal genome of BMA tolerant isolates and *P. putida* KT2440. The pangenome of all three strains was 552 genes. Between the two *P. putida* strains was 3283 genes. MC-G-03 *P.* veronii held 7078 unique genes, RCS09A *P. putida* held 3222 unique genes and KT2440 *P. putida* held 1761 unique genes.

5.6.2 Pangenome of BMA tolerant isolates and *P. putida* KT2440

The core genome of the three BMA-tolerant strains comprised 552 genes. The two *P. putida* strains shared 3283 genes. *P.* veronii MC-G-03 had 7078 unique genes comprising the majority of its genome. RCS09A held 3222 unique genes, and KT2440 held 1761 unique genes (Figure 26). The 522 core genes were analysed, looking for products involved in solvent tolerance, and none were found. To further understand the pangenome concerning efflux and solvent tolerance, the generated Roary database was searched using the keywords: mex, Mex, ttg, Ttg, mdtA, Mdta, rnd and RND.

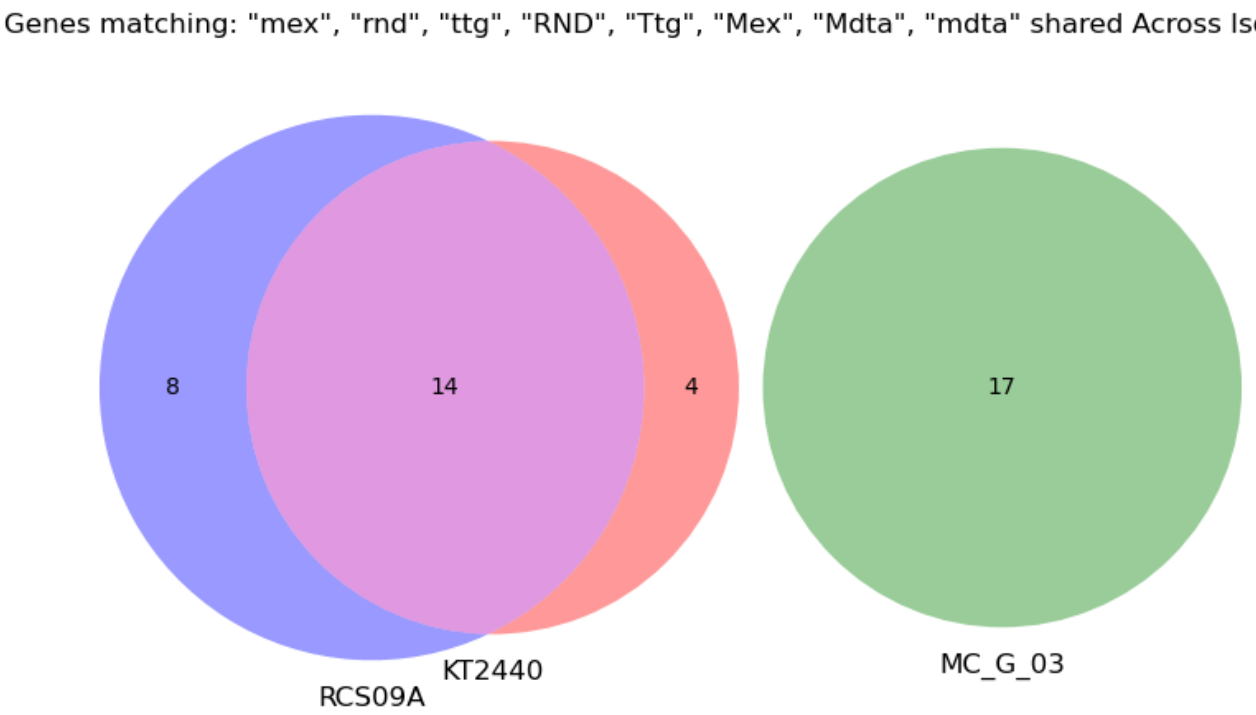


Figure 27 - Shared genes matching with RND Efflux Keywords from Roary pangenome database of RCS09A, MC-G-03 and KT2440. Keywords: mex, Mex, rnd, RND, ttg, Ttg, mdtA and Mdta. The two most tolerant strains, RCS09A and MC-G-03, share no efflux-related genes in their annotations matching the query strings.

5.6.3 RND Efflux Related genes shared between RCS09A, MC-G-03 and KT2440

The two most tolerant strains, RCS09A and MC-G-03, shared no genes containing the queried keywords (Figure 27). As MC-G-03 and RCS09A were the most tolerant strains, their respective database matches were examined further. RCS09A held 8 unique genes matching the queries, and MC-G-03 held 17 (Table *6*, Table *7*).

Table 6 – RCS09A unique RND Efflux matches from Roary Database, **BOLD** genes belong to a complete operon, and **GREEN** Operons were able to have their structure predicted. **Orange** was able to have its structure predicted, but the morphology was unusual compared to other RND efflux pumps. **RED** no template was available for structure prediction.

Gene	Non-unique Gene name	Annotation	RCS09A ID	Structure
ttgC_1	**RCS09A_RND2**	**putative efflux pump outer membrane protein TtgC**	**IGPFHAFJ_00548**	**Yes**
oqxB17		**multidrug efflux RND transporter permease subunit OqxB17**	**IGPFHAFJ_00550**	**Yes**
mdtA_1	**RCS09A_RND3**	**Multidrug resistance protein MdtA**	**IGPFHAFJ_00708**	**Unusual**
ttgI_1		Toluene efflux pump outer membrane protein TtgI	IGPFHAFJ_00711	N/A
mdtA_3	**RCS09A_RND4**	**Multidrug resistance protein MdtA**	**IGPFHAFJ_03794**	**No**
ttgF_2		Toluene efflux pump outer membrane protein TtgF	IGPFHAFJ_04609	N/A
ttgB_1	**RCS09A _RND1**	**putative efflux pump membrane transporter TtgB**	**IGPFHAFJ_01473**	**Yes**
ttgA_1		**Toluene efflux pump periplasmic linker protein TtgA**	**IGPFHAFJ_01470**	**Yes**

5.6.4 RND Efflux Matches from Roary database that are Unique to RCS09A

RCS09A had eight unique matches with the queries for RND efflux genes within the Roary pangenome database. Four of these came from complete RND Efflux operons. Three out of these four operons were able to have their structures predicted using the SWISS model. RCS09A_RND3 predicted structures B subunit appeared to be incomplete (Table 6). RCS09A_RND4 had no template for the SWISS Model to thread the sequence to. The structures were saved to be analysed in the next chapter. The unique matches of MC-G-03 were analysed.

Table 7 – MC-G-03 unique RND Efflux matches from Roary Database.

Gene	Non-unique Gene name	Annotation	MC_G_03 ID
ttgF		Toluene efflux pump outer membrane protein TtgF	BMHMEFFB_04303
mdtA_4		Multidrug resistance protein MdtA	BMHMEFFB_04398
mdtA_5		Multidrug resistance protein MdtA	BMHMEFFB_04399
ttgC		Toluene efflux pump outer membrane protein TtgC	BMHMEFFB_04400
ttgI		Toluene efflux pump outer membrane protein TtgI	BMHMEFFB_06032
mdtA_6		Multidrug resistance protein MdtA	BMHMEFFB_07418
mdtA_7		Multidrug resistance protein MdtA	BMHMEFFB_07498
mdtA_1		Multidrug resistance protein MdtA	BMHMEFFB_00706
mdtA_2		Multidrug resistance protein MdtA	BMHMEFFB_01055
ttgB		Toluene efflux pump membrane transporter TtgB	BMHMEFFB_01207
ttgA		Toluene efflux pump periplasmic linker protein TtgA	BMHMEFFB_01208
ttgR		HTH-type transcriptional regulator TtgR	BMHMEFFB_01209
arnD		putative 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase ArnD	BMHMEFFB_03199
mdtA_3		Multidrug resistance protein MdtA	BMHMEFFB_03778
mexA		Multidrug resistance protein MexA	BMHMEFFB_03807
mexB		Multidrug resistance protein MexB	BMHMEFFB_03809
ttgV		HTH-type transcriptional regulator TtgV	BMHMEFFB_03947

5.6.5 MC-G-03 Unique RND Efflux Matches from Roary database

MC-G-03 contained 17 unique matches for RND efflux-related genes within the Roary pangenome database. As MC-G-03 is not as tolerant as RCS09A to BMA, their structures were not predicted for future analysis but were included just for the record of Mitsubishi Chemical Group (Table 7).

This study aims to provide alternative production strains with increased tolerance to BMA. An alternative production strain would have to have no pathogenic traits, such as virulence factors. To determine if the environmental isolates held any virulence factors, they were analysed using Abricate and the Virulence Factor database. The opportunistic pathogen *P. aeruginosa* was included in the comparison as an example of a BMA-tolerant organism that would not be suitable for development into a production strain.

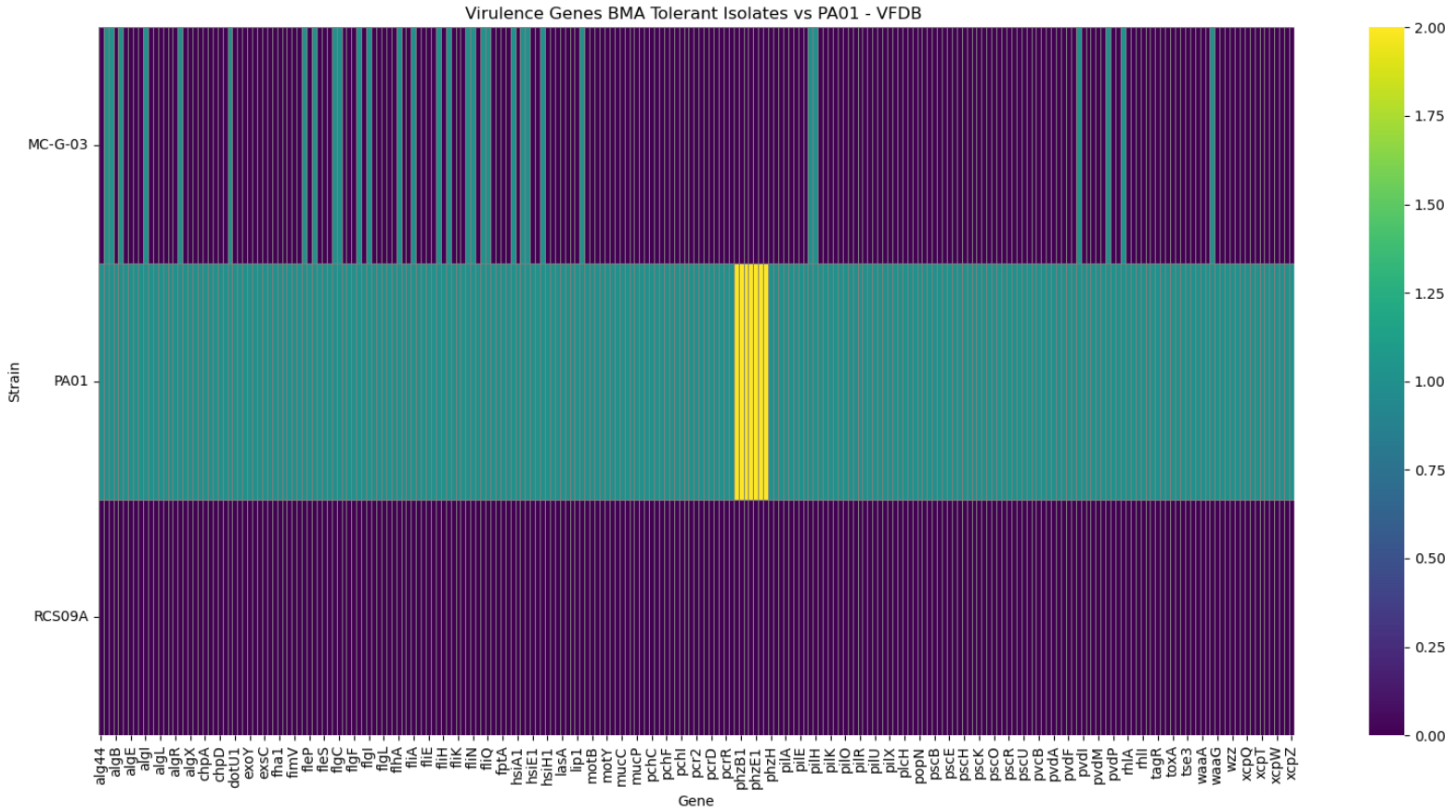


Figure 28 – Abricate VFDB results for MC-G-03, *P. aeruginosa* PA01 and RCS09A. RCS09A contained no virulence factors. MC-G-03 contained Virulence factors mainly contributing to motility, biofilm formation and iron sequestering

5.6.6 Abricate VFDB Analysis of RCS09A, MC-G-03 and PA01 Genomes Identified Virulence Factor Genes in MC-G-03 Genome

Analysis of virulence factors in the genomes of RCS09A performed using Abricate detected no virulence factors, further cementing its candidacy as an alternative production strain. MC-G-03 had 30 virulence genes, which, when compared to PA01 at 248, virulence is minimal (Figure 28). The products were produced in tabular format to further understand the virulence gene content of MC-G-03 (Table *8*).

Table 8 – Abricate Virulence Finder Database Results for MC-G-03 *P. veronii*

GENE	%COVERAGE	%IDENTITY	ACCESSION	PRODUCT
waaF	93.64	80.45	NP_253699	(waaF) heptosyltransferase I [LPS (VF0085)] [Pseudomonas aeruginosa PAO1]
algA	96.89	81.21	NP_252241	(algA) phosphomannose isomerase / guanosine 5'-diphospho-D-mannose pyrophosphorylase [Alginate (VF0091)] [Pseudomonas aeruginosa PAO1]
algI	82.21	84.98	NP_252238	(algI) alginate o-acetyltransferase AlgI [Alginate (VF0091)] [Pseudomonas aeruginosa PAO1]
alg8	97.71	82.43	NP_252231	(alg8) alginate-c5-mannuronan-epimerase AlgG [Alginate (VF0091)] [Pseudomonas aeruginosa PAO1]
algU	98.11	80.56	NP_249453	(algU) alginate biosynthesis protein AlgZ/FimS [Alginate (VF0091)] [Pseudomonas aeruginosa PAO1]
pvdO	90.18	80.23	NP_251085	(pvdO) pyoverdine biosynthesis protein PvdO [pyoverdine (IA001)] [Pseudomonas aeruginosa PAO1]
mbtH-like	93.15	85.29	NP_251102	(mbtH-like) MbtH-like protein from the pyoverdine cluster [pyoverdine (IA001)] [Pseudomonas aeruginosa PAO1]
pvdH	99.29	81.44	NP_251103	(pvdH) diaminobutyrate-2-oxoglutarate aminotransferase PvdH [pyoverdine (IA001)] [Pseudomonas aeruginosa PAO1]
pvdS	90.78	83.59	NP_251116	(pvdS) extracytoplasmic-function sigma-70 factor [Pyoverdine (VF0094)] [Pseudomonas aeruginosa PAO1]
fliA	91.53	80.12	NP_250146	(fliA) flagellar biosynthesis sigma factor FliA [Deoxyhexose linking sugar 209 Da capping structure (AI138)] [Pseudomonas aeruginosa PAO1]
fleN	95.85	82.84	NP_250145	(fleN) flagellar synthesis regulator FleN [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
flhA	99.44	81.59	NP_250143	(flhA) flagellar biosynthesis protein FlhA [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliQ	100	82.22	NP_250138	(fliQ) flagellar biosynthetic protein FliQ [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliP	89.71	83.41	NP_250137	(fliP) flagellar biosynthetic protein FliP [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliN	93.04	81.28	NP_250135	(fliN) flagellar motor switch protein FliN [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliM	96.71	86.73	NP_250134	(fliM) flagellar motor switch protein FliM [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliI	96.76	83.13	NP_249795	(fliI) flagellum-specific ATP synthase FliI [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fliG	99.71	84.79	NP_249793	(fliG) flagellar motor switch protein G [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
fleQ	99.93	81.99	NP_249788	(fleQ) transcriptional regulator FleQ [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
flgI	94.86	84.24	NP_249775	(flgI) flagellar P-ring protein precursor FlgI [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
flgG	98.35	80.82	NP_249773	(flgG) flagellar basal-body rod protein FlgG [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
flgC	99.77	80.9	NP_249769	(flgC) flagellar basal-body rod protein FlgC [Flagella (VF0273)] [Pseudomonas aeruginosa PAO1]
flgB	99.51	80	NP_249768	(flgB) flagellar basal body rod protein FlgB [Deoxyhexose linking sugar 209 Da capping structure (AI138)] [Pseudomonas aeruginosa PAO1]
pilH	90.98	82.58	NP_249100	(pilH) twitching motility protein PilH [Type IV pili (VF0082)] [Pseudomonas aeruginosa PAO1]
pilG	96.81	84.05	NP_249099	(pilG) twitching motility protein PilG [Type IV pili (VF0082)] [Pseudomonas aeruginosa PAO1]
algC	94.76	80.34	NP_254009	(algC) phosphomannomutase AlgC [Alginate biosynthesis (CVF522)] [Pseudomonas aeruginosa PAO1]
hsiB1/vipA	91.33	87.55	NP_248773	(hsiB1/vipA) type VI secretion system tubule-forming protein VipA [HSI-I (VF0334)] [Pseudomonas aeruginosa PAO1]
hsiC1/vipB	96.46	89.14	NP_248774	(hsiC1/vipB) type VI secretion system tubule-forming protein VipB [HSI-I (VF0334)] [Pseudomonas aeruginosa PAO1]
hcp1	95.91	82.38	NP_248775	(hcp1) type VI secretion system substrate Hcp1 [HSI-I (VF0334)] [Pseudomonas aeruginosa PAO1]
hsiG1	99.57	80.75	NP_248778	(hsiG1) type VI secretion system hcp secretion island protein HsiG1 [HSI-I (VF0334)] [Pseudomonas aeruginosa PAO1]
clpV1	97.42	84.65	NP_248780	(clpV1) type VI secretion system AAA+ family ATPase [HSI-I (VF0334)] [Pseudomonas aeruginosa PAO1]

5.6.7 MC-G-03 Virulence Gene Products

Fourteen of the thirty Virulence Factor Genes (VFG) were dedicated to motility and flagella products (Table 8 - Orange). Five VFGs were involved in a type VI secretion system (T6SS) (Table 8 - Purple). Four VFGs were involved in pyoverdine synthesis (Table 8 - Green). Five VFGS were involved in alginate synthesis (Table 8 - Red). Although the flagella and alginate genes may not be indicators of virulence they do contribute to biofilm formation, which is not ideal when considering the strain's applications in a bioreactor environment (Ozer et al., 2021; Valadbeigi et al., 2023). In *P. aeruginosa*, pyoverdine is a virulence determinant. It is used in the acquisition of iron in iron-poor environments. It has been demonstrated that the pyoverdine produced by *P. veronii* is structurally unique, suggesting that this may be an accurate VFG assignment (Meyer et al., 2002, Kang et al., 2019). *P. veronii* has limited research on its use for T6SS, as the organism is generally not considered a pathogen, and the percentage identity is lower than 90% compared to the VFGs of PA01. To further assess their potential as alternative production strains Abricate was used with the MegaRes database, which searches for antimicrobial resistance genes.

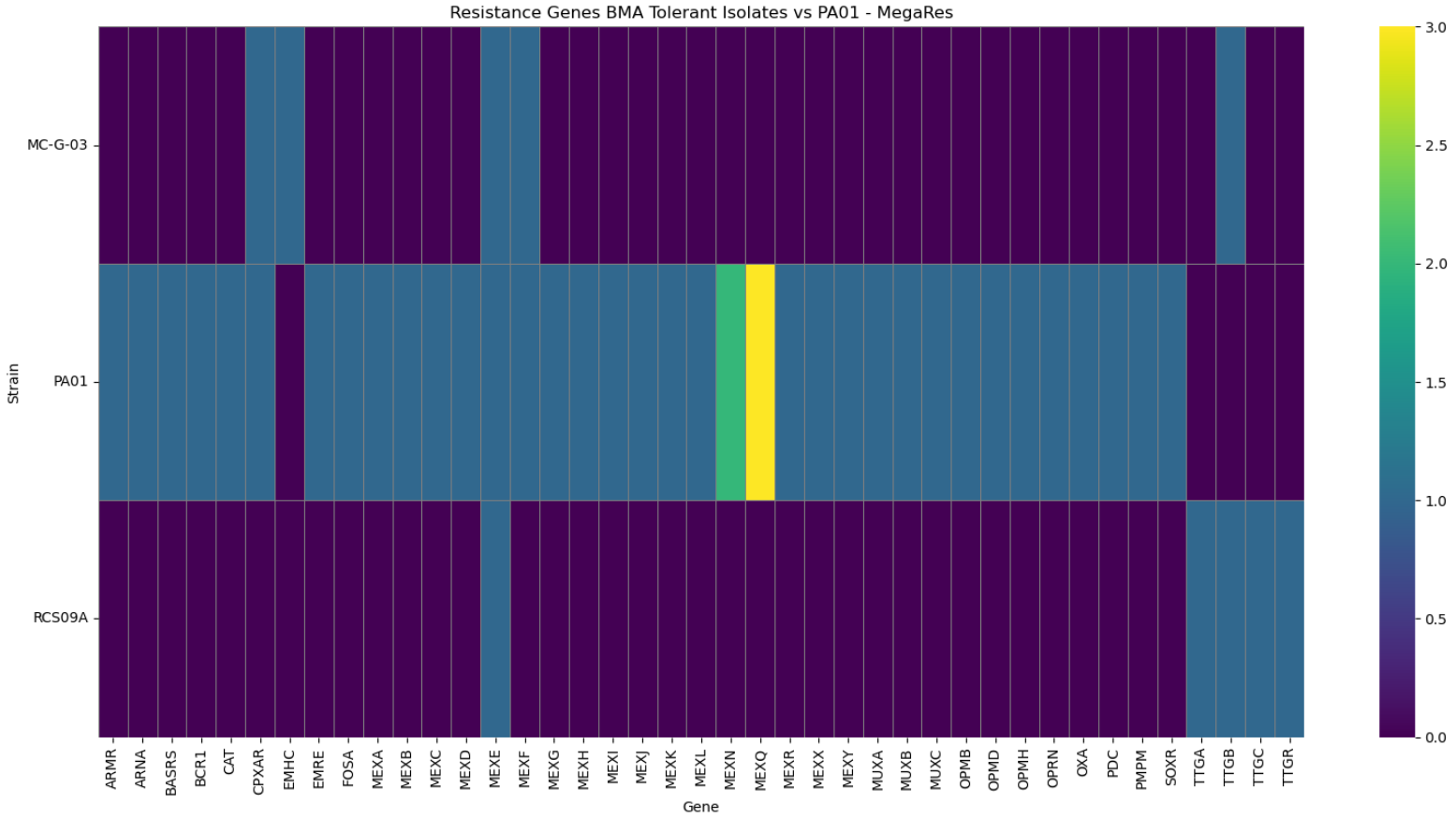


Figure 29 – Abricate MegaRes Results Antimicrobial resistance genes held by MC-G-03, PA01 and RCS09A. RCS09A only held RND efflux pumps which also confer solvent tolerance. MC-G-03 held a *cat* chloramphenicol resistance gene as well as RND efflux genes.

5.6.8 Abricate MegaRes Resistance Gene Analysis

The resistance genes detected by the Abricate MegaRes Analysis conducted on RCS09A were efflux pumps instead of typical antimicrobial resistance genes such as a beta-lactamase gene. The *ttgABC* operon was detected in RCS09A, which has been shown to confer solvent tolerance (Charles Begley – personal communication - Tucker Lab). A chloramphenicol acetyltransferase (*cat*) gene, which confers tolerance to the antibiotic chloramphenicol, was detected in the chromosome of MC-G-03 (Shaw et al., 1979). The RND *EMHC* and *mexEF* were also detected. As it has been shown that RND efflux contributes to solvent tolerance, these resistance genes cannot be discounted as also contributing to solvent tolerance profile without further investigation. *Pseudomonas* species are known to carry more RND pumps than other organisms. *E. coli* typically has 5-7 RND efflux systems compared to *P. aeruginosa*, which typically holds 12-16 (Adamiak et al., 2021; Pugh et al., 2024). As RND efflux pumps have been shown to confer solvent tolerance, the genomes of the BMA tolerant isolates and KT2440 were analysed using BLASTN and *mexAB-orpM*, *ttgABC* and *ttgDEF* as a query to quantify the number of genes that matched give an approximate number of RND efflux systems on the chromosome not including extrachromosomal DNA.

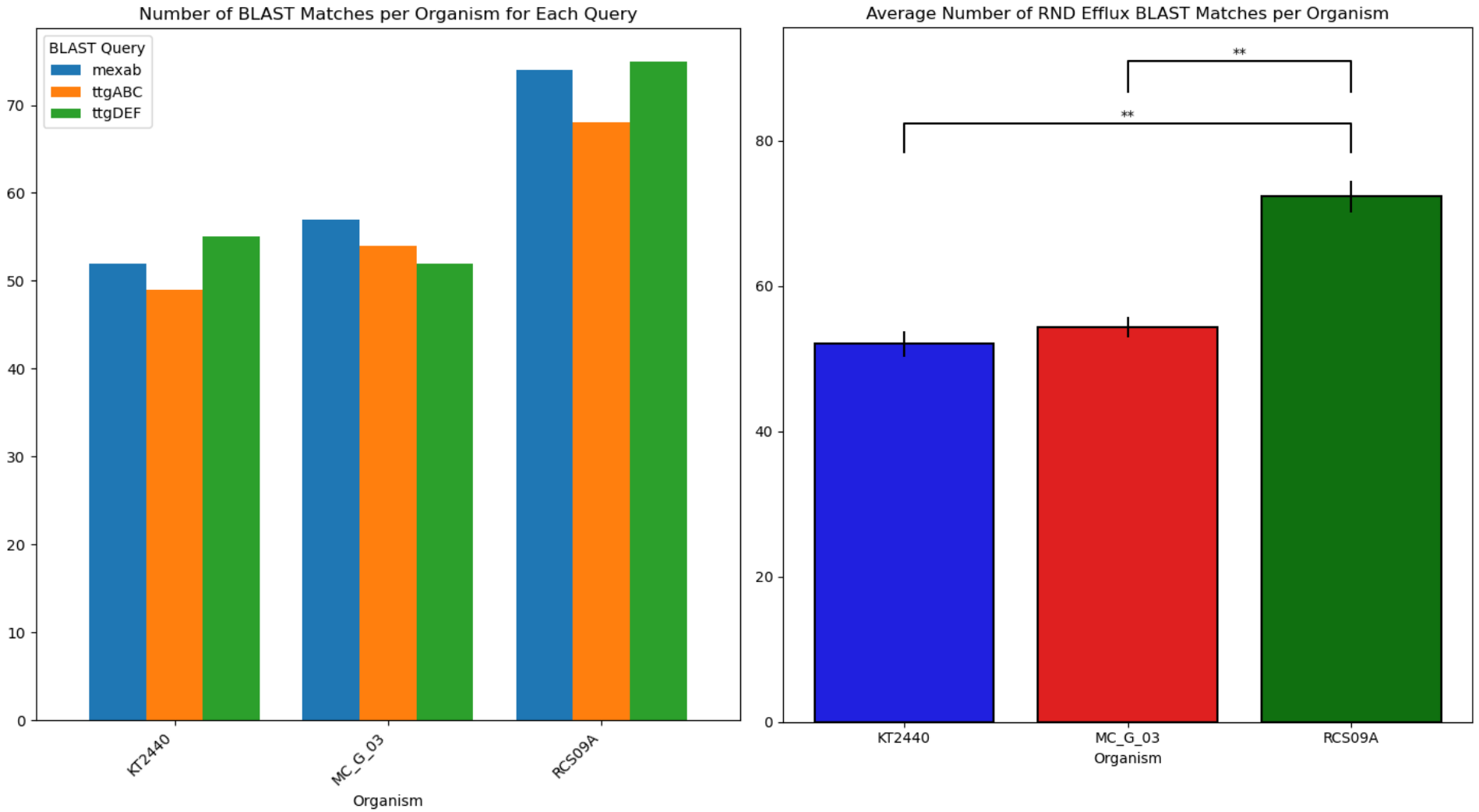


Figure 30 – BLASTN matches for *mexAB-orpM*, *ttgABC* and *ttgDEF* on the chromosome of *P. putida* KT2440, RCS09A and MC-G-03. A one-way ANOVA showed variance in the mean BLAST matches, and a Tukey HSD showed RCS09A to have significantly more RND efflux pump-like sequences than KT2440 and MC-G-03 (P-Value = 0.01). This shows the BLAST matches of three genes per query, so this should not be interpreted as complete operon counts.

5.6.9 BLASTN Matches of *mexAB-orpM*, *ttgABC* and *ttgDEF* on the chromosome of RCS09A, MC-G-03 and KT2440

RCS09A has an average of 72.3 RND efflux pump BLAST matches on its chromosome. This is significantly higher than the BLAST matches for RND efflux pumps on MC-G-03, 54.3, and KT2440, 52 (Figure 30). These are matches for the three separate genes that are used to make up the larger pump structure. This suggests that RCS09A has a more extensive repertoire of pumps for various conditions that it can utilise. RCS09A also had the highest amount of genes matching to the *ttgDEF* RND pump, which has been shown to confer toluene tolerance in *P. putida* a compound with similar hydrophobic properties to BMA (Duque et al., 2001).

5.7 Discussion

The metagenomic study of the RCS09 drain sample showed a dramatic population shift in R-2A treated with BMA to *Pseudomonas* species. Both BMA treated samples saw a shift to the majority of the reads being assigned to *P. putida* compared to the control group which was not treated with BMA. In contrast, the Styrene and Ethyl Benzene groups did not see a shift favouring one organism. From the RCS09 BMA sample, we were able to isolate three significantly more tolerant organisms when compared to *P. putida* KT2440: RCS09A, RCS09A(2) and RCS09A(3). Due to RCS09A(2) and RCS09A(3) being isolated towards the end of the study, we could not characterise these isolates further. RCS09A was characterised as *Pseudomonas putida*, which shared a high nucleotide identity with the strain DZ-C20 isolated in China. RCS09A was significantly more tolerant when compared to RCS09A(2), RCS09A(3) and KT2440. This confirmed that by targeting a solvent-rich environment, you can isolate solvent-tolerant organisms as other less industrious environments failed to yield any tolerant organism.

Once travel restrictions were lifted, the Mitsubishi production facility was sampled, but the samples were not viable for yielding solvent-tolerant organisms by directly culturing them with BMA, as with the Royal Conservatoire samples. They also failed to produce DNA for 16S or whole metagenomic studies. The samples were instead cultured with R-2A, and

these culture-enriched samples were used for 16S metagenomics. The sample sites that contained the 16S genes of the genus of interest, such as Pseudomonas, were plated onto *Pseudomonas* isolation agar and individually challenged with BMA. Site G was the only sample to yield a BMA-tolerant organism, MC-G-03, which was subsequently characterised as *P.* veronii. MC-G-03 was significantly more tolerant when compared to KT2440.

To compare each isolate with each other, we used their percentage increase in CFU when compared to KT2440 as a benchmark. This showed that RCS09A was the most BMA tolerant of the four isolates investigated.

To understand if any shared genes could contribute to solvent tolerance, Roary was used to analyse the pangenome of RCS09A, MC-G-03 and KT2440. The pangenome of all three isolates was 566 genes, none of which were related to RND efflux systems. The Roary results shared genes database was then filtered using keywords related to RND efflux systems that were noted to contribute to solvent tolerance. The two more tolerant isolates shared no matches. KT2440 and RCS09A shared 14 genes, but they were not investigated as they are both *P. putida*. The unique genes of RCS09A had their structures predicted and were analysed in work described in the next chapter, which focuses on RND efflux systems.

To further understand their suitability as production strains, we compared their virulence factors to *P. aeruginosa* PA01, a known opportunistic pathogen, using Abricate. RCS09A contained no virulence genes catalogued in the VF database. MC-G-03 contained 30 virulence genes included in the VF database. Most of these genes contributed to motility, biofilm formation and a T6SS. Although motility and biofilm are not significant causes for concern in *P. veronii*, which is not considered a pathogen, they present some causes for concern when considering its candidacy as an alternative production strain, as biofilm formation inside a bioreactor would not be ideal. A biofilm can disrupt the cleaning and moving parts of a reactor and clog pipelines (Horwood et al., 2013).

Isolates were screened for resistance genes using Abricate and the MegaRes database to further assess their potential as alternative production strains. RCS09A contained no resistance genes in the MegaRes database aside from RND efflux pumps, which are also known to confer solvent tolerance. MC-G-03 also contained RND efflux genes and a *cat* gene that confers chloramphenicol resistance (Shaw et al., 1979). This also supports the notion that MC-G-03 would not be a suitable alternative production strain as it could contribute to the resistome of plant workers.

Pseudomonas species are known to have a larger repertoire of RND efflux pumps when compared to other bacterial species, such as *E. coli* (Adamiak et al., 2021; Pugh et al., 2024). BLASTN was used with *mexAB-*

orpM, *ttgABC* and *ttgDEF* as queries to get an approximate number of genes that matched the three genes in these systems on the chromosomes of RCS09A, MC-G-03 and KT2440. A one-way ANOVA was used to detect a variance in the mean result from all three systems per chromosome, and a Tukey HSD showed that RCS09A had significantly more matches than KT2440 and MC-G-03. This result suggests that RCS09A has a more extensive repertoire of RND efflux pumps than the *P. putida* KT2440 reference strain and *P. veronii* MC-G-03. As RCS09A was isolated from an environment with a highly varied solvent and chemical contamination, this may be why it has evolved/acquired this extensive RND efflux gene collection.

The initial metagenomic study on the RCS09 sample treated with BMA contained reads that assembled into a plasmid-like contig containing an RND efflux operon. This was then confirmed through WGS of RCS09A. Further investigation and analysis of the plasmid are described in the next chapter.

In conclusion, we have successfully isolated four BMA-tolerant organisms, one of which, RCS09A, is significantly more tolerant than the others. It contained no virulence genes and no resistance genes of concern. It also contained significantly more RND efflux pumps than the other organisms studied. These qualities mark RCS09A as a great alternative BMA production strain.

5.8 Future Work

To further assess RCS09A(2) and RCS09A(3) as alternative BMA production strains, their genomes must be sequenced to undergo the same studies as the other isolates.

The initial scope of this study included RNA-Seq of BMA-tolerant organisms under BMA stress conditions versus normal conditions to understand the transcriptional response to the solvent. The original plan was to outsource the sequencing, but the external company's purity specifications were arduous to meet due to the solvent component of the experiment. An in-house RNA-Seq would be the best alternative to this, assuming useful sequence data could be obtained with lower purity. MC-G-03 would be an ideal candidate as it may have a particular RND efflux pump that upregulates, which has evolved in a highly BMA-saturated environment. This pump could then be placed onto a plasmid to determine if high copy number could be used to increase tolerance.

6 Characterisation of the Efflux Operon and Engineering *Pseudomonas putida* for Enhanced Tolerance

6.1 Introduction

Efflux pumps play a vital role in solvent tolerance in microorganisms and have an advantageous application in biotechnology (X. Z. Li et al., 1998b). They confer tolerance by removing toxic molecules from the cell cytosol into the extracellular environment rather than by modifying or degrading the toxin. In an industrial fermentation context, they are also capable of exporting synthesised products 'unchanged' and ready for purification and downstream processing. *Pseudomonas* species are known for their ability to thrive in solvent-rich, contaminated environments and have been found in many hydrocarbon-contaminated sites (Medić & Karadžić, 2022). *P. putida* mediates solvent tolerance through a multilevel response, including efflux. The core *P. putida* genome contains 20 Resistance Nodulation Division (RND) efflux pumps (Ramos et al., 2015; Udaondo et al., 2013). RND efflux pumps span the periplasm and provide cross-membrane removal of compounds from the cytosol into the environment. They confer tolerance to different compounds, e.g. the MexAB-OrpM RND pump of *P. aeruginosa* acts on over 54 antibiotics (Carrara et al., 2022b). The disruption of two RND efflux systems, *ttgABC* and *srpABC*, in *P. putida* B6-2, resulted in increased susceptibility to p-xylene, showing the importance

of SrpABC in conferring tolerance to this solvent (Yao et al., 2017). Similarly, the efflux system *ttgDEF* is expressed at high levels in *P. putida* DOT-T1E treated with toluene: while high levels of TtgR repress the expression of *ttgABC* in the absence of inducer. When compared to DOT-T1E not treated with toluene, *ttgABC* is transcribed at higher levels, suggesting a switch to the TtgDEF system for the efflux of toluene and further evidence of pump specificity of compounds (Duque et al., 2001).

For efflux systems conferring tolerance to BMA, we can use selective screening to isolate BMA-tolerant organisms from environmental samples taken from sites selected based on exposure to solvents and hazardous compounds. For example, a toluene-tolerant *P. putida* T-57 was isolated from an activated sludge sample from a wastewater treatment plant in Kumamoto, Japan. The sample was challenged with toluene for one week, providing the selective screening to isolate the tolerant organism (Faizal et al., 2005a).

In this study, we challenged environmental samples taken from two sites; i) The Royal Conservatoire of Glasgow set design department where multiple paints, solvents, heavy metals, adhesives and other compounds are enriched in the environment; and ii) A Mitsubishi Chemical Methacrylate Production Facility where methacrylate esters, and more specifically BMA, are manufactured. We treated these samples with BMA to provide selective screening to isolate organisms with BMA tolerance.

As described in the previous chapter, four BMA-tolerant organisms were isolated, and their tolerance to BMA was compared using *P. putida* KT2440 as a benchmark. A Royal Conservatoire isolate, RCS09A, was characterised as having a similar sequence identity to *Pseudomonas putida* DZ-C20 and was the most tolerant isolate found. This strain was identified in the initial Metagenomic study of the Royal Conservatoire sample (i.e., sample RCS09), it was then isolated by culturing in the laboratory and identified and further analysed through Whole Genome sequencing (WGS). The metagenomic data identified the presence of a plasmid encoding an RND efflux system in the mixed RCS09 sample. The WGS of the isolated RCS09A strain confirmed that this was the organism carrying the plasmid. The plasmid sequence was analysed and validated using BLAST against the NCBI database. It returned a match for another *Pseudomonas* plasmid, RW109, which is associated with an extreme solvent-tolerant stress response phenotype (Weiser et al., 2019a). The aligned sequences were the RND efflux systems. Due to RCS09A being the most BMA-tolerant isolate and the RW109 plasmid being associated with solvent tolerance, the efflux system encoding the RCS09A plasmid was investigated further.

This Chapter will present the results of metagenomic analysis of the RCS09 sample treated with 50% v/v BMA, and WGS of the cultured isolate RCS09A. The RND efflux pump associated with the plasmid of strain

RCS09A was further characterised and confirmed to contribute to conferring tolerance to BMA. The RCS09A RND efflux pump has a number of characteristics which might be used for identifying other RND systems responsible for efflux of hydrophobic solvents. To understand this the fpocket protein pocket was used to analyse the internal pockets of the RCS09A plasmid efflux system and other pockets known to confer solvent tolerance as well as more broadly acting pumps. To determine if there is metric that can be used to differentiate between pumps suited to solvents versus broadly acting efflux systems.

6.2 Identification of Efflux Operon from Metagenomic Assembly Data

Using the metagenomic reads from sample RCS09 treated with BMA (50% v/v) assemblies were made using the SPAdes metagenomic function (Nurk et al., 2017). Using a custom BLAST DNA database, these were screened for solvent-tolerance conferring systems (Figure 31). Assembly Node 63 was found to have a similar sequence to the *MexB* subunit of the *MexAB-OrpM* efflux system, known to confer BMA tolerance in *P. aeruginosa* (Personal Communication. Dr Terras – Tucker Group). To further understand Node 63, it was annotated using Prokka (Seemann, 2014).

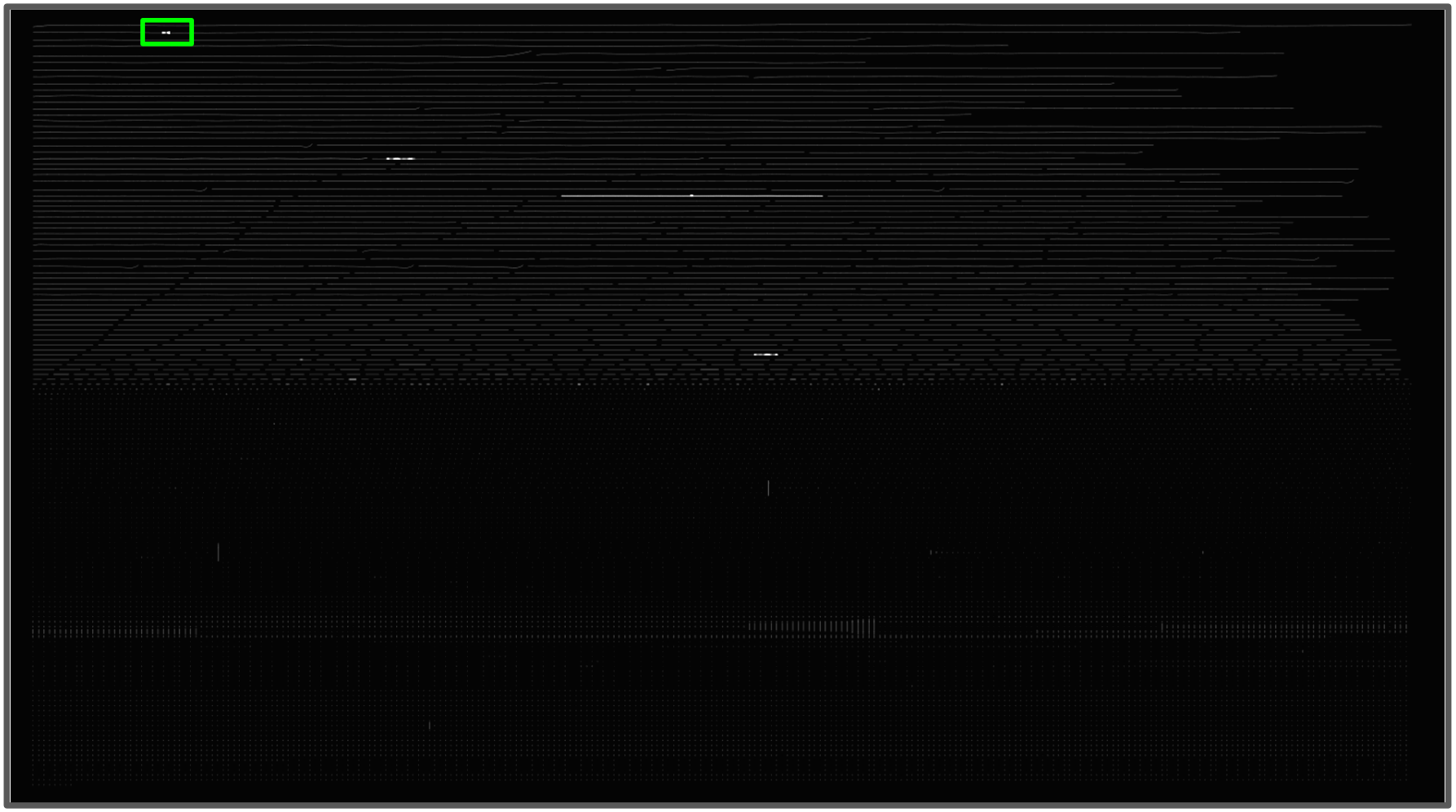


Figure 31 – RCS09 BMA 50% v/v (3.12 M) Metagenomic reads assembled using the SPAdes Metagenomic Function BLAST results visualised using Bandage. Highlighted in the green box is the contig/node that aligned with the *MexB* subunit of the *mexAB-orpM* system.

6.2.1 Prokka Annotation of Node 63 - found to contain a complete RND efflux System contained on a plasmid-like DNA sequence

The Prokka annotation of Node 63 contained a complete operon, characteristic of RND efflux systems comprising the classic three-component system and a transcriptional regulator sequence which could be the main component conferring solvent tolerance. The annotation also included a toxin-antitoxin system VapBC used to maintain plasmid lineage (Figure 32, Pink Gene)and a transfer gene *traD* (Figure 32, Orange Gene), indicating that the assembled sequence could be a transferrable plasmid (Figure 32) (Muñoz-Gómez et al., 2005). The sequence was then queried against the NCBI BLASTN database.

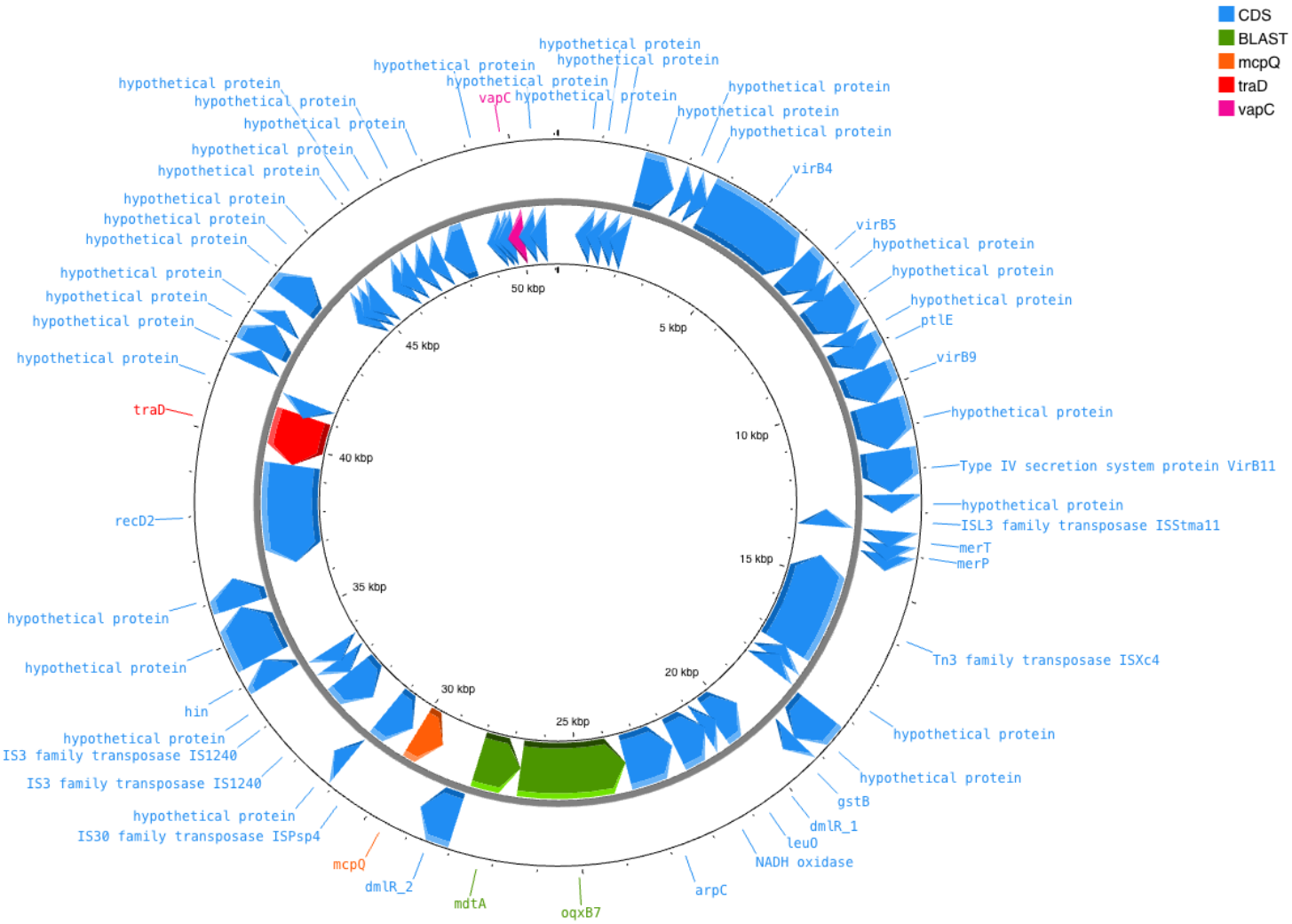


Figure 32 – Node 63 of RCS09 BMA treated 50% 3.12 M Metagenomic Read Assembly. Prokka Annotation (**GREEN**) RND efflux operon B and A subunits and BLAST Match to RW109 plasmid. (**RED**) *traD* plasmid transfer gene. (**PINK**) *vapC* of the vapBC toxin-antitoxin system which is used to maintain plasmid lineage.

6.2.2 BLAST Search Reveals a match to a plasmid associated with solvent tolerance

The BLAST result for Node 63 matched the *P. aeruginosa* plasmid RW109 which is associated with an extreme solvent-tolerance stress response (Weiser et al., 2019c). The primary sequences that aligned were the regions containing the RND efflux operons shown in green in the upper track of Figure 33. The RW109 plasmid is much larger than Node 63 (i.e., 550Kbp compared to 51Kbp for Node 63). The plasmid corresponding to Node 63 was screened using resfinder to investigate if it contained any other RND efflux pumps, but none were found, suggesting the RND efflux pump in question was the primary reason for the solvent tolerance observed.

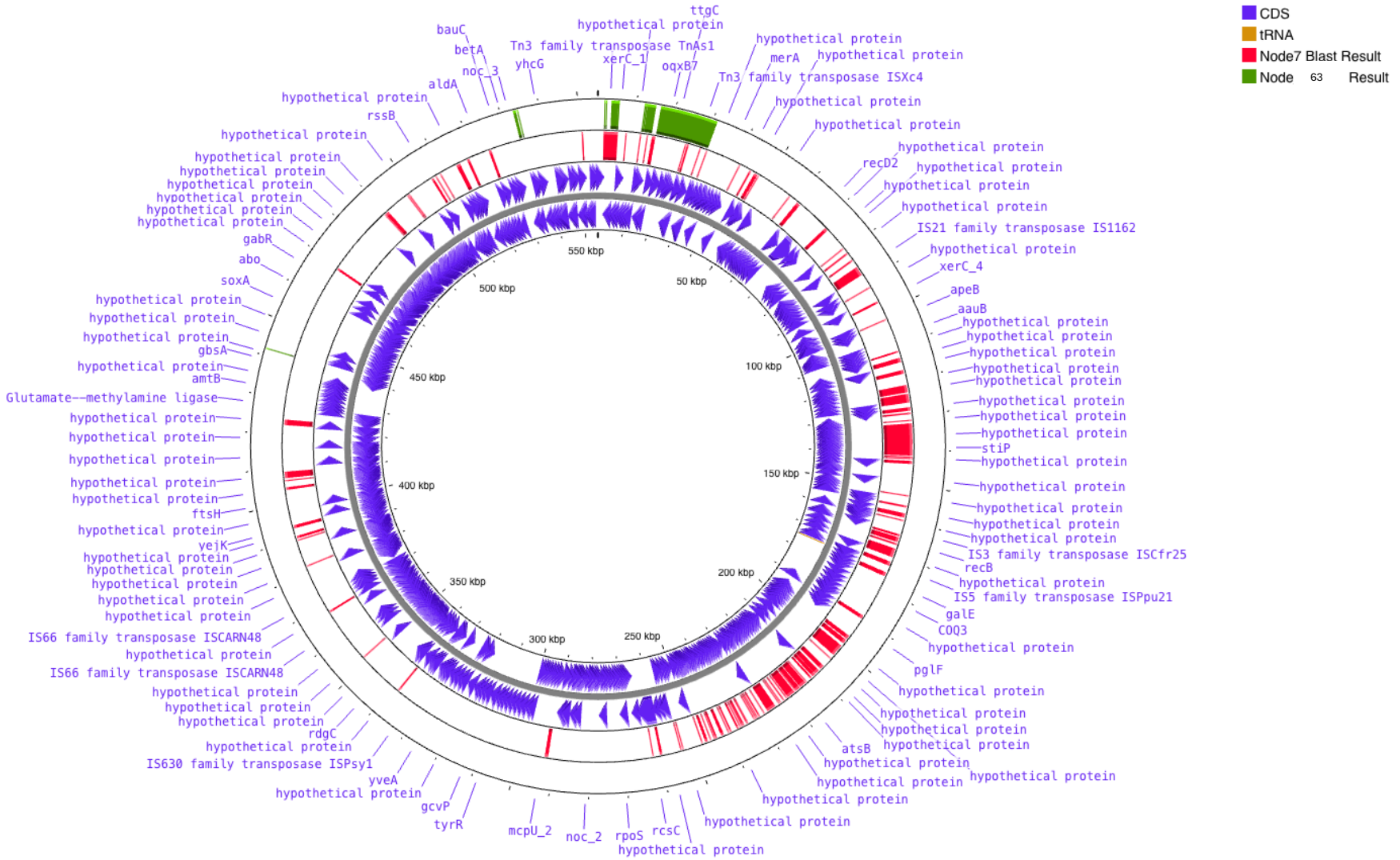


Figure 33 - Plasmid RW109 of *P. aeruginosa* associated with an extreme solvent tolerant phenotype. (**GREEN TRACK**) The BLAST alignment of RCS09A Node63 of RCS09 BMA 50% v/v treatment Metagenomic assembly data

6.2.3 WGS of RCS09A Assembly Contains the Plasmid Identified in the Metagenomic Study

The isolate RCS09A was sequenced using a combination of Nanopore and Illumina sequencing technologies to obtain a high-quality complete genome assembly. The organism was identified as *P. putida*. The sequence assemblies confirmed that the isolate contained a circular plasmid like sequence with an RND efflux pump; although the WGS plasmid was slightly larger than Node 63 (Figure 34, alignment shown in green). The annotated plasmid was named pPp65-MOA (for **p**lasmid ***P. putida*** **65**kb **m**dtA-**o**qxB7-**a**rpC) for its plasmid size and efflux system. To assess the differences between the RCS09A pPp65-MOA efflux pump structure and other RND efflux pumps its structure was predicted using the Swiss model. The Swiss model identified MexAB-OrpM as a guide structure as this was the original BLAST match, and this seemed an appropriate guide. The three-dimensional structures were compared using PyMol to observe topographical differences.

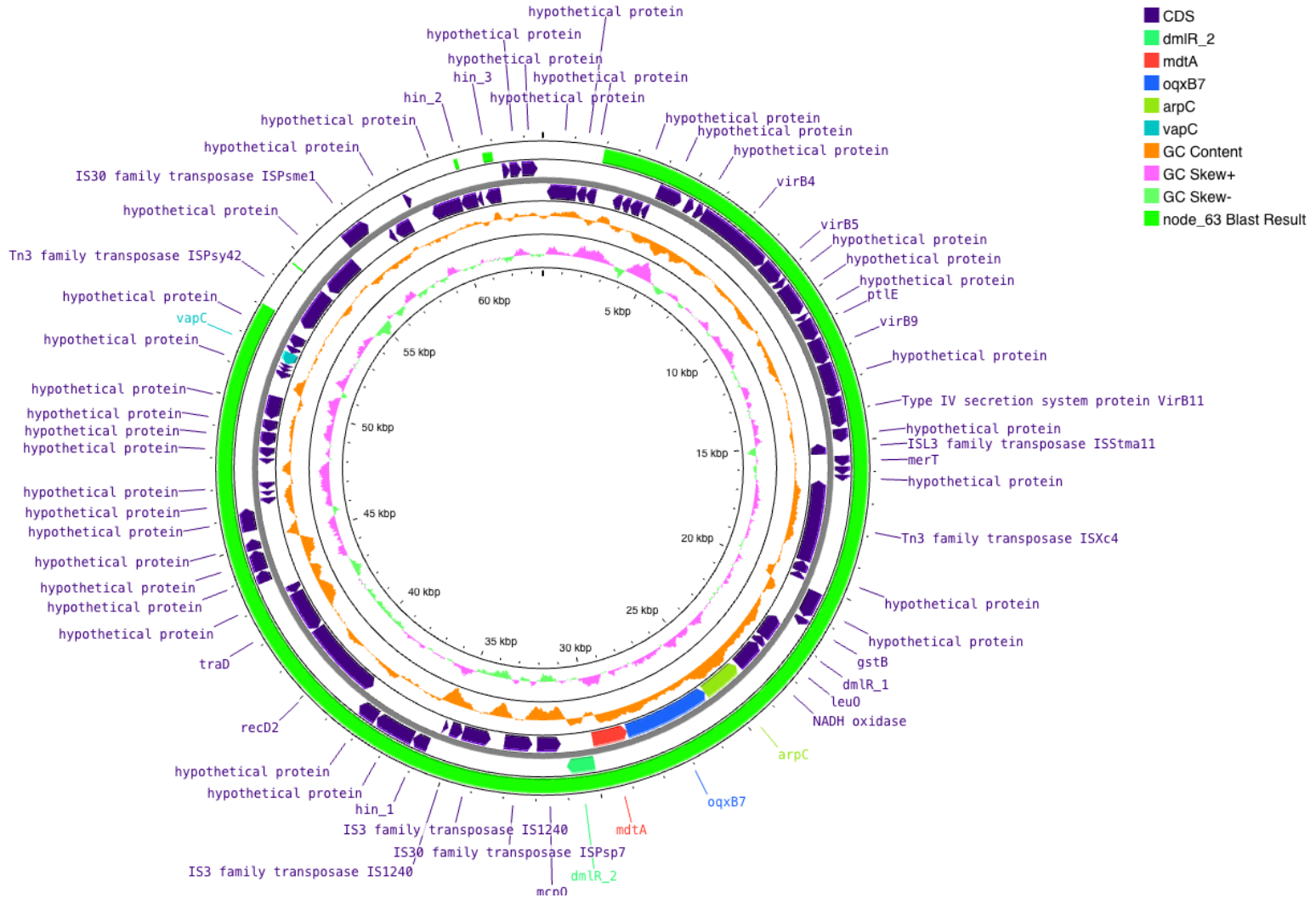


Figure 34 – WGS of RCS09A Plasmid Assembly with RND efflux pump annotated (*mdtA*, *oqxB7*, *arpC*) and the Node63 Metagenomic Assembly alligned in green

6.3 RCS09A RND Efflux mdtA-oqxB7-arpC System

The aligned efflux systems MexAB-OprM and mdtA-oqxB7-arpC (MdtA) were overlayed and viewed in cross-sections to compare the internal structures. The MdtA pump is coloured green, and the Mex pump in red. The MdtA pump has a narrow channel and a ridge above the entry aperture (Figure 35 green box). The MdtA pump also has more open pores at the base of the pump (Figure 35 red box). Both structures can also be viewed independently in Figure 36. The ridge in the channel is a unique feature in all examined RND efflux systems that confer solvent tolerance (TtgDEF, TtgABC, MexAB-OrpM, AcrAB-TolC). To further understand the internal structure of the channels, the electrostatic properties were projected on the protein structures using the APBS electrostatics plugin and visualised using PyMol.

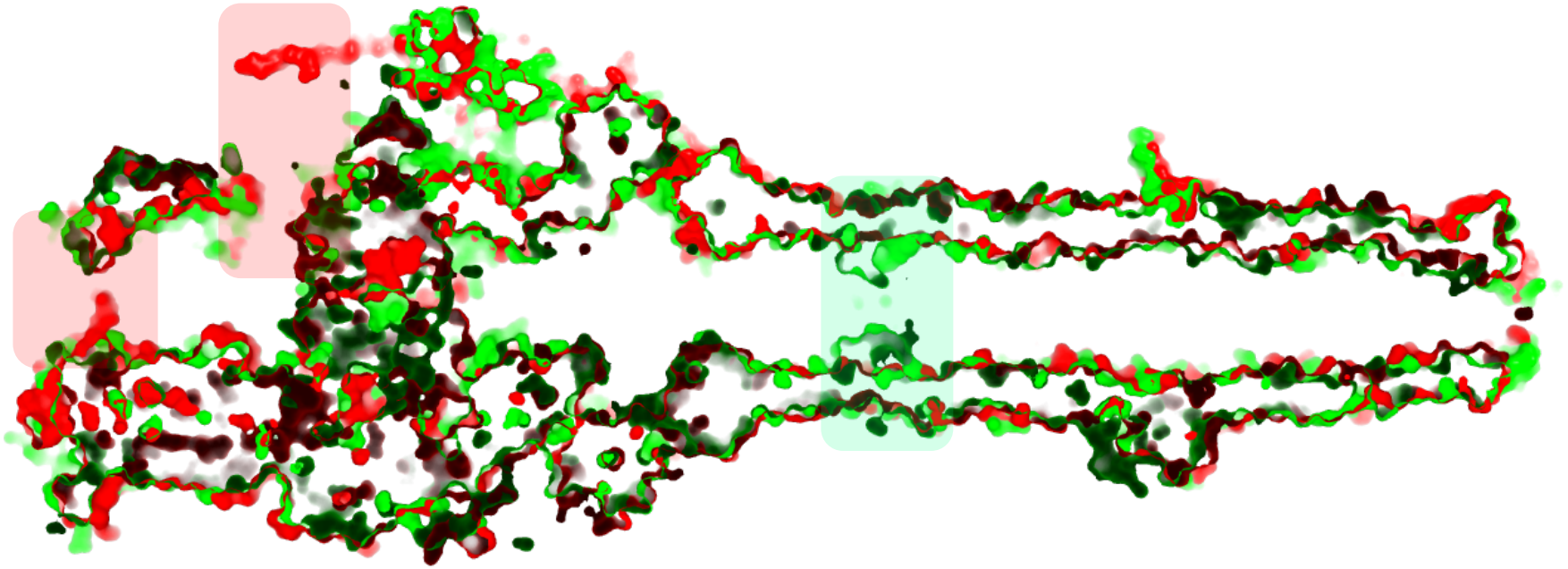


Figure 35 – PyMol Alignment of MOA (MdtA) Efflux System (**GREEN**) and MexAB-OrpM Efflux System (**RED**). The most notable topographical differences have been highlighted in (**GREEN BOX**) for regions unique to MOA (MdtA) and (**RED BOX**) for MexAB-OrpM

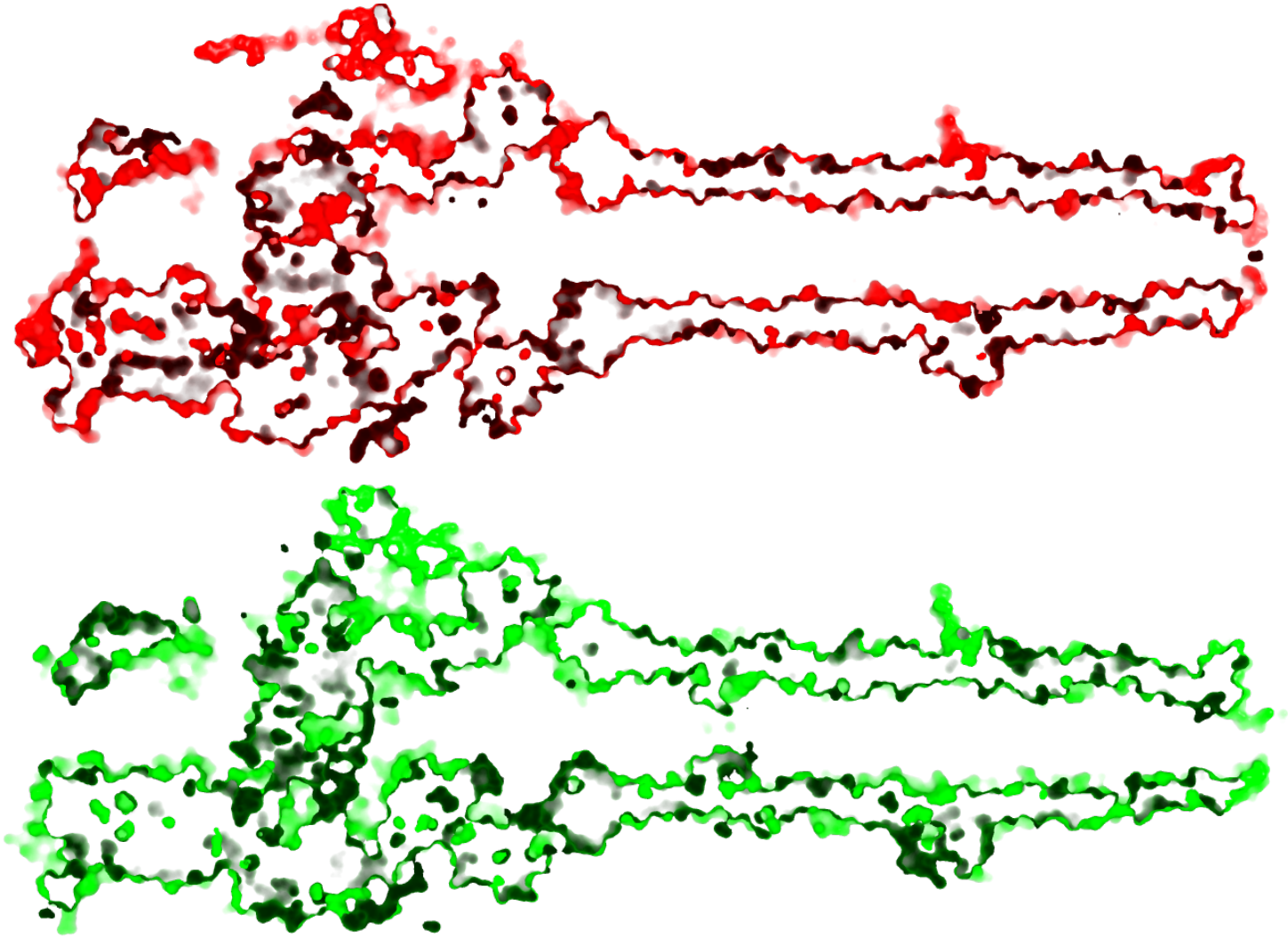


Figure 36 The 3D Pymol cross sections of MexAB-OrpM (**RED**) and the MOA (MdtA) (**GREEN**) efflux systems so the slightly narrower channel of MOA (MdtA) can be seen clearly.

6.3.1 Electrostatic properties of the mdtA and *MexAB-orpM* efflux pumps compared using PyMol.

The internal structures and electrostatic properties of the two efflux pumps are presented in Figure 35. The electrostatic motifs of the upper channels of the aligned structures are quite different. The MdtA channel appears to have a more varied motif of electrostatic change when compared to MexAB-orpM (Figure 37 – A, B), which is more positive and neutral (blue and white, respectively) when compared to the primarily negative channel of MexAB-OrpM (predominantly red) Figure 37 - B). The entry apertures to both channels are also differently charged. The MdtA aperture is positively charged with a neutral core (Figure 37 - C). The entry aperture of *MexAB-orpM* is negatively charged with a more discrete neutral core (Figure 37 - E). The internal biochemical environments of the MdtA pump and MexAB-orpM are electrostatically different, which could be linked to substrate specificity or substrate flexibility. The 54 substrates of MexAB-OrpM have been shown to have a mean XLogP3 of -0.538, suggesting that it is more suited to efflux of hydrophilic compounds (Carrara et al., 2022b).

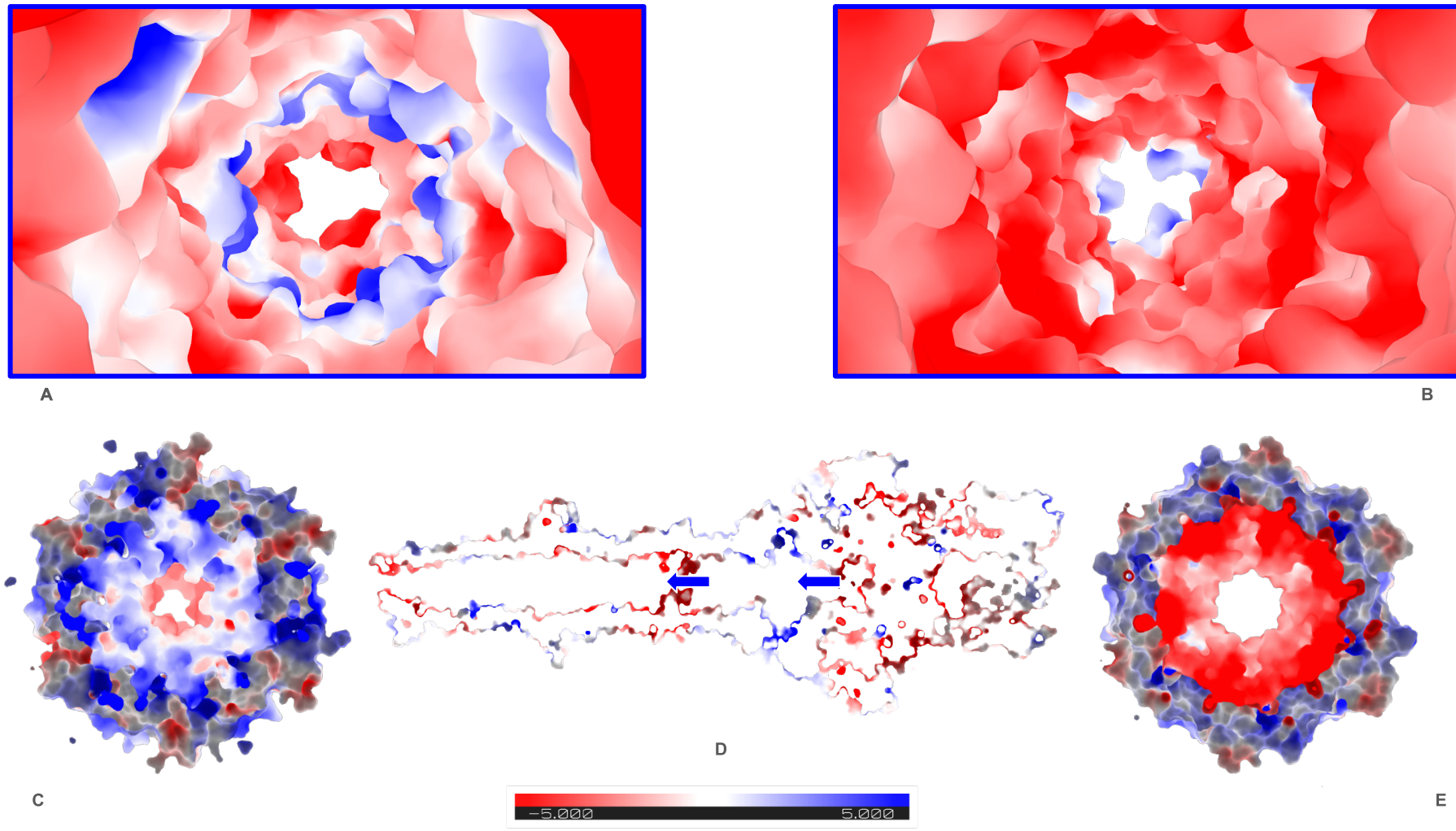


Figure 37 - APBS Electrostatic properties of the MdtA pump channel **(A)**. MexAB-OrpM channel **(B)**. The entry aperture of the MdtA pump **(C)**. The entry aperture MexAB-OrpM **(E)**. The cross-section of the MdtA pump with view arrows and electrostatic range key **(D)**.

6.3.2 Electrostatic properties of the MdtA and AcrAB-TolcC efflux pumps compared using PyMol

The internal electrostatic properties of AcrAB-TolC are similar to MexAB-OrpM. The channel and entry apertures are highly negative (Figure 38). A selection of the compounds that AcrAB-TolC is known to efflux (ivacaftor, Butenafine, Naftifine, Pimozide, Thioridazine, Trifluoperazine, Meloxicam, Ethidium Bromide, Novobiocin, Fluoroquinolones) have a mean XLogP3 of 4.35 which are hydrophobic (Pisoni et al., 2023; Sugawara & Nikaido, 2014). Although this cannot be directly compared to the compound data of MexAB-OrpM, 54 compounds were used in that examination versus 11 for AcrAB-TolC.

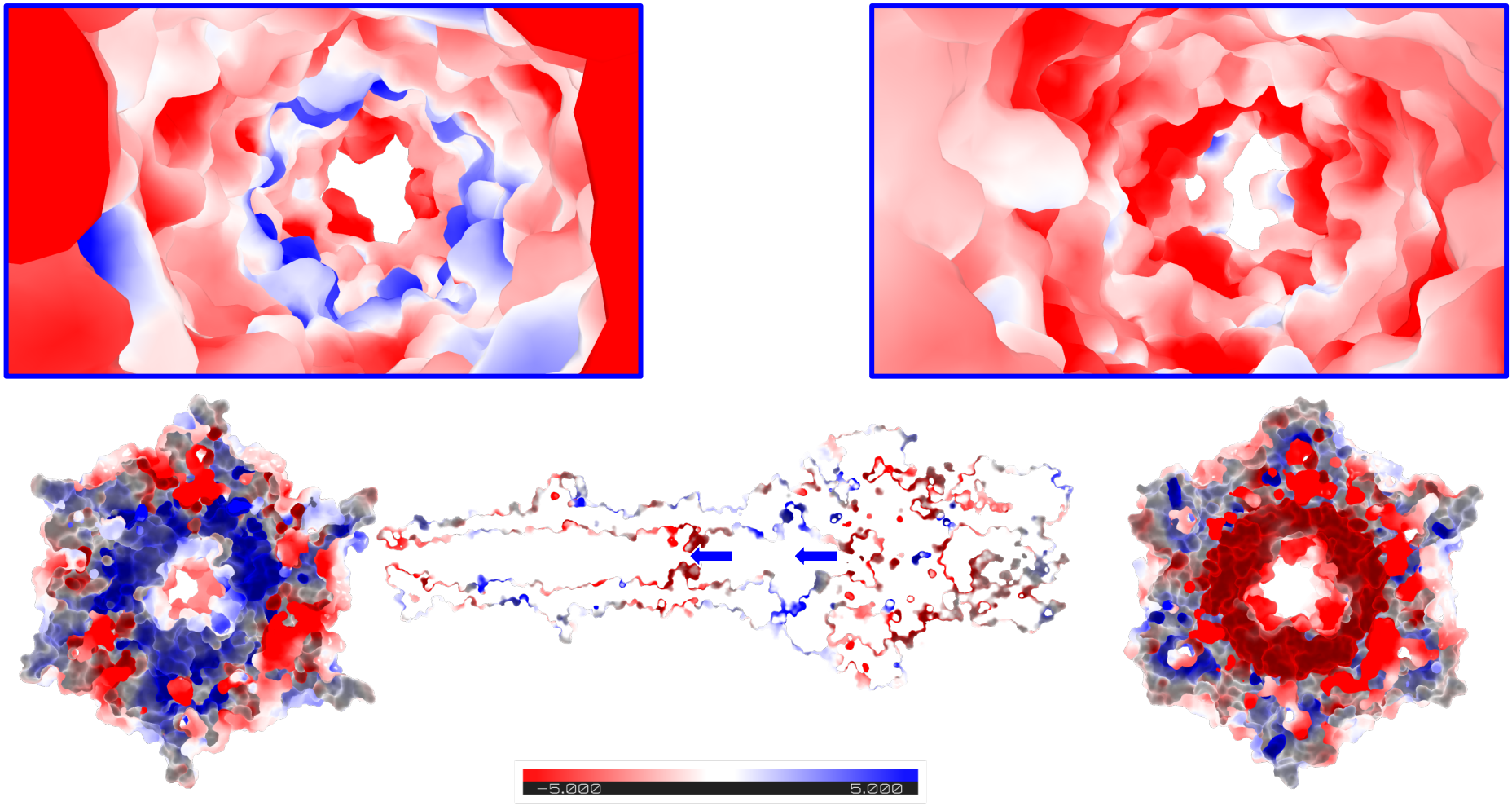


Figure 38 APBS Electrostatic properties of the MdtA pump channel (A). AcrAB-TolC channel (B). The entry aperture of the MdtA pump (C). The entry aperture to AcrAB-TolC (E). The cross-section of the MdtA pump with view arrows and electrostatic range key (D).

6.3.3 Electrostatic properties of the MdtA and TtgDEF efflux pumps compared using PyMol

The genes for pump *ttgDEF* are transcribed in the place of ttgABC when *P. putida* DOT-T1E is exposed to toluene, a hydrophobic molecule with a similar XLogP3 score to BMA (toluene XLogP3 = 2.7, BMA XLogP3 = 2.9). This pump channel also has a varied charge inside; not the same as MdtA, but both MdtA and TtgDEF exhibit different charge motifs inside the channel when compared to the other pumps examined (i.e., MexAB-OrpM, AcrAB-TolC, and ttgABC). To further understand these pumps' internal environments, fpocket was used to profile the electrochemical properties of protein pockets belonging to each MdtA and TtgDEF RND pumps (Le Guilloux et al., 2009b). As has been demonstrated, there are some observational differences between pumps suited to hydrophobic compounds and those that are more generalised. fpocket can determine if there are any quantifiable differences as fpocket profiles, all pockets and the external pockets of an RND efflux must be hydrophobic as it is membrane-bound; a custom Python script was used to remove the external pockets from the dataset and focus on pockets in the internal efflux environment. An example of the exclusion can be seen in Figure 39. A one-way ANOVA was used to analyse variations and differences within measured metrics. If the ANOVA detected variance, a graph was produced.

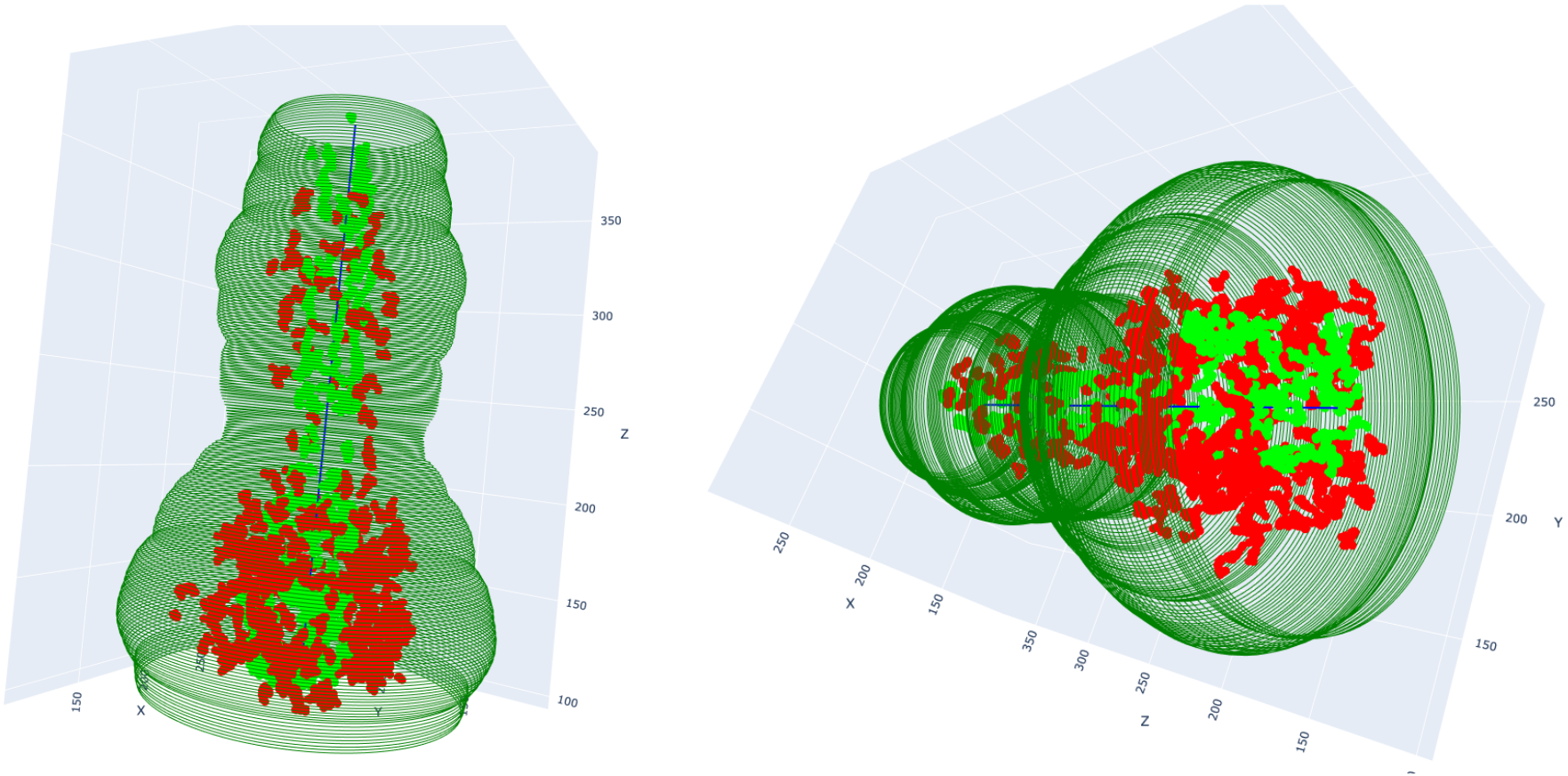


Figure 39 – Python Script Excluded Protein Pockets from Efflux Pump (**RED**) and included Protein Pockets (**GREEN**). The script created a central line through the pore of the efflux pump and then created 300 rings closing in into the central line once it hit a pocket cluster it excluded the pocket and any other the closing circle came in contact with leaving only internal pockets for analysis.

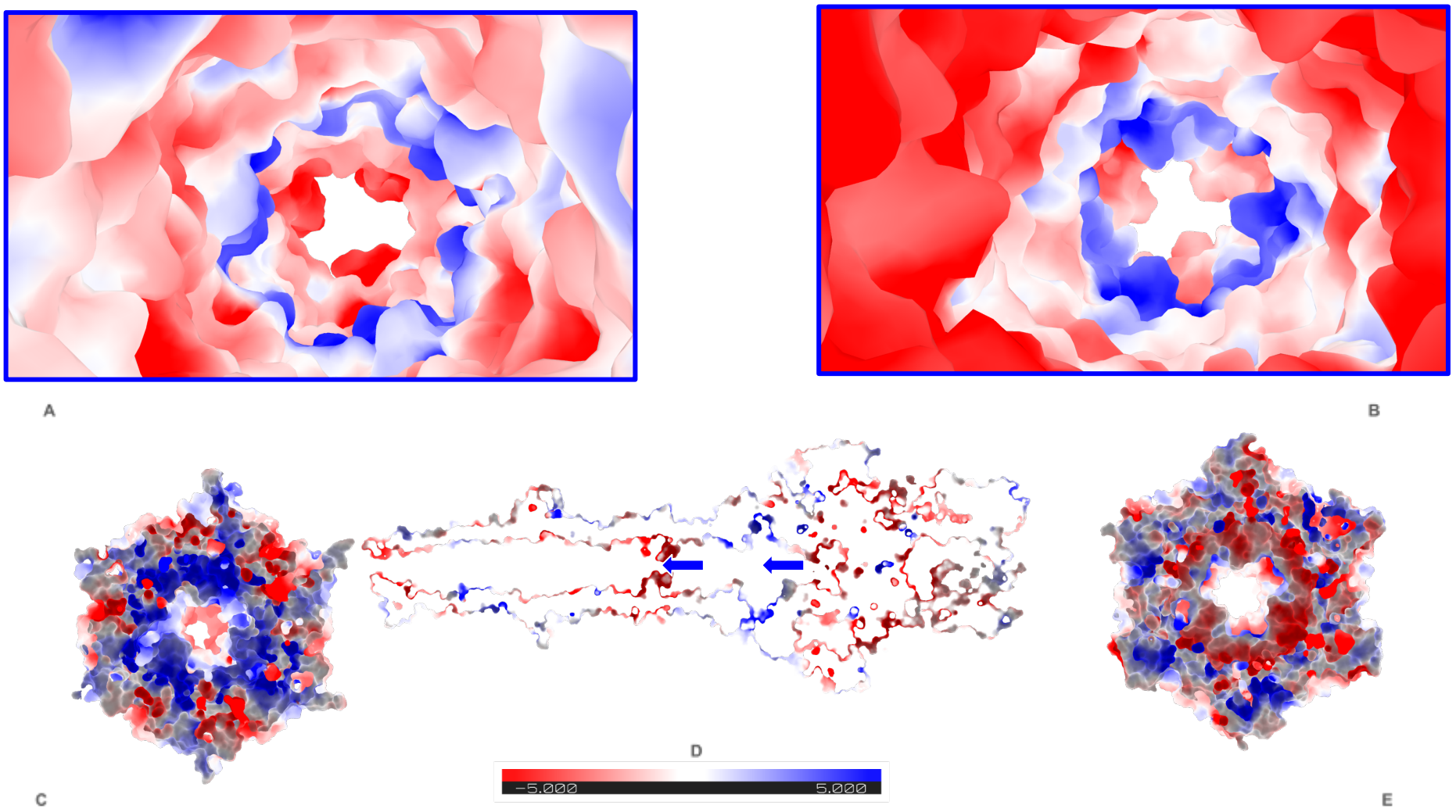


Figure 40 -PBS Electrostatic properties of the MOA (MdtA) efflux pump channel **(A).** TtgDEF channel **(B).** The entry aperture of the mdtA pump **(C).** The entry aperture to TtgDEF **(E).** The cross-section of the MdtA pump with view arrows and electrostatic range key **(D).**

6.3.4 Analysis of Variance in Polar Atom of Internal Protein Pockets of RND Efflux Pumps

There was only one detectable deviation in the mean-variance of polar atom content of all examined RND efflux pumps: the MdtA efflux pump was shown to have significantly fewer polar atoms within its pockets when compared to AcrAB-TolC (P-Value = 0.01) (Figure 41). The pockets in MdtA could be more suited to hydrophobic molecules than AcrAB-TolC (Le Guilloux et al., 2009c). Further analysis of variance of Ligand flexibility per pocket in the fpocket metrics was conducted.

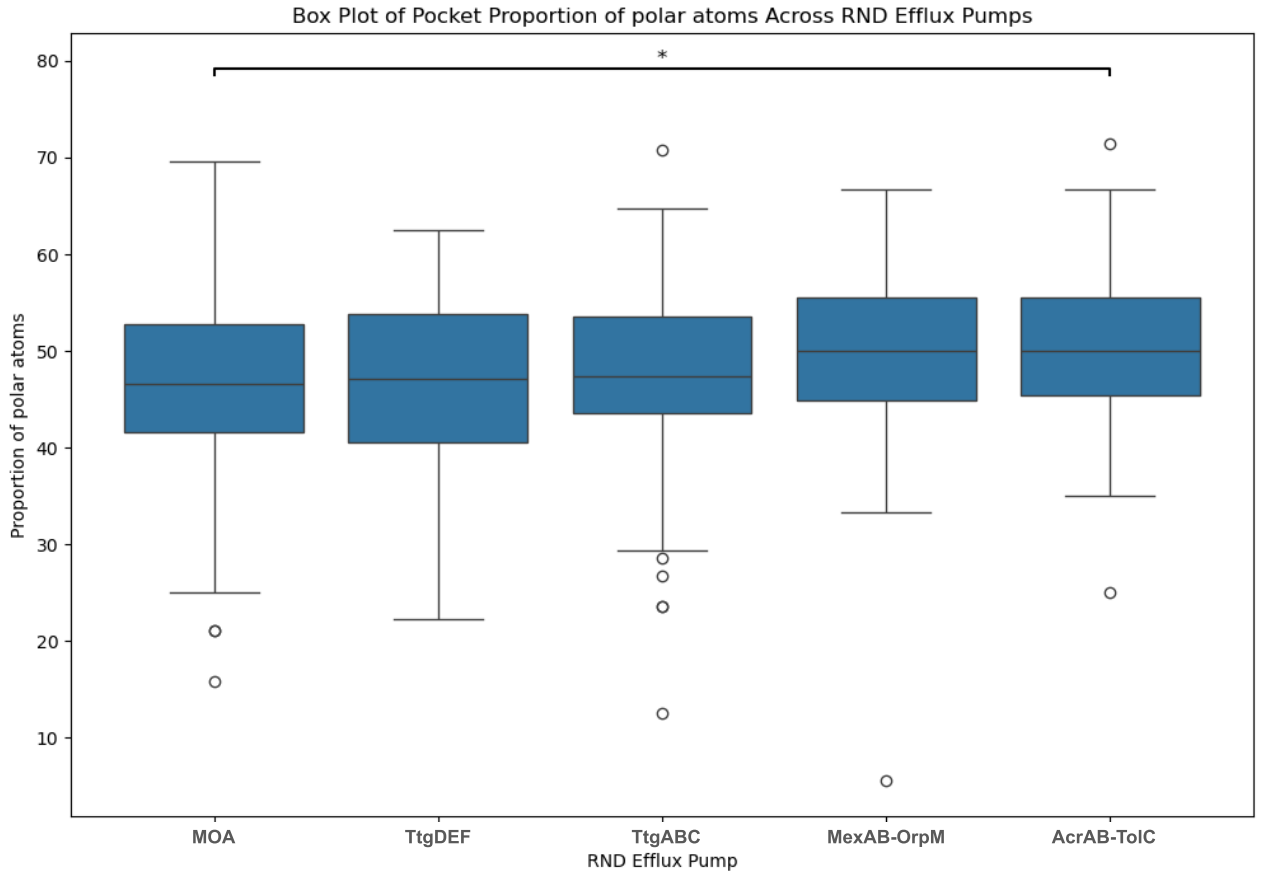


Figure 41 - Analysis of variance of fPocket Polar atoms per pocket. AcrAB-tolC mean pocket was shown to have significantly more polar atoms when compared to MOA (MdtA) (P-Value = <0.05)

6.3.5 Analysis of Variance in Ligand Flexibility of Internal Protein Pockets of RND Efflux Pumps

The variance in ligand flexibility of the RND efflux internal pockets was much more pronounced in the data compared to other compared fpocket results. The two pumps with abnormal internal electrostatic properties, MdtA and TtgDEF, had significantly higher flexibility scores than MexAB-OrpM, AcrAB-TolC, and TtgABC (Figure 42). As both of these pumps are found in organisms that are tolerant to hydrophobic solvents and *P. putida* DOT-T1E switches transcription of TtgABC to TtgDEF when treated with toluene, this could be due to the ligand flexibility of these pumps to accept a varied range of substrates (Udaondo et al., 2013), (Le Guilloux et al., 2009c).

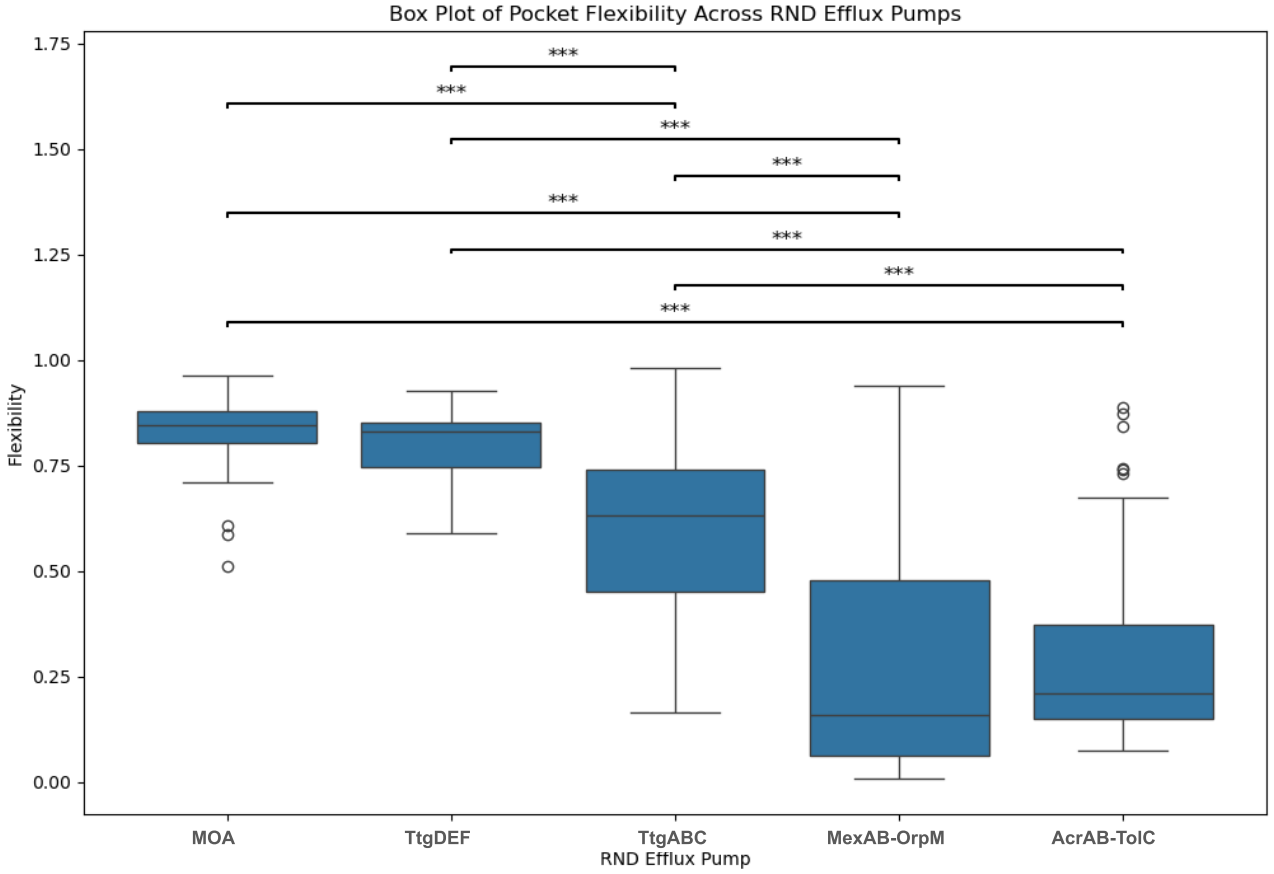


Figure 42 - Analysis of variance in pockets flexibility score MOA (MdtA) and TtgDEF show significantly higher ligand flexibility scores compared to all other systems (P-Value = <0.001).

6.3.6 Analysis of Variance in Volume of Internal Protein Pockets of RND Efflux Pumps

The analysis of pocket volume showed TtgDEF to have significantly higher volumes than other RND efflux systems studied, except TtgABC. The TtgABC mean pocket volume was significantly higher than AcrAB-TolC (Figure 43). As this metric does not trend towards pumps suited to hydrophobic compounds, it cannot be used. The one-way ANOVA also detected variance in the mean hydrophobicity scores, and Tukey HSD analysis was conducted.

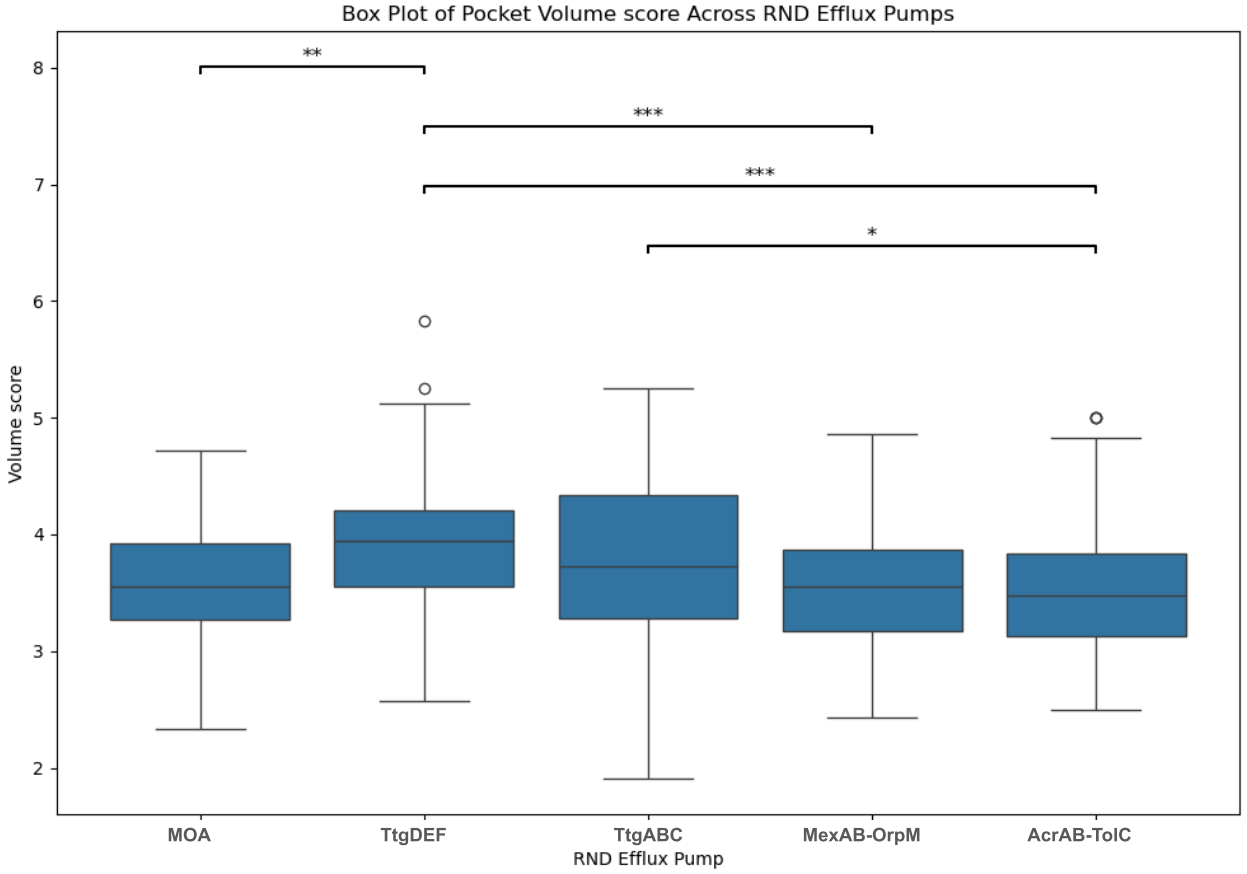


Figure 43 - Analysis of variance in Pocket volume. TtgDEF has a significantly higher mean pocket volume than other studied systems aside from TtgABC (MOA MdtA P-value = <0.01, MexAB-*OrpM* P-value = <0.001, AcrAB-TolC P-value = <0.001). TtgABC was significantly higher than AcrAB-TolC (P-Value = <0.04).

6.3.7 Analysis of Variance in Charge Score of Internal Protein Pockets of RND Efflux Pumps

The RND efflux pumps TtgABC and TtgDEF have significantly more charge variance in pockets than the mostly negatively charged pockets of AcrAB-TolC. The mean value of pocket charge for MdtA, TtgABC and TtgDEF was 0. The mean pocket of these pumps had no charge, suggesting again they might be better suited to dealing with hydrophobic molecules (Sharma et al., 2018).

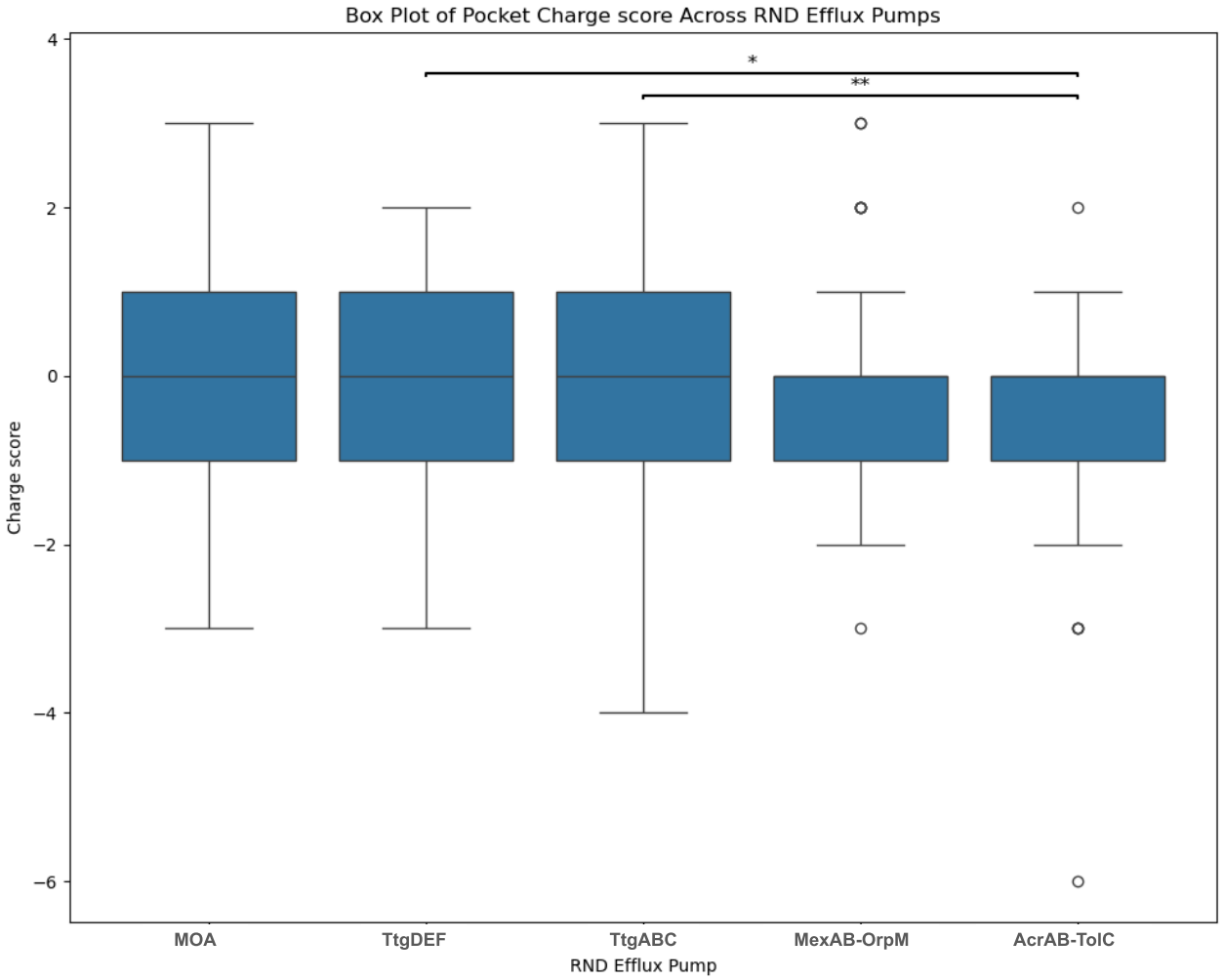


Figure 44 -- Analysis of variance in pockets Charge score. MOA (MdtA), TtgDEF and TtgABC all have a mean charge score of 0. The mean charge score was significantly more negative in AcrAB-TolC when compared to TtgABC (P-Value = <0.01) and TtgDEF (P-Value = <0.05).

6.4 Curing of the plasmid encoding the *mdtA-oqxB7-arpC* RND Efflux Pump in Strain RCS09A

The plasmid found in RCS09A pPp60-MOA contained a vapBC toxin-antitoxin (TA) system, presumably to maintain its lineage through generations (McVicker et al., 2019): VapBC TA toxin can be denatured with heat, allowing cells to divide without being killed by the retained toxin protein in the absence of the TA gene. Accordingly, cells were cultured at 35°C, 5°C above their normal temperature of 30°C, for two days. Cells were plated and colony PCR was performed to assay for *vapBC*, and mdtA was carried out to check for the plasmid presence. A Pp60-MOA negative colony was isolated and named RCS09A35. The RCS09A and RCS09A35 isolates were then subjected to BMA tolerance assays by exposure BMA (25% v/v). They were also grown under normal conditions in LB to check if the plasmid had any significant effect on growth.

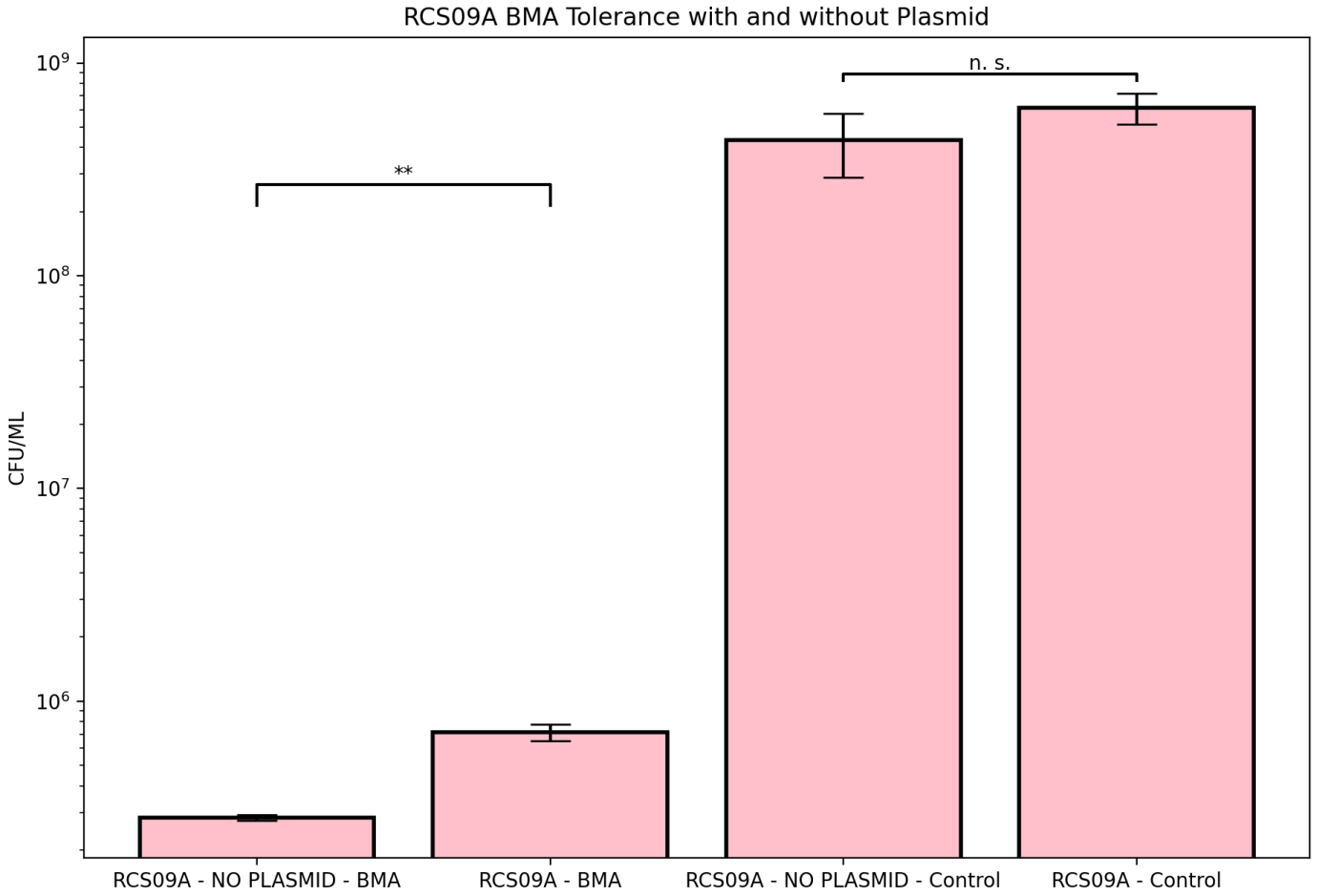


Figure 45 - Cured pPp56-MOA plasmid phenotypes of RCS09A treated with BMA 25% 1.56 M (**LEFT**). RCS09A cured of pPp65-MOA plasmid under normal growth conditions versus RCS09A with pPp65-MOA(**RIGHT**). RCS09A cured of Plasmid lost significant tolerance when cured of pPp65-MOA (P-Value = <0.01). The curing of pPp65-MOA had now significant effect on growth under normal conditions.

6.4.1 RCS09A Cured of Plasmid pPp65-MOA Showed a Significant Loss in of Tolerance to BMA

RCS09A cured of pPp65-MOA shows a significant loss of tolerance to BMA (Figure 46). The cured isolate showed no significant loss of ability to grow under normal conditions with or without the pPp65-MOA (Figure 45). This suggests that this plasmid was not a significant drain on cellular resources under normal conditions but provided a significant advantage when the cells were under solvent stress.

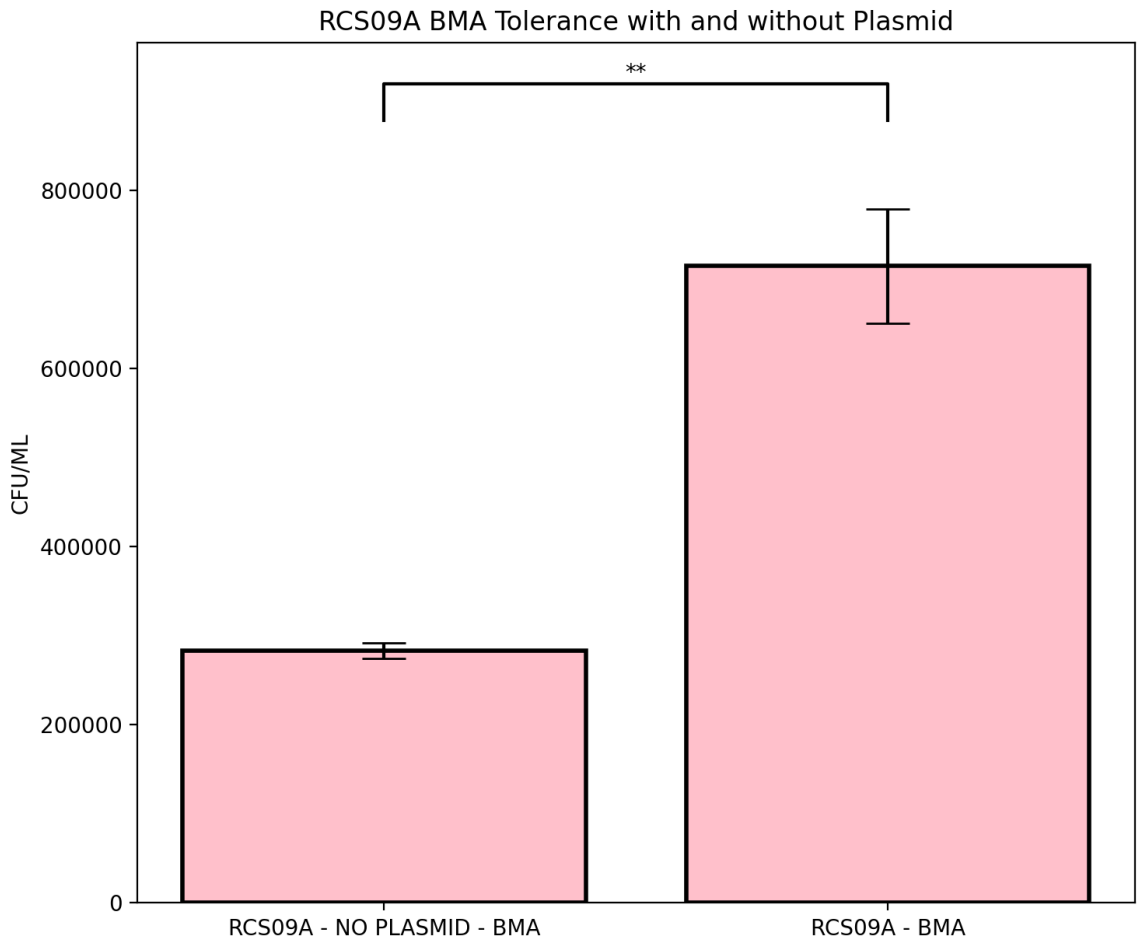


Figure 46 Cured pPp56-MOA plasmid phenotypes of RCS09A treated with BMA 25% v/v (1.56 M). No plasmid was significantly less tolerant than WT (P-Value = <0.01). Suggesting that pPp56-MOA confers a significant amount of tolerance to BMA. This is the same data as Figure 45 but excluding the data with no BMA treatment to show the loss of tolerance of RCS09A cured of plasmid with greater detail.

6.5 Transfer of mdtA-oqxB7-arpC RND Efflux System to the Chromosome of *P. putida* KT2440

To determine if the MdtA-oqxB7-arpC efflux pump is efficient at removing specifically BMA from the cell, it was decided to transfer the operon to the chromosome of *P. putida* KT2440. The wild-type organism, RCS09A, contains the mdtA-oqxB7-arpC operon on a plasmid at a higher copy number than the chromosome (shown in section 5.2.5). However, to test its application without the influence of a higher copy number, it was decided that a single addition to the chromosome would be a better assessment of its ability to confer tolerance to BMA. The entire operon, including the regulator, was constructed in the MCS of pUC18R6KT-mini-Tn7T-Km plasmid, which was confirmed by PCR and nanopore sequencing (Figure 47). The plasmid uses an R6kλ ori, which is not functional in KT2440; hence, the retention of the resistance marker and insert results from incorporation into the chromosome. The transposon insertion (carrying the mdtA-oqxB7-arpC operon) is assisted by helper plasmid pTNS which targets the insertion downstream from the *glmS* gene in a genetically neutral site, preventing disruption of other systems (Choi & Schweizer, 2006b). To further assess this operon's effectiveness in removing BMA via efflux, the resulting *P. putida* KT2440 strain carrying the mdtA-oqxB7-arpC operon was subjected to a BMA tolerance assay.

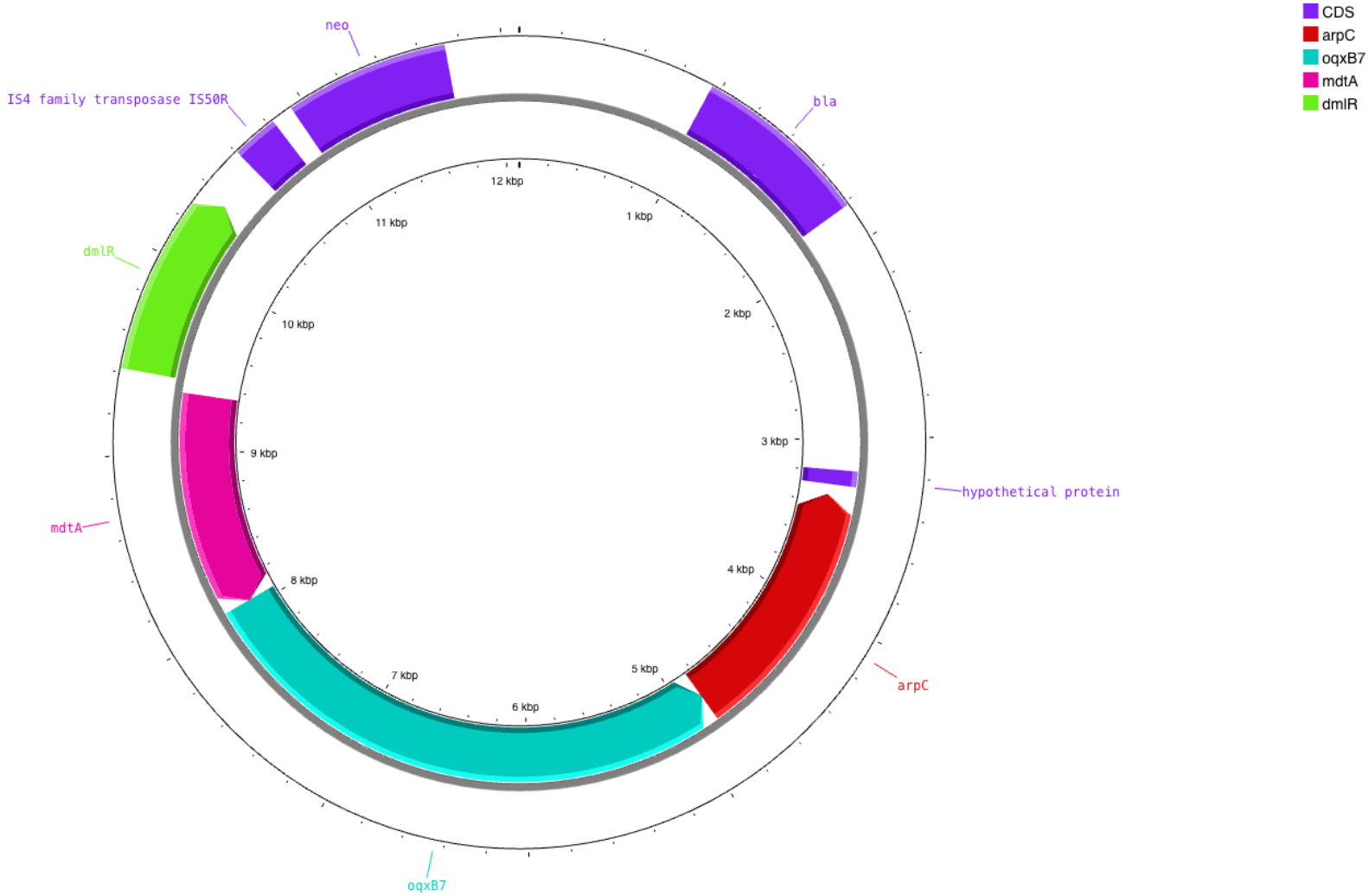


Figure 47 pUC18R6KT-mini-Tn7T-Km - mdtA-oqxB7-arpC construct confirmed by Eurofins plasmid sequencing services. The efflux operon was amplified using PCR from the RCS09A genomic DNA with over hangs for the pUC18R6KT-mini-Tn7T-Km plasmid. The plasmid backbone was also amplified with complimentary overhangs and assembled using Gibson assembly.

6.5.1 Transfer of the MOA efflux system from pPp65-MOA plasmid to the KT2440 chromosome increased tolerance to BMA when compared to KT2440 WT

The Transfer of the mdtA-oqxB7-arpC operon, including the regulator, to the chromosome of *P. putida* KT2440 (confirmed for PCR for the MOA efflux system) saw an increase in BMA tolerance when compared to *P. putida* KT2440 WT. The results were on the border of statistical significance (P-Value = 0.05) (Figure 48). However, a clear difference can be observed when looking at the serial dilutions on agar plates (Figure 49).

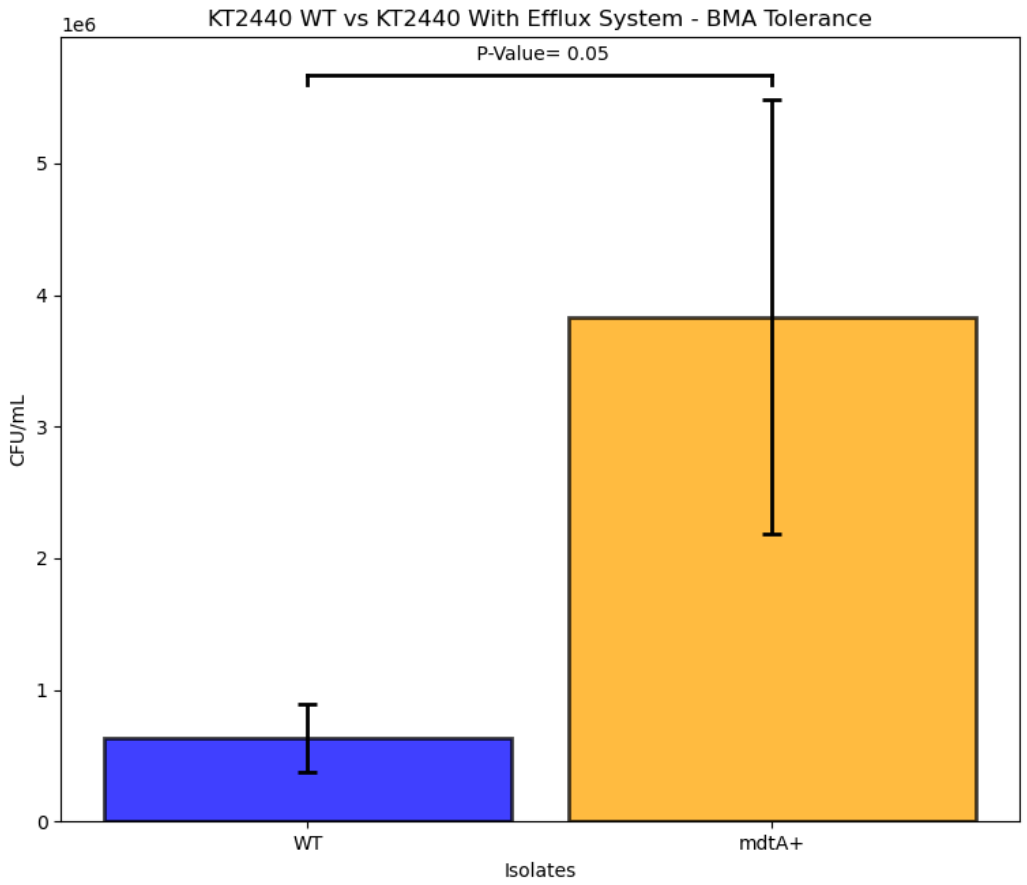


Figure 48 - BMA Tolerance Assay of KT2440 WT versus KT2440 mdtA+ treated with BMA 25% v/v (1.56 M). The efflux operon, including mdtA, introduced a significant increase in tolerance (P-Value = 0.03 *).

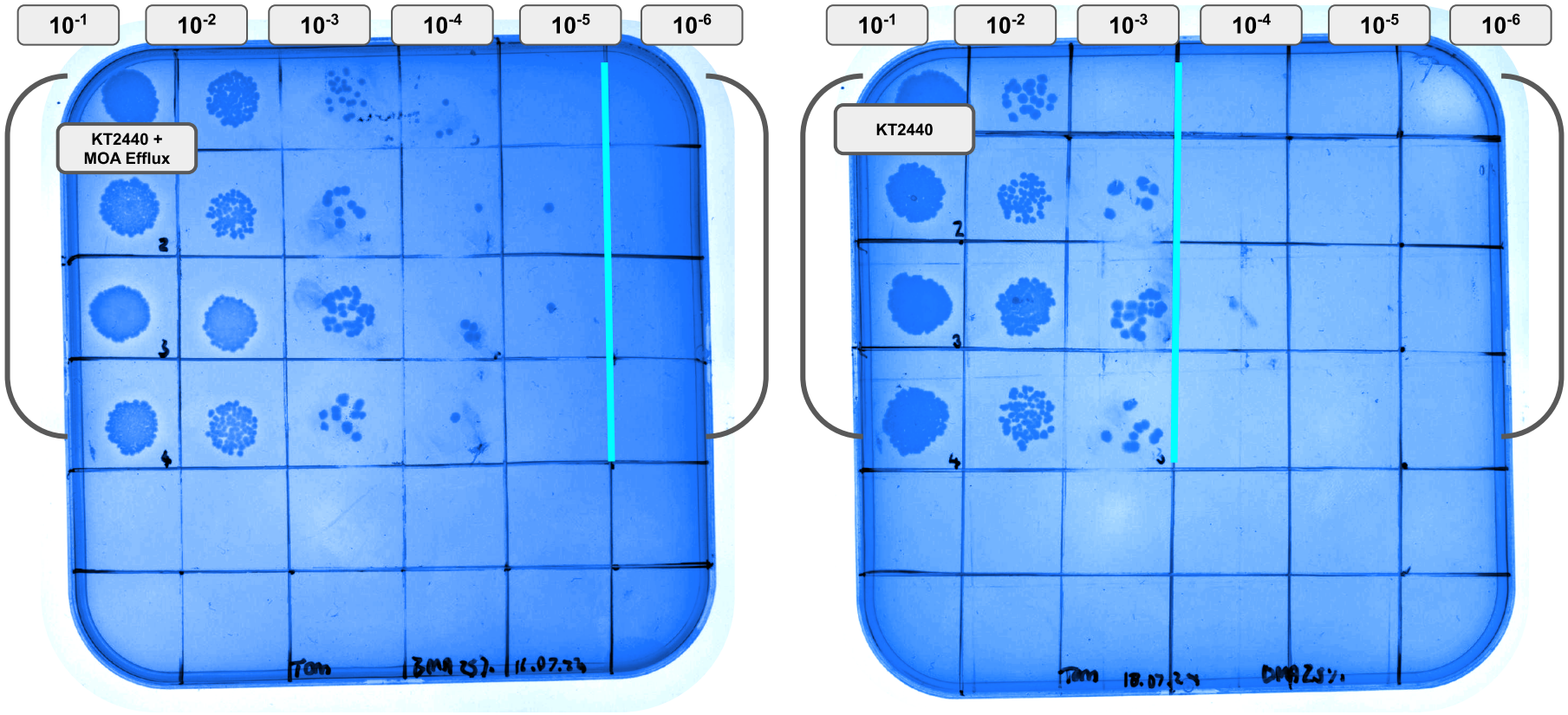


Figure 49- Serial Dilutions of KT2440+ MOA Efflux on Chromosome versus KT2440 WT

6.6 Clustering Outer Membrane subunits of RND Efflux Pumps Found in *P. putida*, veronii, fluorescens and *aeruginosa*. In combination with the outer membrane of subunits of RND pumps known to confer solvent tolerance to identify other pumps suited for industrial solvents

To provide alternative RND efflux pumps to convey tolerance to BMA and hydrophobic industrial solvents without using BLAST, which will return masses of nearly identical sequences of the three components of TtgDEF and MOA (MdtA). Both are shown to contribute to solvent tolerance who score similarly in protein pocket analysis comparison and have similar internal electrostatic properties only share Protein BLAST Identity score of 47.11% for the Outer membrane subunit, periplasmic adaptor 39.03% and 49.16% for the inner membrane permease. Currently, to run this comparison, we must consult the literature to determine if a pump is involved in solvent tolerance, download the resolved structure or predict its structure using the SWIISS MODEL or AlphaFold 2 Multimer (AlphaFold 2 - Google Collabs online resources max out in predictions of efflux pumps and fail) and then run the various analysis and projections which takes hours per pump.

Two-component Principal Component and Agglomerative Clustering Analysis were used to cluster the outer membrane subunits of RND efflux systems from *P. putida*, *P. veronii*, *P. fluorescens* and *aeruginosa*. Protein sequences were downloaded from the NCBI protein database. Due to the over-submission of *P. aeruginosa* sequences (attributable to its common

occurrence in pathology), a Python script was used to access the database and check if the sequences existed within the growing database, ensure unique sequences were downloaded, with a threshold of 95% sequence identity. The outer membrane subunit was chosen as the target as it showed the most observational differences in electrostatic projections inside the channel of pumps that were shown to be more suited to exporting hydrophobic solvents.

Physiochemical analysis was carried out using the Python protein analysis package, which outputs metrics such as molecular weight, aromaticity, instability index, isoelectric point, and gravy (hydrophobicity). The TtgDEF and MOA (MdtA) efflux pumps were both included as examples/references for systems known to confer solvent tolerance.

Two principal component analyses (PCA) and agglomerative clustering of the physiochemical data showed that the closest neighbour of TtgDEF was the MuxAB-OrmB system of *P. aeruginosa* strain A39-1 (QQV51235.1). However, there was no resolved structure for this, and the SWISS model could not find a template to thread the structure to. Structural Prediction using AlphaFold multimer V3 was attempted, but the server did not have enough GPU memory to predict such a complex multimer. The MOA efflux system's two PCA clustering neighbours were an unnamed outer membrane protein (ASJ84935.1) from *P. Aeruginosa* strain 83. In this case,

the SWISS model predicated the structure and compared it to the MOA efflux system (Figure 50).

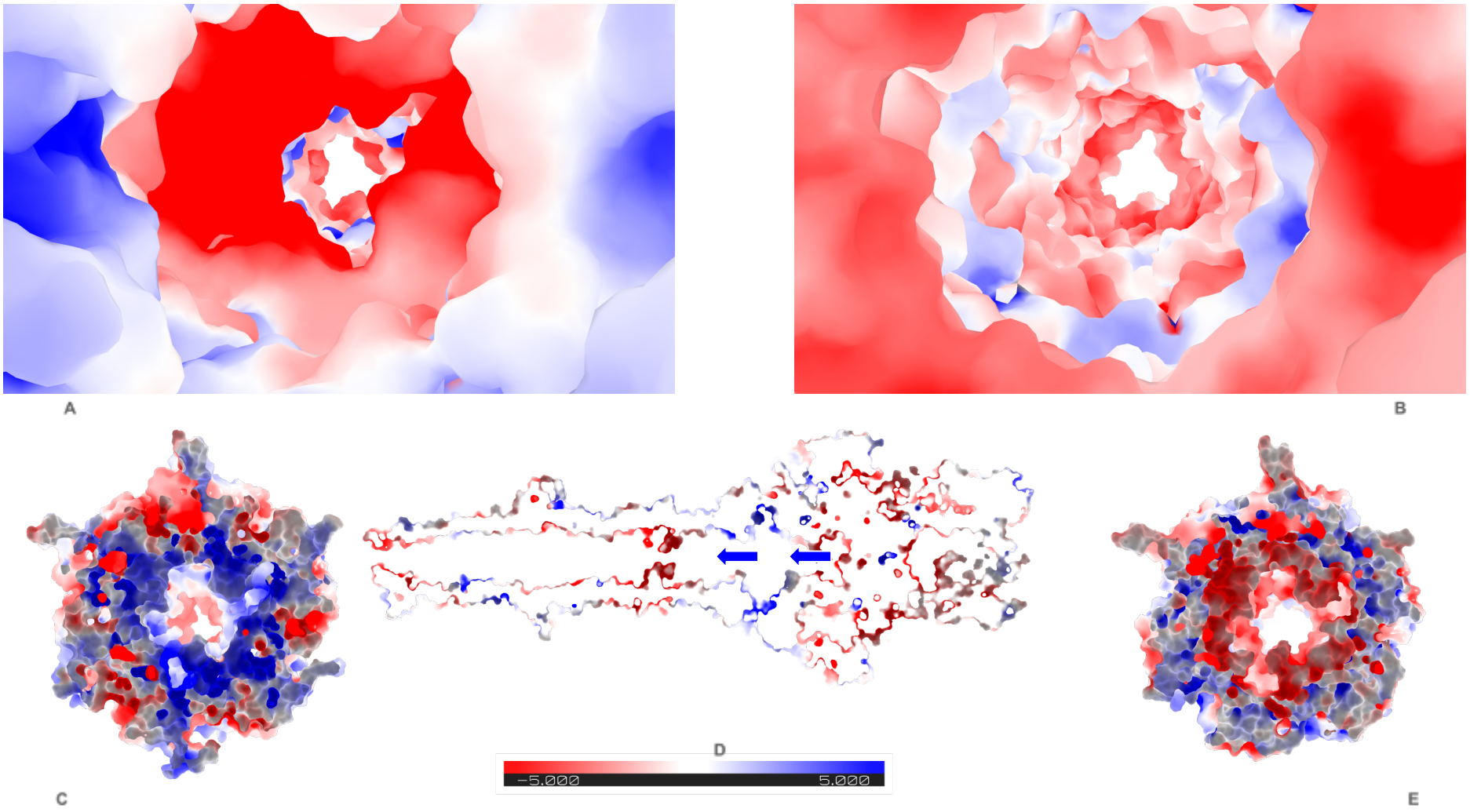


Figure 50- MOA efflux System (**LEFT**) and PA83 RND (**RIGHT**). PA83 has a varied charge motif lower down the channel than MOA (MdtA) (**A, B**).

6.6.1 Comparison MOA Efflux and PA83 Unnamed ASJ84935 RND Efflux Systems Internal Surface Electrostatic Properties

The ABPS Electrostatic projects on the two efflux internal structures are similar. The varied motif in PA83 is lower down the channel, where the MOA efflux has its ridge (Figure 42). The PA83 upper channel is much lighter in colour than MexAB-orpM and AcrAB-Tolc (Figure 31 and Figure 32), suggesting a more neutral overall charge. Unlike MOA, the opening aperture to the channel from the inner membrane subunit was negatively charged. PCA was conducted using three principal components and clustered to assess this approach and identify other potential efflux systems suited to hydrophobic compounds further.

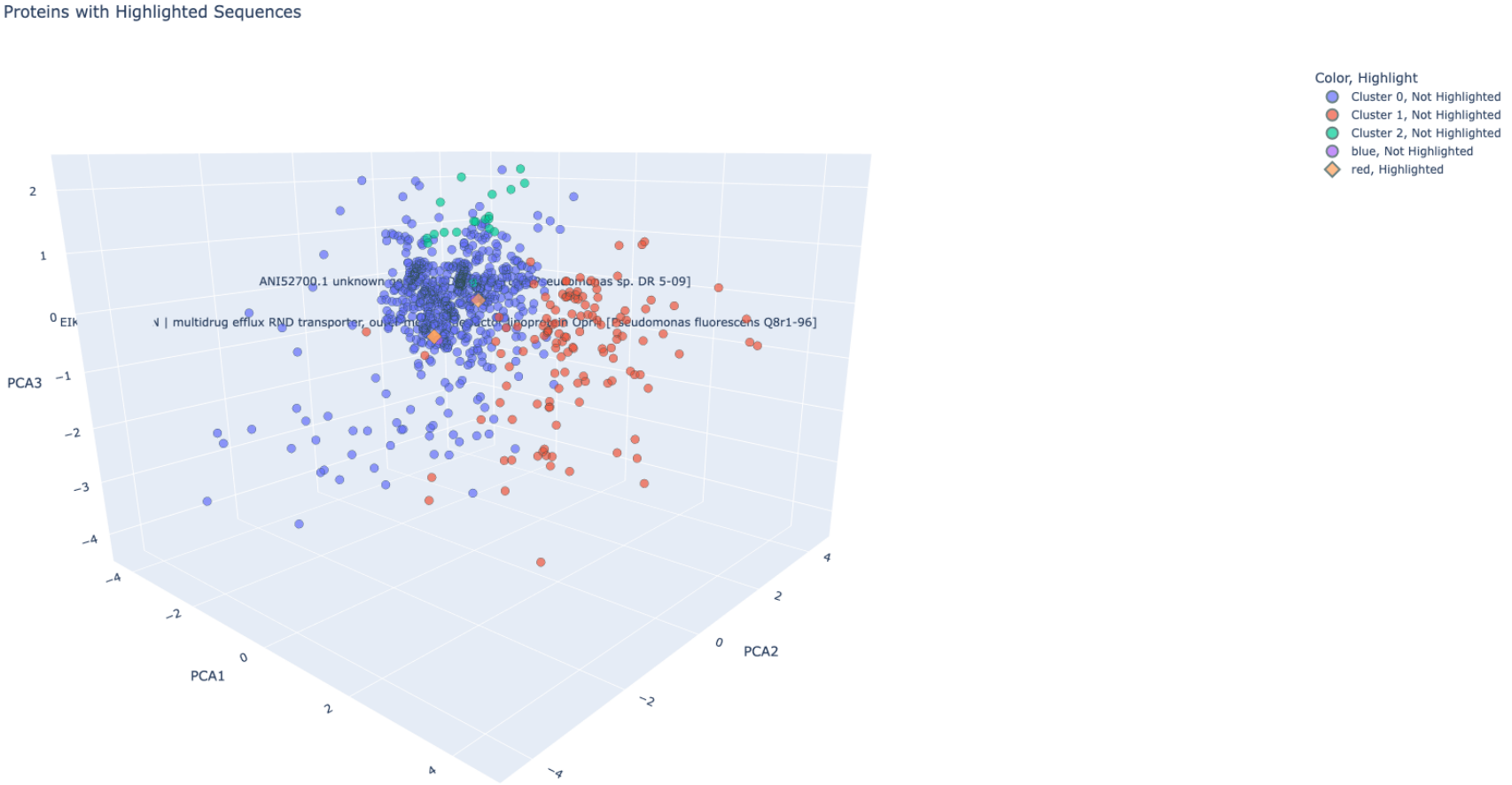


Figure 51 - PCA Three Component Agglomerative Clustering of Physiochemical Properties of RND outer membrane subunits

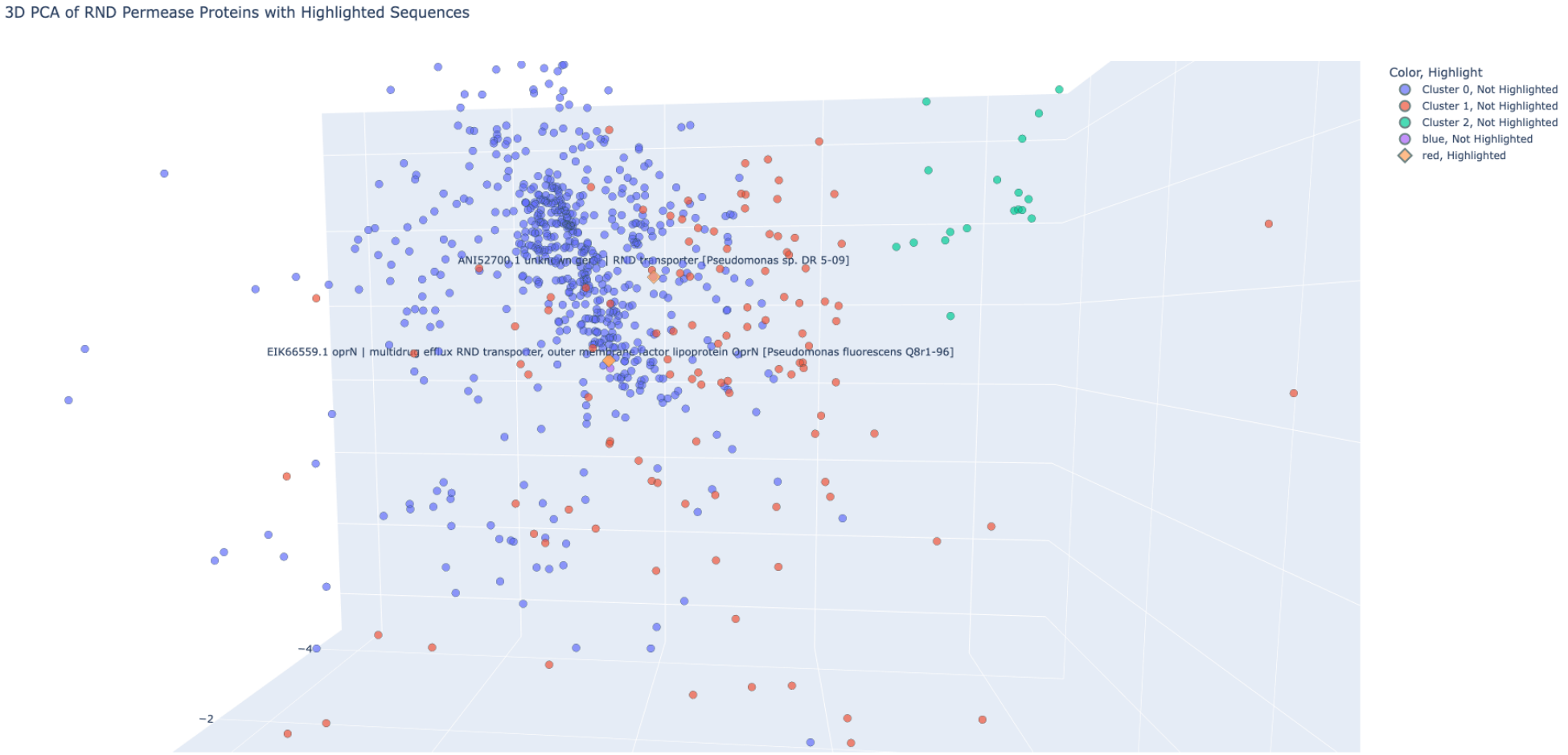


Figure 52 -- PCA Three Component Agglomerative Clustering of Physiochemical Properties of RND Outer Membrane Subunit . Zoomed in to make it more readable

6.6.2 Three-Component Principal Component and Agglomerative Clustering Analysis of RND Outer Membrane Efflux Subunits

The closest neighbour of the MOA efflux outer membrane subunit ArpC was ANI52700.1 from *P.* sp DR5-09, which at first glance seemed to be part of an incomplete RND Operon (containing the three subunits) with just only the outer membrane and a periplasmic adapter gene flanking a fusaric acid resistance gene. After using smart-blast on the fusaric acid resistance gene, a protein sequence of a permease adapter B subunit (ANI52701.1) of an RND belonging to the genus *Escherichia* was returned (blue arrow Figure 53). This suggests that this was not an incomplete operon but lacks characterisation. Fusaric acid is a hydrophobic compound with similar properties to BMA. Unfortunately, the SWISS Model could not find a template to thread the structure to.

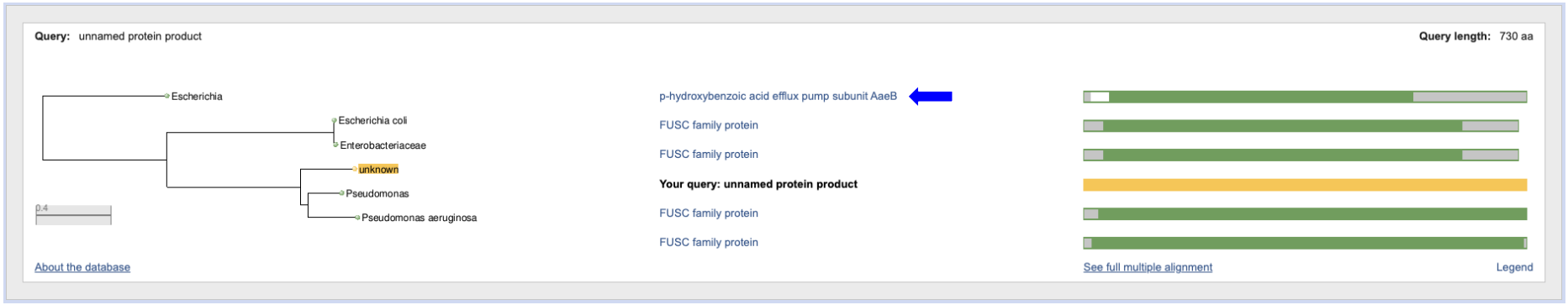


Figure 53 Smart BLAST results of ANI52700.1, a fusaric acid resistance gene.

6.6.3 Comparison of Internal Surface Electrostatic Properties of TtgDEF Efflux and MexFE-OprN EIK66559.1 RND Efflux Systems

The closest neighbour of the TtgDEF outer membrane subunit was EIK66559.1, part of the MexFE-OprN Efflux system of *Pseudomonas* fluorescens strain Q8r1-96. The cross-section of MexFE-OprN showed a ridge in the channel similar to the one observed in the MOA system (Figure 54 Red Arrow).

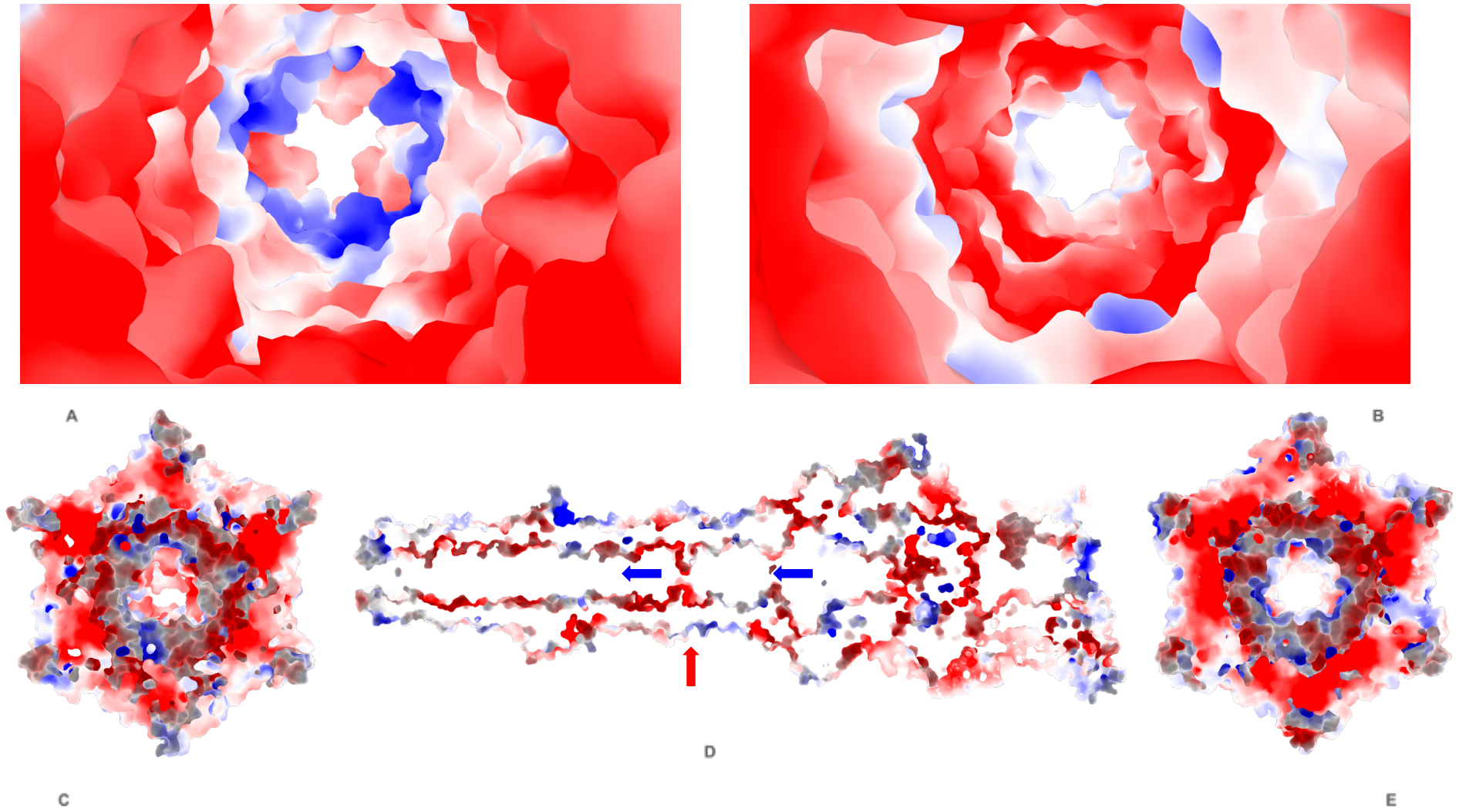


Figure 54 TtgDEF (**LEFT**) and MexFE-OprN (**RIGHT**) internal electrostatic properties. The cross-section shows MexFE-OprN highlighting the viewpoints (**BLUE ARROWS**) and the internal ridge of MexFE-OprN shown above the **RED ARROW** (**D**)**.** MexFE-OprN has a slight variation in charge motif in the channel (**B**) when compared to TtgDEF (**A**).

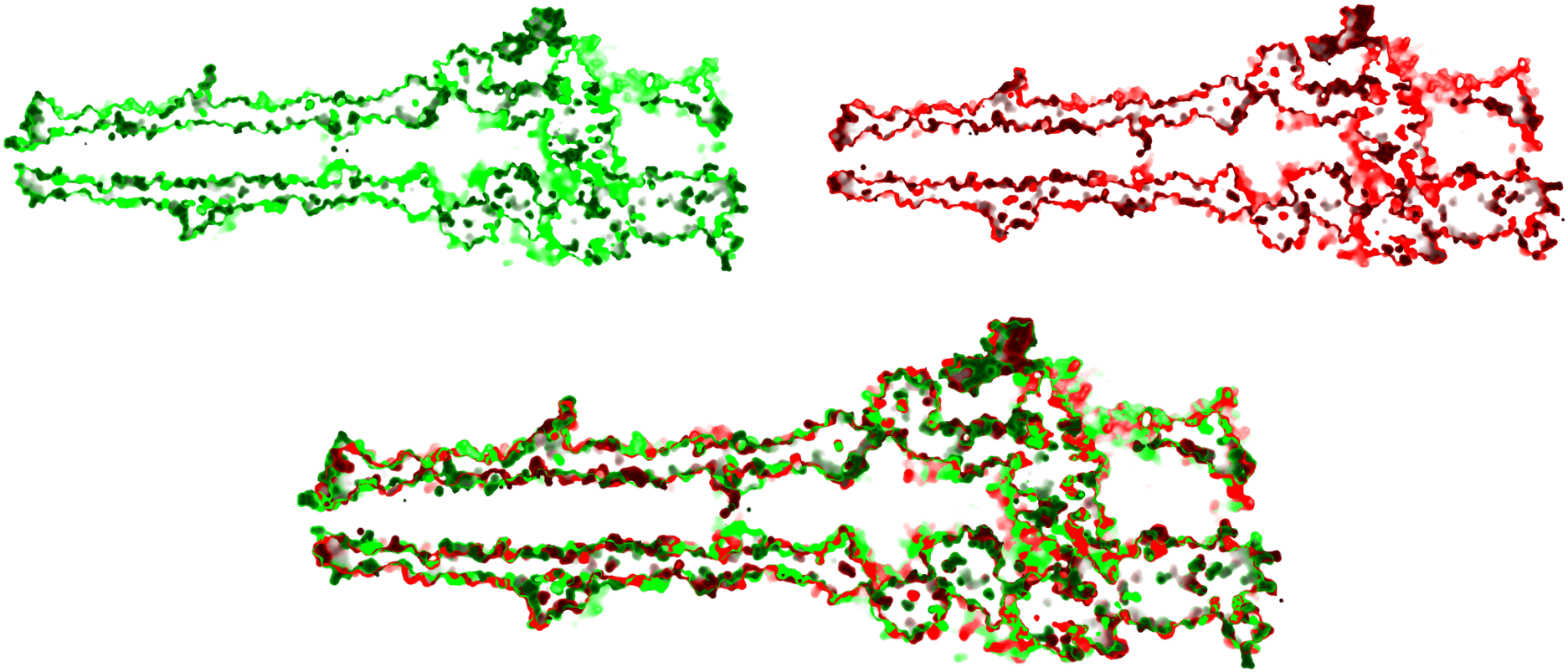
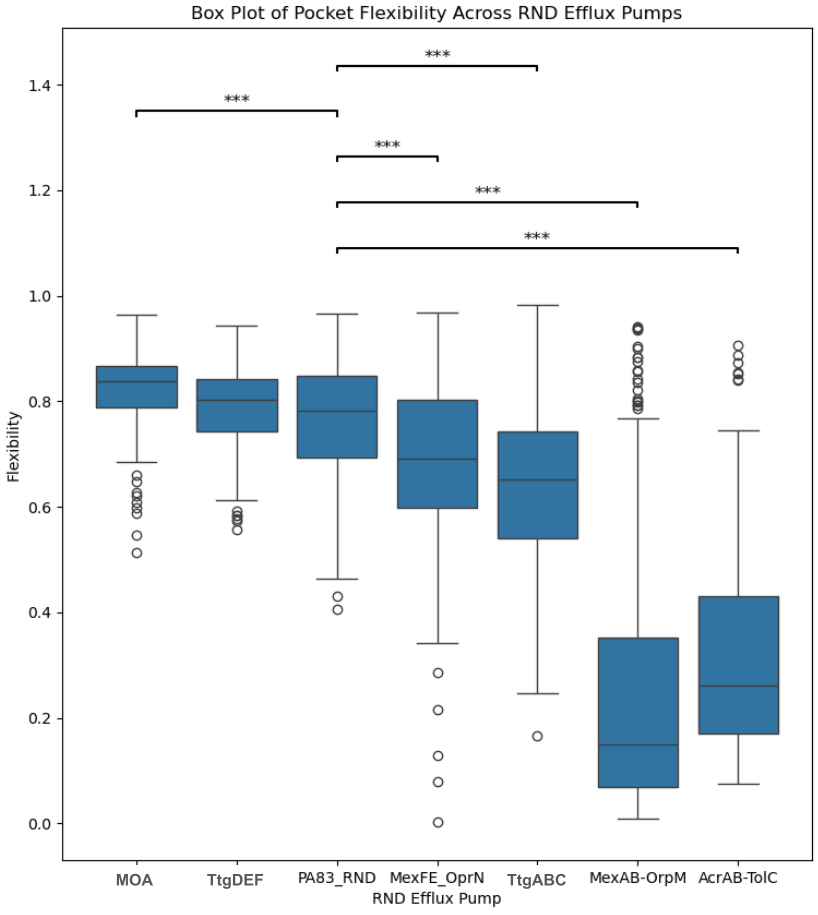


Figure 55 Comparison of MexFE-OprN (**RED**) and MOA (MdtA) (**GREEN**) Efflux systems internal structures. The aligned overlay shows the ridge in the internal channel is in the same position in both systems.

6.6.4 Comparison of internal Fpocket Analysis Results One-Way ANOVA

To understand if these observational similarities of MexFE-OprN and PA 83 RND compared to the MOA and TtgDEF efflux systems also translated into quantifiable similarities, fpocket analysis was conducted, and the external pockets were excluded as before. A one-way ANOVA was performed; if the variance in the means was detected, a graph was generated. As the most varied feature of the pocket analysis for the pumps associated with solvent tolerance was ligand flexibility, this is the focus of the following sections.



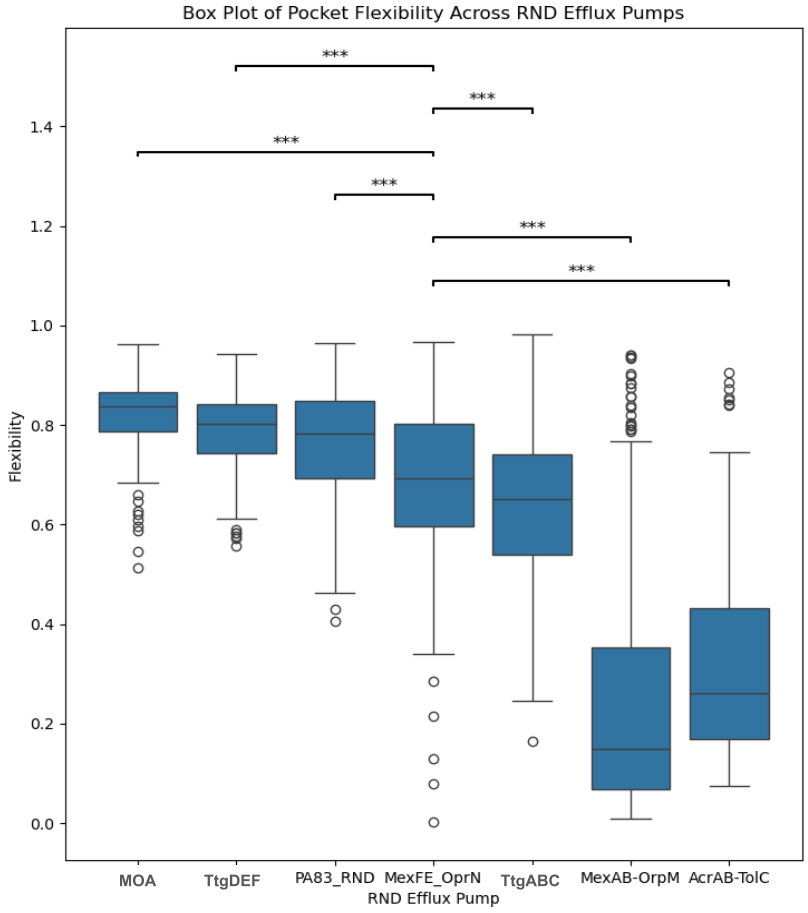


Figure 56 - Flexibility Score with Tukey HSD statistics focused on PA83 RND (**LEFT**) and MexFE-OprN (**RIGHT**). MexFE-OprN (**RIGHT)** is significantly lower in ligand flexibility than MOA (MdtA), TtgDEF, and PA83 RND (P-Value = < 0.001). PA83 RND (**LEFT**) is only significantly lower than MOA (MdtA)

6.6.5 Analysis of Variance in of Means of Protein Pocket Ligand Flexibility

The two newly identified efflux systems scored significantly higher than TtgABC in ligand flexibility but were significantly lower than MdtA (MOA) (Figure 56). How a data set is presented to a machine learning algorithm can impact the output results. Protein sequence data cannot be presented as it consists of letters representing each residue. The physiochemical data is easily generated using Bio-Python and can be clustered quickly due to its simplicity. We presented the data in a different format to get more out of the dataset, converting each residue into two numerical values, one for charge and one for hydrophobicity score (Table 9). This is known as feature engineering.

Table 9 - Feature Engineering Amino Acid Conversion Table ("The Proteomics Protocols Handbook," 2005)

Amino Acid	Charge	Hydrophobicity
A (Alanine)	0	1.8
R (Arginine)	1	-4.5
N (Asparagine)	0	-3.5
D (Aspartic acid)	-1	-3.5
C (Cysteine)	0	2.5
E (Glutamic acid)	-1	-3.5
Q (Glutamine)	0	-3.5
G (Glycine)	0	-0.4
H (Histidine)	1	-3.2
I (Isoleucine)	0	4.5
L (Leucine)	0	3.8
K (Lysine)	1	-3.9
M (Methionine)	0	1.9
F (Phenylalanine)	0	2.8
P (Proline)	0	-1.6
S (Serine)	0	-0.8
T (Threonine)	0	-0.7
W (Tryptophan)	0	-0.9
Y (Tyrosine)	0	-1.3
V (Valine)	0	4.2
X (Unknown)	0	0

6.6.6 Three Principal Component Analysis and DBSCAN Clustering of Converted RND Outer Membrane Subunit Data

The converted features were clustered using DBSCAN, which was chosen for its ability to exclude/identify noise (Ester et al., 1996a). This dataset was automatically generated using broad search terms (“RND” AND “Outer Membrane” AND “PSEUDOMONAS.”). Due to the automatic generation many junk-unrelated sequences were added to the dataset. Three principal components of each converted sequence, including TtgDEF and MOA, were clustered and plotted in three dimensions. For simplicity, the neighbouring proteins are shown in tabular format.

Table 10 - DBSCAN Cluster Neighbours of TtgDEF Outer Membrane Subunit

ttgDEF Distance	NCBI Protein ID	Organism	Junk
0.8296	UXA47019.1	PA01135	Yes (Non RND)
0.9082	UXA41980.1	PA01135	Yes (lone gene)
1.1546	**QDR36816.1**	**PA SCAID WND1-2019**	**No (MexAB-OrpM)**
1.2086	EKM96886.1	Stutzerimonas degradans strain Chol1	Yes
1.2148	UXA45833.1	PA01135	Yes (Non RND)
1.2260	UXA43966.1	PA01135	Yes (Non RND)

6.6.7 DBSCAN Clustering Results of Outer Membrane Subunits Neighbouring to the TtgDEF Outer Membrane Subunit

The six neighbouring outer membrane proteins of the TtgDEF outer membrane subunit were checked, one belonging to the *P. aeruginosa* strain SCAID WND1-2019 isolated in Kazakhstan (Table 10, in bold). The efflux operon was complete, although the annotation split the outer membrane subunit into two parts. This broken OrpM was repaired, and the protein sequences were aligned to the PA01 *MexAB-orpM* used in this study. The MexA subunits aligned with 93.99% identity, *MexB* aligned with 25.43% identity and OrpM aligned with 24.05% identity. This suggests that even though the query proteins genes had been annotated as *MexAB-orpM*, it was a quite different sequence, particularly regarding the outer membrane and permease subunits. The SWISS model was used to predict the structures using *MexAB-orpM* as a template. The structure of TtgDEF and SCAID-WND1 MexAB-OrpM were aligned in PyMol, and their electrostatic properties were projected onto them.

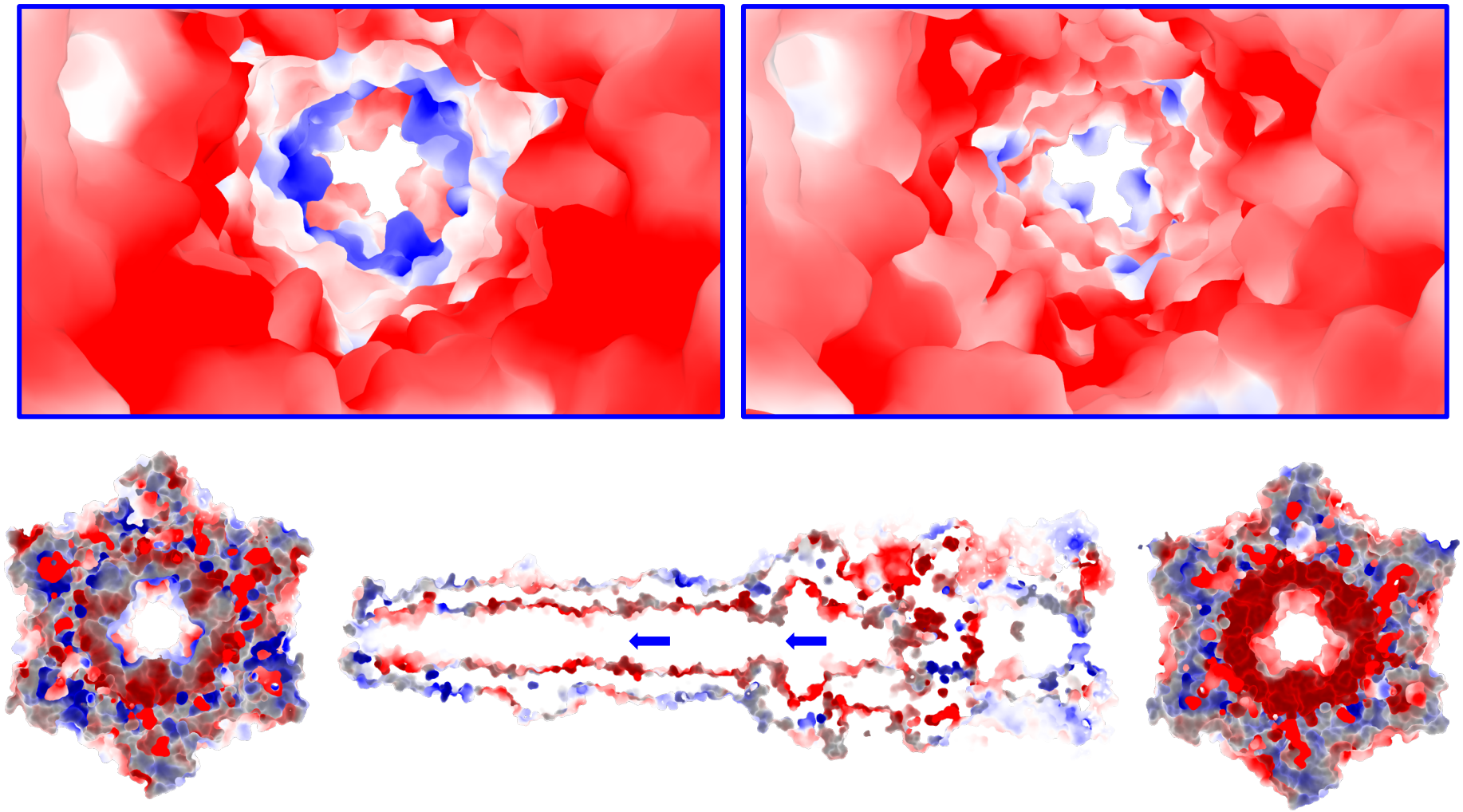


Figure 57 - Internal Electrostatic Proprieties of ttgDEF (**LEFT**) and SCAID MexAB-OrpM (**RIGHT**).

6.6.8 Internal Electrostatic Properties of TtgDEF compared to SCAID *MexAB-orpM* Efflux Systems.

The SCAID MexAB-OrpM Efflux System shows minimal variation in the channel charge motif. It is not as extreme as TtgDEF (Figure 57). The MexFE-OprN system identified in the first clustering experiment showed a more varied charge motif up to the region of its channel than the SCAID system but did not score higher than TtgDEF in the ligand flexibility in the fpocket analysis. To further assess the SCAID system's similarity to the efflux pumps shown to confer solvent tolerance, fpocket analysis was conducted with the external pockets excluded.

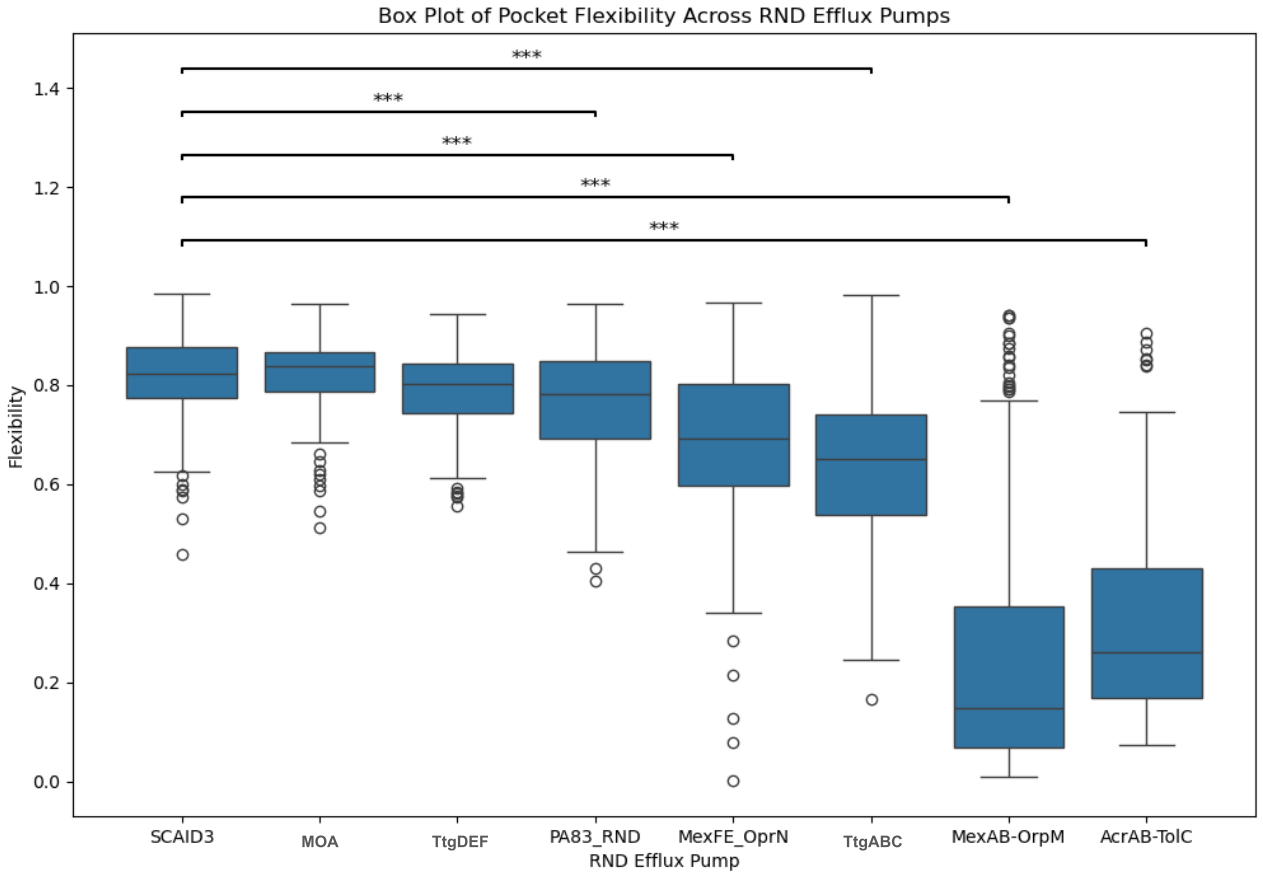


Figure 58 -fPocket Ligand Flexibility Score Focusing on PA SCAID MexAB-orpM Efflux System. The SCAID MexAB-OrpM efflux system did not show a statistical difference in ligand flexibility compared to MOA (MdtA) and TtgDEF efflux systems. It was significantly higher than all other studied systems (P-Value = <0.001).

6.6.9 Analysis of Variance in Means of the fpocket Ligand Flexibility Score of RND Efflux Pump on PA SCAID MexAB-OrpM Efflux System

The SCAID MexAB-OrpM Efflux Systems ligand flexibility score was not significantly different from that of MOA (MdtA) or TtgDEF. Though not significant, the mean score of the SCAID system was higher than that of TtgDEF. The two systems have been shown to confer tolerance to BMA and Toluene. To assess the potential of the SCAID efflux system in conveying tolerance to BMA and toluene, the operon would have to be synthesised and added to the chromosome of KT2440 with or without its regulator. Due to time restrictions for this project, this was not possible. To identify other pump systems the clustering neighbours of the MOA efflux systems' outer membrane subunit were analysed.

Table 11 - DBSCAN Clustering neighbours of MOA Outer Membrane Subunit

MOA Distance	NCBI Protein ID	Organism	Junk
4.2148	UXA42662.1	PA01135	Yes (Non RND)
5.2705	AGI24272.1	Pseudomonas denitrificans ATCC 13867	Maybe (No template)
5.7127	AGA71732.1	PP HB3267	Maybe (No template)
5.9466	EJF71699.1	Pseudomonas sp. Ag1	Maybe (No template)
5.9852	SPO59274.1	SPO59274.1	Yes (Non RND)
6.5085	EFQ62586.1	Pseudomonas fluorescens WH6	Maybe (No template)

6.6.10 DBSCAN Clustering Results for Outer Membrane Subunits that are Neighbours to the MOA Outer Membrane Subunit

The Clustering Neighbours of the MOA outer membrane subunit were much farther away than the clustering neighbours of TtgDEF (Table 10 and Table 11). This suggests there were no protein sequences whose charge and hydrophobic score were similar to that of the MOA outer membrane subunit. To understand this, the DNA sequence of the MOA operon was searched against the NCBI BLAST database. A neighbour-joining tree was generated of the top sequences hits (Figure 59).

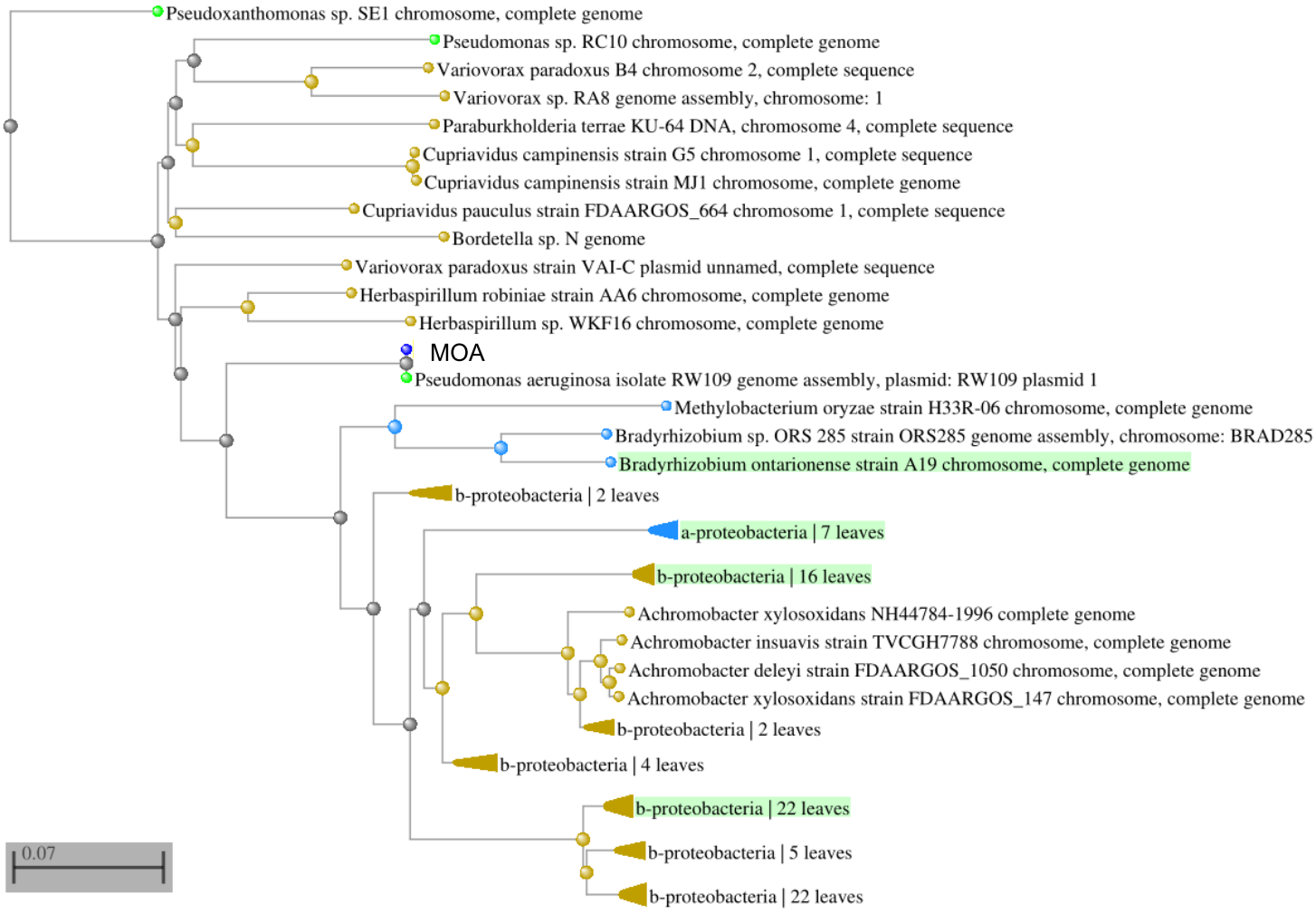


Figure 59 - Neighbour Joining Phylogenetic Tree of Top Returned BLAST Sequences of MOA operon

6.6.11 Analysis of BLAST Results of the MOA Efflux Operon

The closest to the MOA operon DNA sequences was the RW109, previously identified to be associated with a solvent-tolerant phenotype. Other matches aside from two far-removed *Pseudomonas* species (Green branches Figure 59), the closest aligned sequences were from organisms *Alphaproteobacteria Methylobacterium oryzae* strain H33R-06 and *Betaproteobacteria Herbaspirillum* sp. WKF16. This indicates why, when the entire converted sequence data was used to cluster, DBSCAN was unable to place the MOA close to any *Pseudomonas* proteins, as it is likely that, as this efflux operon being held on a plasmid that may not have evolved within a *Pseudomonas* species.

6.6.12 Furthest Points in the Clustering Analysis from TtgDEF

To understand if a high flexibility score was an artefact of being predicted using the SWISS Model, we calculated the furthest points away from the TtgDEF outer membrane subunit calculated. Expecting these pumps to score lower in ligand flexibility. These data points were in cluster 24, and all organisms within that cluster were strains of *P. aeruginosa* (Table 8). Unlike the cluster that included TtgDEF and MOA efflux systems, all operons were complete, and their structures could all be predicted.

Table 12 - Cluster 24 *P. Aeruginosa* strains - Furthest away from TtgDEF outer membrane subunit

ttgDEF Distance	NCBI Protein ID	Organism	Junk
68.3861 (C=24)	QUI58207.1	Pseudomonas aeruginosa strain PAM68	No
68.3536 (C=24)	TEI52613.1	Pseudomonas aeruginosa strain 5985 IPC1306_1.1	No
68.3363 (C=24)	RTU18670.1	Pseudomonas aeruginosa strain MRSN25678	No
68.3243 (C=24)	RRI10072.1	Pseudomonas aeruginosa strain 1713	No
68.3146 (C=24)	WVC66546.1	Pseudomonas aeruginosa strain F006	No
68.3049 (C=24)	UJB97003.1	Pseudomonas aeruginosa strain YTSY4	No

6.6.13 fPocket Analysis of Cluster 24 RND Efflux Pumps

The Analysis of variance in the means of the fPocket report outputted two plots. Plot one showed a slight variance in the hydrophobicity score, suggesting that the internal pockets of the SCAID and MOA RND Efflux Systems were slightly more hydrophobic than those of cluster 24. The Tukey HSD found no significant variance in the means between groups (Figure 60). The ligand flexibility score analysis showed that the internal pockets of the efflux systems represented by cluster 24 were significantly less flexible when compared to the SCAID efflux pump and MOA (Figure 61 - Statistics focusing on MOA MdtA for simplicity). The flexibility score was not as low as the two pumps that were resolved using X-ray crystal, suggesting that there might be a bias towards a more flexible pocket structure in structures predicted by the SWISS Model (Figure 62). To further understand this, an entire comparison study of X-ray-resolved and predicted protein structures would be required. Regardless of this potential bias, there is a statistical difference in ligand flexibility in pumps demonstrated to confer solvent tolerance against others.

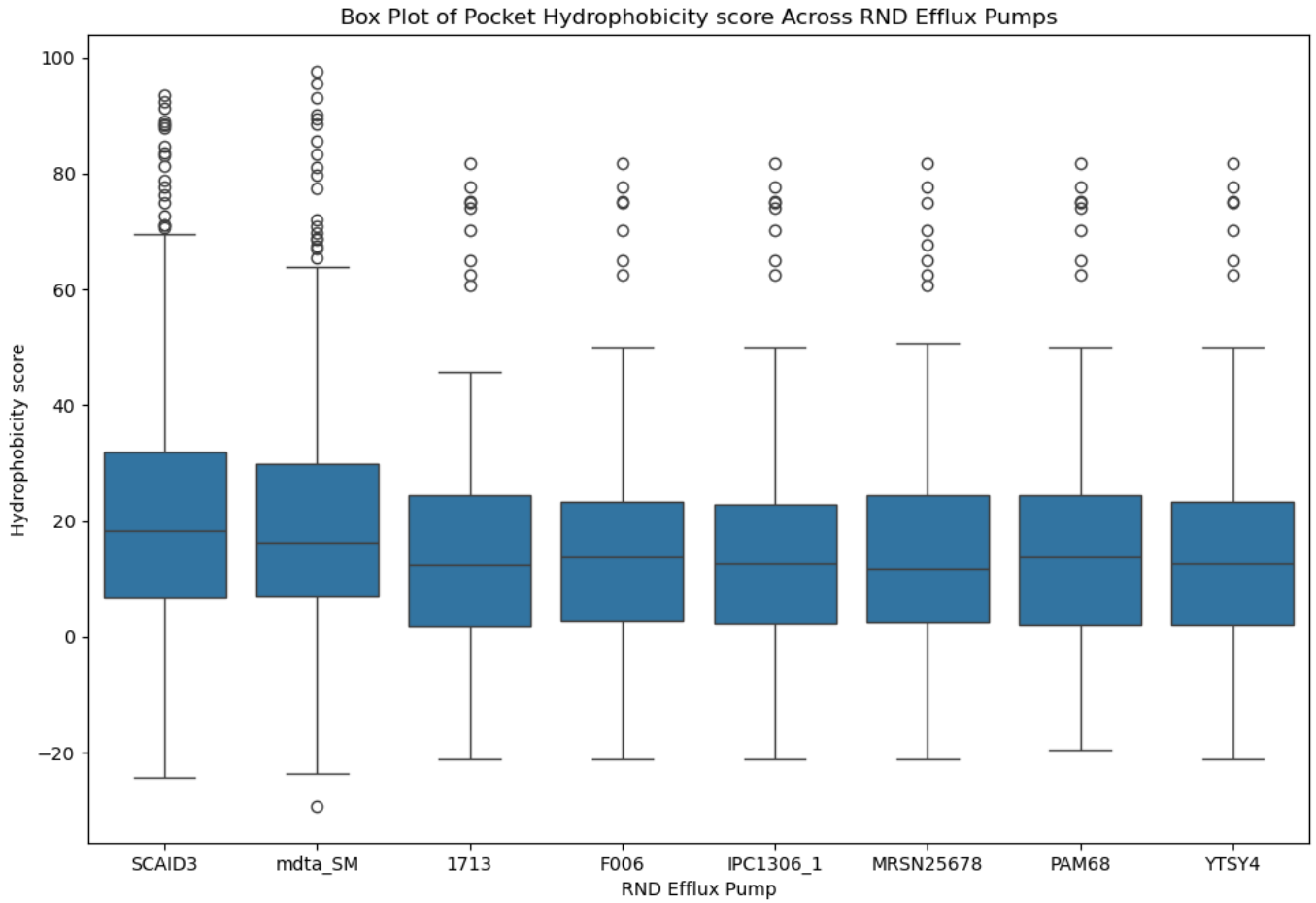


Figure 60 - Analysis of Mean-Variance in fPocket Hydrophobicity Score for Cluster 24,MOA (MdtA) and SCAID RND Efflux Pumps

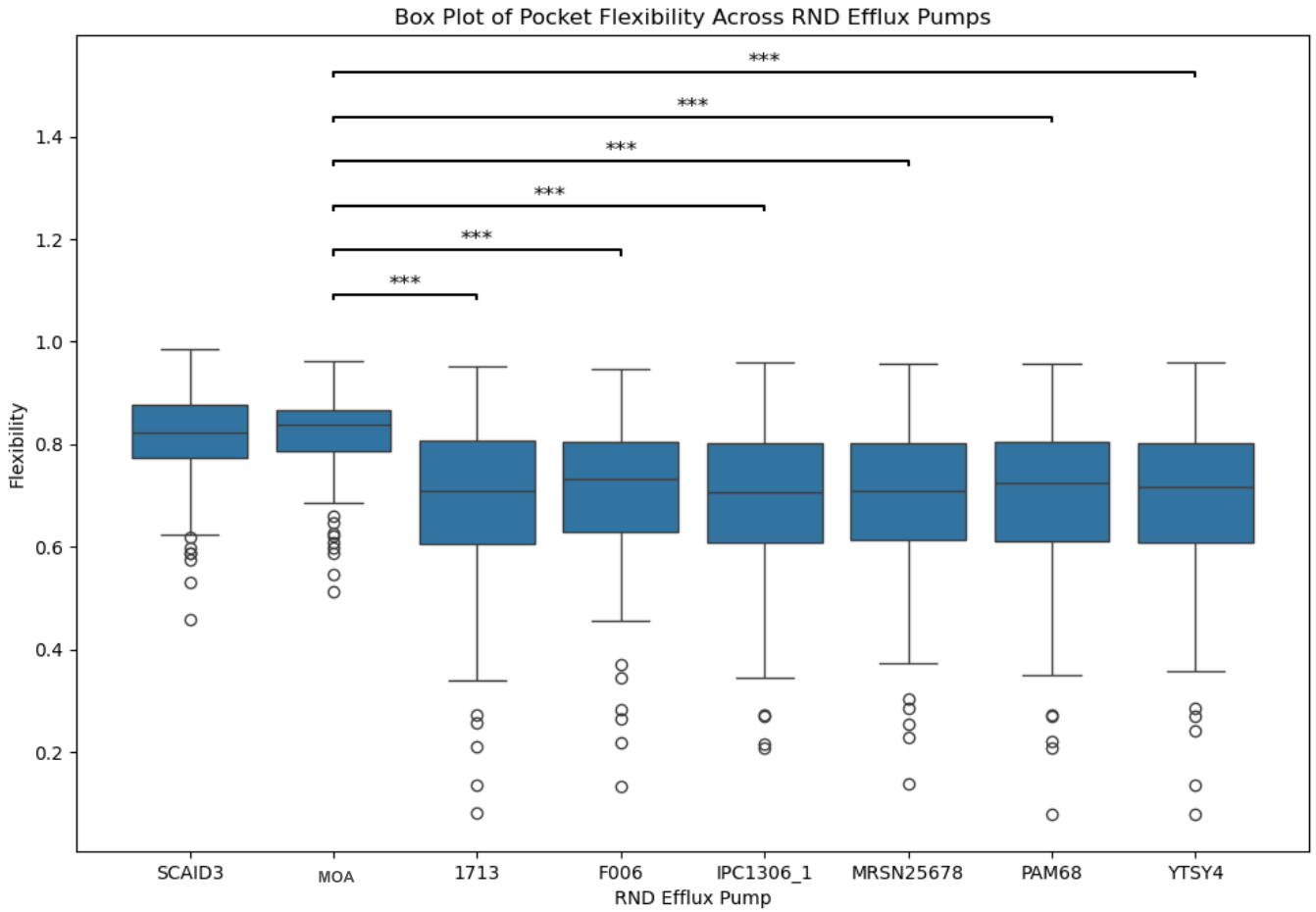


Figure 61 - Analysis of Mean-Variance in fPocket Flexibility Score for Cluster 24, MOA (MdtA) and SCAID RND Efflux Pumps. MOA (MdtA) showed a significant increase in ligand flexibility when compared to the RND efflux pumps of the *P. aeruginosa* strains of cluster 24.

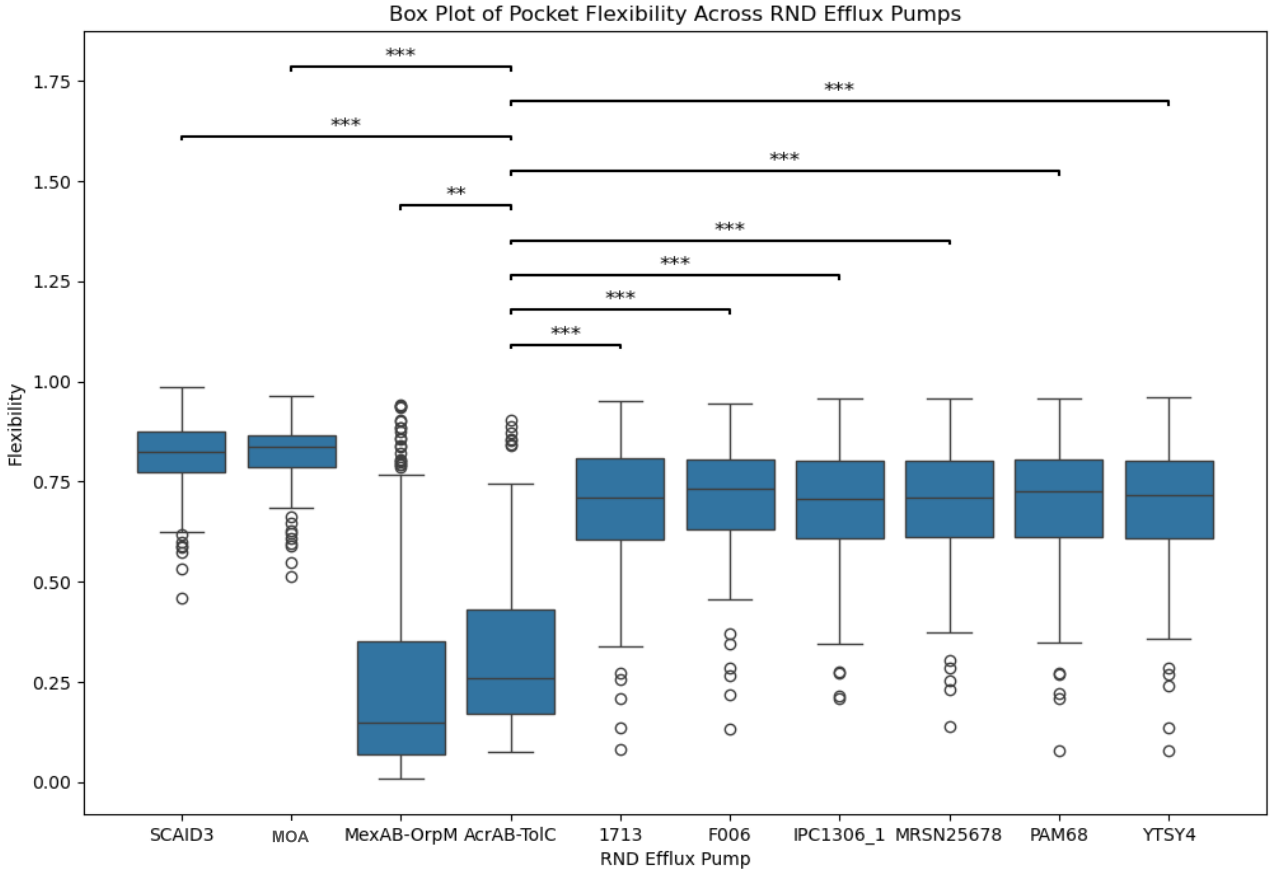


Figure 62 Analysis of Mean-Variance in fPocket Flexibility Score for Cluster 24, MOA (MdtA) and SCAID RND Efflux Pump. Including X-Ray Resolved Structures of MexAB-OrpM and AcrAB-TolC

6.6.14 Fpocket analysis of RND Efflux Pumps Unique to RCS09A Chromosome Identified in Chapter One

In the previous chapter, four RND efflux systems were identified to be unique to the chromosome of RCS09A and not present on the chromosome of MC-G-03. As RCS09A is the most BMA-tolerant isolate, these RND efflux systems had their structures predicted using the SWISS model and were analysed using fpocket to understand if they also showed a significantly higher ligand flexibility score when compared to the other pumps studied. A one-way ANOVA was conducted on the fpocket results with a Tukey HSD to show the significance between efflux systems.

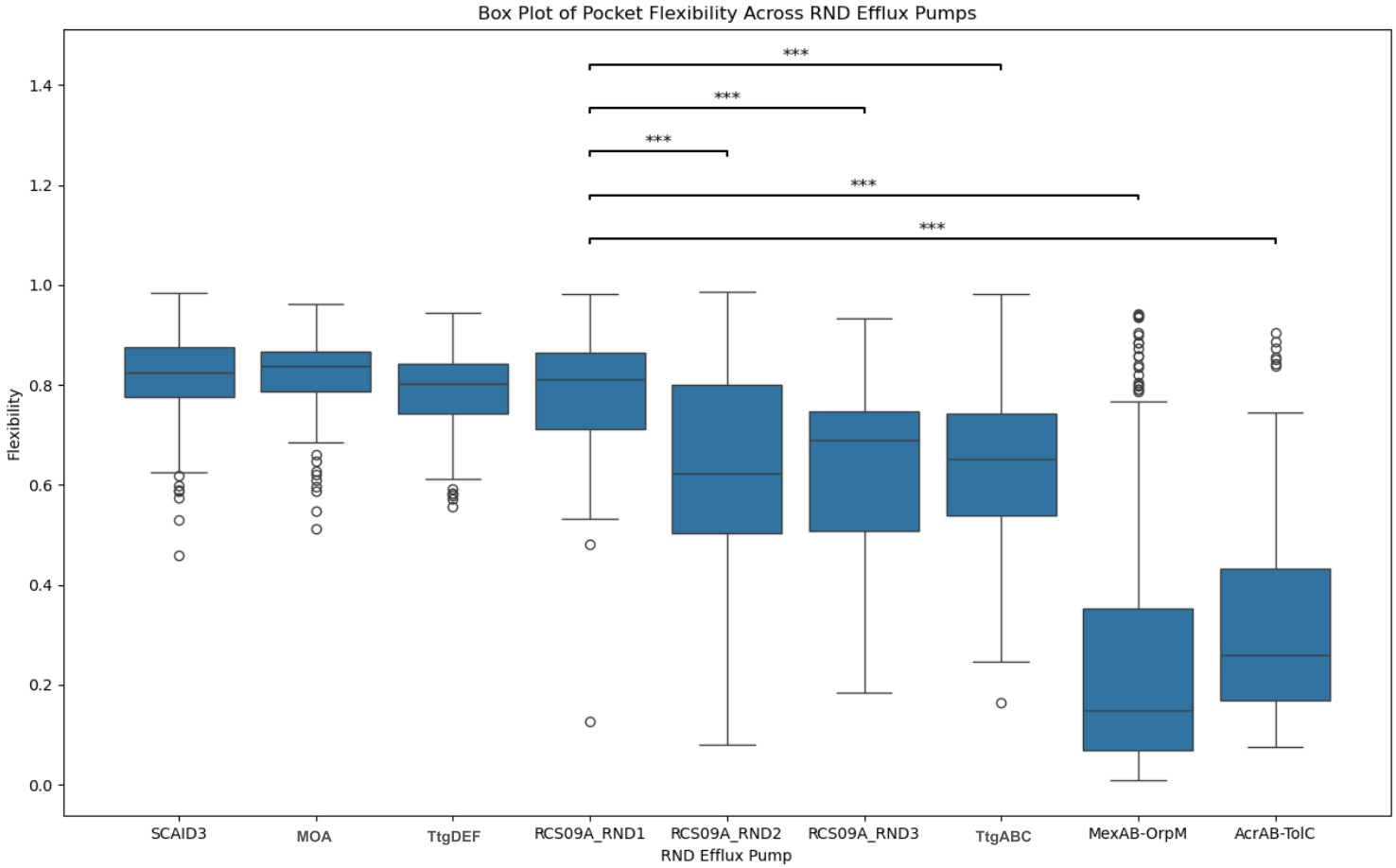


Figure 63 Analysis of Mean-Variance in fPocket Flexibility Score for RND efflux pumps unique to RCS09A. RCS09A_RND1 is shown to have a significantly higher mean flexibility score than other analysed unique pumps RCS09A_RND2, RCS09A_RND3 as well TtgABC, Mex-OrpM and AcrB-TolC

6.6.15 ANOVA and Tukey HSD results of fpocket Flexibility Score - RCS09A Unique RND Efflux Pumps Compared to other studies systems.

Out of all three of the RND efflux systems identified to be unique to RCS09A when compared to other BMA tolerant isolates (Table 6). RCS09A_RND1 was shown to have pockets with significantly higher flexibility scores when compared to RCS09A_RND2, RCS09A_RND3, TtgABC, MexAB-OrpM and AcrAB-TolC. The pumps associated and proposed to confer solvent tolerance showed no significant difference in pocket flexibility score compared to RCS09A_RND1. This identifies RCS09A_RND1 as another RND efflux system of interest.

6.7 Discussion

For the production of BMA and other hydrophobic compounds using a fermentation process, the toxicity issue must be addressed. Hydrophobic compounds interfere with cellular membranes and other internal processes (Sikkema et al., 1995). Hydrophobic properties also make these compounds attractive for bio-reactor production as they form an aqueous phase separate from the main culture body and require dramatically fewer purification steps. RND efflux pumps can contribute to the tolerance of hydrophobic compounds through the removal of toxic compounds into the environment ideal for a bio-fermentation process (Blair et al., 2014; Faizal et al., 2005b). *Pseudomonas* species have been shown to grow well in hydrophobic solvents and have been demonstrated to switch transcription

of one efflux pump in favour of another when challenged with a hydrophobic compound: for example, *P. putida* DOT-T1E upregulates transcription of *ttgDEF* in favour of *ttgABC* when treated with of toluene (Duque et al., 2001; Faizal et al., 2005b). Suggesting that this RND efflux pump has characteristics better suited for toluene.

As the pPp65-MOA plasmids RND efflux operon BLAST NCBI search aligned with another RND efflux operon on a plasmid, which also conferred a solvent-tolerant phenotype (Weiser et al., 2019a), it was investigated as an RND efflux pump suited to hydrophobic compounds such as BMA. Removing the pPp65-MOA plasmid from *P. putida* RCS09A resulted in an impaired BMA tolerance. This suggests that this plasmid is largely responsible for BMA tolerance and mechanisms on encoding on the chromosome provide its base tolerance which could be RCS09A_RND1 efflux pump identified as unique RND efflux pump with feature similar to MOA. To investigate pump influence on tolerance without the rest of the plasmid, it was decided to transfer the whole operon onto the chromosome of *P. putida* KT2440 using a mini-TnT7 system. This resulted in an increased BMA tolerance, also supporting the notion that this pump is responsible for BMA tolerance in RCS09A. This pump would be an ideal candidate operon to increase the tolerance of the tolerance of a BMA production strain and may contribute to improved product titre. This also suggests that the MOA pump has characteristics that are better suited for hydrophobic compounds such as BMA.

Current methods of identifying RND efflux pumps suited to hydrophobic compounds include either removal, silencing, or RND operon Mutant creation and studying sensitivity (Ramos et al., 1998). Such as the methods employed with the MOA pump. Other options are RNA-Seq experiments, which Identified the TtgDEF efflux pump (Duque et al., 2001). These methods can be expensive and labour-intensive and will only produce results if the organism is equipped with such a pump. It was hypothesised that in-silico methods of protein analysis could identify unique features distinguishing the TtgDEF and MOA efflux systems from other less suited to hydrophobic compounds. Once these features are identified, they can be used to identify other RND efflux systems from online databases, which can also be used to enhance tolerance to hydrophobic compounds.

To understand the internal environment of TtgDEF and MOA, they were modelled in PyMol, and their electrostatic properties were projected onto the surface of the proteins. They were then viewed from the bottom through the pore. They were compared with MexAB-orpM and AcrAB-TolC, known to export antimicrobials (Blair et al., 2014; X. Z. Li et al., 1995). They were also compared with TtgABC, which has been shown to facilitate solvent tolerance in *P. putida* KT2440 (Personal communication – Tucker Group); it has been demonstrated to be downregulated in the place of *ttgDEF* in *P. putida* DOT-T1 (Duque et al., 2001). Suggesting that TtgABC might not be the best suited when better alternatives are available.

The internal electrostatic composition of the MOA efflux pump is unique compared to other efflux pumps that have been shown to confer solvent tolerance. The varied motif in the upper channel was most similar to TtgDEF, but they had different electrostatic properties within the lower portion of the channel. When electrostatic properties were projected on the protein surfaces, MexAB-orpM, TtgABC, and AcrAB-TolC harboured a primarily negative or neutral internal environment.

As these differences were observational, fpocket was used to detect pockets within each protein and asses their properties, such as hydrophobicity, charge and polar atoms, which could then be compared across pumps using ANOVA, which would give quantifiable metrics to distinguish TtgDEF and MOA from the other compared (Le Guilloux et al., 2009a).

Fpocket analysis was run to characterise each pump's internal pockets, and the external pockets, which are typically hydrophobic so that they can interact with the two membranes of the Gram-negative cell envelope, were removed (Fuglebakk & Reuter, 2018). The comparison of the measured fpocket metrics significantly showed MOA and ttgDEF internal pockets to be the most versatile in ligand flexibility and with higher scores than MexAB-orpM, TtgABC, and AcrAB-TolC. Suggesting that ligand flexibility could be the metric used to distinguish pumps suited to hydrophobic

compounds from others. The mean value of pocket charge for TtgDEF, TtgABC and MOA were neutral, suggesting they could be better suited for hydrophobic substrates.

To further understand if the characteristics identified in TtgDEF and MOA could be identified in other efflux systems without predicting the structure and applying the various projections and analysis and protein sequence alignment using BLAST will return many similar sequences. The two pumps described herein, associated with solvent tolerance, have little shared sequence identity. As (1) TtgDEF is utilised when *P. putida* DOT-T1E is challenged with toluene, a similarly hydrophobic molecule to BMA, and (2) the addition of the MOA efflux system to KT2440 increased tolerance to BMA, both of these pumps are efflux pumps seem suited to hydrophobic compounds, and they could be used as points of interest in a clustering experiment.

We downloaded a database of unique outer membrane subunits of RND efflux systems from *P. putida*, *P.* veronii, *P. fluorescens*, *P. aeruginosa* and other *Pseudomonas* species.

First, we used the physiochemical properties of these proteins analysed by Bio-Python and clustered them using a two-component principal component analysis clustered using agglomerative clustering. The protein neighbouring the outer membrane subunit of TtgDEF was identified as

MexFE-OprN, belonging to *P.* fluorescens Q8r1-96. The neighbour of the MOA system was ANI52700.1 from *Pseudomonas* sp DR5-09, the structure of which could not be predicted. We then used a three-component principal component and clustering analysis, which Identified an outer membrane subunit of an RND efflux system belonging to *P. Aeruginosa* Strain 83. These pumps share the internal charge motif identified in TtgDEF and MOA (MdtA) and scored significantly higher in ligand flexibility than most other studied systems but not MOA (MdtA).

The physiochemical data is a basic interpretation of the protein sequence with minimal complexity to enable clustering. We converted the protein sequence into a data frame, replacing the protein letter with a numerical value for its charge and hydrophobicity. We clustered the three principal component data using DBSCAN. MOA (MdtA) placed poorly in this experiment, perhaps because its sequences are unrelated to other *Pseudomonas* efflux systems. The neighbour of TtgDEF identified an outer-membrane subunit from a *P. Aeruginosa* SCAID strain annotated as *MexAB-orpM* but shared little sequence identity with the *MexAB-orpM* from *P. aeruginosa* PA01 also used in this study. The ligand flexibility score showed no significant difference between MOA (MdtA) and TtgDEF. If this metric is linked to its ability to efflux solvents, the SCAID Mex system would be an ideal pump for testing this.

To test if the process of predicting structures using the SWISS Model has a bias to predicting structures with pockets that score higher in the ligand flexibility score, we took the outer-membrane subunits that clustered the furthest away from TtgDEF. These systems all belonged to strains of *P. aeruginosa*, and their efflux operons were complete, and structures could be predicted. They score significantly lower in ligand flexibility when compared to MOA (MdtA) and the SCAID efflux pump. However, they scored significantly higher than the two systems whose structure was resolved using X-ray crystallography, suggesting that there could be a bias. Even with this bias, it can still detect a significant increase amongst those demonstrated to contribute to solvent tolerance.

The limitation of this approach is that the database was generated automatically using broad search terms, resulting in junk proteins being added to the database. The species search could also be expanded, which could result in subunits clustering more closely to MOA (MdtA). During the time of this study, AlphaFold 2 will max out a high-spec GPU when trying to predict the structure of an efflux pump, and AlphaFold 3 will only accept protein sequences of less than 5000 amino acids, but as of November 2024, AlphaFold 3 is available to download which could overcome the limitations of the online server (Callaway, 2024). Until this is determined, the SWISS Model is the best alternative, limited to structures of similar sequences identity already being resolved via X-ray crystallography, which would restrict more obscure efflux systems being checked.

In Chapter One, four RND efflux systems were identified as unique to RCS09A in the pan-genome database of RCS09A, MC-G-03 and KT2440. Three of the four efflux pumps had their structures predicted and underwent fpocket analysis, and RCS09A_RND1 was shown to have a similar flexibility score to solvent conferring systems. Its pocket scores were significantly higher than those of other studied systems.

6.8 Future Work

To further profile the MOA (MdtA) efflux pump activity, an Omnilog experiment could be run using a chemical sensitivity plate to screen KT2440 wild type versus the KT2440 MOA (MdtA) positive strain to see which compounds the addition of the efflux system provides protection. Also, testing its ability to efflux other solvents used in the plastics industry, such as styrene and ethyl benzene, would be advantageous to determine if it's a pump that acts broadly on a range of hydrophobic solvents with industrial applications.

The addition of the MOA (MdtA) operon to the chromosome of KT2440 increased BMA tolerance. This demonstrated that this pump, in a single copy, increased tolerance to replicate the tolerance of RCS09A. Further, we could assemble its original plasmid, including a selective marker, so that the entire plasmid could be transformed into KT2440 to see what additional

tolerance this could provide. This would also open the investigation into what other systems the plasmid holds confer tolerance.

Further expanding and curating the database of efflux outer membrane subunits to see if our clustering methods could be used to identify an efflux system with a higher ligand flexibility score than that of MOA (MdtA) to determine if this metric can be linked to the ability to efflux solvents. Synthesising the SCAID *MexAB-OrpM* and RCS09A_RND1 operon into the same TnT7-Mini system we used to add the MOA (mdtA) operon to KT2440 to see if the addition of these operons provides a similar level of increased tolerance would also be another way to validate if the ligand flexibility translates to solvent efflux and tolerance.

7 Results Chapter 3 - Unsupervised Machine Learning for Clustering of Natural and Synthetic Molecules: Identification and Application of BGC Resistance Genes for Enhancing BMA Tolerance

7.1 Introduction

7.2 Clustering of BMA and Biosynthetic Molecules

This Chapter presents the results of using unsupervised machine learning to cluster 3D representations of molecules synthesised by BGC while including BMA. The molecules will be rotated in PyMol around the X and Y axis and converting each rotation into a numerical dataframes and then combining these into a multilayer dataframes which will be then converted into principal components which will be then clustered using various algorithms to try to identify molecules with similar features. Determining if similar molecules to BMA are made in nature and to point to a resistance gene within a BGC that could have some application for intracellular tolerance to BMA.

At the annual Mitsubishi Biotechnology conference, it was suggested that BMA may be 'too toxic' regardless of how well it can be effluxed from cells (Mitsubishi Chemical, Personal Communication). When producing a toxic molecule, nature solves this problem by including a resistance gene within the biosynthetic gene cluster (BGC) that programs the synthesis of the molecule (Stegmann et al., 2015a). Designing a resistance gene for BMA is outside this research group's expertise and would entail another study in

computation biochemistry. However, machine learning might be an alternative approach to finding a solution to this problem.

7.2.1 Biosynthetic Gene Clusters

Biosynthetic gene cluster resources such as MIBig are an ever-growing dataset containing DNA sequence data, protein sequence data, and molecules synthesised by each BGC (Terlouw et al., 2023a). Machine learning, when applied to convert these data into a dynamic tool that derive a DNA sequence as an output to direct the biosynthesis of a specified molecule, would have vast applications for medicine, industry, and society. So far, the number of BGCs that have been wholly categorised is 2502 (as of June 2024) (Terlouw et al., 2023b). Thus, the data needed to train such a model to operate in this fashion is still in its infancy. The training of AlphaFold to predict protein structure from amino acid sequences used a training dataset of over 100,000 proteins (*AlphaFold - Google DeepMind*, n.d.). MIBig product structural data could be used more simply instead of using complex genetic or protein data. Using unsupervised machine learning, we can cluster BCG products' chemical structures and include synthetic molecules such as BMA. We can use this clustering approach to link molecules that share common features across chemical classes (Holliday et al., 2004).

Molecules with similar features to BMA may have resistance genes to protect the producer organism within their associated BGC. Using machine learning protein structure predictions and molecule docking approaches, it may be

possible to identify proteins that bind BMA and assess their potential to confer resistance.

The metallothionein-like proteins bind heavy metals. MymT, found in *M. tuberculosis*, binds copper ions, sequestering the toxic ions and protecting the bacterium engulfed by macrophages (Gold et al., 2008). With this in mind, it may be possible to find a protein that can similarly sequester BMA and reduce its ability to disrupt internal cellular function.

7.2.2 Supervised Machine Learning

There are various machine learning algorithms (MLA), such as 'supervised'. Simplistically, this is when inputs or 'features' and outputs or 'labels' are known and readily available. Most famously used in image recognition, many images (features) for an object are presented to the MLA, and after much training, it can categorise the image as the given label (Krizhevsky et al., n.d.). This does not apply to this study's approach as we do not need to label our features; we only need to know where BMA sits within a larger dataset.

7.2.3 Unsupervised Machine Learning Algorithm

For Unsupervised machine learning (USML), only *'features'* are known and the data are sorted by leveraging inherent patterns and clustering intrinsic similarities. USML is used to highlight outliers within a data set (Singh & Vanschoren, 2022). USML can be used to group all the molecular structures in the MiBig database and incorporate BMA into this dataset.

The USML learning algorithm can group the data in a way a biased human may not.

Sci Kit Learn is a Python package that allows for the simple implementation of various unsupervised machine-learning techniques suited to multiple applications and hardware limitations. These include Incremental Principal Component Analysis (IPCA), K-means and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Pedregosa et al., 2011). Other packages include Torch.

7.2.3.1 Incremental Principal Component Analysis

Due to hardware limitations, Incremental Principal Component Analysis (IPCA) was used to process the data. This works by processing smaller batches of data at a time, rather than the whole dataset in one go. Through trial and error for this study, the processing of batches of 50 molecules was used. IPCA calculates the principal components for each batch of 3D rotations of a molecule converted in a multilayer dataframes. For each batch, the sequential principal component for the data set is updated and allows large datasets to be processed without using vast amounts of computer memory. The number of components that ICPA calculates is predefined by the user. This can influence the results as the number of components determines how the data are transformed to reduce dimensionality. It can also affect how the processed data are visualised. For example, with two principal components, the data can be visualised

along an X and Y axis and for three principal components along the X, Y and Z axes (Pedregosa et al., 2011).

For this study, we used two, three and five components. The five-component analysis can be problematic to visualise as we run out of plottable axes. To overcome this issue, the 4th component is represented as a variance in colour within the primary cluster colour, and the 5th is shown as a variance of point size where a double size point represents the maximum value and the minimum is represented by a half point size. This ensures that no datapoint is unviewable.

A disadvantage of this method is that it is designed to remove dimensionality, and when working with 3D structures, this can remove important information.

7.2.3.2 K-Mean Algorithm

The molecules from MIBig are already categorised into seven groups: 1) polyketides, 2) non-ribosomal peptides, 3) terpenes, 4) saccharides, 5) alkaloids, 6) ribosomally synthesised and post-translationally modified peptides, and 7) others (Terlouw et al., 2023a). K-Means seems an appropriate algorithm as the number of clusters needs to be predefined. After the data is transformed to their reduced dimensions by IPCA, the K-Means algorithm randomly allocates datasets or molecules as the leaders for a category or cluster. As each batch is processed, these are updated dynamically. A

disadvantage is that, due to the number of clusters being predefined by the user, every data point is grouped, leading to poor identification of outliers (Shi et al., 2010).

7.2.3.3 DBSCAN Algorithm

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering algorithm that groups datasets based on their neighbour distance. The user predefines the neighbour distance at which datasets can be considered members of the same group. Within this set distance, the user must also predefine how many minimum samples must be similar to form the core dense region. The dimensionality of the data usually determines this. For this study, with a dataset of over 3000 arrays reduced to 2 dimensions, minimum samples (minimum is two) are multiplied by two (data dimension) and then by 10 to account for the more extensive data set to avoid smaller clusters (min_samples=20). The advantage of this method is that the algorithm will calculate the clusters, which allows for better outlier detection as it will not group these into a cluster. This differs from K-means, where the clusters are predefined, and every array is grouped into a cluster. This might not be advantageous for this study as we want our synthetic compounds to be clustered with BGC products. As the user predefines the neighbour distance and minimum samples, this can result in poor clustering depending on the user's understanding of the data and the overall algorithm (Ester et al., 1996b).

7.2.4 Data preprocessing

To collect data on the molecules produced by biosynthetic gene clusters, the minimal information about a biosynthetic gene cluster (MIBig) database was downloaded in JSON format. This contains a file for each entry in the database, which holds information including chemical products - their PubChem accession number, SMILES chemical structure and chemical name. A custom Python script used this information to check the PubChem accession number for a 3D structure within the PubChem online database. If one exists, it was downloaded and stored in a folder in .sdf format. If no 3D structure was available for a given molecule, the Python package RDKit was used to predict a 3D structure. If both approaches failed, the script would then use the chemical name to search the PubChem database for a 3D .sdf file. If all three approaches failed, the entry would be removed from the database being created.

Once this process was completed, 3788 3D files were generated. Only 879 BGCs failed to generate any 3D structures (Figure 64). To convert each 3D structure into numerical matrixes, they were first converted into images. A Python script utilising PyMol generated 120 images of each molecule, 60 moving across the Y-axis and 60 moving across the X-axis. PyMol must be correctly configured to ensure that each molecule appears from the same perspective and distance; if a molecule is zoomed to fit the aperture, this can warp the results. For example, a larger complex ring-shaped molecule could be grouped with a smaller ringed molecule if it were zoomed to fit the

aperture. It must also be configured to use the same colour scheme for each image generation. To reduce the amount of data being processed, each molecule was coloured so that when reduced to greyscale, each atom was a distinctive grey scale. The image data were then converted into a NumPy array.

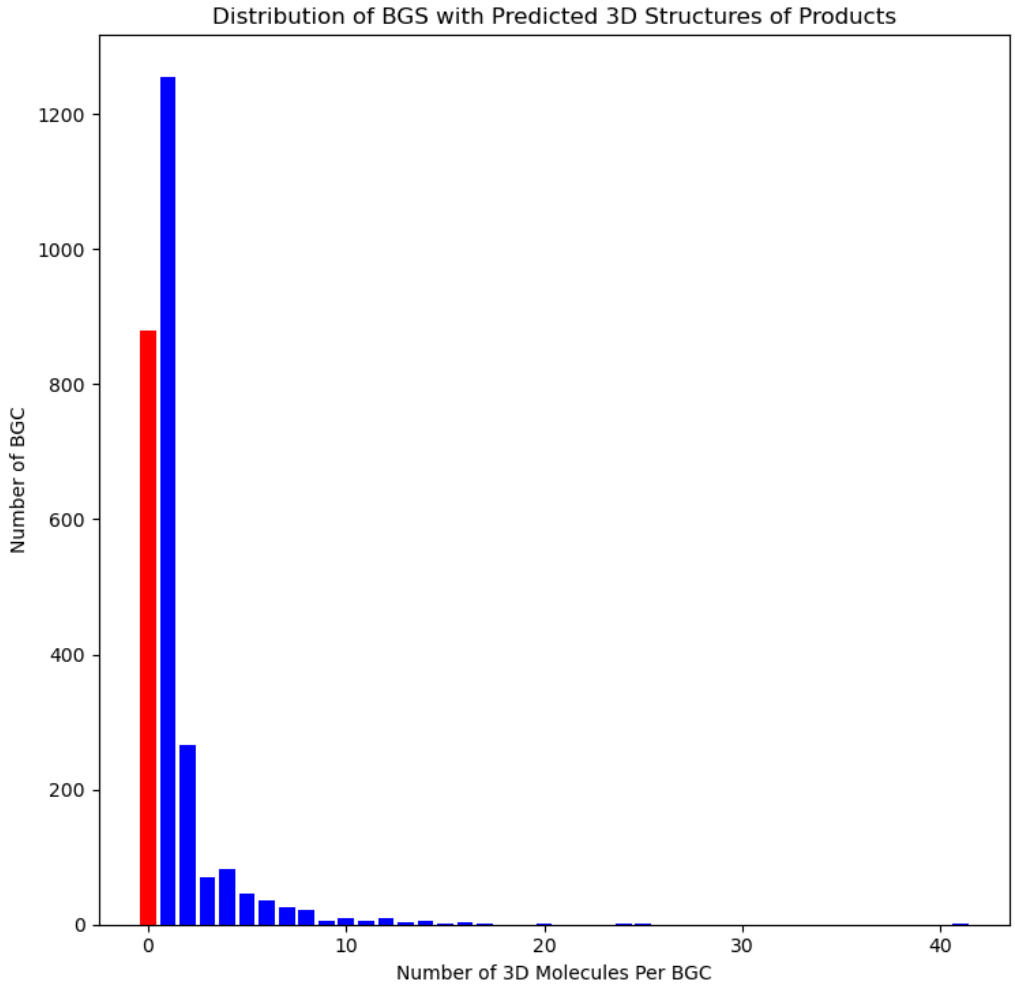


Figure 64 – Distribution of Biosynthetic Gene Clusters and the Number of 3D structures generated from each in Blue (2909). BGC that could not facilitate the generation of a 3D structure are shown in red (879).

7.2.5 Molecules were Clustered using ICPA and K-Means

Using K-Means set to 7 clusters, as determined by the categories of BGC products on MiBig and IPCA to reduce the data to 2 principal components, the data were processed, transformed, clustered, plotted and saved to an HTML file using Bokeh. This interactive plot allows allowed for quick identification of the chemical compound plotted and the BCG to which it belongs (Figure 65).

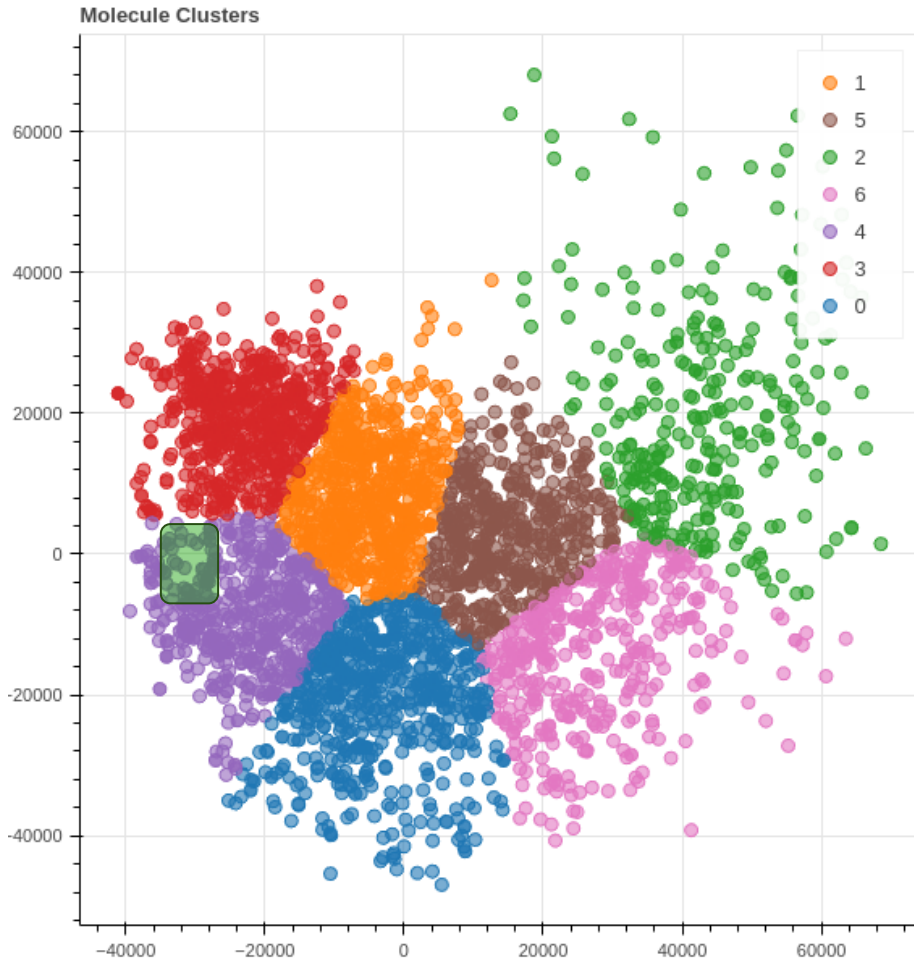


Figure 65 – K-Means 7 Clusters, ICPA 2 principal components analysis. 7 Clusters are labelled in the legend. The location of BMA is highlighted in the green box in cluster 4.

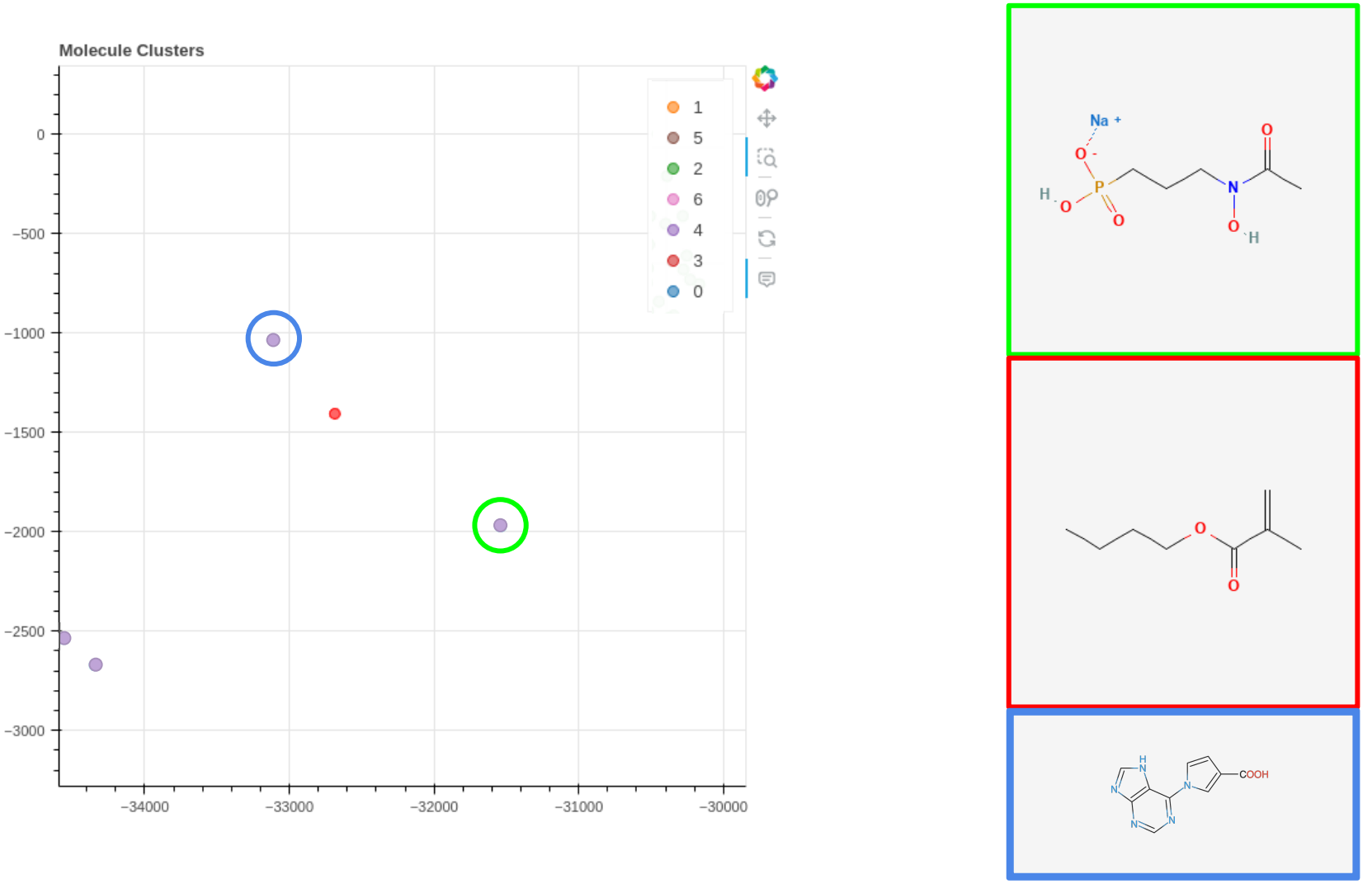


Figure 66 – K-Means 7 Clustering of ICPA 2 principal components analysis. Zoomed on Cluster 4, showing BMA (**RED**), closest neighbour fusatinic acid (**BLUE**) and its second closest neighbour compound FR-900098 (**GREEN**)

7.2.6 ICPA Two Component and K-Means 7 Cluster Analysis

The closest neighbour to BMA was fusatinic acid produced by BGC 0002163 from *Fusarium pseudograminearum*. Observationally, there are no similarities between this product and BMA (Figure 66 - BLUE). Furthermore, the BGC is poorly annotated and only comprises two genes, none annotated to facilitate resistance. The next closest neighbour is FR-900098, an antimalarial compound which visually shares some similarities. It shares a similar configuration and relative size in the 2D representation of the molecule, although the composition is somewhat different (Figure 66 - GREEN). The DBSCAN algorithm was used to validate this clustering further.

7.2.7 Molecules were Clustered using ICPA and DBSCAN

The dataset was reduced dimensionally to two principal components and clustered using DBSCAN that was set to a minimum sample of 20. This reflects a large dataset comprising two components, and the closest neighbour of 0.2 was chosen as the data produced by the K-means study was highly dense. The data were plotted using bokeh, which creates an interactive HTML plot, allowing for easy identification of close-neighbouring molecules and BGC.

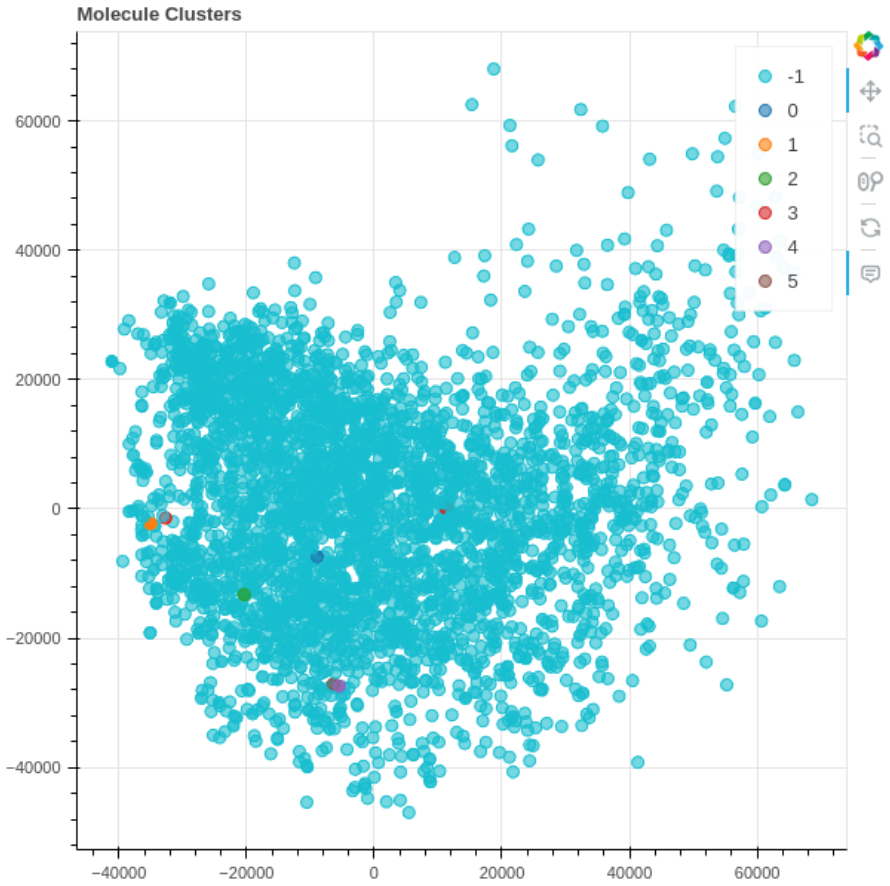


Figure 67 - DBSCAN Clustering of IPCA 2 Principal Components showing the same clustering pattern as K-Means DBSCAN has determined that most analysed molecules are noise shown in **BLUE.** The coloured points are considered groups by the algorithm. **RED** is the location of BMA within the plot.

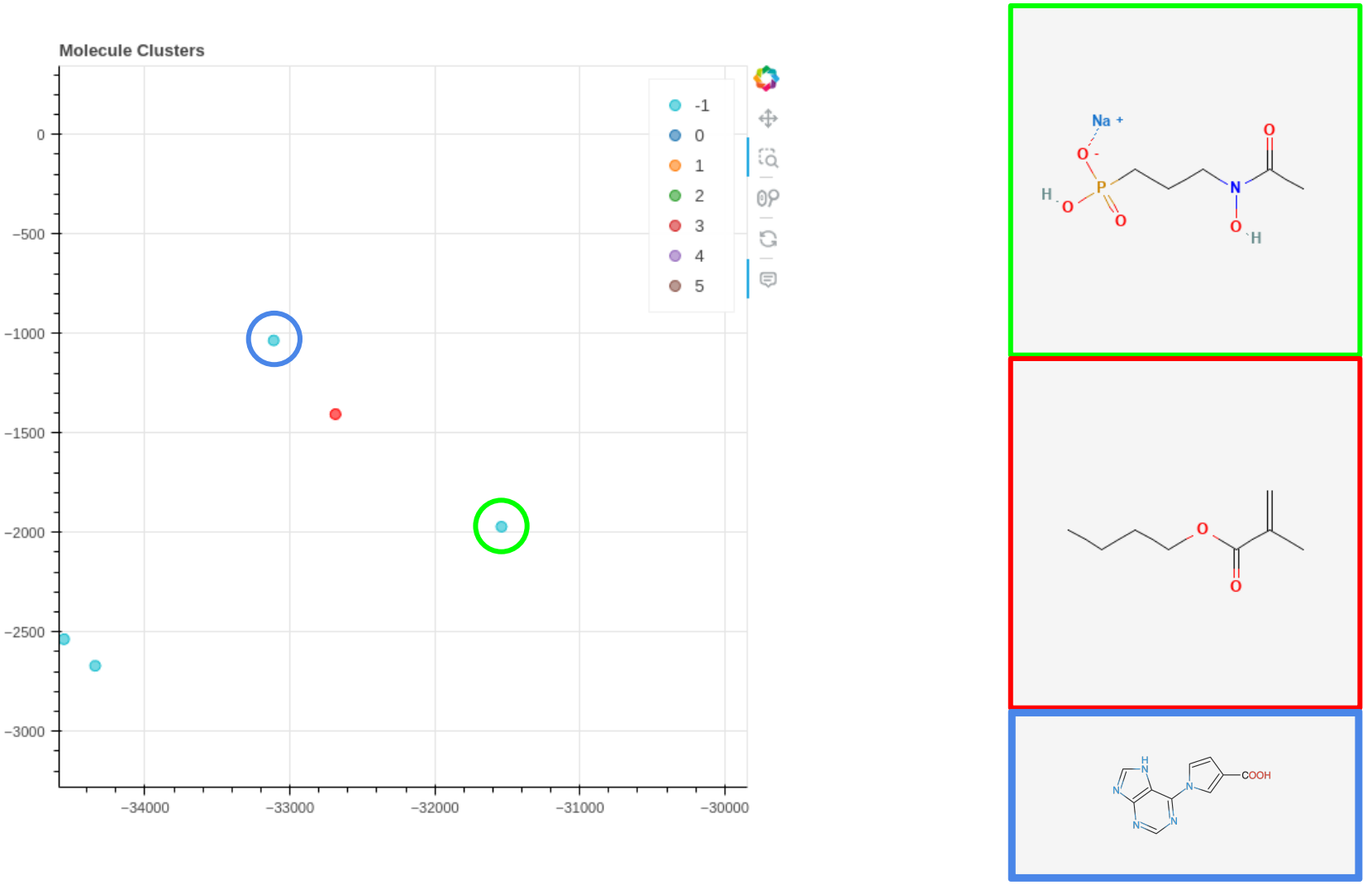


Figure 68 - DBSCAN Clustering of IPCA 2 principal components analysis. Zoomed on Cluster -1 (**NOISE**), showing BMA (**RED**), closest neighbour fusatinic acid (**BLUE**) and its second closest neighbour compound FR-900098 (**GREEN**)

7.2.8 ICPA Two Component and DBSCAN Cluster Analysis

DBSCAN confirmed the same closest molecules as K-Means. The clustering has clustered different BGCs that produce the same product as a cluster and everything else as an outlier. Alternative minimum samples and neighbouring distance experiments were conducted; they did not change the configuration of the plot, just the number of outliers.

7.3 Identification of BMA Structural Relatives in Nature

7.3.1 FR-900098 – An Antimalaria compound

The compound in proximity to BMA in the 2-component K-Means and DBSCAN clustering was identified as FR-900098 (Figure 68 and Figure 66), an antimalarial candidate compound (Wiesner et al., 2016). The compound is similar in size and profile to BMA. However, it has many different features, such as a phosphate group and a slightly different general chemical composition. Concerning protein binding, three-dimensional conformation and electrostatic properties are more critical components. To understand this, APBS Electrostatic Analyses were conducted.

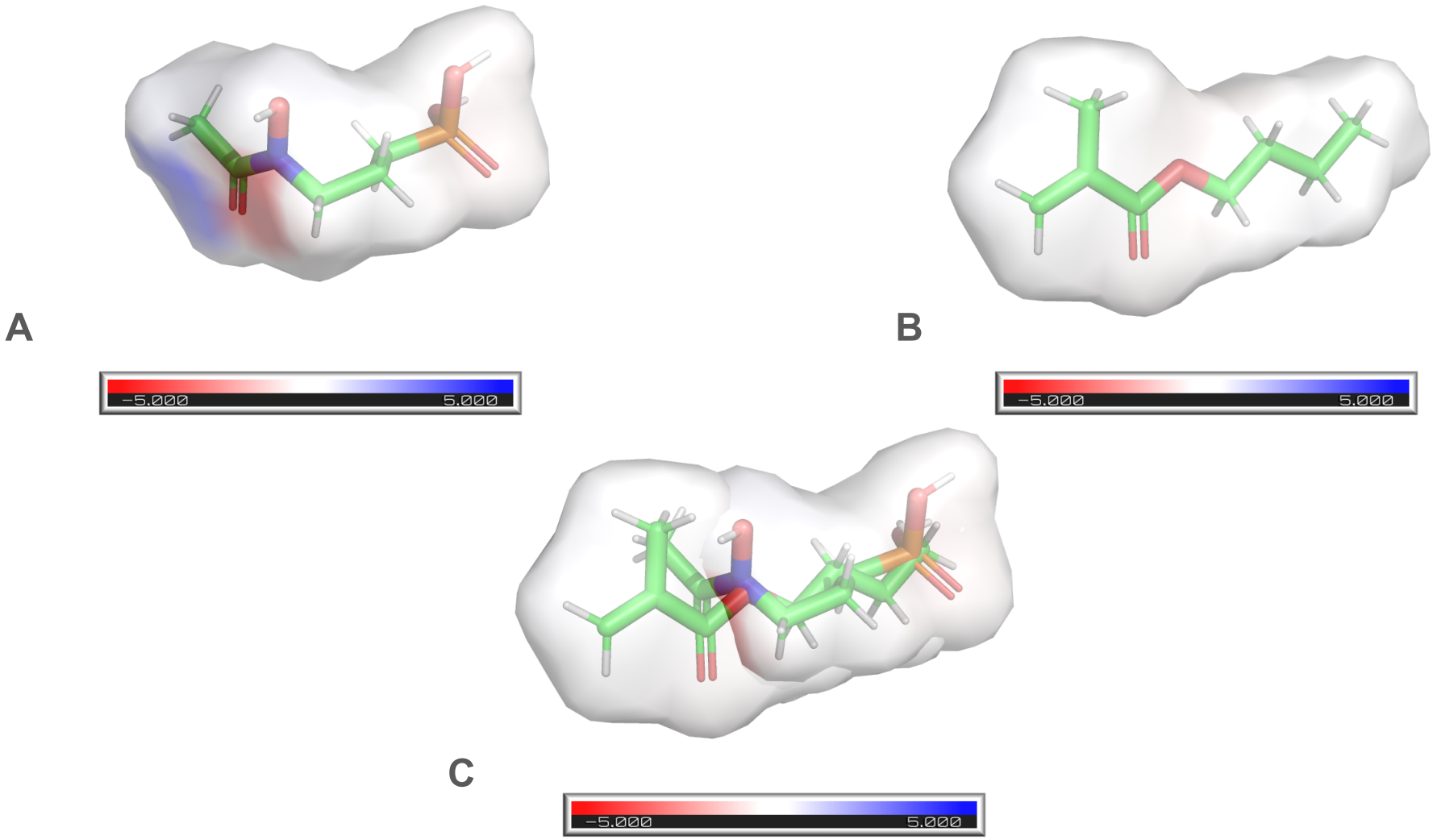


Figure 69 - BMA and FR-900098 Electrostatic Properties - **A)** FR-900098 XLogP3 -2.2 **B)** BMA XLogP3 2.9. Similar electrostatic clouds FR-900098 has a polar region on the left of the molecule (**A**). **C**) represents the two molecules aligned. The electrostatic clouds do not align but each molecule has over 4 rotatable bonds so at a different configuration they could more closely align.

7.3.2 Comparison of Electrostatic Properties of BMA and FR-900098

FR-900098 and BMA were aligned using PyMol, and their electrostatic clouds were calculated using the APBS Electrostatics plug-in (Figure 69 - C) (Jurrus et al., 2018). FR-90098 has a dipole region in the acetamide group and a hydrophilic tail XLogP3 value of -2.2 (Figure 69 - A). BMA has a neutral electrostatic cloud with a hydrophobic XLogP3 value of 2.9 (Figure 69 - B). To further investigate BGG0000904 as a resistance gene donor, the BGC was analysed.

7.3.3 Biosynthetic Gene Cluster – 0000904

The BGC BGG0000904 of *Streptomyces rubellomurinus* comprises the genes *frbA-J*; the resistance gene *frbF* is highlighted in the anti-smash annotation (resistance (resist) AAC3 (Score: 196.7; E-value: 2.5e-60)) on MiBig (Figure 70). FrbF is an N-acetyltransferase involved in the synthesis of FR-900098 (Bae et al., 2011a). To understand how the original ligand binds and how BMA may bind to FrbF, the structure of the protein was retrieved from the PDB and also predicted using AlphaFold2.

Figure 70 - MiBig AntiSMASH Biosynthetic Gene Cluster Annotation with *frbF* highlighted as the resistance gene for the cluster

7.4 FrbF Protein Structure was predicted using AlphaFold

The FrbF protein has a resolved X-ray crystal structure available (Bae et al., 2011a). To compare BMA with other proteins whose X-ray crystal structure is unavailable, we must first compare the binding energies predicted by AutoDock to the X-ray crystal structure and the AlphaFold structures. The X-ray crystal structure of FrbF is a dimer protein consisting of two chains of amino acid sequence (Bae et al., 2011a). The protein was modelled as a dimer using AlphaFold2 multimer.

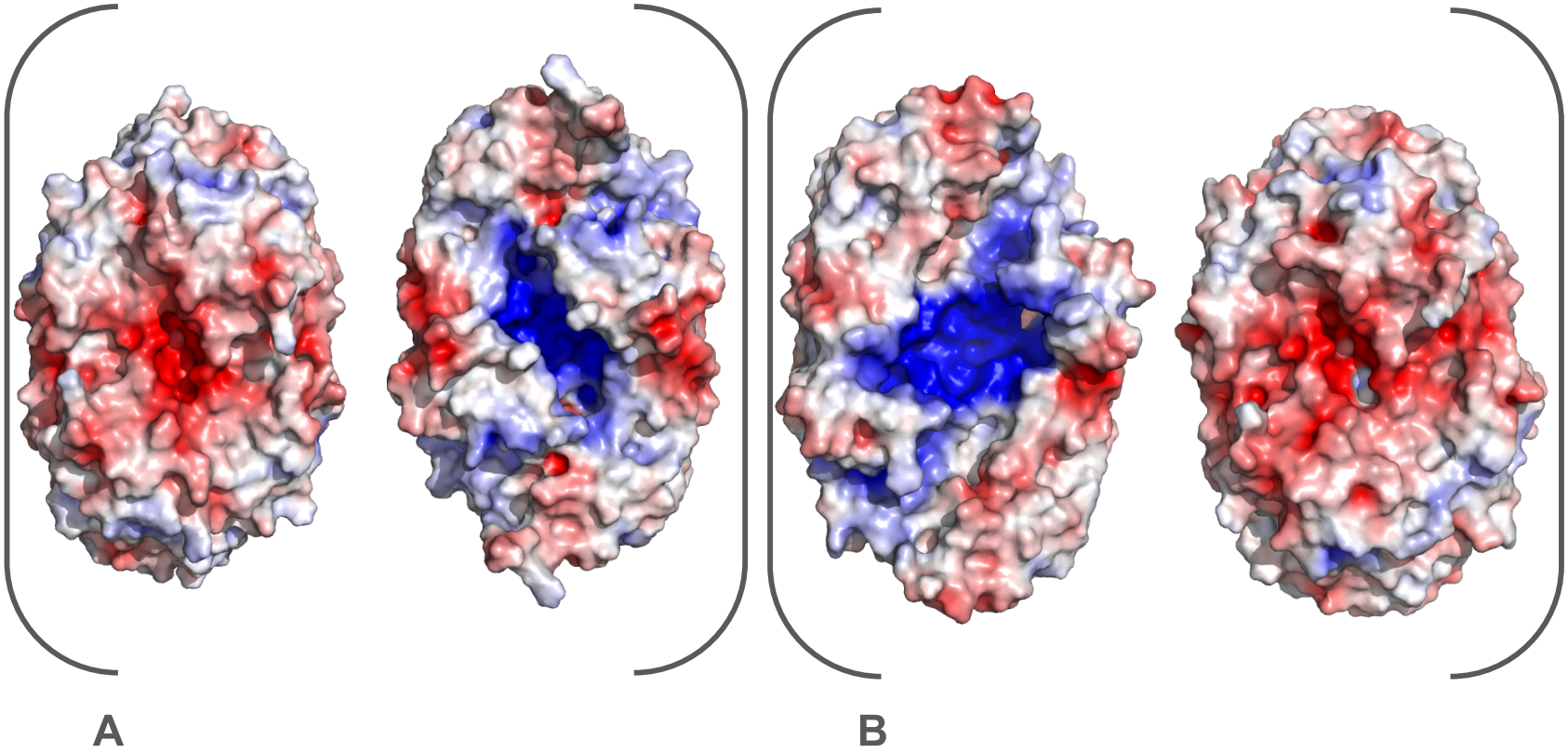


Figure 71 **A)** Alphafold2 Predicated Structure of FrbF. **B)** X-Ray Resolved Crystal Structure of FrbF (Bae et al., 2011b). Showing the visual differences between the two types of structures

Table 13 TM Score and RMSD of Alphafold2 Structure and X-Ray Resolved Structure

FrbF Structure	**AlphaFold2**	**X-Ray**
Name of Chain	A406551	B406551
Length of Chain (residues)	286	263
Aligned Length	263 residues	263 residues
RMSD	-	0.53 Å
Sequence Identity (Seq_ID)	-	1
TM-score (Normalized)	0.9133	0.99266

7.4.1 Comparison of FrbF Alphafold2 and X-Ray Resolved Structures

The protein structures appear slightly different observationally (Figure 71). RMSD (Root Mean Square Deviation) was used to verify the structure's similarity and calculate the average distance between atoms. An RMSD below 2 Å is considered a good alignment (i.e., within observed experimental variations). The two structures compared scored 0.53 Å. A TMscore above 0.5 verifies similarity, and a score of 1 would make the proteins identical. The AlphaFold2 structure scored 0. 9133, and the X-ray resolved scored 0.99266 (Table 13). This indicates that the structures are almost identical (Stevens & He, 2022). Protein-ligand docking simulations were conducted using AutoDock tools to verify that BMA binds with similar energies. AutoDock tools were set up using the blind docking approach, where the binding grid encapsulates the entire protein structure. Lamarckian Genetic Algorithm was used with 50 runs and a population of 300.

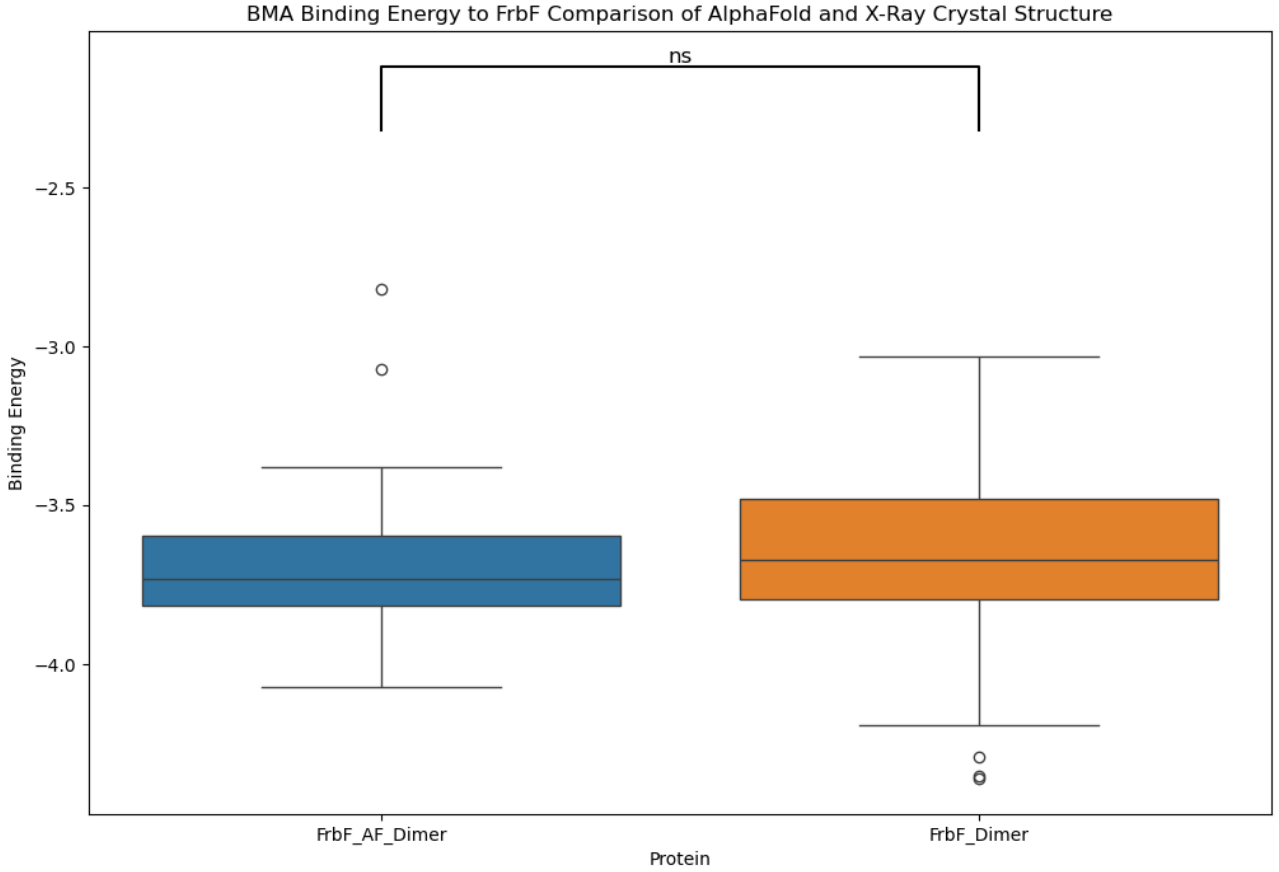


Figure 72 - Comparison of Binding Energies of BMA to FrbF AlphaFold and X-Ray Resolved Structure

7.4.2 Comparison of Binding Energies of BMA to FrbF AlphaFold Structure and X-Ray Resolved Structure

When the means were compared using a 2-sample T-test, there was no significant difference in the binding energies of the two FrbF protein structure types (Figure 72). Docking predictions were made to understand if BMA and FR900098, the BGC's product, bind similarly to the protein.

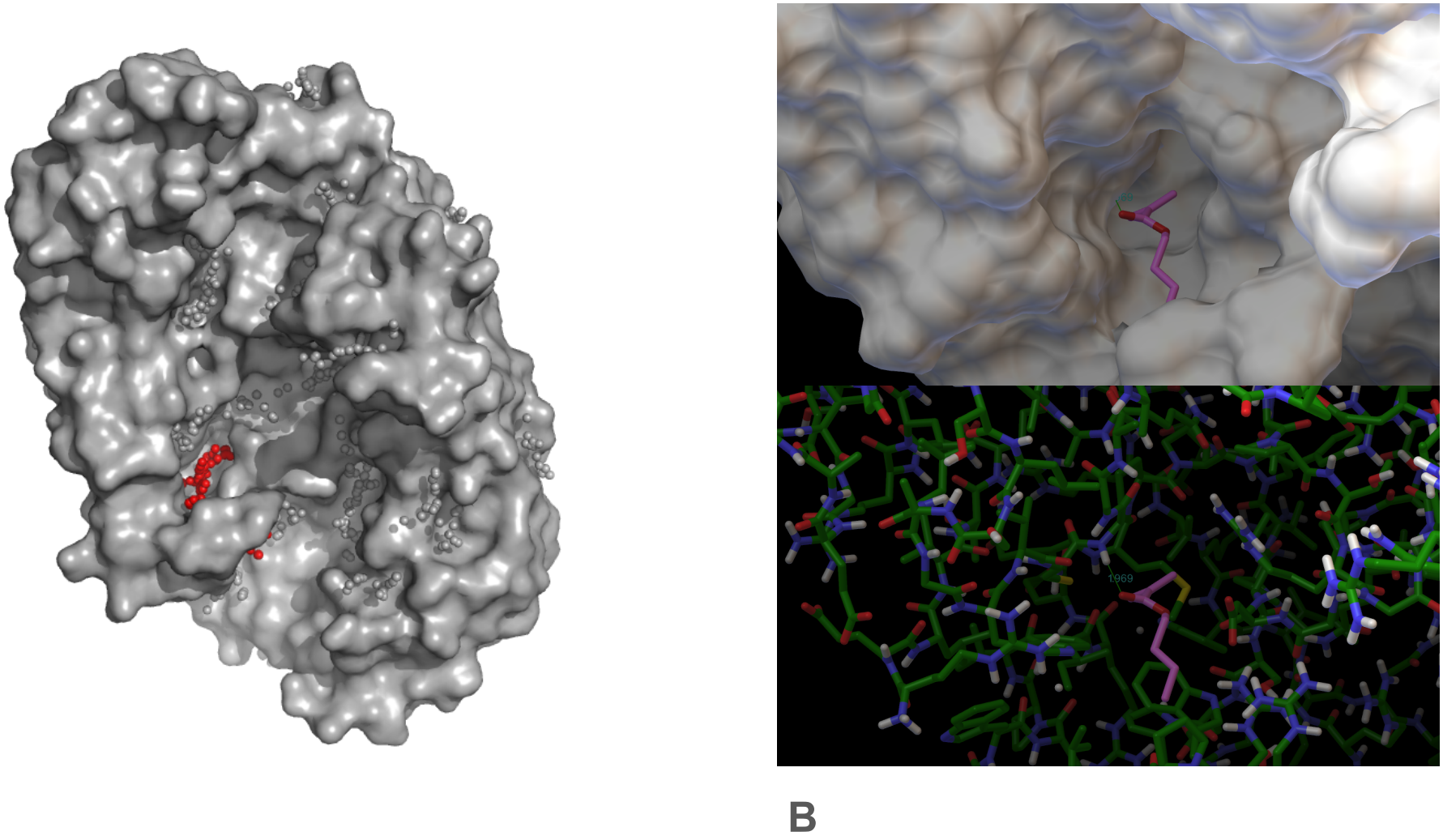


Figure 73 - **A)** Fpocket Protein Pocket Annotation. Pocket 2 in Red **B)** AutoDock Top Ranked Energy Binding Position of BMA in Pocket 2

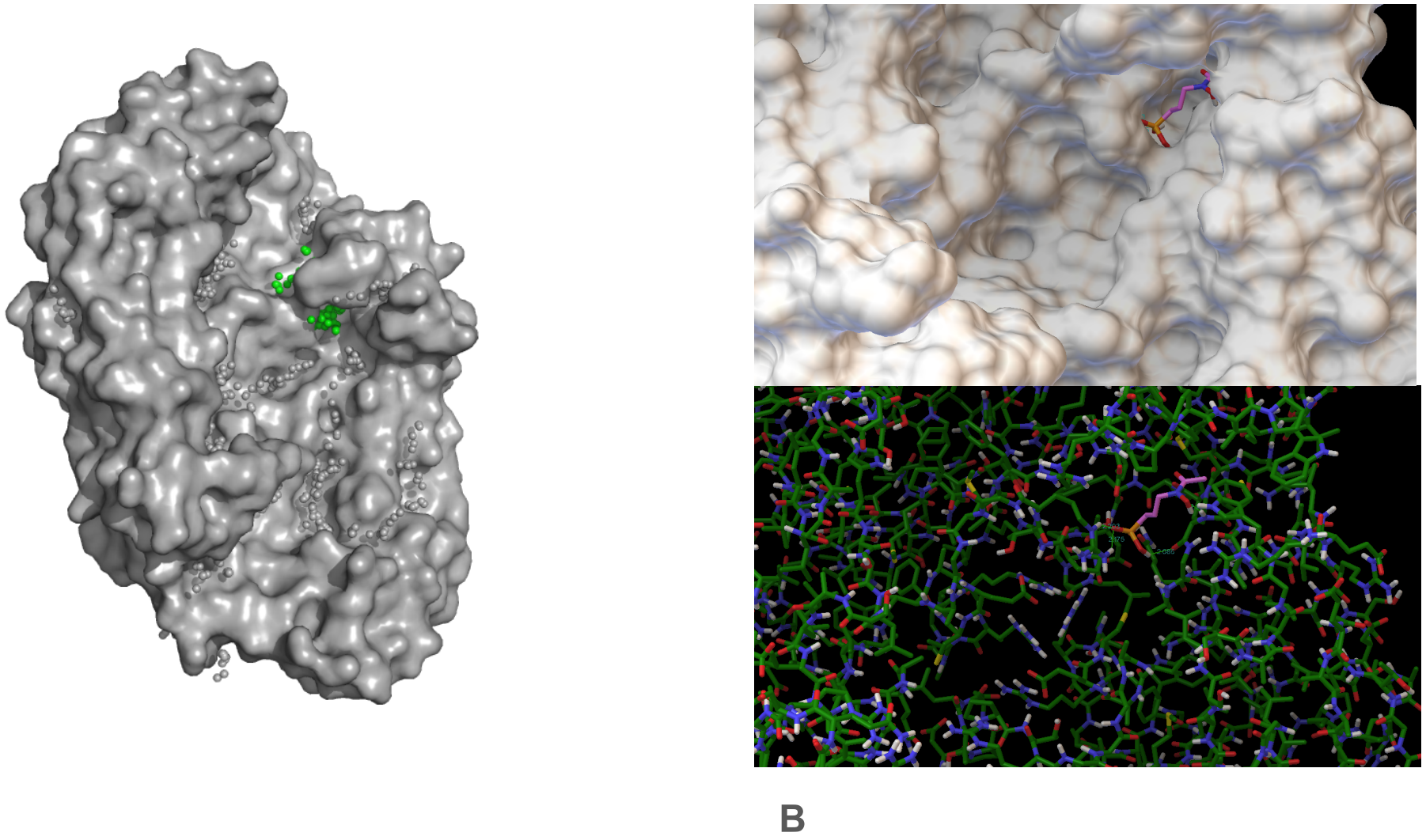


Figure 74- **A)** Fpocket Protein Pocket Annotation. Pocket 1 in Green **B)** AutoDock Top Ranked Energy Binding Position of FR900098 in Pocket 1

Table 14 - FPocket Metrics of the Binding Pockets One and Two

	Pocket 1	Pocket 2
Score	0.292	0.262
Druggability Score	0.21	0.273
Number of Alpha Spheres	98	86
Total SASA (Solvent Accessible Surface Area)	220.79	216.21
Polar SASA	106.065	134.012
Apolar SASA	114.725	82.198
Volume	843.534	831.922
Mean Local Hydrophobic Density	23.784	21.062
Mean Alpha Sphere Radius	3.777	3.871
Mean Alpha Sphere Solvent Access	0.433	0.443
Apolar Alpha Sphere Proportion	0.378	0.372
Hydrophobicity Score	14.652	14.043
Volume Score	3.826	4.087
Polarity Score	11	12
Charge Score	2	3
Proportion of Polar Atoms	40.98%	43.636
Alpha Sphere Density	6.323	6.486
Centroid of Mass - Alpha Sphere Max Distance	14.036	14.866
Flexibility	0.305	0.322

7.4.3 BMA and FR90098 Top Ranked Binding Sites are Two Distinct Pockets

Fpocket was used to predict and annotate the pockets of FrbF and their properties. BMA and FR900098 binding predictions were made using AutoDock. The top-ranked binding site for BMA is in Pocket 2, a hydrophobic pocket (Figure 73). FR900098's top-ranked binding pocket is Pocket 1, a hydrophobic pocket (Figure 74). The two pockets have similar environments regarding hydrophobicity, volume, solvent-accessible surface area, and most other measured metrics (Table 14). Most of the BMA binding sites of the 50 runs are in pocket 2. FR900098 4th ranked binding site was in pocket 2 (Figure 75).

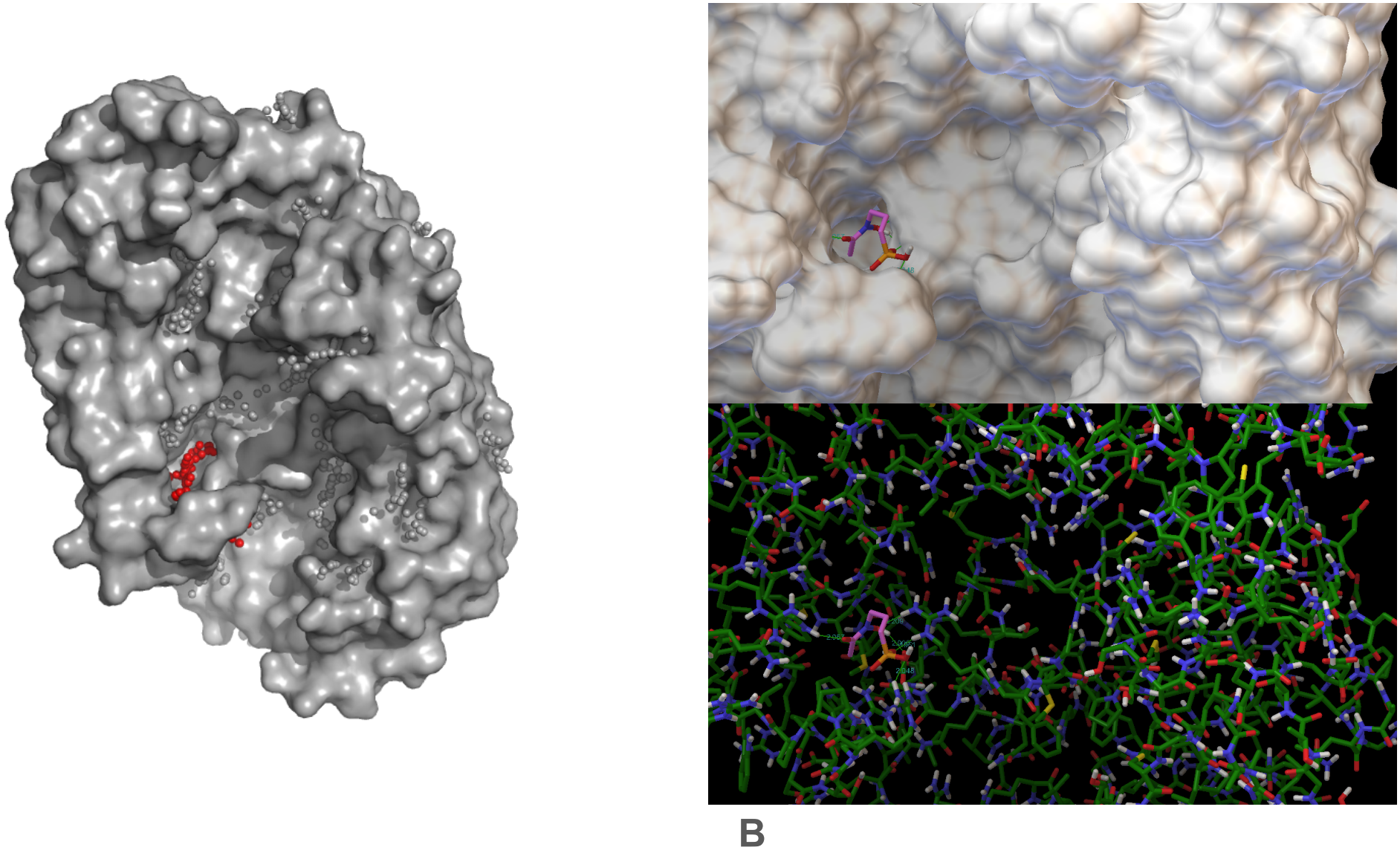


Figure 75-**A)** Fpocket Protein Pocket Annotation. Pocket 2 in Red **B)** AutoDock Forth Ranked Energy Binding Position of FR900098 in Pocket 2. Showing the similar profile of FR900098 and BMA allows them to bind in similar regions/pockets of the protein.

7.4.4 Comparison of Binding Energies of BMA and FR-900098 to FrbF

FR-900098 and BMA bind to FrbF in similar ways. To further investigate their binding energies were compared in a two-sample t-test.

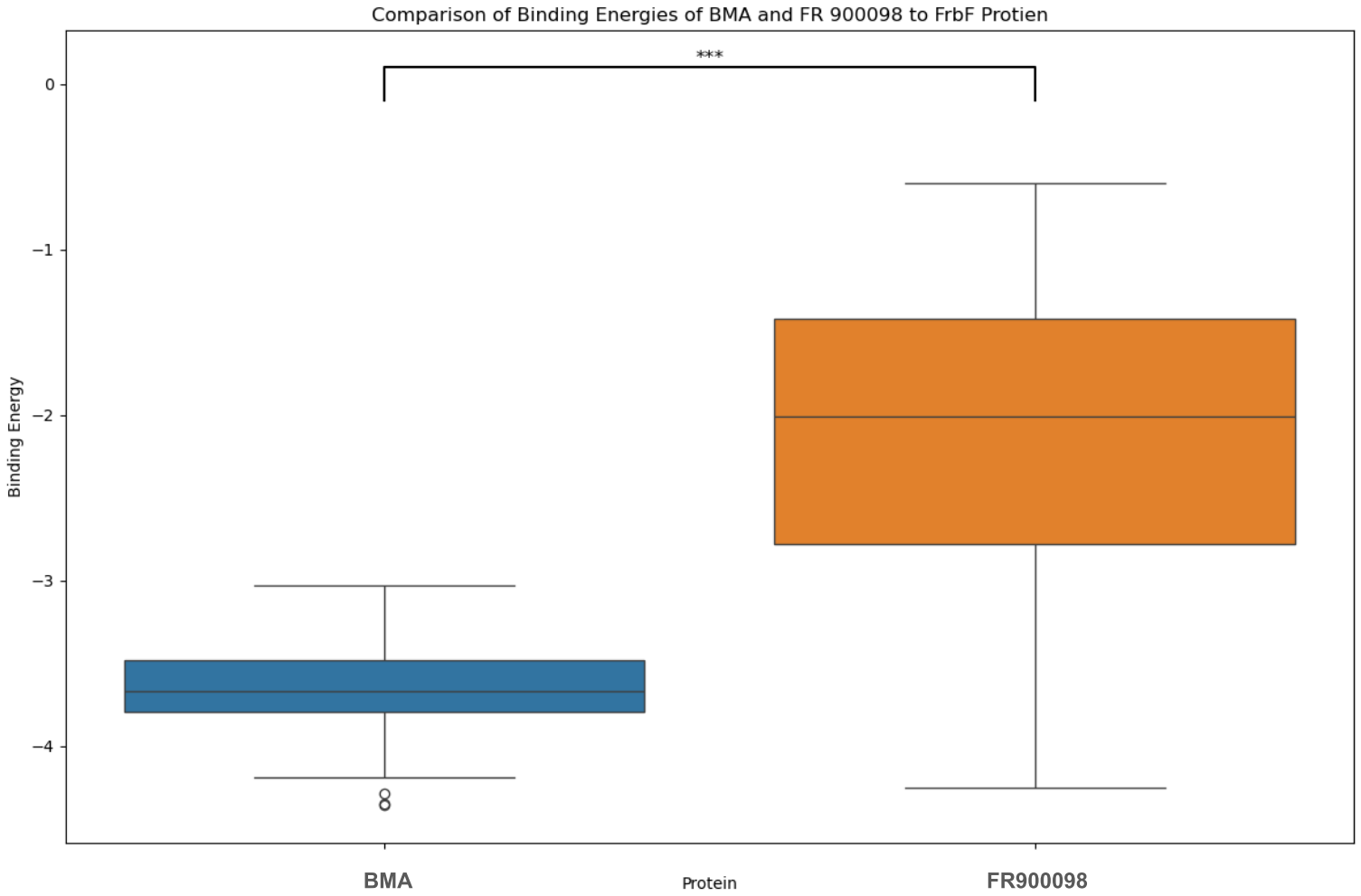


Figure 76 - Comparison of BMA and FR900098 Binding Energies to FrbF – BMA predicted binding energies are significantly higher than those of FR-900098 (Two Sample T-Test P-Value = 0.000)

7.4.5 Comparison of Binding Energies of BMA and FR-900098 to FrbF was conducted

BMA binds to FrbF significantly better when compared to FR-900098 (P-Value = 0.00) (Figure 76). To ascertain if the unsupervised machine learning data could provide more BMA resistance genes / sequestering proteins, we could compare it to FrbF. Further analysis was conducted using three principal ICPA components and K-means.

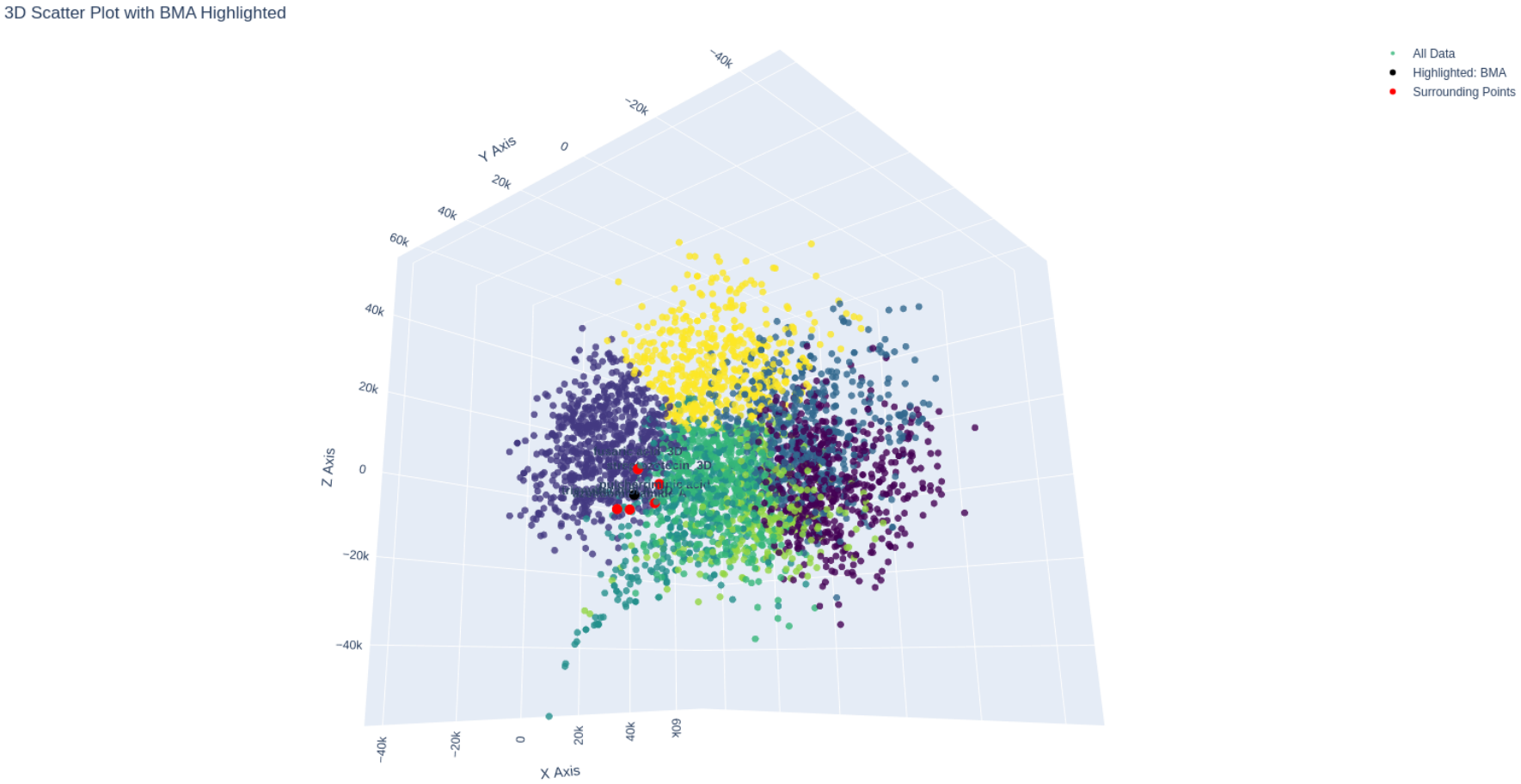


Figure 77 - IPCA Set to 3 Principal Components K-Means clustering Top View

Figure 78 - IPCA Set to 3 Principal Components K-Means clustering Bottom View

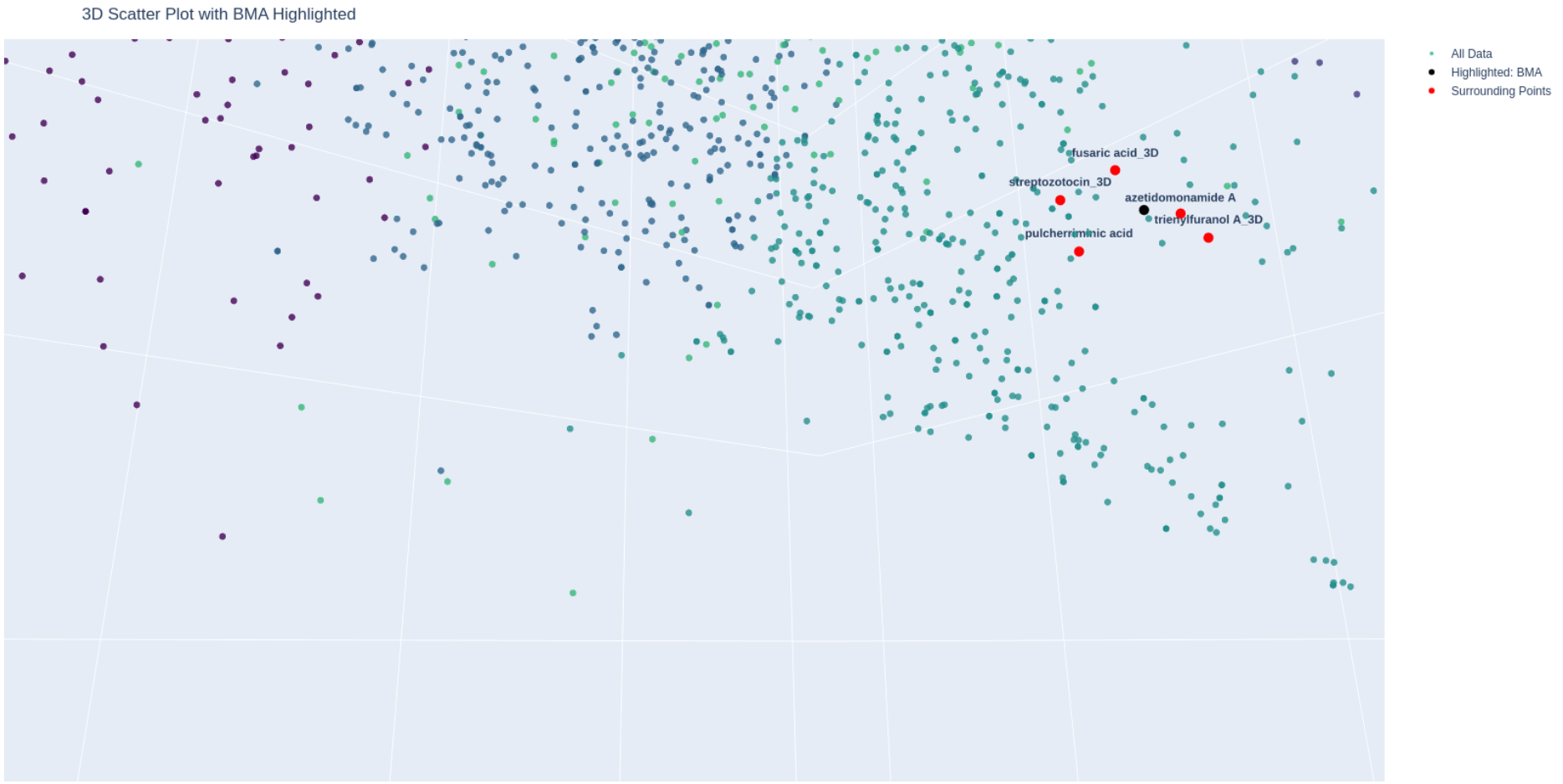


Figure 79 - IPCA Set to 3 Principal Components K-Means clustering - Zoomed on BMA and 5 Surrounding Molecules

7.5 Molecules were Clustered using ICPA Set to 3 Principal Components and K-Means

ICPA set to two principal components highlighted FR-900098, which, in two dimensions, had a similar profile to that of BMA. To understand if the data could reveal a closer three-dimensional relative, the clustering process was repeated with IPCA set to three (as opposed to two) principal components, and the molecules and BGCs were clustered using K-Means (Figure 78). The six closest neighbours were identified, and the biosynthetic gene clusters of these molecules were investigated (Table 15).

Table 15 - Closest Neighbours of BMA in IPCA 3 component K-Means Clustering

Molecule	Biosynthetic Gene Cluster	Resistance Gene	Observed similarity
Trienylfuranol A	BGC0002262	None	No
azetidomonamide A	BGC0002037	None	No
pulcherriminic acid	BGC0002321	None	No
streptozotocin	BGC0002294	QBF29317.1 ABC_efflux	No
streptozotocin	BGC0002313	None	No
fusaric acid	**BGC0001190**	**FVEG_12526**	**alkyl chains (butyl group)**

7.5.1 Fusaric Acid from BGC0001190 was identified to have similar properties to BMA.

Only two of BMA's six neighbours in the clustering analysis held a resistance gene. Only one of these molecules, fusaric acid, bore some resemblance to BMA, which also has a butyl group. The molecules were aligned to assess their further similarity, and the electrostatic properties were superimposed using PyMol and APBS plugin.

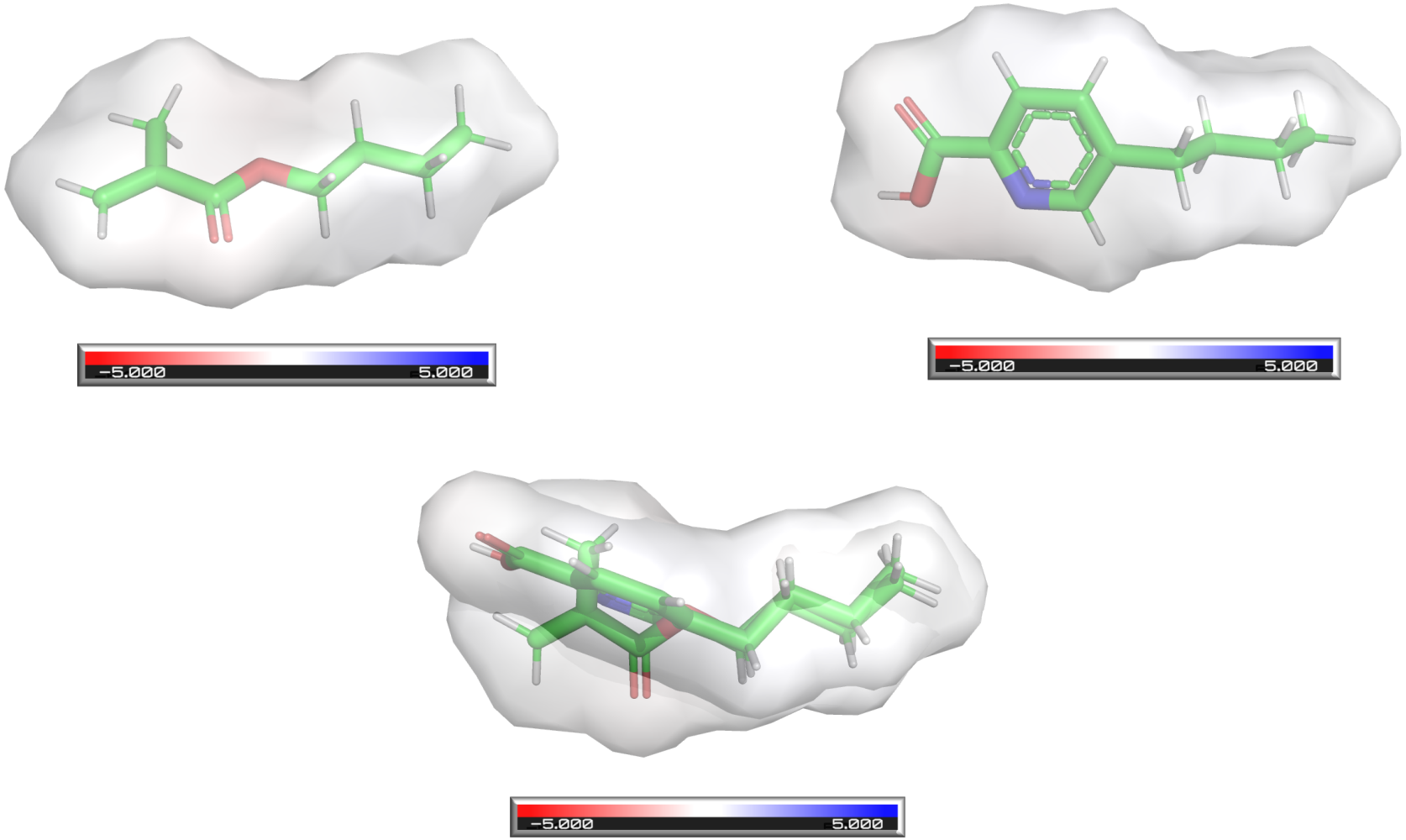


Figure 80 – PyMol ABPS Electrostatic projections of BMA (**LEFT**) and Fusaric Acid (**RIGHT**). The two molecules aligned (**CENTER**)

7.5.2 Comparison of Fusaric Acid and BMA Aligned with PyMol and Electrostatic Properties Superimposed using APBS Plugin

The two molecules have similar electrostatic properties. Both are hydrophobic, with BMA having a calculated partition coefficient (I.e., XLogP3) of 2.9 and Fusaric acid having an XLogP3 of 2.6. BMA has five rotatable bonds, and Fusaric Acid has four. Depending on the conformation of the molecules, the electrostatic profile could be much more closely aligned than shown in Figure 80. Fusaric acid is a mycotoxin produced by the organism *Fusarium moniliforme* (H. Wang & Ng, 1999). The protein structure of FveG, the resistance gene product of the biosynthetic gene cluster, was predicted using AlphaFold, and BMA binding predictions were made using AutoDock.

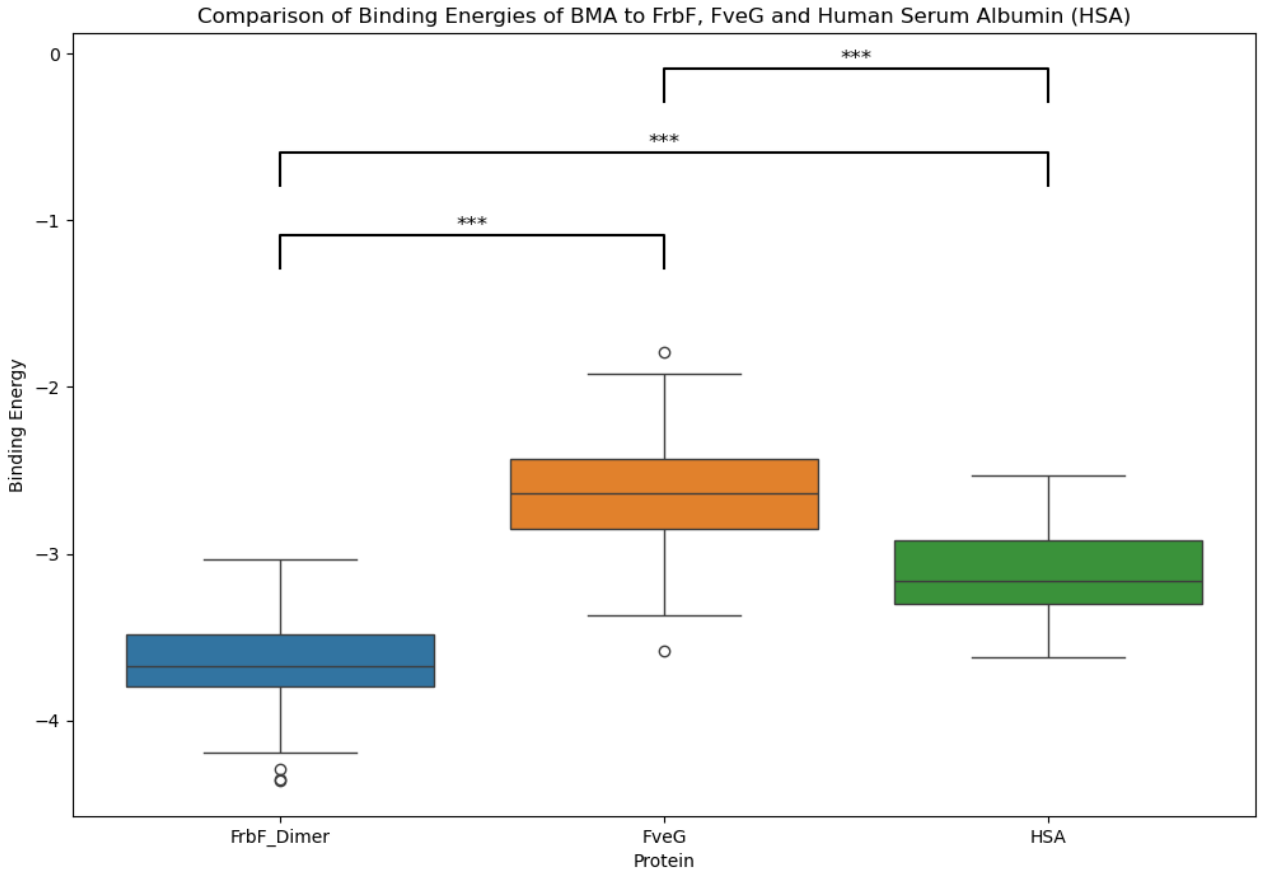


Figure 81 -Comparison of Binding Energies of BMA to USML Identified Proteins with Human Serum Albumin (HSA)

7.5.3 Comparison of Binding Energies of BMA to FrbF, FveG, HSA

Human Serum Albumin (HSA) was also introduced, a protein known for transporting hydrophobic molecules in its large hydrophobic cavity to set some benchmarks for hydrophobic binding to proteins (Yang et al., 2014). A one-way ANOVA was conducted with a Tukey HSD to compare FveG, FrbF (identified in the two PCA approach), and HSA binding energies to BMA. FrbF and HSA outperformed FveG. FrbF was shown to have significantly stronger binding energies to BMA than all tested protein structures. To assess the potential of FrbF experimentally, its gene was first of all, synthesised and cloned onto a plasmid.

7.6 *frbF* Plasmid Design and BMA Challenging of Transformed KT2440

To assess the potential of the FrbF protein for intracellular protection to BMA, the gene was codon optimised for *P. putida* and expressed from a strong constitutive promoter J23119 (Zhang et al., 2024a). A Shine–Dalgarno sequence was also incorporated upstream of the *frbF* coding sequence (Malys, 2011), and a transcriptional terminator B0015 downstream (*Part:BBa B0015 - Parts.Igem.Org*, n.d.). GenScript gene synthesis services synthesised the gene onto the pUC57-Kan plasmid (Figure 82). To assess the potential of FrbF, the plasmid was then transformed into *P. putida* KT2440, which was then challenged with BMA.

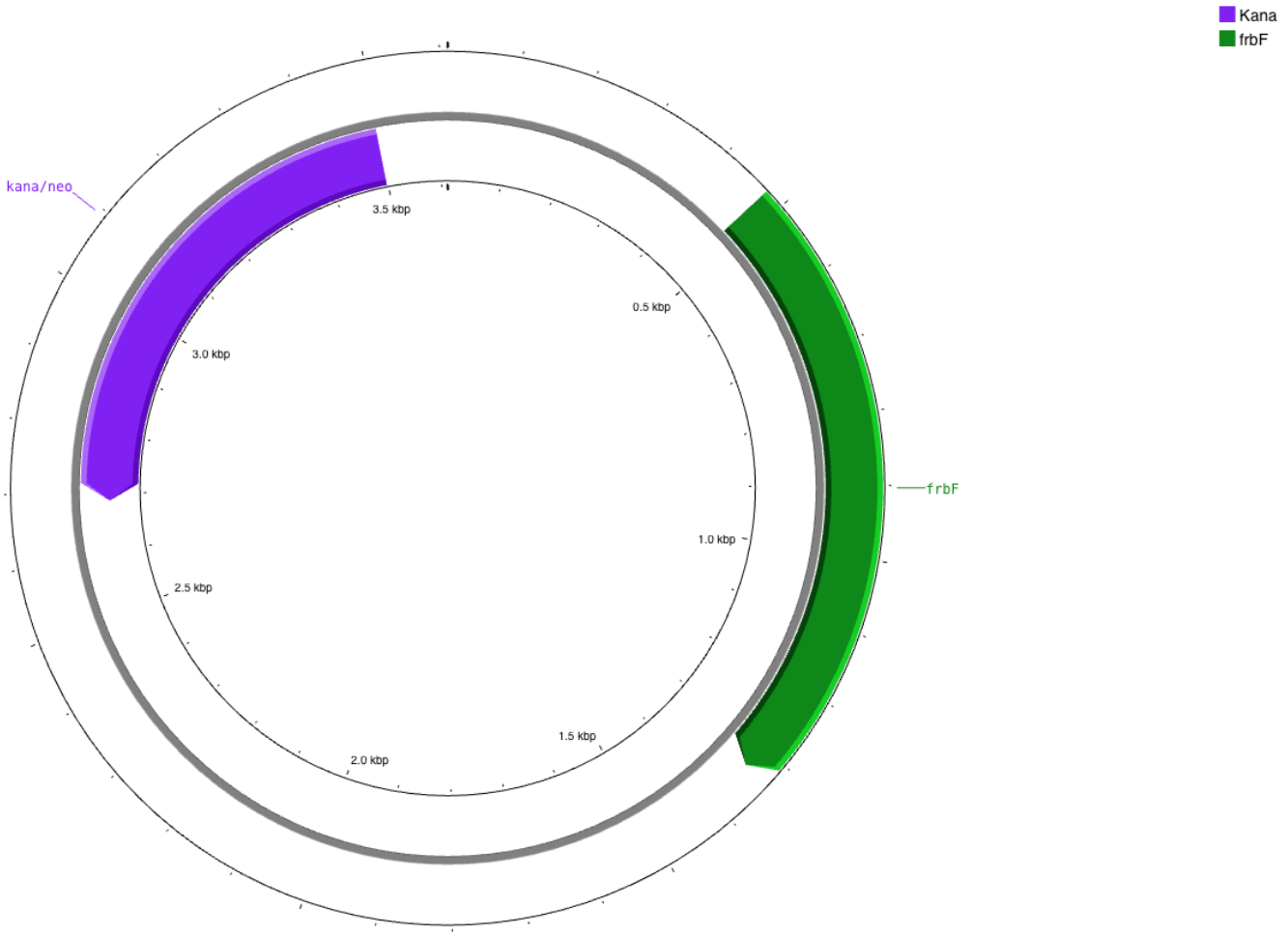


Figure 82 - pUC57-Kan-frbF+ plasmid comprised of a Kanamycin resistance gene and *frbF* with a constitutive promotor J213119 and transcriptional terminator B0015.

7.6.1 BMA Tolerance Assay of *frbF*+ Plasmid Transformed KT2440

KT2440 and transformed KT2440 were inoculated into LB media treated with BMA 25% (v/v). They were cultured for 24 hours, serial dilutions were made, and colonies were counted and analysed. KT2440 holding the *frfB*+ plasmid showed a significant increase in BMA tolerance compared to KT2440 WT (Figure 83 and Figure 84)

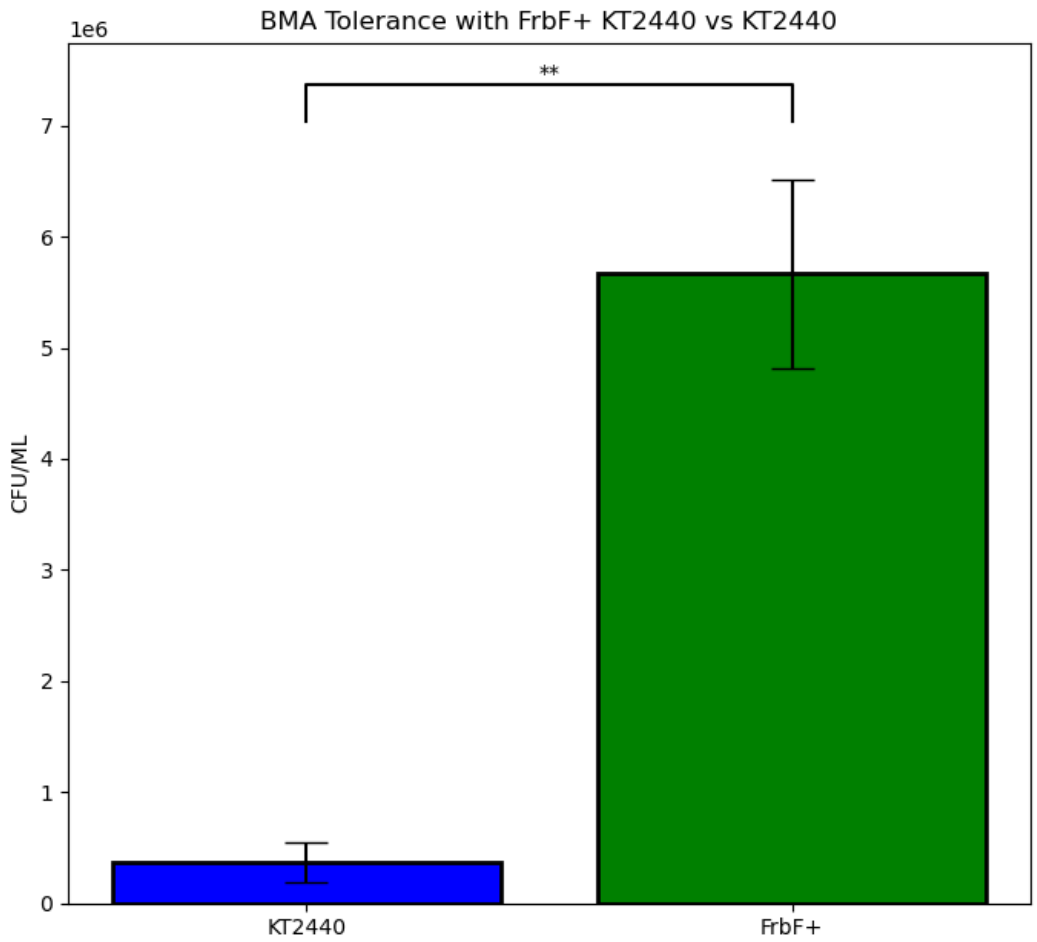


Figure 83 - BMA 25% v/v Assay of *P. putida* KT2440 WT and KT2440 with *frbF* pU57-Kan Plasmid. *frbF*+ KT2440 showed a significant increase in tolerance (P-Value=0.01)

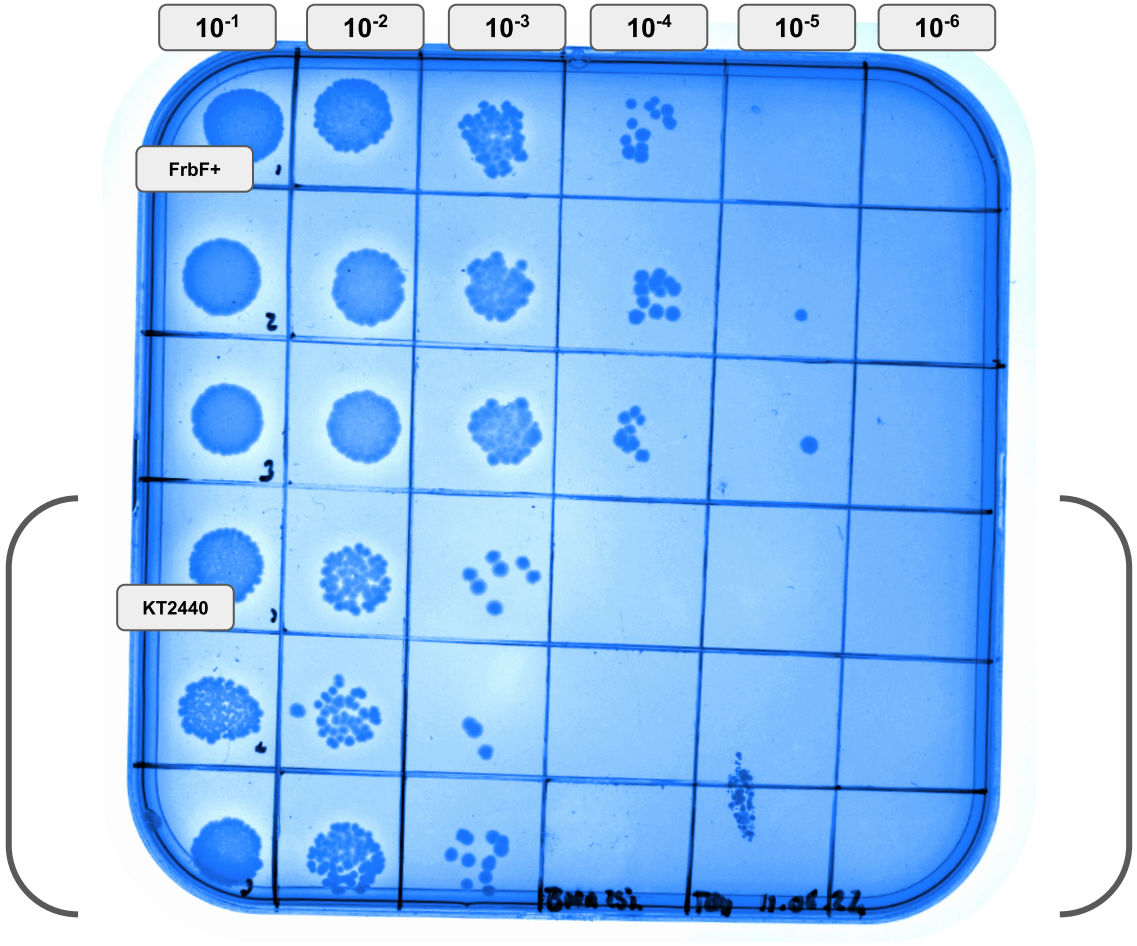


Figure 84 – Plated 20µl of Serial Dilutions of KT2440 and KT2440 *frbF*+ pUC57-Kan Plasmid BMA 25% v/v Tolerance Assay

7.7 Discussion

Both approaches to clustering the molecular structures of the products of BGCS and BMA successfully identified molecules with similar features to BMA. FR-900098 highlighted in the 2 component IPCA, the K-means experiment had a similar profile to BMA, including a hydrophobic tail. The structure of its resistance gene product FrbF, had been resolved. BMA and FR-900098 are bound in similar regions (Pocket 1 and 2), but BMA bound FrbF more favourably with calculated lower binding energies indicative of a high quality and stable binding. When comparing BMA binding energies, there was no significant difference in binding energies when comparing the AlphaFold2 and X-Ray resolved crystal structure. This allowed for the comparison of FrbF with other AlphaFold proteins with confidence.

The 3 component IPCA and K-Means experiment highlighted six more compounds with only one with similar features to BMA. Fusaric acid has a butyl group and has a similar hydrophilic profile with only 0.3 difference in XLogP3. With this similarity in profile and hydrophobic qualities, it would be expected that the resistance gene would be equipped to deal with such a molecule. FveG, the resistance protein, scored poorly in the BMA docking experiment compared to FrbF and HSA. FveG, compared to the NCBI Blast database, comes back with an identity similar to beta-lactamase, suggesting that its specificity to fusaric acid is attributable to the target

molecule having a pyridine ring, as beta-lactamases act on beta-lactam rings.

The introduction of the *frbF* gene with a constitutive promotor on pU57-Kan to KT2440 significantly increased tolerance, showing the success of this machine-learning-guided approach to facilitating intracellular tolerance. Unlike other experiments in this study, where membrane-bound efflux pumps were introduced to pump out BMA / toxic compounds from cells. FrbF protein is soluble within the cytosol, which suggests its promise to confer intracellular tolerance to BMA.

7.8 Future Work

Further validation of the effects of the introduction of the FrbF protein. To assess if the FrbF enzyme modifies BMA, transforming the Mitsubishi production strain with the *frbF* plasmid and then evaluating the quality of BMA produced through mass spectroscopy would be the most viable approach. If the protein acts in the desired way, the BMA will simply be sequestered (rather than modified) by FrbF. This may have implications for downstream purification of BMA but will enable higher production concentrations in a fermentation process. A protein denaturing step may be introduced to free the produced BMA into the culture mix.

The data used to make these clusters was fairly complex. To reduce this complexity, the molecular structure data could be generated differently. The current approach uses a slight rotation along the x and y-axis. Each rotation is saved into a matrix of 120 per array. Instead, a Molecular CT-scan approach might reduce the complexity by saving cross-sections into a matrix and adding these in order like a CT-Scan would drastically reduce data. During this study, we used convolutional neural networks to extract features from the data set for K-means clustering. The neighbours that this study processed were almost like BMA-like fragments within larger molecules. Using the more straightforward CT-Scan data might suit this approach for extracting features.

RT-qPCR of the *frbF* gene to verify its transcription levels. After further research on constitutive promoters in *P. putida*, the literature suggested that the promotor J23111 would result in higher amounts of transcription (Zhang et al., 2024b). A transcription comparison of J23119 used in this study, and J23111 using RT-qPCR combined with a BMA tolerance assay could provide insight into the viability of the gene to be added to the chromosome to determine if there is a link between transcription levels and BMA tolerance.

Computational biochemistry to alter the residues of FrbF protein that BMA interacts with within pocket one to see if they could produce stronger binding energies in silico. Pocket two also had similar properties to pocket

one. This could also undergo in-silico modifications to see if both pockets could be adjusted to favour BMA binding. These could then be synthesised and tested in the lab.

Incorporating the f*rbF* gene into *P. putida* KT2440 on a plasmid with an antibiotic marker would not be suited for large-scale bioreactor fermentation because there are likely to be financial and regulatory constraints around the use of antibiotics and propagation of antibiotic-resistant genes at scale. If strong expression of *frbF* from a high-copy plasmid is advantageous, then an alternative system ensuring plasmid retention may be required (e.g., a toxin-antitoxin system).

8 Discussion

8.1 Successful Isolation of BMA tolerant Organisms from the Royal Conservatoire of Glasgow and Mitsubishi Chemical Production Facility in Teesside could provide strains for industrial use

Mitsubishi Chemical Group are developing a bioprocess for the production of BMA to replace their energy intensive petrochemical process currently in use. Development so far has focused on *P. putida* KT2440 as a production strain due to the strains widely documented tolerance to a wide range of hydrophobic compounds (Faizal et al., 2005b). The hydrophobic nature of BMA has presented as a bottleneck in the development of this bioprocess due to the toxicity of BMA decreasing titres of BMA in fermentation vessels. The principal objective of this study was to isolate organisms that could demonstrate a higher BMA tolerance than *P. putida* K2440 for bioprocess optimisation.

Pseudomonads are highly valuable strains for industrial application due to their widely demonstrated tolerance to many compounds and genetically tractability (Cárdenas Espinosa et al., 2023; Rajagopal, 1996; S. Wang et al., 2020). The approach of targeted environmental sampling has been widely demonstrated in the literature as an effective approach for recovering strains with industrial potential (Chien et al., 1976a; Yoshida et al., 2021). During this study, the Royal Conservatoire was selected for

sampling due to the physiochemical properties caused by its constant exposure to various toxic and hydrophobic compounds. Here, the isolation of solvent-tolerant strains is demonstrated from The Royal Conservatoire. Isolates RCS09A, RCS09A(2), and RCS09A(3) showed a significant increase in BMA tolerance compared to *P. putida* KT2440. RCS09A BMA tolerance was significantly higher than that of the other RCS09 isolates. This higher demonstration of tolerance could lead to higher titres of BMA and lower production costs in a fermentation process. RCS09A was analysed further using WGS, and it was identified as *P. putida*. Furthermore, it carries a plasmid (containing an RND efflux pump MOA) identified through BLAST to have similar sequences to another plasmid RW109 associated with an extreme solvent tolerance stress response phenotype (Weiser et al., 2019a). Tolerance conferring genes contained within plasmids could allow easy transfer of phenotypes to chassis organisms (Hutchison et al., 2016). Here, targeted environmental sampling is demonstrated as a useful tool for the isolation of strains with industrial potential. This approach could be expanded for other compounds and environments to isolate strains relevant to other industrial processes.

As strains which demonstrated a high level of BMA tolerance were isolated from the Royal Conservatoire, a sample site with a specific chemical profile of methacrylate esters was chosen as the next sample site. Industrial environments have been shown in the literature to possess unique microbiomes due to the enrichment of the surrounding

environment with industrial contaminants (Ibarbalz et al., 2013). Sampling the Mitsubishi Chemical site also produced an organism with significantly higher tolerance when compared to *P. putida* KT2440. The isolate MC-G-03 was analysed further through WGS and identified as *P. veronii*, known for its use in bioremediation (Nam et al., 2003b; Onaca et al., 2007b). This work provides further evidence that industrial sites enriched with relevant compounds are a useful tool in strain isolation for bioprocesses.

Utilising comparative genomics to understand the pangenome can identify tolerance conferring systems shared and unique to tolerant organisms (Youenou et al., 2015). Identification of such systems can provide industrial benefit as systems which can be exploited to increase titre can be identified (Faizal et al., 2005b). The pangenomes of the BMA-tolerant organisms isolated from the various sampling sites were studied to determine if they shared any genetic features that may contribute to BMA tolerance. The pangenome was filtered using keywords relating to RND efflux pumps as these have been shown to confer tolerance to a wide range of compounds in *Pseudomonas* species (Blair et al., 2014; Colclough et al., 2020). As RCS09A was the most BMA-tolerant organism, its unique genes were investigated and it held four unique RND efflux operons in addition to the one on its plasmid.

The three strains, RCS09A, MC-G-03 and KT2440, were subjected to virulence factor gene screening to assess these isolates' potential as

alternative BMA production strains. Pathogenic isolates would not make suitable production strains; therefore, identifying virulence genes could eliminate a strain as a potential BMA fermentation candidate. Virulence factors such as biofilm formation can lead to bottlenecks in fermentation processes (Chmielewski & Frank, 2003; Dellias, 2016). RCS09A contained no virulence factors within the screening database. MC-G-03 contained virulence factors that mainly contribute to motility and biofilm formation. Therefore, MC-G-03 would not make an ideal production strain for further development.

To approximate how many RND efflux pumps were held by RCS09A, MC-G-03 and KT2440, they were screened using BLAST and 3 query RND efflux pumps. RCS09A was shown to have significantly more matches compared to MC-G-03 and KT2440. Suggesting that the strain RCS09A has a larger repertoire of RND efflux pumps which could facilitate the efflux of a range of toxic compounds from the cell in fermentation (Blair et al., 2014; Colclough et al., 2020).

BMA-tolerant organisms were successfully isolated by targeting the industrial environment with desired physiochemical properties. The Royal Conservatoire microbiome was sampled due to its constant exposure to chemical concomitants at sustained temperatures. Here, the application of such an environment for industrial strain selection is demonstrated. RCS09A *P. putida*, the most BMA-tolerant and containing no virulence

factor or resistance genes and a higher amount of RND efflux pump genes, is the ideal candidate to be developed as an alternative production strain, which could result in increased yield and lower costs for the production of BMA.

8.2 Demonstration of the MOA Efflux System as Conferring BMA Tolerance

Wild-type environmental strains can be unpredictable in engineering metabolism for the production of platform chemicals due to their uncharacterised genomes and genetic resilience (Chemla et al., 2022). To enhance the BMA tolerance of strains in development for BMA production, the second principal objective of this study was to identify molecular mechanisms in which the environmental isolates confer their tolerance. For this study, a proxy to production strain, *P. putida* KT2440, was used. The plasmid held by RCS09A was identified as a significant contributor to its tolerance, as demonstrated by curing the organism of the plasmid. The main feature was the RND efflux pump MOA, which was hypothesised as the primary molecular mechanism of its tolerance. Transformation of *P. putida* KT2440 with this plasmid significantly increased the BMA tolerance. This is a relativity simple genetic engineering technique. This could provide a 'quick' upgrade for Mitsubishi to determine if increasing tolerance can increase the titres of BMA produced in their strains.

8.3 Identification of Other RND Efflux Pumps with Similar Characteristics to Pumps Suited to Hydrophobic Solvents

From the literature, we found another RND efflux pump suited to the hydrophobic compound toluene, the ttgDEF system belonging to *P. putida* DOT-T1E (Ramos et al., 1998). The internal protein pocket properties of ttgDEF, RCS09A MOA (MdtA) and other commonly studied RND efflux pumps were calculated. Statistical analysis of the variance of all metrics calculated showed that the two pumps demonstrated to facilitate the efflux of hydrophobic compounds scored highly and concisely in ligand flexibility compared to other pumps analysed. Hypothesising that this metric could be linked to the ability to efflux hydrophobic compounds, we could use this as an indicator to identify other pumps for investigation in future work. Unsupervised machine learning was used which identified an RND *MexAB-orpM* efflux operon belonging to *P. aeruginosa* SCAID. This pump scored highly in ligand flexibility with no significant difference to the scores of ttgDEF and MOA (MdtA). Suggesting that this system may also be well suited to the efflux of hydrophobic compounds. Here, a novel approach for the identification of potential efflux systems optimised for hydrophobic compounds is demonstrated. This approach could provide as a useful strain development tool for industrially capable strains.

8.4 Unsupervised Machine Learning Selection for a BMA Resistance / Sequestering Protein

It was determined by Mitsubishi Chemical that BMA was most likely toxic within the cell, regardless of the cell's ability to efflux BMA out of the cytosol (Mitsubishi Chemical – Personal Communication). A new study objective was introduced: to provide intracellular protection from BMA to increase tolerance. In nature, organisms can protect themselves from a toxic compound it produces by including a resistance gene in the BGC (Stegmann et al., 2015b). As BMA is not a naturally occurring compound, a resistance gene for it most likely does not exist. It was hypothesised that machine learning could be used to sort and cluster compounds made in nature, including BMA, to identify if there are any natural compounds made by BGCs that share features with BMA. Here, an unsupervised machine learning approach was successfully demonstrated to identify several BCGs with resistance genes and their products including FrbF which increased BMA tolerance in KT2440. This approach demonstrates huge industrial application to overcome tolerance bottlenecks in a compound specific manner for any bioprocess. Here, we demonstrate the first study of this kind.

8.5 Conclusions

Several organisms with enhanced tolerance to BMA have been identified during this study, as well as several molecular mechanisms to upgrade the BMA tolerance of organisms in development. The findings of this study conclude that the industrial environment can be used as a resource when designing complicated biotechnological processes such as a BMA fermentation process. Here, we present RCS09A *P. putida* as a candidate strain for BMA production. The efflux pump system MOA (MdtA) and the FrbF protein are presented as molecular mechanisms to increase tolerance of current strains in development for the production of BMA. These additions to the Bio-BMA project could increase titres of produced BMA and reduce costs.

9 Future Directions

RCS09A should be engineered to produce BMA and yields compared to current strains in development to ascertain if its increased tolerance increases the yield of BMA produced.

Screening of the alternative RND efflux pumps identified in this study SCAID MexAB-OrpM and RCS09A_RND1. These should be introduced to the chromosome of *P. putida* KT2440 as the MOA (MdtA) operon was. These could be introduced with the regulator and promoter sequences of the MOA (MdtA) operon to profile the pumps, as this has already been demonstrated to increase tolerance in combination with its native efflux pump. This approach would provide a more equal comparison of the pump's suitability to efflux the desired compounds. As the SCAID efflux pump was identified from an online database, it should be synthesised RCS09A_RND1 and can be extracted using PCR from the RCS09A genome into the TnT7 construct.

The FrfB plasmid should be transformed into the Mitsubishi Chemical strains in development for the production of BMA. The strain should then be tested to see if this addition increases tolerance, and then its BMA output should be compared to that of the non-transformed strain using mass spectroscopy. If after a protein denaturing step to release sequestered BMA into the culture mix. If there is more quantifiable BMA, this would suggest that the protein is

sequestering the BMA. If there is less, we could hypothesise that the enzymatic activity of the enzyme is modifying the BMA. The FrbF protein could be modified to be secreted by a secretion system to add additional BMA removal from the cytosol.

10 Bibliography

Adamiak, J. W., Jhawar, V., Bonifay, V., Chandler, C. E., Leus, I. V., Ernst, R. K., Schweizer, H. P., & Zgurskaya, H. I. (2021). Loss of RND-type multidrug efflux pumps triggers iron starvation and lipid a modifications in pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, *65*(10). https://doi.org/10.1128/AAC.00592-21/SUPPL_FILE/AAC.00592-21-S0003.XLSX

Aloush, V., Navon-Venezia, S., Seigman-Igra, Y., Cabili, S., & Carmeli, Y. (2006). Multidrug-resistant Pseudomonas aeruginosa: Risk factors and clinical impact. *Antimicrobial Agents and Chemotherapy*, *50*(1), 43–48. https://doi.org/10.1128/AAC.50.1.43-48.2006/ASSET/FEB4A331-4FC2-472D-ACE2-809A3E757AC4/ASSETS/GRAPHIC/ZAC0010654940001.JPEG

AlphaFold - Google DeepMind. (n.d.). Retrieved August 15, 2024, from https://deepmind.google/technologies/alphafold/

Bae, B., Cobb, R. E., DeSieno, M. A., Zhao, H., & Nair, S. K. (2011a). New N-Acetyltransferase Fold in the Structure and Mechanism of the Phosphonate Biosynthetic Enzyme FrbF. *The Journal of Biological Chemistry*, *286*(41), 36132. https://doi.org/10.1074/JBC.M111.263533

Bae, B., Cobb, R. E., DeSieno, M. A., Zhao, H., & Nair, S. K. (2011b). New N-Acetyltransferase Fold in the Structure and Mechanism of the Phosphonate Biosynthetic Enzyme FrbF. *The Journal of Biological Chemistry*, *286*(41), 36132. https://doi.org/10.1074/JBC.M111.263533

Begmatov, S., Beletsky, A. V., Dorofeev, A. G., Pimenov, N. V., Mardanov, A. V., & Ravin, N. V. (2024). Metagenomic insights into the wastewater resistome before and after purification at large-scale wastewater treatment plants in the Moscow city. *Scientific Reports 2024 14:1*, *14*(1), 1–12. https://doi.org/10.1038/s41598-024-56870-0

Blair, J. M. A., Richmond, G. E., & Piddock, L. J. V. (2014). Multidrug efflux pumps in Gram-negative bacteria and their role in antibiotic resistance. *Future Microbiology*, *9*(10), 1165–1177. https://doi.org/10.2217/FMB.14.66

Callaway, E. (2024). AI protein-prediction tool AlphaFold3 is now more open. *Nature*, *635*(8039), 531–532. https://doi.org/10.1038/D41586-024-03708-4

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 1–9. https://doi.org/10.1186/1471-2105-10-421/FIGURES/4

Cárdenas Espinosa, M. J., Schmidgall, T., Pohl, J., Wagner, G., Wynands, B., Wierckx, N., Heipieper, H. J., & Eberlein, C. (2023). Assessment of New and Genome-Reduced Pseudomonas Strains Regarding Their

Robustness as Chassis in Biotechnological Applications. *Microorganisms*, *11*(4), 837. https://doi.org/10.3390/MICROORGANISMS11040837/S1

Carrara, J. B. ; A. ;, Barroso, C. D. N. ;, Tuon, F. F. ;, Faoro, H., Lorusso, A. B., Carrara, J. A., Deuttner, C., Barroso, N., Tuon, F. F., & Faoro, H. (2022a). Role of Efflux Pumps on Antimicrobial Resistance in Pseudomonas aeruginosa. *International Journal of Molecular Sciences 2022, Vol. 23, Page 15779*, *23*(24), 15779. https://doi.org/10.3390/IJMS232415779

Carrara, J. B. ; A. ;, Barroso, C. D. N. ;, Tuon, F. F. ;, Faoro, H., Lorusso, A. B., Carrara, J. A., Deuttner, C., Barroso, N., Tuon, F. F., & Faoro, H. (2022b). Role of Efflux Pumps on Antimicrobial Resistance in Pseudomonas aeruginosa. *International Journal of Molecular Sciences 2022, Vol. 23, Page 15779*, *23*(24), 15779. https://doi.org/10.3390/IJMS232415779

Chemla, Y., Dorfan, Y., Yannai, A., Meng, D., Cao, P., Glaven, S., Gordon, D. B., Elbaz, J., & Voigt, C. A. (2022). Parallel engineering of environmental bacteria and performance over years under jungle-simulated conditions. *PLOS ONE*, *17*(12), e0278471. https://doi.org/10.1371/JOURNAL.PONE.0278471

Chien, A., Edgar, D. B., & Trela, J. M. (1976a). Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of Bacteriology*, *127*(3), 1550. https://doi.org/10.1128/JB.127.3.1550-1557.1976

Chien, A., Edgar, D. B., & Trela, J. M. (1976b). Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus.

Journal of Bacteriology, *127*(3), 1550. https://doi.org/10.1128/JB.127.3.1550-1557.1976

Chmielewski, R. A. N., & Frank, J. F. (2003). Biofilm Formation and Control in Food Processing Facilities. *Comprehensive Reviews in Food Science and Food Safety*, *2*(1), 22–32. https://doi.org/10.1111/J.1541-4337.2003.TB00012.X

Choi, K. H., & Schweizer, H. P. (2006a). mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nature Protocols 2006 1:1*, *1*(1), 153–161. https://doi.org/10.1038/nprot.2006.24

Choi, K. H., & Schweizer, H. P. (2006b). mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nature Protocols 2006 1:1*, *1*(1), 153–161. https://doi.org/10.1038/nprot.2006.24

Colclough, A. L., Alav, I., Whittle, E. E., Pugh, H. L., Darby, E. M., Legood, S. W., McNeil, H. E., & Blair, J. M. A. (2020). RND efflux pumps in Gram-negative bacteria; Regulation, structure and role in antibiotic resistance. *Future Microbiology*, *15*(2), 143–157. https://doi.org/10.2217/FMB-2019-0235

Corporation, P. (n.d.). *Wizard ® HMW DNA Extraction Kit Instructions for Use of Product A2920*.

Corporation, P. (1999). *Wizard Plus SV Minipreps DNA Purification System Quick Protocol, FB004*.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. https://doi.org/10.1093/GIGASCIENCE/GIAB008

Dellias, M. de T. F. (2016). *Comunidade bacteriana dos biofilmes da fermentação alcoólica: estrutura, composição, suscetibilidade aos antimicrobianos e formação de biofilme em culturas puras*. https://doi.org/10.11606/T.64.2016.TDE-01042016-091529

DNeasy PowerSoil Pro Kits. (n.d.). Retrieved February 9, 2023, from https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/microbial-dna/dneasy-powersoil-pro-kit

Duque, E., Segura, A., Mosqueda, G., & Ramos, J. L. (2001). Global and cognate regulators control the expression of the organic solvent efflux pumps TtgABC and TtgDEF of Pseudomonas putida. *Molecular Microbiology*, *39*(4), 1100–1106. https://doi.org/10.1046/J.1365-2958.2001.02310.X

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996a). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. www.aaai.org

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996b). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. www.aaai.org

Faizal, I., Dozen, K., Hong, C. S., Kuroda, A., Takiguchi, N., Ohtake, H., Takeda, K., Tsunekawa, H., & Kato, J. (2005a). Isolation and characterization of solvent-tolerant Pseudomonas putida strain T-57, and its application to biotransformation of toluene to cresol in a two-phase (organic-aqueous) system. *Journal of Industrial Microbiology and Biotechnology*, *32*(11–12), 542–547. https://doi.org/10.1007/S10295-005-0253-Y

Faizal, I., Dozen, K., Hong, C. S., Kuroda, A., Takiguchi, N., Ohtake, H., Takeda, K., Tsunekawa, H., & Kato, J. (2005b). Isolation and characterization of solvent-tolerant Pseudomonas putida strain T-57, and its application to biotransformation of toluene to cresol in a two-phase (organic-aqueous) system. *Journal of Industrial Microbiology and Biotechnology*, *32*(11–12), 542–547. https://doi.org/10.1007/S10295-005-0253-Y

Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., & Olsen, G. J. (2008). Critical Evaluation of Two Primers Commonly Used for Amplification of Bacterial 16S rRNA Genes. *Applied and Environmental Microbiology*, *74*(8), 2461. https://doi.org/10.1128/AEM.02272-07

Fuglebakk, E., & Reuter, N. (2018). A model for hydrophobic protrusions on peripheral membrane proteins. *PLOS Computational Biology*, *14*(7), e1006325. https://doi.org/10.1371/JOURNAL.PCBI.1006325

Gold, B., Deng, H., Bryk, R., Vargas, D., Eliezer, D., Roberts, J., Jiang, X., & Nathan, C. (2008). Identification of a copper-binding metallothionein

in pathogenic mycobacteria. *Nature Chemical Biology 2008 4:10*, *4*(10), 609–616. https://doi.org/10.1038/nchembio.109

Holliday, J. D., Rodgers, S. L., Willett, P., Chen, M. Y., Mahfouf, M., Lawson, K., & Mullier, G. (2004). Clustering files of chemical structures using the fuzzy k-means clustering method. *Journal of Chemical Information and Computer Sciences*, *44*(3), 894–902. https://doi.org/10.1021/CI0342674/ASSET/IMAGES/LARGE/CI0342674F00006.JPEG

Horwood, C. A., Cramb, D. T., Birss, V., Lawan, M. M., Shek, P. N., & Tahir, M. M. (2013). Electrochemical Removal of Bacteria and Biofilms Formed on Metal Substrates. *ECS Meeting Abstracts*, *MA2013-01*(16), 730. https://doi.org/10.1149/MA2013-01/16/730

Hosseini, R., Kuepper, J., Koebbing, S., Blank, L. M., Wierckx, N., & de Winde, J. H. (2017). Regulation of solvent tolerance in Pseudomonas putida S12 mediated by mobile elements. *Microbial Biotechnology*, *10*(6), 1558–1568. https://doi.org/10.1111/1751-7915.12495

Hutchison, C. A., Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z. Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., … Venter, J. C. (2016). Design and synthesis of a minimal bacterial genome. *Science*, *351*(6280). https://doi.org/10.1126/SCIENCE.AAD6253/SUPPL_FILE/AAD6253-HUTCHISON-SM.PDF

Ibarbalz, F. M., Figuerola, E. L. M., & Erijman, L. (2013). Industrial activated sludge exhibit unique bacterial community composition at high taxonomic ranks. *Water Research*, *47*(11), 3854–3864. https://doi.org/10.1016/j.watres.2013.04.010

Illumina. (2020). *MiniSeq System Guide (1000000002695)*.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., … Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, *27*(1), 112–128. https://doi.org/10.1002/PRO.3280

Kang, D., Revtovich, A. V., Chen, Q., Shah, K. N., Cannon, C. L., & Kirienko, N. V. (2019). Pyoverdine-Dependent Virulence of Pseudomonas aeruginosa Isolates From Cystic Fibrosis Patients. *Frontiers in Microbiology*, *10*, 477286. https://doi.org/10.3389/FMICB.2019.02048/BIBTEX

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology 2019 37:5*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (n.d.). *ImageNet Classification with Deep Convolutional Neural Networks*. Retrieved August 22, 2024, from http://code.google.com/p/cuda-convnet/

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009a). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*(1), 1–11. https://doi.org/10.1186/1471-2105-10-168/TABLES/1

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009b). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*(1), 1–11. https://doi.org/10.1186/1471-2105-10-168/TABLES/1

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009c). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, *10*(1), 1–11. https://doi.org/10.1186/1471-2105-10-168/TABLES/1

Lebeau, J., Efromson, J. P., & Lynch, M. D. (2020). A Review of the Biotechnological Production of Methacrylic Acid. *Frontiers in Bioengineering and Biotechnology*, *8*(March), 1–10. https://doi.org/10.3389/fbioe.2020.00207

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, M., Sun, Y., Pan, S. A., Deng, W. W., Yu, O., & Zhang, Z. (2017). Engineering a novel biosynthetic pathway in: Escherichia coli for the

production of caffeine. *RSC Advances*, *7*(89), 56382–56389. https://doi.org/10.1039/c7ra10986e

Li, X. Z., Nikaido, H., & Poole, K. (1995). Role of MexA-MexB-OprM in antibiotic efflux in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy*, *39*(9), 1948–1953. https://doi.org/10.1128/AAC.39.9.1948

Li, X. Z., Zhang, L., & Poole, K. (1998a). Role of the multidrug efflux systems of Pseudomonas aeruginosa in organic solvent tolerance. *Journal of Bacteriology*, *180*(11), 2987–2991. https://doi.org/10.1128/jb.180.11.2987-2991.1998

Li, X. Z., Zhang, L., & Poole, K. (1998b). Role of the multidrug efflux systems of Pseudomonas aeruginosa in organic solvent tolerance. *Journal of Bacteriology*, *180*(11), 2987–2991. https://doi.org/10.1128/JB.180.11.2987-2991.1998/ASSET/75D2BC3B-363A-46DC-8C3A-71A8CC871ACD/ASSETS/GRAPHIC/JB1181579003.JPEG

Machas, M., Kurgan, G., Abed, O. A., Shapiro, A., Wang, X., & Nielsen, D. (2021). Characterizing Escherichia coli's transcriptional response to different styrene exposure modes reveals novel toxicity and tolerance insights. *Journal of Industrial Microbiology & Biotechnology*, *48*(1–2), 19. https://doi.org/10.1093/JIMB/KUAB019

Malys, N. (2011). Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes. *Molecular Biology Reports*, *39*(1), 33–39. https://doi.org/10.1007/S11033-011-0707-4

Martí, J. M. (2019). Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLOS Computational Biology*, *15*(4), e1006967. https://doi.org/10.1371/JOURNAL.PCBI.1006967

Martinus Pos, K., Yan, A., Dreier, J., & Ruggerone, P. (2015). Interaction of antibacterial compounds with RND efflux pumps in Pseudomonas aeruginosa. *Frontiers in Microbiology*, *6*(JUL), 660. https://doi.org/10.3389/FMICB.2015.00660

McVicker, G., Hollingshead, S., Pilla, G., & Tang, C. M. (2019). Maintenance of the virulence plasmid in Shigella flexneri is influenced by Lon and two functional partitioning systems. *Molecular Microbiology*, *111*(5), 1355–1366. https://doi.org/10.1111/MMI.14225

Medić, A. B., & Karadžić, I. M. (2022). Pseudomonas in environmental bioremediation of hydrocarbons and phenolic compounds- key catabolic degradation enzymes and new analytical platforms for comprehensive investigation. *World Journal of Microbiology and Biotechnology 2022 38:10*, *38*(10), 1–28. https://doi.org/10.1007/S11274-022-03349-7

Meyer, J. M., Geoffroy, V. A., Baida, N., Gardan, L., Izard, D., Lemanceau, P., Achouak, W., & Palleroni, N. J. (2002). Siderophore typing, a powerful tool for the identification of fluorescent and nonfluorescent pseudomonads. *Applied and Environmental Microbiology*, *68*(6), 2745–2753. https://doi.org/10.1128/AEM.68.6.2745-

2753.2002/ASSET/EB6270D1-8777-4DC0-B43C-A7721C559FB4/ASSETS/GRAPHIC/AM062175703T.JPEG

Morris, G. M., Ruth, H., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of Computational Chemistry*, *30*(16), 2785. https://doi.org/10.1002/JCC.21256

Muñoz-Gómez, A. J., Lemonnier, M., Santos-Sierra, S., Berzal-Herranz, A., & Díaz-Orejas, R. (2005). RNase/anti-RNase activities of the bacterial parD toxin-antitoxin system. *Journal of Bacteriology*, *187*(9), 3151–3157. https://doi.org/10.1128/JB.187.9.3151-3157.2005/ASSET/B2069E94-EF09-4DC7-8CD1-B843F40EB6F1/ASSETS/GRAPHIC/ZJB0090546810003.JPEG

Nam, I. H., Chang, Y. S., Hong, H. B., & Lee, Y. E. (2003a). A novel catabolic activity of Pseudomonas veronii in biotransformation of pentachlorophenol. *Applied Microbiology and Biotechnology*, *62*(2–3), 284–290. https://doi.org/10.1007/S00253-003-1255-1/METRICS

Nam, I. H., Chang, Y. S., Hong, H. B., & Lee, Y. E. (2003b). A novel catabolic activity of Pseudomonas veronii in biotransformation of pentachlorophenol. *Applied Microbiology and Biotechnology*, *62*(2–3), 284–290. https://doi.org/10.1007/S00253-003-1255-1/METRICS

Novelli, M., & Bolla, J. M. (2024). RND Efflux Pump Induction: A Crucial Network Unveiling Adaptive Antibiotic Resistance Mechanisms of

Gram-Negative Bacteria. *Antibiotics 2024, Vol. 13, Page 501*, *13*(6), 501. https://doi.org/10.3390/ANTIBIOTICS13060501

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, *27*(5), 824. https://doi.org/10.1101/GR.213959.116

Okabe, M., Lies, D., Kanamasa, S., & Park, E. Y. (2009). Biotechnological production of itaconic acid and its biosynthesis in Aspergillus terreus. *Applied Microbiology and Biotechnology*, *84*(4), 597–606. https://doi.org/10.1007/S00253-009-2132-3/TABLES/5

Onaca, C., Kieninger, M., Engesser, K. H., & Altenbuchner, J. (2007a). Degradation of alkyl methyl ketones by Pseudomonas veronii MEK700. *Journal of Bacteriology*, *189*(10), 3759–3767. https://doi.org/10.1128/JB.01279-06

Onaca, C., Kieninger, M., Engesser, K. H., & Altenbuchner, J. (2007b). Degradation of alkyl methyl ketones by Pseudomonas veronii MEK700. *Journal of Bacteriology*, *189*(10), 3759–3767. https://doi.org/10.1128/JB.01279-06

Ozer, E., Yaniv, K., Chetrit, E., Boyarski, A., Meijler, M. M., Berkovich, R., Kushmaro, A., & Alfonta, L. (2021). An inside look at a biofilm: Pseudomonas aeruginosa flagella biotracking. *Science Advances*, *7*(24), 8581–8592. https://doi.org/10.1126/SCIADV.ABG8581/SUPPL_FILE/SCIADV.ABG8581_SM.PDF

Part:BBa B0015 - parts.igem.org. (n.d.). Retrieved August 26, 2024, from https://parts.igem.org/Part:BBa_B0015

Pearman, J. K., Keeley, N. B., Wood, S. A., Laroche, O., Zaiko, A., Thomson-Laing, G., Biessy, L., Atalah, J., & Pochon, X. (2020). Comparing sediment DNA extraction methods for assessing organic enrichment associated with marine aquaculture. *PeerJ*, *8*, e10231. https://doi.org/10.7717/PEERJ.10231

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot andÉdouardand, M., Duchesnay, andÉdouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

Perera, F. (2018). Pollution from fossil-fuel combustion is the leading environmental threat to global pediatric health and equity: Solutions exist. In *International Journal of Environmental Research and Public*

Health (Vol. 15, Issue 1). MDPI AG. https://doi.org/10.3390/ijerph15010016

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., Kim, E., Cho, H. J., Reynolds, J. M., Song, M. C., Park, S. R., & Yoon, Y. J. (2019). A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology*, *10*(JUN), 1–27. https://doi.org/10.3389/fmicb.2019.01404

Pisoni, L. A., Semple, S. J., Liu, S., Sykes, M. J., & Venter, H. (2023). Combined Structure- and Ligand-Based Approach for the Identification of Inhibitors of AcrAB-TolC in Escherichia coli. *ACS Infectious Diseases*, *9*(12), 2504–2522. https://doi.org/10.1021/ACSINFECDIS.3C00350/SUPPL_FILE/ID3C00350_SI_002.XLSX

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2015). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*, *8*(1), 12–24. https://doi.org/10.1039/C5AY02550H

Pugh, H. L., Darby, E. M., Burgess, L., Colclough, A. L., John, A.-R. M., Dunn, S., Connor, C., Perry, E. A., McNally, A., Bavro, V. N., & Blair, J. M. A. (2024). The RND efflux pump EefABC is highly conserved within lineages of E. coli commonly associated with infection. *BioRxiv*, 2024.08.01.606150. https://doi.org/10.1101/2024.08.01.606150

Rajagopal, A. N. (1996). Growth of Gram-negative bacteria in the presence of organic solvents. *Enzyme and Microbial Technology*, *19*(8), 606–613. https://doi.org/10.1016/S0141-0229(96)00068-3

Ramos, J. L., Cuenca, M. S., Molina-Santiago, C., Segura, A., Duque, E., Ǵomez-Garciá, M. R., Udaondo, Z., & Roca, A. (2015). Mechanisms of solvent resistance mediated by interplay of cellular factors in Pseudomonas putida. *FEMS Microbiology Reviews*, *39*(4), 555–566. https://doi.org/10.1093/FEMSRE/FUV006

Ramos, J. L., Duque, E., Godoy, P., & Segura, A. (1998). Efflux pumps involved in toluene tolerance in Pseudomonas putida DOT- T1E. *Journal of Bacteriology*, *180*(13), 3323–3329. https://doi.org/10.1128/JB.180.13.3323-3329.1998/ASSET/2B711DCC-6B90-4EE9-ADF5-DC7C69FF87DD/ASSETS/GRAPHIC/JB1381737003.JPEG

Reasoner, D. J., & Geldreich, E. E. (1985a). A new medium for the enumeration and subculture of bacteria from potable water. *Applied and Environmental Microbiology*, *49*(1), 1.

Reasoner, D. J., & Geldreich, E. E. (1985b). A new medium for the enumeration and subculture of bacteria from potable water. *Applied and Environmental Microbiology*, *49*(1), 1–7. https://doi.org/10.1128/aem.49.1.1-7.1985

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19126–

19131. https://doi.org/10.1073/PNAS.0906412106/SUPPL_FILE/0906412106SI.PDF

Riggs, A. D. (1981). Bacterial Production of Human Insulin. *Diabetes Care*, *4*(1), 64 LP – 68. https://doi.org/10.2337/diacare.4.1.64

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. https://doi.org/10.1093/BIOINFORMATICS/BTU153

Sharma, R., Kesari, P., Kumar, P., & Tomar, S. (2018). Structure-function insights into chikungunya virus capsid protein: Small molecules targeting capsid hydrophobic pocket. *Virology*, *515*, 223–234. https://doi.org/10.1016/J.VIROL.2017.12.020

Shaw, W. V., Packman, L. C., Burleigh, B. D., Dell, A., Morris, H. R., & Hartley, B. S. (1979). Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. *Nature*, *282*(5741), 870–872. https://doi.org/10.1038/282870A0

Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. *3rd International Symposium on Intelligent Information Technology and Security Informatics, IITSI 2010*, 63–67. https://doi.org/10.1109/IITSI.2010.74

Sikkema, J., Bont, J. A. de, & Poolman, B. (1995). Mechanisms of membrane toxicity of hydrocarbons. *Microbiological Reviews*, *59*(2), 201–222. https://doi.org/10.1128/MR.59.2.201-222.1995

Singh, P., & Vanschoren, J. (2022). *Meta-Learning for Unsupervised Outlier Detection with Optimal Transport*. https://arxiv.org/abs/2211.00372v1

Smith, R. H., Glendinning, L., Walker, A. W., & Watson, M. (2022). Investigating the impact of database choice on the accuracy of metagenomic read classification for the rumen microbiome. *Animal Microbiome*, *4*(1), 1–17. https://doi.org/10.1186/S42523-022-00207-7/FIGURES/6

Spivey, J. J., Gogate, M. R., Zoeller, J. R., & Colberg, R. D. (1997). Novel Catalysts for the Environmentally Friendly Synthesis of Methyl Methacrylate. *Industrial and Engineering Chemistry Research*, *36*(11), 4600–4608. https://doi.org/10.1021/IE970139R

Srikumar, R., Paul, C. J., & Poole, K. (2000). Influence of mutations in the mexR repressor gene on expression of the MexA-MexB-OprM multidrug efflux system of Pseudomonas aeruginosa. *Journal of Bacteriology*, *182*(5), 1410–1414. https://doi.org/10.1128/JB.182.5.1410-1414.2000/ASSET/5ED11891-1F6F-485D-B20A-CBCF0AD0B77C/ASSETS/GRAPHIC/JB0501138002.JPEG

Stegmann, E., Frasch, H. J., Kilian, R., & Pozzi, R. (2015a). Self-resistance mechanisms of actinomycetes producing lipid II-targeting antibiotics. *International Journal of Medical Microbiology*, *305*(2), 190–195. https://doi.org/10.1016/J.IJMM.2014.12.015

Stegmann, E., Frasch, H. J., Kilian, R., & Pozzi, R. (2015b). Self-resistance mechanisms of actinomycetes producing lipid II-targeting antibiotics.

International Journal of Medical Microbiology, *305*(2), 190–195. https://doi.org/10.1016/J.IJMM.2014.12.015

Stevens, A. O., & He, Y. (2022). Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction. *Biomolecules*, *12*(7), 985. https://doi.org/10.3390/BIOM12070985/S1

Striebich, R. C., Smart, C. E., Gunasekera, T. S., Mueller, S. S., Strobel, E. M., McNichols, B. W., & Ruiz, O. N. (2014). Characterization of the F-76 diesel and Jet-A aviation fuel hydrocarbon degradation profiles of Pseudomonas aeruginosa and Marinobacter hydrocarbonoclasticus. *International Biodeterioration & Biodegradation*, *93*, 33–43. https://doi.org/10.1016/J.IBIOD.2014.04.024

Sugawara, E., & Nikaido, H. (2014). Properties of AdeABC and AdeIJK efflux systems of Acinetobacter baumannii compared with those of the AcrAB-TolC system of Escherichia coli. *Antimicrobial Agents and Chemotherapy*, *58*(12), 7250–7257. https://doi.org/10.1128/AAC.03728-14/ASSET/1A15CF86-83CC-49D5-AE6B-F32B9BC094E2/ASSETS/GRAPHIC/ZAC0121434770004.JPEG

Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J. J., Reitz, Z. L., van Santen, J. A., Selem-Mojica, N., Tørring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S. A. A., Augustijn, H. E., … Medema, M. H. (2023a). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids*

Research, *51*(D1), D603–D610. https://doi.org/10.1093/NAR/GKAC1049

Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J. J., Reitz, Z. L., van Santen, J. A., Selem-Mojica, N., Tørring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S. A. A., Augustijn, H. E., … Medema, M. H. (2023b). MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*, *51*(D1), D603–D610. https://doi.org/10.1093/NAR/GKAC1049

The Future of Petrochemicals – Analysis - IEA. (n.d.). Retrieved June 13, 2023, from https://www.iea.org/reports/the-future-of-petrochemicals

The Proteomics Protocols Handbook. (2005). *The Proteomics Protocols Handbook*. https://doi.org/10.1385/1592598900

Tian, Z. X., Yi, X. X., Cho, A., O'Gara, F., & Wang, Y. P. (2016). CpxR Activates MexAB-OprM Efflux Pump Expression and Enhances Antibiotic Resistance in Both Laboratory and Clinical nalB-Type Isolates of Pseudomonas aeruginosa. *PLOS Pathogens*, *12*(10), e1005932. https://doi.org/10.1371/JOURNAL.PPAT.1005932

Udaondo, Z., Molina, L., Daniels, C., Gómez, M. J., Molina-Henares, M. A., Matilla, M. A., Roca, A., Fernández, M., Duque, E., Segura, A., & Ramos, J. L. (2013). Metabolic potential of the organic-solvent tolerant Pseudomonas putida DOT-T1E deduced from its annotated genome.

Microbial Biotechnology, *6*(5), 598–611. https://doi.org/10.1111/1751-7915.12061

Valadbeigi, H., Sadeghifard, N., Kaviar, V. H., Haddadi, M. H., Ghafourian, S., & Maleki, A. (2023). Effect of ZnO nanoparticles on biofilm formation and gene expression of the toxin-antitoxin system in clinical isolates of Pseudomonas aeruginosa. *Annals of Clinical Microbiology and Antimicrobials*, *22*(1), 1–8. https://doi.org/10.1186/S12941-023-00639-2/TABLES/3

Verhoef, S., Wierckx, N., Westerhof, R. G. M., De Winde, J. H., & Ruijssenaars, H. J. (2009). Bioproduction of p-hydroxystyrene from glucose by the solvent-tolerant bacterium Pseudomonas putida S12 in a two-phase water-decanol fermentation. *Applied and Environmental Microbiology*, *75*(4), 931–936. https://doi.org/10.1128/AEM.02186-08/ASSET/482AA561-7936-4AA3-9B6C-AF7FFCDB9E1A/ASSETS/GRAPHIC/ZAM0040996180003.JPEG

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*, *9*(11), e112963. https://doi.org/10.1371/JOURNAL.PONE.0112963

Wang, H., & Ng, T. B. (1999). Pharmacological activities of fusaric acid (5-butylpicolinic acid). *Life Sciences*, *65*(9), 849–856. https://doi.org/10.1016/S0024-3205(99)00083-1

Wang, S., Cui, J., Bilal, M., Hu, H., Wang, W., & Zhang, X. (2020). Pseudomonas spp. as cell factories (MCFs) for value-added products: from rational design to industrial applications. *Critical Reviews in Biotechnology*, *40*(8), 1232–1249. https://doi.org/10.1080/07388551.2020.1809990

Webb, J. P., Arnold, S. A., Baxter, S., Hall, S. J., Eastham, G., & Stephens, G. (2018). Efficient bio-production of citramalate using an engineered escherichia coli strain. *Microbiology (United Kingdom)*, *164*(2), 133–141. https://doi.org/10.1099/mic.0.000581

Weber, F. J., Ooijkaas, L. P., Schemen, R. M. W., Hartmans, S., & De Bont, J. A. M. (1993). Adaptation of Pseudomonas putida S12 to high concentrations of styrene and other organic solvents. *Applied and Environmental Microbiology*, *59*(10), 3502–3504. https://doi.org/10.1128/AEM.59.10.3502-3504.1993

Weiser, R., Green, A. E., Bull, M. J., Cunningham-Oakes, E., Jolley, K. A., Maiden, M. C. J., Hall, A. J., Winstanley, C., Weightman, A. J., Donoghue, D., Amezquita, A., Connor, T. R., & Mahenthiralingam, E. (2019a). Not all Pseudomonas aeruginosa are equal: strains from industrial sources possess uniquely large multireplicon genomes. *Microbial Genomics*, *5*(7). https://doi.org/10.1099/MGEN.0.000276

Weiser, R., Green, A. E., Bull, M. J., Cunningham-Oakes, E., Jolley, K. A., Maiden, M. C. J., Hall, A. J., Winstanley, C., Weightman, A. J., Donoghue, D., Amezquita, A., Connor, T. R., & Mahenthiralingam, E. (2019b). Not all Pseudomonas aeruginosa are equal: strains from

industrial sources possess uniquely large multireplicon genomes. *Microbial Genomics*, *5*(7). https://doi.org/10.1099/MGEN.0.000276

Weiser, R., Green, A. E., Bull, M. J., Cunningham-Oakes, E., Jolley, K. A., Maiden, M. C. J., Hall, A. J., Winstanley, C., Weightman, A. J., Donoghue, D., Amezquita, A., Connor, T. R., & Mahenthiralingam, E. (2019c). Not all Pseudomonas aeruginosa are equal: strains from industrial sources possess uniquely large multireplicon genomes. *Microbial Genomics*, *5*(7). https://doi.org/10.1099/MGEN.0.000276

Wiesner, J., Ziemann, C., Hintz, M., Reichenberg, A., Ortmann, R., Schlitzer, M., Fuhst, R., Timmesfeld, N., Vilcinskas, A., & Jomaa, H. (2016). FR-900098, an antimalarial development candidate that inhibits the non-mevalonate isoprenoid biosynthesis pathway, shows no evidence of acute toxicity and genotoxicity. *Virulence*, *7*(6), 718. https://doi.org/10.1080/21505594.2016.1195537

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology 2019 20:1*, *20*(1), 1–13. https://doi.org/10.1186/S13059-019-1891-0

Yang, F., Zhang, Y., & Liang, H. (2014). Interactive Association of Drugs Binding to Human Serum Albumin. *International Journal of Molecular Sciences 2014, Vol. 15, Pages 3580-3595*, *15*(3), 3580–3595. https://doi.org/10.3390/IJMS15033580

Yao, X., Tao, F., Tang, H., Hu, H., Wang, W., & Xu, P. (2021). Unique regulator SrpR mediates crosstalk between efflux pumps TtgABC and SrpABC in

Pseudomonas putida B6-2 (DSM 28064). *Molecular Microbiology*, *115*(1), 131–141. https://doi.org/10.1111/MMI.14605

Yao, X., Tao, F., Zhang, K., Tang, H., & Xu, P. (2017). Multiple roles for two efflux pumps in the polycyclic aromatic hydrocarbon-degrading Pseudomonas putida strain B6-2 (DSM 28064). *Applied and Environmental Microbiology*, *83*(24). https://doi.org/10.1128/AEM.01882-17/ASSET/E4934E40-96B1-4D18-A5AB-DD2805678C8D/ASSETS/GRAPHIC/ZAM0241782000006.JPEG

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, *351*(6278), 1196–1199. https://doi.org/10.1126/SCIENCE.AAD6359/SUPPL_FILE/AAD6359-YOSHIDA-SM.PDF

Yoshida, S., Hiraga, K., Taniguchi, I., & Oda, K. (2021). Ideonella sakaiensis, PETase, and MHETase: From identification of microbial PET degradation to enzyme characterization. *Methods in Enzymology*, *648*, 187–205. https://doi.org/10.1016/BS.MIE.2020.12.007

Youenou, B., Favre-Bonté, S., Bodilis, J., Brothier, E., Dubost, A., Muller, D., & Nazaret, S. (2015). Comparative Genomics of Environmental and Clinical Stenotrophomonas maltophilia Strains with Different Antibiotic Resistance Profiles. *Genome Biology and Evolution*, *7*(9), 2484–2505. https://doi.org/10.1093/GBE/EVV161

Zhang, L., Fan, T. P., Cai, Y., & Zheng, X. (2024a). Production of mevalonate in Pseudomonas putida via tuning the expression of pathway gene. *Systems Microbiology and Biomanufacturing*, *4*(3), 1162–1173. https://doi.org/10.1007/S43393-023-00225-9/FIGURES/4

Zhang, L., Fan, T. P., Cai, Y., & Zheng, X. (2024b). Production of mevalonate in Pseudomonas putida via tuning the expression of pathway gene. *Systems Microbiology and Biomanufacturing*, *4*(3), 1162–1173. https://doi.org/10.1007/S43393-023-00225-9/FIGURES/4

Zhu, X., Gao, Y., Wang, X., Haribal, V., Liu, J., Neal, L. M., Bao, Z., Wu, Z., Wang, H., & Li, F. (2021). A tailored multi-functional catalyst for ultra-efficient styrene production under a cyclic redox scheme. *Nature Communications*, *12*(1). https://doi.org/10.1038/S41467-021-21374-2

Zou, C., Zhao, Q., Zhang, G., & Xiong, B. (2016). Energy revolution: From a fossil energy era to a new energy era. *Natural Gas Industry B*, *3*(1), 1–11. https://doi.org/10.1016/j.ngib.2016.02.001