Combining Machine Learning and CRISPRi to Discover Novel Antimicrobials Against *Pseudomonas aeruginosa*

by

Julieta Novomisky Nechcoff

A Thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfilment of the requirement of the degree of

MASTER OF SCIENCE

Department of Microbiology

University of Manitoba

Winnipeg, Manitoba, Canada

Copyright © 2025 by Julieta Novomisky Nechcoff

DEDICATION

To Viviana and Sergio, because without them I would not be here.

To Matías and Tomás, because I always want to make them proud.

ABSTRACT

The rise in antibiotic-resistant infections has led to an increasing need for novel antibiotics. However, traditional methods are becoming obsolete, as they provide hit rates lower than 1%. Moreover, these methods do not provide any insight into the mechanism of action of the discovered drugs, hindering compound prioritization and further increasing the costs and time associated with drug discovery.

Machine learning approaches offer a promising alternative to traditional methods, increasing hit rates and reducing costs. In this thesis, we use these methods to find promising growth inhibitor compounds against *Pseudomonas aeruginosa.* Nonetheless, because the predictive accuracy of machine learning models depends on the quality of the training dataset, we also analyzed how different datasets and model architectures influence the model's performance in antibiotic discovery.

Additionally, we validated the results of a previous chemical-genetic interaction profile in *Burkholderia cenocepacia,* which identified active compounds against *Pseudomonas aeruginosa* and other high-public health threat pathogens. Using a rhamnose-inducible CRISPRi system in *P. aeruginosa* PA14, we explored the mechanism of action of ZR50, a promising compound that inhibits *fda* knockdown mutants, suggesting it targets glycolysis.

By integrating ML-driven prediction with CRISPRi-based validation, this study advances antibiotic discovery and the determination of mechanisms of action, providing a comprehensive strategy for combating antimicrobial resistance.

ACKNOWLEDGEMENTS

I would like to start by thanking my supervisor, Dr. Silvia T. Cardona, for her unwavering support inside and outside the lab. Silvia, you have mentored me to become a better researcher, you have supported my decisions even when you did not completely agree, and you gave me a place to grow professionally without being judged. I would be forever grateful to you for giving me the opportunity to work in your lab.

I would also like to provide a heartfelt thank you to my committee member, Dr. Rebecca Davis and Dr. Michael Becker, for your insightful feedback and support during my degree.

I extend my appreciation to all my collaborators for their invaluable help to this project. Your knowledge and dedication made the completion of this project possible, and I am thankful I had the opportunity to learn and work alongside all of you.

Thank you to all past and present members of the Cardona lab for being awesome lab mates and friends. A huge shoutout to Dr. Zisanur Rahman for being my mentor since day one, for welcoming me, being patient, and for becoming a dear friend. Thank you to all the students I have mentored, your contribution to the completion of this thesis has been invaluable. Special thanks to Dr. Zhong-Ling Yap, Juliana Osler, Sion Yi, Miranda Palumbo, and Sarah Iqbal for being there for me, for the chats, and the science discussions that made every day more enjoyable. I have a full heart working alongside all of you, and I am thankful for the memories we made together.

A heartfelt thank you to everyone in the Department of Microbiology, especially the members of the Wieden and Kumar lab, for always being there when I needed a hand or a sounding

board. Thank you for being such a great department, for all the fun events, the advice, and the happy memories we made together.

I would like to thank my funding sources over the years, without them a lot of this would not have been possible: CIHR, University of Manitoba Faculty of Science and Faculty of Graduate Studies, and the Molly Weinberg Soudack Fund.

Thank you to all my friends and family in Argentina, for always supporting me and welcoming with open arms whenever we see each other, particularly Juan Mac Donagh because even 11 thousand kilometers could not break our friendship. And an especial thank you to all my friends in Winnipeg, for making these years less lonely, for filling my life with laughs and love, and for making even the coldest winters warm.

A huge thank you to Amanda Vazquez Gonzalez, for being an awesome housemate and an even better friend. Your steadfast support these last 3 years have meant the world to me. Thank you for being there in the tough times, for all the coffees and dinners, and for always giving great life and science advice.

And finally, thank you to my parents Sergio and Viviana, and my brothers Matías and Tomás, because they make me want to be better, and because without their unwavering support – even at a distance – I would not be the person I am today.

TABLE OF CONTENTS

DEDICATION ..... ***II***

ABSTRACT ..... ***III***

ACKNOWLEDGEMENTS ..... ***IV***

TABLE OF CONTENT ..... ***VI***

CONTRIBUTIONS ..... ***IX***

LIST OF TABLES ..... ***X***

LIST OF FIGURES ..... ***XI***

LIST OF ABBREVIATIONS ..... ***XII***

CHAPTER I ***1***

1. Introduction **1**

1.1- Pseudomonas aeruginosa 2

1.2- Antibiotic Discovery 5

1.3- Machine Learning in Antibiotic Discovery 6

1.3-1. Machine Learning (ML) and Deep Learning (DL) 6

1.4- Essential Genes and Mechanism of Action of Novel Antibiotics 9

1.4-1. CRISPRi 11

1.5- Rationale 14

1.6- Hypothesis 15

1.7- Objectives 15

CHAPTER II *16*

2. Results **16**

2.1- Applying ML to drug discovery increases hit rates by 8% 17

2.2- Imbalanced training datasets decrease the prediction performance of compound activity 22

2.3- Benchmarking predictive performance using "spike" molecules 26

2.4- CRISPRi in P. aeruginosa PA14 selectively silence native gene expression 28

2.5- CRISPRi silencing of fda confers hypersusceptibility to ZR50 37

2.6- Hit compounds have pharmacokinetic properties comparable to other known antibiotics 39

CHAPTER III *40*

3. Discussion and Future Directions **40**

3.1- Discussion 41

3.2- Future directions 45

CHAPTER IV *47*

4. Methods and Materials **47**

4.1- Bacterial strains and growth conditions 48

4.2- Compound formulation 54

4.3- Compound susceptibility testing using minimal inhibitory concentration assays 55

4.4- Small molecule library optimization 56

4.5- t-SNE visualization 56

4.6- **Creation of the discovery set** 57

4.7- Machine Learning methods 58

4.8- Hyperparameter optimization 59

4.9- Evaluation methods 59

4.10- Construction of a *P. aeruginosa* dCas9 insertional mutant 60

4.11- Construction of the sgRNA-expressing plasmids 60

4.12- Insertion of sgRNAs into pJO4 62

4.13- Extraction of plasmid and genomic DNA 63

4.14- Genome sequencing 63

4.15- SDS-PAGE and immunoblotting 63

4.16- Growth of *P. aeruginosa* in dilutions of CAMHB and biofilm quantification 64

4.17- Growth curves 65

4.18- Rhamnose dose-response curves 66

4.19- Hypersusceptibility assay with glucose as the sole carbon source 66

CHAPTER V ***67***

5. Appendix **67**

5.1- Supplemental tables 68

5.2- Supplemental figures 74

CHAPTER VI ***76***

6. References **76**

CONTRIBUTIONS

Multiple authors have contributed to the work presented in Chapter II.

H. Hadipour trained the machine learning models. H. Sturm cleaned the training datasets and prepared the discovery set. I.I. Oyshe and R. Seales contributed to the minimal inhibitory concentration assays of JN19 and JN25.

S. Patel aided in the construction of the dCas9 insertional mutant. J. Osler and S. Yi contributed to the design and creation of the sgRNA-expressing plasmids. I.I. Oyshe and R. Seales contributed to the minimal inhibitory concentration assays of ZR50.

While these individuals performed select experiments under my supervision, I designed and performed all other experiments listed experiments and was responsible for the data analysis, figure preparation, and writing and editing the manuscript under the guidance of Dr. Silvia Cardona.

LIST OF TABLES

Table 1. Different models were trained with various configurations to extract molecular-level features. 18

Table 2. Original training datasets that were used in this study. 24

Table 3. Table of mutants depleted in the CGIP experiment against *B. cenocepacia* K56-2 and their homologous genes in *P. aeruginosa* PA14 34

Table 4. List of strains used in this study 48

Table 5. Plasmids used in study..... 51

Table 6. Primers and oligonucleotides used in study 52

Table S1. Training datasets used in this study..... 68

Table S2. Discovery set 1 used for predictions. 25 compounds (JN01-JN25) were acquired and experimentally tested. Attached Excel file..... 68

Table S3. Acquired compounds for experimental confirmations 68

Table S4. Data cleaning results 68

Table S5. Discovery set 2 used for predictions. Attached Excel file..... 69

Table S6. Spike compounds 70

Table S7. IC50 of rhamnose for the different mutants 73

LIST OF FIGURES

Figure 1. Electronic microscopies of *Pseudomonas aeruginosa* ATCC27853 3

Figure 2. Hypersusceptibility effect in knockdown mutants 10

Figure 3. Overview of CRISPR interference 12

Figure 4. Machine learning driven predictions of active compounds 19

Figure 5. JN25 has inhibitory activity against MDR strains of *P. aeruginosa* 21

Figure 6. Visualization of the chemical space and model training 25

Figure 7. "Spike" molecules are an internal benchmark to assess model performance 28

Figure 8. Creation of a CRISPRi system in *P. aeruginosa* PA14 30

Figure 9. ZR50 has inhibitory activity against different strains of *P. aeruginosa* 33

Figure 10. sgRNAs targeting the initiation of transcription show the strongest conditional growth phenotypes 36

Figure 11. CRISPRi in *P. aeruginosa* PA14 effectively silences native gene expression and confers hypersusceptibility to ZR50 38

Figure 12. Hit compounds have comparable ADMET properties to other common antibiotics 40

Figure S1. Two-dimensional visualization of training datasets 74

Figure S2. Diluting the nutrients in CAMHB inhibits biofilm formation without affecting cell growth 75

LIST OF ABBREVIATIONS

ADMET	Absorption, Distribution, Metabolism, Excretion, And Toxicity
Amp	Ampicillin
AMR	Antimicrobial Resistant
ATG	Translation Start Site
AUPRC	Area Under the Precision-Recall Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
BBB	Blood-Brain Barrier
BLAST	Basic Local Alignment Search Tool
CAMHB	Cation Adjusted Mueller-Hinton Broth
Cas	CRISPR-Associated Proteins
CF	Cystic Fibrosis
CGIP	Chemical-Genetic Interaction Profile
CLSI	Clinical And Laboratory Standards Institute
CNN	Convolutional Neural Network
CO-ADD	Community For Open Antimicrobial Drug Discovery
CRISPR	Clustered Regularly Interspaced Palindromic Repeats
CRISPRi	Clustered Regularly Interspaced Palindromic Repeats Interference
D-MPNN	Directed Message Passing Neural Network
dCas9	Catalytically Dead Cas9
DL	Deep Learning
DMSO	Dimethyl Sulfoxide

dsDNA	Double-Stranded DNA
ECFP	Extended-Connectivity Fingerprints
ESKAPE	*Enterobacter cloacae, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, Enterococcus faecium*
FBA	Fructose-1,6-bisphosphate aldolase
FBP	Fructose-1,6-bisphosphate
FCN	Fully Connected Network
FFNN	Feed Forward Neural Network
GCN	Graph Convolutional Networks
HAP	Hospital-Acquired Pneumonia
HCl	Hydrochloric Acid
hERG	Human Ether-a-go-go Related Gene
HTS	High-Throughput Screen
IC50	Inhibitory Concentration 50%
Kan	Kanamycin
KPGT	Knowledge-Guided Pre-training Of Graph Transformers
LESB58	*Pseudomonas aeruginosa* LESB58
LES400	*Pseudomonas aeruginosa* LES400
LESlike1	*Pseudomonas aeruginosa* LESlike1
LESlike5	*Pseudomonas aeruginosa* LESlike5
LB	Lysogeny Broth
MDR	Multidrug-Resistant

MHB	Mueller-Hinton Broth
MIC	Minimal Inhibitory Concentration
ML	Machine Learning
MoA	Mechanism Of Action
NEB	New England Biolabs
NTC	Non-Targeting Control
OD600	Optical Density Measured At 600 Nm
OMPs	Outer Membrane Protein
PA14	*Pseudomonas aeruginosa* PA14
PAM	Protospacer Adjacent Motif
PAO1	*Pseudomonas aeruginosa* PAO1
PBS	Phosphate-Buffered Saline Solution
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
Rha	Rhamnose
SDS	Sodium Dodecyl Sulfate
sgRNA	Single-Guide RNA
SMILES	Simplified Molecular Input Line Entry System
SMOTE	Synthetic Minority Over-Sampling Technique
SOC	Super Optimal Broth with Catabolite Repression
Spy	Streptococcus Pyogenes
t-SNE	T-Distributed Stochastic Neighbor Embedding

Tet	Tetracycline
Tp	Trimethoprim
tracrRNA	Trans-Acting RNA
WHO	World Health Organization
WT	Wild Type

CHAPTER I

1. Introduction

1.1- Pseudomonas aeruginosa

In the last decades, overpopulation, globalization, and the increased use and misuse of antimicrobials have led to an accelerated spread of antibiotic-resistant infections (Aslam et al. 2018; Asokan et al. 2019), exemplified by the growth of carbapenem-resistant strains of *Pseudomonas aeruginosa.* At least 700,000 people die each year due to antibiotic-resistant infections – a number that reached 4.95 million deaths globally in 2019 (Murray et al. 2022). Looking forward, projections estimate that by 2050, antimicrobial resistance (AMR) could be responsible for 10 million annual deaths (Miethke et al. 2021). Yet despite these alarming numbers, the pace of antibiotic discovery has not kept pace with the rapid evolution of resistance, creating an unfavorable market and discouraging pharmaceutical companies from investing in antimicrobial development (Cardona et al. 2015; Miethke et al. 2021).

Pseudomonas aeruginosa has been listed as a priority pathogen by the World Health Organization, due the rapid development of carbapenem-resistant strains. This bacterium stands as a notable member of the *Pseudomonas* genus which is comprised of more than 120 rod-shaped, Gram-negative species ubiquitous to moist environments and capable of establishing infections in both plants and animals, especially humans (Streeter et al. 2016; Langendonk et al. 2021; Grace et al. 2022). Its widespread presence and ability to infect diverse hosts have established *P. aeruginosa* as a bacterium of exceptional ecological and medical concern.

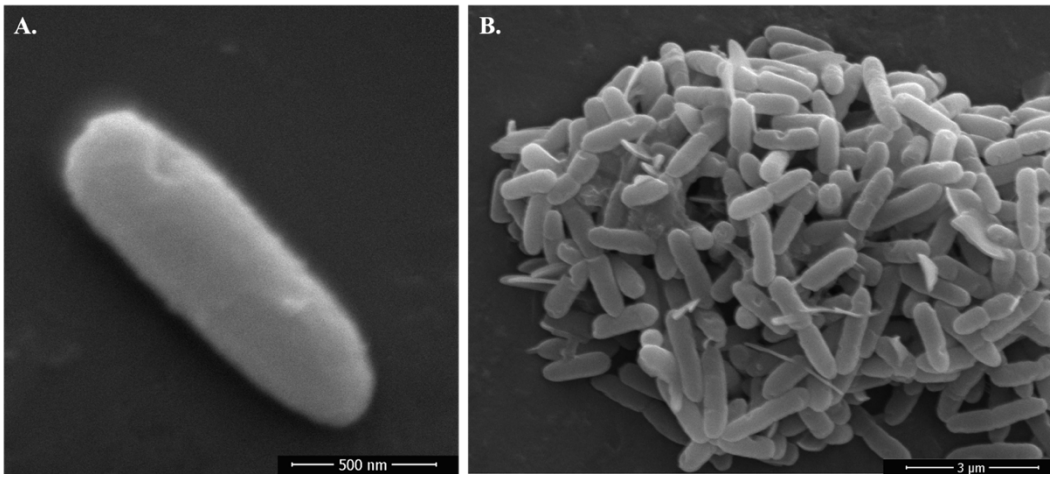


Figure 1. ***Scanning electron micrographs of Pseudomonas aeruginosa*** **ATCC27853.** *(A)* Individual cell of *P. aeruginosa* grown in Lysogeny broth, an*d (B)* agglomerate of cells grown in Cation-Adjusted Mueller-Hinton broth.

One of the reasons for *P. aeruginosa*'s clinical relevance is its metabolic versatility and natural antibiotic resistance. This resistance is greatly attributed to the species' significantly large genome of 5.5-7.0 Mbp and its remarkable ease for acquiring resistance genes from nosocomial reservoirs (Stover et al. 2000). Moreover, the bacterium's resistance is even more pronounced in infections where there is biofilm formation, as the extracellular matrix formed acts as a barrier that stops antibiotics from reaching the bacterial cells (Grace et al. 2022).

Clinically, *P. aeruginosa* is most often associated with nosocomial infections, particularly infecting immunocompromised individuals. This bacterium is particularly linked to respiratory tract infections, especially in cystic fibrosis (CF) patients, where it can colonize the airways and cause persistent, chronic lung infections that contribute to elevated rates of morbidity and mortality (Nixon et al. 2001); nonetheless, non-CF patients who are immunocompromised are also vulnerable, with hospital-acquired pneumonia (HAP) occurring often in patients that use mechanical ventilation, a situation especially affected by the ability of the is bacterium to form biofilms in medical equipment (Parker et al. 2008; Fujitani et al. 2011).

In addition to respiratory issues, catheterized patients are at high risk of urinary tract infections caused by *P. aeruginosa* biofilms on catheter surfaces (Mittal et al. 2009). This bacterium is also commonly isolated from severe burn and wound infections, which carry a high global burden due to their potential to enter the bloodstream and cause sepsis (Gang et al. 1999; Armour et al. 2007; Ressner et al. 2008).

Another factor enhancing virulence is the bacterium's production of pyocyanin, a redox-active blue/green pigment that acts as an antimicrobial against competing bacteria, enabling *P. aeruginosa* to dominate both infection sites and environmental niches (Zhou et al., 2022c; Grace et al., 2022). *P. aeruginosa*'s molecular versatility and the early knowledge on their genome sequence had led researchers to focus on the PAO1 strain of *P. aeruginosa.* However, there has been a shift toward the use of *P. aeruginosa* PA14 – a burn wound isolate - in pathogenesis studies due to its expanded host range and increased virulence (Mathee 2018; Grace et al. 2022). The heightened pathogenicity of PA14 is attributed to the presence of multiple pathogenicity islands. Moreover, it contains around 150 genes for outer membrane proteins (OMPs), which play roles in antimicrobials efflux and contribute to the secretion of biofilm-forming components (Grace et al. 2022).

Emerging strains, such as the Liverpool Epidemic Strains (LES) and LESlike strains, are also gaining attention in antibiotic research. These strains were the first epidemic variants identified in CF clinics in the UK and later in Ontario, Canada. Their resistance profiles are extensive, and they have been linked to increased transmission among CF patients and their families (Goodyear et al. 2021, 2022)

Due to the high incidence of fatal infections, combined with both innate and acquired resistance to conventional antibiotics – including the rise of strains resistant to the last-resort antibiotic carbapenem – *P. aeruginosa* has been classified by the World Health Organization (WHO) as one of the pathogens most urgently required new antimicrobial therapies (*WHO Bacterial Priority Pathogens List 2024*).

1.2- Antibiotic Discovery

The Golden Age of antibiotics – spanning the 1940s to 1960s – was a period of remarkable innovation, with new antibiotics being discovered almost yearly (Gould 2016). However, in the past 50 years, only one truly novel antibiotic, the narrow-spectrum daptomycin, has reached clinical use (Lewis 2012). As of 2021, only 40 new antibiotics were undergoing clinical trials, and a significant portion of these are derivatives of existing classes. More importantly, most of these drugs do not target the WHO's priority pathogens, leaving critical gaps in our treatment arsenal (Miethke et al. 2021; Cardona et al. 2025).

This slow progression in antibiotic discovery is further compounded by the inefficiency of current discovery approaches. Although science has advanced dramatically, the techniques used to uncover new compounds remain largely unchanged. Both target-based and whole-cell assays depend on high-throughput screenings (HTS), which tests large libraries of compounds for activity. Unfortunately, these methods yield active compounds at a rate of just 0.1%, and the volume of compounds that can be screened is heavily constrained by time and resources (Payne et al. 2007; Miethke et al. 2021; Cardona et al. 2025).

Even once a promising compound is identified, the development process demands substantial investments – often millions of dollars and decades of research – to bring a candidate

to market. These downstream challenges further delay the antibiotic pipeline and limit the global response to resistance infections (Miethke et al. 2021).

1.3- Machine Learning in Antibiotic Discovery

One of the most pressing bottlenecks in antibiotic discovery lies in the limitations of traditional screening methods. HTS, while essential, are constrained by cost-time, and low success rates (Payne et al. 2007; Miethke et al. 2021; Cardona et al. 2025). To solve these issues, researchers have turned to artificial intelligence (AI), particularly machine learning (ML), to revolutionize how antibiotics are discovered. AI-driven approaches offer a way to merge conventional experimentation with computational power, drastically accelerating early-stage discovery (Angermueller et al. 2016; Chan et al. 2019; Stokes et al. 2020; Arnold et al. 2023; Wong et al. 2023b).

ML models have already demonstrated their utility in identifying antimicrobial compounds effective against numerous multidrug-resistant (MDR) pathogens, including *Escherichia coli* (Wong et al. 2022)*, Staphylococcus aureus* (Wong et al. 2023b), *Acinetobacter baumannii* (Stokes et al. 2020), and *Burkholderia cenocepacia* (Rahman et al. 2022). By integrating predictive analysis with experimental data, these models can boost hit rates – sometimes by up to 12-fold – and uncover compounds with novel mechanisms of action (MoA), capable of sidestepping existing resistance strategies (Camacho et al. 2018).

1.3-1. Machine Learning (ML) and Deep Learning (DL)

ML algorithms represent a critical turning point in the drug discovery landscape, offering ways to investigate massive and complex datasets that would be unfeasible to analyze manually.

At their core, ML models are trained to recognize patterns within data – such as chemical structures or biological activity – and use those insights to predict outcomes in new, unseen compounds. This is especially powerful in drug discovery, where identifying therapeutic candidates requires navigating immense chemical diversity (Tarca et al. 2007; Lluka and Stokes 2023).

The typical ML workflow begins with preparing and processing input data, followed by iterative training where the model fine-tunes its internal parameters to make increasingly accurate predictions. Once trained, the model is applied to new data to predict features of interest, such as antimicrobial activity (Camacho et al. 2018). ML approaches can be broadly categorized into supervised and unsupervised learning. Supervised learning uses labeled data, where known outcomes (e.g., inhibitory activity) are linked to specific compound features. The model learns to associate features to labels, enabling it to predict the label of unknown molecules with similar characteristics (Tarca et al. 2007; Camacho et al. 2018; Lluka and Stokes 2023). In drug development, this method has already led to significant discoveries – for example, PHAR(A), a novel peptidyl-tRNA synthetase inhibitor (Rahman et al. 2022), and Halicin, a broad-spectrum antimicrobial were identified through ML-guided screening (Stokes et al. 2020).

Unsupervised learning, in contrast, does not rely on labeled data. Instead, it groups compounds based on feature similarities, allowing researchers to identify structural or functional trends within novel chemical spaces. Common methods include clustering algorithms like k-means, or dimension-reduction tools such as principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE), which helps visualize high-dimensional data patterns (Tarca et al. 2007; Camacho et al. 2018; Lluka and Stokes 2023).

A particularly transformative subclass of ML is deep learning (DL), which excels at uncovering pattern recognition within complex datasets (Hochreiter et al. 2018; Liu et al. 2022). DL models employ neural networks formed by at least three layers: input, hidden, and output. In the input layer, the training data is given to the model as vectors; the hidden layers act as engines of analysis, progressively extracting relevant features from molecular data and correlating them with the labels; and finally, in the output layer, the model encounters the new and unseen data and performs the predictions (Liu et al. 2022; Lluka and Stokes 2023).

The neurons encoded in the hidden layers of the DL can transmit information to each other, forming what is called a neural network. The greater the number of hidden layers, the deeper the neural network is. More properties can be extracted from the features – as each hidden layer learns the properties of the features by taking as input the output of the previous layer; moreover, these hidden layers can iterate through the data passage, optimizing their internal parameters to minimize errors during the prediction stage (Camacho et al. 2018).

To make chemical data understandable for DL models, molecules are encoded using formats such as SMILES (Simplified Molecular Input Line Entry System, Weininger 1988). These representations are then translated into graph structures where atoms are nodes and bonds are edges. Graph convolutional networks (GCNs), a type of DL model, process these molecular graphs and learn which structural elements are associated with specific antimicrobial (Camacho et al. 2018; Chan et al. 2019; Zhou et al. 2020). Because DL models are typically supervised, they thrive on the ability to link learned features to biological activity, offering researchers a powerful tool to confront the growing threat of antibiotic resistance accelerating drug discovery.

1.4- Essential Genes and Mechanism of Action of Novel Antibiotics

Even as ML accelerates the early stages of drug discovery, a persistent bottleneck remains: the limited understanding of how newly identified compounds work. Mechanism of action (MoA) studies are crucial not only for elucidating how a drug inhibits growth but also for anticipating how resistance might emerge. Without this knowledge, drug approval processes are delayed, and development costs rise – adding to the already strained pipeline (Miethke et al. 2021; Wong et al. 2023a; Halawa et al. 2024).

The classification of gene essentiality plays a crucial role in drug development and MoA investigation, as most antibiotics target the product of essential genes (Mobegi et al. 2014; Cardona et al. 2015; Hogan and Cardona 2022). Though essentiality is influenced by environmental and nutritional factors, researchers generally categorize genes into three tiers: essential (necessary for growth in rich media), conditionally essential (required only under specific conditions), and non-essential genes (not vital for survival but offering competitive advantages) (D'Elia et al. 2009; Cardona et al. 2015; Hogan and Cardona 2022).

One particularly insightful approach to MoA investigation is analyzing hypersusceptibility; by modulating the expression, and therefore the bioavailability, of an essential gene whose product is targeted by a drug, researchers can evaluate how changes in protein abundance affect drug susceptibility. When expression is knocked down, fewer copies of the target protein are present, and a smaller concentration of the compound is sufficient to inhibit growth, indicating potential target specificity (**Figure 2**) (Cardona et al. 2015). This dose-susceptibility shift is a powerful clue in mapping the biochemical action of unknown drugs.

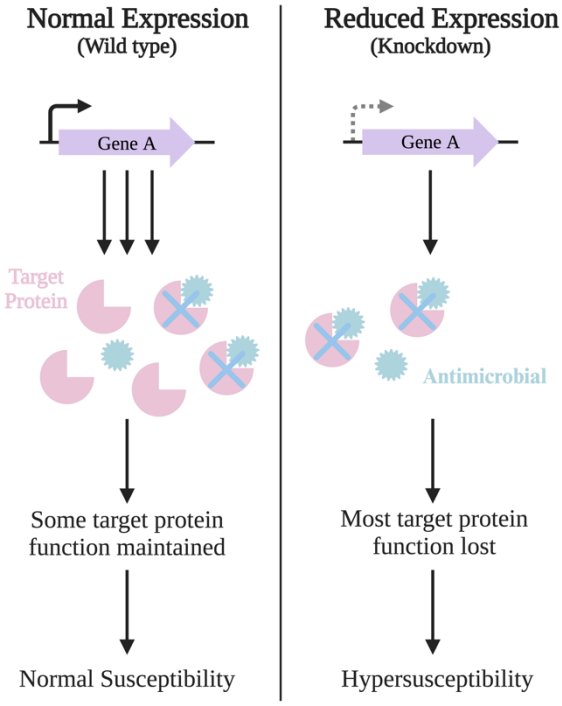


Figure 2. ***Hypersusceptibility effect in knockdown mutants***. In wild-type conditions, there is normal expression of an essential gene (purple arrow), and a certain concentration of inhibitory compound (light blue circles) is needed to inhibit protein (pink shapes) activity and detect cell susceptibility; however, when a knockdown mutation generates reduced expression of the same essential gene, a lower concentration of compound is needed to see cell susceptibility.

Most antibiotics target the product of essential genes, like the anti-tuberculosis drug rifampicin, which targets the β-subunit of the RNA polymerase, and is encoded by the essential gene *rpoB* (Campbell et al. 2001); novobiocin, an antimicrobial used to treat infections by *Staphylococcus aureus,* targets the GyrB subunit of the DNA gyrase, product of the essential *gyrB* gene (Gellert et al. 1976), or the antimicrobial C109 inhibits growth by inhibiting the assembly of the Z-ring during cellular division by targeting FtsZ, product of the essential gene *ftsZ* (Hogan et al. 2018). However, the essential nature of these genes prevents their complete study, indicating roadblocks in their investigation. Several methods have been develop to downregulate the expression of essential genes, one being CRISPR interference (CRISPRi), which allows for the

reversible and tunable gene regulation and can be used to observe hypersusceptibility without compromising cell viability (Hogan and Cardona 2022).

1.4-1. CRISPRi

The clustered regularly interspaced palindromic repeats (CRISPR) system, originally discovered as a microbial defense mechanism, has been adapted into one of the most versatile tools in modern genetic engineering. In bacteria and archaea, CRISPR functions like an immune system, allowing cells to recognize and neutralize invading genetic elements such as phages (Makarova et al. 2011). Following its breakthrough adaptation for genome editing in 2012 (Jinek et al. 2012; Gasiunas et al. 2012), CRISPR's potential expanded rapidly, leading us into a new era of RNA-programmable genetic manipulation (Backes and Phillips 2021).

The commonly used Type II CRISPR system from *Streptococcus pyogenes* involves the CRISPR-associated protein 9 (Cas9), which requires two non-coding RNAs: the trans-activating CRISPR RNA (tracrRNA) and the CRISPR RNA (crRNA). The tracrRNA is structurally inherent to the host system, while the crRNA provides specificity through a "memory" sequence derived from previous genetic invaders. Together, these RNAs hybridize to form a guide RNA (gRNA), which directs the Cas9 to a target DNA sequence located adjacent to a protospacer adjacent motif (PAM) – typically 5'-NGG-3' for *S. pyogenes* Cas9. Upon binding, Cas9 unwinds the DNA and induces a double-stranded break (Backes and Phillips 2021; Bendixen et al. 2023).

While CRISPR-Cas9 was initially designed for gene editing; researchers have since adapted it for transcriptional regulation through a method known as CRISPRi (Qi et al. 2013). This technique involves mutation the RuvC-like and HNH catalytic domains of the Cas9 to create a "dead" Cas9 (dCas9), which lacks nuclease activity but still binds tightly to the target DNA. When complexed with a gRNA, the dCas9 physically blocks the RNA polymerase from initiating or elongating transcription, effectively silencing gene expression without altering the genome **(Figure 3)** (Backes and Phillips 2021; Bendixen et al. 2023; Cai et al. 2023).

CRISPRi's power lies in its specificity, tunability, and reversibility. By placing the *dcas9*

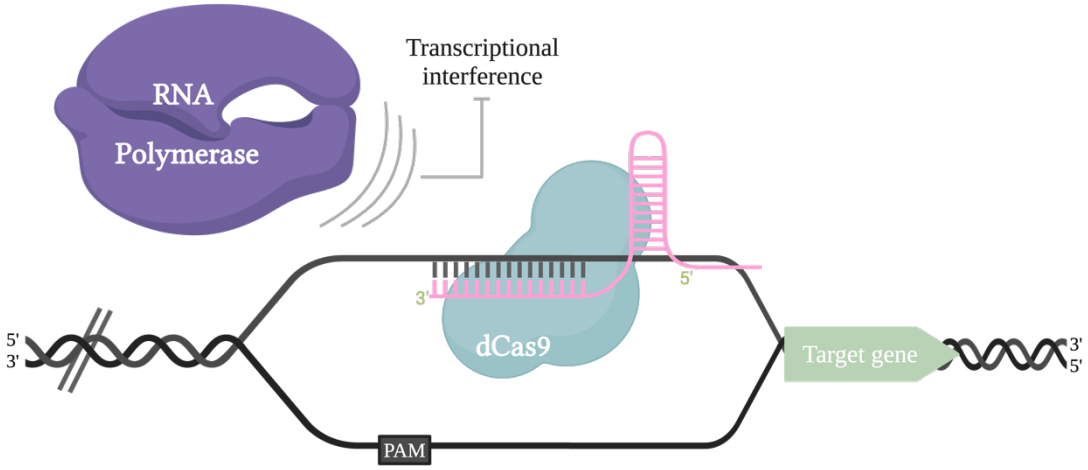


Figure 3. ***Overview of CRISPR interference.*** Here, a sgRNA containing a 20-nucleotide complexes with a dCas9 to bind to the targeted gene and sterically block the RNA polymerase, thus transiently inhibiting transcription.

under an inducible promoter, the timing and strength of gene silencing can be precisely controlled (Larson et al. 2013; Qi et al. 2013; Hogan et al. 2019; Backes and Phillips 2021). The high degree of control makes CRISPRi especially valuable for studying essential genes, where complete knockouts would be lethal, and for dissecting the MoA of novel drugs (Backes and Phillips 2021; Bendixen et al. 2023; Cai et al. 2023).

To simplify employment, the crRNA and tracrRNA are often fused into a single guide RNA (sgRNA), which maintains targeting specificity while streamlining experimental design. The tracrRNA sequence – determined by the dCas9 variant used – is typically fixed in the expression vector, while the customizable crRNA segment can be cloned using various methods (Backes and Phillips 2021).

CRISPRi can be used in both arrayed and pooled screening formats. In array assays, individual mutants are isolated and studied under controlled conditions, allowing to measure effects like fitness changes with direct methods such as optical density tracking. In contrast, pooled assays combine multiple mutants in a single culture, introducing competitions and interaction effects. These more dynamic environments require barcoding or sequencing techniques to resolve individual mutant responses and can uncover complex biological relationships that arrayed methods may miss. (Gislason et al. 2017; Bock et al. 2022; Rahman et al. 2024).

1.5- Rationale

The rapid increase in multidrug-resistant *Pseudomonas aeruginosa* infections – particularly those resistant to carbapenems, a last-resort class of antibiotics (*WHO Bacterial Priority Pathogens List 2024* 2024) – has led to an urgent need to find novel antimicrobials and innovative methods to accelerate early-stage drug development (Miethke et al. 2021). However, conventional discovery methods continue to fall short in keeping pace with the rapid emergence of resistance (Lewis 2020; Miethke et al. 2021).

In this context, machine learning offers a promising solution, increasing hit rates by 12-fold (Stokes et al. 2020; Rahman et al. 2022) and significantly accelerating molecular screening (Camacho et al. 2018; Miethke et al. 2021; Lluka and Stokes 2023). Additionally, exploiting knockdown mutants of essential genes to study compound hypersusceptibility can also accelerate drug discovery by providing insight into the mechanism of action (MoA) (Pathania et al. 2009; Johnson et al. 2019; Warrier et al. 2022; Rahman et al. 2024).

Since most antibiotics target the products of essential genes (Cardona et al. 2015), the ability to selectively silence these genes in *P. aeruginosa* PA14 would aid to elucidate and validate the MoA of potential candidates. Although CRISPRi – renowned for its tunability and specificity – has been explored in this context (Qu et al. 2019; Xiang et al. 2020; Prasad 2022; Zhou et al. 2022a), its widespread implementation is hindered by system leakiness, the cytotoxicity of plasmid-expressed dCas9, and the need for multiple selection markers.

1.6- Hypothesis

Applying a machine learning model trained on inhibitory activity for compounds against *P. aeruginosa* will increase the hit rates of finding active molecules against this bacterium. A CRISPRi system in *P. aeruginosa* can be used to silence essential genes, facilitating the subsequent characterization and confirmation of the mechanism of action of potential hits.

1.7- Objectives

I- Use a machine learning model to predict inhibitory activity of compounds against *P. aeruginosa* PA14.

II- Create a CRISPRi system in *P. aeruginosa* PA14 to characterize and confirm the mechanism of action of compounds with predicted inhibitory activity.

We addressed the two objectives of this thesis using *in silico* approaches (Objective 1) and "wet-lab" demonstrations (Objective 2) in Chapter II. Briefly, we evaluated the different parameters affecting ML predictions and identified the optimal model-training dataset combination (2.2-1 and 2.2-2). This approach led to the discovery of JN19 and JN25 as potential growth inhibitors of *P. aeruginosa* PA14 (2.2-3). In parallel, we developed a tightly regulated CRISPRi system in *P. aeruginosa* PA14 and demonstrated its ability to selectively silence native gene expression (2.2-4), enabling us to give insight into the mechanism of action of the potential growth inhibitor ZR50 (2.2-5), advancing the search for novel antimicrobials.

CHAPTER II

2. Results

2.1- Applying ML to drug discovery increases hit rates by 8%

To predict new inhibitory compounds against *Pseudomonas* aeruginosa, we repurposed our previously published D-MPNN model which was able to predict PHAR(A) as a growth inhibitor against *Burkholderia cenocepacia* (Rahman et al. 2022). We reasoned that we could reproduce previous results by training this same model with compound inhibitory activity against *P. aeruginosa* **(Figure 4A)**. To ensure that the training dataset was in the same order of magnitude to the high-throughput screen previously used, we extracted our training data from ChEMBL, a literature-based database that compiles medicinal chemistry data, such as their mechanism of action or therapeutic indications (Mendez et al. 2019); before its utilization this dataset was curated to contain only information about our pathogen of interest and to remove any redundant entries, where multiple entries with contradicting labels were given to the same compound (**Figure 4B, Table S1**).

The training dataset, referred to as Training_A, was split into training (80%), testing (10%), and validation (10%) subsets. Next, to explore whether higher molecule feature extraction could enhance the accuracy of the predictions, we supplemented the model with different feature-processing modules **(Table 1).** From the four different combinations tested, Model 4 – which incorporates a transformer module that processes all molecules simultaneously to learn structural representations – achieved the highest area under the receiver operating characteristic (AUROC = 0.906), area under the precision-recall curve (AUPRC = 0.874), and F1 score (F1 = 0.830) **(Figure 4C-D**). To further assess Model 4's performance, we created a confusion matrix using the validation set and demonstrated that the model correctly predicted 78% of the labels (**Figure 4E**).

Table 1. Different models were trained with various configurations to extract molecular-level features.

Model	Description	AUROC	AUPRC	F1 Score
Model 1	DMPNN + FFNN	0.900	0.841	0.827
Model 2	DMPNN + RDKit + FFNN	0.895	0.828	0.820
Model 3	DMPNN + RDKit + Transformers + FFNN	0.895	0.828	0.820
Model 4	DMPNN + Transformers + FFNN	**0.906**	**0.874**	**0.830**

[1]Bold numbers indicate the best scores
[2]All scores reported are after hyperparameter optimization.

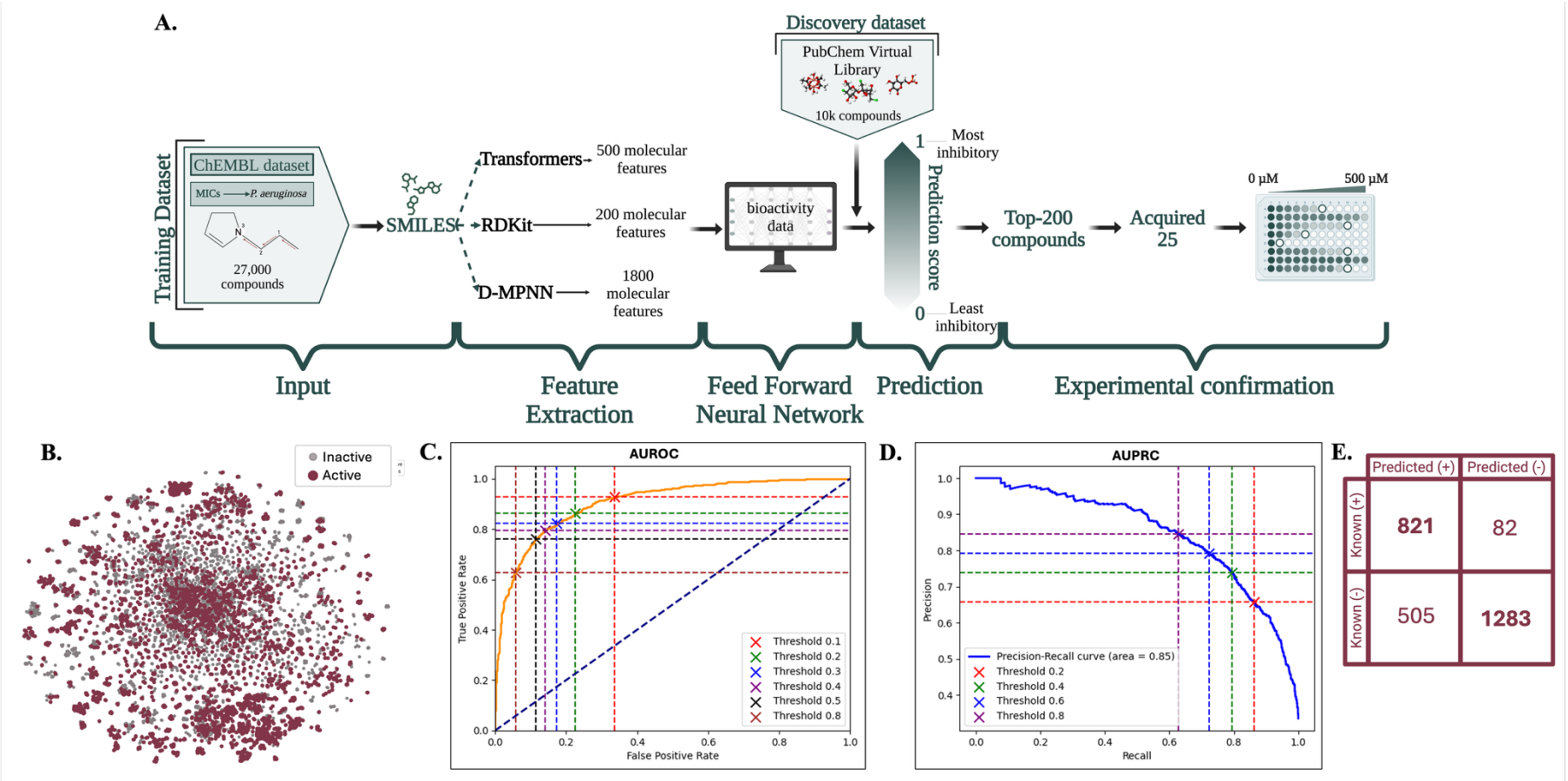


Figure 4. ***Machine learning driven predictions of active compounds***. *(A)* Overview of machine learning applied to drug discovery. *(B)* t-SNE analysis of the training dataset. The plot shows the chemical similarity or dissimilarity of all the active (red) and inactive (grey) compounds in Training_A *(C-D)* AUROC and AUPRC plots evaluating the performance of Model 4 after being trained; the curves show the effect of the different selected thresholds on the maximization of the plotted score. Model 4 had an AUROC score of 0.906 and an AUPRC score of 0.874. *(E)* Confusion matrix showing the binarization results of the validation set with prediction made with Model 4 after selection of the best threshold. 78.19% of the compounds had their label predicted correctly.

Once we had confirmed the performance of Model 4, we proceeded to apply this model to identify growth inhibitors in a previously unseen discovery set, called Discovery set 1 **(Table S2)**. This library consists of 11,000 compounds selected from the PubChem library to cover a bast chemical space without including compounds sharing scaffolds with known antibiotics nor compounds already included in the Training_A dataset **(Figure 5A)**. Here, the model generated scores between 0 and 1 for each molecule, with 1 indicating the highest probability for growth inhibitory activity. We used these scores to rank the compounds in the discovery set and we selected the top 200 predicted compounds. Next, we acquired 25 of these compounds based on availability and experimentally validated their activity against *P. aeruginosa* PA14 **(Figure 5B, Table S3).** Two compounds, called herein JN19 and JN25 (**Figure 5C**) inhibited the growth of *P. aeruginosa* PA14 by 40% at 128 μg/mL (equivalent to 371 μM and 382 μM, respectively)**.** To assess broader-spectrum potential, JN19 and JN25 were also tested against five ESKAPE pathogens – *Acinetobacter baumannii, Enterobacter cloacae,* Methicillin-resistant *Staphylococcus aureus, Klebsiella pneumoniae,* and *Pseudomonas aeruginosa* - and *Escherichia coli* **(Figure 5D)**. Moreover, we also included other *P. aeruginosa* strains, showing that JN25 inhibited the growth of the MDR *P. aeruginosa* strains LES400 and LESlike5 by 65% at 250 μM (83.8 μg/mL) (**Figure 5E).**

Applying a ML model trained with inhibitory activity against *P. aeruginosa*, we were able to identify two compounds with inhibitory activity against this pathogen. However, we achieved a hit-rate of 8%, which is lower than previous results using similar approaches (Stokes et al. 2020;

Rahman et al. 2022), showcasing the need to better understand the parameters affecting ML model predictions.

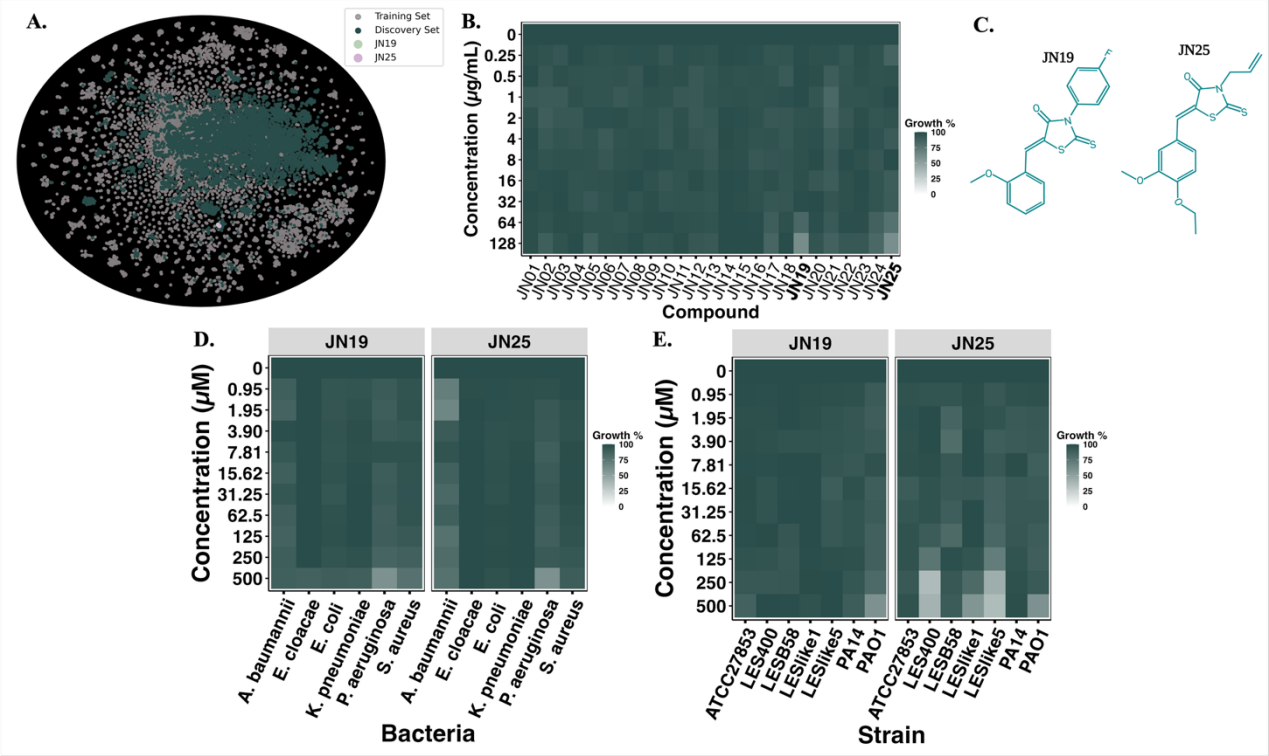


Figure 5. ***Applying a ML model to drug discovery increases the hit rate by 8%.*** *(A)* t-SNE plot showing the chemical similarity or dissimilarity of all the compounds in the training and discovery datasets; the discovery set (green) and training set (grey) occupy the same chemical space, compounds JN19 and JN25, which demonstrated activity against *P. aeruginosa* are denoted with light green and pink, respectively. *(B)* Inhibitory activity of the acquired compounds, JN19 and JN25, bolded, inhibit the growth of *P. aeruginosa* PA14 by 40%. *(C)* Chemical structure of the tested compounds JN19 and JN25. *(D)* Inhibitory activity of JN19 and JN25 against five ESKAPE pathogens and *E. coli*. *(E)* Inhibitory activity of JN19 and JN25 against different strains of *P. aeruginosa. (B), (D),* and *(E)* show the average of three independent biological replicates.

2.2- Imbalanced training datasets decrease the prediction performance of compound activity

To better understand the key components influencing ML-driven predictions, we investigated how the origin of training datasets and the choice of model architecture impact a model's performance. Training datasets are typically generated through HTS or compiled from several experimental assays published in the literature (literature-based datasets); to evaluate the effect of dataset origin, we selected five training sets derived from these distinct sources. Three datasets were obtained from HTS against multiple bacteria: the multidrug-resistant (MDR) Gram-negative *B. cenocepacia* K56-2 (Selin et al. 2013); the Gram-negative model *E. coli* BW25113 (Wong et al. 2022); and the also MDR Gram-positive *St. aureus* RN4220 (Wong et al. 2023b). The remaining two were derived from public repositories: the Community for Open Antimicrobial Drug Discovery (CO-ADD), which curates compound bioactivity against a broad panel of pathogens (Desselle et al. 2017), and the ChEMBL library; because these two datasets were not a direct HTS, they were downloaded and curated to only contain values related to our pathogens of interest. Redundant entries caused by stereoisomeric representations of the same compound (SMILES) with contradictory labels can reduce model accuracy; to address this issue we curated each dataset following a rigorous pipeline (See Methods and Material section) ensuring that each SMILE was only related to a single label in the training datasets. Using this cleanup approach, we filtered out between 3% and 30% of the compounds from the different datasets, producing five training datasets with various sizes and balances (**Table 2, Table S1**, and **Table S4**).

We used t-distributed stochastic neighbour embedding (t-SNE) to cluster chemical molecules based on similarity by extracting molecular features and calculating the similarity of their molecular scaffolds and are regularly represented as two dimensional distributions (van der Maaten

and Hinton 2008). We used this visualization tool to evaluate the chemical space encompassed and ensure that no bias towards a specific class of antibiotic was observed in the training sets after their cleaning was completed **(Figure 6A**). We were able confirm that both active and inactive compounds were well distributed throughout the chemical space and did not cluster together (**Figure S1**), indicating scaffold diversity in the datasets. Training_5 exhibited the most even distribution, with its active compounds occupying most of the dataset's chemical space.

Table 2. Original training datasets that were used in this study.

Training dataset	Source	Type	Biological assay	Target	Before cleanup			After cleanup		
					Total compounds	% active	% inactive	Total compounds	% active	% inactive
Training_1	Selin et al. 2013	High-throughput screening	Residual growth at 50 µM (RG ≤ 0.7)	*Burkholderia cenocepacia* K56-2	29,568	0.75	99.25	28,777	0.71	99.29
Training_2	Wong et al. 2023	High-throughput screening	Residual growth at 50 µM (RG ≤ 0.2)	*Staphylococcus aureus* RN4220	39,312	1.30	98.70	13,067	2.43	97.57
Training_3	Wong et al. 2022	High-throughput screening	Residual growth at 50 µM (RG ≤ 0.2)	*Escherichia coli* BW25113	39,128	0.56	99.44	13,139	1.29	98.71
Training_4	CO-ADD	High-throughput screening	Minimal inhibitory concentration (MIC ≤ 16 µg/mL)	*Pseudomonas aeruginosa* ATCC27853	4,112	2.09	97.91	3,336	1.35	98.65
Training_5	ChEMBL	Literature	Minimal inhibitory concentration (MIC ≤ 16 µg/mL)	*Pseudomonas aeruginosa*	20,018	22.04	77.96	18,403	20.45	79.55

[1]Before cleanup denotes the original size and balance of the training datasets, as published; after cleanup shows the size and balance of the different training datasets after the different sanitization steps were conducted.

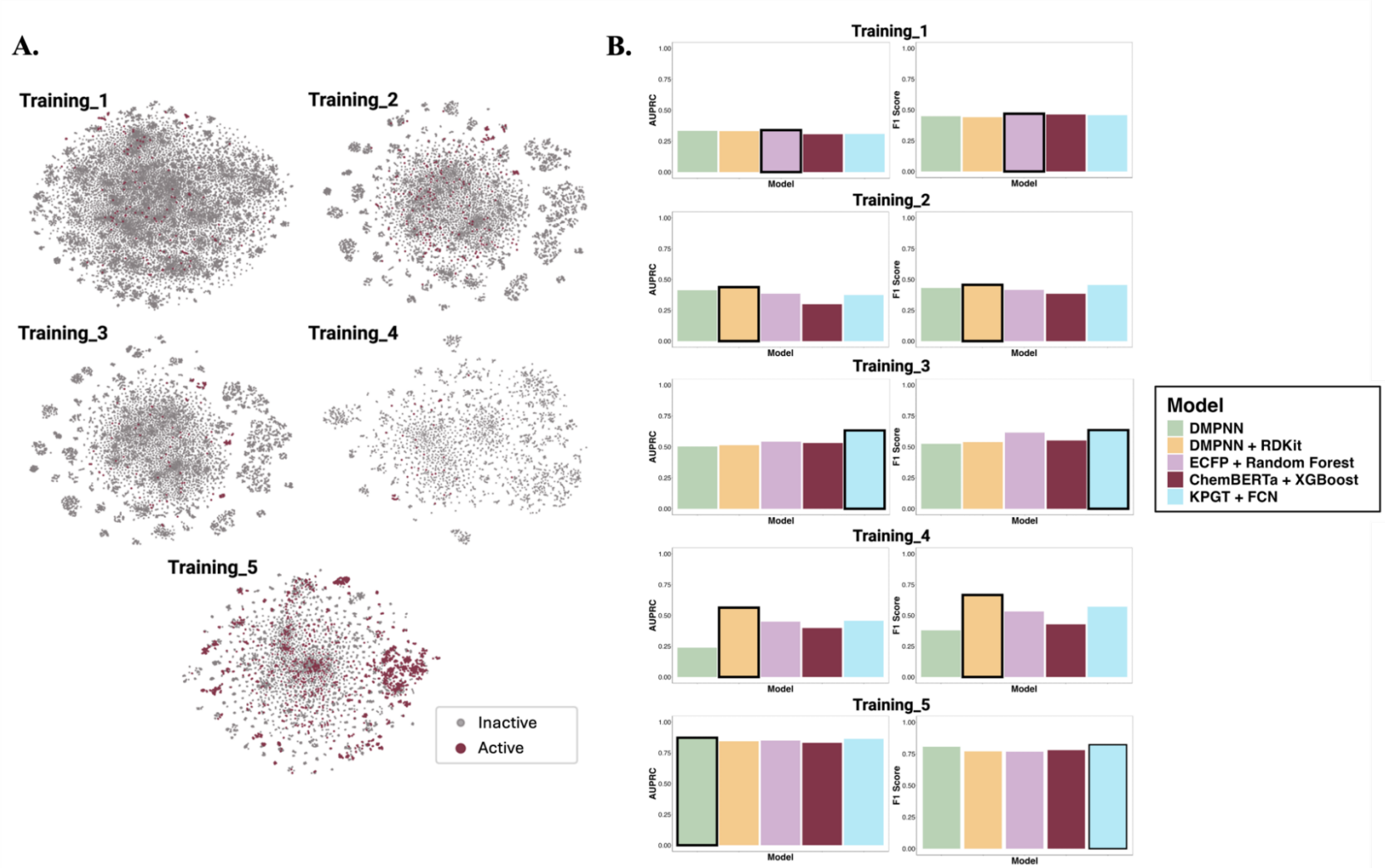


Figure 6. ***Visualization of the chemical space and model training.*** *(A)* Each t-SNE plot shows the chemical similarity or dissimilarity of all the compounds in each training dataset, with the active compounds shown in red and the inactive compounds shown in grey. *(B)* All five models were trained with the different training sets and the AUPRC and F1 scores were calculated for each combination, the best combination (highlighted by the black box) was used to run the predictions.

After finishing the assessment of the training datasets, we evaluated different models in combination with the various training datasets to observe the impact of model architecture on prediction performance. Five distinct ML models with different architectures and supplemented with different molecular descriptors, were independently trained on each of the five datasets. Following the same 80-10-10 data split as described in section 2.1, we used the validation subset to assess performance using the AUPRC and F1 score (Leisman et al. 2020; Hicks et al. 2022); in this occasion, AUROC scores were disregarded due to its unsuitability for highly imbalanced datasets (Leisman et al. 2020). Our results showed that Trianing_5 consistently produced the highest scores across all models, with AUPRC and F1 values exceeding 0.7 (**Figure 6B)**.

In conclusion, we found that models trained on moderately imbalanced datasets – such as Training_5 – outperformed those trained on highly imbalanced datasets, highlighting the importance of data composition in ML-guided antimicrobial discovery.

2.3- Benchmarking predictive performance using "spike" molecules

To further analyze the predicting capabilities of the models, we selected the best-performing model-training set combinations – based on the AUPRC and F1 scores – and used them to predict compound activity in a new and unseen dataset, referred to as Discovery Set 2 (**Table S5**). This discovery dataset contained a subset of one million compounds from the PubChem library. As with Discovery set 1, the compounds were specifically chosen to uniformly represent the chemical space, differ structurally from known antibiotics, and ensure no overlap exists with the training datasets.

To evaluate the prediction capabilities of ML models in a real-world scenario, Discovery Set 2 was supplemented with the chemical structures of 47 "spike" molecules, which are compounds

with previously experimentally validated activity labels against the organisms in each specific training set (**Table S6)**. Based on the "spike" molecules' predicted score, we calculated an optimal threshold to distinguish between "active" and "inactive" predicted compounds using the F1 score (Michelucci 2024), which was then used to generate confusion matrices and evaluate model classification (**Figure 7A**). The resulting F1 scores ranged from 0.4 and 0.5; given the pronounced class imbalance and the relatively small proportion of spike molecules, these values indicate a reasonable trade-off between precision and recall for identifying active compounds, a key metric for imbalanced datasets. Although Training_5 was the most balanced dataset, its associated model yielded a high number of false positives – i.e., inactive compounds predicted as active – a pattern also observed in models trained with more imbalanced datasets. This suggests that Training_5 still exhibited a level of imbalance greater than ideal for these prediction tasks. Models trained with Training_4 and Trianing_1 correctly predicted most active "spike" molecules but had false positive rates around 50%. The model trained with Training_2 showed a clear bias towards predicting compounds as active, resulting in a false positive rate of ~93%, whereas Training_3 showed the opposite trend, incorrectly classifying 86% of the active spike molecules as inactive **(Figure 7A)**.

Further analysis of the binarized predictions revealed that certain models – particularly the one trained with Training_5 – classified an unexpectedly large number of compounds as active, with approximately 800,000 compounds flagged using the calculated threshold (**Figure 7B**). To obtain a number of active compounds that could be experimentally tested, we applied a secondary binarization step based on the hit rate of traditional drug discovery methods (e.g., HTS), selecting the top 1% (10,000) predicted compounds as active. A Venn diagram of this narrowed set revealed that 139 compounds were predicted as active across all model outputs (**Figure 7C**).

Interestingly, the model trained on Training_5 – which is derived from Training_A – was able to re-discover JN25 as a growth inhibitor for *P. aeruginosa* in the top 10% active compounds (100,000 compounds). Moreover, these findings demonstrate that the inclusion of "spike" molecules with known activity can serve as an effective internal benchmark to assess the predictive capabilities of ML models.

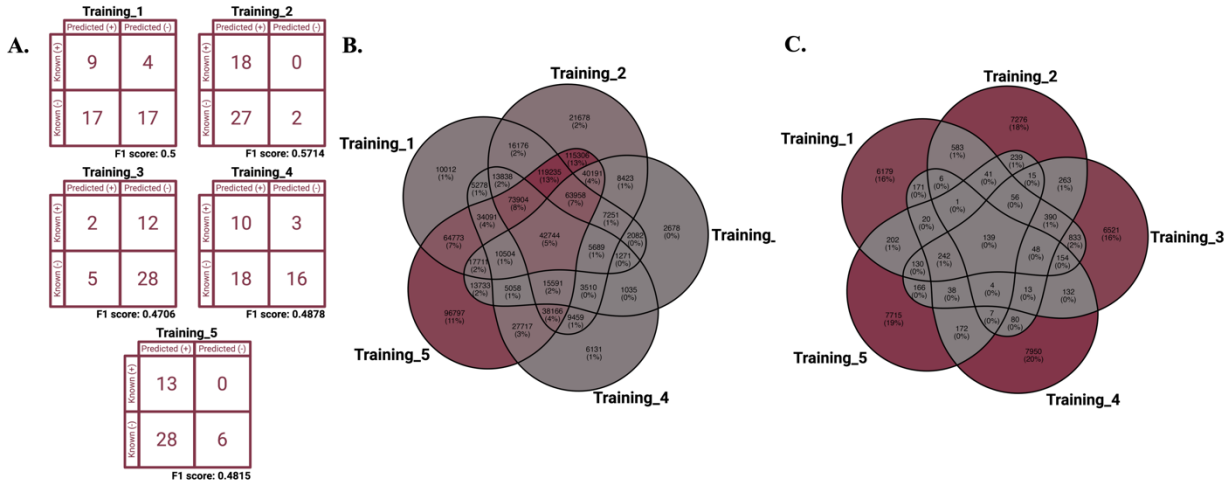


Figure 7. ***"Spike" molecules are an internal benchmark to assess model performance.*** *(A)* Confusion matrices of the "spike" molecules included in Discovery Set 2. These compounds were also used to calculate the F1 score for each prediction and establish the best threshold to binarize the data. *(B)* Venn diagram showing the predicted active compounds by all 5 models when the binarization of the predictions was done using the threshold calculated with the "spike" molecules. *(C)* Venn diagram showing the predicted active compounds by all 5 models when only the top 1% of compounds were considered active.

2.4- CRISPRi in P. aeruginosa PA14 selectively silence native gene expression

Another major bottleneck in drug discovery is the limited understanding of the mechanism of action (MoA) of new molecules, hindering compound prioritization. Most antibiotics target the products of essential genes, so developing a method to downregulate gene expression can provide valuable insight into the likely targets of potential antimicrobial candidates. Mutants with reduced

expression of these essential genes will exhibit hypersusceptibility to antimicrobials that act on their gene products, thereby facilitating target identification (Cardona et al. 2015). To investigate this, we developed a CRISPRi system in *P. aeruginosa* PA14 for the selective silencing of native gene expression. The commonly used dCas9 originates from *Streptococcus pyogenes*, a low GC-content bacterium (~40%) (Qi et al. 2013) which can pose problems when expressed in high-GC organism (>50%) such as *P. aeruginosa* (Hogan et al. 2019; Grace et al. 2022), as codon utilization may vary. To overcome this, we employed high-GC codon-optimized $dCas9_{Spy}$ previously validated in *B. cenocepacia* (Hogan et al. 2019). From previous studies we knew that a single copy of this dCas9 variant generated enough silencing effect in Gram-negative bacteria, including *P. aeruginosa* (Tan et al. 2018; Peters et al. 2019; Hogan et al. 2019); we therefore inserted as a single chromosomal copy of this gene into the genome of *P. aeruginosa* PA14 under the control of a rhamnose-inducible promoter using the mini-CTX system (Hoang et al. 2000) (**Figure 8A)**. Whole-genome sequencing confirmed successful insertion, and immunoblotting showed a tightly regulated and titratable dCas9 expression at various levels of rhamnose (**Figure 8B)**.

To allow CRISPRi-mediated gene silencing, we engineered a sgRNA-expressing plasmid using pSCB2-gRNA as the backbone (Hogan et al. 2019) and replacing the trimethoprim resistance cassette with carbenicillin, to allow for its selection in *P. aeruginosa*; we also added Golden Gate sites between the sgRNA promoter and sgRNA scaffold for modular sgRNAs cloning (**Figure 8C**), streamlining the exchange of sgRNA and putting on par to other available methods (Tan et al. 2018; Peters et al. 2019; Noirot-Gros et al. 2019; Zhang et al. 2024).

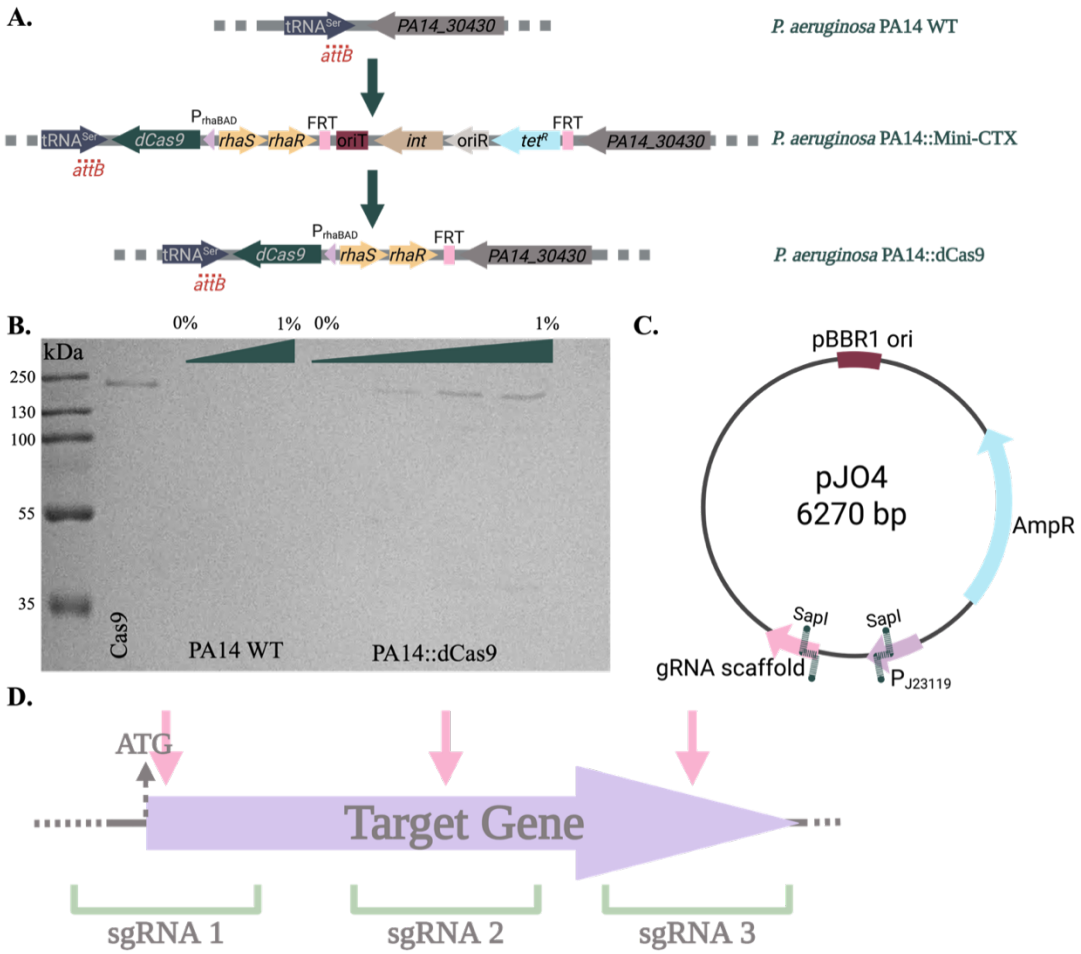


***Figure 8. Creation of a CRISPRi system in P. aeruginosa* PA14.** *(A)* Workflow for the insertion of a single copy of a codon-optimized dCas9 into the chromosome of *P. aeruginosa* PA14 under the control of a rhamnose-inducible promoter using the Mini-CTX system. *(B)* dCas9 was induced from this chromosomal insertion with different concentrations of rhamnose and detected by immunoblotting. The lane labeled Cas9 was loaded with 20 ng of purified Cas9. The rhamnose concentrations tested were 0%, 0.05 %, 0.5%, and 1%. *(C)* The sgRNA-expressing plasmid pJO4 was adapted from pSCB2-sgRNA by exchanging the resistance cassette and adding Golden Gate-compatible sites for the insertion of the sgRNAs. *(D)* Position of the sgRNAs designed for all the targeted genes. sgRNA 1 was designed to target adjacent regions to the start codon, sgRNA 2 targeted the middle of the coding region, and sgRNA 3 was designed to target regions (or sites) adjacent to the stop codon.

As a proof-of-principle of the system, we investigated the compound ZR50 **(Figure 9A)**. This compound was initially discovered through a chemical-genetic interaction profile (CGIP) on *B. cenocepacia* K56-2 (Cardona Lab, unpublished results), where a CRISPRi-pooled library of essential genes in *B. cenocepacia* K56-2 was exposed to 5,000 compounds specifically selected to be chemically different from known antibiotic; the depletion of each mutant was assessed and compared to the untreated control in an effort to identify possible target-compound interactions. This CGIP assay revealed that *fda* and *groES* CRISPRi mutants were significantly depleted in the presence of ZR50 compared to the DMSO control, suggesting their gene products as potential targets. In this assay, a CRISPRi-pooled library of essential genes in The *fda* gene encodes the fructose-1,6-bisphosphate aldolase (FBA), a key enzyme in carbohydrate metabolisms that catalyzes the reversible cleavage of fructose-1,6-bisphosphate (FBP) into two triose sugars (Cooper et al. 1996). In contrast, *groES* encodes the co-chaperonin component of the GroEL/GroES complex, which is essential for protein folding (Xu and Sigler 1998) **(Table 3)**.

After confirming the inhibitory activity of ZR50 against *B. cenocepacia* K56-2 (data not shown), we investigated its effect against the ESKAPE pathogens. MIC assays against five of these pathogens and *E. coli*, observed a 30% inhibition in *P. aeruginosa* PAO1 growth (**Figure 9B**); moreover, further growth inhibition testing revealed that ZR50 inhibits multiple *P. aeruginosa* strains – including PA14, PAO1, ATCC27853, and LESB58 – at 500 μM (246.3 μg/mL) **(Figure 9C).** Once the activity of ZR50 was confirmed, we explored whether the putative targets of ZR50 in *B. cenocepacia* K56-2 were also present in *P. aeruginosa* PA14. A bi-directional nucleotide BLAST analysis confirmed their presence in *P. aeruginosa* PA14, with an identity of 77.42% (E-value 2e-143) and 58.95% (E-value 4e-33) for *fda* and *groES,* respectively. Moreover, other

bioinformatic approaches (e.g., checking genes 1000 bp upstream and downstream of studied gene) were used to confirm the identity of our targets.

Next, to evaluate the efficacy of CRISPRi-mediated gene silencing in *P. aeruginosa* PA14, we decided to target these homologous genes. We designed sgRNAs against *fda* and *groES* **,** and additional sgRNAs targeting the gene *rpoB* – whose product is the known target of rifampicin and was not downregulated by ZR50 (Campbell et al. 2001; Turner et al. 2015) – as well as a non-targeting sgRNA (NTC), were also generated and used as the negative and non-targeting control, respectively. Additionally, to evaluate whether targeting the initiation or elongation steps of transcription was more efficient, we designed between one and three sgRNAs per gene using a computational pipeline (van Gestel et al. 2021): sgRNA 1 targeted a region adjacent to the start codon, sgRNA 2 targeted the middle of the translated region, and sgRNA 3's target was placed in the proximity of the stop codon (**Figure 8D, Table 6**). These guides were cloned into the pJO4 plasmid via Golden Gate and conjugated into *P. aeruginosa* PA14::dCas9 using tri-parental mating.

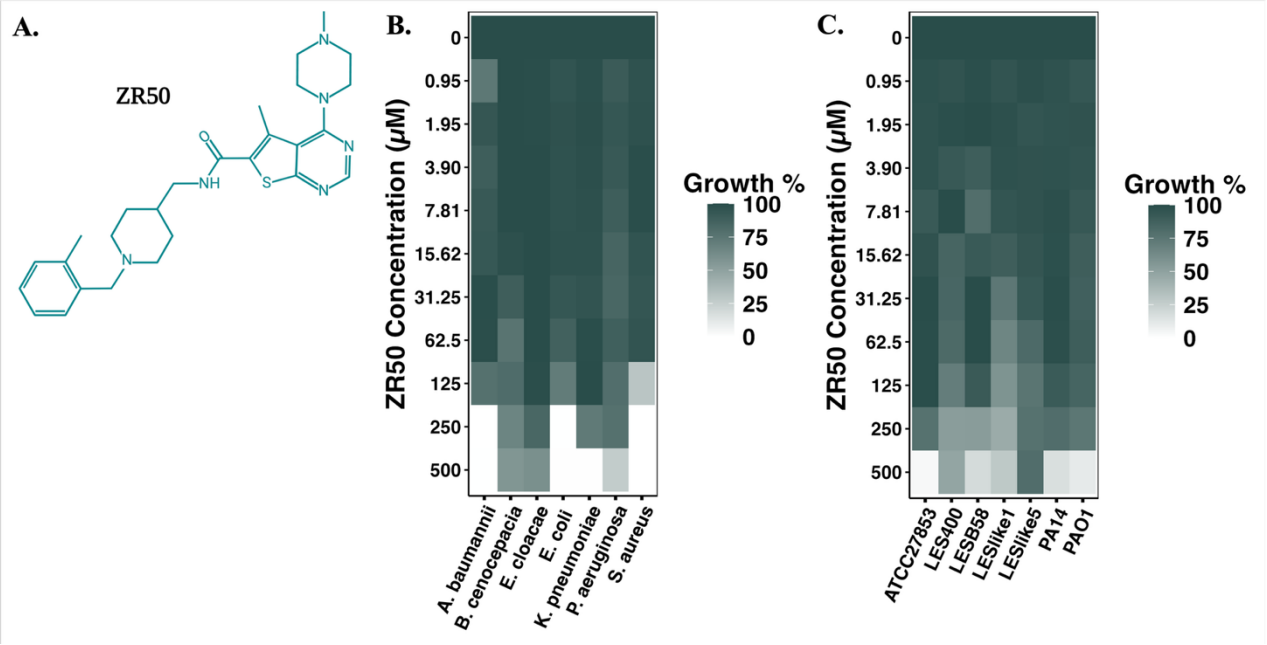


Figure 9. ***ZR50 has inhibitory activity against different strains of P. aeruginosa.*** *(A)* Chemical structure of ZR50, a potential growth inhibitor identified through a CGIP assay in *B. cenocepacia K56-2*. *(B)* ZR50 was tested against the ESKAPE pathogens, showing an MIC for both *E. coli* and *S. aureus*, as well as displaying a moderate inhibitory activity against *P. aeruginosa*. *(C)* The compound ZR50 was tested against different isolates of *P. aeruginosa,* exhibiting a MIC of 500 µM against *P. aeruginosa* PAO1, *P. aeruginosa* PA14, and the clinical isolates *P. aeruginosa*

Table 3. Table of mutants depleted in the CGIP experiment against *B. cenocepacia* K56-2 and their homologous genes in *P. aeruginosa* PA14.

	Mutant depleted in *B. cenocepacia K56-2*	Locus Tag in *B. cenocepacia* K56-2	Gene in *P. aeruginosa* PA14	Locus Tag in *P. aeruginosa* PA14	COG	Product
ZR50	*cbbA*	WQ49_RS02275	*fda*	PA14_07230	G	Fructose-1,6-bisphosphate aldolase
	groES	WQ49_RS18705	*groES*	PA14_57020	O	Molecular chaperone

[1]For each compound, we show the targeted gene in each mutant that showed significant depletion compared to the DMSO control, the homologous gene in *P. aeruginosa* PA1, and their COG categories (G, carbohydrate transport and metabolism; O: post-translational modification, protein turnover, and chaperones).

To assess dCas9-associated toxicity, we studied its growth by titrating rhamnose and confirmed that at concentrations up to 1% rhamnose there was no growth defect in *P. aeruginosa* PA14::dCas9 harbouring the non-targeting control **(Figure 10A)**. Furthermore, to investigate if a growth defect was observed in the absence of inducer, we measure the growth of the CRISPRi mutants over time without inducer and observed no growth defect, suggesting the tight regulation of the dCas9 in non-inducing conditions **(Figure 10B)**. Phenotypic assays – conducted via growth curve analyses – confirmed that PA14::dCas9-fda_sgRNA1, PA14::dCas9-groES_sgRNA, and PA14::dCas9-rpoB_sgRNA1, which targeted sequences near the start codon, consistently produced the strongest silencing phenotypes **(Figures 10C-F),** corroborating previous findings

(Hogan et al. 2019; Rahman et al. 2024). Due to the short length of *groES* (294 bp), only one sgRNA close to the start codon was designed.

Together, these results demonstrate that a chromosomically integrated, rhamnose-inducible CRISPRi system allows for the observation of conditional growth phenotypes in a concentrations-dependant manner in *P. aeruginosa* PA14, providing a valuable tool for hypersusceptiblity studies.

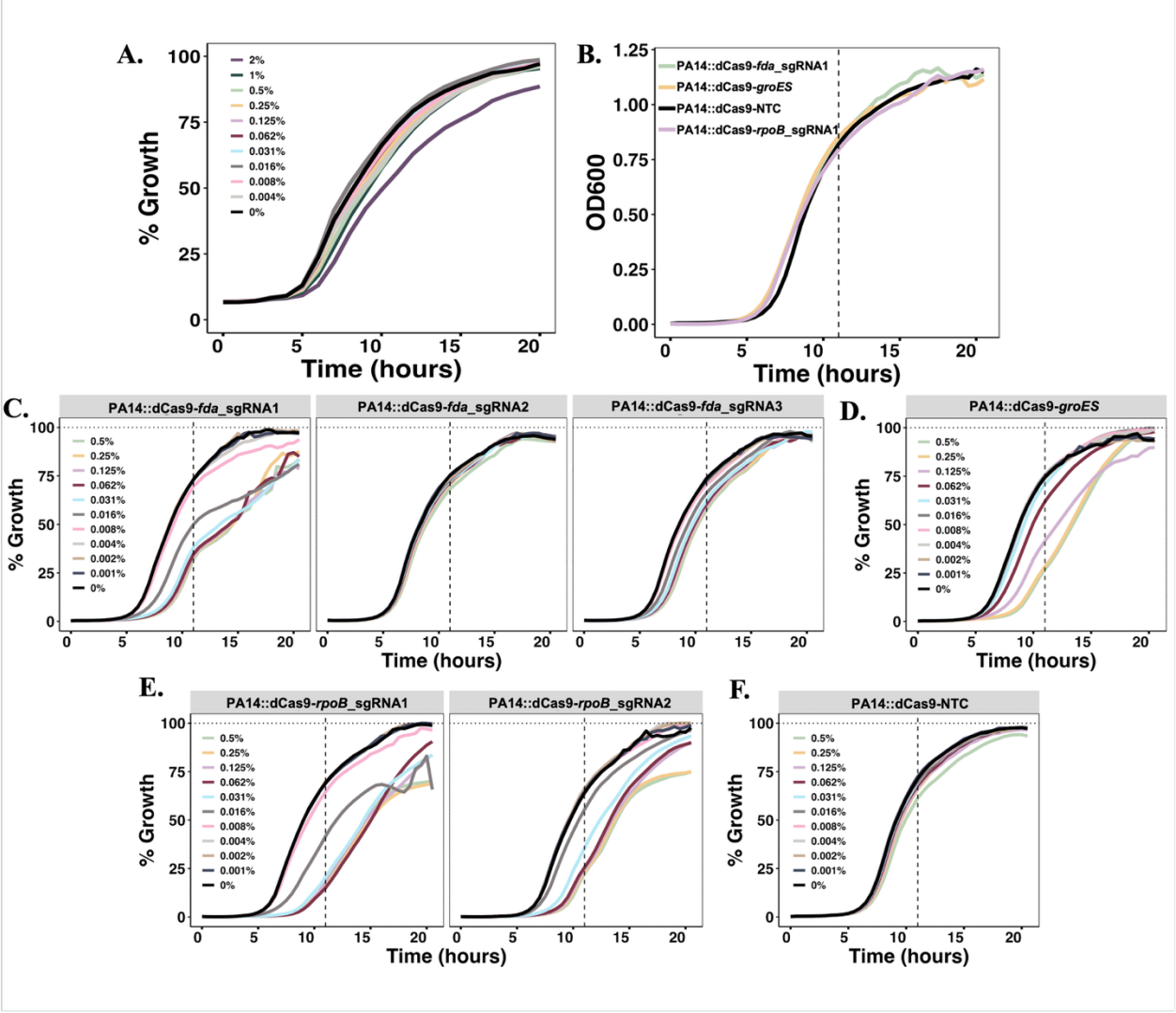


Figure 10. ***sgRNAs targeting the initiation of transcription show the strongest conditional growth phenotypes.*** *(A)* Growth curves of the non-targeting control at different concentrations of rhamnose. *(B)* Growth curves of PA14::dCas9-mutant_sgRNA1 for all mutants in the absence of rhamnose. Growth curves for PA14::dCas9-*fda*_sgRNAs *(C)*, PA14::dCas9-*groES* *(D),* PA14::dCas9-*rpoB*_sgRNAs *(E),* and PA14::dCas9-NTC *(F)*. For all growth curves, overnight cultures of the mutants were OD600 adjusted to 0.001 and grown in a 50% dilution of CAMHB and PBS supplemented with different concentrations of rhamnose ranging from 2% to 0% in a plate spectrophotometer. Readings were taken at 30-minute intervals over a 20-hour period. Results are the mean of four independent biological replicates. The dotted horizontal line shows 100% growth, while the dashed vertical line denotes the middle of the exponential phase at 11 hours.

2.5- CRISPRi silencing of fda confers hypersusceptibility to ZR50

The previously mentioned CGIP in *B. cenocepacia* K56-2 demonstrated that study, ZR50 depleted the mutants targeting *fda* and *groES.* To confirm that our CRISPRi system could downregulate the expression these two genes conferring hypersusceptiblity to ZR50, mutants and non-targeting controls were grown at rhamnose concentrations that inhibit 50% growth **(Figure 11A, Table S7)** in a 50% dilution of CAMHB-PBS. We were not able to observe hypersusceptibility to ZR50 for any of the mutants tested **(Figure 11B)**.

fda encodes for the fructose-1,6-bisphosphate aldolase, an essential enzyme in rich media involved in the glycolysis cycle (Turner et al. 2015) **(Figure 11C)**; previous findings have demonstrated that *fda* mutants are unable to grow in minimal media with glucose as the sole carbon source (Puckett et al. 2014). To investigate if the lack of hypersusceptiblity seen in rich media might relate to the action of ZR50 in the glycolysis pathway – which might be escaped by other intracellular pathways in rich media, as there are multiple carbon sources available – we decided to test the conditional phenotype of the mutants as described before but in M9 minimal media with glucose as the sole carbon source. We were able to demonstrate that the *fda*-silenced strain exhibited hypersusceptibility to ZR50 compared to *groES* and the controls (**Figure 11D**).

These results confirm that our CRISPRi system can downregulate essential genes to detect hypersusceptibility to compounds, and that ZR50 might be involved in disrupting the glycolysis pathway.

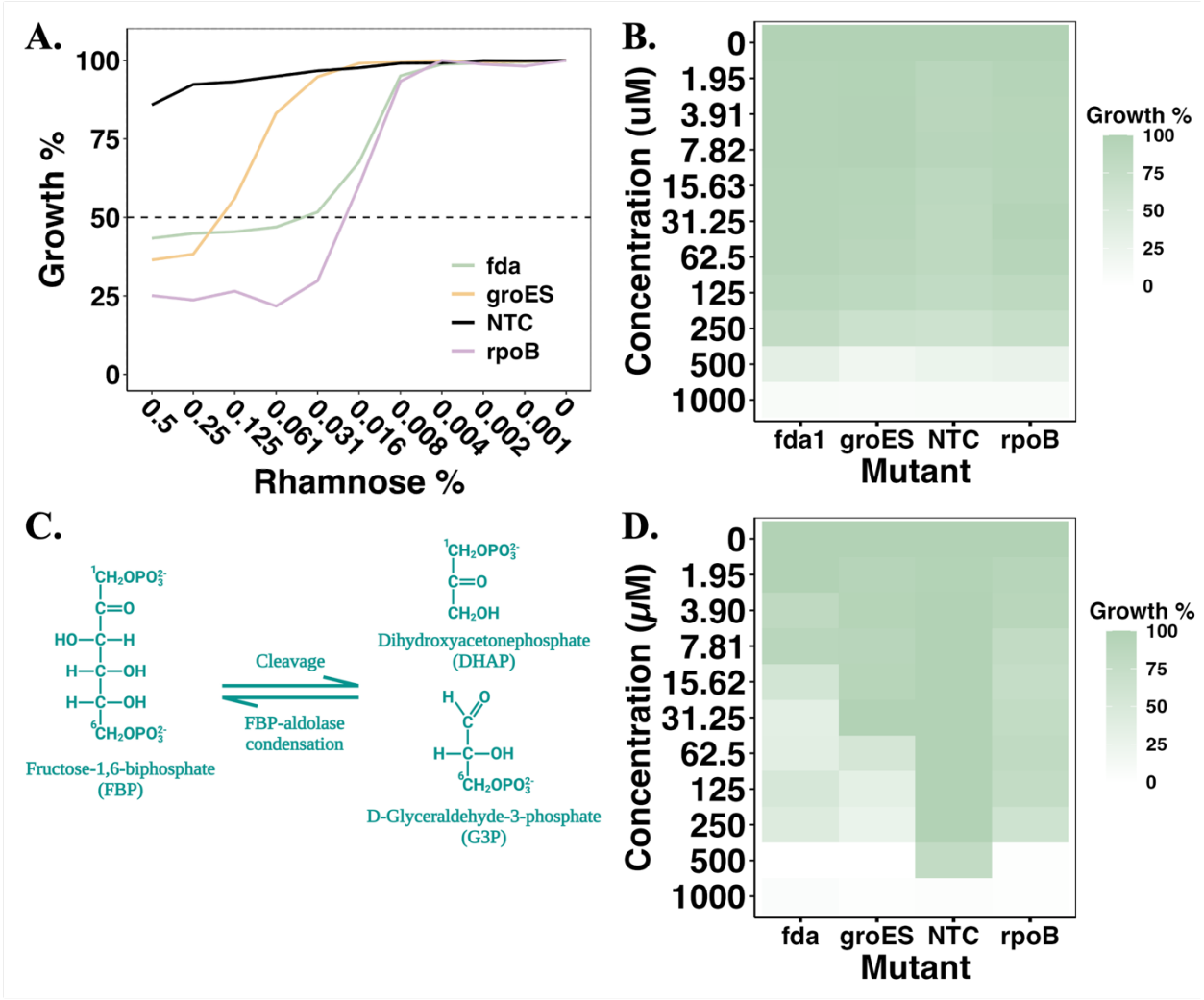


Figure 11. ***CRISPRi in P. aeruginosa*** **PA14** ***effectively silences native gene expression and confers hypersusceptibility to ZR50.*** *(A)* Rhamnose dose-response curves for the tested mutants. Dashed line shows 50% growth. Results shown are the average of three independent biological replicates done in a 50% dilution of CAMHB and PBS. *(B)* Hypersusceptiblity assay of CRISPRi knockdown mutants to ZR50 in a 50% dilution of CAMHB and PBS. Green indicates more growth (less inhibition) and white indicates less growth (more inhibition). Results are the average of two independent biological replicates. *(C)* The fructose-1,6-bisphosphate aldolase catalyzes the reversible cleavage FBA into two triose sugars; the components are shown in the Fischer depiction. *(D)* Hypersusceptiblity assay of CRISPRi knockdown mutants to ZR50 in M9 minimal media with glucose as the sole carbon source. Green indicates more growth (less inhibition) and white indicates less growth (more inhibition). Results are the average of two independent biological replicates.

2.6- Hit compounds have pharmacokinetic properties comparable to other known antibiotics

When a drug is discovered, concerns about their undesirable pharmacokinetic properties often arise, as these are a major factor contributing to drug disapproval (Speck-Planche and Cordeiro 2014). To evaluate whether the pharmacokinetic profiles of the hit compounds JN19, JN25, and ZR50 align with those of established antibiotics, we conducted an *in silico* assessment of their absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties (Ferreira and Andricopulo 2019). Predictions were performed using two publicly available machine learning platforms: ProTox 3.0 (Banerjee et al. 2024) and ADMET-AI (Swanson et al. 2024), which together had been trained with close to 100 different datasets.

We compared the ADMET profiles of the three candidates with those of antibiotics commonly used to treat *P. aeruginosa* infections including cefepime, ciprofloxacin, gentamicin, levofloxacin, meropenem, tetracycline, and tobramycin. JN19, JN25, and ZR50 were similar to known antibiotics respect to predicted respiratory toxicity **(Figure 12A),** toxicity class **(Figure 12B),** cytotoxicity **(Figure 12C),** hepatotoxicity **(Figure 12D),** clinical toxicity **(Figure 12E),** and neurotoxicity **(Figure 12F).** However, they had a different blood-brain barrier penetration score than the control antibiotics **(Figure 12G)**. Moreover, ZR50 presented a higher predicted hERG blocking than all other tested compounds **(Figure 12H)**.

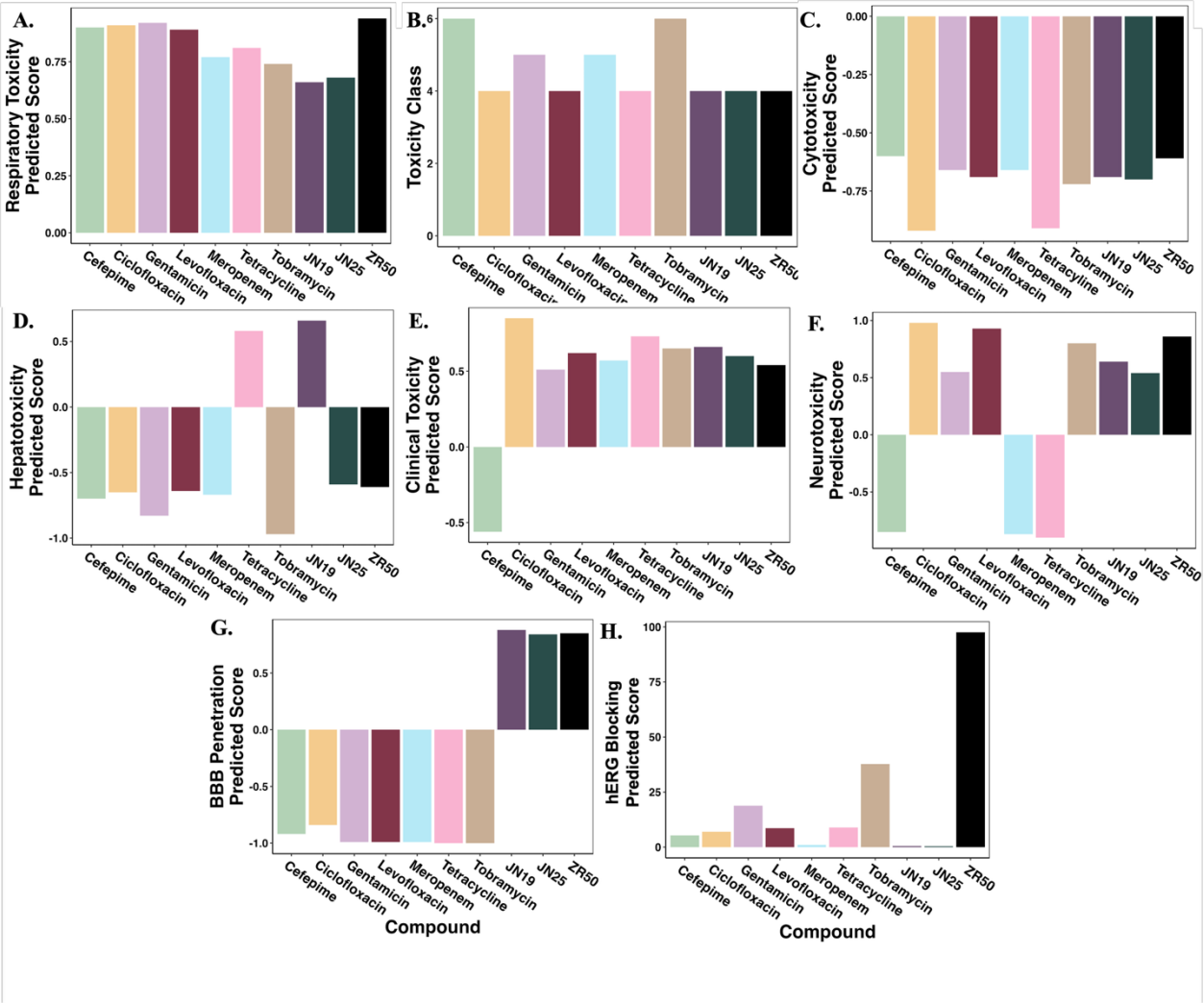


Figure 12. ***Hit compounds have comparable ADMET properties to other common antibiotics.*** ProTox 3.0 and ADMET-AI were used to predict the scores for *(A)* blood-brain barrier penetration, *(B)* clinical toxicity, *(C)* cytotoxicity, *(D)* hepatotoxicity, *(E)* hERG blocking, *(F)* neurotoxicity, *(G)* respiratory toxicity, and *(H)* toxicity class of JN19 (dark purple), JN25 (dark green), and ZR50 (black) and other known antibiotics.

CHAPTER III

3. Discussion and Future Directions

3.1- Discussion

The increase in antibiotic-resistant infections, particularly those caused by multidrug-resistant *Pseudomonas aeruginosa,* underscores the urgent need for new antimicrobial strategies (Lewis 2020; World Health Organization 2024). Despite this growing need, drug discovery remains hindered by bottlenecks such as low hit rates in early stages of discovery and the challenge of compound prioritization due to limited insights into the mechanisms of action (Miethke et al. 2021). This thesis aimed to accelerate early-stage antimicrobial discovery by developing a machine learning (ML) approach to identify novel compounds and designing a gene silencing platform capable of inferring mechanisms of action early in the discovery pipeline.

To address the first objective, we retrained a previously published deep neural network that had demonstrated a 10-fold increase in the likelihood of identifying active compounds (Rahman et al. 2022). We trained this ML model with inhibitory activity information for more than 20,000 compounds against *P. aeruginosa* and predicted antibacterial activity *in silico* in a subset of compounds from the PubChem library. Excitingly, during the validation step of the training, the ML model was able to predict activity with an accuracy of 78.2%. Nonetheless, experimental confirmation yielded activity in only two of the 25 tested compounds (obtained from the top-200 predicted molecules, JN19 and JN2, which possessed a unique scaffold, including a rhodanine and anisole groups connected by an allyl group. The resulting overall hit rate was 8%, notably lower than prior ML-guided approaches ; we hypothesize that this lower hit rate stems from the nature of the training data: Training_A, although moderately balanced, was solely derived from the literature. Literature-based datasets are known to suffer from annotation inconsistencies and experimental variability, with reported error rates between 30% and 50% (Tiikkainen et al. 2013). As such, the risk of false positive predictions is still high, which we corroborated with our

experimental validations, suggesting that future model performance might be improved through training on rigorously curated experimental datasets.

To further investigate the impact of dataset characteristics on ML performance, we trained five models with distinct architectures using datasets sourced from both high-throughput screenings (HTS) and literature. We also developed a novel evaluation framework by augmenting a discovery set with 47 "spike" molecules of known activity to assess prediction reliability *in silico.* Our results revealed that training data imbalance significantly affected prediction outcomes. Models trained on moderately imbalanced datasets (e.g., Training_5) outperformed those trained on severely imbalanced datasets (Training_1-4), which skewed predictions towards the majority (Azlim Khan and Ahamed Hassain Malim 2023; Narykov et al. 2024). Although various strategies exist to mitigate this imbalance issue, a particularly promising solution is self-supervised learning. This approach allows models to learn generalizable representations from large collections of unlabelled chemical structures, thereby enhancing downstream classification tasks (Gui et al. 2024). Several groups have employed this approach to enhance prediction performance (Wang et al. 2023; Xie et al. 2023). Olayo-Alarcon *et al.* (Olayo-Alarcon et al. 2025) demonstrated that a pre-trained model incorporating chemical structures from PubChem enhanced the predictive capabilities, successfully re-predicting active compounds such as Halicin (Stokes et al. 2020) and Abaucin (Liu et al. 2023). Furthermore, employing such techniques simultaneously broaden the chemical space beyond traditionally curated datasets.

The second objective of this thesis tackled another major barrier in antimicrobial development: elucidating the mechanism of action (MoA) of promising compounds (Miethke et al. 2021). To this end, we developed a robust CRISPRi system for gene repression in *P. aeruginosa* PA14 via a mini-CTX platform. Our system allowed for fine-tuned repression of native genes and

facilitated hypersusceptibility profiling of gene knockdown mutants exposed to candidate inhibitors, demonstrated by the hypersusceptibility of probable targets to the identified inhibitor ZR50 in minimal media with glucose as the sole carbon source. While CRISPRi tools have previously been introduced in *Pseudomonas* species, our system addressed leaky dCas9 expression (Tan et al. 2018; Noirot-Gros et al. 2019; Xiang et al. 2020), enabling precise and tunable gene knockdown regulation. Unlike other platforms, our system integrates the dCas9 into the chromosome of *P. aeruginosa* PA14, eliminating the need for its selection and conserving selection markers – a considerable advantage in multidrug-resistant (Tan et al. 2018; Peters et al. 2019; Noirot-Gros et al. 2019; Xiang et al. 2020; Zhou et al. 2022a; Zhang et al. 2024). We additionally streamlined our system with Golden Gate-compatible cloning sites for sgRNA insertion, improving previous platforms (Hogan et al. 2019; Noirot-Gros et al. 2019), and incorporated flanking restriction sites to facilitate flexible resistance cassette exchange. As mentioned above, using this platform we successfully confirmed that ZR50 targets the glycolysis pathway, as hypersusceptiblity was only observed with glucose as the sole carbon source. This provides compelling evidence that our CRISPRi system can be used for mechanistic studies of antimicrobial candidates targeting essential genes.

Finally, we assessed the pharmacokinetic profiles of compounds JN19, JN25, and ZR50 to evaluate their therapeutic potential. *In silico* ADMET analyses revealed that these compounds exhibit similar pharmacokinetic properties to existing antibiotics used for *P. aeruginosa* treatment. However, suboptimal blood-brain barrier penetration scores and elevated hERG blocking potential were observed in some compounds. Literature suggests that adjusting the basicity of nitrogen atoms, particularly within heteroaromatic or aliphatic amine groups, can mitigate hERG-related cardiotoxicity (Carvalho et al. 2013; Lagu et al. 2020). Similarly, reducing lipophilicity and

enhancing hydrogen bonding capacity can decrease central nervous system penetration while preserving cellular uptake (Banks et al. 2024);

In summary, this thesis advances two key areas of antimicrobial discovery: improving ML prediction accuracy, and mechanistic validation using a customized CRISPRi tool in *P. aeruginosa* PA14. Together, these strategies offer scalable and adaptable frameworks for identifying and characterizing novel antimicrobial agents against multidrug-resistant pathogens.

3.2- Future directions

In objective I, we predicted activity on 1 million compounds with five different models trained with various training datasets. 139 compounds were predicted active by all 5 model-training dataset combination. At the time of writing this thesis, we have acquired a subset of these compounds but have yet not received them; we will experimentally test their activity to further assess the predictive capabilities of the models and determine if any new hits can be identified. Moreover, we would like to apply explainable AI (xAI) on these 139 compounds to investigate which part of the different molecules confers the growth inhibitory activity.

As described in Section 3.1, pre-training the models with compound structures can enhance their predictive capabilities. Therefore, we will pre-train our DMPNN model – as it showed the best performance – with chemical structures and will then use it to predict activity in the Discovery set, using the spike molecules to assess whether the predictive capabilities of the models have increased.

Another important point that we would like to add into our machine learning-driven discovery pipeline is the analysis of the ADMET properties of the predicted compounds. As it was indicated in section 2.6, the pharmacokinetic properties of any identified hits are a key factor in the promotion of the compound into further development stages. Filtering out compounds that will not pass the pharmacokinetic filter early on would also decrease the costs and times associated with drug discovery.

In objective II, we were able to demonstrate that our CRISPRi system can be used to study compound hypersusceptiblity in essential gene knockdown mutants. Our lab has previously created a CRISPRi-mediated Pooled Library of Essential Genes (CIMPLE) in *B. cenocepacia*

K56-2, which was used to create chemical-genetic interactions profiles (CGIP) with 5,000 compounds (Cardona lab, unpublished results). We would like to construct a similar CIMPLE library in *P. aeruginosa* PA14 to study the chemical-genetic interactions with any potential identified hits from the experiments mentioned above.

CHAPTER IV

4. Methods and Materials

4.1- Bacterial strains and growth conditions

Strains and CRISPRi mutants **(Table 4)** were grown in LB-Lennox medium (Difco) or a 50% dilution of Cation-Adjusted Mueller-Hinton broth (Oxoid) and Phosphate Saline Buffer (Sigma) at 37ºC with shaking at 230 rpm and supplemented with antibiotics as required for plasmid **(Table 5)** selection. The induction of CRISPRi mutants was performed in CAMHB media supplemented with their IC_{50} concentration of rhamnose. A stock solution of L-rhamnose (Sigma) was prepared at 20% w/v in deionized water, then filtered and sterilized.

Primers and oligonucleotides (**Table 6**) were ordered from Integrated DNA Technologies.

The following selective antibiotics were used: kanamycin (Fisher Scientific, 40 µg/mL for *E. coli*), tetracycline (Sigma, 60 µg/mL for *P. aeruginosa,* 20 µg/mL for *E. coli*), carbenicillin (BioSynth, 300 µg/mL for *P. aeruginosa*), ampicillin (Sigma, 100 µg/mL for *E. coli*), chloramphenicol (Sigma, 20 µg/mL for *E. coli*), and trimethoprim (Sigma. 50 µg/mL for *E. coli*).

Table 4. List of strains used in this study.

Strain	Features	Source
***Pseudomonas aeruginosa* PA14**	Human wound isolate	(Mathee 2018)
***Burkholderia cenocepacia* K56-2**	Clinical isolate from a cystic fibrosis patient; ET12 lineage	(Darling et al. 1998)
***P. aeruginosa* PAO1**	Human wound isolate	(Holloway and Zhang 1990)
***Acinetobacter baumannii* 1225**	Clinical isolate, coccyx	(Weber et al. 2013)
***Enterobacter cloacae* ENT001_EB001**	Wild-type	From Dr. Ayush Kumar's lab at the University of Manitoba

Methicillin-resistant *Staphylococcus aureus* (MRSA) subsp. aureus ATCC33592	Gentamicin and methicillin resistant, pvl negative	ATCC
***Klebsiella pneumoniae* ESBL_120310**	Extended-spectrum β-lactamase positive	From Dr. George G. Zhanel's lab at the University of Manitoba
***Escherichia coli* BW25113**	*rrn*B3 Δ*lac*Z4787 *hsd*R514 Δ(*ara*BAD)567 Δ(*rha*BAD)568 *rph*-1	Cardona lab collection
***P. aeruginosa* ATCC27853**	Clinical blood isolate	ATCC
P. aeruginosa LESB58	Liverpool epidemic strain	From Dr. Cezar Khursigara's lab at the University of Guelph
***P. aeruginosa* LES400**	Liverpool epidemic strain	From Dr. Cezar Khursigara's lab at the University of Guelph
***P. aeruginosa* LESlike1**	Liverpool epidemic strain-like	From Dr. Cezar Khursigara's lab at the University of Guelph
***P. aeruginosa* LESlike5**	Liverpool epidemic strain-like	From Dr. Cezar Khursigara's lab at the University of Guelph
***E. coli* DH5α**	F^- Φ80 *lac*ZΔM15 Δ(*lac*ZYA*arg*F) U169 r*ec*A1 *end*A1 *hsd*R17 (r_H^+ , m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	Cardona lab collection
***E. coli* MM290**	F^- Φ80 *lac*ZΔM15 *end*A1 *rec*A1 *hsd*R17 (r_H^+ , m_K^+)	Cardona lab collection

	*sup*E44 thi-1 Δ*gyr*A96 (Δ*lac*ZYA-*arg*F) U169 *rel*A1	
***P. aeruginosa* PA14::dCas9**	Derived from PA14; pAH-CTX1-rhadCas9 was integrated at the *attB* site; clean deletion of plasmid accessory genes	This study
***P. aeruginosa* PA14::dCas9-NTC**	CRISPRi mutant with non-targeting sgRNA	This study
***P. aeruginosa* PA14::dCas9-*fda*_sgRNA1**	CRISPRi mutant targeting *fda* (PA14_07230) – 43 bp downstream of the ATG site	This study
***P. aeruginosa* PA14::dCas9-*fda*_sgRNA2**	CRISPRi mutant targeting *fda* (PA14_07230) – 344 bp downstream of the ATG site	This study
***P. aeruginosa* PA14::dCas9-*fda*_sgRNA3**	CRISPRi mutant targeting *fda* (PA14_07230) – 925 bp downstream of the ATG site	This study
P. aeruginosa* PA14::dCas9-*groES	CRISPRi mutant targeting *groES* (PA14_57020)	This study
***P. aeruginosa* PA14::dCas9-*rpoB*_sgRNA1**	CRISPRi mutant targeting *rpoB* (PA14_08760) – 36 bp downstream of ATG site	This study
***P. aeruginosa* PA14::dCas9-*rpoB*_sgRNA2**	CRISPRi mutant targeting *rpoB* (PA14_08760) – 396 bp downstream of ATG site	This study

Table 5. Plasmids used in this study.

Plasmid	Features	Source
pAH-CTX1-rhadCas9	Derived from mini-CTX1; contains a codon optimized *dcas9* under the control of a rhamnose-inducible promoter (*rhaS rhaR* P_{rhaB} Tet^R)	(Hogan et al. 2019)
pRK2013	ori_{colE1} RK2 derivative Kan^R mob^+ tra^+	(Figurski and Helinski 1979)
pFLP2	Amp^R $sacB^+$ flp^+	(Choi et al. 2005)
pSCB2-sgRNA	Derived from pSCrhaB2-sgRNA, contains scaffold and promoter for sgRNA expression. Tp^R	(Hogan et al. 2019)
pJO1	Derived from pSCB2-sgRNA, with SapI sites for sgRNA added by inverse PCR. Tp^R	This study
pJO4	Derived from pJO1 with Amp^R cloned into *XhoI* and *XmaI* sites	This study
pJO4-NTC	pJO4 expressing a non-targeting sgRNA	This study
pJO4-*fda1*	pJO4 expressing a sgRNA targeting *fda* 43 bp downstream of the ATG site	This study
pJO4-*fda2*	pJO4 expressing a sgRNA targeting *fda* 344 bp downstream of the ATG site	This study

pJO4-*fda3*	pJO4 expressing a sgRNA targeting *fda* 925 bp downstream of the ATG site	This study
pJO4-*groES*	pJO4 expressing a sgRNA targeting *groES*	This study
pJO4-*rpoB1*	pJO4 expressing a sgRNA targeting *rpoB* 36 bp downstream of the ATG site	This study
pJO4-*rpoB2*	pJO4 expressing a sgRNA targeting *rpoB* 396 bp downstream of the ATG site	This study

Table 6. Primers and oligonucleotides used in study.

Primer	Sequence (5'-3')	Notes
3925	GCTGGGTGGTGACTGACTC	Forward primer for colony PCR for mini-CTX1::dCas9 insertion
3926	ACAGGCCGAAGAACACCAG	Reverse primer for colony PCR for mini-CTX1::dCas9 insertion
1292	GCTAACGCAGTCAGGCAC	Forward primer for colony PCR to test Tet^R cassette
1293	CCATTCAGGTCGAGGTGG	Reverse primer for colony PCR to test Tet^R cassette
1026	CCCTGGAAGCTCCCTCGTG	Forward primer for colony PCR to test pFLP2 excision
1284	ATAATAGGTACCTCTTAGACGTCAGGTGGCAC	Reverse primer for colony PCR to test pFLP2 excision
3990	GGACGCTCCTAGTCGAAGCTCTTCTGTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGC	Forward primer for reverse PCR of pSCB2-gRNA to add SapI sites for Golden Gate

3991	CGACTAGGAGCGTCCATGGCTACGCTCTTCTACTAGTATTATACCTAGGACTGAGCTAGC	Reverse primer for reverse PCR of pSCB2-gRNA to add SapI sites for Golden Gate
3992	GCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC	Reverse primer for colony PCR of pJO# plasmids to check for sgRNA insertion
3993	GCTAGCTCAGTCCTAGGTATAATACTAGT	Forward primer for colony PCR of pJO# plasmids to check for sgRNA insertion
3916	ACTACGCTCGAGCATAACATCAAACATCGACCCAC	Reverse primer to remove Tp^R from pJO1 with *XhoI* restriction site
3921	TCTATGCCCGGGCGGTGATGTCGGCGATATAG	Forward primer to remove Tp^R from pJO1 with *XmaI* restriction site
4010	ACTACGCTCGAGTCTTAGACGTCAGGTGGCAC	Reverse primer to amplify Amp^R from pFLP2 with *XhoI* restriction site
4011	AATATCCCCGGGTGGTCTGACAGTTACCAATGCT	Forward primer to amplify Amp^R from pFLP2 with *XmaI* restriction site
3922	GAACGCAGCGGTGGTAAC	Forward primer for colony PCR of pJO4
4016	GGGATAATACCGCGCCACA	Reverse primer for colony PCR of pJO4
4014	TGCTCGATGGCTCTTCGAGT	Forward primer to amplify sgRNA
4015	CCAGTGAGAGCTCTTCGAAC	Reverse primer to amplify sgRNA

NTC	CCAGTGAGAGCTCTTCGAGT**ATGAGTTAGTAT GAGTTAGT**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with non-targeting sgRNA
4047	CCAGTGAGAGCTCTTCGAGT**GTCGTAACTCTC CCAAACAG**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *groES*
4051	CCAGTGAGAGCTCTTCGAGT**CGGCACGCCGTA GCCGAATT**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *fda1*
4052	CCAGTGAGAGCTCTTCGAGT**CGTTGTAGTCGT AGTCGGCA**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *fda2*
4053	CCAGTGAGAGCTCTTCGAGT**AATGTCGCGCAT GGCCTCGA**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *fda3*
4072	CCAGTGAGAGCTCTTCGAGT**GCAACTTGCTAA AGTCCTTG**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *rpoB1*
4073	CCAGTGAGAGCTCTTCGAGT**TACCGTTCTCGG TCATCAGG**GTTCGAAGAGCCATCGAGCA	Oligonucleotide with sgRNA targeting *rpoB2*

[1]Underline sections correspond to restriction recognition sites
[2]Bolded sections correspond to sgRNA sequences

4.2- Scanning Electron Microscopy

Prior to Scanning Electron Microscopy (SEM), the samples were chemically fixed as follows: a 10 µL aliquot of overnight cultures of *P. aeruginosa* ATCC27853, grown in LB or CAMBH broth at 37°C with shaking at 140 rpm, were placed in round glass coverslips (10 mm diameter, Chance Proper Ltd., UK) with a 10 µl of fixative (3% glutaraldehyde in 0.1% sodium cacodylate, pH 7.3). The coverslips were previously coated with polylysine (Sigma-Aldrich) to assist adherence of bacterial cells. The cells were allowed to fix for 10 min following which the supernatant was discarded, and excess fixative was added to the coverslip for 5 min followed by a final washing step with 0.1 M sodium cacodylate, pH 7.3. To dehydrate the bacteria, the coverslips were successively placed for 10 min in each one of the following solutions: 30%, 50%, 70%, 90%,

and 100% (twice) (v/v) acetone. The coverslips were then dried with a critical point drier and sputter coated with Au: Pt, 60:40 in argon (Polarow E5100).

Light and Epifluorescence microscopy examination of P. aeruginosa cells was performed using a Nikon Eclipse E800 microscope equipped with 40 × and 60 × water objectives, differential interference contrast (DIC) polarizing filters and reflectance optics. For epifluorescence microscopy, the microscope was equipped with a 100 W Hg-vapour discharge lamp and fluorescent images were obtained using the following filters: B-2A blue excitation filter with excitation wavelength 470-490 nm, (Nikon) and a Red excitation filter: Cy5 HYQ (Nikon). Images were captured by a Micromax RTE/CCD-732-7 (Princeton Instruments, Trenton, NJ, USA) camera and MetaVue 5.0 software (Universal Imaging Co., Downingtown, PA, USA).

4.3- Compound formulation

All compounds listed in Table S3, as well as ZR50, were acquired from VitasM or MolPort in their powder form and were later diluted in dimethyl sulfoxide (DMSO, Fisher) to a concentration of 100 mM to prepare working solutions.

4.4- Compound susceptibility testing using minimal inhibitory concentration assays

To determine the minimum inhibitory concentration of any predicted active compounds, the broth microdilution method was employed, following the guidelines provided by the Clinical and Laboratory Standards Institute (Clinical and Laboratory Standards Institute CLSI 2012). All inoculums were prepared in Phosphate-buffered saline solution (PBS, Dulbecco) from overnight

cultures diluted to a turbidity equal to a 0.5 McFarland standard and then diluted in cation-adjusted Mueller-Hinton broth (CAMHB, Oxoid) to produce $5x10^5$ CFU/mL in each well.

4.5- Small molecule library optimization

The datasets used to train the various ML models were extracted from the web, as described in the preceding text. These datasets underwent a cleaning process as follows: firstly, all compounds' SMILES were canonicalized using RDKit (Landrum et al. 2025), and the largest fragment was retained, ensuring the elimination of any counterions. Any SMILES that were not considered valid by the program were also removed. Secondly, any compound with a molecular weight lower than 150 g/mol was also removed. Thirdly, compounds that had conflicting labels were also discarded. Finally, all duplicate compounds were identified, and if they had the same activity label, all but one were deleted. The final training dataset size and the final number of active compounds are reported in Table S2.

4.6- t-SNE visualization

To create the t-SNE (t-Distributed Stochastic Neighbour Embedding) plots, the libraries pandas, RDKit, t-SNE and matplotlib were used for data manipulation, extraction of chemoinformatic functionalities, dimensionality reduction, and plotting, respectively.

The Extended-Connectivity Fingerprints (Rogers and Hahn 2010) (ECFP, also known as Morgan Fingerprints) were calculated from SMILES strings, where each SMILES is converted into a molecular object, and a bit-vector fingerprint is calculated using a radius of 2 and a bit size of 2048.

The high-dimensional ECFP features were reduced to two dimensions using t-SNE with default parameters and a random state of 42 to ensure reproducibility. The t-SNE results were visualized as a 2D scatter plot.

4.7- Creation of the discovery set

The discovery sets used for this study are a small subset of molecules selected from a larger parent library of compounds, comprising sixty million compound SMILES downloaded from the PubChem Database (Kim et al. 2023). The creation of these datasets began by canonicalizing the SMILES strings using RDKit, retaining the largest fragment, which removes counter ions and facilitates a more accurate comparison of the SMILES to one another (Weininger 1988; Landrum et al. 2025). After canonicalization, duplicate compounds are removed, ensuring that each compound is represented only once in the database. Finally, the last sanitization step involved filtering out FDA-approved antibiotics to ensure they were not present in the final dataset. An important filter applied during the creation of the discovery set was the filtering of any compounds with scaffolds similar to known antibiotics, in an effort to promote chemical diversity and ensure the identification of novel antimicrobials.

After cleaning, 59,574,674 compounds remained and were converted to Morgan fingerprints (radius = 3, nbits = 2048) using CHEMFP, a free software that converts SMILES to fingerprints and can also perform clustering (Rogers and Hahn 2010; Dalke 2019). The dataset was divided randomly into four datasets of equal size, and using CHEMFP, each dataset was clustered using the distance-based Butina clustering method. A cutoff of 0.7 was used, and the molecule at the center of each cluster was selected and compiled into a list, which included compounds from all four datasets. This list ultimately contained 5 million molecules. These molecules were then

clustered one last time, and the center molecule of each cluster was kept, making up the Discovery sets. The number of clusters used depended on the desired final size of the discovery set.

4.8- Machine Learning methods

In this study, we evaluated five distinct ML architectures for molecular activity prediction: D-MPNN, D-MPNN + RDKit, ECFP + Random Forest, ChemBERTa + XGBoost, and KPGT + Fully Connected Network (FCN).

a. D-MPNN (Directed Message Passing Neural Network): Based on the Message Passing Neural Network (MPNN) framework (Gilmer et al. 2017), D-MPNN enhances message propagation by using bond-centred, directional message passing, reducing redundant loops and improving the signal-to-noise ratio during training. It encodes atom- and bond-level features and then aggregates hidden states to form a molecule-level representation (Heid et al. 2024).

b. D-MPNN + RDKit Features: This variant extends D-MPNN by incorporating 200 RDKit molecular descriptors – physicochemical, topological, and electronic features - into learned embeddings, enhancing expressiveness with domain knowledge.

c. ECFP + Random Forest: This approach utilizes Extended-Connectivity Fingerprints (ECFP, radius = 4 and bit = 2048) to capture molecular substructures and classify binary fingerprints using a Random Forest model to establish a robust and interpretable baseline.

d. ChemBERTa + XGBoost: ChemBERTa, a transformer-based model pre-trained with 77 million SMILES strings, learns chemical patterns via contextual embeddings. These embeddings are classified using KGBoost for high-performance tabular predictions (Chithrananda et al. 2020).

e. KPGT + FCN: The Knowledge-guided Pretraining of Graph Transformers (KPGT) model is trained on 2D molecular graphs of approximately two million compounds to learn enriched chemical representations (Li et al. 2023). These embeddings are then passed through a fully connected network (FCN) for downstream classification.

4.9- Hyperparameter optimization

We conducted systematic hyperparameter optimization to enhance model performance using tailored strategies for each model architecture. For the DMPNN and DMPNN + RDKit models, we employed the Optuna optimization framework (Akiba et al. 2019), integrated with the Ray Tune package (Liaw et al. 2018), to efficiently explore the hyperparameter space through adaptive sampling techniques. In contrast, we applied a randomized grid search strategy to identify optimal configurations for the ECFP + Random Forest, ChemBERTa + XGBoost, and KPGT + FCN models.

4.10- Evaluation methods

To binarize the obtained predictions, we applied F1 score maximization using a subset of small molecules with known experimental labels, hereafter referred to as spike molecules. For this, we used a Python-based approach using the roc_curve function from the scikit-learn library to derive the thresholds based on the false positive and true positive rates.

The spike molecules in the file were binarized based on each obtained threshold, and the F1 score for that specific cutoff point was calculated using the f1_score function, which accounts for precision and recall. The script iterated through all the possible thresholds to find the one that yielded in the highest F1 score.

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall}$$

4.11- Construction of a P. aeruginosa dCas9 insertional mutant

pAH-CTX1-rhadCas9 was introduced into *P. aeruginosa* PA14 by tri-parental mating. Briefly, equal volumes of the *E. coli* donor containing the plasmid to be mobilized, the *E. coli* MM290/pRK2013 helper, and the *P. aeruginosa* PA14 recipient were mated on an LB plate and incubated overnight (18-20 hours) at 37ºC. Successful conjugants were selected on LB plates containing the appropriate antibiotic and 20 µg/mL of chloramphenicol to counterselect for the *E. coli* donor and helper. Colonies that appeared within 24 hours at 37ºC were screened by colony PCR using primers 3925-3926.

To remove the tetracycline resistance and integrase genes from the insertional mutants, the pFLP system (Choi et al. 2008). *E. coli* DH5α/pFLP2 was inserted into *P. aeruginosa* PA14::CTX1-rhadCas9 via triparental mating using *E. coli* MM290/pRK2013 as the helper. Tetracycline-sensitive colonies were screened by PCR using primers 1292 and 1293. pFLP2 contains the *sacB* genes; therefore, conjugants were grown in LB supplemented with 15% sucrose at 37ºC overnight to cure the plasmid. Single colonies were tested for carbenicillin sensitivity and then confirmed by colony PCR with primers 1026 and 1284.

4.12- Construction of the sgRNA-expressing plasmids

sgRNAs were designed as described previously (van Gestel et al. 2021). Obtained sgRNAs were further filtered to discard sgRNAs associated with PAMs beginning with an adenine (‘AGG’) and sgRNAs that had thymine in the first three positions at the 5’-end. We then manually selected

between 1 and 3 sgRNAs for each targeted gene, covering different distances from the ATG site (**Figure 9D**).

pSCB2-sgRNA was then modified to include Golden-Gate compatible sites and to exchange the resistance cassette, as trimethoprim is not usable in *P. aeruginosa.* Briefly, pSCB2-sgRNA was amplified by inverse PCR using Q5 polymerase (NEB) and primers 3990 and 3991. The resulting fragment was ligated by blunt-end ligation using 1 µL of PCR product incubated with 1 µL of *DpnI,* 1 µL of T4 polynucleotide kinase, and 1 µL of T4 ligase (NEB) with quick ligation buffer (NEB) at 16ºC overnight.

Chemically competent *E. coli* DH5α were transformed with 2 µL of ligation mix using a standard heat-shock method. After a 2-hour recovery period with shaking at 230 rpm in SOC medium (Difco), the cells were diluted and plated on LB agar supplemented with trimethoprim and incubated overnight at 37°C. Transformant colonies were screened for pJO1 by colony PCR with primers 3992 and 3993, which span the ligated junction.

pJO1 was then further modified to exchange its resistance cassette from trimethoprim to ampicillin. pFLP2 was used to amplify the ampicillin resistance cassette by PCR using Q5 polymerase (NEB) and primers 4010 and 4011, which included *XmoI* and *XhaI* restriction sites at their 5' ends, respectively. The trimethoprim resistance cassette was eliminated from pJO1 by PCR amplification using Q5 polymerase (NEB) and primers 3916 and 3921, which added *XmoI* and *XmaI* restriction sites. Both PCR products were combined by restriction cloning using *XhoI* and *XmaI* (NEB). The resulting plasmid, pJO4, was transformed into *E. coli* DH5α as described above. Transformants were selected in LB supplemented with 100 µg/mL of ampicillin and confirmed by colony PCR using primers 3922 and 4016.

4.13- Insertion of sgRNAs into pJO4

sgRNAs were designed using a previously published Python script (van Gestel et al. 2021). Whenever possible, multiple sgRNAs were chosen for each gene: sgRNA 1 was designed to bind near the gene's start, sgRNA 2 targeted the middle region, and sgRNA 3 was selected to bind as far from the ATG site as possible. To minimize off-target gene silencing, any sgRNA with an 11 bp seed region matching another gene site adjacent to a PAM site was excluded. Additionally, sgRNAs were removed if they exhibited any of the following characteristics, which are thought to reduce silencing efficiency: 1) a PAM site of 'ANN' or 'TNN', 2) thymine in the first three positions at the 5' end, and 3) a GC content below 20% or above 80%.

The selected sgRNAs were ordered as 60 bp single-stranded oligos from Integrated DNA Technologies and subsequently PCR-amplified to create double-stranded DNA using Q5 polymerase (NEB) and primers 4014 and 4015. The PCR products were introduced into pJO4 by Golden Gate cloning with SapI, with the following cycle parameters: (5 mins 37ºC → 5 min 16ºC) x 50 cycles followed by 5 mins at 60ºC and 5 mins at 85ºC.

The newly constructed plasmids were transformed into *E. coli* DH5α as described above, and ampicillin-resistant colonies were screened by colony PCR with primers 3992 and 3993. Later, triparental mating with *E. coli* DH5α/pJO4-sgRNA as the donor and *E. coli* MM290/pRK2013 as the helper was conducted as described previously to introduce pJO4-sgRNA into *P. aeruginosa* PA14::dCas9. Conjugants were selected in LB supplemented with 300 µg/mL carbenicillin and screened by colony PCR with primers 3992 and 3993.

4.14- Extraction of plasmid and genomic DNA

Plasmid DNA was routinely extracted using the EZNA Plasmid DNA Mini Kit (Omega Bio-tek) or the Monarch® Plasmid Miniprep Kit (NEB) and eluted in 10 mM Tris-HCl pH 8.5. Genomic DNA was isolated from *P. aeruginosa* with the PureLink Genomic DNA Mini Kit (Invitrogen) and eluted in 10 mM Tris-HCl pH 8.5.

4.15- Genome sequencing

Genomic DNA was extracted as described above. Long-read sequencing was performed by Plasmidsaurus using Oxford Nanopore Technologies (ONT). Briefly, the genomic DNA was minimally fragmented, and the amplification-free long-read sequencing library was constructed using the v14 library prep chemistry. The library was then sequenced using R10.4.1 flow cells (ONT). Genome assembly was performed using Flye v2.9.1 (Kolmogorov et al. 2019) with high-quality ONT reads and subsequently polished with Medaka v1.8.0. Geneious 8.1.9 was used to analyze and annotate the sequences using the parental strains as a source.

4.16- SDS-PAGE and immunoblotting

Cells from overnight cultures of *P. aeruginosa* PA14::dCas9 and *P. aeruginosa* PA14 WT were subcultured into fresh LB medium and grown at 37ºC with shaking at 230 rpm to an OD600 of 0.3; they were then exposed to various concentrations of rhamnose (0%, 0.05%, 0.5% and 1%) and grown for 3 hours to induce the production of dCas9. The cells were then normalized to an OD600 of 0.1, and soluble protein was isolated by sonication in TNG Buffer (100 mM Tris-HCl, 150 mM NaCl, 10% glycerol, pH 7.4) and centrifugation at 15,000 × g for 5 minutes. Following boiling denaturation in SDS loading buffer (50 mM Tris-HCl, 2% SDS, 0.2% bromophenol blue,

20% glycerol, pH 7.4), the samples were run on an 8% Tris/Glycine gel in duplicate, one for immunoblotting and one for Coomassie staining. As a control, we loaded 25 ng of purified Cas9.

Protein was transferred by iBlot (Invitrogen) to a PVDF membrane, blocked in 5% skim milk-TBST buffer (150 mM NaCl, 10 mM Tris-HCl, 0.5% Tween-20-, pH 7.5) at room temperature for 1 hour, then probed with a 1:2000 dilution of a primary α-Cas9 antibody (ThermoFisher 10C11-A12) in 5% skim milk-TBST overnight at 4ºC. Following washes, the blot was probed with a 1:20,000 dilution of the secondary antibody, linked to alkaline phosphatase (ThermoFisher G-21060), in 5% skim milk-TBST for 1 hour at room temperature. Protein was detected by incubation with a solution of NBT/BCIP (Roche) according to the manufacturer's instructions.

4.17- *Growth of P. aeruginosa in dilutions of CAMHB and biofilm quantification*

P. aeruginosa is known for its biofilm formation capabilities (Grace et al. 2022), which can impede optical density (OD) readings when assays are done in a microwell plate format. To minimize biofilm formation but ensure consistency with previous assays (e.g., MIC assays), we conducted growth tests in cation-adjusted Mueller-Hinton broth (CAMHB) diluted to varying concentrations (25%-100%) with phosphate-buffered saline (PBS). We observed that a 50% dilution reduced biofilm formation without compromising viability (**Figure S2**).

P. aeruginosa PA14::dCas9-NTC was grown overnight in LB media with the appropriate antibiotic. The overnight cultures were diluted to 0.001 in CAMHB with the proper antibiotic and arrayed in 96-well plates at the appropriate CAMHB + PBS dilutions. The mutant was grown at 37ºC with continuous shaking at 230 rpm, and the OD600 was measured in real-time every 30

minutes using a BioTek Synergy 2 microplate reader. The data from the microplate reader was processed in R.

The biofilm formation in the 96-well plate was measured according to the protocol detailed in Chapter 48 of "*Pseudomonas*: Methods and Protocols" (Filloux and Ramos 2014). Briefly, planktonic cells were removed from the microplate, and 0.1% crystal violet was added to each well to cover the biofilm formed. After ensuring absorption, the crystal violet was rinsed, and the plate was left to air dry completely. Later, 30% acetic acid was added to solubilize the formed biofilm, and 200 µL of each well was transferred to a new optically clear, flat-bottom 96-well plate. The OD550 of the samples was measured on the BioTek Synergy 2 microplate reader, and the data were processed in R version 4.4.2.

Statistical analysis of the biofilm formation assay was performed on R using base R and the ggpbur package. Here, a Kruskal-Wallis test was conducted to determine if significant differences existed in biofilm measurements across the four tested conditions. We chose this non-parametric method because the data did not meet the assumption of normality. After this, pairwise comparisons between conditions were conducted using the Wilcoxon rank-sum test with a Bonferroni correction for multiple comparisons. This test determined specific group differences following the Kruskal-Wallis analysis.

4.18- Growth curves

The mutants were grown overnight in LB medium supplemented with the appropriate antibiotic (300 µg/mL carbenicillin) at 37ºC with shaking at 230 rpm. The overnight cultures were then diluted in PBS to an OD600 of 0.001 and arrayed in a 96-well plate in duplicate with serial dilutions of rhamnose in CAMHB, ranging from 0.5% to 0%. The mutants were grown at 37ºC

with continuous shaking at 230 rpm, and their OD600 was measured in real-time every 30 minutes for 20 hours using a BioTek Synergy 2 microplate reader. The data was then processed in R version 4.4.2.

4.19- Rhamnose dose-response curves

From the growth curves, we extracted the growth values for the different concentrations of rhamnose at 11 hours of growth (middle of the exponential phase for all mutants). This data was then used to plot growth vs rhamnose concentration and compute the dose-response curves. The data were analyzed using R version 4.4.2, and the dcr package was employed for fitting Hill's equation and calculating IC_{50} values. Any fitting with a Hill's coefficient ≤ 1 was deemed as not having a dose-response relationship, and therefore, the highest concentration of rhamnose was used as the IC_{50}.

4.20- Hypersusceptibility assay with glucose as the sole carbon source

Overnight cultures of the CRISPRi mutants were diluted 1:100 and subcultured in CAMHB supplemented with mutant-specific Rha IC_{50} and 300 µg/mL of carbenicillin for 2 hours at 37°C with shaking at 230 rpm. After 2 hours, the cells' OD600 was adjusted to 0.001 and grown in a 96-well format containing M9 media (Difco) with glucose as the sole carbon source, supplemented with 300 µg/mL of carbenicillin, Rha IC_{50}, and various concentrations of ZR50. The plates were incubated for 20 hours at 37ºC with continuous shaking at 230 rpm. OD600 readings were obtained using a BioTek Synergy 2 microplate reader, and the data were analyzed with R version 4.4.2.

CHAPTER V

5. Appendix

5.1- Supplemental tables

Table S1. *Training datasets used in this study.*
Attached Excel file.

Table S2. *Discovery set 1 used for predictions.*
25 compounds (JN01-JN25)were acquired and experimentally tested. Attached Excel file.

Table S3. *Acquired compounds for experimental confirmations*

Compound	Compound CID	Smile
JN01	326810	Cc1nn(-c2ccccc2)c2c1C(c1ccccc1)C(C#N)=C(N)O2
JN02	717919	O=C(c1ccccc1)N1CCN(CC=Cc2ccccc2)CC1
JN03	2815057	N#CC1=C(N)Oc2c(c(=O)oc3ccccc23)C1c1ccccc1
JN04	772424	COc1ccc(C=NNC(=O)c2cccnc2)cc1OC
JN05	1092379	O=C(c1ccccc1)N1CCN(CCN2CCN(C(=O)c3ccccc3)CC2)CC1
JN06	2182759	COc1cc(C=C2N=C(c3cccc(F)c3)OC2=O)cc(Cl)c1OCc1ccccc1
JN07	3325640	COc1cc(C=NNC(=O)c2ccc(F)cc2)ccc1OCc1ccccc1
JN08	4440943	O=C(Nc1ccccc1)C(Cc1ccccc1)N1C(=O)C2C3C=CC(C4CC34)C2C1=O
JN09	2457533	O=C(COc1ccc2c(c1)OCO2)Nc1ccc2c(c1)OCCO2
JN10	85270	CCOC(=O)c1cnn(-c2ccccc2)c1N
JN11	414907	c1cc(C2=NCCN2)ccc1C1=NCCN1
JN12	3092607	COc1cc(C=O)ccc1OCCOc1ccc(C=O)cc1OC
JN13	587080	O=C(NCCNC(=O)c1ccc(Cl)cc1)c1ccc(Cl)cc1
JN14	272218	O=C(NCc1ccco1)c1ccc(C(=O)NCc2ccco2)cc1
JN15	895580	O=C(C=Cc1ccc2c(c1)OCO2)Nc1ccc2c(c1)OCCO2
JN16	68740078	COc1cc(C=Cc2ccc(OC(C)=O)c(OC)c2)ccc1OC(C)=O
JN17	4621453	CCCCCN1C(=O)C(=C2C(=O)N(CCC)c3ccccc32)SC1=S
JN18	1536029	O=c1nc2sc(=Cc3ccc(F)cc3)c(=O)n2nc1Cc1ccccc1
JN19	2859944	COc1ccccc1C=C1SC(=S)N(c2ccc(F)cc2)C1=O
JN20	3796517	COc1ccc(C2C(C#N)=C(N)Oc3n[nH]c(-c4cccc(OC)c4)c32)cc1
JN21	4164008	Cc1cc(C=C2SC(=O)N(Cc3ccc(F)cc3)C2=O)c(C)n1-c1ccc(F)cc1
JN22	2849619	CCOc1ccc(C=C2C(=O)NN(c3ccccc3)C2=O)cc1OC
JN23	1349631	O=C(Nc1ccc(NC(=O)c2ccccc2C(F)(F)F)cc1)c1ccccc1C(F)(F)F
JN24	2272528	O=C(NCCNC(=O)c1ccccc1C(F)(F)F)c1ccccc1C(F)(F)F
JN25	4086918	C=CCN1C(=O)C(=Cc2ccc(OCC)c(OC)c2)SC1=S
ZR50	46344733	CN1CCN(CC1)c1ncnc2sc(C(=O)NCC3CCN(Cc4ccccc4C)CC3)c(C)c12

[1]Compounds JN01 to JN25 were acquired from VistasM
[2]Compound ZR50 was acquired from MolPort

Table S4. *Data cleaning results.*

	Number of compounds				
	Training_1	**Training_2**	**Training_3**	**Training_4**	**Training_5**
Initial	29,568	39,313	39,128	4,112	20,018

Initial Number of Actives	222	512	218	86	4,411
Step 1	28,951	39,302	39,128	3,356	18,495
Step 2	28,777	39,302	39,128	3,356	18,403
Step 3	28,777	38,021	38,308	3,336	18,403
Step 4	28,777	13,068	13,139	3,336	18,403
Final Size	28,777	13,068	13,139	3,336	18,403
Final number of actives	204	318	169	45	3,764
Activity cutoff	RG < 0.7	RG ≤ 0.2	RG ≤ 0.2	MIC ≤ 16 μg/mL	MIC ≤ 16 μg/mL

Table S5. *Discovery set 2 used for predictions.*
Attached Excel file.

Table S6. *Spike compounds.*

SMILE	Concentration (µM)	*B. cenocepacia* K56-2	*S. aureus* ATCC33592	*E. coli* BW25113	*P. aeruginosa* PAO1
O=[N+]([O-])c1ccc(NC2CC2)nc1	500	-	-	-	-
COC1OC2COC(c3ccccc3)OC2C(OC)C1OC	500	-	-	-	-
Clc1ccccc1-c1nc2ccccc2s1	500	-	-	-	-
Cc1cc(C)c(NC(=O)C(C)Oc2ccc(Cl)cc2Cl)c(C)c1	500	-	-	-	-
O=C(Nc1ccccc1Oc1ccccc1)c1cccs1	500	-	-	+	-
CCN(CC)C(=O)Nc1ccc(OC)cc1	500	-	-	+	-
CCCOC(=O)c1ccc(NC(=O)c2ccccc2C)cc1	500	-	-	-	-
COc1ccc(-c2cc(C(N)=O)n[nH]2)c(F)c1	250	-	-	-	-
O=S(=O)(Nc1nccs1)c1ccc(N=Cc2cc(Br)ccc2O)cc1	250	-	-	+	+
Clc1ccnc(-c2ccccc2)n1	500	-	-	-	-
O=c1nc(C2CCCN2)[nH]c2c1oc1ccc(Cl)cc12	500	-	-	-	-
CC(C)CC1C(=O)N2CCCC2C2(O)OC(NC(=O)C3C=C4c5cccc6[nH]c(Br)c(c56)CC4N(C)C3)(C(C)C)C(=O)N12	500	-	+	-	-
CCN1C(=O)c2cccc3c(NC(=O)c4ccccc4Cl)ccc1c23	250	-	-	+	-
Cc1ccc(C(C)C)c(OCC(O)C[NH2+]C(C)(C)C)c1	500	+	-	+	+
Cc1c(C(=O)NCCc2ccc(Cl)cc2)sc2ncnc(N3CCN(C)CC3)c12	500	-	-	-	-
CC(NC(=O)C1CCN(c2nc3cccnc3s2)CC1)c1ccc2c(c1)OCCO2	500	+	+	+	+
O=c1cnn(C2OC(CO)C(O)C2O)c(=S)[nH]1	500	-	-	-	-
CCCc1c(C(=O)Nc2ccccc2Cl)nnn1-c1ccccc1	500	-	-	-	-
NC(=O)c1ccccc1NC(=O)C1CC(=O)N(C2CCCC2)C1	500	-	-	-	-
CCOC(=O)N1CCN(CC2=CCC3CC2C3(C)C)CC1	500	+	-	+	-
NNC(=O)c1cc(=O)[nH]c2ccccc12	500	-	-	-	-
Brc1csc(CN2CCC(Cc3ccccc3)CC2)c1	500	-	+	+	-
COc1cc(N)c(C)cc1NC(C)=O	500	+	+	+	-
O=C(CN1CCN(c2ncccn2)CC1)c1ccc(F)cc1	500	-	-	-	-
CCOc1ccc(CN2CCN(Cc3cccc4ccccc34)CC2)cc1	125	+	-	-	-
N=c1n(CCOc2ccc(Cl)cc2Cl)c2ccccc2n1CC(N)=O	125	-	-	-	-

CCC1=CC2CN3CCc4c([nH]c5ccccc45)C(C(=O)OC)(C2)C13	500	+	+	+	-
C#CCOc1ccc(Br)cc1C(c1c(C)n[nH]c1O)c1c(C)n[nH]c1O	500	-	+	-	-
Cc1ccc(Sc2ccc([N+](=O)[O-])cc2)cc1	500	+	+	+	+
Cc1ccccc1CN1CCC(CNC(=O)c2sc3ncnc(N4CCN(C)CC4)c3c2C)CC1	250	+	+	+	+
Cc1c(C(=O)NCC2CCN(CCc3cccs3)CC2)sc2ncnc(N3CCN(C)CC3)c12	500	-	-	+	-
Cc1c(C(=O)NCCCc2ccccc2)sc2ncnc(N3CCN(C)CC3)c12	250	-	-	-	-
CCCCCCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)C	50	+	+	+	+
CC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NC1CCCC1	50	+	+	-	+
CCCCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)CC	50	+	+	-	+
CCC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NCCCOC(C)C	50	-	+	-	+
CC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NCC1CCCO1	50	-	-	-	-
CC(C)OCCCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)C	50	-	+	-	-
COCCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)C	50	-	-	-	-
CC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)N1CCCC1C(=O)O	50	-	-	-	-
CC(C)CCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)C	50	+	+	-	+
CCCCCCNC(=O)C(C)NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12	50	+	+	-	+
CCC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NC1CCCCC1	50	-	+	-	+
CC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NC1CCCCC1	50	-	+	-	+

CCC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NCC1CCCO1	50	-	+	-	-
CCC(C)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(=O)NCCCOC	50	-	-	-	-
COCCCNC(=O)C(NC(=O)C1Cc2c([nH]c3ccccc23)C2c3ccccc3C(=O)N12)C(C)C	50	-	-	-	-

[1]All compounds were experimentally tested in the lab and binarized as follows: for *B. cenocepacia, E. coli, and S. aureus* active compounds signify an RG ≤ 0.7; for *P. aeruginosa* active compounds represent an MIC ≤ 16 μg/mL. Active compounds are signaled as "+" while "–" denotes inactive compounds.

Table S7. *IC50 of rhamnose for the different mutants*

Targeted gene	Hill's coefficient	[Rhamnose IC50] (%)
Non-targeting control	0.65	0.5
fda	3.30	0.01
groES	2.56	0.09
rpoB	3.76	0.02

[1]Values were obtained from growth curves at time 11 hours, equivalent to middle of the exponential phase

[2]A Hill's coefficient greater to 1 infers a relationship between the concentration rhamnose and the conditional growth phenotype exhibit by the mutant

[3]These IC50 were calculated from the average of three independent biological replicates with two technical replicates each

5.2- Supplemental figures

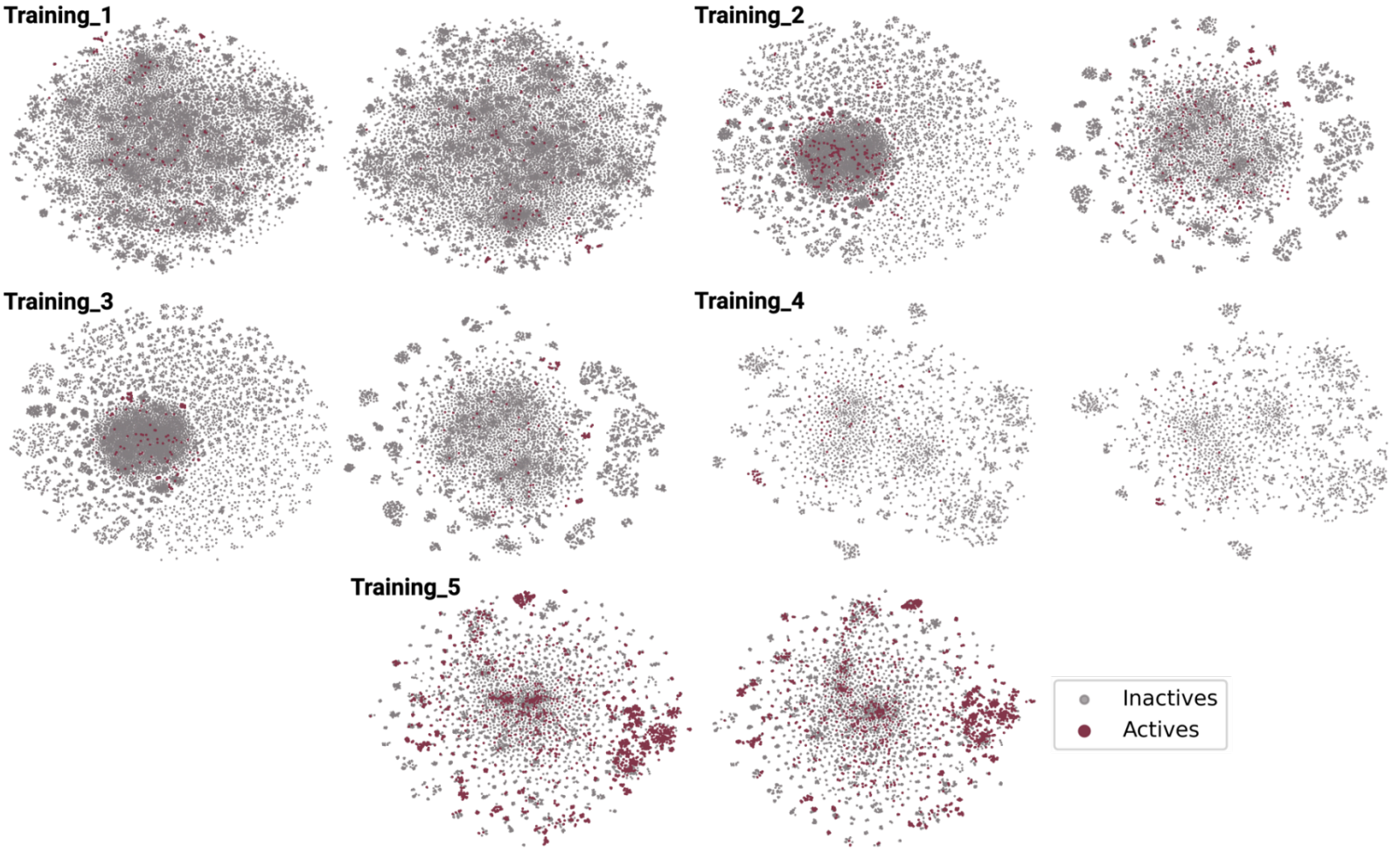


Figure S1. ***Two-dimensional visualization of training datasets.*.** t-SNE analysis was used to visualize the training datasets before (left) and after (right) the clean-up process Although the overall size of the datasets changes, we can observe that the chemical space covered did not.

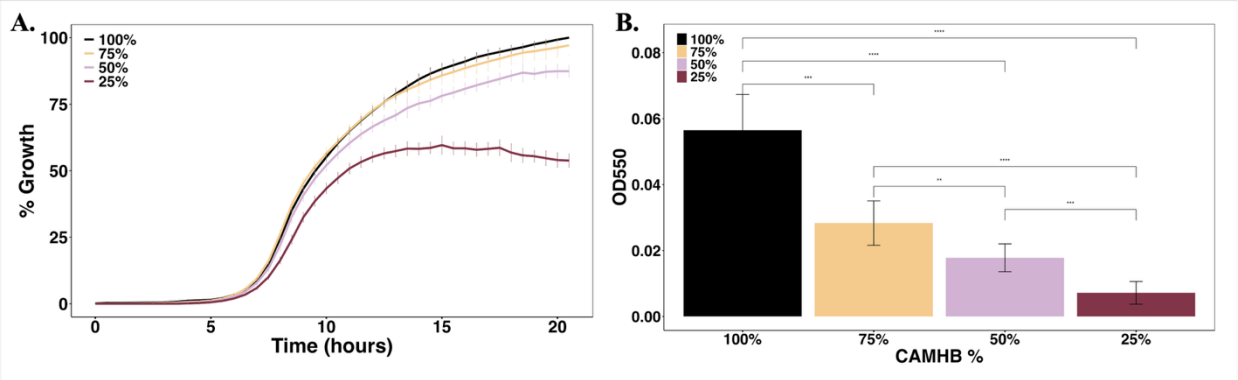


Figure S2. ***Diluting the nutrients in CAMHB inhibits biofilm formation without affecting cell growth.*** *(A)* Growth curves of *P. aeruginosa* PA14::dCas9-NTC at different dilutions of CAMHB with PBS. Results show the average of three biological replicates with sixteen technical replicates each, error bars representing the standard deviation. *(B)* Bar graph showing biofilm production – measured as optical density at 550 nm by the crystal violet method (Filloux and Ramos 2014) – at different dilutions of CAMHB with PBS. Results show the average of three biological replicates with sixteen technical replicates each, error bars representing the standard deviation.

CHAPTER VI

6. References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv. doi:10.48550/ARXIV.1907.10902.

Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. 2016. Deep learning for computational biology. Mol. Syst. Biol. **12**(7): 878. doi:10.15252/msb.20156651.

Armour, A.D., Shankowsky, H.A., Swanson, T., Lee, J., and Tredget, E.E. 2007. The Impact of Nosocomially-Acquired Resistant Pseudomonas aeruginosa Infection in a Burn Unit. J. Trauma Inj. Infect. Crit. Care **63**(1): 164–171. doi:10.1097/01.ta.0000240175.18189.af.

Arnold, A., Alexander, J., Liu, G., and Stokes, J.M. 2023. Applications of machine learning in microbial natural product drug discovery. Expert Opin. Drug Discov. **18**(11): 1259–1272. doi:10.1080/17460441.2023.2251400.

Aslam, B., Wang, W., Arshad, M.I., Khurshid, M., Muzammil, S., Rasool, M.H., Nisar, M.A., Alvi, R.F., Aslam, M.A., Qamar, M.U., Salamat, M.K.F., and Baloch, Z. 2018. Antibiotic resistance: a rundown of a global crisis. Infect. Drug Resist. **11**: 1645–1658. doi:10.2147/IDR.S173867.

Asokan, G.V., Ramadhan, T., Ahmed, E., and Sanad, H. 2019. WHO Global Priority Pathogens List: A Bibliometric Analysis of Medline-PubMed for Knowledge Mobilization to Infection Prevention and Control Practices in Bahrain. Oman Med. J. **34**(3): 184–193. doi:10.5001/omj.2019.37.

Azlim Khan, A.K., and Ahamed Hassain Malim, N.H. 2023. Comparative Studies on Resampling Techniques in Machine Learning and Deep Learning Models for Drug-Target Interaction Prediction. Molecules **28**(4): 1663. doi:10.3390/molecules28041663.

Backes, N., and Phillips, G.J. 2021. Repurposing CRISPR-Cas Systems as Genetic Tools for the Enterobacteriales. EcoSal Plus **9**(2): eESP00062020. doi:10.1128/ecosalplus.ESP-0006-2020.

Banerjee, P., Kemmler, E., Dunkel, M., and Preissner, R. 2024. ProTox 3.0: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. **52**(W1): W513–W520. doi:10.1093/nar/gkae303.

Banks, W.A., Rhea, E.M., Reed, M.J., and Erickson, M.A. 2024. The penetration of therapeutics across the blood-brain barrier: Classic case studies and clinical implications. Cell Rep. Med. **5**(11): 101760. Elsevier BV. doi:10.1016/j.xcrm.2024.101760.

Bendixen, L., Jensen, T.I., and Bak, R.O. 2023. CRISPR/Cas-mediated transcriptional modulation: The therapeutic promises of CRISPRa and CRISPRi. Mol. Ther.: S1525001623001454. doi:10.1016/j.ymthe.2023.03.024.

Bock, C., Datlinger, P., Chardon, F., Coelho, M.A., Dong, M.B., Lawson, K.A., Lu, T., Maroc, L., Norman, T.M., Song, B., Stanley, G., Chen, S., Garnett, M., Li, W., Moffat, J., Qi, L.S., Shapiro, R.S., Shendure, J., Weissman, J.S., and Zhuang, X. 2022. High-content CRISPR screening. Nat. Rev. Methods Primer **2**(1): 1–23. Nature Publishing Group. doi:10.1038/s43586-021-00093-4.

Cai, R., Lv, R., Shi, X., Yang, G., and Jin, J. 2023. CRISPR/dCas9 Tools: Epigenetic Mechanism and Application in Gene Transcriptional Regulation. Int. J. Mol. Sci. **24**(19): 14865. doi:10.3390/ijms241914865.

Camacho, D.M., Collins, K.M., Powers, R.K., Costello, J.C., and Collins, J.J. 2018. Next-Generation Machine Learning for Biological Networks. Cell **173**(7): 1581–1592. Elsevier. doi:10.1016/j.cell.2018.05.015.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. 2001. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. Cell **104**(6): 901–912. doi:10.1016/S0092-8674(01)00286-0.

Cardona, S.T., Rahman, A.S.M.Z., and Novomisky Nechcoff, J. 2025. Innovative perspectives on the discovery of small molecule antibiotics. Npj Antimicrob. Resist. **3**(1): 1–11. Nature Publishing Group. doi:10.1038/s44259-025-00089-0.

Cardona, S.T., Selin, C., and Gislason, A.S. 2015. Genomic tools to profile antibiotic mode of action. Crit. Rev. Microbiol. **41**(4): 465–472. Taylor & Francis. doi:10.3109/1040841X.2013.866073.

Carvalho, J.F.S., Louvel, J., Doornbos, M.L.J., Klaasse, E., Yu, Z., Brussee, J., and IJzerman, A.P. 2013. Strategies To Reduce hERG K^+ Channel Blockade. Exploring Heteroaromaticity and Rigidity in Novel Pyridine Analogues of Dofetilide. J. Med. Chem. **56**(7): 2828–2840. American Chemical Society (ACS). doi:10.1021/jm301564f.

Chan, H.C.S., Shan, H., Dahoun, T., Vogel, H., and Yuan, S. 2019. Advancing Drug Discovery via Artificial Intelligence. Trends Pharmacol. Sci. **40**(8): 592–604. doi:10.1016/j.tips.2019.06.004.

Chithrananda, S., Grand, G., and Ramsundar, B. 2020. ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction. arXiv. doi:10.48550/ARXIV.2010.09885.

Choi, K.H., Gaynor, J.B., White, K.G., Lopez, C., Bosio, C.M., Karkhoff-Schweizer, R.R., and Schweizer, H.P. 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat. Methods **2**(6): 443–448. doi:10.1038/NMETH765.

Choi, K.H., Mima, T., Casart, Y., Rholl, D., Kumar, A., Beacham, I.R., and Schweizer, H.P. 2008. Genetic tools for select-agent-compliant manipulation of *Burkholderia pseudomallei*. Appl. Environ. Microbiol. **74**(4): 1064–1075. doi:10.1128/AEM.02430-07.

Clinical and Laboratory Standards Institute CLSI. 2012. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically; approved standard—ninth edition. CLSI, Pennsylvania, USA.

Cooper, S.J., Leonard, G.A., McSweeney, S.M., Thompson, A.W., Naismith, J.H., Qamar, S., Plater, A., Berry, A., and Hunter, W.N. 1996. The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold. Structure **4**(11): 1303–1315. doi:10.1016/S0969-2126(96)00138-4.

Dalke, A. 2019. The chemfp project. J. Cheminformatics **11**(1): 76. doi:10.1186/s13321-019-0398-8.

Darling, P., Chan, M., Cox, A.D., and Sokol, P.A. 1998. Siderophore production by cystic fibrosis isolates of *Burkholderia cepacia*. Infect. Immun. **66**(2): 874–877.

D'Elia, M.A., Pereira, M.P., and Brown, E.D. 2009. Are essential genes really essential? Trends Microbiol. **17**(Journal Article): 433–438. doi:10.1016/j.tim.2009.08.005.

Desselle, M.R., Neale, R., Hansford, K.A., Zuegg, J., Elliott, A.G., Cooper, M.A., and Blaskovich, M.A. 2017. Institutional profile: Community for Open Antimicrobial Drug Discovery – crowdsourcing new antibiotics and antifungals. Future Sci. OA **3**(2): FSO171. doi:10.4155/fsoa-2016-0093.

Ferreira, L.L.G., and Andricopulo, A.D. 2019. ADMET modeling approaches in drug discovery. Drug Discov. Today **24**(5): 1157–1165. doi:10.1016/j.drudis.2019.03.015.

Figurski, D.H., and Helinski, D.R. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. Proc. Natl. Acad. Sci. U. S. A. **76**(4): 1648–1652.

Filloux, A., and Ramos, J.-L. (*Editors*). 2014. Pseudomonas Methods and Protocols. Springer New York, New York, NY. doi:10.1007/978-1-4939-0473-0.

Fujitani, S., Sun, H.-Y., Yu, V.L., and Weingarten, J.A. 2011. Pneumonia Due to Pseudomonas aeruginosa. Chest **139**(4): 909–919. doi:10.1378/chest.10-0166.

Gang, R.K., Bang, R.L., Sanyal, S.C., Mokaddas, E., and Lari, A.R. 1999. Pseudomonas aeruginosa septicaemia in burns. Burns **25**(7): 611–616. doi:10.1016/S0305-4179(99)00042-X.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. **109**(39): E2579-2586. doi:10.1073/pnas.1208507109.

Gellert, M., O'Dea, M.H., Itoh, T., and Tomizawa, J. 1976. Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. Proc. Natl. Acad. Sci. U. S. A. **73**(12): 4474–4478.

van Gestel, J., Hawkins, J.S., Todor, H., and Gross, C.A. 2021. Computational pipeline for designing guide RNAs for mismatch-CRISPRi. STAR Protoc. **2**(2): 100521. doi:10.1016/j.xpro.2021.100521.

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., and Dahl, G.E. 2017. Neural message passing for Quantum chemistry. *In* Proceedings of the 34th International Conference on Machine Learning - Volume 70. JMLR.org, Sydney, NSW, Australia. pp. 1263–1272. [accessed 13 August 2024].

Gislason, A.S., Choy, M., Bloodworth, R.A.M., Qu, W., Stietz, M.S., Li, X., Zhang, C., and Cardona, S.T. 2017. Competitive growth enhances conditional growth mutant sensitivity to antibiotics and exposes a two-component system as an emerging antibacterial target in *Burkholderia cenocepacia*. Antimicrob. Agents Chemother. **61**(1): e00790-16. doi:10.1128/AAC.00790-16.

Goodyear, M.C., Garnier, N., Krieger, J.R., Geddes-McAlister, J., and Khursigara, C.M. 2021. Label-free quantitative proteomics identifies unique proteomes of clinical isolates of the Liverpool Epidemic Strain of *Pseudomonas aeruginosa* and laboratory strain PAO1. PROTEOMICS – Clin. Appl. **15**(6): 2100062. doi:10.1002/prca.202100062.

Goodyear, M.C., Garnier, N.E., Levesque, R.C., and Khursigara, C.M. 2022. Liverpool Epidemic Strain Isolates of Pseudomonas aeruginosa Display High Levels of Antimicrobial Resistance during Both Planktonic and Biofilm Growth. Microbiol. Spectr. **10**(3): e01024-22. doi:10.1128/spectrum.01024-22.

Gould, K. 2016. Antibiotics: from prehistory to the present day. J. Antimicrob. Chemother. **71**(3): 572–575. doi:10.1093/jac/dkv484.

Grace, A., Sahu, R., Owen, D.R., and Dennis, V.A. 2022. Pseudomonas aeruginosa reference strains PAO1 and PA14: A genomic, phenotypic, and therapeutic review. Front. Microbiol. **13**: 1023523. doi:10.3389/fmicb.2022.1023523.

Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., and Tao, D. 2024. A Survey on Self-Supervised Learning: Algorithms, Applications, and Future Trends. IEEE Trans. Pattern Anal. Mach. Intell. **46**(12): 9052–9071. doi:10.1109/TPAMI.2024.3415112.

Halawa, E.M., Fadel, M., Al-Rabia, M.W., Behairy, A., Nouh, N.A., Abdo, M., Olga, R., Fericean, L., Atwa, A.M., El-Nablaway, M., and Abdeen, A. 2024. Antibiotic action and resistance:

updated review of mechanisms, spread, influencing factors, and alternative approaches for combating resistance. Front. Pharmacol. **14**: 1305294. doi:10.3389/fphar.2023.1305294.

Heid, E., Greenman, K.P., Chung, Y., Li, S.-C., Graff, D.E., Vermeire, F.H., Wu, H., Green, W.H., and McGill, C.J. 2024. Chemprop: A Machine Learning Package for Chemical Property Prediction. J. Chem. Inf. Model. **64**(1): 9–17. doi:10.1021/acs.jcim.3c01250.

Hicks, S.A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M.A., Halvorsen, P., and Parasa, S. 2022. On evaluation metrics for medical applications of artificial intelligence. Sci. Rep. **12**(1): 5979. doi:10.1038/s41598-022-09954-8.

Hoang, T.T., Kutchma, A.J., Becher, A., and Schweizer, H.P. 2000. Integration-Proficient Plasmids for Pseudomonas aeruginosa: Site-Specific Integration and Use for Engineering of Reporter and Expression Strains. Plasmid **43**(1): 59–72. doi:10.1006/plas.1999.1441.

Hochreiter, S., Klambauer, G., and Rarey, M. 2018. Machine Learning in Drug Discovery. J. Chem. Inf. Model. **58**(9): 1723–1724. American Chemical Society. doi:10.1021/acs.jcim.8b00478.

Hogan, A.M., and Cardona, S.T. 2022. Gradients in gene essentiality reshape antibacterial research. FEMS Microbiol. Rev.: fuac005. doi:10.1093/femsre/fuac005.

Hogan, A.M., Rahman, A.S.M.Z., Lightly, T.J., and Cardona, S.T. 2019. A broad-host-range CRISPRi toolkit for silencing gene expression in *Burkholderia*. ACS Synth. Biol. **8**(10): 2372–2384. doi:10.1021/acssynbio.9b00232.

Hogan, A.M., Scoffone, V.C., Makarov, V., Gislason, A.S., Tesfu, H., Stietz, M.S., Brassinga, A.K.C., Domaratzki, M., Li, X., Azzalin, A., Biggiogera, M., Riabova, O., Monakhova, N., Chiarelli, L.R., Riccardi, G., Buroni, S., and Cardona, S.T. 2018. Competitive fitness of essential gene knockdowns reveals a broad-spectrum antibacterial inhibitor of the cell

division protein FtsZ. Antimicrob. Agents Chemother. **62**(12): e01231-18. doi:10.1128/AAC.01231-18.

Holloway, B.W., and Zhang, C. 1990. Genetic maps. In: Locus maps of complex organisms. *In* 5th ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**(6096): 816–821. doi:10.1126/science.1225829.

Johnson, E.O., LaVerriere, E., Office, E., Stanley, M., Meyer, E., Kawate, T., Gomez, J.E., Audette, R.E., Bandyopadhyay, N., Betancourt, N., Delano, K., Silva, I.D., Davis, J., Gallo, C., Gardner, M., Golas, A.J., Guinn, K.M., Kennedy, S., Korn, R., McConnell, J.A., Moss, C.E., Murphy, K.C., Nietupski, R.M., Papavinasasundaram, K.G., Pinkham, J.T., Pino, P.A., Proulx, M.K., Ruecker, N., Song, N., Thompson, M., Trujillo, C., Wakabayashi, S., Wallach, J.B., Watson, C., Ioerger, T.R., Lander, E.S., Hubbard, B.K., Serrano-Wu, M.H., Ehrt, S., Fitzgerald, M., Rubin, E.J., Sassetti, C.M., Schnappinger, D., and Hung, D.T. 2019. Large-scale chemical–genetics yields new *M. tuberculosis* inhibitor classes. Nature **571**(7763): 72–78. doi:10.1038/s41586-019-1315-z.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., Zaslavsky, L., Zhang, J., and Bolton, E.E. 2023. PubChem 2023 update. Nucleic Acids Res. **51**(D1): D1373–D1380. doi:10.1093/nar/gkac956.

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

Lagu, B., Senaiar, R.S., Kluge, A.F., Mallesh, B., Ramakrishna, M., Bhat, R., and Patane, M.A. 2020. Addressing hERG activity while maintaining favorable potency, selectivity and pharmacokinetic properties of PPARδ modulators. Bioorg. Med. Chem. Lett. **30**(4): 126928. Elsevier BV. doi:10.1016/j.bmcl.2019.126928.

Landrum, G., Tosco, P., Kelley, B., Rodriguez, R., Cosgrove, D., Vianello, R., and Gedeck, P. 2025, March 3. RDKit. Python. Available from https://www.rdkit.org.

Langendonk, R.F., Neill, D.R., and Fothergill, J.L. 2021. The Building Blocks of Antimicrobial Resistance in Pseudomonas aeruginosa: Implications for Current Resistance-Breaking Therapies. Front. Cell. Infect. Microbiol. **11**: 665759. doi:10.3389/fcimb.2021.665759.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. **8**(11): 2180–2196. doi:10.1038/nprot.2013.132.

Leisman, D.E., Harhay, M.O., Lederer, D.J., Abramson, M., Adjei, A.A., Bakker, J., Ballas, Z.K., Barreiro, E., Bell, S.C., Bellomo, R., Bernstein, J.A., Branson, R.D., Brusasco, V., Chalmers, J.D., Chokroverty, S., Citerio, G., Collop, N.A., Cooke, C.R., Crapo, J.D., Donaldson, G., Fitzgerald, D.A., Grainger, E., Hale, L., Herth, F.J., Kochanek, P.M., Marks, G., Moorman, J.R., Ost, D.E., Schatz, M., Sheikh, A., Smyth, A.R., Stewart, I., Stewart, P.W., Swenson, E.R., Szymusiak, R., Teboul, J.-L., Vincent, J.-L., Wedzicha, J.A., and Maslove, D.M. 2020. Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. Crit. Care Med. **48**(5): 623–633. doi:10.1097/CCM.0000000000004246.

Lewis, K. 2012. Antibiotics: Recover the lost art of drug discovery. Nature **485**(7399): 439–440. doi:10.1038/485439a.

Lewis, K. 2020. The Science of Antibiotic Discovery. Cell **181**(1): 29–45. doi:10.1016/j.cell.2020.02.056.

Li, H., Zhang, R., Min, Y., Ma, D., Zhao, D., and Zeng, J. 2023. A knowledge-guided pre-training framework for improving molecular representation learning. Nat. Commun. **14**(1): 7568. doi:10.1038/s41467-023-43214-1.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., and Stoica, I. 2018, July 13. Tune: A Research Platform for Distributed Model Selection and Training. arXiv. doi:10.48550/arXiv.1807.05118.

Liu, C., Hogan, A.M., Sturm, H., Khan, M.W., Islam, Md.M., Rahman, A.S.M.Z., Davis, R., Cardona, S.T., and Hu, P. 2022. Deep learning-driven prediction of drug mechanism of action from large-scale chemical-genetic interaction profiles. J. Cheminformatics **14**(1): 12. doi:10.1186/s13321-022-00596-6.

Liu, G., Catacutan, D.B., Rathod, K., Swanson, K., Jin, W., Mohammed, J.C., Chiappino-Pepe, A., Syed, S.A., Fragis, M., Rachwalski, K., Magolan, J., Surette, M.G., Coombes, B.K., Jaakkola, T., Barzilay, R., Collins, J.J., and Stokes, J.M. 2023. Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii. Nat. Chem. Biol. **19**(11): 1342–1350. doi:10.1038/s41589-023-01349-8.

Lluka, T., and Stokes, J.M. 2023. Antibiotic discovery in the artificial intelligence era. Ann. N. Y. Acad. Sci. **1519**(1): 74–93. doi:10.1111/nyas.14930.

van der Maaten, L., and Hinton, G. 2008. Visualizing Data using t-SNE. JMLR **9**: 2579–2605.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., van der Oost, J., and Koonin, E.V. 2011.

Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. **9**(6): 467–477. Nature Publishing Group. doi:10.1038/nrmicro2577.

Mathee, K. 2018. Forensic investigation into the origin of Pseudomonas aeruginosa PA14 — old but not lost. J. Med. Microbiol. **67**(8): 1019–1021. doi:10.1099/jmm.0.000778.

Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., Gordillo-Marañón, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C.J., Segura-Cabrera, A., Hersey, A., and Leach, A.R. 2019. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. **47**(D1): D930–D940. doi:10.1093/nar/gky1075.

Michelucci, U. 2024. Fundamental Mathematical Concepts for Machine Learning in Science. Springer International Publishing, Cham. doi:10.1007/978-3-031-56431-4.

Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P.B., Glaser, P., Aigle, B., Bode, H.B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M.H., Pernodet, J.-L., Stadler, M., Tormo, J.R., Genilloud, O., Truman, A.W., Weissman, K.J., Takano, E., Sabatini, S., Stegmann, E., Brötz-Oesterhelt, H., Wohlleben, W., Seemann, M., Empting, M., Hirsch, A.K.H., Loretz, B., Lehr, C.-M., Titz, A., Herrmann, J., Jaeger, T., Alt, S., Hesterkamp, T., Winterhalter, M., Schiefer, A., Pfarr, K., Hoerauf, A., Graz, H., Graz, M., Lindvall, M., Ramurthy, S., Karlén, A., van Dongen, M., Petkovic, H., Keller, A., Peyrane, F., Donadio, S., Fraisse, L., Piddock, L.J.V., Gilbert, I.H., Moser, H.E., and Müller, R. 2021. Towards the sustainable discovery and development of new antibiotics. Nat. Rev. Chem. **5**(10): 726–749. Nature Publishing Group. doi:10.1038/s41570-021-00313-1.

Mittal, R., Aggarwal, S., Sharma, S., Chhibber, S., and Harjai, K. 2009. Urinary tract infections caused by Pseudomonas aeruginosa: A minireview. J. Infect. Public Health **2**(3): 101–111. doi:10.1016/j.jiph.2009.08.003.

Mobegi, F.M., van Hijum, S.A., Burghout, P., Bootsma, H.J., de Vries, S.P., van der Gaast-de Jongh, C.E., Simonetti, E., Langereis, J.D., Hermans, P.W., de Jonge, M.I., and Zomer, A. 2014. From microbial gene essentiality to novel antimicrobial drug targets. BMC Genomics **15**: 958. doi:10.1186/1471-2164-15-958.

Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Aguilar, G.R., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Hamadani, B.H.K., Kumaran, E.A.P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., Luca, M.D., Dokova, K., Dramowski, A., Dunachie, S.J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M.M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A.Y., Perrone,

C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rudd, K.E., Russell, N., Schnall, J., Scott, J.A.G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., Doorn, H.R. van, Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C., and Naghavi, M. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet **399**(10325): 629–655. Elsevier. doi:10.1016/S0140-6736(21)02724-0.

Narykov, O., Zhu, Y., Brettin, T., Evrard, Y.A., Partin, A., Xia, F., Shukla, M., Vasanthakumari, P., Doroshow, J.H., and Stevens, R.L. 2024, March 15. Data Imbalance in Drug Response Prediction - Multi-Objective Optimization Approach in Deep Learning Setting. Cancer Biology. doi:10.1101/2024.03.14.585074.

Nixon, G.M., Armstrong, D.S., Carzino, R., Carlin, J.B., Olinsky, A., Robertson, C.F., and Grimwood, K. 2001. Clinical outcome after early Pseudomonas aeruginosa infection in cystic fibrosis. J. Pediatr. **138**(5): 699–704. doi:10.1067/mpd.2001.112897.

Noirot-Gros, M.-F., Forrester, S., Malato, G., Larsen, P.E., and Noirot, P. 2019. CRISPR interference to interrogate genes that control biofilm formation in Pseudomonas fluorescens. Sci. Rep. **9**(1): 15954. doi:10.1038/s41598-019-52400-5.

Olayo-Alarcon, R., Amstalden, M.K., Zannoni, A., Bajramovic, M., Sharma, C.M., Brochado, A.R., Rezaei, M., and Müller, C.L. 2025. Pre-trained molecular representations enable antimicrobial discovery. Nat. Commun. **16**(1): 3420. doi:10.1038/s41467-025-58804-4.

Parker, C.M., Kutsogiannis, J., Muscedere, J., Cook, D., Dodek, P., Day, A.G., and Heyland, D.K. 2008. Ventilator-associated pneumonia caused by multidrug-resistant organisms or

Pseudomonas aeruginosa: Prevalence, incidence, risk factors, and outcomes. J. Crit. Care **23**(1): 18–26. doi:10.1016/j.jcrc.2008.02.001.

Pathania, R., Zlitni, S., Barker, C., Das, R., Gerritsma, D.A., Lebert, J., Awuah, E., Melacini, G., Capretta, F.A., and Brown, E.D. 2009. Chemical genomics in *Escherichia coli* identifies an inhibitor of bacterial lipoprotein targeting. Nat Chem Biol **5**(11): 849–856.

Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat. Rev. Discov. **6**(1): 29–40. doi:10.1038/nrd2201.

Peters, J.M., Koo, B.-M., Patino, R., Heussler, G.E., Hearne, C.C., Qu, J., Inclan, Y.F., Hawkins, J.S., Lu, C.H.S., Silvis, M.R., Harden, M.M., Osadnik, H., Peters, J.E., Engel, J.N., Dutton, R.J., Grossman, A.D., Gross, C.A., and Rosenberg, O.S. 2019. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. Nat. Microbiol. **4**(2): 244–250. doi:10.1038/s41564-018-0327-z.

Prasad, N.K. 2022. A pooled CRISPRi screen to probe Pseudomonas aeruginosa gene vulnerability during murine lung infection and phage predation. UCSF. Available from https://escholarship.org/uc/item/962907p5 [accessed 22 February 2023].

Puckett, S., Trujillo, C., Eoh, H., Marrero, J., Spencer, J., Jackson, M., Schnappinger, D., Rhee, K., and Ehrt, S. 2014. Inactivation of Fructose-1,6-Bisphosphate Aldolase Prevents Optimal Co-catabolism of Glycolytic and Gluconeogenic Carbon Substrates in Mycobacterium tuberculosis. PLoS Pathog. **10**(5): e1004144. doi:10.1371/journal.ppat.1004144.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell **152**(5): 1173–1183. doi:10.1016/j.cell.2013.02.022.

Qu, J., Prasad, N.K., Yu, M.A., Chen, S., Lyden, A., Herrera, N., Silvis, M.R., Crawford, E., Looney, M.R., Peters, J.M., and Rosenberg, O.S. 2019. Modulating pathogenesis with mobile-CRISPRi. J. Bacteriol. **201**(202): e00304-19. doi:10.1128/JB.00304-19.

Rahman, A.S.M.Z., Liu, C., Sturm, H., Hogan, A.M., Davis, R., Hu, P., and Cardona, S.T. 2022. A machine learning model trained on a high-throughput antibacterial screen increases the hit rate of drug discovery. PLOS Comput. Biol. **18**(10): e1010613. doi:10.1371/journal.pcbi.1010613.

Rahman, A.S.M.Z., Syroegin, E.A., Novomisky Nechcoff, J., Devarajan, A., Polikanov, Y.S., and Cardona, S.T. 2024. Rationally designed pooled CRISPRi-seq uncovers an inhibitor of bacterial peptidyl-tRNA hydrolase. Cell Rep. **43**(11): 114967. doi:10.1016/j.celrep.2024.114967.

Ressner, R.A., Murray, C.K., Griffith, M.E., Resnake, M.S., Hospenthal, D.R., and Wolf, S.E. 2008. Outcomes of Bacteremia in Burn Patients Involved in Combat Operations Overseas. J. Am. Coll. Surg. **206**(3): 439–444. doi:10.1016/j.jamcollsurg.2007.09.017.

Rogers, D., and Hahn, M. 2010. Extended-Connectivity Fingerprints. J. Chem. Inf. Model. **50**(5): 742–754. doi:10.1021/ci100050t.

Selin, C., Brown, E.D., and Cardona, S.T. 2013. High-throughput screening for growth inhibitors of *Burkholderia cenocepacia*. *In* PEDIATRIC PULMONOLOGY. WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA. pp. 331–331.

Speck-Planche, A., and Cordeiro, M.N.D.S. 2014. Simultaneous Virtual Prediction of Anti-*Escherichia coli* Activities and ADMET Profiles: A Chemoinformatic Complementary Approach for High-Throughput Screening. ACS Comb. Sci. **16**(2): 78–84. doi:10.1021/co400115s.

Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C.R., French, S., Carfrae, L.A., Bloom-Ackerman, Z., Tran, V.M., Chiappino-Pepe, A., Badran, A.H., Andrews, I.W., Chory, E.J., Church, G.M., Brown, E.D., Jaakkola, T.S., Barzilay, R., and Collins, J.J. 2020. A deep learning approach to antibiotic discovery. Cell **180**(4): 688-702.e13. doi:10.1016/j.cell.2020.01.021.

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., and Olson, M.V. 2000. PAO1, an opportunistic pathogen. Nature **406**.

Streeter, K., Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia, Katouli, M., and Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, Queensland, Australia. 2016. Pseudomonas aeruginosa: A review of their Pathogenesis and Prevalence in Clinical Settings and the Environment. Infect. Epidemiol. Med. **2**(1): 25–32. doi:10.18869/modares.iem.2.1.25.

Swanson, K., Walther, P., Leitz, J., Mukherjee, S., Wu, J.C., Shivnaraine, R.V., and Zou, J. 2024. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics **40**(7): btae416. doi:10.1093/bioinformatics/btae416.

Tan, S.Z., Reisch, C.R., and Prather, K.L.J. 2018. A robust CRISPRi gene repression system in *Pseudomonas*. J. Bacteriol. **200**(7): e00575-17. doi:10.1128/JB.00575-17.

Tarca, A.L., Carey, V.J., Chen, X., Romero, R., and Drăghici, S. 2007. Machine Learning and Its Applications to Biology. PLoS Comput. Biol. **3**(6): e116. doi:10.1371/journal.pcbi.0030116.

Tiikkainen, P., Bellis, L., Light, Y., and Franke, L. 2013. Estimating Error Rates in Bioactivity Databases. J. Chem. Inf. Model. **53**(10): 2499–2505. doi:10.1021/ci400099q.

Turner, K.H., Wessel, A.K., Palmer, G.C., Murray, J.L., and Whiteley, M. 2015. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc. Natl. Acad. Sci. **112**(13): 4110–4115. Proceedings of the National Academy of Sciences. doi:10.1073/pnas.1419677112.

Wang, J., Guan, J., and Zhou, S. 2023. Molecular property prediction by contrastive learning with attention-guided positive sample selection. Bioinformatics **39**(5): btad258. doi:10.1093/bioinformatics/btad258.

Warrier, T., Romano, K.P., Clatworthy, A.E., and Hung, D.T. 2022. Integrated genomics and chemical biology herald an era of sophisticated antibacterial discovery, from defining essential genes to target elucidation. Cell Chem. Biol. **29**(5): 716–729. doi:10.1016/j.chembiol.2022.04.006.

Weber, B.S., Miyata, S.T., Iwashkiw, J.A., Mortensen, B.L., Skaar, E.P., Pukatzki, S., and Feldman, M.F. 2013. Genomic and Functional Analysis of the Type VI Secretion System in Acinetobacter. PLoS ONE **8**(1): e55142. doi:10.1371/journal.pone.0055142.

Weininger, D. 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. **28**(1): 31–36. doi:10.1021/ci00057a005.

WHO Bacterial Priority Pathogens List 2024: Bacterial Pathogens of Public Health Importance, to Guide Research, Development, and Strategies to Prevent and Control Antimicrobial Resistance. 2024. *In* 1st ed. World Health Organization, Geneva.

Wong, F., de la Fuente-Nunez, C., and Collins, J.J. 2023a. Leveraging artificial intelligence in the fight against infectious diseases. Science **381**(6654): 164–170. doi:10.1126/science.adh1114.

Wong, F., Krishnan, A., Zheng, E.J., Stärk, H., Manson, A.L., Earl, A.M., Jaakkola, T., and Collins, J.J. 2022. Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery. Mol. Syst. Biol. **18**(9): e11081. John Wiley & Sons, Ltd. doi:10.15252/msb.202211081.

Wong, F., Zheng, E.J., Valeri, J.A., Donghia, N.M., Anahtar, M.N., Omori, S., Li, A., Cubillos-Ruiz, A., Krishnan, A., Jin, W., Manson, A.L., Friedrichs, J., Helbig, R., Hajian, B., Fiejtek, D.K., Wagner, F.F., Soutter, H.H., Earl, A.M., Stokes, J.M., Renner, L.D., and Collins, J.J. 2023b. Discovery of a structural class of antibiotics with explainable deep learning. Nature. doi:10.1038/s41586-023-06887-8.

World Health Organization. 2024. WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. Available from https://www.who.int/publications/i/item/9789240093461 [accessed 27 August 2024].

Xiang, L., Qi, F., Jiang, L., Tan, J., Deng, C., Wei, Z., Jin, S., and Huang, G. 2020. CRISPR-dCas9-mediated knockdown of prtR, an essential gene in Pseudomonas aeruginosa. Lett. Appl. Microbiol. **71**(4): 386–393. doi:10.1111/lam.13337.

Xie, A., Zhang, Z., Guan, J., and Zhou, S. 2023. Self-supervised learning with chemistry-aware fragmentation for effective molecular property prediction. Brief. Bioinform. **24**(5): bbad296. doi:10.1093/bib/bbad296.

Xu, Z., and Sigler, P.B. 1998. GroEL/GroES: Structure and Function of a Two-Stroke Folding Machine. J. Struct. Biol. **124**(2–3): 129–141. doi:10.1006/jsbi.1998.4060.

Zhang, Y., Zhang, T., Xiao, X., Kawalek, A., Ou, J., Ren, A., Sun, W., De Bakker, V., Liu, Y., Li, Y., Yang, L., Ye, L., Jia, N., Veening, J.-W., Wang, Y., and Liu, X. 2024, September 1. Genome-wide CRISPRi-seq identified ferredoxin-NADP reductase FprB as a synergistic target for gallium therapy in *Pseudomonas aeruginosa*. doi:10.1101/2024.09.01.610675.

Zhou, D., Huang, G., Xu, G., Xiang, L., Huang, S., Chen, X., Zhang, Y., and Wang, D. 2022a. CRISPRi-Mediated Gene Suppression Reveals Putative Reverse Transcriptase Gene PA0715 to Be a Global Regulator of Pseudomonas aeruginosa. Infect. Drug Resist. **15**: 7577–7599. Dove Medical Press. doi:10.2147/IDR.S384980.

Zhou, H., Yang, Y., Shang, W., Rao, Y., Chen, J., Peng, H., Huang, J., Hu, Z., Zhang, R., and Rao, X. 2022b. Pyocyanin biosynthesis protects *Pseudomonas aeruginosa* from nonthermal plasma inactivation. Microb. Biotechnol. **15**(6): 1910–1921. doi:10.1111/1751-7915.14032.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., and Sun, M. 2020. Graph neural networks: A review of methods and applications. AI Open **1**: 57–81. doi:10.1016/j.aiopen.2021.01.001.